PARTICIPATORY APPROACHES TO A NEW ETHICAL AND LEGAL FRAMEWORK FOR ICT

Leitlinien zu ethischen und rechtlichen Fragen des Datenschutzes in der IKT-Forschung und -Innovation.

KÜNSTLICHE INTELLIGENZ

Dieses Werk ist lizenziert unter einer Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Dieses Projekt wurde aus Mitteln des Forschungs- und Innovationsprogramms Horizont 2020 der Europäischen Union unter der Finanzhilfevereinbarung Nr. 788039 finanziert. Die Verantwortung für den Inhalt dieses Dokuments tragen allein die Verfasser; die Agentur haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

KI: Anforderungen für Entwickler und Innovatoren

Iñigo de Miguel Beriain[385] (UPV/EHU), Felix Schaber[386] (OEAW), Gianclaudio Malgieri und Andres Chomczyk Penedo[387] (VUB)

Danksagung: Die Autoren danken José Antonio Castillo Parrilla, Eduard Fosch Villaronga und Lorena Perez Campillo für ihre freundliche Unterstützung beim Redigieren dieses Dokuments. Selbstverständlich tragen wir für alle Fehler die volle Verantwortung.

Dieser Teil des Leitfadens wurde von Marko Sijan, Senior Advisor Specialist, (HR DPA) überprüft und bestätigt.

EINLEITUNG: TEIL A

Der erste Teil dieses Kapitels befasst sich mit den sieben ethischen Anforderungen, die in den von der hochrangigen Expertengruppe für künstliche Intelligenz[388] in ihren „Ethik-Leitlinien für eine vertrauenswürdige KI“ veröffentlichten Empfehlungen enthalten sind.[389] Diese Empfehlungen wurden vor kurzem im Rahmen des SHERPA-Projekts[390] analysiert, das auch eine umfassende Analyse der ethischen Fragen bei der Entwicklung geeigneter Instrumente zur Bewältigung dieser Herausforderungen umfasst. Angesichts dieser lobenswerten Arbeit

---

[385] Autor des gesamten Dokuments mit Ausnahme der Abschnitte 2 und 7 dieses Teils.

[386] Autor von Abschnitt 2 dieses Teils.

[387] Autor von Abschnitt 7 dieses Teils.

[388] Die hochrangige Expertengruppe für künstliche Intelligenz wurde von der Europäischen Kommission 2018 eingesetzt. Ihr allgemeines Ziel ist die Unterstützung der Umsetzung der europäischen Strategie für künstliche Intelligenz. Dazu gehört auch die Ausarbeitung von Empfehlungen für die künftige politische Entwicklung und für ethische, rechtliche und gesellschaftliche Fragen im Zusammenhang mit der künstlichen Intelligenz, einschließlich sozioökonomischer Herausforderungen. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence (abgerufen am 15. Mai 2020).

[389] Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 15 ff. Brüssel, Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

[390] SHERPA (2019) Guidelines for the ethical use of AI and big data systems. SHERPA, www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf (abgerufen am 5. Mai 2020). – Hochrangige Expertengruppe für künstliche Intelligenz, 390

wäre es überflüssig, hier eine ausführliche Analyse desselben Themas (KI) aus derselben Perspektive (Ethik) vorzunehmen. Stattdessen haben wir in diesem Dokument versucht, eine ergänzende Analyse bereitzustellen. Diese Leitlinien sollen die Überschneidungen zwischen den ethischen Empfehlungen der hochrangigen Expertengruppe für künstliche Intelligenz und dem durch die Datenschutz-Grundverordnung (DSGVO) geschaffenen Rechtsrahmen für Datenschutzfragen zu ermitteln.

Vor dem Start der Analyse sind jedoch einige Vorbemerkungen erforderlich. Erstens konzentriert sich dieser Bericht hauptsächlich auf KI-Entwickler: Unternehmen, die bereit sind, KI-Tools zu entwickeln. Diese Unternehmen werden zu Verantwortlichen, sobald sie mit der Verarbeitung personenbezogener Daten beginnen. In diesem Sinne sollten die Begriffe „Tool“, „Lösung“, „Modell“ und „Entwicklung“ im Rahmen dieser Analyse als synonym betrachtet werden.

Zweitens kann dieser Teil der Leitlinien nur im Zusammenhang mit dem gesamten Instrument (den Leitlinien) verstanden werden. Es gibt mehrere Konzepte, die in diesem Dokument nicht untersucht werden, weil sie in anderen Abschnitten der Leitlinien behandelt werden. Wir haben gegebenenfalls auf diese verwiesen (die Verweise sind gelb hervorgehoben). In Zukunft werden alle Abschnitte auf einer Website zur Verfügung stehen, wodurch die Leitlinien wesentlich benutzerfreundlicher sein werden.

Die einzelnen Abschnitte dieses Dokuments enthalten nur das, was wir für unbedingt notwendig erachten, um die Kernargumente der betreffenden ethischen und rechtlichen Fragen zu verstehen. Wir haben Checklisten eingefügt, damit die Verantwortlichen leichter erkennen können, ob sie die Themen richtig angehen, und einen Abschnitt mit weiterführender Literatur, die Leser bei Bedarf konsultieren können. Fußnoten geben weitere Hinweise auf die wichtigsten Aussagen.

Schließlich wurde dieses Dokument so strukturiert, dass es leicht verständlich ist. Wie bereits erwähnt, basiert es auf den sieben, von der hochrangigen Expertengruppe für künstliche Intelligenz beschriebenen Anforderungen. Wir beginnen unsere Analyse mit einer kurzen Beschreibung der zentralen ethischen Fragen und fassen dann die wichtigsten ethischen Probleme und Herausforderungen im Zusammenhang mit ihnen zusammen. Diese dienen als gemeinsame Grundlage und Kontext, auf der bzw. dem unsere rechtliche Analyse aufbaut.

1 Vorrang menschlichen Handelns und menschlicher Aufsicht

„KI-Systeme sollten die menschliche Autonomie und Entscheidungsfindung, wie es der Grundsatz der Achtung der menschlichen Autonomie vorsieht, unterstützen. Dies erfordert, dass KI-Systeme sowohl einer demokratischen, florierenden und gerechten Gesellschaft dienen, indem sie das menschliche Handeln und die Wahrung der Grundrechte fördern, als auch die menschliche Aufsicht ermöglichen.“

1.1 Ethische Grundsätze

Diese erste Anforderung für die Entwicklung von KI umfasst drei unterschiedliche Hauptgrundsätze:[391]

[391] Ibid., Seite 15.

- **Grundrechte.** KI-Systeme können für Grundrechte entweder förderlich oder hinderlich sein. In solchen Situationen sollte eine Folgenabschätzung der Auswirkungen auf die Grundrechte vor der Entwicklung einer KI-Lösung vorgenommen werden.
- **Vorrang menschlichen Handelns.** Die Benutzer von KI-Systemen sollten in der Lage sein, informierte, autonome Entscheidungen zu treffen. KI-Systeme sollten die einzelnen Menschen dabei unterstützen, im Einklang mit ihren eigenen Zielen bessere, fundiertere Entscheidungen zu treffen. Die Richtschnur für das Funktionieren von KI-Systemen muss der Grundsatz der Selbstbestimmung des Nutzers sein. Die betroffenen Personen müssen beispielsweise gegebenenfalls wissen, dass ihre Daten zum Profiling verwendet werden könnten. Zudem muss das Recht einer Person gewahrt werden, nicht einer ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt. Es ist jedoch zu beachten, dass sich dies im Allgemeinen auf kommerzielle Zwecke bezieht. Dies gilt also nicht für Strafverfolgungsbehörden, die personenbezogene Daten auf gesetzlicher Grundlage verarbeiten und KI zur wirksamen Bekämpfung verschiedener Straftaten und zur Erfüllung der gesetzlich vorgeschriebenen Pflichten einsetzen können.
- **Vorrang menschlicher Aufsicht.** Die menschliche Aufsicht hilft, dafür zu sorgen, dass ein KI-System die menschliche Autonomie nicht untergräbt oder sich sonst nachteilig auswirkt. Die Aufsicht kann durch verschiedene Lenkungs- und Kontrollmechanismen erreicht werden. Für ein ansonsten gleiches System muss gelten: Je weniger Aufsicht ein Mensch über ein KI-System ausüben kann, desto ausführlicher muss es zuvor getestet werden und desto strenger muss die Lenkung und Kontrolle sein.

1.2

Das Erfordernis menschlichen Handelns und einer menschlichen Aufsicht bei der Entwicklung von KI-Tools steht in klarem Zusammenhang mit dem Recht auf menschliches Eingreifen, dem Recht, keiner ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung unterworfen zu werden, dem Recht auf Auskunft über die automatische Entscheidungsfindung und der damit verbundenen Logik, die alle in der Datenschutz-Grundverordnung verankert sind. Diese Rechte werden durch den Einsatz von KI-Tools einem Risiko ausgesetzt. KI beinhaltet oft eine Form der automatisierten Verarbeitung, und in einigen Fällen werden Entscheidungen direkt von dem KI-Modell getroffen. In der Tat lernen und entscheiden KI-Tools manchmal ohne menschliche Aufsicht, und gelegentlich ist die mit ihrer Leistung verbundene Logik schwer zu verstehen.[392]

In dieser Hinsicht **ist das Profiling im Rahmen der KI-Entwicklung besonders problematisch** (siehe Kasten 1), denn vom Profiling „bekommen die betroffenen Personen selbst oft gar nichts mit. Denn dabei werden aus Rückschlüssen erzeugte oder hergeleitete Daten zu einzelnen Personen erzeugt, d. h. „neue“ personenbezogene Daten, die von den Betroffenen selbst gar nicht direkt zur Verfügung gestellt wurden. Die Menschen haben in dieser Thematik einen unterschiedlichen Kenntnisstand und für manche sind die

[392] Burrell, J. (2016), How the machine ‘thinks’: understanding opacity in machine learning algorithms, *Big Data & Society* 3(1): 1–12.

komplizierten Techniken, die bei Profiling und automatisierten Entscheidungen zur Anwendung kommen, möglicherweise schwer verständlich.“[393] (siehe Abschnitt „

Verständnis von Transparenz und Undurchsichtigkeit“).

Die DSGVO verhindert natürlich **nicht jede Form von Profiling und/oder einer automatisierten Entscheidungsfindung**: Sie gewährt dem Einzelnen lediglich ein qualifiziertes Recht, darüber informiert zu werden, und ein Recht darauf, keiner ausschließlich auf einer automatisierten Verarbeitung einschließlich Profiling beruhenden Entscheidung unterworfen zu werden. Ihr Recht auf Auskunft muss in Anwendung des Grundsatzes der Rechtmäßigkeit, der Verarbeitung nach Treu und Glauben und der Transparenz gewahrt werden. Somit muss der Verantwortliche zumindest „der betroffenen Person mitteilen, dass er dies tut, aussagekräftige Informationen über die involvierte Logik liefern und die Tragweite und die angestrebten Auswirkungen der Verarbeitung erläutern“[394] (siehe Artikel 13 und 14 DSGVO).

Die **Informationen über die Logik eines Systems** und die Erläuterungen zu den Entscheidungen sollten der betroffenen Person den notwendigen Kontext für die Entscheidung übr die Verarbeitung sie betreffender personenbezogener Daten geben. In einigen Fällen können unzureichende Erläuterungen dazu führen, dass die betroffene Person unnötigerweise auf andere Rechte zurückgreift. Anträge auf Eingreifen, Darlegung des eigenen Standpunkts oder Einwände gegen die Verarbeitung sind wahrscheinlicher, wenn die betroffene Person das Gefühl hat, nicht ausreichend zu verstehen, wie die Entscheidung zustande gekommen ist. Betroffene Personen müssen in der Lage sein, **ihre Rechte auf einfache und benutzerfreundliche Weise wahrzunehmen.** Wenn beispielsweise „das Ergebnis einer ausschließlich automatisierten Entscheidung über eine Website mitgeteilt wird, sollte die Seite einen Link oder klare Informationen enthalten, anhand der sich die Person ohne unangemessene Verzögerungen oder Komplikationen an einen Mitarbeiter wenden kann, der dann eingreift.“[395] Der volle Umfang der bereitzustellenden Informationen lässt sich jedoch schwer konkret angeben. In der Tat gibt es derzeit eine lebhafte akademische Diskussion über dieses Thema.[396]

---

[393] Artikel-29-Datenschutzgruppe (2017), Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679, WP 251, Seite 9. Europäische Kommission, Brüssel.

[394] Ibid., Seiten13–14.

[395] ICO (2020) AI auditing framework: draft guidance for consultation, Seite 94. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[396] S. Wachter, B. Mittelstadt und L. Floridi, (2017), Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation, *International Data Privacy Law.* Verfügbar unter: https://ssrn.com/abstract=2903469 or http://dx.doi.org/10.2139/ssrn.2903469 (abgerufen am 15. Mai 2020); A.D. Selbst und J. Powles (2017), Meaningful information and the right to explanation, *International Data Privacy Law* 7(4): 233–242, https://doi.org/10.1093/idpl/ipx022 (abgerufen am 15. Mai 2020).

Kasten 0. Die Frage des Rankings

Anbieter von Dienstleistungen oder Waren, die an der sogenannten „kollaborativen Wirtschaft“ (oder „Plattformwirtschaft“) teilnehmen, müssen die Funktionsweise des Rankings im Rahmen der von ihnen vorgenommenen Nutzung spezifischer Online-Vermittlungsdienste oder Online-Suchmaschinen verstehen. Dies könnte zum Beispiel ein Hotel sein – ob groß oder klein – das seine Unterkünfte über Booking.com oder TripAdvisor anbietet. Damit Unternehmen als Anbieter auf der Plattform teilnehmen können, sollten Plattformen nicht verpflichtet sein, die detaillierte Funktionsweise ihrer Rankingmethoden – einschließlich der Algorithmen – offenzulegen. Eine allgemeine Erläuterung der für das Ranking verwendeten Hauptparameter genügt (einschließlich der Möglichkeit, das Ranking gegen eine vom Anbieter gezahlte direkte oder indirekte Vergütung zu beeinflussen), solange diese Beschreibung leicht und öffentlich zugänglich und in klarer und verständlicher Sprache verfasst ist.[397]

Darüber hinaus haben betroffene Personen gemäß Artikel 22 Absatz 1 das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt. Daher sollten Verantwortliche stets darauf achten, dass die von ihnen verwendeten oder entwickelten KI-Tools **in keiner Weise eine unvermeidbare automatische Entscheidungsfindung fördern.** Tatsächlich führt die Artikel-29-Datenschutzgruppe an: „Schwebt dem Verantwortlichen somit ein „Modell“ mit ausschließlich automatisierten Entscheidungen vor, die erhebliche Auswirkungen auf die jeweils betroffenen Personen haben und sich auf zu diesen erstellte Profile stützen, und kann er sich nicht auf die Einwilligung der betroffenen Person, einen Vertrag mit ihr oder ein Gesetz berufen, wonach dies zulässig ist, sollte er von diesem Modell Abstand nehmen. Er kann dennoch ein „Modell“ mit auf Profiling beruhenden Entscheidungen ins Auge fassen, wenn er den Umfang des Eingreifens einer Person so deutlich erhöht, dass das Modell nicht mehr auf einer vollständig automatisierten Entscheidungsfindung beruht, auch wenn die Verarbeitung noch Risiken für die Grundrechte und Grundfreiheiten der betroffenen Person bergen könnte.“[398]

Kasten 1. Profiling verstehen

Untersuchungen von Kosinski et al. (2013)[399] haben gezeigt, dass im Jahr 2011 zugängliche digitale Verhaltensaufzeichnungen (z. B. „gelikte“ Seiten auf Facebook) zur genauen Vorhersage einer Reihe hochsensibler personenbezogener Merkmale genutzt werden konnten. Dazu gehörten: sexuelle Orientierung, ethnische Zugehörigkeit, religiöse

[397] Verordnung (EU) 1159/2019 vom 20. Juni zur Förderung von Fairness und Transparenz für gewerbliche Nutzer von Online- Vermittlungsdiensten, Artikel 5 and Erwägungsgrund 27. Verfügbar unter: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1150&from=EN

[398] Artikel-29-Datenschutzgruppe (2018), Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679. Europäische Kommission, Brüssel, S. 30. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053.

[399] M. Kosinski, D. Stillwell und T. Graepel (2013), Digital records of behavior expose personal traits, *Proceedings of the National Academy of Sciences* 110(15): 5802–5805, DOI:10.1073/pnas.1218772110

und politische Ansichten, Persönlichkeitsmerkmale, Intelligenz, Glück, Konsum von Suchtmitteln, Trennung der Eltern, Alter und Geschlecht. Die Analyse basierte auf einem Datensatz von über 58.000 Freiwilligen und deren „Likes“ bei Facebook, detaillierten demografischen Profilen und den Ergebnissen mehrerer psychometrischer Tests.

Das Modell unterschied in 88 % der Fälle korrekt zwischen homosexuellen und heterosexuellen Männern, in 95 % der Fälle zwischen Afroamerikanern und kaukasischen Amerikanern und in 85 % der Fälle zwischen Wählern der Demokraten und Republikaner. In Bezug auf das Persönlichkeitsmerkmal „Offenheit“ lag die Vorhersagegenauigkeit nahe an der Test-Retest-Genauigkeit eines Standard-Persönlichkeitstests. Die Autoren lieferten zudem Beispiele für Assoziationen zwischen Attributen und Likes und erörtern die Auswirkungen auf die Online-Personalisierung und den Datenschutz.

Dieser Fall ist ein hervorragendes Beispiel dafür, wie Profiling funktioniert: Die Informationen betroffener Personen dienten dazu, sie zu klassifizieren und Vorhersagen über sie zu treffen.

Darüber hinaus muss ein Verantwortlicher immer beachten, dass gemäß Artikel 9 Absatz 2 Buchstabe a DSGVO automatisierte Entscheidungen, die die Verarbeitung besonderer Kategorien personenbezogener Daten beinhalten, nur dann zulässig sind, wenn die betroffene Person ausdrücklich in die Verarbeitung dieser personenbezogenen Daten für einen oder mehrere festgelegte Zwecke eingewilligt hat (siehe Abschnitt „Einwilligung“ im Kapitel „Konzepte“), oder wenn es eine Rechtsgrundlage für die genannte Verarbeitung gibt. Diese Ausnahme gilt nicht nur, wenn die entsprechenden Daten in diese Kategorie fallen, sondern auch dann, wenn der Abgleich verschiedener Arten personenbezogener Daten sensible Informationen über Personen offenbaren kann oder wenn hergeleitete Daten dieser Kategorie entsprechen. Eine Studie, **die besondere Datenkategorien herleiten kann, unterliegt nämlich denselben rechtlichen Verpflichtungen gemäß der DSGVO, als ob von Anfang an sensible personenbezogene Daten verarbeitet worden wären.** In all diesen Fällen müssen die für die Verarbeitung besonderer Kategorien personenbezogener Daten geltenden Vorschriften und die notwendige Anwendung geeigneter Garantien zum Schutz der Rechte, Interessen und Freiheiten der betroffenen Personen berücksichtigt werden. Die Verhältnismäßigkeit zwischen dem Ziel der Forschung und der Verwendung besonderer Kategorien von Daten muss gewährleistet sein. Die Verantwortlichen sollten sich auch darüber im Klaren sein, dass ihre Mitgliedstaaten zusätzliche Bedingungen und Beschränkungen für die Verarbeitung von genetischen Daten, biometrischen Daten oder Gesundheitsdaten einführen bzw. anwenden können (Artikel 9 Absatz 4 DSGVO).

Wenn das Profiling auf personenbezogene Daten zurückgreift, die nicht von der betroffenen Person zur Verfügung gestellt wurden, müssen die Verantwortlichen sicherstellen, dass die Verarbeitung nicht mit dem ursprünglichen Zweck unvereinbar ist (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung" im Kapitel „Konzepte"), dass sie eine Rechtsgrundlage für die Verarbeitung der besonderen Datenkategorie ermittelt haben und dass sie die betroffene Person über die Verarbeitung unterrichten[400] (siehe Abschnitt „Zweckbindung" im Kapitel „Grundsätze").

[400] Artikel-29-Datenschutzgruppe (2017), Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679, WP 251, Seite 15. Europäische Kommission, Brüssel.

Die Durchführung einer „**Datenschutz-Folgenabschätzung“ (DSFA)** ist **zwingend erforderlich, wenn ein reales Risiko einer unzulässigen Profilerstellung oder automatisierten Entscheidungsfindung besteht.** Gemäß Artikel 35 Absatz 3 Buchstabe a DSGVO muss der Verantwortliche im Falle einer systematischen und umfassenden Bewertung persönlicher Aspekte natürlicher Personen eine Datenschutz-Folgenabschätzung durchführen. Dies ist für Tools erforderlich, die sich auf eine automatisierte Verarbeitung einschließlich Profiling gründen, sowie für solche, die als Grundlage für Entscheidungen dienen, die Rechtswirkung gegenüber natürlichen Personen entfalten oder diese in ähnlich erheblicher Weise beeinträchtigen.

Gemäß Artikel 37 Absatz 1 Buchstabe b DSGVO ist eine zusätzliche Rechenschaftspflicht die **Benennung eines Datenschutzbeauftragten (DSB)**, wenn die Kerntätigkeit des Verantwortlichen im Profiling oder der automatisierten Entscheidungsfindung besteht, welche aufgrund ihres Umfangs eine umfangreiche regelmäßige und systematische Überwachung von betroffenen Personen erforderlich macht. Die Verantwortlichen müssen außerdem **im Rahmen ihrer Rechenschafts- und Dokumentationspflicht alle Entscheidungen, die von einem KI-System getroffen werden**, aufzeichnen (siehe Abschnitt „Rechenschaftspflicht“ im Kapitel „Grundsätze“). Darin sollte festgehalten werden, ob eine Person um das Eingreifen einern Person gebeten, einen eigenen Standpunkt dargelegt, die Entscheidung angefochten hat und ob eine Entscheidung daraufhin geändert worden ist.[401]

Die folgenden zusätzlichen Maßnahmen können ebenfalls äußerst nützlich sein, um eine automatisierte Entscheidungsfindung zu vermeiden.[402]

- Berücksichtigen Sie bereits in der Designphase die Systemanforderungen, die zur Unterstützung einer aussagegkräftigen Überprüfung durch Personen erforderlich sind.
- Berücksichtigen Sie insbesondere die Anforderungen an die Interpretierbarkeit und die effektive Gestaltung der Benutzeroberfläche zur Unterstützung von Überprüfungen und Eingriffen durch Personen.
- Gestalten und setzen Sie eine geeignete Schulung und Unterstützung für menschliche Prüfer um.
- Geben Sie dem Personal die entsprechenden Befugnisse, Anreize und Unterstützung, damit es die Bedenken einzelner Personen ansprechen oder eskalieren und gegebenenfalls die Entscheidung des KI-Systems außer Kraft setzen kann.

In jedem Fall sollten sich die Verantwortlichen darüber im Klaren sein, dass die Mitgliedstaaten einige konkrete Hinweise zu diesem Thema in ihre nationalen Vorschriften aufnehmen **und dabei verschiedene Instrumente einsetzen, um eine angemessene Einhaltung zu gewährleisten.**[403]

---

[401] ICO (2020) AI auditing framework: draft guidance for consultation, Seite 94–95. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[402] Ibid., Seite 95.

[403] G. Malgieri (2018), Automated decision-making in the EU Member States laws: the right to explanation and other 'suitable safeguards' for algorithmic decisions in the EU national legislations. Verfügbar unter: https://ssrn.com/abstract=3233611 or http://dx.doi.org/10.2139/ssrn.3233611 (abgerufen am 2. Mai 2020).

Checkliste: Profiling und automatisierte Entscheidungsfindung[404]

□ Die Verantwortlichen verfügen über eine Rechtsgrundlage für die Durchführung von Profiling und/oder automatisierter Entscheidungsfindung und dokumentieren dies in ihrer Datenschutzerklärung.

□ Die Verantwortlichen senden den Personen einen Link zu ihrer Datenschutzerklärung, wenn sie sie betreffende personenbezogene Daten indirekt erhalten haben.

□ Die Verantwortlichen erläutern, wie Personen auf Details der Informationen zugreifen können, die sie zur Erstellung ihres Profils verwendet haben.

□ Die Verantwortlichen informieren Personen, die ihnen sie betreffende personenbezogene Daten zur Verfügung stellen, darüber, wie sie dem Profiling widersprechen können.

□ Die Verantwortlichen verfügen über Verfahren, anhand der Kunden auf die in ihre Profile eingegebenen personenbezogenen Daten zuzugreifen und diese im Hinblick auf etwaige Fehler überprüfen und bearbeiten können.

□ Die Verantwortlichen haben zusätzliche Kontrollen für ihre Systeme zur Profilerstellung/automatisierten Entscheidungsfindung eingeführt, um schutzbedürftige Gruppen (einschließlich Kinder) zu schützen.

□ Die Verantwortlichen erheben nur die erforderliche Mindestmenge an Daten und verfügen über eine klare Speicherungsrichtlinie für die von ihnen erstellten Profile.

Ein bewährtes Modell

□ Die Verantwortlichen führen eine DSFA durch, um die Risiken bei Einführung einer neuen automatisierten Entscheidungsfindung oder eines neuen Profilings zu berücksichtigen und zu bewältigen.

□ Die Verantwortlichen informieren ihre Kunden über das durchgeführte Profiling und die vorgenommene automatisierte Entscheidungsfindung, welche Informationen sie zur Erstellung der Profile verwenden und woher sie diese Informationen erhalten.

□ Die Verantwortlichen verwenden bei ihren Profiling-Aktivitäten anonymisierte Daten.

□ Die Verantwortlichen garantieren das Recht auf Verständlichkeit algorithmischer Entscheidungen.

□ Die Entscheidungsträger verfügen über einen Mechanismus, mit dem sie die Gründe mitteilen und erläutern können, wenn eine Anfechtung der algorithmischen Entscheidung aufgrund fehlender menschlicher Intervention nicht akzeptiert wird.

□ Die Entscheidungsträger verfügen über ein Modell zur Bewertung der Menschenrechte bei automatisierten Entscheidungen.

□ Eine qualifizierte menschliche Aufsicht ist von der Designphase an vorhanden, insbesondere in Bezug auf die Auslegungsanforderungen und die effektive Gestaltung der Schnittstelle, und die Prüfer sind geschult.

□ Es werden Audits im Hinblick auf mögliche Abweichungen von den Ergebnissen der

[404] ICO (kein Datm) Rights related to automated decision making including profiling. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/ (abgerufen am 15. Mai 2020).

Herleitungen in adaptiven oder evolutionären Systemen durchgeführt.

 Das KI-System wird bzw. wurde zertifiziert.

Zusätzliche Informationen

Artikel-29-Datenschutzgruppe (2018), Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

ICO (2020) AI auditing framework: draft guidance for consultation, Seite 94–95. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

ICO (2019) Data Protection Impact Assessments and AI. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-data-protection-impact-assessments-and-ai/

M. Kosinski, D. Stillwell und T. Graepel (2013), Digital records of behavior expose personal traits, *Proceedings of the National Academy of Sciences 110(15): 5802–5805, DOI:10.1073/pnas.1218772110*

G. Malgieri (2018), Automated decision-making in the EU Member States laws: the right to explanation and other 'suitable safeguards' for algorithmic decisions in the EU national legislations. Verfügbar unter: https://ssrn.com/abstract=3233611 oder http://dx.doi.org/10.2139/ssrn.3233611

Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

A.D. Selbst und J. Powles, (2017), Meaningful information and the right to explanation, International Data Privacy Law 7(4): 233–242, https://doi.org/10.1093/idpl/ipx022

S. Wachter, B. Mittelstadt und L. Floridi, (2017), Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation, International Data Privacy Law. Verfügbar unter: https://ssrn.com/abstract=2903469 or http://dx.doi.org/10.2139/ssrn.2903469

2 Technische Robustheit und Sicherheit

„Eine entscheidende Komponente bei der Verwirklichung einer vertrauenswürdigen KI ist die technische Robustheit, die mit dem Grundsatz der Schadensverhütung eng verbunden ist. Die technische Robustheit macht es erforderlich, dass KI-Systeme mit einem präventiven Herangehen an Risiken und so entwickelt werden, dass sie sich zuverlässig gemäß ihrer Bestimmung verhalten, bei gleichzeitiger Minimierung von unbeabsichtigtem und unerwartetem Schaden und unter Verhinderung von inakzeptablem Schaden. Dies sollte auch für potenziell veränderte Betriebsumgebungen oder die Anwesenheit von anderen (menschlichen oder künstlichen) wirkenden Kräften gelten, die gegebenenfalls in feindlicher

Art und Weise mit dem System interagieren. Darüber hinaus sollte die körperliche und geistige Unversehrtheit von Menschen gewährleistet sein."

– *Hochrangige Expertengruppe für künstliche Intelligenz*[405]

2.1 Ethische Grundsätze und Bestimmungen der DSGVO

Die hochrangige Expertengruppe für künstliche Intelligenz unterteilt die Anforderungen an die technische Robustheit und Sicherheit in vier Teilkomponenten: (1) Widerstandsfähigkeit gegen Angriffe und Sicherheitsverletzungen; (2) ein Auffangplan und allgemeine Sicherheit; (3) Präzision; sowie (4) Zuverlässigkeit und Reproduzierbarkeit.

Zur leichteren Bezugnahme spiegelt dieser Abschnitt diese Struktur wider, wobei diese Unterkomponenten mit den rechtlichen (DSGVO-) Anforderungen und Empfehlungen verbunden werden. Dies ist wichtig, denn während die DSGVO-Anforderungen im Allgemeinen nur bei Verarbeitung personenbezogener Daten gelten, sind viele praktische KI-Systeme darauf ausgelegt, ein personalisiertes Ergebnis zu erzielen (z. B. Empfehlungssysteme). Sie müssen daher irgendwann personenbezogene Daten verarbeiten.

2.1.1 Widerstandsfähigkeit gegen Angriffe und Sicherheitsverletzungen

Die Widerstandsfähigkeit gegen Angriffe sollte ein Ziel aller IKT-Systeme sein, darunter auch von KI-Systemen. Bei der Verarbeitung personenbezogener Daten verlangt Artikel 32 DSGVO ausdrücklich die Umsetzung geeigneter technischer und organisatorischer Maßnahmen zur Gewährleistung der Datensicherheit (siehe Unterabschnitt „Maßnahmen zur Unterstützung der Vertraulichkeit", Abschnitt „Integrität und Vertraulichkeit", Kapitel „Grundsätze").

Die erforderlichen Sicherheitsmaßnahmen hängen von den wahrscheinlichen Auswirkungen einer Fehlfunktion des KI-Systems ab. Diese Maßnahmen sollten auch Schritte umfassen, die die Belastbarkeit der Verarbeitungssysteme gewährleisten.[406] Bei bestimmten Arten von KI-Systemen kann der Entscheidungsprozess besonders anfällig gegenüber Angriffen sein. So könnte ein böswilliger Akteur beispielsweise eine irreführende Eingabe erstellen, um die grundlegenden Wahrnehmungsunterschiede zwischen Menschen und KI-Systemen auszunutzen, wie das Beispiel in Kasten 2 zeigt.

Kasten 2. Beispiel für den Bedarf an Sicherheit in KI-Systemen

Ein autonomes Fahrzeug soll mithilfe von Bordkameras automatisch Straßenschilder erkennen und seine Geschwindigkeit entsprechend anpassen. KI-Algorithmen, die auf tiefen neuronalen Netzen basieren, können diese Aufgabe zwar hervorragend bewältigen, doch muss insbesondere darauf geachtet werden, dass das System vor gezielten Angriffen eines böswilligen Gegners geschützt wird. So könnten beispielsweise kleine, gezielte Änderungen an Straßenschildern dazu führen, dass das KI-System ein Stoppschild mit dem einer Geschwindigkeitsbegrenzung verwechselt, was zu potenziell gefährlichen Situationen führen kann. Einem zufälligen menschlichen Beobachter kann die Änderung dagegen als

[405] Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 16 ff. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 28. Mai 2020).

[406] Artikel 32 Absatz 1 Buchstabe b DSGVO.

einfaches Graffiti erscheinen. Es ist daher extrem wichtig, ein zu diesem Zweck eingesetztes KI-System gegen solche Angriffe zu schützen und damit seine Belastbarkeit zu erhöhen.[407]

Trainierte KI-Modelle können zudem eine wertvolle Datenquelle sein. Unter bestimmten Umständen kann es möglich sein, Einblicke in die ursprünglichen Eingabedaten zu gewinnen, indem nur das trainierte Modell verwendet wird.[408] Ein solches „Informationsleck" könnte sowohl von internen als auch von externen Akteuren ausgenutzt werden. Die Verantwortlichen müssen daher unbedingt **Maßnahmen ergreifen, um den Zugang zum Modell und den zugrunde liegenden Trainingsdaten zu begrenzen, und zwar für alle Kategorien von Akteuren** (siehe Unterabschnitt „Maßnahmen zur Unterstützung der Vertraulichkeit“, Abschnitt „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“).

Einmal trainiert, kann das resultierende KI-System für ganz andere Zwecke als die ursprünglich vorgesehen verwendet werden. Ein KI-System zur Gesichtserkennung kann beispielsweise dazu verwendet werden, Fotos mit einer bestimmten Person in einem Online-Fotoalbum zu erkennen und zu gruppieren. Dasselbe KI-System könnte auch dazu dienen, das Internet nach Fotos einer bestimmten Person zu durchsuchen und dabei möglicherweise sensible personenbezogene Daten preiszugeben (z. B. anhand des Aufnahmeorts oder des Aufnahmekontexts). Diese Art von Mehrzwecknutzung ist bei KI-Systemen häufig möglich. Es ist somit Aufgabe des Systementwicklers, eine *mögliche rechtswidrige Verarbeitung personenbezogener Daten vorherzusehen und Sicherheitsmaßnahmen zu ergreifen, die dies verhindern oder minimieren*. Dies könnte durch Maßnahmen wie die Beschränkung der nutzbaren Datenquellen oder das Verbot bestimmter Nutzungsmuster durch Lizenzbedingungen geschehen. Der Datenschutz-Rechtsrahmen kann solche Beschränkungen ergänzen, jedoch keineswegs ersetzen.

2.1.2 Auffangplan und allgemeine Sicherheit

„Irren ist menschlich, aber um wirklich etwas zu vermasseln, braucht man einen Computer.“

– Paul Ehrlich / Bill Vaughan[409]

Wie alle IKT-Systeme können auch KI-Systeme versagen und falsche Ergebnisse oder Vorhersagen liefern. Im Falle von KI-Systemen kann es jedoch besonders schwierig sein zu erklären, warum eine bestimmte (falsche) Schlussfolgerung auf greifbare, menschliche Art und Weise erreicht wurde. Ein Beispiel für ein unerwünschtes Verhalten wäre ein KI-System, das Entscheidungen trifft, die eine Person erheblich beeinträchtigen (z. B. die automatische Ablehnung eines Kreditantrags). Die DSGVO verlangt von den Verantwortlichen die Umsetzung geeigneter Auffangpläne, um die betroffenen Personen vor solchen Situationen zu schützen, einschließlich des Rechts, eine KI-Entscheidung anzufechten und ein menschliches

---

[407] Eykholt, K. et al. (2018), Robust physical-world attacks on deep learning models, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 10.4.2018, arXiv:1707.08945.

[408] Fredrikson, M. et al. (2015), Model inversion attacks that exploit confidence information and basic countermeasure’, CCS '15: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Oktober 2015. Cornell University, Ithaca. Verfügbar unter: https://rist.tech.cornell.edu/papers/mi-ccs.pdf (abgerufen am 20. Mai 2020).

[409] Die Urheberschaft des Zitats scheint umstritten zu sein, siehe z. B. https://quoteinvestigator.com/2010/12/07/foul-computer/#note-1699-18 (abgerufen am 2. Juni 2020).

Eingreifen zu erwirken, das die Sichtweise der betroffenen Personen berücksichtigt.[410] Solche Garantien sollten bei der Systemgestaltung berücksichtigt werden. Selbst in Fällen, in denen die DSGVO einen solchen Auffangplan nicht ausdrücklich vorschreibt, sollten die Verantwortlichen die Einführung eines solchen Plans in Betracht ziehen.

Die Verantwortlichen sollten sich auch der Sicherheitsaspekte bewusst sein. Neue Technologien führen oft zu neuen Risiken. In diesem Zusammenhang sollte nie vergessen werden, dass der Schutz personenbezogener Daten von IT-Sicherheitsmaßnahmen abhängt und daher die Risiken im Zusammenhang mit personenbezogenen Daten auch die Risiken im Zusammenhang mit der IT sind. Folglich werden geeignete technische und organisatorische Maßnahmen, die in der IT implementiert werden, den Datenschutz gemäß der DSGVO gewährleisten. Diese sollten zudem regelmäßig getestet und aktualisiert werden, um Sicherheitsrisiken zu verhindern oder zu minimieren. (siehe Unterabschnitt „Hauptunterschied zu anderen Risiken in der DSGVO und zu Risiken in der IT-Sicherheit“, Abschnitt „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“).

Um diese Risiken zu bewerten und geeignete Garantien abzuleiten, schreibt die DSGVO vor, dass vor der Verarbeitung eine Datenschutz-Folgenabschätzung durchgeführt werden muss, wenn ein hohes Risiko für die Rechte und Freiheiten einer natürlichen Person besteht[411] (siehe Abschnitt „Datenschutz-Folgenabschätzung“, Kapitel „Maßnahmen und Tools“). Der Einsatz neuer Technologien wie KI erhöht die Wahrscheinlichkeit, dass die Verarbeitung in die Hochrisikokategorie fällt. Einige nationale Datenschutzbehörden haben Richtlinien erlassen, die bei der Verwendung bestimmter KI-Algorithmen eine DSFA vorschreiben.[412] Im Zweifelsfall wird den Verantwortlichen empfohlen, eine DSFA durchzuführen.[413]

2.1.3 Präzision

Eine hohe Systempräzision ist in der Regel eines der Entwicklungsziele von KI-Systemen. Viele KI-Systeme benötigen genaue und zuverlässige Trainingsdaten, um optimale Ergebnisse zu erzielen. Bei der Verarbeitung personenbezogener Daten ist es außerdem gesetzlich vorgeschrieben, diese auf dem neuesten Stand zu halten und falsche Eingaben zu berichtigen.[414] Die betroffene Person kann zudem die Berichtigung unrichtiger personenbezogener Daten verlangen.[415] Bei der Entwicklung von KI-Systemen sollte daher die Notwendigkeit einer Nachschulung berücksichtigt werden, bei der Daten nicht nur hinzugefügt, sondern auch gelöscht werden können (siehe Abschnitte „

d) Recht auf Berichtigung“ und „Fairness, Vielfalt und Nichtdiskriminierung“).

Darüber hinaus sollte ein KI-System nicht nur ein Ergebnis ausgeben, sondern auch anzeigen, mit welcher Wahrscheinlichkeit das Ergebnis korrekt ist. Eine solche Anzeige ist nicht nur ein technischer Indikator für die Präzision des Systems, sondern auch ein wertvoller Hinweis darauf, ob ein menschliches Eingreifen erforderlich sein könnte (siehe Abschnitt „Genauigkeit“ im Kapitel „Grundsätze“).

---

[410] Artikel 33 Absatz 3 DSGVO.

[411] Artikel 35 Absatz 1 DSGVO.

[412] Siehe z. B. die Rechtslage in Österreich § 2(2)(4) DSFA-V.

[413] Artikel-29-Datenschutzgruppe, Leitlinien zur Datenschutz-Folgenabschätzung (DSFA) und Beantwortung der Frage, ob eine Verarbeitung im Sinne der Verordnung 2016/679 „wahrscheinlich ein hohes Risiko mit sich bringt“, Seite 8. Europäische Kommission, Brüssel.

[414] Artikel 5 Absatz 1 Buchstabe d DSGVO.

[415] Artikel 16 DSGVO.

2.1.4 Zuverlässigkeit und Reproduzierbarkeit

Viele KI-Systeme werden mit Blick auf einen bestimmten Anwendungsfall entwickelt. Wie bereits erwähnt, kann sich dieser jedoch im Laufe der Zeit weiterentwickeln und sich langsam von den ursprünglichen Absichten der Entwickler entfernen. Daher ist es wichtig, die ursprünglichen Annahmen und Bedingungen, unter denen das KI-System eingesetzt werden sollte, klar zu dokumentieren. Erwartet das KI-System beispielsweise eine bestimmte Umgebung, oder enthält die Trainingsmenge bekannte Verzerrungen? Wenn ein KI-System öffentlich zugänglich ist, sollte auch die Zuverlässigkeit des Systems dokumentiert werden.

Neben der Zuverlässigkeit ist auch die Reproduzierbarkeit der Ergebnisse eines KI-Systems wichtig. Reproduzierbarkeit ist nicht nur eine wünschenswerte technische Eigenschaft eines KI-Systems (z. B. um die Ursache fehlerhafter Ergebnisse zu untersuchen), sondern auch eine wichtige Voraussetzung für Vertrauen. Wenn ein Ergebnis nicht reproduzierbar ist, kann seine Erklärbarkeit – und damit das Vertrauen in das KI-System – leiden.

Checkliste: Technische Robustheit und Sicherheit[416]

Widerstandsfähigkeit gegen Angriffe und Sicherheitsverletzungen

 Der Verantwortliche hat potenzielle Angriffsformen bewertet, für die das KI-System anfällig sein könnte.

 Der Verantwortliche hat verschiedene Arten und Formen von Schwachstellen wie Datenverunreinigung, physische Infrastruktur und Cyberangriffe berücksichtigt.

 Der Verantwortliche hat Maßnahmen oder Systeme eingeführt, um die Integrität und Belastbarkeit des KI-Systems gegen mögliche Angriffe zu gewährleisten.

 Der Verantwortliche hat überprüft, wie sich das System in unerwarteten Situationen und Umgebungen verhält.

 Der Verantwortliche hat geprüft, inwieweit das System einen doppelten Verwendungszweck haben könnte. In diesem Fall ergreift der Verantwortliche geeignete vorbeugende Maßnahmen für diesen Fall (z. B. Verzicht auf die Veröffentlichung der Forschung oder Verzicht auf den Einsatz des Systems).

Auffangplan und allgemeine Sicherheit

 Der Verantwortliche hat sichergestellt, dass das System im Falle feindlicher Angriffe oder anderer unerwarteter Situationen über einen ausreichenden Auffangplan verfügt (z. B. technische Umschaltverfahren oder das Heranziehen eines menschlichen Bedieners, bevor die Arbeit fortgesetzt wird).

 Der Verantwortliche hat das vom KI-System verursachte Risiko in diesem speziellen Anwendungsfall berücksichtigt.

 Der Verantwortliche hat ein Verfahren zur Messung und Bewertung von Risiken und Sicherheit eingeführt.

---

[416] Diese Checkliste wurde anhand der von der hochrangigen Expertengruppe für künstliche Intelligenz (2019) (Ethik-Leitlinien für eine vertrauenswürdige KI) ausgearbeitete Liste erstellt und angepasst. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 20. Mai 2020).

 Der Verantwortliche hat die notwendigen Informationen für den Fall der Bedrohung der körperlichen Unversehrtheit von Menschen bereitgestellt.

 Der Verantwortliche hat eine Versicherung in Betracht gezogen, die mögliche, vom KI-System verursachte Schäden deckt.

 Der Verantwortliche hat die potenziellen Sicherheitsrisiken infolge der vorhersehbaren (und anderer) Anwendungen der Technologie, einschließlich eines versehentlichen oder böswilligen Missbrauchs, ermittelt. Gibt es einen Plan zur Begrenzung oder Minderung dieser Gefahren?

 Der Verantwortliche hat beurteilt, ob die Möglichkeit besteht, dass das KI-System den Anwendern oder Dritten Schaden zufügt. Der Verantwortliche hat die Wahrscheinlichkeit, den potenziellen Schaden, den betroffenen Personenkreis und die Schwere des Schadens eingeschätzt.

 Der Verantwortliche hat die Haftungs- und Verbraucherschutzbestimmungen berücksichtigt.

 Der Verantwortliche hat die möglichen Auswirkungen oder Sicherheitsrisiken für Umwelt oder Tiere in Betracht gezogen.

 Der Verantwortliche hat bei der Risikoanalyse berücksichtigt, ob Sicherheits- oder Netzwerkprobleme (z. B. Bedrohungen der Cybersicherheit) ein Sicherheitsrisiko oder Schäden durch ein unbeabsichtigtes Verhalten des KI-Systems Schäden verursachen könnten.

 Der Verantwortliche hat die möglichen Auswirkungen eines Ausfalls des KI-Systems eingeschätzt und ob dieser zu falschen Ergebnissen, zur Nichtverfügbarkeit des Systems oder zu gesellschaftlich inakzeptablen Ergebnissen (z. B. diskriminierende Praktiken) führen könnte.

 Der Verantwortliche hat Schwellenwerte sowie Maßnahmen zur Lenkung und Kontrolle festgelegt, damit alternative oder Auffangpläne aktiviert werden.

 Der Verantwortliche hat Auffangpläne festgelegt und erprobt.

Präzision

 Der Verantwortliche hat beurteilt, welcher Grad an Präzision und welche Definition von Genauigkeit im Zusammenhang mit dem KI-System und dem jeweiligen Anwendungsfall erforderlich sind.

 Der Verantwortliche hat beurteilt, wie die Präzision gemessen und gewährleistet wird.

 Der Verantwortliche hat Maßnahmen ergriffen, um sicherzustellen, dass die verwendeten Daten umfassend und aktuell sind.

 Der Verantwortliche hat Maßnahmen ergriffen, um zu beurteilen, ob zusätzliche Daten erforderlich sind, z. B. um die Präzision zu verbessern oder Verzerrungen zu vermeiden.

 Der Verantwortliche hat den Schaden bewertet, der entstehen würde, wenn das KI-System ungenaue Vorhersagen machen würde.

 Der Verantwortliche hat Methoden eingeführt, um zu messen, ob das System eine inakzeptable Anzahl nicht präziser Vorhersagen ausgibt.

 Der Verantwortliche hat eine Reihe von Maßnahmen ergriffen, um die Präsizion des Systems zu erhöhen.

Zuverlässigkeit und Reproduzierbarkeit

 Der Verantwortliche hat eine Strategie zur Überwachung und Erprobung eingeführt, ob das KI-System die Ziele, Zwecke und vorgesehenen Anwendungen erfüllt.

 Der Verantwortliche hat getestet, ob bestimmte Kontexte oder Bedingungen berücksichtigt werden müssen, damit die Wiederholbarkeit gewährleistet ist.

 Der Verantwortliche hat Prüfverfahren oder -methoden zur Messung und Sicherstellung verschiedener Aspekte der Zuverlässigkeit und Wiederholbarkeit eingeführt.

 Der Verantwortliche hat Verfahren eingeführt, die beschreiben, wann ein KI-System bei bestimmten Einstellungen ausfällt.

 Der Verantwortliche hat diese Prozesse zur Erprobung und Prüfung der Zuverlässigkeit von KI-Systemen eindeutig dokumentiert und operativ umgesetzt.

 Der Verantwortliche hat ein Kommunikationssystem eingerichtet, mit dem er den (End-)Nutzern die Zuverlässigkeit des KI-Systems garantieren kann.

Zusätzliche Informationen

Artikel-29-Datenschutzgruppe, Leitlinien zur Datenschutz-Folgenabschätzung (DSFA) und Beantwortung der Frage, ob eine Verarbeitung im Sinne der Verordnung 2016/679 „wahrscheinlich ein hohes Risiko mit sich bringt“. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/newsroom/document.cfm?doc_id=47711

K. Eykholt et al. (2018), Robust physical-world attacks on deep learning models, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition 10.4.2018, arXiv:1707.08945

M. Fredrikson et al. (2015), Model inversion attacks that exploit confidence information and basic countermeasure’, CCS '15: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Oktober 2015. Cornell University, Ithaca. Verfügbar unter: https://rist.tech.cornell.edu/papers/mi-ccs.pdf

3 Schutz der Privatsphäre und Datenqualitätsmanagement

„Eng verbunden mit dem Grundsatz der Schadensverhütung ist der Schutz der Privatsphäre, ein Grundrecht, auf das sich KI-Systeme insbesondere auswirken. Die Schadensverhütung in Bezug auf den Schutz der Privatsphäre erfordert außerdem ein angemessenes Datenqualitätsmanagement, wozu auch die Qualität und Integrität der verwendeten Daten, ihre Relevanz gegenüber dem Bereich, in dem die KI-Systeme eingesetzt werden, die Zugangsprotokolle sowie die Fähigkeit zur Datenverarbeitung unter Wahrung des Datenschutzes gehören.“

– *Hochrangige Expertengruppe für künstliche Intelligenz*[417]

[417] Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 17. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 28. Mai 2020).

3.1 Ethische Grundsätze

Diese Anforderung umfasst drei unterschiedliche Hauptgrundsätze, und zwar:[418]

- **Schutz der Privatsphäre und Datenschutz.** KI-Systeme müssen den Schutz der Privatsphäre und den Datenschutz in allen Phasen des Lebenszyklus eines Systems gewährleisten. Dies umfasst auch anfänglich vom Benutzer angegebene Informationen sowie die im Verlauf der Interaktion mit dem System über den Benutzer erzeugten Informationen (z. B. Ergebnisse, die das KI-System für bestimmte Benutzer erzeugt oder die Reaktion von Benutzern auf bestimmte Empfehlungen). Damit die Menschen Vertrauen in die Datenverarbeitung haben können, muss sichergestellt sein, dass die über sie gesammelten Daten nicht dazu verwendet werden, sie unrechtmäßig oder unfair zu diskriminieren.
- **Qualität und Integrität der Daten.** Die Qualität der verwendeten Datensätze ist für die Leistungsfähigkeit von KI-Systemen von entscheidender Bedeutung. Bei der Erfassung der Daten können sozial bedingte Verzerrungen, Ungenauigkeiten, Fehler und andere Mängel auftreten. Solche Probleme müssen vor dem Training mit einem bestimmten Datensatz behoben werden. Darüber hinaus muss die Integrität der Daten gewährleistet sein. Die Eingabe schädlicher Daten in ein KI-System kann sein Verhalten verändern, insbesondere bei selbstlernenden Systemen. Die verwendeten Prozesse und Datensätze müssen in allen Schritten wie Planung, Training, Erprobung und Einsatz getestet und dokumentiert werden. Dies sollte auch für KI-Systeme gelten, die nicht intern entwickelt, sondern von außerhalb erworben werden.
- **Datenzugriff.** In jeder Organisation, die mit personenbezogenen Daten umgeht, müssen interne Dokumente/Richtlinien vorhanden sein, die festlegen, wer und unter welchen Bedingungen auf personenbezogene Daten zugreifen darf, einschließlich organisatorischer und technischer Maßnahmen zur Zugangskontrolle. Nur entsprechend qualifiziertes Personal sollte auf die Daten zugreifen dürfen, wenn die Kompetenz und die Notwendigkeit dafür gegeben sind. Außerdem müssen alle Mitarbeiter, denen Zugang gewährt wird, eine Vertraulichkeitserklärung unterzeichnen.

3.2

Die DSGVO bezieht sich auf die Verarbeitung der personenbezogenen Daten von betroffenen Personen. Es gibt einige Bestimmungen, die für den Datenschutz und das Datenqualitätsmanagement besonders relevant sind. Da die Qualität und Integrität von Daten sowie der Zugriff auf Daten im vorherigen Abschnitt analysiert wurden, konzentrieren wir uns hier auf vier Konzepte, die für die Gewährleistung eines angemessenen Datenqualitätsmanagements äußerst wichtig sind. Diese sind: (1) Zweckbindung; (2) Rechtmäßigkeit; (3) Datenminimierung sowie (4) Verarbeitung nach Treu und Glauben, ein umfassender Grundsatz, der den Schutz der Rechte der betroffenen Personen erfordert.

Es ist sinnlos, über Datenschutz zu sprechen, wenn die Verarbeitung nicht rechtmäßig ist, wobei ein eindeutiger und ausdrücklicher Zweck eine Voraussetzung für eine rechtmäßige

---

[418] Ibid., Seite15 ff.

Verarbeitung ist. Doch selbst wenn die Verarbeitung zulässig (d. h. rechtmäßig und legitim) ist, lässt sich der Datenschutz nicht umsetzen, wenn die Zwecke der Verarbeitung unklar sind. Außerdem ist die Verarbeitung nicht rechtmäßig, wenn sie nicht mit den Zwecken zusammenhängt, für die die Daten erhoben wurden. Daher steht der Grundsatz der Zweckbindung in direktem Zusammenhang mit dem Datenqualitätsmanagement.

Die Datenminimierung ist derweil der Schlüssel zum Schutz der Privatsphäre. Der beste Weg zur Sicherstellung, dass „Daten, die über [die betroffenen Personen] gesammelt werden, nicht dazu verwendet werden, sie unrechtmäßig oder unfair zu diskriminieren", besteht darin, die Menge und den Umfang der gesammelten personenbezogenen Daten zu minimieren. Schließlich ist eine angemessene Umsetzung der Rechte der betroffenen Personen, wie sie in der DSGVO verankert sind, von wesentlicher Bedeutung, um diese zu befähigen und das Datenqualitätsmanagement zu stärken. [419]

3.2.1 Zweckbindung

Der Grundsatz der Zweckbindung beschränkt die Verwendung personenbezogener Daten auf den/die ursprünglichen Zweck(e) oder auf solche Zwecke, die damit vereinbar sind. Die Entwicklung einer KI erfordert jedoch, dass Daten recht häufig wiederverwendet werden. Darüber hinaus kann es vorkommen, dass das KI-Tool die Daten automatisch wiederverwendet (was im Falle des Deep Learning sicherlich der Fall ist). Diese Situationen führen zu Spannungen zwischen den KI-Trainingstechniken und dem Grundsatz der Zweckbindung.

Um eine unrechtmäßige Datenverarbeitung zu vermeiden, sollten Verantwortliche, die KI-Systeme einsetzen, **den Zweck der Verarbeitung „zu Beginn des Trainings oder Einsatzes festlegen und diese Festlegung neu bewerten, falls die Verarbeitung des Systems unerwartete Ergebnisse liefert**, da er verlangt, dass personenbezogene Daten nur für „festgelegte, eindeutige und legitime Zwecke" erhoben und nicht in einer Weise verwendet werden, die mit dem ursprünglichen Zweck unvereinbar ist"[420] (siehe Abschnitt „Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen" im Kapitel „Konzepte").

Die **Weiterverwendung von Daten** bei der Entwicklung eines KI-Tools bringt große Herausforderungen in Bezug auf die Zweckbindung mit sich. KI-Systeme verarbeiten personenbezogene Daten in verschiedenen Phasen und für eine Vielzahl von Zwecken. Infolgedessen kann ein Verantwortlicher unter Umständen einzelne Verarbeitungsvorgänge nicht unterscheiden und Daten für andere Zwecke verarbeiten als die, für die sie ursprünglich erhoben wurden. Die Verantwortlichen sollten besonders auf diese Situationen achten, da sie zu einem Verstoß gegen den Datenschutzgrundsatz der Rechtmäßigkeit führen könnten[421]

---

[419] Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 17. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 28. Mai 2020).

[419] Ibid., Seite15 ff.

[420] CIPL (2020) Artificial intelligence and data protection: how the GDPR regulates AI. Centre for Information Policy Leadership, Washington DC / Brüssel / London, Seite 6. Hervorgehoben durch den Autor. Verfügbar unter: www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (abgerufen am 17. Mai 2020).

[421] ICO (2020) Guidance on the AI auditing framework: draft guidance for consultation. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

(siehe Unterabschnitt „Verwendung für unvereinbare Zwecke“, Abschnitt „Zweckbindung“, Kapitel „Grundsätze“).

Die Verantwortlichen müssen berücksichtigen, dass die Bestimmung der angemessenen Rechtsgrundlage an die Grundsätze von Treu und Glauben (Fairness) und der Zweckbindung gebunden ist.[422] Sie müssen die Rechtsgrundlage wählen, die der wahren Natur ihrer Beziehung zu der betroffenen Person und dem Zweck der Verarbeitung am ehesten entspricht. Diese Entscheidung ist von grundlegender Bedeutung, da die Rechtsgrundlage für die Verarbeitung infolge des Grundsatzes der Zweckbindung nur dann geändert werden darf, wenn es stichhaltige Gründe gibt, die dies rechtfertigen. Wenn die KI-Entwickler planen, einen Datensatz in verschiedenen Phasen (z. B. Training, Validierung oder Einsatz) zu verwenden, sollten sie diese Schritte als unterschiedliche und separate Zwecke betrachten.[423] Darüber hinaus **müssen sie die Art der Beziehung berücksichtigen, die sie mit der betroffenen Person unterhalten.** So könnte beispielsweise die Einwilligung eine geeignete Rechtsgrundlage für die Verarbeitung sein, wenn ein ständiger Kontakt mit den betroffenen Personen besteht und die Verantwortlichen in der Lage sind, von den betroffenen Personen nacheinander Einwilligungen für verschiedene Verwendungszwecke oder für mehrere verschiedene Verarbeitungsvorgänge einzuholen, bevor die Verarbeitung beginnt. Im Falle der KI ist es jedoch oft schwierig, diese Art von Beziehung aufrechtzuerhalten, da künstliche Intelligenz oft durch die Aggregation und Zusammenführung großer Datensätze entsteht.

Nicht zuletzt sollten sich die Verantwortlichen darüber im Klaren sein, dass die Rechtsvorschriften der Union oder der Mitgliedstaaten für die Verarbeitung personenbezogener Daten zu wissenschaftlichen oder historischen Forschungszwecken oder zu statistischen Zwecke Ausnahmen von den in den Artikeln 15, 16, 18, 21 und 21 festgelegten Rechten der betroffenen Personen vorsehen können. Die Verarbeitung dieser Daten zu anderen Zwecken als denen, für die sie ursprünglich erhoben wurden, sollte daher rechtmäßig sein, sofern geeignete technische und organisatorische Maßnahmen, insbesondere zur Datenminimierung, getroffen werden. (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“ in diesen Leitlinien).

Checkliste: Zweckbindung[424]

 Die Verantwortlichen haben ihren Zweck oder ihre Zwecke für die Verarbeitung eindeutig festgelegt.

 Die Verantwortlichen haben diese Zwecke dokumentiert.

 Die Verantwortlichen haben Details zu ihren Zwecken in die Datenschutzerklärungen für natürliche Personen aufgenommen.

---

[422] EDSA (2018), Leitlinien 2/2019 für die Verarbeitung personenbezogener Daten gemäß Artikel 6 Absatz 1 Buchstabe b DSGVO im Zusammenhang mit der Erbringung von Online-Diensten für betroffene Personen, angenommen am 9. April 2019, Seite 6. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_de_0.pdf (abgerufen am 15. Mai 2020).

[423] ICO (2020) Guidance on the AI auditing framework: draft for consultation. 2020. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[424] ICO (keine Daten) Principle (b): purpose limitation. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/ (abgerufen am 17. Mai 2020).

 Die Verantwortlichen überprüfen regelmäßig die Verarbeitung und aktualisieren gegebenenfalls ihre Dokumentation und die Informationen zum Schutz der Privatsphäre von Personen.

 Wenn die Verantwortlichen beabsichtigen, personenbezogene Daten für einen neuen Zweck zu verwenden, der nicht auf einer rechtlichen Verpflichtung oder einer gesetzlich festgelegten Funktion beruht, prüfen sie, ob dies mit dem ursprünglichen Zweck vereinbar ist, oder sie holen eine spezifische Einwilligung für den neuen Zweck ein.

Zusätzliche Informationen

Artikel-29-Datenschutzgruppe (2013) Stellungnahme 03/2013 zur Zweckbindung. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

CIPL (2020) Artificial intelligence and data protection: how the GDPR regulates AI. Centre for Information Policy Leadership, Washington DC / Brüssel / London. Verfügbar unter: www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf

EDSA (2018), Leitlinien 2/2019 für die Verarbeitung personenbezogener Daten gemäß Artikel 6 Absatz 1 Buchstabe b DSGVO im Zusammenhang mit der Erbringung von Online-Diensten für betroffene Personen, angenommen am 9. April 2019, Seite 6. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_de_0.pdf

ICO (2020) Guidance on the AI auditing framework: draft guidance for consultation. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

ICO (keine Daten) Principle (b): purpose limitation. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/

3.2.2 Rechtmäßigkeit

Die Rechtmäßigkeit ist ein wesentlicher Grundsatz des Datenschutzes. Er besagt, dass die Verantwortlichen sicherstellen müssen, dass sie über eine **Rechtsgrundlage für die Verarbeitung personenbezogener Daten verfügen**. **Anderenfalls** darf die Verarbeitung **nicht durchgeführt werden**.[425] Die Rechtsgrundlagen für die Verarbeitung von Daten, einschließlich besonderer Datenkategorien, sind in Artikel 6 und 9 DSGVO beschrieben. Im Falle der künstlichen Intelligenz werden in der Regel folgende Rechtsgrundlagen zur Rechtfertigung der Verarbeitung herangezogen: Einwilligung, berechtigtes Interesse, vertragliche Notwendigkeit, rechtliche Verpflichtung oder lebenswichtiges Interesse. Die

[425] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, Seite 20. Agencia Espanola Proteccion Datos, Madrid. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

Verarbeitung im öffentlichen Interesse kann ebenfalls eine Rechtsgrundlage sein. Wr werden uns hier jedoch nicht darauf konzentrieren, da wir dieses Thema im Abschnitt „Datenschutz und wissenschaftliche Forschung“ im Kapitel „Konzepte“ ausführlich behandeln. Im Mittelpunkt hier stehen daher die vier aufgeführten Rechtsgrundlagen.

a) Einwilligung

Die Datenverarbeitung beruht häufig auf der Einwilligung der betroffenen Personen (siehe Abschnitt „Einwilligung“ im Kapitel „Konzepte“). Die Einwilligung passt jedoch nicht gut zum Wesen der meisten KI-Entwicklungen, und zwar aus einem einfachen Grund: Die Einwilligung ist von Natur aus an einen genau definierten, konkreten Zweck gebunden.[426] Im Fall der künstlichen Intelligenz entsteht durch die Verwendung von Big Data und die Aggregation, gemeinsame Nutzung oder Neuausrichtung von Daten ein Szenario, das nicht zu den zugrundeliegenden Grundsätzen des Konzepts der Einwilligung und dem Grundsatz der Zweckbindung passt.

Die Einwilligung kann eine nützliche Rechtsgrundlage für die Datenverarbeitung im Rahmen einer KI-Entwicklung sein, insbesondere wenn die Verantwortlichen eine direkte Beziehung zur betroffenen Personen haben, die die erforderlichen Daten für das Training, die Validierung und den Einsatz des Modells bereitstellt.[427] Wenn das KI-Tool beispielsweise darauf abzielt, Diagnosen von Lungenentzündungen zu erstellen, und Ärzte Daten von Patienten in ihrer Gesundheitseinrichtung erhalten, könnte die Einwilligung als Rechtsgrundlage für die Verarbeitung gut geeignet sein. Beinhaltet die Verarbeitung jedoch den Einsatz eines komplexen KI-Tools, das die Daten umfassender verwenden kann (beispielweise, wenn Profiling und automatisierte Entscheidungsfindung unbeabsichtigt erfolgen könnten, Daten wahrscheinlich während der Verarbeitung hergeleitet werden, diese hergeleiteten Daten für verschiedene Zwecke verwendet werden können usw.), ist es schwer vorstellbar, wie eine einzige Einwilligung die gesamte Verarbeitung rechtfertigen könnte. Zu diesem Zweck müssen die Verantwortlichen die Leitlinien der Artikel-29-Datenschutzgruppe zur Einwilligung sorgfältig beachten[428] (siehe auch Abschnitt „Einwilligung“ im Kapitel „Konzepte“).

Im Rahmen der wissenschaftlichen Forschung (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“ im Kapitel „Konzepte“) sieht die DSGVO spezifische Ausnahmen von den Einwilligungsattributen vor, anhand der die Verantwortlichen eine **breite Einwilligung** als Rechtsgrundlage für die Verarbeitung verwenden können. Die breite Einwilligung ist im Zusammenhang mit Erwägungsgrund 33 der Datenschutz-Grundverordnung zu verstehen, in dem es heißt: „Oftmals kann der Zweck der Verarbeitung personenbezogener Daten für Zwecke der wissenschaftlichen Forschung zum Zeitpunkt der Erhebung der personenbezogenen Daten nicht vollständig angegeben werden. Daher sollte es betroffenen Personen erlaubt sein, ihre Einwilligung für bestimmte Bereiche

---

[426] International Bioethics Committee (2017) Report of the IBC on big data and health, Seite 20. UNESCO. Verfügbar unter: http://unesdoc.unesco.org/images/0024/002487/248724e.pdf (abgerufen am 13. März 2020).

[427] ICO (kein Datum) How do we apply legitimate interests in practice? Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/ (abgerufen am 15. Mai 2020). Darüber hinaus muss die Bewertung der Art dieser Beziehung eine Untersuchung des Machtgefälles zwischen der betroffenen Person und dem Verantwortlichen umfassen.

[428] Artikel 29-Datenschutzgruppe, Leitlinien 05/2020 zur Einwilligung gemäß Verordnung 2016/679. Europäische Kommission, Brüssel, Seite 29. Verfügbar unter: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_de.pdf (abgerufen am 5. Mai 2020).

wissenschaftlicher Forschung zu geben, wenn dies unter Einhaltung der anerkannten ethischen Standards der wissenschaftlichen Forschung geschieht."

Die breite Einwilligung ist jedoch keine pauschale Einwilligung oder gleichbedeutend mit einer offenen Einwilligung. Sie ist ein **außergewöhnliches Instrument, das nur bei Vorliegen mehrerer Bedingungen akzeptabel sein kann.** Wenn eine breite Einwilligung für besondere Datenkategorien verwendet wird, sollten die Verantwortlichen sicherstellen, dass ihre nationalen Vorschriften dies zulassen. Sie sollten sich auch über die zu treffenden Sicherheitsvorkehrungen im Klaren sein. Die Verhältnismäßigkeit zwischen dem Ziel der Forschung und der Verwendung besonderer Kategorien von Daten muss gewährleistet sein. Darüber hinaus müssen die Verantwortlichen sicherstellen, dass die Vorschriften ihrer Mitgliedstaaten genetische, biometrische und Gesundheitsdaten nicht durch die Einführung zusätzlicher Bedingungen oder Beschränkungen schützen, da dies nach der Datenschutz-Grundverordnung zulässig ist.

Wenn eine breite Einwilligung zum Erreichen des Forschungszwecks verwendet wird, sollten außerdem einige wesentliche Maßnahmen erwogen werden, um die abstrakte Definition des Forschungszwecks auszugleichen. Die Einhaltung der anerkannten ethischen Standards der wissenschaftlichen Forschung gemäß Erwägungsgrund 33 der Datenschutz-Grundverordnung scheint für diesen Zweck besonders wichtig zu sein.

Kasten 3: Breite Einwilligung und zusätzliche Sicherungsmaßnahmen

Die deutsche Datenschutzaufsichtsbehörde hat vor Kurzem einige zusätzliche Sicherungsmaßnahmen aufgelistet, die im Falle einer breiten Einwilligung zu treffen sind.[429] Diese sind:

1. Zusätzliche Sicherungsmaßnahmen zur Gewährleistung von Transparenz:

- Verwendung einer für den Einwilligenden zugänglichen Nutzungsordnung oder eines einsehbaren Forschungsplanes, der die geplanten Arbeitsmethoden und die Fragen, die Gegenstand der Forschung sein sollen, beleuchtet
- Ausarbeitung und Dokumentation im Hinblick auf das konkrete Forschungsprojekt, wieso in diesem Fall eine nähere Konkretisierung der Forschungszwecke nicht möglich ist
- Einrichten einer Internetpräsenz, durch die die Studienteilnehmer über laufende und künftige Studien informiert werden

2. Zusätzliche Sicherungsmaßnahmen zur Vertrauensbildung:

- Positives Votum eines Ethikgremiums vor der Nutzung für weitere Forschungszwecke
- Prüfung, ob das Arbeiten mit einem „dynamic consent" möglich ist bzw. Einräumung einer Widerspruchsmöglichkeit vor der Verwendung der Daten für neue Forschungsfragen

3. Zusätzliche Garantiemaßnahmen zur Datensicherheit:

[429] DSK, Beschluss der 97. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder zu Auslegung des Begriffs „bestimmte Bereiche wissenschaftlicher Forschung" im Erwägungsgrund 33 der DS-GVO 3. April 2019, unter: www.datenschutzkonferenz-online.de/media/dskb/20190405_auslegung_bestimmte_bereiche_wiss_forschung.pdf (abgerufen am 20. Mai 2020). Eine englische Zusammenfassung dieser Maßnahmen finden Sie hier: www.technologylawdispatch.com/2019/04/privacy-data-protection/german-dpas-publish-resolution-on-concept-of-broad-consent-and-the-interpretation-of-certain-areas-of-scientific-research/

- Keine Datenweitergabe in Drittländer mit geringerem Datenschutzniveau
- Gesonderte Zusagen zur Datenminimierung, Verschlüsselung, Anonymisierung oder Pseudonymisierung
- Spezifische Vorschriften für die Begrenzung des Zugriffs auf die erhobenen Daten

In jedem Fall müssen die Forschungsteilnehmer die Möglichkeit haben, ihre Einwilligung zurückzuziehen und sich für oder gegen bestimmte Forschungsarbeiten oder Teile davon zu entscheiden. Sie müssen sich auch sicher sein, dass ihre Rechte durch die Einhaltung der ethischen Standards der wissenschaftlichen Forschung gewahrt werden.[430] Mitunter kann dies der KI-Lösung schaden oder die Verantwortlichen zu komplexen Maßnahmen zwingen. Daher sollten die Verantwortlichen prüfen, ob alternative Rechtsgrundlagen besser geeignet sind, um das Tool unter Einhaltung der Gesetze zu entwickeln.

Zusammenfassend lässt sich sagen, dass die Verantwortlichen bei der Verwendung der Einwilligung als Rechtsgrundlage der Datenverarbeitung Vorsicht walten lassen sollten, denn die Einwilligung befreit sie nicht von ihren Pflichten in Bezug auf Gerechtigkeit, Notwendigkeit und Verhältnismäßigkeit der Verarbeitung.[431] Darüber lässt es sich im Fall der künstlichen Intelligenz, die Big Data verwendet, oft nur schwer rechtfertigen, dass die Einwilligung alle notwendigen Anforderungen erfüllt: freiwillig, für einen konkreten Fall, nach ausreichender Information des Betroffenen und unmissverständlich abgegeben sowie eine klare bestätigende Handlung seitens der betroffenen Person. Generell gilt: Je mehr Dinge die KI-Entwickler mit den Daten machen wollen, desto schwieriger ist es sicherzustellen, dass die Einwilligung wirklich spezifisch und in Kenntnis der Sachlage erfolgt. Dies alles sollte bei der Wahl der Einwilligung als Rechtsgrundlage für die Datenverarbeitung berücksichtigt werden.

Kasten 4. Einwilligung als Rechtsgrundlage: der Fall OkCupid

2016 veröffentlichte eine Gruppe dänischer Forscher einen Datensatz von etwa 70.000 Nutzern. Diese Daten stammten von der Online-Dating-Website OkCupid[432] und umfassten Datenkategorien wie Nutzernamen, Alter, Geschlecht, Standort, die Art der Beziehung (oder des Geschlechts), an der die betroffenen Personen interessiert waren, ihre Persönlichkeitsmerkmale usw.

Die Forscher waren der Ansicht, dass allein die Tatsache, dass diese Daten öffentlich zugänglich waren (auf den Dating-Profilen der Nutzer), eine Rechtsgrundlage für die weitere Verarbeitung darstellte. Dies ist ein hervorragendes Beispiel für die schrecklichen Folgen des Arguments, dass „die Daten bereits öffentlich sind“. Die personenbezogenen Daten der betroffenen Personen wurden ohne deren Einwilligung verarbeitet und sehr sensible Informationen der Öffentlichkeit zugänglich gemacht.

Leider ist diese Verknüpfung zwischen öffentlichen und offenen Daten immer noch zu weit verbreitet. Forscher sollten sich darüber im Klaren sein, dass eine für eine konkrete

[430] A. Kuyumdzhieva, (2018), Ethical challenges in the digital era: focus on medical research, Seiten 45–62 in: Z. Koporc (ed.) *Ethics and integrity in health and life sciences research*. Bingley Emerald.

[431] Artikel 29-Datenschutzgruppe, Leitlinien 05/2020 zur Einwilligung gemäß Verordnung 2016/679. WP259. Europäische Kommission, Brüssel, Seite 3. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 (abgerufen am 15. Mai 2020).

[432] www.okcupid.com (abgerufen am 5. Mai 2020).

Verarbeitung erteilte Einwilligung keine Rechtsgrundlage für weitere Verarbeitungen darstellt und dass „öffentlich zugänglich“ nicht gleichbedeutend ist mit „offenen Daten“, d. h. mit Daten und Inhalten, die von jedermann zu jedem Zweck frei verwendet, geändert und weitergegeben werden können, wie vom Open Data Institute definiert.[433]

Checkliste: Einwilligung

 Die Verantwortlichen haben geprüft, ob die Einwilligung die am besten geeignete Rechtsgrundlage für die Verarbeitung ist.

 Die Verantwortlichen holen die Einwilligung der betroffenen Personen freiwillig, für einen konkreten Fall, nach ausreichender Information und unmissverständlich ein.

 Eine breite Einwilligung wird nur dann verwendet, wenn es schwierig oder unwahrscheinlich ist, vorherzusehen, wie diese Daten in Zukunft verarbeitet werden.

 Die für die Verarbeitung besonderer Datenkategorien verwendete breite Einwilligung ist mit den nationalen Vorschriften vereinbar.

 Wenn eine breite Einwilligung verwendet wird, haben die betroffenen Personen die Möglichkeit, ihre Einwilligung zu widerrufen und zu entscheiden, ob sie an bestimmten Forschungsarbeiten oder Teilen davon teilnehmen wollen oder nicht.

 Die Verantwortlichen stehen in einer direkten Beziehung zu der betroffenen Person, die die Daten für die das Training, die Validierung und den Einsatz des IA-Modells zur Verfügung stellt.

 Es besteht kein Machtgefälle zwischen den Verantwortlichen und den betroffenen Personen.

 Die Verantwortlichen fordern die Personen auf, aktiv zuzustimmen.

 Die Verantwortlichen verwenden keine Markierungskästchen oder eine andere Art von Standardeinwilligung.

 Die Verantwortlichen verwenden eine klare, einfache und leicht verständliche Sprache.

 Die Verantwortlichen geben an, wozu sie die Daten benötigen und was sie mit ihnen vorhaben.

 Die Verantwortlichen bieten separate („granulare”) Optionen zur Einwilligung in verschiedene Zwecke und Arten der Verarbeitung.

 Die Verantwortlichen teilen den betroffenen Personen mit, dass sie ihre Einwilligung widerrufen können und wie sie dies tun können.

 Die Verantwortlichen stellen sicher, dass die betroffenen Personen ihre Einwilligung ohne Nachteile verweigern können.

 Die Verantwortlichen vermeiden es, die Einwilligung zur Vorbedingung für einen Dienst zu machen.

b) Berechtigte Interessen

Das berechtigte Interesse kann als Rechtsgrundlage für die Verarbeitung zum Zwecke der KI-Entwicklung verwendet werden, sofern das Ergebnis der Abwägungsprüfung dies rechtfertigt.

[433] http://opendefinition.org/ abgerufen am 5. Mai 2020)

Dies kann bedeuten, dass das Ziel der KI-Verarbeitung zu Beginn festgelegt und zudem sichergestellt werden muss, dass der ursprüngliche Zweck der Verarbeitung neu bewertet wird, wenn das KI-System ein unerwartetes Ergebnis liefert, sodass entweder die verfolgten berechtigten Interessen ermittelt oder eine gültige Einwilligung der betroffenen Personen eingeholt werden kann.[434] Die Abwägungsprüfung **sollte in den Aufzeichnungen über die Verarbeitung angemessen dokumentiert werden.** In einigen Fällen kann es jedoch sein, dass das berechtigte Interesse für die Zwecke der KI-Verarbeitung nicht ausreicht. Wenn die Verantwortlichen beispielsweise planen, eine beträchtliche Menge personenbezogener Daten „für den Fall der Fälle" zu sammeln, sollten sie das berechtigte Interesse nicht als Rechtsgrundlage für die Datenverarbeitung in Betracht ziehen, da die Abwägung zwischen der Notwendigkeit der Verarbeitung und den möglichen Auswirkungen der Verarbeitung auf Personen dies kaum rechtfertigen würde.[435]

Checkliste: Berechtigtes Interesse als Rechtsgrundlage

 Die Verantwortlichen haben geprüft, ob das berechtigte Interesse die am besten geeignete Grundlage ist.

 Die Verantwortlichen sind sich ihrer Verantwortung für den Schutz der Interessen der betroffenen Personen bewusst.

 Die Verantwortlichen führen Aufzeichnungen über die getroffenen Entscheidungen und die ihnen zugrunde liegenden Überlegungen, um sicherzustellen, dass sie ihre Entscheidung rechtfertigen können.

 Die Verantwortlichen haben die relevanten berechtigten Interessen ermittelt.

 Die Verantwortlichen haben sich vergewissert, dass die Verarbeitung notwendig ist und dass es keinen weniger in die Privatsphäre eingreifenden Weg gibt, um das gleiche Ergebnis zu erzielen.

 Die Verantwortlichen haben eine Abwägung vorgenommen und sind zuversichtlich, dass die Interessen des Einzelnen diese berechtigten Interessen nicht überwiegen.

 Die Verantwortlichen verwenden die Daten von Personen nur in einer Weise, die diese vernünftigerweise erwarten würden, es sei denn, die Verantwortlichen haben einen triftigen Grund.

 Die Verantwortlichen verwenden die Daten von Personen nicht in einer Weise, die diese als in die Privatsphäre eingreifend empfinden würden oder die ihnen Schaden zufügen könnte, es sei denn, die Verantwortlichen haben einen sehr triftigen Grund.

 Wenn die Verantwortlichen Daten von Kindern verarbeiten, achten sie besonders darauf, dass die Interessen der Kinder geschützt werden.

 Die Verantwortlichen haben Garantien in Betracht gezogen, um die Auswirkungen nach Möglichkeit zu verringern.

---

[434] CIPL (2020) Artificial intelligence and data protection. How the GDPR regulates AI. Centre for Information Policy Leadership, Washington DC / Brüssel / London, Seite 5. Verfügbar unter: www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (abgerufen am 15. Mai 2020).

[435] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 22. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

- Die Verantwortlichen haben geprüft, ob sie ein Opt-out anbieten können.
- Die Verantwortlichen haben geprüft, ob sie zudem eine Datenschutz-Folgenabschätzung durchführen müssen.

c) Erfüllung eines Vertrags

Die Erfüllung eines Vertrags, bei dem die betroffene Person Vertragspartei ist, oder die Durchführung von Maßnahmen auf Antrag der betroffenen Person vor Abschluss eines Vertrags kann als Rechtsgrundlage für die Verarbeitung dienen, wenn die Verwendung künstlicher Intelligenz für einen dieser Zwecke objektiv erforderlich ist. Dies könnte bei Entwicklern der Fall sein, die betroffene Personen einstellen, um deren personenbezogenen Daten in der Trainingsphase des Systems zu verwenden. Der Verantwortliche, der interessierten Dritten einen die KI-Lösung umfassenden Dienst anbietet, könnte die Daten dieser Personen auch im Rahmen des Dienstleistungsvertrags verwendet.[436] Dieser Rechtsgrund sollte jedoch nach dem Grundsatz der Zweckbindung nicht für andere Zwecke (wie z. B. Systemverbesserung o. ä.) verwendet werden, da die zur Vertragserfüllung verwendeten Daten für diese alternativen Ziele nicht erforderlich sind.[437] So können Verantwortliche Daten, die für die Erfüllung eines Vertrags unabdingbar sind, laut dieser Rechtsgrundlage verarbeiten, wenn sie für die Erfüllung des Vertrags objektiv erforderlich sind, nicht aber für andere Zwecke.[438] Zusammenfassend lässt sich sagen, dass es nur schwer vorstellbar ist, wie die Erfüllung eines Vertrags als Rechtsgrundlage für KI-Forschung und -Innovation dienen könnte.

d) Rechtliche Verpflichtung oder lebenswichtiges Interesse

Gemäß Artikel 6 Absatz 1 Buchstabe d DSGVO können Daten verarbeitet werden, wenn die „Verarbeitung erforderlich ist, um lebenswichtige Interessen der betroffenen Person oder einer anderen natürlichen Person zu schützen“. Ebenso ist die Verarbeitung rechtmäßig, wenn sie „zur Erfüllung einer rechtlichen Verpflichtung erforderlich, der der Verantwortliche unterliegt“ (Artikel 6 Absatz 1 Buchstabe c). Wenn es sich um besondere Datenkategorien handelt, gibt es alternative rechtliche Gründe für die Verarbeitung, wie in Artikel 9 Absatz 2 dargelegt. Auch hier **ist es schwierig, sich einen einzigen Fall vorzustellen, in dem eine dieser Grundlagen** zum jetzigen Zeitpunkt eine Rechtsgrundlage für das Training eines KI-Systems darstellen könnte, auch wenn sich dies durch die Überarbeitung bestehender Vorschriften auf nationaler und europäischer Ebene in Zukunft ändern könnte. Für das Training potenziell lebensrettender KI-Systeme wäre es in jedem Fall besser, sich auf andere Rechtsgrundlagen zu stützen, wie etwa die Einwilligung oder das öffentliche Interesse.[439]

---

[436] Ibid., Seite 20.

[437] Artikel-29-Datenschutzgruppe (2014), Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG. Europäische Kommission, Brüssel, Seiten 16–17. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (abgerufen am 16. Mai 2020).

[438] EDSA (2019), Leitlinien 2/2019 für die Verarbeitung personenbezogener Daten gemäß Artikel 6 Absatz 1 Buchstabe b DSGVO im Zusammenhang mit der Erbringung von Online-Diensten für betroffene Personen. Europäischer Datenschutzbeauftragter, Brüssel, Seite 14. Verfügbar unter: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_de_0.pdf (abgerufen am 15. Mai 2020).

[439] Artikel-29-Datenschutzgruppe (2014), Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG. Europäische Kommission, Brüssel, Seite 20. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (abgerufen am 15. Mai 2020).

Kasten 5. Beispiele für ein lebenswichtiges Interesse als Rechtsgrundlage für die Datenverarbeitung durch ein KI-Tool

Stellen Sie sich vor, ein Unternehmen entwickelt während der COVID-19-Pandemie ein KI-Tool, das die Krankheit mithilfe der Radiologie diagnostizieren kann. In solchen Fällen könnten die Daten von Patienten auf der Grundlage eines lebenswichtigen Interesses verarbeitet werden, wie in Artikel 9 Absatz 2 Buchstabe c DSGVO festgelegt. Allerdings könnten alternative Rechtsgründe wie etwa ein erhebliches öffentliches Interesse (Artikel 9 Absatz 2 Buchstabe g oder i), angemessener sein.

Zusätzliche Informationen

AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 20. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

Artikel-29-Datenschutzgruppe (2014), Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG. Europäische Kommission, Brüssel. Verfügbar unter: www.dataprotection.ro/servlet/ViewDocument?id=1086

CIPL (2020) Artificial intelligence and data protection. How the GDPR regulates AI. Centre for Information Policy Leadership, Washington DC / Brüssel / London. Verfügbar unter: www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf

EDSA (2019), Leitlinien 2/2019 für die Verarbeitung personenbezogener Daten gemäß Artikel 6 Absatz 1 Buchstabe b DSGVO im Zusammenhang mit der Erbringung von Online-Diensten für betroffene Personen. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_de_0.pdf

EDSA (2020), Leitlinien 05/2020 zur Einwilligung gemäß Verordnung 2016/679, Version 1.1, angenommen am 4. Mai 2020 Verfügbar unter: https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf

EDPS (2017) EDSB (2017), Beurteilung der Erforderlichkeit von Maßnahmen, die das Grundrecht auf Schutz personenbezogener Daten einschränken: Ein Toolkit. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edps.europa.eu/data-protection/our-work/publications/papers/necessity-toolkit_en

Weitere Informationen über das berechtigte Interesse, mit praktischen Fällen und mehreren Verweisen auf die Urteile des Gerichtshofs der Europäischen Union, finden Sie in den nachstehenden Dokumenten.

Future of Privacy Forum (no date) Processing personal data on the basis of legitimate interests under the GDPR. European Judicial Training Network, Brüssel. Verfügbar unter: www.ejtn.eu/PageFiles/17861/Deciphering_Legitimate_Interests_Under_the_GDPR%20(1).pdf

ICO (kein Datum) How do we apply legitimate interests in practice? Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/

ICO (kein Datum) Lawful basis for processing. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/

A. Kuyumdzhieva, (2018), Ethical challenges in the digital era: focus on medical research, Seiten 45–62 in: Z. Koporc (ed.), Ethics and integrity in health and life sciences research. Bingley Emerald.

Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

3.2.3 Datenminimierung

Der Grundsatz der Datenminimierung besagt, dass personenbezogene Daten „dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein müssen".[440] Im Zusammenhang mit künstlicher Intelligenz bedeutet dies in erster Linie, dass **Verantwortliche personenbezogene Daten nicht verwenden sollten, wenn dies nicht erforderlich ist**, d. h. wenn das von dem Verantwortlichen angestrebte Ziel auch ohne die Verarbeitung personenbezogener Daten erreicht werden kann (siehe Abschnitt „Rechtmäßigkeit und Verarbeitung nach Treu und Glauben"). Manchmal können personenbezogene Daten durch nicht-personenbezogene Daten ersetzt werden, ohne dass der Forschungszweck beeinträchtigt wird. Unter solchen Umständen ist die Verwendung anonymisierter Daten gemäß Artikel 89 Absatz 1 DSGVO obligatorisch.

Wenn eine Anonymisierung nicht möglich ist, sollten die Verantwortlichen zumindest versuchen, mit pseudonymisierten Daten zu arbeiten. Letztlich muss jeder Verantwortliche festlegen, welche personenbezogenen Daten für den Zweck der Verarbeitung tatsächlich benötigt werden (und welche nicht), einschließlich der entsprechenden Datenspeicherfristen. Verantwortliche müssen nämlich bedenken, dass die Erforderlichkeit der Verarbeitung im Falle der meisten Rechtsgrundlagen nachgewiesen werden muss – einschließlich aller in Artikel 6 DSGVO genannten Grundlagen mit Ausnahme der Einwilligung und der meisten in Artikel 9 Absatz 2 genannten Grundlagen in Bezug auf besondere Kategorien von Daten. Anders ausgedrückt: Bei den meisten Rechtsgrundlagen für die Verarbeitung personenbezogener Daten müssen die Verantwortlichen sowohl nach dem Grundsatz der Datenminimierung als auch nach dem Grundsatz der Rechtmäßigkeit sicherstellen, dass die KI-Entwicklung nicht ohne die Verwendung personenbezogener Daten erfolgen kann.

Der Begriff der Erforderlichkeit ist jedoch komplex und hat im Recht der Europäischen Union eine eigenständige Bedeutung.[441] Im Allgemeinen wird verlangt, dass die Verarbeitung ein gezieltes und verhältnismäßiges Mittel ist, um einen bestimmten Zweck zu erreichen. Das Argument, dass die Verarbeitung erforderlich ist, weil die Verantwortlichen sich entschieden haben, ihr Geschäft auf eine bestimmte Weise zu betreiben, reicht nicht aus. Sie müssen nachweisen können, dass die Verarbeitung **für das angestrebte Ziel erforderlich** ist und **weniger in die Privatsphäre eingreift als andere Möglichkeiten,** um das gleiche Ziel zu erreichen. Es genügt nicht, dass die Verarbeitung notwendiger Bestandteil der von ihnen

[440] Artikel 5 Absatz 1 Buchstabe c DSGVO.

[441] Siehe CJEU, Urteil C-524/06, Heinz Huber gegen Bundesrepublik Deutschland, 18. Dezember 2008, Abs. 52.

gewählten Methoden ist.[442] Wenn es realistische, weniger in die Privatsphäre eingreifende Alternativen gibt, wird die Verarbeitung personenbezogener Daten nicht als notwendig erachtet.[443]

Daher verlangt der Grundsatz der Datenminimierung von KI-Entwicklern, solche Tools zu wählen, deren Entwicklung im Vergleich zu den verfügbaren Alternativen eine minimale Verwendung personenbezogener Daten erfordert. Sobald dies erreicht ist, sollten spezifische Prozesse eingerichtet werden, um die Erhebung und/oder Übermittlung unnötiger personenbezogener Daten auszuschließen, Datenfelder zu reduzieren und automatische Löschmechanismen vorzusehen.[444] Die Datenminimierung kann insbesondere im Falle des Deep Learning komplex sein, da eine Unterscheidung nach Merkmalen unter Umständen nicht möglich ist. Wenn alternative Lösungen zu den gleichen Ergebnissen führen könnten, sollte Deep Learning daher besser vermieden werden.

Das CIPL stellt ferner fest: „Welche personenbezogenen Daten genau als ‚erforderlich' erachtet werden, variiert je nach KI-System und dem Ziel, für das es verwendet wird. Die Regelung der DSGVO in diesem Bereich sollte verhindern, dass der Perfektionismus zum Feind des Guten für KI-Designer wird – die Tatsache, dass die personenbezogenen Daten begrenzt werden müssen, bedeutet nicht, dass das KI-System selbst unbrauchbar wird, insbesondere da nicht alle KI-Systeme ein präzises Ergebnis liefern müssen."[445] Um den Umfang und die Menge der benötigten personenbezogenen Daten genau zu bestimmen, **ist ein Experten extrem nützlich, der in der Lage ist, die relevanten Merkmale auszuwählen.** Dies sollte das Risiko für die Privatsphäre der betroffenen Personen deutlich verringern – ohne Qualitätseinbußen.

Es gibt jedoch ein effizientes Instrument, die Menge der erhobenen Daten zu regulieren und sie nur dann zu erhöhen, wenn es notwendig erscheint: die **Lernkurve**.[446] Der Verantwortliche sollte damit beginnen, eine begrenzte Menge an Trainingsdaten zu erheben und zu verwenden, und dann die Genauigkeit des Modells überwachen, während es mit neuen Daten gefüttert wird. Dies hilft dem Verantworlichen auch, den „Fluch der Dimensionalität" zu vermeiden, d. h. „Algorithmen mit einer schlechten Leistung und hohen Komplexität aufgrund von Datenrahmen mit zahlreichen Dimensionen/Merkmalen, die die Zielfunktion häufig recht komplex machen und zu Modell-Overfitting führen können, solange der Datensatz häufig auf der unteren Dimensionalitätsebene liegt."[447]

---

[442] EDSB (2017) Beurteilung der Erforderlichkeit von Maßnahmen, die das Grundrecht auf Schutz personenbezogener Daten einschränken: Ein Toolkit. Seite 5. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edps.europa.eu/data-protection/our-work/publications/papers/necessity-toolkit_en (abgerufen am 15. Mai 2020); ICO (keine Datum) Lawful basis for processing. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/ (abgerufen am 15. Mai 2020).

[443] Siehe CJEU, Urteil in den verbundene Rechtssachen C-92/09 and C-93/09, Volker und Markus Schecke GbR and Hartmut Eifert gegen Land Hessen, 9. November 2010.

[444] ENISA (2015) Privacy by design in big data. European Union Agency for Cybersecurity, Athen/Heraklion, Seite 23. Verfügbar unter: www.enisa.europa.eu/publications/big-data-protection (abgerufen am 28. Mai 2020).

[445] CIPL (2020) Artificial intelligence and data protection: how the GDPR regulates AI. Centre for Information Policy Leadership, Washington DC / Brüssel / London, Seite 13. Verfügbar unter: www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf (abgerufen am 15. Mai 2020).

[446] Siehe: www.ritchieng.com/machinelearning-learning-curve/ (abgerufen am 28. Mai 2020).

[447] H. Oliinyk (2018), Why and how to get rid of the curse of dimensionality right (with breast cancer dataset visualization). Towards Data Science, 20. März. Verfügbar unter: https://towardsdatascience.com/why-and-

Einige zusätzliche Maßnahmen im Zusammenhang mit dem Grundsatz der Minimierung sind:

- Begrenzung der Ausdehnung der Datenkategorien (z. B. Name, Anschrift und Adresse, Angaben zu Gesundheit, Beschäftigungssituation, Überzeugungen, Ideologie usw.)
- Begrenzung des Grads an Detailliertheit oder Genauigkeit der Daten, der Granularität der Erhebung in Bezug auf Zeit und Häufigkeit sowie des Alters der verwendeten Daten
- Begrenzung der Ausdehnung der Anzahl interessierter Parteien auf diejenigen, die die Daten verarbeiten
- Begrenzung der Zugänglichkeit der verschiedenen Datenkategorien auf das Personal des Verantwortlichen/Verwalters oder sogar den Endnutzer (wenn in den KI-Modellen Daten von Dritten enthalten sind) in allen Verarbeitungsphasen[448]

Natürlich kann die Annahme dieser Maßnahmen unter Umständen mit einem enormen Aufwand für die Vereinheitlichung, Homogenisierung usw. der Daten verbunden sein, wird jedoch dazu beitragen, den Grundsatz der Datenminimierung wesentlich effizienter umzusetzen.[449]

Schließlich ist zu bedenken, dass die Verantwortlichen es **vermeiden sollten, lange Datenbanken mit historischen Daten zu speichern**, z. B. über den Zeitraum hinaus, der für normale Geschäftszwecke oder zur Erfüllung rechtlicher Verpflichtungen erforderlich ist, oder einfach nur, weil ihr Analysewerkzeug in der Lage ist, eine große Anzahl von Daten zu erzeugen, und ihre Speicherkapazität dies ermöglicht. Stattdessen müssen Unternehmen, die Big Data nutzen, angemessene Speicherfristen durchsetzen (siehe Abschnitt „Speicherbegrenzung“ im Kapitel „Grundsätze“).

Kasten 6. Ein Beispiel für die Vorteile der Datenminimierung in der KI

Ein von der norwegischen Steuerverwaltung entwickeltes Tool zum Filtern von Steuererklärungen auf Fehler testete in der Trainingsphase 500 Variablen. In das endgültige KI-Modell wurden jedoch nur 30 aufgenommen, da sie sich für die gestellte Aufgabe als besonders relevant erwiesen hatten. Wahrscheinlich hätten die Entwickler des Tools die Erhebung derart vieler personenbezogener Daten vermeiden können, wenn sie zu Beginn des Entwicklungsprozesses eine bessere Auswahl der relevanten Variablen getroffen hätten.

Quelle: Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

how-to-get-rid-of-the-curse-of-dimensionality-right-with-breast-cancer-dataset-7d528fb5f6c0 (abgerufen am 15. Mai 2020).

[448] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 39–40. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

[449] Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 15. Mai 2020).

Checkliste: Datenminimierung

- Die Verantwortlichen haben sichergestellt, dass sie personenbezogene Daten nur bei Bedarf verwenden.
- Die Verantwortlichen haben die Verhältnismäßigkeit zwischen der Datenmenge und der Genauigkeit des KI-Tools berücksichtigt
- Die Verantwortlichen überprüfen regelmäßig die in ihrem Besitz befindlichen Daten und löschen alle, die nicht benötigt werden.
- Die Verantwortlichen bereinigen in der Trainingsphase des KI-Systems alle Informationen, die für das Training nicht unbedingt erforderlich sind.
- Die Verantwortlichen prüfen, ob personenbezogene Daten in der Verteilungsphase des KI-Systems verarbeitet werden, und löschen diese, sofern keine begründete Notwendigkeit und Rechtmäßigkeit besteht, sie für andere vereinbare Zwecke zu behalten.

Zusätzliche Informationen

ENISA (2015) Privacy by design in big data. European Union Agency for Cybersecurity, Athen/Heraklion, Seite 23. Verfügbar unter: www.enisa.europa.eu/publications/big-data-protection

ICO (kein Datum) Principle (c): data minimisation. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/

Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

Pure Storage (2015) Big data's big failure: the struggles businesses face in accessing the information they need. Pure Storage, Mountain View, CA. Verfügbar unter: http://info.purestorage.com/rs/225-USM-292/images/Big%20Data%27s%20Big%20Failure_UK%281%29.pdf

3.2.4 Treu und Glauben im Hinblick auf die Rechte der betroffenen Personen

Die Verarbeitung nach Treu und Glauben ist ein wesentliches Konzept des Datenschutzes (siehe Unterabschnitt „Verarbeitung nach Treu und Glauben", Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz", Kapitel „Grundsätze" und Kapitel „Rechte der betroffenen Personen"). Es kann nur schwer erreicht werden, wenn man sich nicht bewusst ist, dass die Entwicklung von KI-Tools die Interessen, Rechte und Freiheiten der betroffenen Personen beeinträchtigen kann. Deshalb ist es sinnvoll, für die Umsetzung angemessener Garantien zu sorgen, nicht nur um unfaire Folgen zu vermeiden, sondern auch um den betroffenen Personen durchsetzbare Rechte zu gewähren, die einen angemessenen Schutz vor einer unfairen Verarbeitung gewährleisten.

In diesem Abschnitt wird untersucht, wie die in der DSGVO anerkannten Schlüsselrechte auf den KI-Entwicklungsrahmen anzuwenden sind. Zu diesem Zweck werden wir uns auf einige Rechte konzentrieren, die in diesem Bereich besonders relevant sind: (a) Recht auf Information, (b) Recht auf Auskunft, (c) Recht auf Datenübertragbarkeit, (d) Recht auf Berichtigung, (e) Recht auf Löschung und (f) Recht auf Widerspruch.

Vor der Berücksichtigung dieser Rechte sollten Forscher jedoch prüfen, ob ihre Forschung **als wissenschaftliche Forschung gemäß Artikel 89 DSGVO angesehen werden kann**. Dies ist äußerst wichtig. Denn in diesem Fall können die Rechtsvorschriften der EU oder der Mitgliedstaaten Ausnahmen von den in den Artikeln 15, 16, 18 und 21 genannten Rechten (Auskunft, Berichtigung, Einschränkung und Widerspruch – und indirekt auch Übertragbarkeit) vorsehen. Diese unterliegen den in Artikel 89 Absatz 1 genannten Bedingungen und Garantien, sofern diese Rechte die Verwirklichung der spezifischen Zwecke unmöglich machen oder ernsthaft beeinträchtigen könnten und solche Ausnahmen für die Erfüllung dieser Zwecke erforderlich sind (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“ im Kapitel „Konzepte“).

a) Recht auf Information

Gemäß Artikel 13 DSGVO muss der Verantwortliche den betroffenen Personen vor der Verarbeitung personenbezogener Daten vollständige Informationen über die Verarbeitung und ihre Rechte in einem verständlichen Format zur Verfügung stellen. „Beabsichtigt der Verantwortliche, die personenbezogenen Daten für einen anderen Zweck weiterzuverarbeiten als den, für den die personenbezogenen Daten erhoben wurden, so stellt er der betroffenen Person vor dieser Weiterverarbeitung Informationen über diesen anderen Zweck und alle anderen maßgeblichen Informationen zur Verfügung“ (Artikel 13 Absatz 3).

Die Verantwortlichen sind jedoch davon befreit, den betroffenen Personen Informationen bereitzustellen, wenn sich die Bereitstellung derselben als unmöglich erweist; einen unverhältnismäßigen Aufwand erfordern würde, insbesondere bei Verarbeitungen für im öffentlichen Interesse liegende Archivzwecke, für wissenschaftliche oder historische Forschungszwecke oder für statistische Zwecke; oder wenn die Verpflichtung zur Bereitstellung der Informationen die Verwirklichung der Ziele dieser Verarbeitung voraussichtlich unmöglich machen oder ernsthaft beeinträchtigen würde. In diesen Fällen sollte der Verantwortliche geeignete Maßnahmen zum Schutz der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person, einschließlich der Bereitstellung dieser Informationen für die Öffentlichkeit ergreifen. Diese Ausnahmeregelung ist jedoch an die Annahme der in Artikel 89 vorgesehenen Garantien geknüpft (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“ im Kapitel „Konzepte“).

b) Recht auf Auskunft

Das Recht der betroffenen Personen auf Auskunft über ihre Daten muss in allen Phasen des Lebenszyklus eines KI-Tools gewährleistet sein. Die Verantwortlichen sind angehalten, geeignete technische Maßnahmen zur Gewährleistung zu ergreifen, dass die betroffene Person leicht Zugang zu ihren Daten erhält. Gemäß Artikel 15 DSGVO hat die betroffene Person das Recht, Einzelheiten zu allen personenbezogenen Daten zu erfahren, die für das Profiling verwendet werden, einschließlich der zur Erstellung eines Profils verwendeten Datenkategorien. Darüber hinaus ist der Verantwortliche gemäß Artikel 15 Absatz 3 verpflichtet, die zur Erstellung des Profils verwendeten Daten zur Verfügung zu stellen und Auskunft über das Profil sowie darüber zu erteilen, in welche Segmente die betroffene Person

eingeordnet wurde. Ebenso heißt es in Erwägungsgrund 63 der Datenschutz-Grundverordnung: „Nach Möglichkeit sollte der Verantwortliche den Fernzugang zu einem sicheren System bereitstellen können, der der betroffenen Person direkten Zugang zu ihren personenbezogenen Daten ermöglichen würde“. **Dies schließt beobachtete, aus Rückschlüssen erzeugte und hergeleitete Daten ein**.[450]

Kasten 7. Das Problem der hergeleiteten Daten

Eine der dringlichsten Fragen, mit denen wir im Bereich der KI konfrontiert sind, ist der konkrete Status der hergeleiteten Daten. Dabei handelt es sich um Daten, die nicht von den betroffenen Personen bereitgestellt, sondern diesen aus verfügbaren Daten „zugewiesen“ werden, die entweder von betroffenen oder von anderen Personen stammen. Manchmal liefern diese hergeleiteten Daten Informationen über eine identifizierbare Person. Unabhängig davon, ob diese Informationen genau sind oder nicht, sind diese Daten als personenbezogene Daten zu betrachten und fallen daher unter die Datenschutz-Grundverordnung. Infolgedessen sollten die Rechte der betroffenen Personen strengstens eingehalten werden, einschließlich des Rechts auf Auskunft über diesen Daten.[451] Wie jedoch bereits an anderer Stelle in diesen Leitlinien erörtert wurde, fallen hergeleitete Daten nicht unter das Recht auf Übertragbarkeit (siehe Abschnitt „Recht auf Datenübertragbarkeit“).

Eines der Hauptprobleme bei der Verarbeitung von KI und Big Data ist, dass das Auskunftsrecht zuweilen mit dem Interesse eines Unternehmens an der Wahrung seiner Geschäftsgeheimnisse kollidieren kann. Durch Erwägungsgrund 63 werden Verantwortliche tatsächlich geschützt, die Sorge haben, Geschäftsgeheimnisse oder Rechte des geistigen Eigentums offenzulegen, was bei Profiling besonders relevant sein könnte.[452] Allerdings können KI-Entwickler **den Schutz ihrer Geschäftsgeheimnisse nicht als Rechtfertigung dafür verwenden, der betroffenen Person Auskünft über ihre Daten oder die Übermittlung von Informationen zu verweigern.** Stattdessen müssen die Unternehmen pragmatische Lösungen finden.[453]

Das Recht auf Auskunft kann mehr oder weniger anwendbar sein, je nachdem, in welchem Stadium des Lebenszyklus sich die KI-Entwicklung befindet. So könnte es beispielsweise schwierig sein, einer einzelnen betroffenen Person Auskunft zu Trainingsdaten bereitzustellen, da diese in der Regel nur Informationen enthalten, die für Vorhersagen

---

[450] ICO (2014) Big data and data protection. Information Commissioner’s Office, Wilmslow, Seiten 99–10. Verfügbar unter: https://rm.coe.int/big-data-and-data-protection-ico-information-commissioner-s-office/1680591220 (abgerufen am 28. Mai 2020).

[451] Siehe: B. Custers, (2018), Profiling as inferred data. Amplifier effects and positive feedback loops, Seiten 112–115 in Bayamlıoğlu, E. et al. (eds) *Being profiled: cogitas ergo sum. 10 years of profiling the European Citizen*. Amsterdam University Press, Amsterdam. DOI 10.5117/9789463722124/CH19. Verfügbar unter: https://ssrn.com/abstract=3466857 or http://dx.doi.org/10.2139/ssrn.3466857 (abgerufen am 28. Mai 2020).

[452]Artikel-29-Datenschutzgruppe (2016), Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679. Europäische Kommission, Brüssel, Seite 17. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (abgerufen am 28. Mai 2020).

[453] Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo, Seite 19. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 28. Mai 2020).

relevant sind (z. B. frühere Transaktionen, demografische Daten, Standort), aber keine Kontaktangaben oder eindeutige Kundenkennungen. Darüber hinaus werden sie häufig vorverarbeitet, um sie für die Algorithmen des maschinellen Lernens besser nutzbar zu machen. Dies bedeutet jedoch nicht, dass all diese Daten als anonymisiert betrachtet werden können. Es handelt sich somit weiterhin um personenbezogene Daten. Im Falle eines Modells zur Kaufvorhersage könnte das Training beispielsweise ein für einen Kunden spezifisches Kaufmuster enthalten. Wenn in diesem Beispiel ein Kunde eine Liste seiner letzten Einkäufe im Rahmen seiner Anfrage zur Verfügung stellt, könnte das Unternehmen möglicherweise den Teil der Trainingsdaten identifizieren, der sich auf diese Person bezieht.

Unter diesen Umständen sollten **Verantwortliche auf Antrag der betroffenen Personen in Bezug auf Auskunft über sie betreffende personenbezogenen Daten reagieren, vorausgesetzt, sie haben angemessene Maßnahmen ergriffen, um die Identität der betroffenen Person zu überprüfen, und es gelten keine weiteren Ausnahmen.** Wie das ICO zudem feststellt, „sollten Anträge auf Auskunft, Berichtigung oder Löschung von Trainingsdaten nicht als offensichtlich unbegründet oder übertrieben angesehen werden, nur weil sie möglicherweise schwieriger zu erfüllen sind oder die Motivation der Anfrage im Vergleich zu anderen Auskunftsanfragen, die ein Unternehmen in der Regel erhält, unklar sein könnte.“[454] Allerdings **sind Unternehmen nicht verpflichtet, zusätzliche personenbezogene Daten zwecks Identifizierung der betroffenen Personen in den Trainingsdaten zu erheben oder zu speichern, nur um die Verordnung einzuhalten.** Wenn die Verantwortlichen eine betroffene Person in den Trainingsdaten nicht identifizieren können und die betroffene Person keine zusätzlichen Informationen bereitstellen kann, die ihre Identifizierung ermöglichen würden, sind sie nicht verpflichtet, einem Antrag nachzukommen, der nicht erfüllt werden kann.[455]

Checkliste: Recht auf Auskunft[456]

Vorbereitung auf Auskunftsanfragen betroffener Personen

 Die Verantwortlichen wissen, wie eine Aufkunftsanfrage zu personenbezogenen Daten zu erkennen ist wann wann das Auskunftsrecht anwendbar ist.

 Die Verantwortlichen wissen, dass das Auskunftsrecht in jeder Phase des Lebenszyklus der KI-Lösung anzuwenden ist, wenn diese personenbezogene Daten verwendet.

 Die Verantwortlichen haben eine Richtlinie, wie sie mündliche Anfragen aufzeichnen.

 Die Verantwortlichen wissen, wann sie eine Anfrage ablehnen können, und kennen die Informationen, die sie den Personen in diesem Fall zur Verfügung stellen müssen.

 Die Verantwortlichen kennen die Art der zusätzlichen Informationen, die sie als Antwort auf eine Auskunftsanfrage zu personenbezogenen Daten bereitstellen müssen.

Erfüllung von Auskunftsanfragen betroffener Personen

[454] ICO (2019) Enabling access, erasure, and rectification rights in AI systems. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (abgerufen am 28. Mai 2020).

[455] Ibid.

[456] ICO (kein Datum) Right of access. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/ (abgerufen am 28. Mai 2020).

- Die Verantwortlichen verfügen über Verfahren, die sicherstellen, dass sie eine Auskunftsanfrage zu personenbezogenen Daten ohne unangemessene Verzögerung und innerhalb eines Monats nach Eingang beantworten.
- Die Verantwortlichen sind sich der Umstände bewusst, unter denen sie die Frist für die Beantwortung einer Anfrage verlängern können.
- Die Verantwortlichen wissen, dass es besonders wichtig ist, eine klare und verständliche Sprache zu verwenden, wenn sie Informationen an ein Kind weitergeben.
- Die Verantwortlichen wissen, was sie beachten müssen, wenn eine Anfrage Informationen über andere Personen enthält.
- Die Verantwortlichen wissen, wie sie das Auskunftsrecht in Trainingsphasen anwenden können.

Zusätzliche Informationen

Artikel-29-Datenschutzgruppe (2014), Stellungnahme 06/2014 zum Begriff des berechtigten Interesses des für die Verarbeitung Verantwortlichen gemäß Artikel 7 der Richtlinie 95/46/EG. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

ICO (2013), Big data, artificial intelligence, machine learning and data protection. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

ICO (kein Datum) Right of access. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access/

Norwegische Datenschutzbehörde. (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

c) Recht auf Datenübertragbarkeit

Durch Artikel 20 DSGVO wurde ein neues Recht begründet: das Recht auf Datenübertragbarkeit.[457] Dieses Recht gibt den betroffenen Personen die Kontrolle über die Verwendung ihrer Daten, indem sie diese dorthin weiterleiten, wo sie am nützlichsten sind (siehe Abschnitt „Recht auf Datenübertragbarkeit" im Kapitel „Rechte der betroffenen Personen"). Das Recht auf Datenübertragbarkeit könnte jedoch im Bereich der KI aus mehreren Gründen schwer umzusetzen sein. Man darf die Kosten und die Durchführbarkeit der Bereitstellung extrem großer, komplexer Datensätze nicht vergessen, die sich über viele Jahre angesammelt haben. Dies könnte es einem Unternehmen schwer machen, sein Recht auf Datenübertragbarkeit zu erfüllen.

Es gibt verschiedene Arten von personenbezogenen Daten, die ein System des maschinellen Lernens verarbeiten kann. Laut der Artikel-29-Datenschutzgruppe sind einige

[457] Artikel-29-Datenschutzgruppe (2015), Leitlinien zum Recht auf Datenübertragbarkeit. Europäische Kommission, Brüssel. Verfügbar unter: http://ec.europa.eu/newsroom/document.cfm?doc_id=45685 (abgerufen am 28. Mai 2020).

Datenkategorien mit dem Recht auf Datenübertragbarkeit verbunden, nämlich: personenbezogene Daten, die die betroffene Person betreffen und Daten, die sie einem Verantwortlichen bereitgestellt hat. Im Allgemeinen ist die Formulierung „von der betroffenen Person bereitgestellt“ **weit** auszulegen. Daher schließt sie Daten ein, die das Ergebnis einer Beobachtung des Verhaltens einer Person (z. B. Rohdaten, die von intelligenten Messgeräten verarbeitet werden, Aktivitätsprotokolle oder Website-Historie) sind. Ausgeschlossen werden sollten lediglich „hergeleitete Daten“ und „aus Rückschlüssen erzeugte Daten“, die personenbezogene Daten beinhalten, welche von einem Dienstanbieter erzeugt werden (z. B. algorithmische Ergebnisse). Im Gegensatz zu beobachteten oder erhobenen Daten werden **hergeleitete Daten von dem Dienst selbst auf der Grundlage der beobachteten Daten erstellt und nicht von der betroffenen Person bereitgestellt.**[458] Das Recht auf Datenübertragbarkeit **gilt daher nicht für Daten, die durch ein Verfahren des maschinellen Lernens hergeleitet wurden.**

Checkliste: Datenübertragbarkeit[459]

Vorbereitung auf Anfragen bezüglich der Datenübertragbarkeit

 Die Verantwortlichen wissen, wie sie eine Anfrage auf Datenübertragbarkeit erkennen können und wann dieses Recht anwendbar ist.

 Die Verantwortlichen berücksichtigen das Erfordernis der Datenübertragbarkeit bereits in den frühesten Phasen der Konzeption und Gestaltung der KI-Verarbeitung.

 Die Verantwortlichen haben eine Richtlinie, wie sie mündliche Anfragen aufzeichnen.

 Die Verantwortlichen wissen, wann sie eine Anfrage ablehnen können, und kennen die Informationen, die sie den Personen zur Verfügung stellen müssen, wenn sie eine solche Ablehnung vornehmen.

Erfüllung von Anfragen bezüglich der Datenübertragbarkeit

 Die Verantwortlichen können personenbezogene Daten in strukturierten, gängigen und maschinenlesbaren Formaten übermitteln.

 Die Verantwortlichen informieren die Nutzer im Voraus, wenn es technisch nicht möglich ist, das Recht auf Übertragbarkeit mittels eines Protokolls auszuüben.

 Die Verantwortlichen verwenden sichere Verfahren zur Übermittlung personenbezogener Daten.

 Die Verantwortlichen verfügen über Verfahren, die sicherstellen, dass sie auf Anfrage auf Datenübertragbarkeit ohne unangemessene Verzögerung und innerhalb eines Monats nach Eingang beantworten.

 Die Verantwortlichen sind sich der Umstände bewusst, unter denen sie die Frist für die Beantwortung einer Anfrage verlängern können.

---

[458] Artikel-29-Datenschutzgruppe (2015), Leitlinien zum Recht auf Datenübertragbarkeit. Europäische Kommission, Brüssel, Seite 8. Verfügbar unter: http://ec.europa.eu/newsroom/document.cfm?doc_id=45685 (abgerufen am 28. Mai 2020).

[459] ICO (kein Datum) Right to data portability. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/ (abgerufen am 28. Mai 2020).

Zusätzliche Informationen

Artikel-29-Datenschutzgruppe (2016), Leitlinien zum Recht auf Datenübertragbarkeit. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/information_society/newsroom/image/document/2016-51/wp242_en_40852.pdf

EBF (2017) European Banking Federation's comments to the Working Party 29 guidelines on the right to data portability. European Banking Federation, Brüssel, Seite 4. Verfügbar unter: www.ebf.eu/wp-content/uploads/2017/04/EBF_025448E-EBF-Comments-to-the-WP-29-Guidelines_Right-of-data-portabi.._.pdf

ICO (kein Datum) Right to data portability. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/

N. Wallace und D. Castro (2018), The impact of the EU's new data protection regulation on AI. Center for Data Innovation, Washington DC / Brüssel / London. Verfügbar unter: www2.datainnovation.org/2018-impact-gdpr-ai.pdf

d) Recht auf Berichtigung

Das Recht auf Berichtigung unrichtiger Daten ist im Falle künstlicher Intelligenz besonders wichtig, da Algorithmen für maschinelles Lernen häufig Daten herleiten. Diese Daten können die betroffene Person betreffen, insbesondere wenn sie in fortgeschrittenen Stadien des KI-Lebenszyklus erzeugt werden. Unrichtige Daten, die in der Trainingsphase hergeleitet werden, sind nicht so besorgniserregend wie die in den Endphasen. Da der Zweck von Trainingsdaten darin besteht, Modelle zu trainieren, die auf allgemeinen Mustern in großen Datensätzen beruhen, ist es weniger wahrscheinlich, dass einzelne Ungenauigkeiten direkte Auswirkungen auf eine betroffene Person haben.[460] Wenn beispielsweise personenbezogene Daten, die zur Bereitstellung von Kundeninformationen verwendet werden, nicht korrekt sind (wie etwa eine falsche Telefonnummer in einem Datensatz), könnte die betroffene Person einen größeren Schaden erleiden, als wenn eine hergeleitete Telefonnummer zum Trainieren eines Modells verwendet wird. Dies bedeutet jedoch keineswegs, dass das Recht auf Berichtigung in diesem Stadium nicht gilt.

Manche konkrete Art von Algorithmen wie z. B. Support Vector Machines (SVMs) verwenden einige Schlüsselbeispiele aus den Trainingsdaten, um bei der Anwendung zwischen neuen Beispielen zu unterscheiden. Fordert die betroffene Person die Berichtigung oder Löschung dieser Daten an, wäre dies nicht möglich, ohne das Modell mit den berichtigten Daten neu zu trainieren oder das Modell ganz zu löschen.[461] Dies bedeutet jedoch nicht, dass das Recht auf Berichtigung nicht anwendbar ist.

---

[460] R. Binns (2019), Enabling access, erasure, and rectification rights in AI systems. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (abgerufen am 15. Mai 2020).

[461] Ibid.

Es ist insbesondere zu beachten, dass der Verantwortliche die Daten nicht berichtigen muss, wenn er feststellt, dass die Daten entgegen der Auffassung der betroffenen Person im Hinblick auf die Zwecke der Verarbeitung nicht unrichtig sind.[462] Die Beweislast liegt jedoch bei den Verantwortlichen. Sie müssen einen guten Grund für die Ablehnung der Berichtigung anführen, wobei es nur schwer vorstellbar ist, dass der Schaden, den dies für das KI-System bedeuten könnte, als überzeugender Grund dienen könnte. Der EDSB hat Systeme kritisiert, die nicht die Möglichkeit bieten, einzelne personenbezogene Daten zu berichtigen, ohne das gesamte System erheblich zu schädigen.[463] Entscheidet sich der Verantwortliche dafür, die Anfrage der betroffenen Person abzulehnen, muss er der betroffenen Person in jedem Fall eine Begründung für die Nichtberichtigung der Daten geben. Die betroffene Person kann sich dann auf Wunsch an die Aufsichtsbehörde wenden.[464]

Checkliste: Recht auf Berichtigung[465]

Vorbereitung auf Berichtigungsanfragen

 Die Verantwortlichen wissen, wie sie eine Berichtigungsanfrage erkennen können und wann dieses Recht anwendbar ist.

 Die Verantwortlichen haben eine Richtlinie, wie sie mündliche Anfragen aufzeichnen.

 Die Verantwortlichen wissen, wann sie eine Anfrage ablehnen können, und kennen die Informationen, die sie Einzelpersonen auf Anfrage zur Verfügung stellen müssen.

Erfüllung von Berichtigungsanfragen

 Die Verantwortlichen sind darauf vorbereitet, dem Recht auf Berichtigung der Daten betroffener Personen stattzugeben, insbesondere der Daten, die durch die von der KI-Lösung erstellten Herleitungen und Profile erzeugt wurden.

 Die Verantwortlichen verfügen über Verfahren, die sicherstellen, dass sie Berichtigungsanfragen ohne unangemessene Verzögerung und innerhalb eines Monats nach Eingang beantworten.

 Die Verantwortlichen sind sich der Umstände bewusst, unter denen sie die Frist für die Beantwortung einer Anfrage verlängern können.

 Die Verantwortlichen verfügen über geeignete Systeme, um Informationen zu berichtigen oder zu vervollständigen bzw. eine ergänzende Erklärung abzugeben.

---

[462] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 27. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 28. Mai 2020).

[463] EDSB (2014), Leitlinien zu den Rechten natürlicher Personen in Bezug auf die Verarbeitung personenbezogener Daten. Europäischer Datenschutzbeauftragter, Brüssel, Seite 18. Verfügbar unter: https://edps.europa.eu/sites/edp/files/publication/14-02-25_gl_ds_rights_de.pdf (abgerufen am 10. Mai 2020).

[464] Büro des Datenschutzbeauftragten (Ombudsmann) (kein Datum), Right to Rectification. Büro des Datenschutzbeauftragten (Ombudsmann), Helsinki Verfügbar unter: https://tietosuoja.fi/en/right-to-rectification (abgerufen am 28. Mai 2020).

[465] ICO (kein Datum) Right to rectification. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/ (abgerufen am 28. Mai 2020).

 Die Verantwortlichen verfügen über Verfahren, um alle Empfänger zu informieren, wenn sie Daten, die sie ihnen mitgeteilt haben, berichtigen.

Zusätzliche Informationen

R. Binns (2019), Enabling access, erasure, and rectification rights in AI systems. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/

ICO (kein Datum) Right to rectification. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-rectification/

d) Recht auf Löschung

Die betroffenen Personen haben das ständige Recht, von dem Verantwortlichen die Löschung sie betreffender personenbezogener Daten zu verlangen. Dies kann jedoch in einigen Fällen äußerst kompliziert sein.[466] Denn gelegentlich kann es unmöglich sein, die rechtlichen Ziele des Rechts auf Löschung – auch bekannt als das Recht auf Vergessenwerden – in KI-Umgebungen zu erfüllen, da die Undurchsichtigkeit der Verarbeitung einige personenbezogene Daten für den Verarbeiter verbergen könnte (siehe „

Verständnis von Transparenz und Undurchsichtigkeit").

Das Hauptproblem beim Recht auf Löschung besteht jedoch darin, dass es **ein ganzes KI-System blockieren könnte, das auf der Grundlage der Daten trainiert wurde, deren Löschung eine betroffene Person verlangt.** Einfach ausgedrückt: Algorithmen müssen die für das Training verwendet en Daten behalten. Die Löschung dieser Daten könnte dazu führen, dass die Algorithmen weniger genau sind oder sogar ganz versagen. Daher sollten die Verantwortlichen bedenken, dass die Änderung einer Datenbank, die von einer Datenlöschung ernsthaft betroffen ist, unter Umständen unmöglich sein kann.

Die Verantwortlichen könnten dies für inakzeptabel halten. Tatsache ist jedoch, dass die Datenschutz-Grundverordnung keine Ausnahme vom Recht auf Löschung aufgrund der Schädigung einer Datenbank mit personenbezogenen Daten vorsieht. Einige Autoren wie z. B. Humerick haben in diesem Zusammenhang Folgendes vorgeschlagen: „Anstatt die vollständige Löschung personenbezogener Daten zu fordern, sollten Verantwortliche und Auftragsverarbeiter in der Lage sein, Informationen bis zum Zeitpunkt der Löschung zu bewahren. Auf diese Weise würde das maschinelle Lernen der KI an dem Punkt verbleiben, an dem es fortgeschritten ist, und nicht zu einer erzwungenen Amnesie führen." Seiner Meinung nach könnte dies die Interessen der betroffenen Personen ausreichend schützen, ohne dass die KI einen Datenrückschritt erleidet. Es ist jedoch nicht unbedingt sicher, ob diese Lösung den Anforderungen der Datenschutz-Grundverordnung entspricht.

Jede Empfehlung sollte sich immer auf die ersten Schritte im Lebenszyklus des Produkts konzentrieren. Technisch gesehen ist es schwierig, sichere Lösungen für das Dilemma des Rechts auf Löschung zu finden, nachdem eine Datenbank erstellt wurde. Daher sollten die

[466] E. Fosch-Villaronga, P. Kieseberg und T. Li (2018), Humans forget, machines remember: artificial intelligence and the right to be forgotten, *Computer Law & Security Review* 34(2): 304–313.

Verantwortlichen stets einer einfachen Schlussfolgerung folgen: **Der beste Weg, um katastrophale Schäden zu vermeiden, besteht darin, sich von Anfang an auf einen möglichen Datenverlust vorzubereiten.**

Schließlich muss der Verantwortliche immer die Einschränkungen des Rechts auf Löschung gemäß Artikel 17 Absatz 3 DSGVO im Auge behalten. Darüber hinaus können nationale Behörden zusätzliche Beschränkungen auferlegen, die berücksichtigt werden müssen.

Checkliste: Recht auf Löschung[467]

Vorbereitung auf Löschanfragen

 Die Verantwortlichen wissen, wie sie eine Anfrage auf Löschung erkennen können und wann dieses Recht anwendbar ist.

 Die Verantwortlichen haben eine Richtlinie, wie sie mündliche Anfragen aufzeichnen.

 Die Verantwortlichen wissen, wann sie eine Anfrage ablehnen können, und kennen die Informationen, die sie den Personen in diesem Fall zur Verfügung stellen müssen.

Erfüllung von Löschanfragen

 Die Verantwortlichen verfügen über Verfahren, die sicherstellen, dass sie Löschanfragen ohne unangemessene Verzögerung und innerhalb eines Monats nach Eingang beantworten.

 Die Verantwortlichen sind sich der Umstände bewusst, unter denen sie die Frist für die Beantwortung einer Anfrage verlängern können.

 Die Verantwortlichen wissen, dass das Recht auf Löschung besonders wichtig ist, wenn sich die Anfrage auf Daten bezieht, die von Kindern erhoben wurden.

 Die Verantwortlichen verfügen über Verfahren, um alle Empfänger zu informieren, wenn sie Daten, die sie ihnen mitgeteilt haben, löschen.

 Die Verantwortlichen verfügen über geeignete Verfahren zur Löschung von Informationen.

Zusätzliche Informationen

Ein Interview mit Tiffany Li über das Recht auf Löschung und KI finden Sie hier: www.youtube.com/watch?v=Sozg6yJJkHk

R. Binns (2019), Enabling access, erasure, and rectification rights in AI systems. ICO-Blog, 15. Oktober. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/

E. Fosch-Villaronga, P. Kieseberg und T. Li (2018), Humans forget, machines remember: artificial intelligence and the right to be forgotten, *Computer Law & Security Review* 34(2) 304–313.

M. Humerick (2018), Taking AI personally: how the E.U. must learn to balance the

[467] ICO (kein Datum) Right to erasure. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasurerectification/ (abgerufen am 28. Mai 2020).

interests of personal data privacy & artificial intelligence, 34 Santa Clara High Tech. L.J.393. Verfügbar unter: https://digitalcommons.law.scu.edu/chtlj/vol34/iss4/3

ICO (kein Datum) Right to erasure. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-erasure/

N. Wallace und D. Castro (2018), The impact of the EU's new data protection regulation on AI. Center for Data Innovation, Washington DC / Brüssel / London. Verfügbar unter: www2.datainnovation.org/2018-impact-gdpr-ai.pdf

e) Recht auf Widerspruch

Betroffene Personen haben das Recht, der Verarbeitung sie betreffender personenbezogener Daten zu widersprechen, wenn der Verantwortliche sie auf der Grundlage eines berechtigten Interesses oder für eine im öffentlichen Interesse liegende Aufgabe verarbeitet. Dies gilt nicht für Fälle, in denen die Rechtsgrundlage für die Verarbeitung die Einwilligung in Kenntnis der Sachlage war, da die betroffenen Personen in diesen Fällen einfach ihre Einwilligung widerrufen können und der Verantwortliche ihre Daten nicht mehr verarbeiten darf. Auf Anfrage der betroffenen Personen müssen die Verantwortlichen die Verarbeitung der Daten einstellen, es sei denn, sie können nachweisen, dass sie zwingende und berechtigte Gründe für die Fortsetzung der Verarbeitung haben und dass diese Gründe die Interessen, Rechte und Freiheiten der betroffenen Personen überwiegen.[468]

Sobald die Verantwortlichen einen Widerspruch gegen die Verarbeitung personenbezogener Daten erhalten, müssen sie, sofern keine Gründe für eine Ablehnung vorliegen, die Verarbeitung der Daten unverzüglich einstellen. Dies kann bedeuten, dass sie gespeicherte personenbezogene Daten löschen müssen, da die weit gefasste Definition der Verarbeitung nach der DSGVO auch die Speicherung von Daten umfasst.

Checkliste: Recht auf Widerspruch

Vorbereitung auf Widersprüche gegen die Verarbeitung

 Die Verantwortlichen wissen, wie sie einen Widerspruch erkennen können und wann dieses Recht anwendbar ist.

 Die Verantwortlichen haben eine Richtlinie, wie sie mündliche Widersprüche aufzeichnen.

 Die Verantwortlichen wissen, wann sie einen Widerspruch ablehnen können, und kennen die Informationen, die sie den Personen in diesem Fall zur Verfügung stellen müssen.

 Die Verantwortlichen haben in ihrer Datenschutzerklärung klare Informationen über das Widerspruchsrecht von Personen getrennt von anderen Informationen über deren Rechte angegeben.

 Die Verantwortlichen wissen, wann sie Personen zusätzlich zur Aufnahme in die Datenschutzerklärung über ihr Widerspruchsrecht informieren müssen.

[468] Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo, Seite 29. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 28. Mai 2020).

Erfüllung von Widersprüchen gegen die Verarbeitung

 Die Verantwortlichen verfügen über Verfahren, die sicherstellen, dass sie Widersprüche ohne unangemessene Verzögerung und innerhalb eines Monats nach Eingang beantworten.

 Die Verantwortlichen sind sich der Umstände bewusst, unter denen sie die Frist für die Beantwortung eines Widerspruchs verlängern können.

 Die Verantwortlichen verfügen über geeignete Verfahren, um personenbezogene Daten zu löschen, zu unterdrücken oder anderweitig die Verarbeitung einzustellen.

Zusätzliche Informationen

EDSB (2020) A preliminary opinion on data protection and scientific research. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf

ICO (kein Datum), The right to object to use your data. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/your-data-matters/the-right-to-object-to-the-use-of-your-data/

Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

4 Transparenz

„Diese Anforderung ist eng mit dem Grundsatz der Erklärbarkeit verbunden und bezieht sich auf die Transparenz der für ein KI-System relevanten Komponenten: die Daten, das System und die Geschäftsmodelle."

– *Hochrangige Expertengruppe für künstliche Intelligenz*[469]

4.1 Ethische und rechtliche Fragen

Diese Anforderung umfasst drei unterschiedliche Hauptgrundsätze, und zwar: Rückverfolgbarkeit, Erklärbarkeit und Kommunikation-[470]

Rückverfolgbarkeit

Die Datensätze und Prozesse, die zu der Entscheidung des KI-Systems geführt haben, einschließlich der Datenerfassung und -kennzeichnung sowie der verwendeten Algorithmen, sollten so gut wie möglich dokumentiert werden, um deren Rückverfolgbarkeit sicherzustellen und die Transparenz zu erhöhen. Dies gilt auch für die vom KI-System

[469] Hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 18. Brüssel, Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

[470] Ibid.

getroffenen Entscheidungen. So können die Gründe für eine fehlerhafte KI-Entscheidung ermittelt werden, was wiederum zur Vermeidung zukünftiger Fehler beitragen kann. Rückverfolgbarkeit erleichtert somit die Nachprüfbarkeit und Erklärbarkeit.

Erklärbarkeit

Dies bezieht sich auf die Möglichkeit, sowohl die technischen Prozesse eines KI-Systems als auch die damit verbundenen menschlichen Entscheidungen (z. B. Anwendungsbereiche eines KI-Systems) zu erklären. Technische Erklärbarkeit setzt voraus, dass die von einem KI-System getroffenen Entscheidungen vom Menschen verstanden und rückverfolgt werden können. Darüber hinaus müssen möglicherweise Kompromisse zwischen einer verbesserten Erklärbarkeit eines Systems (was die Präzision beeinträchtigen kann) und mehr Präzision (auf Kosten der Erklärbarkeit) eingegangen werden. Wann immer ein KI-System das Leben von Menschen entscheidend beeinflusst, muss es möglich sein, eine geeignete Erklärung für den Entscheidungsprozess des KI-Systems zu erhalten. Eine solche Erklärung sollte rechtzeitig erfolgen und auf die jeweilige Sachkenntnis des betroffenen Interessenträgers (z. B. Laie, Regulierungsbehörde oder Forscher) zugeschnitten sein. Darüber hinaus sollten Erläuterungen darüber vorliegen, inwieweit ein KI-System die Entscheidungsprozesse einer Organisation beeinflusst und gestaltet, aber auch über die Entwurfsentscheidungen und die Gründe für die Einführung des Systems (zur Gewährleistung der Transparenz des Geschäftsmodells).

Kommunikation

KI-Systeme dürfen gegenüber den Nutzern nicht als Menschen auftreten. Menschen haben das Recht, darüber informiert zu werden, dass sie mit einem KI-System interagieren. Dies bedeutet, dass KI-Systeme als solche erkennbar sein müssen. Zur Gewährleistung der Einhaltung der Grundrechte sollte darüber hinaus bei Bedarf die Möglichkeit bestehen, sich gegen diese Interaktion und zugunsten einer zwischenmenschlichen Interaktion zu entscheiden. Darüber hinaus sollten die Fähigkeiten und Einschränkungen des KI-Systems den Anwendern der KI und den Endnutzern in einer dem jeweiligen Anwendungsfall angemessenen Weise mitgeteilt werden. Das könnte Informationen zur Präzision des KI-Systems sowie seiner Grenzen umfassen.

4.2 Bestimmungen der DSGVO: Transparenz

4.2.1 Verständnis von Transparenz und Undurchsichtigkeit

Im Allgemeinen bedeutet Transparenz, dass die betroffenen Personen klare Informationen über die Datenverarbeitung erhalten (siehe Unterabschnitt „Transparenz“, Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“, Kapitel „Grundsätze“). Sie müssen darüber informiert werden, wie und zu welchen Zwecken sie betreffende Informationen (einschließlich der durch Beobachtung gewonnenen und hergeleiteten Daten) verwendet werden, unabhängig davon, ob diese Informationen von den betroffenen Personen selbst oder von anderen erhoben wurden.[471] Betroffene Personen sollten immer wissen, wie und warum eine KI-gestützte Entscheidung über sie getroffen wurde oder wo ihre personenbezogenen Daten zum Trainieren und Testen eines KI-Systems verwendet wurden. Die Verantwortlichen müssen bedenken, dass Transparenz in solchen Fällen sogar noch wichtiger ist, als wenn sie keine direkte Beziehung zu den betroffenen Personen haben.

[471] Artikel 13 und 14 DSGVO.

Im Allgemeinen muss die Transparenz durch den Einsatz einer Reihe von ergänzenden Instrumenten gewährleistet werden. Die Benennung eines Datenschutzbeauftragten, der dann als einziger Ansprechpartner für Anfragen der betroffenen Personen dient, ist eine ausgezeichnete Option. Die Erstellung angemessener Aufzeichnungen über die Verarbeitung für die Aufsichtsbehörden oder die Durchführung von Datenschutz-Folgenabschätzungen sind ebenfalls sehr empfehlenswerte Maßnahmen zur Förderung der Transparenz. Auch die Durchführung von Analysen zur Bewertung der Wirksamkeit und Zugänglichkeit der den betroffenen Personen zur Verfügung gestellten Informationen trägt zur effizienten Umsetzung dieses Grundsatzes bei.

Die größte Herausforderung bei künstlicher Intelligenz besteht darin, dass sie eine Reihe stark unterschiedlicher Techniken umfasst. Einige sind sehr einfach, sodass es für den Verantwortlichen leicht ist, alle erforderlichen Informationen bereitzustellen. Andere wie z. B. das Deep Learning weisen ernsthafte Probleme in Bezug auf die Transparenz auf. Dies wird oft als „Black Box"-Problem bezeichnet – eine Undurchsichtigkeit im Rahmen der KI, die es schwierig macht, Transparenz zu erreichen. Undurchsichtigkeit ist in der Tat eine der größten Bedrohungen für eine faire KI, da sie dem Erfordernis der Transparenz direkt zuwiderläuft. Es gibt mindestens drei Arten von Undurchsichtigkeit, die der KI mehr oder weniger stark innewohnen: (1) absichtliche Geheimhaltung durch Unternehmen oder Staaten, (2) technischer Analphabetismus und (3) epistemische Intransparenz.

4.2.1.1 Undurchsichtigkeit als absichtliche Geheimhaltung durch Unternehmen oder Staaten

Diese Art der Undurchsichtigkeit kann durch den Schutz von Betriebsgeheimnissen berechtigt sein. Sie kann auch berechtigten Interessen entsprechen, wie z. B. der Wahrung von Wettbewerbsvorteilen, der Erhaltung der Sicherheit des Systems oder der Verhinderung von böswilligen Nutzern, die das System missbrauchen. Sie sollte jedoch mit der Einbeziehung unabhängiger Zertifizierungssysteme vereinbar sein, die in der Lage sind, zu bestätigen, dass das Instrument die Anforderungen der Datenschutz-Grundverordnung erfüllt. In den meisten Fällen sollte es unproblematisch sein, der betroffenen Person die Informationen zu geben, die sie zum Schutz ihrer Interessen benötigt, ohne gleichzeitig Geschäftsgeheimnisse preiszugeben.[472] Dies liegt an der einfachen Tatsache, dass die betroffenen Personen nicht im Detail verstehen müssen, wie das System funktioniert, sondern nur, inwieweit es ihre Interessen, Rechte und Freiheiten beeinträchtigen könnte.

4.2.1.2 Undurchsichtigkeit als technischer Analphabetismus

Diese Art der Undurchsichtigkeit ergibt sich aus den spezifischen Fähigkeiten, die für die Entwicklung und Programmierung der Algorithmen erforderlich sind, sowie aus der Fähigkeit, den Code zu lesen und zu schreiben. Man könnte sagen, dass die in der KI verwendeten Codes für die überwiegende Mehrheit der Bevölkerung ein Rätsel sind, da ihr das entsprechende spezifische Wissen fehlt. Dies sollte jedoch kein Hindernis für die Erfüllung der in der Datenschutz-Grundverordnung festgelegten Informationspflicht sein. Die Fähigkeit, die Computersprache zu verstehen, darf keine Barriere für eine verständliche Erklärung des Zwecks eines KI-Systems sein, und zwar nicht nur für die Betroffenen, die einem Profiling oder einer automatisierten Entscheidung unterworfen werden, sondern für alle anderen.

---

[472] Norwegische Datenschutzbehörde (2018) Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 20. Mai 2020).

4.2.1.3 Epistemische Intransparenz

Diese Undurchsichtigkeit ergibt sich aus den Merkmalen der Algorithmen des maschinellen Lernens und dem Umfang, der für ihre sinnvolle Anwendung erforderlich ist. Sie hängt mit der Tatsache zusammen, dass bestimmte algorithmische Modelle von Menschen nicht interpretiert werden können. Vereinfacht ausgedrückt ist der Übergang zwischen den Eingaben, die das Modell erhält, und den bereitgestellten Ausgaben für den Menschen nicht nachvollziehbar. Auf regulatorischer Ebene gibt es kein Verbot für die Verwendung dieser Art von Modellen. Es empfiehlt sich jedoch, es nach dem Grundsatz der Vorsicht zu verwenden, da die fehlende Interpretierbarkeit die Schwierigkeiten bei der Identifizierung von Verzerrungen im Modell verschärfen könnte, was wiederum zu diskriminierenden Ergebnissen oder falschen oder unechten Korrelationen führen könnte. Natürlich sind nicht alle Modelle des maschinellen Lernens in diesem Sinne undurchsichtig.

4.2.1.4 Vorrang für transparente Werkzeuge

Im Allgemeinen sollten Verantwortliche stets verständlichere Algorithmen gegenüber weniger verständlichen Algorithmen bevorzugen. Zielkonflikte zwischen der **Erklärbarkeit/Transparenz und der besten Leistung des Systems** müssen je nach Nutzungskontext angemessen abgewogen werden. So können beispielsweise im Gesundheitswesen die Genauigkeit und Leistung eines Systems wichtiger sein als seine Erklärbarkeit. Im Polizeiwesen dagegen ist die Erklärbarkeit viel wichtiger, um Verhaltensweisen und Ergebnisse der Strafverfolgung zu rechtfertigen. In anderen Bereichen wie beispielweise der Personalbeschaffung werden sowohl der Genauigkeit als auch der Erklärbarkeit ein ähnlicher Wert beigemessen.[473] Wenn eine Dienstleistung sowohl durch einen leicht verständlichen als auch durch einen undurchsichtigen Algorithmus angeboten werden kann – d. h. wenn es keinen Zielkonflikt zwischen Erklärbarkeit und Leistung gibt – sollte sich der Verantwortliche für den besser interpretierbaren Algorithmus entscheiden.

Wenn die Verantwortlichen keine andere Wahl haben, als ein undurchsichtiges Modell zu verwenden, sollten sie zumindest versuchen, technische Lösungen für den Mangel an Interpretierbarkeit zu finden. Natürlich ist es äußerst schwierig, genau zu messen, inwieweit eine Erweiterung der Erklärbarkeit erreicht wird. Weitere Informationen finden Sie im Abschnitt „Das Recht auf Information über eine automatisierten Entscheidungsfindung und die damit verbundene Logik, das Recht, keiner automatisierten Entscheidungsfindung unterworfen zu werden, und das Recht auf menschliches Eingreifen“. Wenn es für die Verantwortlichen schwierig ist, Erklärungen zu finden, sollten sie externen Rat einholen. Auch die Möglichkeit, unabhängige Audits durchzuführen, kann eine sinnvolle Option sein.

Zusätzliche Informationen

EDSB (2015) Stellungnahme 7/2015. Bewältigung der Herausforderungen in Verbindung mit Big Data: Ein Ruf nach Transparenz, Benutzerkontrolle, eingebautem Datenschutz und Rechenschaftspflicht. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf

Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[473] SHERPA (2019) Guidelines for the ethical use of AI and big data systems. SHERPA-Projekt, Seite 16. Verfügbar unter: www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf (abgerufen am 15. Mai 2020).

ICO (2020) Explaining decisions made with AI. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf

Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

SHERPA (2019), Guidelines for the ethical use of AI and big data systems. SHERPA-Projekt. Verfügbar unter: https://www.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdfwww.project-sherpa.eu/wp-content/uploads/2019/12/use-final.pdf

5 Fairness, Vielfalt und Nichtdiskriminierung

„Zur Schaffung einer vertrauenswürdigen KI müssen Inklusion und Vielfalt während des gesamten Lebenszyklus des KI-Systems garantiert sein. Neben der Berücksichtigung und Einbindung aller betroffenen Interessenträger in den gesamten Prozess setzt dies auch die Sicherstellung eines gleichberechtigten Zugangs durch inklusive Gestaltungsprozesse sowie Gleichbehandlung voraus. Diese Anforderung ist eng mit dem Grundsatz der Fairness verbunden."

– *Hochrangige Expertengruppe für künstliche Intelligenz*[474]

5.1 Ethische Grundsätze

Vermeidung unfairer Verzerrungen

Die von KI-Systemen (sowohl zur Ausbildung als auch für den Einsatz) verwendeten Datensätze können unbeabsichtigte historische Verzerrungen und schlechte Lenkungs- und Kontrollmodelle aufweisen oder unvollständig sein. Die Fortschreibung solcher Verzerrungen könnte (in)direkte Vorurteile und die Diskriminierung bestimmter Gruppen oder Personen zur Folge haben und Vorurteile und Marginalisierung potenziell verschärfen. Schaden kann aber auch aus einer beabsichtigten Ausnutzung von Vorurteilen (der Verbraucher) oder durch unlauteren Wettbewerb entstehen, wie z. B. durch die Vereinheitlichung der Preise durch geheime Absprachen oder einen undurchsichtigen Markt.

Erkennbare diskriminierende Verzerrungen sollten nach Möglichkeit in der Phase der Datenerhebung beseitigt werden. Auch die Art und Weise, wie KI-Systeme entwickelt werden (z. B. wie der Programmcode eines Algorithmus geschrieben wird), kann durch unfaire Einflüsse beeinträchtigt werden. Dem könnte durch die Einführung von Aufsichtsverfahren entgegengewirkt werden, mit deren Hilfe der Zweck, die Einschränkungen, Anforderungen

[474] Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 15 ff. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

und Entscheidungen des Systems klar und transparent analysiert und angegangen werden könnten. Darüber hinaus sollte die Beschäftigung von Personen mit unterschiedlichen kulturellen Hintergründen, die aus verschiedenen Fachrichtungen stammen, gefördert werden, da dies zur Gewährleistung der Meinungsvielfalt beitragen könnte.

Barrierefreiheit und universeller Entwurf

Die Systeme sollten benutzerorientiert und so gestaltet sein, dass alle Menschen unabhängig von ihrem Alter, Geschlecht, ihren Fähigkeiten oder Merkmalen KI-Produkte oder -dienstleistungen nutzen können. Die barrierefreie Zugänglichkeit dieser Technologie für Kinder oder Menschen mit Behinderungen, die in allen gesellschaftlichen Gruppen präsent sind, ist von besonderer Bedeutung. KI-Systeme sollten keinen Einheitsansatz verfolgen und die Grundsätze eines universellen Entwurfs44 sollten einen möglichst breiten Nutzerkreis ansprechen und sich an die einschlägigen Barrierefreiheitsnormen halten. So wird ein gerechter Zugang und eine aktive Beteiligung aller Menschen an bestehenden und neu hinzukommenden computergestützten menschlichen Tätigkeiten, auch im Hinblick auf assistierende Technologien, möglich sein.

Beteiligung der Interessenträger

Zur Entwicklung vertrauenswürdiger KI-Systeme ist eine Konsultation der Interessenträger ratsam, die möglicherweise während des gesamten Lebenszyklus des Systems direkt oder indirekt von diesem betroffen sind. Es ist von Vorteil, auch nach der Einführung eines Systems regelmäßige Rückmeldungen einzuholen und längerfristige Vorkehrungen zur Beteiligung der Interessenträger zu schaffen, beispielsweise durch die Gewährleistung der Schulung, Anhörung und Beteiligung der Arbeitnehmerinnen und Arbeitnehmer während des gesamten Prozesses der Implementierung von KI-Systemen in einem Unternehmen.

5.2 Bestimmungen der DSGVO

Laut Erwägungsgrund 71 DSGVO sollte der Verantwortliche „geeignete mathematische oder statistische Verfahren für das Profiling verwenden, technische und organisatorische Maßnahmen treffen, mit denen in geeigneter Weise insbesondere sichergestellt wird, dass Faktoren, die zu unrichtigen personenbezogenen Daten führen, korrigiert werden und das Risiko von Fehlern minimiert wird, und personenbezogene Daten in einer Weise sichern, dass den potenziellen Bedrohungen für die Interessen und Rechte der betroffenen Person Rechnung getragen wird und unter anderem verhindern, dass es gegenüber natürlichen Personen aufgrund von Rasse, ethnischer Herkunft, politischer Meinung, Religion oder Weltanschauung, Gewerkschaftszugehörigkeit, genetischer Anlagen oder Gesundheitszustand sowie sexueller Orientierung zu diskriminierenden Wirkungen oder zu einer Verarbeitung kommt, die eine solche Wirkung hat.“

Dieser Absatz steht in engem Zusammenhang mit **Artikel 21 der Charta der Grundrechte der Europäischen Union**, in dem es heißt: „Diskriminierungen, insbesondere wegen des Geschlechts, der Rasse, der Hautfarbe, der ethnischen oder sozialen Herkunft, der genetischen Merkmale, der Sprache, der Religion oder der Weltanschauung, der politischen oder sonstigen Anschauung, der Zugehörigkeit zu einer nationalen Minderheit, des Vermögens, der Geburt, einer Behinderung, des Alters oder der sexuellen Ausrichtung, sind verboten.“ Der Europäische Datenschutzbeauftragte definiert den Begriff der Verarbeitung nach Treu und Glauben in seinen Leitlinien zum Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen wie folgt: „Die Verarbeitung nach Treu und Glauben ist ein übergeordneter Grundsatz, nach dem personenbezogene Daten nicht auf eine

Weise verarbeitet werden dürfen, die für die betroffene Person in nicht gerechtfertigter Weise schädlich, widerrechtlich diskriminierend, unerwartet oder irreführend ist.“[475]

Die Diskriminierung ist daher ein dramatischer Verstoß gegen den Grundsatz der Verarbeitung nach Treu und Glauben. Im Bereich der künstlichen Intelligenz stellen Verzerrungen jedoch eine enorme Bedrohung für diesen Grundsatz dar, da sie zu einer potenziellen Stigmatisierung oder Diskriminierung einzelner Personen oder ganzer Gemeinschaften führen können.[476]

5.2.1.1 Verzerrungen: die Ursachen

Verzerrungen können verschiedene Ursachen haben. Bei der Erfassung der Daten können **sozial bedingte Verzerrungen, Ungenauigkeiten, Fehler und andere Mängel auftreten.** Es gibt mehrere Gründe, die diese Verzerrungen erklären. Manchmal kann es auch vorkommen, dass Datensätze durch **böswillige Handlungen verzerrt werden.** Die Eingabe schädlicher Daten in ein KI-System kann sein Verhalten verändern, insbesondere bei selbstlernenden Systemen.[477] Im Fall des Chatbots Tay von Microsoft beispielsweise begann eine große Zahl von Internetnutzern rassistische und sexistische Kommentare zu posten, die dem Algorithmus als Nahrung dienten. Das Ergebnis war, dass Tay nach nur wenigen Stunden Betrieb anfing, ebenfalls rassistische und sexistische Tweets zu senden. In anderen Fällen **sind die Daten einfach von schlechter Qualität**, was zu Verzerrungen führt. So bergen beispielsweise Daten von Social-Media-Plattformen aufgrund der Merkmale der Online-Umgebung, die die Genauigkeit und Repräsentativität der Daten nicht garantieren, ernsthafte Risiken für Forscher.

Ein weiterer Grund für Verzerrungen sind **unausgewogene Trainingsdaten** (siehe Kasten 8), die entstehen, wenn der Anteil der verschiedenen Kategorien in den Trainingsdaten nicht ausgewogen ist. Im Zusammenhang mit klinischen Versuchen könnten beispielsweise viel mehr Daten von Männern als von Frauen vorhanden sein. In solchen Fällen ist es wahrscheinlich, dass Frauen durch das resultierende KI-Modell diskriminiert werden. Daher werfen Fragen im Zusammenhang mit der Zusammensetzung der für das Training verwendeten Datenbanken entscheidende ethische und rechtliche Fragen auf, nicht nur Fragen in Bezug auf die Effizienz oder technische Natur.

Kasten 8. Verzerrungen durch unausgewogenes Datentraining

Der Fall Beauty.AI

Das 2016 eingeführte Tool Beauty.AI wurde entwickelt, um mithilfe von Algorithmen zur Alters- und Gesichtserkennung die „erste Schönheitskönigin oder den ersten Schönheitskönig zu wählen, die bzw. der von Robotern beurteilt wird“. Siebentausend Menschen schickten ihre Bilder über eine App ein, aber die meisten der 44 Gewinner

[475] EDSA (2019) Leitlinien 4/2019 zu Artikel 25 Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen (Version für die öffentliche Konsultation). Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2019/guidelines-42019-article-25-data-protection-design_es (abgerufen am 20. Mai 2020).

[476] B. Mittelstadt und L. Floridi (2016), The ethics of big data: current and foreseeable issues in biomedical context, *Science and Engineering Ethics* 22(2): 303–341.

[477] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI. Europäische Kommission, Brüssel, Seite 17. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 20. Mai 2020).

waren weiß – nur eine Handvoll war asiatisch und nur eine Person hatte dunkle Haut. Und das, obwohl die Mehrheit der Teilnehmer zwar weiß war, aber auch viele farbige Menschen Fotos einreichten, darunter große Gruppen aus Afrika und Indien. Dies wurde sofort als rassistisches Ergebnis gewertet, das auf eine schlechte Auswahl des Trainingsdatensatzes zurückzuführen war. Das Hauptproblem bestand darin, dass die Daten, die das Projekt zur Festlegung von Schönheitsstandards verwendet hatte, hauptsächlich von weißen Menschen stammten. Obwohl die Entwickler den Algorithmus nicht so konzipiert hatten, dass helle Haut ein Zeichen von Schönheit ist, verleiteten die Eingabedaten die Roboterrichter tatsächlich zu diesem Schluss.[478]

Das Amazon-Rekrutierungstool

Im Dezember 2018 rangierte Amazon sein KI-Rekrutierungstool aus, als das Unternehmen entdeckte, dass das KI-System Frauen benachteiligte. Amazon hatte seit 2014 Computerprogramme zur Prüfung der Lebensläufe von Bewerbern mit dem Ziel entwickelt, die Suche nach Top-Talenten zu mechanisieren. Das Tool nutzte KI, um Bewerbern eine Bewertung von einem bis fünf Sternen zu geben. 2015 entdeckte Amazon jedoch, dass das Tool Bewerber für Softwareentwickler-Jobs und andere technische Stellen nicht geschlechtsneutral bewertete. Dies lag daran, dass Amazon die Computermodelle darauf trainiert hatte, Bewerber zu überprüfen, indem sie Muster in Lebensläufen beobachteten, die dem Unternehmen über einen Zeitraum von 10 Jahren vorgelegt wurden. Die meisten stammten von Männern, was die männliche Dominanz in der Tech-Branche widerspiegelt.[479]

Drittens können die Trainingsdaten **Diskriminierungen aus der Vergangenheit widerspiegeln, die durch gesellschaftliche Trends verursacht wurden** (siehe Kasten 9). Wenn Verantwortliche historische Daten verwenden, sollten sie sich der wahrscheinlichen Unterschiede zwischen den sozialen Kontexten im Vergleich zur Gegenwart bewusst sein. Andernfalls sind Verzerrungen unvermeidlich. Manchmal ergeben sich die Verzerrungen aus den unterschiedlichen sozialen Kontexten der Gemeinschaft, die die Daten geliefert hat, gegenüber der Gemeinschaft, die den Algorithmus verwenden soll. Wenn der Verantwortliche nicht sorgfältig darauf achtet, wird das Tool wahrscheinlich ebenfalls Verzerrungen enthalten.

Kasten 10. Verzerrungen aufgrund gesellschaftlicher Trends

In der Vergangenheit wurden Kreditanträge von Frauen aufgrund von Vorurteilen häufiger abgelehnt als solche von Männern. In diesem Fall wird jedes KI-Modell, das auf historischen Daten trainiert wurde, voraussichtlich das gleiche Diskriminierungsmuster reproduzieren. Diese Probleme können auch dann auftreten, wenn die Trainingsdaten keine geschützten Merkmale wie Geschlecht oder ethnische Zugehörigkeit enthalten. Eine Vielzahl von Merkmalen in den Trainingsdaten sind oft eng mit geschützten Merkmalen verbunden (z. B. Beruf, ethnische Zugehörigkeit usw.). Aufgrund dieser „Näherungsvariablen“ kann Modell mit diesen Merkmalen verbundene

[478] S. LEVIN (2016), A BEAUTY CONTEST WAS JUDGED BY AI AND THE ROBOTS DIDN'T LIKE DARK SKIN, *THE GUARDIAN*, 8. SEPTEMBER. VERFÜGBAR UNTER: WWW.THEGUARDIAN.COM/TECHNOLOGY/2016/SEP/08/ARTIFICIAL-INTELLIGENCE-BEAUTY-CONTEST-DOESNT-LIKE-BLACK-PEOPLE (ABGERUFEN AM 26. MAI 2020).

[479] J. Dastin J. (2018), Amazon scraps secret AI recruiting tool that showed bias against women, *Reuters*, 10. Oktober. Verfügbar unter: www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Diskriminierungsmuster reproduzieren, selbst wenn die Entwickler dies nicht beabsichtigt haben.

Diese Probleme können bei jedem statistischen Modell auftreten. Sie sind jedoch bei KI-Systemen wahrscheinlicher, da sie eine größere Anzahl von Merkmalen umfassen und komplexe Kombinationen von Merkmalen identifizieren können, die stellvertretend für geschützte Merkmale stehen. Viele moderne Methoden des maschinellen Lernens sind leistungsfähiger als traditionelle statistische Ansätze, weil sie besser in der Lage sind, nichtlineare Muster in hochdimensionalen Daten aufzudecken. Dazu gehören jedoch auch Muster, die Diskriminierung widerspiegeln.[480]

Schließlich können Verzerrungen **durch ein schlecht konzipiertes KI-Tool verursacht werden** (siehe Kasten 11). Es kann vorkommen, dass der Entwickler Korrelationen durch Proxys einführt, die nur schlecht mit der Realität übereinstimmen. In diesem Fall macht das Modell ungenaue Vorhersagen, da seine konzeptionelle Grundlage nicht solide ist.

Kasten 11. Verzerrung durch ein schlecht konzipiertes KI-Werkzeug: Algorithmen

Das US-Gesundheitssystem verwendet kommerzielle Algorithmen, um Gesundheitsentscheidungen zu treffen. Obermeyer et al. [481]fanden Hinweise auf rassistische Verzerrungen in einem weit verbreiteten Algorithmus. Bei schwarzen und weißen Patienten, denen der Algorithmus dasselbe Risiko zuwies, waren die schwarzen Patienten tatsächlich kränker waren als die weißen. Die Autoren schätzten, dass aufgrund dieser rassistischen Verzerrung die Zahl der schwarzen Patienten, die für eine zusätzliche Behandlung ausgewählt wurden, um mehr als die Hälfte zurückging. Die Verzerrung entstand, weil der Algorithmus die Gesundheitskosten als Indikator für den Gesundheitsbedarf verwendete. Für schwarze Patienten wurde bei gleichem Bedarf weniger Geld ausgegeben als für weiße Patienten, sodass der Algorithmus fälschlicherweise zu dem Schluss kam, schwarze Patienten seien gesünder als gleich kranke weiße Patienten. In Wirklichkeit wurden die geringeren Ausgaben durch eine Reihe von rassisch bedingten Faktoren verursacht, darunter der unterschiedliche Zugang zur Behandlung, ein unterschiedliches Maß an Vertrauen in das System, Ungleichgewichte, die von den Leistungserbringern verursacht werden, usw.

5.2.1.2 Beseitigung von Verzerrungen

Es gibt verschiedene Strategien, die dazu beitragen können, Verzerrungen zu vermeiden oder zu korrigieren. Bei der Erstellung der Datenbanken zum Aufbau eines KI-Modells sollten die Verantwortlichen **alle Anstrengungen unternehmen, um unausgewogene oder falsche Daten zu vermeiden.** Erkennbare und diskriminierende Verzerrungen sollten nach Möglichkeit bereits in der Phase der Datensatzerstellung eliminiert werden.[482]

---

[480] ICO (2020) AI auditing framework: draft guidance for consultation, Seite 54. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 26. Mai 2020).

[481] Z. Obermeyer et al. (2019), Dissecting racial bias in an algorithm used to manage the health of populations, *Science,* 25. Oktober, 447–453.

[482] Erwägungsgrund 71 DSGVO.

Wenn der Ursprung der Verzerrung mit dem Trainingsdatensatz zusammenhängt, sollte der Verantwortliche nach einer **angemessenen Auswahl von Daten suchen, die in der Trainingsphase verwendet werden**, um zu vermeiden, dass die Ergebnisse des nachfolgenden Modells falsch oder diskriminierend sind.[483] Ein KI-Modell muss „**anhand relevanter und korrekter Daten trainiert werden und lernen, welche Daten es betonen sollte.** Das Modell darf keine Informationen über ethnische Herkunft, politische Meinung, Religion oder Weltanschauung, Gewerkschaftszugehörigkeit, genetischen Status, Gesundheitszustand oder sexuelle Orientierung betonen, wenn dies zu einer willkürlichen diskriminierenden Behandlung führen würde (Hervorhebung hinzugefügt).“[484]

Darüber hinaus sollten Menschen mit Behinderungen bei der Beschaffung von Daten zur Erstellung von Modellen und bei Tests einbezogen werden, um ein integrativeres und robusteres System zu schaffen. Wenn dieser Prozess angemessen durchgeführt wird, wird die Verzerrung wahrscheinlich verschwinden. In der Fallstudie zu rassistischen Verzerrungen in Gesundheitsalgorithmen (siehe Kasten 11) war es beispielsweise möglich, den Algorithmus neu zu formulieren (in diesem Fall so, dass er nicht mehr die Kosten als Proxy für den Bedarf verwendete) und rassistische Verzerrungen bei der Vorhersage von Pflegebedürftigen zu beseitigen. Durch die Änderung des Gesundheitsindikators von den prognostizierten Kosten auf die Anzahl der chronischen Erkrankungen stieg der Anteil der schwarzen Patienten, die eine bessere Gesundheitsversorgung erhielten, von 17 % auf 46 %. Dies ist ein hervorragendes Beispiel für mehr Fairness durch die Neuformulierung eines Algorithmus.

Die Verantwortlichen sollten jedoch stets bedenken, dass die Bekämpfung von Verzerrungen so komplex ist, weil die Auswahl eines Datensatzes Entscheidungen und Auswahlmöglichkeiten beinhaltet, die manchmal fast **unbewusst getroffen werden können. Im Gegensatz dazu ist die Programmierung eines herkömmlichen, deterministischen Algorithmus immer ein bewusster Vorgang. In der Tat ist der Mensch immer die Intelligenz hinter einer Entwicklung – selbst dann, wenn sie in Algorithmen eingebettet ist, die wir für neutral halten. Derjenige, der einen Datensatz aufbaut, baut ihn bis zu einem gewissen Grad nach seinem eigenen Bild auf, um seine eigene Weltsicht, seine Werte oder zumindest die Werte widerzuspiegeln, die mehr oder weniger in den aus der Vergangenheit gesammelten Daten enthalten sind.**[485]

Vor diesem Hintergrund ist es wichtig, dass die Teams, die für die Auswahl der in einen Datensatz zu integrierenden Daten zuständig sind, aus Personen bestehen, die die Vielfalt widerspiegeln, die von der KI-Entwicklung erwartet wird. Dies stellt derzeit eine große Herausforderung dar. Was das Geschlecht anbelangt, so machen Frauen beispielsweise nur 15 % des KI-Forschungspersonals bei Facebook und 10 % bei Google aus. Darüber hinaus gibt es keine öffentlichen Daten über transsexuelle Arbeitnehmer oder andere geschlechtsspezifische Minderheiten. In Bezug auf die ethnische Zugehörigkeit ist das Gefälle sogar noch extremer: Nur 2,5 % der Google-Mitarbeiter sind schwarz, während Facebook und

---

[483] Für eine Definition von unmittelbarer und mittelbarer Diskriminierung siehe beispielsweise Artikel 2 der Richtlinie 2000/78/EG des Rates vom 27. November 2000 zur Festlegung eines allgemeinen Rahmens für die Verwirklichung der Gleichbehandlung in Beschäftigung und Beruf. Siehe auch Artikel 21 der Charta der Grundrechte der Europäischen Union.

[484] Norwegische Datenschutzbehörde (2018) Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo, Seite 16. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 15. Mai 2020).

[485] CNIL (2017), How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence. Commission Nationale de l'Informatique et des Libertés, Paris, Seite 34. Verfügbar unter: www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (abgerufen am 15. Mai 2020).

Microsoft jeweils 4 % aufweisen.[486] Verantwortliche sollten alle Anstrengungen zur Gewährleistung unternehmen, **dass ihre Teams die Vielfalt besser widerspiegeln und genaue Daten implementieren, die dies widerspiegeln.**

Zusammenfassend lässt sich sagen, **dass die Entwicklungsprozesse von Algorithmen immer eine sorgfältige Überwachung möglicher Verzerrungen beinhalten sollten.** Interne und externe Überprüfungen sollten dieser Frage besondere Aufmerksamkeit widmen. Datensätze, die zu Validierungszwecken erstellt werden, sollten sorgfältig ausgewählt werden, um eine angemessene Einbeziehung von Daten zu gewährleisten, die sich auf Personen aus verschiedenen Bereichen der Gesellschaft beziehen, z. B. in Bezug auf Alter, ethnische Zugehörigkeit, Geschlecht, Behinderung usw. Glücklicherweise gibt es viele technische Hilfsmittel, die sich mit der Beseitigung von Verzerrungen in KI-Modellen befassen.[487] Der Norm P7003TM „Algorithmic Bias Considerations" der IEEE ist im Moment besonders interessant.[488]

Keine diese Optionen bietet jedoch eine magische Lösung oder ein „Allheilmittel", das für alle Arten von Algorithmen gilt. In den meisten Fällen hängt die richtige Lösung von den zahlreichen Variablen ab, die in den Algorithmus einfließen. Verantwortliche sollten bestrebt sein, Verzerrungen so weit wie möglich zu eliminieren, und aufrichtig in Bezug auf die Endergebnisse ihrer Bemühungen sein. Wenn Verzerrungen aufgedeckt werden, sollte die KI-Lösung erneut trainiert werden. Wenn **unfaire Verzerrungen nicht aus dem Modell entfernt werden können, sollte es nicht eingesetzt werden.**

Checkliste: Verzerrungen

- Der Verantwortliche hat eine Strategie oder eine Reihe von Verfahren festgelegt, um zu vermeiden, dass im KI-System unfaire Verzerrungen entstehen oder verstärkt werden, und zwar sowohl bei der Verwendung von Eingabedaten als auch bei der Gestaltung des Algorithmus.
- Der Verantwortliche bewertet und erkennt die möglichen Einschränkungen an, die sich aus der Zusammensetzung der verwendeten Datensätze ergeben.
- Der Verantwortliche hat die Vielfalt und Repräsentativität der verwendeten Daten berücksichtigt.
- Der Verantwortliche hat in Bezug auf spezielle Populationen oder problematische Anwendungsfälle Tests durchgeführt.
- Der Verantwortliche hat die verfügbaren technischen Hilfsmittel genutzt, um sein Verständnis für die Daten, das Modell und die Leistung zu verbessern.
- Der Verantwortliche hat Verfahren zur Prüfung und Überwachung potenzieller Verzerrungen während der Entwicklungs-, Einführungs- und Nutzungsphase des KI-Systems vorgesehen.

---

[486] S. M. West, M., Whittaker und K. Crawford (2019), Discriminating systems: gender, race and power in AI. AI Now Institute, New York, Seite 3. Verfügbar unter: https://ainowinstitute.org/discriminatingsystems.html (abgerufen am 15. Mai 2020).

[487] ICO (2020) AI auditing framework: draft guidance for consultation. Information Commissioner's Office, Wilmslow, Seiten 55–56. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[488] Siehe: https://ethicsinaction.ieee.org/ (abgerufen am 17. Mai 2020).

 Der Verantwortliche hat einen Mechanismus eingeführt, der es anderen ermöglicht, Probleme im Zusammenhang mit Verzerrungen, Diskriminierung oder schlechter Leistung des KI-Systems zu melden.

 Der Verantwortliche hat klare Schritte und Kommunikationswege festgelegt, wie und wem gegenüber solche Probleme angesprochen werden können.

 Der Verantwortliche hat neben den (End-)Nutzern auch andere Personen berücksichtigt, die möglicherweise indirekt von dem KI-System betroffen sind.

 Der Verantwortliche hat geprüft, ob es eine mögliche Entscheidungsvariabilität gibt, die unter den gleichen Bedingungen auftreten kann.

 Für den Fall einer Variabilität hat der Verantwortliche einen Mess- oder Bewertungsmechanismus für die potenziellen Auswirkungen einer solchen Variabilität auf die Grundrechte eingerichtet.

 Der Verantwortliche hat eine quantitative Analyse oder Metrik zur Messung und Prüfung der angewandten Definition von Verarbeitung nach Treu und Glauben eingeführt.

 Der Verantwortliche hat Mechanismen zur Gewährleistung einer Verarbeitung nach Treu und Glauben in den KI-Systemen eingerichtet und andere mögliche Mechanismen in Betracht gezogen.

Zusätzliche Informationen

CNIL (2017) How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence. Commission Nationale de l'Informatique et des Libertés, Paris. Verfügbar unter: www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

EDSA (2019) Leitlinien 4/2019 zu Artikel 25 Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen (Version für die öffentliche Konsultation). Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2019/guidelines-42019-article-25-data-protection-design_es

ICO (2020) AI auditing framework: draft guidance for consultation. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf

B. Mittelstadt und L. Floridi (2016), The ethics of big data: current and foreseeable issues in biomedical context, *Science and Engineering Ethics* 22(2): 303–341.

Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf

S. M. West, M., Whittaker und K. Crawford (2019), Discriminating systems: gender, race and power in AI. AI Now Institute, New York, Seite 3. Verfügbar unter: https://ainowinstitute.org/discriminatingsystems.html

6. Gesellschaftliches und ökologisches Wohlergehen

„Im Einklang mit den Grundsätzen der Fairness und Schadensverhütung sollten während des gesamten KI-Lebenszyklus auch die breitere Gesellschaft, andere fühlende Wesen und die Umwelt als Akteure berücksichtigt werden. Die Nachhaltigkeit und ökologische Verantwortung der KI-Systeme sollten gefördert werden und die Erforschung von KI-Lösungen in Bezug auf globale Belange, wie z. B. die Ziele für eine nachhaltige Entwicklung, sollte ausgebaut werden. Im Idealfall sollte die KI zum Wohle aller Menschen, auch künftiger Generationen, eingesetzt werden."

– *Hochrangige Expertengruppe für künstliche Intelligenz*[489]

5.3 Ethische Grundsätze

Nachhaltige und umweltfreundliche KI

KI-Systeme versprechen, einen Beitrag zur Bewältigung einiger der drängendsten gesellschaftlichen Probleme zu leisten, doch muss sichergestellt sein, dass dies auf möglichst umweltfreundliche Art und Weise geschieht. Der Prozess der Entwicklung, Einführung und Nutzung des Systems sowie dessen gesamte Lieferkette sollten diesbezüglich bewertet werden, z. B. anhand einer kritischen Untersuchung des Verbrauchs von Ressourcen und Energie während des Trainings, wobei weniger umweltschädliche Optionen gewählt werden sollten. Maßnahmen zur Sicherung der Umweltverträglichkeit der gesamten Lieferkette des KI-Systems sollten gefördert werden.

Soziale Auswirkungen

Die Omnipräsenz sozialer KI-Systeme in allen Lebensbereichen (ob in Bildung, Arbeit, Pflege oder Unterhaltung) kann unsere Vorstellung von sozialer Handlungsfähigkeit verändern oder unsere sozialen Beziehungen und Bindungen beeinflussen. So wie KI-Systeme zur Verbesserung sozialer Kompetenzen eingesetzt werden können, können sie auch zu deren Verschlechterung beitragen. Dies könnte sich ebenso auf das körperliche und geistige Wohlergehen der Menschen auswirken. Die Auswirkungen dieser Systeme müssen daher sorgfältig überwacht und berücksichtigt werden.

Gesellschaft und Demokratie

Neben der Bewertung der Auswirkungen der Entwicklung, Einführung und Nutzung eines KI-Systems auf den Einzelnen sollten die Systeme auch aus gesellschaftlicher Sicht unter Berücksichtigung der Auswirkungen auf Institutionen, Demokratie und die Gesellschaft insgesamt bewertet werden. Der Einsatz von KI-Systemen sollte insbesondere im Zusammenhang mit dem demokratischen Prozess – einschließlich der politischen Entscheidungsfindung und in Bezug auf Wahlen – sorgfältig geprüft werden.

5.4 Bestimmungen der DSGVO: Legitimität

Die DSGVO enthält keine spezifischen Bestimmungen über das gesellschaftliche und ökologische Wohlbefinden. Gemäß Artikel 5 Absatz 1 Buchstabe b DSGVO müssen personenbezogene Daten „für festgelegte, eindeutige und legitime Zwecke erhoben werden".

---

[489] Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 19. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

Mit dieser Klausel führt die DSGVO das Konzept der Legitimität in den Datenschutzkontext ein.

Legitimität ist jedoch ein unscharfer Begriff, der in der DSGVO überhaupt nicht definiert ist (siehe Unterabschnitt „Legitimität und Rechtmäßigkeit“, Abschnitt „Rechtmäßigkeit, Fairness und Transparenz“, Kapitel „Grundsätze“). Laut der Artikel 29-Datenschutzgruppe beinhaltet die Legitimität jedoch, dass die Daten „im Einklang mit dem Gesetz“ verarbeitet werden müssen, wobei „Gesetz“ als weit gefasster Begriff verstanden werden sollte, der „alle Formen des geschriebenen Rechts und des Gewohnheitsrechts, primäre und sekundäre Rechtsvorschriften, kommunale Verordnungen, gerichtliche Präzedenzfälle, Verfassungsgrundsätze, Grundrechte, andere Rechtsgrundsätze sowie die Rechtsprechung umfasst, d. h. wie ein solches „Gesetz“ von einem zuständigen Gericht ausgelegt und berücksichtigt würde.“ Es muss daher als ein sehr weit gefasster Begriff verstanden werden, der Überlegungen zum sozialen und ökologischen Wohlergehen einbezieht.

Im Weißbuch „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“ stellen die Autoren fest: „Angesichts der zunehmenden Bedeutung der KI muss den Umweltauswirkungen von KI-Systemen von Anfang bis Ende ihres Lebenszyklus und entlang der gesamten Lieferkette gebührend Rechnung getragen werden, beispielsweise hinsichtlich des Ressourcenverbrauchs beim Trainieren von Algorithmen und der Speicherung von Daten.“[490]

Weitere konkrete Empfehlungen für die Entwicklung künstlicher Intelligenz, die sich am gesellschaftlichen und ökologischen Wohlergehen orientiert, finden sich im „Bericht der Kommission an das Europäische Parlament, den Rat und den Europäischen Wirtschafts- und Sozialausschuss: Bericht über die Auswirkungen von künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung.“[491] Diese Art von ethischen Empfehlungen sollten von KI-Entwicklern sorgfältig geprüft werden, bevor sie personenbezogene Daten verarbeiten, da sie eindeutig mit ihrer Legitimität verbunden sind.

6 Rechenschaftspflicht

„Mit der Rechenschaftspflicht werden die oben genannten Anforderungen, die eng mit dem Grundsatz der Fairness verbunden sind, ergänzt. Zu diesem Zweck müssen Vorkehrungen getroffen werden, die die Verantwortlichkeit und Rechenschaftspflicht für KI-Systeme und deren Ergebnisse vor und nach deren Umsetzung gewährleisten.“

– Hochrangige Expertengruppe für künstliche Intelligenz[492]

---

[490] Europäische Kommission (2020), Weißbuch, Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen. Europäische Kommission, Brüssel, Seite 3. Verfügbar unter: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf (abgerufen am 26 Mai 2020).

[491] Europäische Kommission (2020), Bericht der Kommission an das Europäische Parlament, den Rat und den Europäischen Wirtschafts- und Sozialausschuss: Bericht über die Auswirkungen von künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/info/sites/default/files/report-safety-liability-artificial-intelligence-feb2020_de.pdf (abgerufen am 26 Mai 2020).

[492] Hochrangige Expertengruppe für künstliche Intelligenz (2019), Ethik-Leitlinien für eine vertrauenswürdige KI, Seite 19. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

6.1 Ethische Grundsätze

Nachprüfbarkeit.

Nachprüfbarkeit bedeutet, dass Algorithmen, Daten und das Entwurfsverfahren einer Bewertung unterzogen werden können. Die Bewertung von KI-Systemen durch interne und externe Prüfer und das Vorliegen solcher Bewertungsberichte kann beträchtlich zur Vertrauenswürdigkeit der Technik beitragen. Die externe Nachprüfbarkeit sollte insbesondere bei Anwendungen sichergestellt sein, die sich auf die Grundrechte auswirken, sowie bei sicherheitskritischen Anwendungen. Das heißt nicht unbedingt, dass Informationen über Geschäftsmodelle und geistiges Eigentum im Zusammenhang mit dem KI-System immer öffentlich verfügbar sein müssen.

Minimierung negativer Auswirkungen und Meldung

Sowohl die Möglichkeit der Berichterstattung über Handlungen oder Entscheidungen, die zu einem bestimmten Systemergebnis beitragen als auch die Reaktionsfähigkeit auf die Folgen eines solchen Ergebnisses müssen gewährleistet sein. Die Ermittlung, Bewertung, Berichterstattung und Minimierung potenziell negativer Auswirkungen von KI-Systemen ist für (in)direkt Betroffene besonders wichtig. Ein angemessener Schutz für Informanten, Nichtregierungsorganisationen, Gewerkschaften und andere Stellen, die berechtigte Bedenken hinsichtlich eines KI-gestützten Systems äußern, muss gewährleistet sein. Der Einsatz von Folgenabschätzungen (z. B. „Red Teaming“ oder Formen der Algorithmen-Folgenabschätzungen) sowohl vor als auch während der Entwicklung, Einführung und Nutzung von KI-Systemen kann hilfreich sein, um negative Folgen möglichst gering zu halten. Diese Bewertungen müssen in einem angemessenen Verhältnis zu dem Risiko stehen, das die KI-Systeme darstellen.

Kompromisse

Bei der Umsetzung der oben genannten Anforderungen können Spannungen zwischen ihnen auftreten, die Kompromisse erforderlich machen. Solche Kompromisse sollten nach dem neuesten Stand der Technik rational und methodisch angegangen werden. Dies bedeutet, dass relevante Interessen und Werte, die vom KI-System betroffen sind, ermittelt werden sollten und dass im Falle von Konflikten zwischen diesen beiden ausdrückliche Kompromisse eingegangen werden müssen, die im Hinblick auf ihr Risiko für ethische Grundsätze, einschließlich der Grundrechte, bewertet werden müssen. In Situationen, in denen keine ethisch vertretbaren Kompromisse möglich sind, sollte das KI-System in einer anderen Form entwickelt, eingeführt und genutzt werden. Jede Entscheidung darüber, welcher Kompromiss einzugehen ist, sollte begründet und ordnungsgemäß dokumentiert werden. Der Entscheidungsträger ist für die Art und Weise des Zustandekommens eines angemessenen Kompromisses rechenschaftspflichtig und er sollte die Angemessenheit der daraus resultierenden Entscheidung kontinuierlich überprüfen, um sicherzustellen, dass bei Bedarf notwendige Änderungen am System vorgenommen werden können.

Rechtsmittel

Sollte es schließlich doch zu ungerechten und nachteiligen Auswirkungen kommen, sollten Vorkehrungen für einen angemessenen Rechtsschutz getroffen werden. Zu wissen, dass Rechtsbehelfe möglich sind, wenn etwas schief geht, ist der Schlüssel zur Vertrauensbildung. Schutzbedürftigen Personen oder Gruppen sollte besondere Aufmerksamkeit geschenkt werden.

6.2 Bestimmungen der DSGVO

6.2.1 Rechenschaftspflicht

Gemäß Artikel 5 Absatz 2 DSGVO ist der Verantwortliche für die Einhaltung aller in Artikel 5 Absatz 1 genannten Grundsätze der Datenschutz-Grundverordnung verantwortlich und muss dessen Einhaltung nachweisen können. Dazu gehört auch der Grundsatz der Rechenschaftspflicht.

Der Grundsatz der Rechenschaftspflicht der DSGVO ist risikobasiert: Je höher das Risiko der Datenverarbeitung für die Grundrechte und -freiheiten der betroffenen Personen ist, desto umfassender sind die Maßnahmen, die zur Minderung dieser Risiken erforderlich sind.[493] Der Grundsatz der Rechenschaftspflicht beruht auf mehreren Pflichten der Verantwortlichen, darunter: Transparenzpflichten (Artikel 12–14); Gewährleistung der Ausübung der Datenschutzrechte (Artikel 15–22); Führung eines Verzeichnisses von Datenverarbeitungstätigkeiten (Artikel 30); Meldung etwaiger Datenschutzverletzungen an eine nationale Aufsichtsbehörde (Artikel 33) und Benachrichtigung der betroffenen Personen (Artikel 34); sowie, in Fällen mit höherem Risiko, Beauftragung eines Datenschutzbeauftragten und Durchführung einer Datenschutz-Folgenabschätzung (Artikel 35).

Da die Verarbeitung personenbezogener Daten in KI-Systemen häufig als risikoreich angesehen werden kann, [494]muss der Entwickler von KI häufig einen Datenschutzbeauftragten benennen und eine Datenschutz-Folgenabschätzung durchführen. Die nächsten beiden Abschnitte befassen sich mit diesen beiden spezifischen Rechenschaftspflichten.

6.2.2 Risikobewertung und DSFAs

Eine DSFA ist ein Verfahren, bei dem der Verantwortliche vor der Einleitung eines Datenverarbeitungsvorgangs mit hohem Risiko für die Grundrechte und Freiheiten der betroffenen Personen vorab eine Abschätzung der Folgen der vorgesehenen Verarbeitungsvorgänge für den Schutz personenbezogener Daten durchführt (Artikel 35 Absatz 1).

Die Feststellung, ob die Datenverarbeitung mit einem hohen Risiko verbunden ist, ist jedoch keine leichte Aufgabe. Artikel 35 Absatz 3 führt drei Fälle auf: (1) systematische und umfassende Bewertung persönlicher Aspekte natürlicher Personen, die sich auf automatisierte Verarbeitung einschließlich Profiling gründet und die ihrerseits als Grundlage für Entscheidungen dient, die Rechtswirkung gegenüber natürlichen Personen entfalten oder diese in ähnlich erheblicher Weise beeinträchtigen; (2) umfangreiche Verarbeitung besonderer Kategorien von personenbezogenen Daten gemäß Artikel 9 Absatz 1 oder von personenbezogenen Daten über strafrechtliche Verurteilungen und Straftaten gemäß Artikel 10 oder; und (3) systematische umfangreiche Überwachung öffentlich zugänglicher Bereiche.

In Bezug auf innovative Technologien erläuterte die Artikel-29-Datenschutzgruppe einige Beispiele, darunter die „Kombination aus Fingerabdruck- und Gesichtserkennung zum

---

[493] Siehe Artikel 24, 25 und 32 DSGVO, die von den Verantwortlichen verlangen, unter Berücksichtigung der „Eintrittswahrscheinlichkeit und Schwere der Risiken für die Rechte und Freiheiten natürlicher Personen geeignete technische und organisatorische Maßnahmen umzusetzen“.

[494] Siehe insbesondere Artikel 35 Absatz 3 Buchstabe a, wonach die Datenverarbeitung unter anderem in folgenden Fällen als risikoreich gilt: „wenn es sich um eine systematische und umfassende Bewertung persönlicher Aspekte natürlicher Personen handelt, die sich auf automatisierte Verarbeitung einschließlich Profiling gründet und die ihrerseits als Grundlage für Entscheidungen dient, die Rechtswirkung gegenüber natürlichen Personen entfalten oder diese in ähnlich erheblicher Weise beeinträchtigen.“

Zwecke einer verbesserten Zugangskontrolle“ und „einige Anwendungen des „Internet der Dinge““. Diese Datenverarbeitungsvorgänge werden als risikoreich eingestuft, „der Einsatz einer solchen Technologie mit neuartigen Formen der Datenerfassung und -nutzung einhergehen kann, was möglicherweise ein hohes Risiko für die Rechte und Freiheiten von Personen mit sich bringt. Schließlich sind die persönlichen und gesellschaftlichen Folgen, die der Einsatz einer neuen Technologie haben kann, kaum absehbar.“[495]

Wenn die Verarbeitung mit einem hohen Risiko verbunden ist, sollte eine Datenschutz-Folgenabschätzung gemäß Artikel 35 Absatz 7 DSGVO durchgeführt werden. In Erwägungsgrund 90 der DSGVO wird weiter ausgeführt, dass die Risikobewertung anhand von zwei Parametern erfolgen sollte: **Eintrittswahrscheinlichkeit** und die **Schwere** dieses hohen Risikos unter Berücksichtigung der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung und der Ursachen des Risikos. Mehrere nationale Aufsichtsbehörden haben Leitlinien für die Bewertung dieser Risiken herausgegeben, darunter die Agencia Española de Protección de Datos Personales, das Information Commissioner's Office, die irische Data Protection Commission, die Commission Nationale de l'Informatique et des Libertés und andere (siehe Abschnitt „Datenschutz-Folgenabschätzung“ im Kapitel „Maßnahmen und Tools“).

Wenn die DSFA ergibt, dass der geplante Verarbeitungsvorgang ein hohes Risiko für die Grundrechte und Freiheiten der betroffenen Personen zur Folge hat, sollte der Verantwortliche in bestimmten Situationen die nationale Aufsichtsbehörde konsultieren, wie in Artikel 36 DSGVO vorgesehen. Einige Mitgliedstaaten haben Listen mit Beispielen für Datenverarbeitungsvorgänge herausgegeben, die diese obligatorische Konsultation auslösen würden. Diese umfassen Situationen, die mit KI-Techniken übereinstimmen und in einigen Fällen sogar ausdrücklich KI einschließen. Die Aufsichtsbehörden können verlangen, dass bestimmte Maßnahmen zur Risikominderung ergriffen werden, wenn dies möglich ist, bzw. die Verwendung der KI verbieten, wenn dies nicht möglich ist.

Checkliste: Ist eine Datenschutz-Folgenabschätzung notwendig?

ü Der Verantwortliche hat die Zuständigkeitsbereiche bestimmt, in denen die Datenverarbeitung stattfinden wird.

ü Der Verantwortliche hat geprüft, ob diese Länder Listen mit Verarbeitungen erlassen haben, die eine DSFA erfordern, und ob die beabsichtigten Datenverarbeitungsvorgänge, bei denen künstliche Intelligenz zum Einsatz kommt, unter diese Bestimmungen fallen.

ü Ist sich der Verantwortliche nicht sicher, ob eine DSFA erforderlich ist, konsultiert er den Datenschutzbeauftragten oder stattdessen die Rechtsabteilung des Verantwortlichen.

ü Gegegenenfalls hat der Verantwortliche eine DSFA durchgeführt.

ü Gegebenenfalls hat der Verantwortliche vorab die zuständige Aufsichtsbehörde konsultiert.

ü Falls änderungen vorgeschlagen wurden, hat der Verantwortliche den Rat der Aufsichtsbehörde befolgt.

[495] Artikel-29-Datenschutzgruppe (2017) Leitlinien zur Datenschutz-Folgenabschätzung (DSFA), WP248, Seite 10. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=6112362053 (abgerufen am 20. Mai 2020).

6.2.3 Sorgfaltspflicht des Auftragsverarbeiters

Der Grundsatz der Rechenschaftspflicht gilt auch, wenn ein Verantwortlicher die Dienste eines Auftragsverarbeiters in Anspruch nehmen will. In diesem Zusammenhang verlangt Artikel 28 Absatz 1 DSGVO [496]von Verantwortlichen, dass sie bestimmte Sorgfaltspflichten erfüllen, bevor sie Auftragsverarbeitern Zugang zu personenbezogenen Daten für die Durchführung von Datenverarbeitungstätigkeiten gewähren. Wie bei anderen Bestimmungen der DSGVO wird nicht angegeben, welche spezifischen Maßnahmen ein Verantwortlicher bei der Bewertung von Auftragsverarbeitern durchführen sollte. Als einziges Kriterium gibt die DSGVO vor, dass die **Verantwortlichen die Auftragsverarbeiter auf der Grundlage ihrer Fähigkeit beurteilen sollten, nachzuweisen, dass sie die Verarbeitungstätigkeiten im Einklang mit der DSGVO durchführen können.**

Ein Forscher, der KI entwickelt und für bestimmte Verarbeitungsvorgänge einen Dritten beauftragen muss, muss sich daher zwei Fragen stellen: (1) welche Art von Verhalten wird erwartet, um die Einhaltung dieser Verpflichtung nachzuweisen; und (2) wenn eine Form von positiven Maßnahmen erwartet wird, wie sollten die Verantwortlichen dann vorgehen, um diese Sorgfaltspflicht zu erfüllen?

In Bezug auf die erste Frage weist die DSGVO darauf hin, dass die Verantwortlichen, wenn sie die DSGVO einhalten wollen, nur einen Auftragsverarbeiter beauftragen dürfen, der die Einhaltung der DSGVO nachweisen kann. Somit müssen die Verantwortlichen Informationen anfordern, um dies zu beurteilen. Anders ausgedrückt: Die DSGVO erwartet von den Verantwortlichen, dass sie ihren potenziellen Auftragsverarbeiter aktiv danach fragen. Es reicht nicht aus, sich auf eine Zusicherungs- und Gewährleistungsklausel in der Datenverarbeitungsvereinbarung zu verlassen (siehe Abschnitt „Integrität und Vertraulichkeit“ im Kapitel „Grundsätze“).

Wie die Verantwortlichen diese Sorgfaltspflichten erfüllen sollten, wird in der DSGVO ebenfalls nicht konkret erläutert. Dennoch haben einige nationale Aufsichtsbehörden zu berücksichtigende Themen vorgeschlagen, z. B. ob der Auftragsverarbeiter Branchenstandards einhält, ob er sowohl rechtliche als auch technische Informationen über die Verarbeitung personenbezogener Daten durch den Auftragsverarbeiter bereitstellt, ob er sich an einen Verhaltenskodex hält oder ob er ein Zertifizierungssystem durchlaufen hat.[497]

Neben diesen allgemeinen Überlegungen und je nachdem, wie sich die bei diesem Dritten beantragte Verarbeitung in den Rahmen der entwickelten KI einfügt, sollten weitere Fragen gestellt werden. Diesbezüglich sollten **alle Fragen, die sich die Verantwortlichen bei der Entwicklung der KI stellen würden, auch an den Auftragsverarbeiter gestellt werden.** Zur weiteren Orientierung verweisen wir auf die in der Checkliste aufgeworfenen Fragen.

Checkliste: Sorgfaltspflicht des Auftragsverarbeiters

---

[496] Artikel 28 Absatz 1. „Erfolgt eine Verarbeitung im Auftrag eines Verantwortlichen, so arbeitet dieser nur mit Auftragsverarbeitern, die hinreichend Garantien dafür bieten, dass geeignete technische und organisatorische Maßnahmen so durchgeführt werden, dass die Verarbeitung im Einklang mit den Anforderungen dieser Verordnung erfolgt und den Schutz der Rechte der betroffenen Person gewährleistet.“

[497] ICO (kein Datum) Guide to the General Data Protection Regulation (GDPR), What responsibilities and liabilities do controllers have when using a processor? Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/contracts-and-liabilities-between-controllers-and-processors-multi/responsibilities-and-liabilities-for-controllers-using-a-processor/ (abgerufen am 20. Mai 2020).

ü Die Verantwortlichen haben Informationen darüber angefordert, wo die Datenverarbeitungsvorgänge stattfinden werden, und: (1) die unten vorgeschlagene Überprüfung der Rechtsprechung durchgeführt und (2) beurteilet, ob die Gerichtsbarkeit im Falle von Nicht-EU-Ländern von der EU-Kommission als angemessen angesehen wird.

ü Die Verantwortlichen haben die Rechtsprechung der nationalen Aufsichtsbehörden, in denen der Auftragsverarbeiter tätig ist, geprüft, um mögliche Sanktionen zu ermitteln.

ü Die Verantwortlichen haben einen Nachweis über die Einhaltung eines Verhaltenskodex oder eine Zertifizierung verlangt.

ü Die Verantwortlichen haben einen Nachweis einer einschlägigen ISO-Zertifizierung verlangt.

ü Die Verantwortlichen haben eine Kopie der Verzeichnisse über die Verarbeitungsvorgänge angefordert.

ü Die Verantwortlichen haben sich nach dem Entwicklungsprozess der KI erkundigt, insbesondere nach der Art der Daten, die zum Training der KI verwendet wurden, und nach den vom Tool benötigten Daten, um zu funktionieren und ein nützliches Ergebnis zu liefern.

6.2.4 DSB

Die Datenschutzbeauftragten spielen eine entscheidende Rolle bei der Gestaltung und Umsetzung von Datenverarbeitungstätigkeiten auf DSGVO-konforme Weise. Sie sind eine weitere Schutzmaßnahme, die die DSGVO in bestimmten Fällen vorschreibt. Im Allgemeinen wird daher die Ernennung einer solchen Person empfohlen. Die Artikel-29-Datenschutzgruppe ist der Ansicht, dass „dass der DSB ein wichtiger Akteur im Bereich der Rechenschaftspflicht sei und dass die Ernennung eines DSB die Einhaltung der Vorschriften erleichtere“.[498]

Artikel 37 Absatz 1 DSGVO [499]legt fest, wann die Verantwortlichen und die Auftragsverarbeiter einen Datenschutzbeauftragten benennen sollten. Im Falle der KI-Entwicklung ist, wie bereits erläutert**, die Bestellung eines DSB (fast) sicher erforderlich, da viele KI-Systeme personenbezogene Daten verarbeiten und damit in den meisten Fällen unter die in Artikel 37 Absatz 1 Buchstaben a und b beschriebenen Bedingungen fallen.** Diese Ansicht wird beispielsweise von der spanischen Aufsichtsbehörde geteilt.[500] Allerdings haben weder die Artikel-29-Datenschutzgruppe noch der Europäische

[498] Artikel-29-Datenschutzgruppe (2017) Leitlinien in Bezug auf Datenschutzbeauftragte („DSB“), Seite 4. Europäische Kommission, Brüssel.

[499] Artikel 37. Bennenung eines Datenschutzbeauftragten: 1. Der Verantwortliche und der Auftragsverarbeiter benennen auf jeden Fall einen Datenschutzbeauftragten, wenn a) die Verarbeitung von einer Behörde oder öffentlichen Stelle durchgeführt wird, mit Ausnahme von Gerichten, soweit sie im Rahmen ihrer justiziellen Tätigkeit handeln, b) die Kerntätigkeit des Verantwortlichen oder des Auftragsverarbeiters in der Durchführung von Verarbeitungsvorgängen besteht, welche aufgrund ihrer Art, ihres Umfangs und/oder ihrer Zwecke eine umfangreiche regelmäßige und systematische Überwachung von betroffenen Personen erforderlich machen, oder c) die Kerntätigkeit des Verantwortlichen oder des Auftragsverarbeiters in der umfangreichen Verarbeitung besonderer Kategorien von Daten gemäß Artikel 9 oder von personenbezogenen Daten über strafrechtliche Verurteilungen und Straftaten gemäß Artikel 10 besteht.

[500] Agencia Española de Protección de Datos Personales (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, Seite 35. Agencia Española de Protección de Datos Personales, Madrid. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 20. Mai 2020).

Datenschutzausschuss ausdrücklich erklärt, dass ein Datenschutzbeauftragter zwingend erforderlich ist, wenn ein Verantwortlicher oder ein Auftragsverarbeiter Datenverarbeitungsvorgänge durchführt, die KI umfassen. Die Artikel-29-Datenschutzgruppe hat jedoch darauf hingewiesen, dass **Profiling-Vorgänge als Tätigkeiten angesehen werden können, die die obligatorische Benennung eines DSB auslösen**[501], wenn diese Profiling-Vorgänge wie oben dargelegt KI betreffen.

Es wäre sinnvoll, wenn die Verordnungen der einzelnen Mitgliedstaaten über die Notwendigkeit eines DSB die Liste der Tätigkeiten, die die Benennung eines DSB erfordern, erweitern oder zumindest klare Beispiele anführen würden. Dies würde die Auslegung erleichtern, welche der von Verantwortlichen und Auftragsverarbeitern durchgeführten Datenverarbeitungstätigkeiten eine solche Benennung erfordern.

Wenn ein DSB aus einem der oben genannten Gründe benannt werden muss, ist es notwendig, dass er an der Datenschutz-Folgenabschätzung (gemäß Artikel 39 Absatz 1 Buchstabe c)) sowie bei allen anderen datenschutzrelevanten Fragen innerhalb der Einrichtung (gemäß Artikel 39 Absatz 1 Buchstabe a)) beteiligt wird. Dazu kann auch die Überprüfung eines potenziellen Auftragsverarbeiters gehören, wie im vorherigen Punkt beschrieben. Daher sollten die an der Entwicklung der KI beteiligten Forscher den DSB zu den Datenschutzfragen konsultieren, die sich bei der Entwicklung der KI ergeben könnten. Beispielsweise ist die Rolle des DSB im Zusammenhang mit KI-Systemen auch für die Zusammenarbeit bei der Abfassung einer angemessenen Information gemäß Artikel 13 und 14 relevant, um die betroffenen Personen ordnungsgemäß darüber zu informieren, wie die KI funktioniert und welche Folgen sie für sie haben könnte.

Checkliste: DSB

ü Die Verantwortlichen haben geprüft, ob die Einrichtung bereits einen DSB benannt hat.

ü Anderenfalls haben sie mit der Rechtsabteilung geprüft, ob die geplanten Datenverarbeitungsvorgänge die Benennung eines DSB erforderlich machen, indem sie entweder die einschlägigen europäischen Auslegungen, die lokalen Vorschriften, die einschlägigen lokalen Auslegungen, die Rechtsprechung – sowohl die lokale als auch die europäische – und schließlich die akademischen Auslegungen heranzogen haben.

ü Die Verantwortlichen haben gegebenenfalls die Benennung eines DSB und dessen Beteiligung an der Entwicklung der KI verlangt.

ü Generell sollte der DSB über jeden Schritt informiert sein, damit er eingreifen kann, wenn er dies für erforderlich hält.

Zusätzliche Informationen

Agencia Española de Protección de Datos Personales (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, Seite 35. Agencia Española de Protección de Datos Personales, Madrid. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

Artikel-29-Datenschutzgruppe (2013) Stellungnahme 3/2010 zum Grundsatz der Rechenschaftspflicht. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-

[501] Artikel-29-Datenschutzgruppe (2017) Leitlinien in Bezug auf Datenschutzbeauftragte („DSB“), Seite 4. Europäische Kommission, Brüssel.

recommendation/files/2010/wp173_en.pdf

Artikel-29-Datenschutzgruppe (2017) Leitlinien zur Datenschutz-Folgenabschätzung (DSFA), Seiten 9–10. Europäische Kommission, Brüssel. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

KI: Schritt-für-Schritt-Prozess

Iñigo de Miguel Beriain (UPV/EHU)

Danksagung: Der Autor dankt Andres Chomsky, Oliver Feeney, Gianclaudio Malgieri Aurélie Pols und Marko Sijan für ihre Ratschläge, Beiträge und Rückmeldungen zu den Entwürfen. Selbstverständlich trage ich für alle Fehler die volle Verantwortung.

Dieser Teil des Leitfadens wurde von Marko Sijan, Senior Advisor Specialist, (HR DPA) überprüft und bestätigt.

EINLEITUNG TEIL B

Dieser zweite Teil der Leitlinien basiert auf den Phasen des CRISP-DM-Modells,[502] das weithin zur Erläuterung der Phasen bei der Entwicklung von Datenanalysen und datenintensiven KI-Tools verwendet wird. Es wurde vom SHERPA-Projekt für die Entwicklung seiner Guidelines for the Ethical Development of AI and Big Data Systems (Leitlinien für die ethische Entwicklung von KI- und Big-Data-Systemen) ausgewählt.[503] Die sechs Phasen umfassen: Geschäftsverständnis, Datenverständnis, Datenaufbereitung, Modellierung, Bewertung und Bereitstellung. Es handelt sich dabei nicht um eine feste Klassifizierung, da Entwickler häufig einige dieser Schritte miteinander vermischen. So kann zum Beispiel ein trainierter Algorithmus nach der Validierungsphase durch ein erneutes Training verbessert werden.

Dennoch ist hervorzuheben, dass einige der ethischen und rechtlichen Anforderungen an die Entwicklung künstlicher Intelligenz während des gesamten Lebenszyklus einer KI-

[502] C. Shearer (2000),The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing* 5(4): 13–23. Verfügbar unter: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (abgerufen am 15. Mai 2020).

[503] SHERPA-Projekt (2019), Guidelines for the ethical development of AI and big data systems: an ethics by design approach. SHERPA-Projekt. Verfügbar unter: www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (abgerufen am 15. Mai 2020).

Entwicklung kontinuierlich bewertet werden müssen. Verantwortliche müssen die ethische Zulässigkeit der Verarbeitung und deren unerwarteten Folgen überwachen. Zudem sollten sie die möglichen Nebeneffekte einer solchen Verarbeitung in einem sozialen Umfeld bewerten, und zwar über die ursprünglich gedachten Grenzen des Zwecks, der Dauer und der Ausdehnung hinaus.[504] Und dies muss laut Artikel 25 DSGVO während des gesamten Lebenszyklus eines KI-Tools geschehen. Die Artikel-29-Datenschutzgruppe erklärte dazu:

> *„Die Verantwortlichen sollten die von ihnen verarbeiteten Datensätze häufigen Bewertungen unterziehen, um Verzerrungen aufzuspüren und Möglichkeiten zu entwickeln, benachteiligende Elemente wie ein zu starkes Verlassen auf Zusammenhänge anzugehen. Weitere sinnvolle Maßnahmen sind Systeme, die Algorithmen prüfen, sowie regelmäßige Überprüfungen der Richtigkeit und Relevanz automatisierter Entscheidungen einschließlich Profiling. Die Verantwortlichen sollten geeignete Verfahren und Maßnahmen einführen, um Fehler, Unrichtigkeiten oder Diskriminierung auf der Grundlage besonderer Datenkategorien zu vermeiden. Diese Maßnahmen sollten zyklisch angewandt werden, nicht nur in der Planungsphase, sondern auch durchgehend, wenn Profiling auf Personen angewandt wird. Das Ergebnis dieser Prüfung sollte wieder in den Systemaufbau einfließen.“*[505]

Darüber hinaus ist zu berücksichtigen, dass KI eine allgemeine Bezeichnung ist, die eine Vielzahl verschiedener Technologien umfasst. Grundsätzlich muss zwischen überwachtem maschinellem Lernen (von Menschen gekennzeichnete Eingabedaten werden einem Algorithmus übergeben, der dann die Regeln beruhend auf validierten Beispielfällen festlegt) und unüberwachtem Lernen (nicht gekennzeichnete Eingabedaten werden einem Algorithmus übergeben, der seine eigene Klassifizierung durchführt und eigene Ausgabedaten frei erzeugt, wenn ihm ein Muster oder eine Variable präsentiert wird) unterschieden werden. Beim überwachten Lernen muss der Überwacher der Maschine beibringen, welchen Output sie produzieren soll, d.h. sie „trainieren“. Grundsätzlich ist das überwachte Lernen leichter zu verstehen und zu überwachen.[506] Da die ür das Training verwendeten Datensätze von den Trainern ausgewählt werden, können einige der beunruhigendsten Herausforderungen, die diese Technologien mit sich bringen, recht vernünftig bewältigt werden. Unüberwachte KI hingegen und insbesondere Techniken wie Deep Learning bedürfen einer ausgefeilteren Überwachung und Kontrolle, da Unklarheiten, Verzerrungen oder Profiling viel schwieriger zu erkennen sind, zumindest in einigen Phasen des Lebenszyklus der KI-Entwicklung.

In diesem Teil der Leitlinien versuchen wir, sowohl überwachte als auch unüberwachte KI zu unterstützen. Wir sind uns bewusst, dass es fast unmöglich ist, Empfehlungen für alle möglichen Situationen auszusprechen. Wir hoffen jedoch, die Grundlagen hervorheben und nützliche zusätzliche Informationsquellen einbeziehen zu können. Schließlich sind wir uns darüber im Klaren, dass manche Experten einige unserer Empfehlungen von einer Phase zur

---

[504] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 7. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

[505] Artikel-29-Datenschutzgruppe (2017) Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679. Angenommen am 3. Oktober 2017. Zuletzt überarbeitet und angenommen am 6. Februar 2018. Europäische Kommission, Brüssel, Seite 28. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (abgerufen am 15. Mai 2020).

[506] CNIL (2017), How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence. Commission Nationale de l'Informatique et des Libertés, Paris, Seite 17. Verfügbar unter: www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (abgerufen am 15. Mai 2020).

nächsten verschieben würden. Außerdem könnten einige der Empfehlungen für mehrere Phasen gelten. Daher empfehlen wir dringend, diese Leitlinien nach bestem Wissen und Gewissen anzupassen.

Dieses Dokuments ist einfach gegliedert und zu verfolgen. Zunächst wird das Kapitel von Colin Shearer zitiert,[507] gefolgt von einer Beschreibung der Aufgaben, die laut diesem Autor in jeder konkreten Phase des Prozesses anfallen. Anschließend werden einige Empfehlungen vorgestellt, die zu diesem Zeitpunkt umgesetzt werden sollten. Verweise auf andere Kapitel der Leitlinien sind hervorgehoben, während Verweise auf andere Teile dieses Kapitels mit Querverweisen versehen sind. Die Anhänge schließlich enthalten Verweise auf einige Tools, die für die Zwecke dieses Teils der Leitlinien nützlich sein könnten. Anhang I enthält die von der spanischen Datenschutzbehörde erarbeiteten Empfehlungen in Bezug auf die Prüfung von KI-Tools. Anhang II ist spezifischer, da er sich auf den Einsatz der KI im Gesundheitswesen bezieht. Er ist jedoch ein hervorragender Leitfaden für alle, die ein KI-Tool in diesem Sektor entwickeln wollen. In der Zukunft werden wir nach Möglichkeit weitere Anhänge aufnehmen, sobald ein effizienter Mechanismus dafür geschaffen ist.

1 Geschäftsverständnis (Business Understanding)

„Im Mittelpunkt der ersten Phase steht ein präzises Verständnis der Projektziele aus geschäftlicher Perspektive, der Abbildung dieses Wissen in einer Data Mining-Problemdefinition und der anschließenden Entwicklung eines vorläufigen Plans, um die Ziele zu erreichen. Um zu verstehen, welche Daten später wie analysiert werden sollen, benötigen Data Mining-Fachleute ein umfassendes Verständnis des Unternehmens, für das sie eine Lösung finden wollen. Die Phase des Geschäftsverständnisses umfasst mehrere wichtige Schritte, darunter die Bestimmung der Geschäftsziele, die Bewertung der Situation, die Bestimmung der Data Mining-Ziele und die Erstellung des Projektplans."[508]

1.1 Beschreibung

Dieses allgemeine Ziel umfasst vier Hauptaufgaben:

1. Bestimmung der Geschäftsziele:
    a. Ermittlung des primären Geschäftsziels sowie der damit verbundenen Fragen, die das Unternehmen angehen möchte.
    b. Bestimmung des Erfolgsmaßstabs.
2. Bewertung der Situation:
    a. Ermittlung der für das Projekt verfügbaren materiellen und personellen Ressourcen.

[507] C. Shearer (2000),The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing* 5(4): 13–23. Verfügbar unter: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (abgerufen am 15. Mai 2020).

[508] C. Shearer (2000),The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing* 5(4): 13–23, Seite 14. Verfügbar unter: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (abgerufen am 15. Mai 2020).

b. Ermittlung der verfügbaren Daten verfügbar, um das primäre Geschäftsziel zu erreichen.

c. Auflistung der dem Projekt zugrunde liegenden Annahmen.

d. Auflistung der Projektrisiken, Auflistung möglicher Lösungen für diese Risiken, Erstellung eines Glossars der Geschäfts- und Data Mining-Begriffe und Erstellung einer Kosten-Nutzen-Analyse für das Projekt.

3. Bestimmung der Data Mining-Ziele:

   a. Festlegung, welcher Grad an Vorhersagegenauigkeit erwartet wird, um das Projekt als erfolgreich zu betrachten.

4. Erstellung eines Projektplans:

   a. Beschreibung des beabsichtigten Plans zur Erreichung der Data Mining-Ziele, einschließlich der Beschreibung spezifischer Schritte und eines vorgeschlagenen Zeitplans, einer Bewertung potenzieller Risiken und einer ersten Bewertung der zur Unterstützung des Projekts erforderlichen Tools und Techniken.

1.2 Wichtigste zu ergreifende Maßnahmen

1.2.1 Festlegung der Geschäftsziele

KI-Entwickler sollten von Anfang an wissen, welche Leistungen sie vom Tool erwarten. Je ungenauer diese Erwartungen sind, desto schwieriger wird es, die genauen Zwecke der Verarbeitung zu bestimmen (siehe Unterabschnitt „Voraussetzungen für die Rechtmäßigkeit – festgelegte, eindeutige Zwecke“, Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“, Kapitel „Grundsätze“). Wenn wir uns vor Augen halten, dass die Verantwortlichen die Zwecke der Verarbeitung eindeutig angeben müssen, d. h. „in verständlicher Weise offenlegen, erläutern oder ausdrücken müssen“,[1] sind genaue Erwartungen sehr empfehlenswert. Allerdings muss man zwischen den verschiedenen Phasen des Lebenszyklus einer KI-Entwicklung unterscheiden. In der Trainingsphase kann die Verwendung großer Datenmengen unerlässlich sein, um den konkreten Nutzen des Werkzeugs abzuschätzen. Daher könnte die Verarbeitung großer Datenmengen akzeptabel sein, auch wenn das spezifische Ziel (die Entwicklung des KI-Tools) nicht so präzise ist. Dies wäre natürlich in der letzten Phase des Prozesses nicht so einfach zu akzeptieren, d. h. bei der Einführung und Nutzung des Tools. Wenn der Verantwortliche zu diesem Zeitpunkt eine große Menge an Daten verwenden muss, ist eine viel detailliertere Begründung erforderlich.

In jedem Fall ist hervorzuheben, dass einige Schlüsselgedanken von Anfang an im Auge behalten werden müssen. So ist beispielsweise die Festlegung des erwarteten Grads an Vorhersagegenauigkeit, damit das Projekts als Erfolg angesehen werden kann, von wesentlicher Bedeutung, um die für die Entwicklung des KI-Tools benötigte Datenmenge oder die Art dieser Daten bewerten zu können. Der Grad der Vorhersagbarkeit oder Präzision des Algorithmus, die Validierungskriterien der Tests, die maximale Menge oder Mindestqualität der Daten, die für den Einsatz in der realen Welt erforderlich sind, sind nur einige der grundlegenden Merkmale einer KI-Entwicklung. Diese wichtigen Entscheidungen sollten bereits in der ersten Phase des Lebenszyklus einer Lösung berücksichtigt werden. Dies ist äußerst hilfreich für die Umsetzung einer Richtlinie des Datenschutzes durch Technikgestaltung.

So sollten KI-Entwickler je nach Anwendungsfall akzeptable Schwellenwerte für falsche positive/negative Ergebnisse oder Bereiche festlegen und dann eine Nutzenbilanz erstellen. KI-Entwickler sollten sich darüber im Klaren sein, dass die Bestimmung des erwarteten Genauigkeitsgrads zweifellos mit der Menge der benötigten Daten zusammenhängt. Es ist nicht dasselbe, beispielsweise ein Produkt für das Gesundheitswesen oder für die Empfehlung von Fernsehserien zu entwickeln. Darüber hinaus ist es selbst im Gesundheitswesen nicht dasselbe, ein Tool zu entwickeln, das eine erste Triage durchführen kann (d. h. die Empfehlung, ob ein Hausarzt oder ein Spezialist eingreifen sollte), oder eine Lösung, die Radiologen bei der Diagnose unterstützen soll. Je nach den spezifischen Aufgaben des Tools werden höhere oder niedrigere Genauigkeitsanforderungen gestellt.

Wenn ein akzeptables Genauigkeitsniveau bei Verwendung deutlich weniger personenbezogener Daten erreicht werden kann als für ein höheres Genauigkeitsniveau erforderlich sind, sollte dies unbedingt in Betracht gezogen werden. Darüber hinaus müssen KI-Entwickler bedenken, dass jede geringfügige Verbesserung der Vorhersagegenauigkeit mitunter eine erhebliche Zunahme der benötigten personenbezogenen Daten erfordert.[2] Wenn sie also eine grundlegende Änderung der erforderlichen Vorhersagegenauigkeit in Betracht ziehen, sollten sie sorgfältig prüfen, ob dies mit dem Grundsatz der Datenminimierung vereinbar ist.

1.2.2 Die Wahl der technischen Lösung

Im Allgemeinen sollten KI-Entwickler stets verständlichere Algorithmen gegenüber weniger verständlichen Algorithmen bevorzugen (siehe Abschnitt **Bestimmungen der DSGVO: Transparenz**). Zielkonflikte zwischen der Erklärbarkeit/Transparenz und der besten Leistung des Systems müssen je nach Nutzungskontext angemessen abgewogen werden. So können beispielsweise im Gesundheitswesen die Genauigkeit und Leistung eines Systems wichtiger sein als seine Erklärbarkeit, während im Polizeiwesen die Erklärbarkeit viel wichtiger ist, um Verhaltensweisen und Ergebnisse der Strafverfolgung zu rechtfertigen. In anderen Bereichen wie beispielweise der Personalbeschaffung werden sowohl der Genauigkeit als auch der Erklärbarkeit ein ähnlicher Wert beigemessen.[509] Wenn eine Dienstleistung sowohl durch einen leicht verständlichen als auch durch einen undurchsichtigen Algorithmus angeboten werden kann, d. h. wenn es keinen Zielkonflikt zwischen Erklärbarkeit und Leistung gibt, sollte sich der Verantwortliche für den besser interpretierbaren Algorithmus entscheiden (siehe Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“, Kapitel „Grundsätze“).

Kasten 13: Interpretation der Interpretierbarkeit

Auch wenn die Interpretierbarkeit empfohlen wird, muss eingeräumt werden, dass es sich hierbei nicht um ein eindeutiges Konzept handelt. In der wissenschaftlichen Literatur finden sich unterschiedliche Begründungen für die Interpretierbarkeit und sogar unzählige Vorstellungen darüber, welche Attribute Modelle interpretierbar machen. Es ist immer noch unklar, welche Vorteile „Interpretation“ an einen Tisch bringt. Auf den ersten Blick scheint die Annahme vernünftig, dass einfache, lineare Algorithmen leichter zu verstehen sind. Doch „bei einigen Arten der Post-hoc-Interpretation zeigen tiefe neuronale Netze einen klaren Vorteil. Sie lernen vielfältige Repräsentationen, die visualisiert, verbalisiert

[509] SHERPA-Projekt (2019), Guidelines for the ethical development of AI and big data systems: an ethics by design approach. SHERPA, Seite 26. Verfügbar unter: www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (abgerufen am 15. Mai 2020).

oder zum Clustering verwendet werden können. In Anbetracht der Wünsche in Bezug auf die Interpretierbarkeit scheinen lineare Modelle eine bessere Erfolgsbilanz bei der Untersuchung der natürlichen Welt zu haben. Allerdings ist uns kein theoretischer Grund bekannt, warum das so sein muss. Es ist denkbar, dass sich Post-hoc-Interpretationen in ähnlichen Szenarien als nützlich erweisen könnten.“ Daher lassen sich nur schwer spezifische Empfehlungen abgeben, welche Art von Modell aufgrund ihrer „Interpretierbarkeit“ bevorzugt werden sollte.[510]

1.2.3 Implementierung eines Schulungsprogramms

Diese Maßnahme ist eine der wichtigsten Ratschläge, der vom ersten Moment einer geschäftlichen KI-Entwicklung an berücksichtigt werden sollte. Designer von Algorithmen (Entwickler, Programmierer, Datenwissenschaftler, Ingenieure), die das erste Glied in der algorithmischen Kette darstellen, sind sich der ethischen und rechtlichen Auswirkungen ihres Handelns wahrscheinlich nicht bewusst. Eines der Hauptprobleme der KI besteht darüber hinaus darin, dass sie im Allgemeinen personenbezogene Daten verwendet, die in großen Datensätzen enthalten sind. Dadurch wird die Beziehung zwischen den Daten und der betroffenen Person in gewisser Weise verwischt, sodass es zu Verletzungen von Vorschriften kommt, die bei einer direkten Beziehung zwischen dem Verantwortlichen und der betroffenen Person nur selten auftreten.[511] Dies könnte Folgen für die angemessene Einhaltung der Datenschutzstandards haben. Diese Schlüsselpersonen müssen sich somit unbedingt der ethischen und sozialen Auswirkungen ihrer Arbeit und der Tatsache umfassend bewusst sein, dass diese sogar bis hin zu gesellschaftlichen Entscheidungen reichen können, [512] auch wenn das Alibi des „Rogue Engineering“ nach dem Fall Google Street View kaum noch vorgebracht werden kann.[513]

Um zu vermeiden, dass die falsche Darstellung ethischer und rechtlicher Fragen unerwünschte Folgen nach sich zieht, gibt es zwei Hauptmaßnahmen, die ergriffen werden können. Erstens könnten die Entwickler sicherstellen, dass Designer von Algorithmen die Auswirkungen ihres Handelns sowohl für den Einzelnen als auch für die Gesellschaft verstehen und sich ihrer Verantwortung bewusst sind, indem sie lernen, kontinuierlich aufmerksam und wachsam zu sein.[514] In diesem Zusammenhang könnte eine optimale Schulung aller am Projekt beteiligten Personen (Entwickler, Programmierer, Datenwissenschaftler, Ingenieure, Forscher) noch vor Projektbeginn eines der effizientesten Instrumente sein, um Zeit und Ressourcen im Hinblick auf die Einhaltung der Datenschutzvorschriften zu sparen. So sollten grundlegende Schulungsprogramme implementiert werden, die zumindest die Grundlagen der Charta der Grundrechte, die in Artikel 5 DSGVO dargelegten Grundsätze, die Notwendigkeit einer Rechtsgrundlage für die Verarbeitung (einschließlich Verträge zwischen den Parteien) usw. umfassen.

---

[510] Z.C. Lipton (2017), The mythos of model interpretability, 2016 ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. Verfügbar unter: https://arxiv.org/pdf/1606.03490.pdf (abgerufen am 15. Mai 2020).

[511] A. Kuyumdzhieva (2018), Ethical challenges in the digital era: focus on medical research, Seiten 45–62 in: Z. Koporc (ed.) *Ethics and integrity in health and life sciences research*. Bingley Emerald.

[512] CNIL (2017), How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence. Commission Nationale de l'Informatique et des Libertés, Paris, Seite 55. Verfügbar unter: www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (abgerufen am 15. Mai 2020).

[513] Siehe beispielsweise: https://www.slashgear.com/googles-rogue-engineer-street-view-excuse-blown-apart-30225200/

[514] Ibid., Seite 55.

Die Schulung von Personen, die noch nie mit Datenschutzfragen in Berührung gekommen sind, könnte jedoch schwierig sein. Eine alternative Maßnahme ist die Einbeziehung eines Experten für Datenschutz sowie ethische und rechtliche Fragen in das Entwicklungsteam, um ein disziplinübergreifendes Team zu bilden. Zu diesem Zweck könnte ein Experte (ein interner Mitarbeiter oder ein externer Berater) eingestellt werden, der die Strategie und die nachfolgenden Entscheidungen über personenbezogene Daten, die für die Entwicklung der Toole erforderlich sind, in enger Zusammenarbeit mit dem Datenschutzbeauftragten ausarbeitet.

Die Umsetzung angemessener Maßnahmen zur Gewährleistung der Vertraulichkeit, Integrität und Verfügbarkeit von Daten ist ebenfalls sehr zu empfehlen (siehe Unterabschnitt „Maßnahmen zur Unterstützung der Vertraulichkeit", Abschnitt „Integrität und Vertraulichkeit", Kapitel „Grundsätze").

1.2.4 **Entwurf legitimer Datenverarbeitungstools** Gemäß Artikel 5 Absatz 1 Buchstabe b DSGVO müssen personenbezogene Daten „für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden". Das Konzept der Legitimität ist in der DSGVO nicht genau definiert. Laut der Artikel 29-Datenschutzgruppe beinhaltet die Legitimität jedoch, dass die Daten „im Einklang mit dem Gesetz" verarbeitet werden müssen, wobei „Gesetz" als weit gefasster Begriff verstanden werden sollte, der „alle Formen des geschriebenen Rechts und des Gewohnheitsrechts, primäre und sekundäre Rechtsvorschriften, kommunale Verordnungen, gerichtliche Präzedenzfälle, Verfassungsgrundsätze, Grundrechte, andere Rechtsgrundsätze sowie die Rechtsprechung umfasst, d. h. wie ein solches „Gesetz" von einem zuständigen Gericht ausgelegt und berücksichtigt würde".[515]

Es handelt sich also um ein umfassenderes Konzept als das der Rechtmäßigkeit. Es umfasst die Einhaltung der wichtigsten Werte der geltenden Verordnung und der wichtigsten ethischen Grundsätze. So werden beispielsweise einige konkrete KI-Entwicklungen das Eingreifen einer Ethik-Kommission erfordern. In einigen anderen Fällen könnten Richtlinien oder andere weiche Regelungen anwendbar sein. KI-Entwickler sollten eine angemessene Erfüllung dieser Anforderungen gewährleisten, indem sie bereits während dieser Vorstufe des Tool-Lebenszyklus einen entsprechenden Plan erstellen (siehe Unterabschnitt „Legitimität und Rechtmäßigkeit", Abschnitt „Rechtmäßigkeit, Fairness und Transparenz", Kapitel „Grundsätze").

1.2.5 Annahme eines risikobasierten Denkansatzes

Die Verantwortlichen sollten die Risiken in Bezug auf die Rechte, Interessen und Freiheiten der betroffenen Personen minimieren. Zu diesem Zweck sollten sie nach einem risikobasierten Ansatz arbeiten (siehe Abschnitt „Integrität und Vertraulichkeit", Kapitel „Grundsätze").

Der risikobasierte Datenschutzansatz verlangt von den Verantwortlichen, dass sie ihren Verpflichtungen nachkommen und geeignete Maßnahmen im Zusammenhang mit ihren besonderen Umständen ergreifen – Art, Umfang, Umstände und Zwecke der von ihnen beabsichtigten Verarbeitung sowie die Risiken, die dies für die Rechte und Freiheiten der betroffenen Personen darstellt. Bei den Überlegungen zur Einhaltung der Vorschriften geht es also darum, die Risiken für die Rechte und Freiheiten der betroffenen Personen zu bewerten

---

[515] Artikel-29-Datenschutzgruppe (2013) Stellungnahme 03/2013 zur Zweckbindung, angenommen am 2. April 2013, WP203. Europäische Kommission, Brüssel, Seite 20. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf (abgerufen am 15. Mai 2020)

und zu beurteilen, was unter diesen Umständen angemessen ist. In jedem Fall müssen die Verantwortlichen sicherstellen, dass sie die Datenschutzanforderungen erfüllen (siehe Abschnitt „Rechenschaftspflicht“, Kapitel „Grundsätze“).

Risikobasiertes Denken in Bezug auf die Vertraulichkeit von Daten oder ein risikobasierter Ansatz in Bezug auf die Frage, welcher Schaden den Menschen zugefügt werden könnte, muss von den ersten Schritten des Prozesses an berücksichtigt werden. Wenn dies erst später in Betracht gezogen wird, könnte es zu spät sein. Um die Risiken für die betroffenen Personen infolge der Verarbeitung personenbezogener Daten in KI-Tools in den Griff zu bekommen, ist es wichtig, dass die Verantwortlichen ein ausgereiftes Verständnis und eine klare Formulierung der Grundrechte, Risiken und der Abwägung dieser und anderer Interessen entwickeln. Letztlich müssen die Verantwortlichen die Risiken für die Rechte der betroffenen Personen infolge des Einsatzes von KI bewerten und festlegen, wie sie diese Risiken bewältigen werden und welche Auswirkungen dies auf den Einsatz der KI hat.[516] Dazu müssen zwei Schlüsselfaktoren berücksichtigt werden:[517]

- Risiken, die sich aus der Verarbeitung selbst ergeben, beispielsweise das Auftreten von Verzerrungen im Zusammenhang mit Profiling oder automatisierten Entscheidungsfindungssystemen (siehe **Bestimmungen der DSGVO**).
- Risiken, die sich aus der Verarbeitung in Bezug auf den sozialen Kontext ergeben, sowie die Nebenwirkungen, die indirekt mit dem Gegenstand der Verarbeitung zusammenhängen und auftreten können.

Kasten 14: Die Bedeutung von Softwareanbietern in Bezug auf die Sicherheit

Im August 2015 meldete ein Unternehmen für medizinische Software aus Indiana der US-Regierung, dass seine Netzwerke gehackt worden waren und die privaten Daten von 3,9 Millionen Menschen offengelegt wurden. Dazu gehörten personenbezogene Daten wie Namen, Adressen, Geburtsdaten, Sozialversicherungsnummern und Gesundheitsdaten. Laut IBM X-Force Research handelte es sich um eine der größten Datenschutzverletzungen im Gesundheitswesen der letzten Jahre. Nach Angaben des Unternehmens wurde der Angriff neunzehn Tage, nachdem sich die Täter unbefugten Zugang zu dem Netzwerk verschafft hatten, entdeckt. Die Kunden wurden erst fast einen Monat nach Beginn des Angriffs benachrichtigt.[518]

Darüber hinaus müssen die Verantwortlichen sicherstellen, dass geeignete technische und organisatorische Maßnahmen getroffen werden, um das Sicherheitsrisiko zu eliminieren oder zumindest abzuschwächen, sodass die Eintrittswahrscheinlichkeit der festgestellten Bedrohungen verringert wird oder deren Auswirkungen gemindert werden. Die allgemeine Beschreibung der technischen und organisatorischen Sicherheitsmaßnahmen muss nach

[516] ICO (2020) AI auditing framework – draft guidance for consultation. Information Commissioner’s Office, Wilmslow, Seiten 13–14. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[517] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 30. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

[518] IBM X-Force® Research (2017) Security trends in the healthcare industry: Data theft and ransomware plague healthcare organizations. IBM Security, Somers, NY, Seite 7. Verfügbar unter: www.ibm.com/downloads/cas/PLWZ76MM (abgerufen am 17. Mai 2020).

Möglichkeit Teil des Verzeichnisses von Verarbeitungstätigkeiten werden (Artikel 30 Absatz 1 Buchstabe g für Verantwortliche und Artikel 30 Absatz 2 Buchstabe d für Auftragsverarbeiter), und alle durchgeführten Maßnahmen sind Teil der DSFA als unterstützende Abhilfemaßnahmen zur Risikobegrenzung. Sobald die ausgewählten Maßnahmen umgesetzt sind, sollte das verbleibende Restrisiko schließlich bewertet und weiterhin kontrolliert werden. Sowohl die Risikoanalyse als auch die DSFA sind die Instrumente, die dabei zum Einsatz kommen.

Eine DSFA ist bei der Entwicklung von KI oft zwingend erforderlich (siehe Unterabschnitt „In welchen Fällen muss eine Datenschutz-Folgenabschätzung durchgeführt werden?“, Abschnitt „Datenschutz-Folgenabschätzung”, Kapitel „Maßnahmen udn Tools“). Dies hängt gemäß Artikel 35 Absatz 3 DSGVO davon ab, ob die mit der Verarbeitung verbundenen Risiken hoch sind oder nicht. Sie wird jedoch dringend empfohlen, da sie die Rechenschaftspflicht unterstützt. Im Zweifelsfall wird eine Konsultation der zuständigen Aufsichtsbehörde vor der Verarbeitung dringend empfohlen. Schließlich darf nicht vergessen werden, dass sich bei der Nutzung von Big Data und KI die künftigen Risiken nur schwer vorhersehen lassen, sodass eine Bewertung der ethischen Auswirkungen nicht ausreicht, um alle möglichen Risiken zu berücksichtigen. Daher ist es wichtig, eine Neubewertung der Risiken in Erwägung zu ziehen, und es ist auch sehr empfehlenswert, eine dynamischere Methode zur Bewertung von Forschungsrisiken zu integrieren. Zögern Sie nicht, in anderen Phasen des Prozesses bei Bedarf zusätzliche Datenschutz-Folgenabschätzungen durchzuführen.

1.2.6 Vorbereiten der Dokumentation der Verarbeitung

Wer auch immer personenbezogene Daten verarbeitet (einschließlich Verantwortliche[519] und Auftragsverarbeiter[520]), muss seine Tätigkeiten in erster Linie zur Verwendung durch qualifizierte/zuständige Aufsichtsbehörden dokumentieren.[521] Dies muss durch Aufzeichnungen über die Verarbeitung geschehen, die von dem Unternehmen zentral für alle Verarbeitungstätigkeiten geführt werden, sowie durch zusätzliche Dokumentation, die sich auf eine einzelne Datenverarbeitungstätigkeit bezieht (siehe Abschnitt „Dokumentation der Verarbeitung”, Kapitel „Maßnahmen und Tools“). Diese Vorbereitungsphase ist der ideale Zeitpunkt, um eine systematische Erfassung der erforderlichen Dokumentation festzulegen, da das Unternehmen zu diesem Zeitpunkt die Verarbeitungstätigkeit konzipiert und plant[522].

Die Verantwortlichen sollten eine Datenschutzrichtlinie erstellen, die die Rückverfolgbarkeit von Informationen ermöglicht (im Falle genehmigter Verhaltenskodizes sollten diese umgesetzt werden; siehe Unterabschnitt „Größenvorteile bei der Einhaltung der Vorschriften und deren Nachweis“, Abschnitt „Rechenschaftspflicht“, Kapitel „Grundsätze“). Diese Richtlinie sollte auch die den Auftragsverarbeitern zugewiesenen Verantwortlichkeiten deutlich machen und in der Verarbeitungsvereinbarung die Aufgaben enthalten, die ihnen im Zusammenhang mit der Ausübung der Rechte der betroffenen Personen übertragen werden. KI-Entwickler sollten immer daran denken, dass Artikel 32 Absatz 4 DSGVO festlegt, dass ein wichtiges Sicherheitselement in der Gewährleistung besteht, dass Mitarbeiter nur auf

---

[519] Siehe Artikel 30 Absatz 1 DSGVO.

[520] Siehe Artikel 30 Absatz 2 DSGVO.

[521] Siehe Artikel 58 Absatz 1 Buchstabe a, Artikel 30 Absatz 4 und Artikel 5 Absatz 2 DSGVO.

[522] In Artikel 25 Absatz 1 DSGVO wird dies als „Zeitpunkt der Festlegung der Mittel für die Verarbeitung“ bezeichnet.

Anweisung und gemäß den Weisungen des Verantwortlichen handeln (siehe Abschnitt „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“).

Die Verantwortlichen müssen stets bedenken, dass bei der Entwicklung von KI-Tools häufig verschiedene Datensätze verwendet werden. Die Rückverfolgbarkeit der Verarbeitung, Informationen über die mögliche Wiederverwendung von Daten und die Verwendung von Daten, die zu verschiedenen Datensätzen gehören, in verschiedenen oder denselben Phasen des Lebenszyklus müssen durch die Aufzeichnungen sichergestellt werden.

1.2.7 Dokumentation der Verarbeitung

Wie in den Anforderungen und Akzeptanztests für den Kauf und/oder die Entwicklung der verwendeten Software, Hardware und Infrastruktur (siehe Unterabschnitt des Abschnitts „Dokumentation der Verarbeitung“) dargelegt, müssen die Risikobewertung und die getroffenen Entscheidungen dokumentiert werden, um die Anforderung „Datenschutz durch Technikgestaltung“ (Artikel 25 DSGVO) zu erfüllen.

Schließlich sollten sich die Verantwortlichen stets bewusst sein, dass gemäß Artikel 32 Absatz 1 Buchstabe d DSGVO der Datenschutz ein Verfahren ist. Daher **sollten sie die Wirksamkeit der technischen und organisatorischen Maßnahmen regelmäßig überprüfen, bewerten und evaluieren.** Zu diesem Zeitpunkt sollten Verfahren geschaffen werden, dank der die Verantwortlichen Änderungen erkennen können, die eine erneute Überprüfung der DSFA auslösen würden. Die Verantwortlichen sollten nach Möglichkeit versuchen, ein dynamisches Modell zur Überwachung der betreffenden Maßnahmen einzuführen (siehe „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“).

Kasten 15: Die extreme Problematik der Rechenschaftspflicht bei der KI-Entwicklung

Auch wenn die Rechenschaftspflicht ein notwendiges Ziel ist und auf die Zuweisung von Verantwortlichkeiten an einen bestimmten Auftragsverarbeiter nicht verzichtet werden kann, müssen sich die Verantwortlichen immer bewusst sein, dass die Funktionsweise von KI die Überwachung eines Systems extrem erschweren kann. Wie die CNIL feststellte, „muss die Frage, wo Rechenschaftspflicht und Entscheidungsfindung angesiedelt werden können, bei maschinellen Lernsystemen etwas anders angegangen werden“. Daher sollten Verantwortliche besser den Ansatz einer Kette der Rechenschaftspflicht vom Systemdesigner bis hin zum Benutzer verfolgen, und zwar über die Person, die die Trainingsdaten in das System einspeist. Letztere wird in Abhängigkeit der Eingabedaten unterschiedlich vorgehen.

Zu diesem Thema könnte man den Chatbot Tay von Microsoft anführen. Er wurde nur vierundzwanzig Stunden nach seiner Einführung abgeschaltet, nachdem er aus den Beiträgen von Social-Media-Nutzern gelernt und begonnen hatte, selbst rassistische und sexistische Kommentare zu twittern. Es erübrigt sich zu erwähnen, dass die genaue Aufteilung der Verantwortung zwischen diesen verschiedenen Kettengliedern sehr mühsam sein kann.[523]

[523] CNIL (2017), How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence. Commission Nationale de l'Informatique et des Libertés, Paris, Seite 29. Verfügbar unter: www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (abgerufen am 15. Mai 2020).

1.2.8 Überprüfung des Rechtsrahmens

Die DSGVO enthält einen spezifischen Rechtsrahmen für die Verarbeitung zu wissenschaftlichen Forschungszwecken (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“, Kapitel „Konzepte“).[524] Wenn die KI-Entwicklung als wissenschaftliche Forschung angesehen werden kann, „können vorbehaltlich der Bedingungen und Garantien gemäß Absatz 1 des vorliegenden Artikels im Unionsrecht oder im Recht der Mitgliedstaaten insoweit Ausnahmen von den Rechten gemäß der Artikel 15, 16, 18 und 21 vorgesehen werden, als diese Rechte voraussichtlich die Verwirklichung der spezifischen Zwecke unmöglich machen oder ernsthaft beeinträchtigen und solche Ausnahmen für die Erfüllung dieser Zwecke notwendig sind“ (Artikel 89 Absatz 2). Darüber hinaus gilt gemäß Artikel 5 Buchstabe b „eine Weiterverarbeitung der erfassten Daten gemäß Artikel 89 Absatz 1 nicht als unvereinbar mit den ursprünglichen Zwecken („Zweckbindung“)". Es gibt einige andere besondere Ausnahmen von dem allgemein gültigen Rahmenwerk für die Verarbeitung zu Forschungszwecken (z. B. die Beschränkung der Speicherung), die ebenfalls berücksichtigt werden sollten. KI-Entwickler sollten sich jedoch des konkreten Rechtsrahmens bewusst sein, der für ihre Forschung gilt. Dieser kann je nach den nationalen Vorschriften wichtige Änderungen enthalten. Zu diesem Zweck wird eine Konsultation mit dem jeweiligen Datenschutzbeauftragten dringend empfohlen.

1.2.9 Definition der Speicherungsrichtlinien

Gemäß Artikel 5 Absatz 1 Buchstabe e DSGVO sollten personenbezogene Daten „in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es für die Zwecke, für die sie verarbeitet werden, erforderlich ist“. Dieses Erfordernis umfasst zwei Aspekte. Auf der einen Seite geht es um die Identifizierung: Personenbezogene Daten sollten in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es erforderlich ist. Folglich sollten die KI-Entwickler Richtlinien festlegen, um eine Identifizierung zu vermeiden, sobald sie für die Verarbeitung nicht mehr erforderlich ist. Dies umfasst die Ergreifung geeigneter Maßnahmen, um sicherzustellen, dass zu jedem Zeitpunkt nur **das Mindestmaß an Identifizierung verwendet wird, das für die Erfüllung der Zwecke erforderlich ist** (siehe Abschnitt „Speicherbegrenzung“, Kapitel „Grundsätze“).

Andererseits bedeutet Datenspeicherung, dass Daten nur für einen **begrenzten Zeitraum** gespeichert werden dürfen: den Zeitraum, der für die Zwecke, für die die Daten verarbeitet werden, unbedingt erforderlich ist. Die DSGVO erlaubt jedoch eine „längere Speicherung“, wenn der ausschließliche Zweck die wissenschaftliche Forschung ist (oder für im öffentlichen Interesse liegende Archivzwecke oder für wissenschaftliche und historische Forschungszwecke oder für statistische Zwecke) (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“ Kapitel „Konzepte“).

Im Falle der KI-Entwicklung birgt diese Ausnahme das Risiko, dass die Entwickler beschließen, die Daten länger als unbedingt erforderlich zu speichern und so sicherzustellen, dass sie für andere Zwecke als die ursprünglichen, für die sie erhoben wurden, zur Verfügung stehen. Die Verantwortlichen sollten sich darüber im Klaren sein, dass sie, auch wenn die DSGVO eine längere Speicherung zulässt, einen guten Grund für einen solchen längeren Zeitraum haben müssen (siehe Unterabschnitt „Zeitlicher Aspekt“, Abschnitt

[524] Dieses besondere Rahmenwerk umfasst auch historische Forschungszwecke oder statistische Zwecke. Die IKT-Forschung ist jedoch in der Regel nicht auf diese Zwecke ausgerichtet. Daher werden wir sie nicht analysieren.

„Speicherbegrenzung“, Kapitel „Grundsätze“). Sofern sich aus der Nichtbeachtung des Grundsatzes der Zweckbindung ein echtes Risiko ergibt, **sollte die Prüfung der Vereinbarkeit Teil jeder möglichen Weiterverwendung der Daten sein**.

Es scheint die Absicht des Gesetzgebers gewesen zu sein, von einer unbegrenzten Speicherung selbst im Falle dieser Sonderregelung abzuraten, um zu verhindern, dass die wissenschaftliche Forschung als Vorwand für eine längere Speicherung zu anderen, privaten Zwecken dient. Im Zweifelsfall sollte der Verantwortliche prüfen, ob eine neue Rechtsgrundlage angemessen ist. Daher sollten die Speicherfristen in einem angemessenen Verhältnis zu den Zielen der Verarbeitung stehen: „Bei der Festlegung der Speicherfristen sollten Kriterien wie Länge und Zweck der Forschung berücksichtigt werden. Es sei darauf hingewiesen, dass nationale Vorschriften ebenfalls Bestimmungen über die Speicherfrist enthalten können.“[525]

Wenn die Verantwortlichen die Daten somit nicht benötigen und es keine zwingenden rechtlichen Gründe gibt, die sie zur Speicherung derselben verpflichten, sollten sie diese besser anonymisieren oder löschen. Forscher sollten ihre Datenschutzbeauftragten konsultieren, wenn sie Daten über einen längeren Zeitraum speichern wollen, und die geltenden nationalen Vorschriften berücksichtigen. Dies könnte auch ein hervorragender Zeitpunkt sein, **um Fristen für die Löschung der verschiedenen Datenkategorien vorzusehen und diese Entscheidungen zu dokumentieren** (siehe Abschnitt „Rechenschaftspflicht“, Kapitel „Grundsätze“).

1.2.10 Benennung eines Datenschutzbeauftragten

Die Benennung eines DSB ist eine der besten Maßnahmen, die Verantwortliche ergreifen können, um die Einhaltung der Rechte der betroffenen Personen zu gewährleisten. Die Benennung eines DSB ist keine zwingende Folge des Einsatzes von KI-Tools. Die Benennung eines DSB ist unbestreitbar nur dann obligatorisch, wenn die in Artikel 37 Absatz 1 Buchstabe b) oder c) festgelegten Bedingungen erfüllt sind. Es ist also nicht so, dass alle KI-Entwickler einen DSB benennen müssen. **Eine derartige Benennung empfiehlt sich jedoch immer, zumindest aus Gründen der Transparenz** (siehe Abschnitt „Transparenz“, Kapitel „Grundsätze“)**.**

In jedem Fall sollten die Verantwortlichen die Rolle des DSB in Bezug auf das Gesamtmanagement des Projekts erläutern und sicherstellen, dass die Rolle des DSB nicht am Rande steht, sondern in die Entscheidungsprozesse der Organisation/des Projekts eingebunden ist. Sie sollten deutlich machen, wie diese Rolle in Bezug auf Aufsicht, Entscheidungsfindung und Ähnliches aussehen könnte.

2 Datenverständnis

[525] Leitlinien 3/2020 für die Verarbeitung von Gesundheitsdaten für wissenschaftliche Forschungszwecke im Zusammenhang mit dem COVID-19-Ausbruch. Angenommen am 21. April 2020- Europäischer Datenschutzbeauftragter, Brüssel, Seite 10 Verfügbar unter https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf (abgerufen am 23. April 2020).

„Die Phase des Datenverständnisses beginnt mit einer ersten Datenerhebung. Der Analytiker fährt dann fort, sich mit den Daten vertraut zu machen, Probleme in Bezug auf die Datenqualität zu erkennen, erste Erkenntnisse über die Daten zu gewinnen oder interessante Teilmengen zu entdecken, um Hypothesen über verborgene Informationen zu bilden. Die Phase des Datenverständnisses umfasst vier Schritte, darunter die Erhebung von Ausgangsdaten, die Beschreibung der Daten, die Untersuchung der Daten und die Verifizierung der Datenqualität.“[526]

2.1 Beschreibung

In dieser Phase werden die ersten Daten erhoben und eine erste Untersuchung derselben durchgeführt. Sie umfasst vier aufeinanderfolgende Aufgaben:

- Erhebung von Ausgangsdaten
- Beschreibung der Daten
- Untersuchung der Daten
- Verifizierung der Datenqualität

Alle diese Aufgaben zielen darauf ab, die verfügbaren Daten zu ermitteln. In dieser Phase müssen sich die Entwickler über die Daten, mit denen sie arbeiten müssen, im Klaren sein und Entscheidungen über die Art und Weise treffen, in der die wichtigsten Grundsätze des Datenschutzes umgesetzt werden sollen.

2.2 Wichtigste zu ergreifende Maßnahmen

In diesem Stadium gibt zahlreiche grundlegende Fragen im Zusammenhang mit dem Schutz personenbezogener Daten, die behandelt werden müssen. Je nach den getroffenen Entscheidungen müssen Grundsätze wie Datenminimierung, Datenschutz durch Technikgestaltung oder durch datenschutzfreundliche Voreinstellungen, Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz usw. in angemessener Weise geregelt werden.

2.2.1 Art der gesammelten Daten

Laut DSGVO trifft der Verantwortliche „geeignete technische und organisatorische Maßnahmen, die sicherstellen, dass durch Voreinstellung nur personenbezogene Daten, deren Verarbeitung für den jeweiligen bestimmten Verarbeitungszweck erforderlich ist, verarbeitet werden. Diese Verpflichtung gilt für die Menge der erhobenen personenbezogenen Daten, den Umfang ihrer Verarbeitung, ihre Speicherfrist und ihre Zugänglichkeit. Solche Maßnahmen müssen insbesondere sicherstellen, dass personenbezogene Daten durch Voreinstellungen nicht ohne Eingreifen der Person einer unbestimmten Zahl von natürlichen Personen zugänglich gemacht werden.”[527] (siehe Abschnitt „Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen” Kapitel „Konzepte“). Dies ist insbesondere

---

[526] C. Shearer (2000), The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing* 5(4): 13–23, Seite 15. Verfügbar unter: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (abgerufen am 15. Mai 2020).

[527] Artikel 25 DSGVO.

während dieser Phase zu beachten, da zu diesem Zeitpunkt häufig Entscheidungen über die Art der zu verwendenden Daten getroffen werden.

Die Verantwortlichen sollten in diesem Zusammenhang bedenken, dass es immer besser ist, die Verwendung personenbezogener Daten nach Möglichkeit zu vermeiden. Nach dem Grundsatz der Datenminimierung sollte die Verwendung personenbezogener Daten dem Zweck angemessen und erheblich sowie auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein. **Wenn also derselbe Zweck auch ohne die Verwendung personenbezogener Daten erreicht werden könnte, sollte die Verarbeitung vermieden werden.**

Eine zweite Vorsichtsmaßnahme besteht darin, **dass Entwickler, wenn sie personenbezogene Daten verwenden müssen, die Verwendung von Daten besonderer Kategorien vermeiden sollten.** Manchmal ist dies machbar, manchmal nicht. Dies hängt oft vom Anwendungsbereich des Modells ab. Die Arbeit an einem Modell, das für die Analyse des Einflusses der Epigenetik auf die menschliche Gesundheit verwendet oder das zur Überwachung eines Epidemieausbruchs eingesetzt werden soll, unterscheidet sich von der an einem Modell, das dazu dienen soll, Werbung zielgenau zu platzieren. Sollten Daten solcher besonderer Kategorien schließlich verwendet werden, müssen die Verantwortlichen die für ihre Verarbeitung geltenden Vorschriften und die notwendige Anwendung geeigneter Garantien berücksichtigen, die die Rechte, Interessen und Freiheiten der betroffenen Personen schützen können. Die Verhältnismäßigkeit zwischen dem Ziel der Forschung und der Verwendung besonderer Kategorien von Daten muss gewährleistet sein. Darüber hinaus müssen die Verantwortlichen sicherstellen, dass die Vorschriften ihrer Mitgliedstaaten genetische, biometrische und Gesundheitsdaten nicht durch die Einführung zusätzlicher Bedingungen oder Beschränkungen schützen, da die DSGVO sie dazu ermächtigt.

Wenn personenbezogene Daten erforderlich sind, sollte der KI-Entwickler zumindest versuchen, die Menge der zu berücksichtigenden Daten so weit wie möglich zu reduzieren (siehe Abschnitt „Datenminimierung“, Kapitel „Grundsätze“). Sie sollten immer daran denken, dass sie Daten nur dann verarbeiten dürfen, wenn die Verarbeitung angemessen und erheblich ist. Daher sollten sie es vermeiden, übermäßig viele personenbezogene Daten zu verwenden. Allzu oft ist dies leichter zu bewerkstelligen, als es scheint. Die norwegische Datenschutzbehörde stellte fest: „Es ist erwähnenswert, dass die Qualität der Trainingsdaten sowie die verwendeten Merkmale in vielen Fällen wesentlich wichtiger sein können als die Quantität. Beim Training eines Modells ist es wichtig, dass die Auswahl der Trainingsdaten repräsentativ für die später zu lösende Aufgabe ist. Riesige Datenmengen sind wenig hilfreich, wenn sie nur einen Bruchteil dessen abdecken, was das Modell später bearbeiten soll.“[528] Deshalb ist es besonders wichtig, **keine unnötigen Daten zu erheben**. Eine korrekte Kennzeichnung könnte ein gutes Mittel gegen unnötige Datenerfassung sein. **Wenn die Daten bereits gespeichert sind, beinhaltet die Auswahl zudem das Löschen unnötiger Datenelemente.**

Entwickler sollten nach Möglichkeit den „Fluch der Dimensionalität“ vermeiden, d. h. „Algorithmen mit einer schlechten Leistung und hohen Komplexität aufgrund von Datenrahmen mit zahlreichen Dimensionen/Merkmalen, die die Zielfunktion häufig recht komplex machen und zu Modell-Overfitting führen können, solange der Datensatz häufig auf

[528] Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 15. Mai 2020).

der unteren Dimensionalitätsebene liegt."[529] Zu diesem Zweck **könnte ein Expert extrem hilfreich sein, der in der Lage ist, relevante Merkmale auszuwählen.** So könnte die Menge der verwendeten personenbezogenen Daten ohne Qualitätseinbußen erheblich reduziert werden. Dies sollte nicht schwierig sein, wenn der Datenwissenschaftler mit dem Datensatz und der Bedeutung seiner numerischen Merkmale gut vertraut ist. Unter solchen Bedingungen ließe sich einfach bestimmen, ob einige der Variablen benötigt werden oder nicht. Ein solcher Ansatz ist jedoch nur dann möglich, wenn der Datensatz leicht zu interpretieren ist und die Abhängigkeiten zwischen den Variablen bekannt sind. Daher benötigt der Entwickler eine geringere Menge an Daten, wenn diese angemessen klassifiziert wurden. Intelligente Daten könnten viel nützlicher sein als Big Data. Natürlich ist dies unter Umständen mit einem enormen Aufwand für die Vereinheitlichung, Homogenisierung usw. verbunden, aber es wird die Umsetzung des Grundsatzes der Datenminimierung wesentlich effizienter gestalten.

Darüber hinaus sollten die Verantwortlichen versuchen, die **Auflösung der Daten auf das für die mit der Verarbeitung verfolgten Zwecke erforderliche Mindestmaß zu beschränken**. Sie sollten zudem **ein optimales Niveau der Datenaggregation bestimmen,** bevor sie mit der Verarbeitung beginnen (siehe Abschnitt „Angemessen, erheblich und beschränkt", Abschnitt „Datenminimierung", Kapitel „Grundsätze").

Die Datenminimierung kann im Falle des Deep Learning komplex sein, da eine Unterscheidung nach Merkmalen unter Umständen nicht möglich ist. Es gibt jedoch eine effiziente Möglichkeit, die Menge der erhobenen Daten zu regulieren und sie nur dann zu erhöhen, wenn es notwendig erscheint: die Lernkurve[530]. Der Entwickler sollte damit beginnen, eine begrenzte Menge an Trainingsdaten zu erheben und zu verwenden, und dann die Genauigkeit des Modells überwachen, während es mit neuen Daten gefüttert wird.

Kasten 16: Nicht angemessen umgesetzte Datenminimierung

Ein von der norwegischen Steuerverwaltung entwickeltes Tool zum Filtern von Steuererklärungen auf Fehler testete in der Trainingsphase 500 Variablen. In das endgültige KI-Modell wurden jedoch nur 30 aufgenommen, da sie sich für die gestellte Aufgabe als besonders relevant erwiesen hatten. Somit hätte man wahrscheinlich die Erhebung derart vieler personenbezogener Daten vermeiden können, wenn man von Anfang an eine bessere Auswahl der relevanten Variablen getroffen hätte.[531]

2.2.2 Auswahl der geeigneten Rechtsgrundlage für die Verarbeitung

Die Verantwortlichen sollten sich vor Beginn der Verarbeitung für eine Rechtsgrundlage entscheiden, ihre Entscheidung in einer Datenschutzerklärung (zusammen mit den Zwecken) dokumentieren und die Gründe für ihre Entscheidung angeben (siehe Abschnitt „Rechenschaftspflicht", Kapitel „Grundsätze"). Grundsätzlich sollten sie die **Rechtsgrundlage wählen, die der wahren Natur ihrer Beziehung zu der betroffenen Person und dem Zweck der Verarbeitung am ehesten entspricht.** Diese Entscheidung ist

---

[529] H. Oliinyk (2018), Why and how to get rid of the curse of dimensionality right (with breast cancer dataset visualization). Towards Data Science, 20. März. Verfügbar unter: https://towardsdatascience.com/why-and-how-to-get-rid-of-the-curse-of-dimensionality-right-with-breast-cancer-dataset-7d528fb5f6c0 (abgerufen am 15. Mai 2020).

[530] R. Ng (kein Datum) Learning curve. Verfügbar unter: www.ritchieng.com/machinelearning-learning-curve/ (abgerufen am 15. Mai 2020).

[531] Norwegische Datenschutzbehörde (2018), Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 15. Mai 2020).

von grundlegender Bedeutung, da die Rechtsgrundlage für die Verarbeitung nur dann geändert werden darf, wenn es stichhaltige Gründe gibt, die dies rechtfertigen (siehe Abschnitt „Zweckbindung“, Kapitel „Grundsätze“).

Grundsätzlich ist die Einwilligung eine der häufigsten Rechtsgrundlagen für die Verarbeitung (siehe Abschnitt „Einwilligung“, Kapitel „Hauptkonzepte“). Dies ist jedoch mit gewissen Risiken verbunden. Die Einwilligung ist nämlich immer auf einen bestimmten Zweck beschränkt. Eine „Ausweitung“ der Verarbeitungszwecke über die ausdrückliche Einwilligung der betroffenen Personen hinaus stellt daher eine unrechtmäßige Verarbeitung dar. Um festzustellen, ob eine weitere Verarbeitung mit der ursprünglichen Verarbeitung vereinbar ist oder nicht, sollten die Verantwortlichen die in Artikel 6 Absatz 4 DSGVO angegebenen Kriterien heranziehen (siehe Unterabschnitt „Wann sind Zwecke vereinbar?“, Abschnitt „Zweckbindung“). Wie bereits erwähnt, gilt die Verarbeitung zu wissenschaftlichen oder historischen Forschungszwecken oder zu statistischen Zwecken nicht als unvereinbar mit den ursprünglichen Zwecken (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“, Abschnitt „Konzepte“).

Die häufigsten alternativen Gründe für die Verarbeitung von Daten im Rahmen der künstlichen Intelligenz sind berechtigte Interessen, Vertragserfüllung und rechtliche Verpflichtungen oder lebenswichtige Interessen. Alle diese Gründe weisen spezifische Merkmale auf, die sorgfältig analysiert werden müssen.

2.2.3 Überprüfung der rechtmäßigen Verwendung von Datensätzen

Datensätze können auf unterschiedliche Weise beschafft werden. Als erste Option kann sich der Entwickler dafür entscheiden, eine bereits von jemand anderen erstellte Datenbank oder den Zugriff darauf zu erwerben. In diesem Fall ist besondere Vorsicht seitens des Verantwortlichen geboten, da der Erwerb des Zugangs zu einer Datenbank mit zahlreichen rechtlichen Fragen verbunden ist (siehe Abschnitt „Erwerb des Zugangs zu einer Datenbank”, Kapitel „Maßnahmen und Tools”).[532]

Eine zweite und sehr gängige Alternative darin, eine Datenbank zu erstellen. Natürlich müssen die Verantwortlichen in diesem Fall sicherstellen, dass sie bei der Erstellung dieser Datenbank alle rechtlichen Anforderungen der DSGVO erfüllen (siehe Abschnitt „Erstellen einer Datenbank“, Kapitel „Maßnahmen und Tools“).

Drittens können Entwickler auch einen alternativen Weg wählen. Sie **mischen lizenzierte Daten von Dritten miteinander oder mit dem eigenen Datensatz der Verantwortlichen, um einen großen Trainingsdatensatz und einen weiteren für Validierungszwecke zu erstellen.** Dies könnte einige Probleme mit sich bringen, darunter die Möglichkeit, dass die Kombination dieser personenbezogenen Daten zusätzliche Informationen über die betroffenen Personen liefert. Unter Umständen könnte der Verantwortliche dadurch die betroffenen Personen identifizieren, was bisher nicht möglich war. Dies könnte bedeuten, dass anonymisierte Daten deanonymisiert und neue personenbezogene Informationen geschaffen werden, die im ursprünglichen Datensatz nicht enthalten waren, ein Umstand, der dramatische ethische und rechtliche Fragen aufwerfen würde. Daher muss die Re-Identifizierung mit Maßen wie k-Anonymity, l-Diversity oder t-Closeness getestet werden.[533]

---

[532] Z. K. Yeong (2019), Legal issues in AI deployment. Law Gazette, Februar. Verfügbar unter: https://lawgazette.com.sg/feature/legal-issues-in-ai-deployment/ (abgerufen am 15. Mai 2020).

[533] K. Rajendran, M. Jayabalan und M. E. Rana (2017), A study on k-anonymity, l-diversity, and t-closeness techniques focusing medical data, International Journal of Computer Science and Network Security 17(12): 172–177.

Ein weiteres häufiges Problem ist, dass die jeweilige Rechtsgrundlage für die Verarbeitung der in den einzelnen Datensätzen erfassten Daten unterschiedlich sein kann. Wenn ein Verantwortlicher die Datensätze zusammenführt und dann eine der Rechtsgrundlagen nicht mehr anwendbar ist, steht er vor einer schwierigen Situation. Wenn beispielsweise eine der Datenbanken auf der Grundlage einer Einwilligung aufgebaut wurde und einige der betroffenen Personen ihre Einwilligung widerrufen, muss der Verantwortliche sie aus dem zusammengeführten Datensatz löschen. Dies könnte in der Praxis sehr schwer zu bewerkstelligen sein.

Darüber hinaus können die auf diese Weise gewonnenen neuen Informationen auch auf Wahrscheinlichkeiten oder Vermutungen beruhen und daher falsch sein oder Verzerrungen bei der Darstellung von Personen enthalten (siehe Abschnitt „**Bestimmungen der DSGVO**“).[534] Daher sollten die Verantwortlichen solche Folgen nach Möglichkeit vermeiden, indem sie sicherstellen, dass die Zusammenführung von Datensätzen nicht gegen die Rechte und Interessen der betroffenen Personen verstößt.

Wenn die Verantwortlichen schließlich mehrere Datensätze verwenden, die unterschiedlichen Zwecken dienen, sollten sie geeignete Maßnahmen ergreifen, um die verschiedenen Verarbeitungstätigkeiten zu trennen. Andernfalls könnten sie leicht Daten, die zu einem bestimmten Zweck erhoben wurden, für andere Tätigkeiten verwenden, was zu Problemen im Zusammenhang mit dem Grundsatz der Zweckbindung führen könnte.

3 Datenaufbereitung

„Die Phase der Datenaufbereitung umfasst alle Aktivitäten zur Erstellung des endgültigen Datensatzes bzw. der Daten, die aus den anfänglichen Rohdaten in das/die Modellierungswerkzeug(e) eingespeist werden. Zu den Aufgaben gehören die Auswahl von Tabellen, Datensätzen und Attributen sowie die Transformation und Bereinigung der Daten für die Modellierungswerkzeuge. Die fünf Schritte der Datenaufbereitung umfassen die Auswahl der Daten, die Bereinigung der Daten, den Aufbau der Daten, die Integration der Daten und die Formatierung der Daten.“[535]

3.1 Beschreibung

Diese Phase umfasst alle erforderlichen Aktivitäten, um aus den anfänglichen Rohdaten den endgültigen Datensatz zu erstellen, der dann in das Modell eingespeist wird. Sie umfasst die folgenden fünf Aufgaben, die nicht unbedingt nacheinander ausgeführt werden müssen.

1. Auswahl der Daten. Festlegung der für die Analyse zu verwendenden Daten, auf der Grundlage der Relevanz für die Data-Mining-Ziele, der Qualität sowie der technischen Beschränkungen, wie z. B. Einschränkungen des Datenvolumens oder der Datentypen.

---

[534] SHERPA-Projekt (2019), Guidelines for the ethical development of AI and big data systems: an ethics by design approach. SHERPA, Seite 38. Verfügbar unter: www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (abgerufen am 15. Mai 2020).

[535] C. Shearer (2000), The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing* 5(4): 13–23, Seite 16. Verfügbar unter: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (abgerufen am 15. Mai 2020).

2. Bereinigung der Daten. Anhebung der Datenqualität auf das erforderliche Niveau, beispielsweise durch Auswahl sauberer Teilmengen der Daten, Einfügen von Standardwerten und Schätzung fehlender Daten durch Modellierung.
3. Aufbau der Daten. Der Aufbau neuer Daten durch die Erstellung abgeleiteter Attribute, neuer Datensätze oder transformierter Werte für vorhandene Attribute.
4. Integration der Daten. Kombinieren von Daten aus mehreren Tabellen oder Datensätzen, um neue Datensätze oder Werte zu erstellen.
5. Formatierung der Daten. Syntaktische Änderungen an Daten, die vom Modellierungswerkzeug benötigt werden.

3.2 Wichtigste zu ergreifende Maßnahmen

3.2.1 Gewährleistung der sachlichen Richtigkeit personenbezogener Daten

Laut der DSGVO müssen die Daten sachlich richtig sein (siehe Abschnitt „Richtigkeit“, Kapitel „Grundsätze“). Dies bedeutet, dass die Daten korrekt und auf dem neuesten Stand sind, bezieht sich aber auch auf die Genauigkeit der durchgeführten Analysen. Der EDSB hat die Bedeutung der Richtigkeit der Profilerstellung oder des (nicht ausschließlich) automatisierten Entscheidungsfindungsprozesses in allen Phasen (von der Erhebung der Daten bis zur Anwendung des Profils auf die Person) hervorgehoben.[536]

Die Verantwortlichen müssen die Richtigkeit personenbezogener Daten gewährleisten. Sobald sie die Erhebung personenbezogener Daten abgeschlossen haben, sollten sie daher geeignete Instrumente einsetzen, um die Richtigkeit der Daten zu gewährleisten. Dies beinhaltet im Wesentlichen die Umsetzung technischer und organisatorischer Maßnahmen, die die Anwendung dieses Grundsatzes gewährleisten (siehe Unterabschnitt „Verbundene technische und organisatorische Maßnahmen“, Abschnitt „Richtigkeit“, Kapitel „Grundsätze“). Wenn personenbezogene Daten von betroffenen Personen stammen, kann der Verantwortliche von deren Richtigkeit ausgehen (es sei denn, der Verantwortliche ist der Ansicht, dass die betroffene Person einen Grund haben könnte, unrichtige Daten anzugeben). Wurden die personenbezogenen Daten nicht bei der betroffenen Person erhoben, sind die Verantwortlichen verpflichtet, „die Richtigkeit der erhaltenen Daten zu überprüfen, zumindest im Hinblick auf ihre Eignung für die angegebenen Verarbeitungszwecke und auf etwaige negative Folgen, die Ungenauigkeiten für die betroffenen Personen haben können.“ (siehe Unterabschnitt „Wie wird die Ungenauigkeit von Daten entdeckt?“, Abschnitt „Richtigkeit“, Kapitel „Grundsätze“). In jedem Fall erfordert die Richtigkeit der Daten eine angemessene Umsetzung von Maßnahmen, damit betroffene Personen ihr Recht auf Berichtigung in

[536] *Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679* (wp251rev.01). 22.08.*2018, Seite* 13; Ross Ducato, Private Ordering of Online Platforms in Smart Urban Mobility The Case of Uber’s Rating System, CRIDES Working Paper Series no. 3/20202 Februar 2020 Aktualisiert am 26 Juli 2020, Seite 20–21, verfügbar unter: https://poseidon01.ssrn.com/delivery.php?ID=247104118003073117118086021112071111102048023015008020118084071112086000027097102088036101006014057116105116119119026079007006118044033055000114023106007076115096073024007094081002078064098028091093003078095099082108113086098120001079015123027083125024&EXT=pdf&INDEX=TRUE

Anspruch nehmen können (siehe Abschnitt „Recht auf Berichtigung“, Kapitel „Rechte der betroffenen Person“).

3.2.2 Konzentration auf Profiling-Probleme

Im Falle einer Datenbank, die ein KI-Tool trainieren oder validieren soll, besteht eine besonders wichtige Verpflichtung, die betroffenen Personen darüber zu informieren, **dass sie betreffende Daten zu einer automatisierten Entscheidungsfindung oder einem Profiling führen könnten, es sei denn, die Verantwortlichen können gewährleisten, dass das Tool in keiner Weise zu diesen Folgen führen wird.**

Auch wenn eine automatisierte Entscheidungsfindung im Rahmen der Forschung kaum vorkommen kann, sollten die Entwickler diesem Punkt Aufmerksamkeit schenken.

Profiling hingegen kann bei der KI-Entwicklung einige Probleme mit sich bringen.

Dafür gibt es einen einfachen Grund: „Vom Profiling bekommen die betroffenen Personen selbst oft gar nichts mit. Denn dabei werden abgeleitete oder hergeleitete Daten zu einzelnen Personen erzeugt, d. h. „neue“ personenbezogene Daten, die von den Betroffenen selbst gar nicht direkt zur Verfügung gestellt wurden. Die Menschen haben in dieser Thematik einen unterschiedlichen Kenntnisstand und für manche sind die komplizierten Techniken, die bei Profiling und automatisierten Entscheidungen zur Anwendung kommen, möglicherweise schwer verständlich.“[537] Daher gilt: „Schwebt dem Verantwortlichen somit ein „Modell“ mit ausschließlich automatisierten Entscheidungen vor, die erhebliche Auswirkungen auf die jeweils betroffenen Personen haben und sich auf zu diesen erstellte Profile stützen, und kann er sich nicht auf die Einwilligung der betroffenen Person, einen Vertrag mit ihr oder ein Gesetz berufen, wonach dies zulässig ist, sollte er von diesem Modell Abstand nehmen.“[538] Das Risiko für die Rechte, Interessen und Freiheiten der betroffenen Person ist ein sehr wichtiger Faktor, der immer berücksichtigt werden muss. Es ist nicht dieselbe Art von Profiling, eine Entscheidung über beliebte Fernsehserien einer Person zu treffen, im Vergleich zu einem Profiling, das der Genehmigung ihrer Krankenversicherungspolice gewidmet ist. Wenn also die Verarbeitung Risiken für die Grundrechte und -freiheiten von Personen birgt, müssen die Verantwortlichen sicherstellen, dass sie diese Risiken bewältigen und die Anforderungen erfüllen können.

Wenn eine Verarbeitung und/oder automatisierte Entscheidungsfindung stattfindet, **müssen die betroffenen Personen angemessen über diese Verarbeitung und die Funktionsweise des Algorithmus informiert werden.** Mit anderen Worten: Ihr Recht auf Auskunft muss in Anwendung des Grundsatzes der Rechtmäßigkeit, der Verarbeitung nach Treu und Glauben und der Transparenz gewahrt werden. Somit muss der Verantwortliche zumindest „der betroffenen Person mitteilen, dass er dies tut, aussagekräftige Informationen über die involvierte Logik liefern und die Tragweite und die angestrebten Auswirkungen der Verarbeitung erläutern.“[539]

Die Informationen über die Logik eines Systems und die Erläuterungen zu den Entscheidungen sollten der betroffenen Person den notwendigen Kontext für die Entscheidung geben, ob und aus welchen Gründen das Eingreifen einer Person wünscht. In

---

[537] Artikel-29-Datenschutzgruppe (2017) Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679. Angenommen am 3. Oktober 2017. Zuletzt überarbeitet und angenommen am 6. Februar 2018. Europäische Kommission, Brüssel, Seite 9. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (abgerufen am 15. Mai 2020)

[538] Ibid., Seite 30.

[539] Ibid., Seiten13–14.

einigen Fällen können unzureichende Erläuterungen dazu führen, dass die betroffene Person unnötigerweise auf andere Rechte zurückgreift oder ihre Einwilligung widerruft. Anträge auf Eingreifen, Darlegung des eigenen Standpunkts oder Anfechtung sind wahrscheinlicher, wenn die betroffene Person das Gefühl hat, nicht ausreichend zu verstehen, wie die Entscheidung zustande gekommen ist.[540]

Schließlich müssen Verantwortliche stets bedenken, dass gemäß Artikel 22 Absatz 3 automatisierte Entscheidungen, die besondere Kategorien personenbezogener Daten betreffen, nur dann zulässig sind, wenn die betroffene Person eingewilligt hat oder wenn sie auf einer gesetzlichen Grundlage erfolgen (siehe Abschnitt „**Vorrang menschlichen Handelns und menschlicher Aufsicht**"). Diese Ausnahme gilt nicht nur, wenn die entsprechenden Daten in diese Kategorie fallen, sondern auch dann, **wenn der Abgleich verschiedener Arten personenbezogener Daten sensible Informationen über Personen offenbaren kann oder wenn abgeleitete Daten in diese Kategorie fallen.** In all diesen Fällen muss man von der Verarbeitung besonderer Kategorien personenbezogener Daten sprechen. Eine Studie, die in der Lage ist, besondere Datenkategorien herzuleiten, unterliegt nämlich denselben rechtlichen Verpflichtungen gemäß der DSGVO, als ob von Anfang an sensible personenbezogene Daten verarbeitet worden wären. Wenn bei der Profilerstellung personenbezogene Daten hergeleitet werden, die nicht von der betroffenen Person zur Verfügung gestellt wurden, sollten die Verantwortlichen sicherstellen, dass die Verarbeitung nicht mit dem ursprünglichen Zweck unvereinbar ist, dass eine Rechtsgrundlage für die Verarbeitung der besonderen Datenkategorie ermittelt wurde und dass die betroffene Person über die Verarbeitung unterrichtet wird.[541]

Kasten 17: Beispiel für die Herleitung von Daten besonderer Kategorien

Eine Untersuchung aus dem Jahr 2011 zeigte, dass leicht zugängliche digitale Verhaltensaufzeichnungen – „Likes" bei Facebook – zur automatischen und genauen Vorhersage einer Reihe hochsensibler persönlicher Merkmale verwendet werden können, darunter: sexuelle Orientierung, ethnische Zugehörigkeit, religiöse und politische Ansichten, Persönlichkeitsmerkmale, Intelligenz, Glück, Konsum von Suchtmitteln, Trennung der Eltern, Alter und Geschlecht. Die Analyse basierte auf einem Datensatz von über 58.000 Freiwilligen und deren „Likes" bei Facebook, detaillierten demografischen Profilen und den Ergebnissen mehrerer psychometrischer Tests. Das Modell unterschied in 88 % der Fälle korrekt zwischen homosexuellen und heterosexuellen Männern, in 95 % der Fälle zwischen Afroamerikanern und kaukasischen Amerikanern und in 85 % der Fälle zwischen Demokraten und Republikanern. In Bezug auf das Persönlichkeitsmerkmal „Offenheit" lag die Vorhersagegenauigkeit nahe an der Test-Retest-Genauigkeit eines Standard-Persönlichkeitstests. Die Autoren lieferten Beispiele für Assoziationen zwischen Attributen und Likes und erörtern die Auswirkungen auf die Online-Personalisierung und den Datenschutz.[542]

---

[540] ICO (2020) AI auditing framework – draft guidance for consultation. Information Commissioner's Office, Wilmslow, Seiten 94. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[541] Article 29 Working Party (2018) Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679. Angenommen am 3. Oktober 2017. Zuletzt überarbeitet und angenommen am 6. Februar 2018. Europäische Kommission, Brüssel, Seite 15. Verfügbar unter: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053 (abgerufen am 15. Mai 2020).

[542] M. Kosinski, D. Stillwell und T Graepel (2013), Digital records of behavior expose personal traits, *Proceedings of the National Academy of Sciences* 110 (15): 5802-5805, DOI: 10.1073/pnas.1218772110.

Die Durchführung einer DSFA ist zwingend erforderlich, wenn ein reales Risiko einer unzulässigen Profilerstellung oder automatisierten Entscheidungsfindung besteht. Gemäß Artikel 35 Absatz 3 Buchstabe a DSGVO muss der Verantwortliche eine DSFA durchführen, wenn es sich um eine systematische und umfassende Bewertung persönlicher Aspekte natürlicher Personen handelt, die sich auf automatisierte Verarbeitung einschließlich Profiling gründet und die ihrerseits als Grundlage für Entscheidungen dient, die Rechtswirkung gegenüber natürlichen Personen entfalten oder diese in ähnlich erheblicher Weise beeinträchtigen. Die Verantwortlichen sollten sich darüber im Klaren sein, dass zum gegenwärtigen Zeitpunkt jedes Land dem EDSA eine Liste vorgelegt hat, in der festgelegt ist, wann eine DSFA erforderlich ist. Wenn sich der Verantwortliche innerhalb des EWR befindet, sollte diese Liste auch vor Ort überprüft werden.[543]

Gemäß Artikel 37 Absatz 1 Buchstabe b und Absatz 5 DSGVO müssen Verantwortliche einen Datenschutzbeauftragten benennen, wenn „die Kerntätigkeit des Verantwortlichen oder des Auftragsverarbeiters in der Durchführung von Verarbeitungsvorgängen besteht, welche aufgrund ihrer Art, ihres Umfangs und/oder ihrer Zwecke eine umfangreiche regelmäßige und systematische Überwachung von betroffenen Personen erforderlich machen.“ Die Verantwortlichen müssen außerdem im Rahmen ihrer Rechenschafts- und Dokumentationspflicht alle Entscheidungen, die von einem KI-Tool getroffen werden, aufzeichnen. Darin sollte auch festgehalten werden, ob eine Person um das Eingreifen einern Person gebeten, einen eigenen Standpunkt dargelegt, die Entscheidung angefochten hat und ob eine Entscheidung daraufhin geändert worden ist.[544] (siehe Abschnitt „Rechenschaftspflicht“, Kapitel „Grundsätze“).

Folgende zusätzliche Maßnahmen können ebenfalls äußerst nützlich sein, um eine automatisierte Entscheidungsfindung zu vermeiden:[545]

- Berücksichtigen Sie bereits **in der Designphase die Systemanforderungen, die zur Unterstützung einer aussagekräftigen Überprüfung durch Personen erforderlich sind.** Insbesondere die Anforderungen an die Interpretierbarkeit und die effektive Gestaltung der Benutzeroberfläche zur Unterstützung von Überprüfungen und Eingriffen durch Personen.
- Gestalten und setzen Sie eine geeignete Schulung und Unterstützung für menschliche Prüfer um.
- Geben Sie dem Personal die entsprechenden Befugnisse, Anreize und Unterstützung, damit es die Bedenken einzelner Personen ansprechen oder eskalieren und gegebenenfalls die Entscheidung des KI-Tools außer Kraft setzen kann.

---

[543] EDSA (2019) Data Protection Impact Assessment. Europäischer Datenschutzbeauftragter, Brüssel. Verfügbar unter: https://EDSA.europa.eu/our-work-tools/our-documents/topic/data-protection-impact-assessment-dpia_de (abgerufen am 3. Juni 2020).

[544] ICO (2020) Guidance on the AI auditing framework – draft guidance for consultation. Information Commissioner's Office, Wilmslow, Seiten 94–95. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[545] Ibid., Seite 95.

3.2.3 Auswahl unverzerrter Daten

Verzerrungen sind eines der Hauptprobleme bei der Entwicklung von KI, ein Problem, das dem Grundsatz der Fairness zuwiderläuft. Verzerrungen können durch viele verschiedene Faktoren verursacht werden. Bei der Erfassung der Daten können sozial bedingte Verzerrungen, Ungenauigkeiten, Fehler und andere Mängel auftreten. Manchmal kann es auch vorkommen, dass Datensätze durch böswillige Handlungen verzerrt werden. Die Eingabe schädlicher Daten in ein KI-Tool kann sein Verhalten verändern, insbesondere bei selbstlernenden Systemen.[546] Im Fall des Chatbot Tay von Microsoft beispielsweise begann eine große Zahl von Internetnutzern rassistische und sexistische Kommentare zu posten, die dem Algorithmus als Nahrung dienten. Das Ergebnis war, dass Tay nach nur wenigen Stunden Betrieb anfing, ebenfalls rassistische und sexistische Tweets zu senden. In anderen Fällen besteht das Hauptproblem darin, dass der Datensatz die betrachtete Population und den beabsichtigten Zweck nicht gut abbildet. Daher enthält er versteckte Verzerrungen, die auf das trainierte Tool übertragen werden, das diese Verzerrungen widerspiegelt, was dazu führen kann, dass die Ergebnisse des Modells falsch oder diskriminierend sind.[547]

Daher werfen Fragen im Zusammenhang mit der Zusammensetzung der für das Training verwendeten Datenbanken entscheidende ethische und rechtliche Fragen auf, nicht nur Fragen der Effizienz oder technischer Natur. Diese Fragen müssen somit vor dem Training des Algorithmus geklärt werden. KI-Modelle müssen „anhand relevanter und korrekter Daten trainiert werden und lernen, welche Daten sie betonen sollten. Das Modell darf keine Informationen über Rasse oder ethnische Herkunft, politische Meinung, Religion oder Weltanschauung, Gewerkschaftszugehörigkeit, genetischen Status, Gesundheitszustand oder sexuelle Orientierung betonen, wenn dies zu einer willkürlichen diskriminierenden Verarbeitung führen würde.“[548] Erkennbare und diskriminierende Verzerrungen sollten nach Möglichkeit bereits in der Phase der Datensatzerstellung eliminiert werden.

Kasten 18: Verzerrungen verstehen: der Gorilla-Fall

2015 machte der Software-Ingenieur Jacky Alciné darauf aufmerksam, dass die in Google Photos verwendeten Bilderkennungsalgorithmen einige schwarze Menschen als „Gorillas“ identifizierten. Google erkannte das Problem sofort und versprach, es zu beheben.

Diese Panne beruhte auf einem schwerwiegenden Fehler in der Trainingsphase. Der Algorithmus wurde darauf trainiert, Menschen anhand eines Datensatzes zu erkennen, der hauptsächlich aus Fotos weißer Menschen bestand. Infolgedessen nahm der Algorithmus an, dass eine schwarze Person dem Objekt „Gorilla“, auf das er trainiert worden war, sehr viel ähnlicher sei als dem Objekt „Mensch“. Dieses Beispiel zeigt sehr gut, wie wichtig die Auswahl der Daten für das Training ist.

---

[546] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI. Europäische Kommission, Brüssel, Seite 17. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

[547] Für eine Definition von unmittelbarer und mittelbarer Diskriminierung siehe beispielsweise Artikel 2 der Richtlinie 2000/78/EG des Rates vom 27. November 2000 zur Festlegung eines allgemeinen Rahmens für die Verwirklichung der Gleichbehandlung in Beschäftigung und Beruf. Siehe auch Artikel 21 der Charta der Grundrechte der Europäischen Union.

[548] Norwegische Datenschutzbehörde (2018) Artificial intelligence and privacy. Norwegische Datenschutzbehörde, Oslo. Verfügbar unter: https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf (abgerufen am 15. Mai 2020).

Damit ethische Anforderungen in dieser Phase integriert werden können, sollte der KI-Entwickler somit die ethischen Folgen der Datenauswahl in Bezug auf die Diversität bewerten und gegebenenfalls Änderungen vornehmen. Der Verantwortliche sollte nämlich „geeignete mathematische oder statistische Verfahren für das Profiling verwenden, technische und organisatorische Maßnahmen treffen, mit denen in geeigneter Weise insbesondere sichergestellt wird, dass **Faktoren, die zu unrichtigen personenbezogenen Daten führen, korrigiert werden und das Risiko von Fehlern minimiert wird, und personenbezogene Daten in einer Weise sichern,** dass den potenziellen Bedrohungen für die Interessen und Rechte der betroffenen Person Rechnung getragen wird und unter anderem verhindern, dass es gegenüber natürlichen Personen aufgrund von Rasse, ethnischer Herkunft, politischer Meinung, Religion oder Weltanschauung, Gewerkschaftszugehörigkeit, genetischer Anlagen oder Gesundheitszustand sowie sexueller Orientierung zu diskriminierenden Wirkungen oder zu einer Verarbeitung kommt, die eine solche Wirkung hat.“[549]

Die Verantwortlichen sollten sich stets vor Augen halten, dass dieses Problem sich insbesondere dadurch auszeichnet, dass die Auswahl eines Datensatzes für das Training mitunter fast unbewusste Entscheidungen beinhaltet (während die Programmierung eines traditionellen, deterministischen Algorithmus immer ein bewusster Vorgang ist). Derjenige, der einen Algorithmus trainiert, baut in gewisser Weise seine eigene Weltanschauung, seine Werte oder zumindest die Werte in ihn ein, die mehr oder weniger direkt in den aus der Vergangenheit gesammelten Daten enthalten sind.[550] **Die mit der Auswahl der in die Datensätze zu integrierenden Daten betrauten Teams sollten somit aus Personen bestehen, die die Diversität gewährleisten, die von der KI-Entwicklung erwartet wird.** In jedem Fall könnte juristisches Fachwissen über Antidiskriminierungsvorschriften in diesem Punkt von Bedeutung sein.

4 Modellierung (Training)

„In dieser Phase werden verschiedene Modellierungstechniken ausgewählt und angewandt sowie deren Parameter auf optimale Werte kalibriert. In der Regel gibt es für ein und dieselbe Art von Data-Mining-Problem mehrere Techniken. Einige Techniken haben spezifische Anforderungen an die Form der Daten. Daher kann es notwendig sein, zur Phase der Datenaufbereitung zurückzukehren. Die Modellierungsschritte umfassen die Auswahl der Modellierungstechnik, die Erstellung des Testdesigns, die Erstellung der Modelle und die Bewertung der Modelle.“[551]

4.1 Beschreibung

Diese Phase umfasst mehrere wichtige Aufgaben. Im Allgemeinen muss der Entwickler:

---

[549] Erwägungsgrund 71 DSGVO.

[550] CNIL (2017), How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence. Commission Nationale de l'Informatique et des Libertés, Paris, Seite 34. Verfügbar unter: www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (abgerufen am 15. Mai 2020).

[551] C. Shearer (2000), The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing* 5(4): 13–23, Seite 17. Verfügbar unter: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (abgerufen am 15. Mai 2020).

- **Die zu verwendende Modellierungstechnik auswählen.** Je nach Art der Technik sind Folgen wie Herleitung, Unklarheit oder Verzerrungen von Daten mehr oder weniger wahrscheinlich.
- **Das zu verwendende Trainingstool festlegen.** Mit diesem Tool kann der Entwickler messen, wie gut das Modell die Vergangenheit vorhersagen kann, bevor er es für die Vorhersage der Zukunft einsetzt. Zum Training gehören immer empirische Tests mit personenbezogenen Daten. Manchmal testen die Entwickler das Modell mit Daten, die sich von den zur Erstellung des Modells verwendeten Daten unterscheiden. Daher kann man in dieser Phase von verschiedenen Arten von Datensätzen sprechen. Unter Umständen kann es schwierig sein, die Personen zu identifizieren, auf die sich die Trainingsdaten beziehen. Dies führt zu Problemen bei der Wahrung der Rechte dieser Personen, was geeigneter Weise angegangen werden sollte.

4.2 Wichtigste zu ergreifende Maßnahmen

4.2.1 Umsetzung des Grundsatzes der Datenminimierung

Nach dem Grundsatz der Zweckbindung sollten die Verantwortlichen, die KI-Tools einsetzen, den Zweck der Nutzung des KI-Tools zu Beginn des Trainings oder Einsatzes festlegen und diese Festlegung neu bewerten, falls die Verarbeitung des Systems unerwartete Ergebnisse liefert, da er verlangt, dass personenbezogene Daten nur für „festgelegte, eindeutige und legitime Zwecke“ erhoben und nicht in einer Weise verwendet werden, die mit dem ursprünglichen Zweck unvereinbar ist.

Nach dem Grundsatz der Datenminimierung müssen die Verantwortlichen die Menge der bereitgestellten Daten und/oder den Umfang der Informationen über die betroffene Person so schnell wie möglich auf das notwendige Maß beschränken. Folglich müssen die in der Trainingsphase verwendeten Daten von allen Informationen bereinigt werden, die nicht unbedingt für das Training des Modells erforderlich sind (siehe Unterabschnitt „Zeitlicher Aspekt“, Abschnitt „Datenminimierung“, Kapitel „Grundsätze“). Es gibt mehrere Strategien, um die Datenminimierung in der Trainingsphase zu gewährleisten. Die Techniken entwickeln sich ständig weiter. Im Folgenden werden jedoch einige der gängigsten genannt;[552] siehe auch Abschnitt „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“.):

- Analyse der von den Daten zu erfüllenden Bedingungen, um als qualitativ hochwertig und mit einer hohen Vorhersagekraft für die spezifische Anwendung zu gelten.
- Kritische Analyse des Umfangs der in jeder Phase des KI-Tools verwendeten Datentypologie.
- Löschung von unstrukturierten Daten und unnötigen Informationen, die bei der Vorverarbeitung der Informationen gesammelt wurden.
- Identifizierung und Unterdrückung der Datenkategorien, die keinen signifikanten Einfluss auf das Lernen oder das Ergebnis der Herleitung haben.

[552] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 40. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

- Unterdrückung irrelevanter Schlussfolgerungen im Zusammenhang mit personenbezogenen Informationen während des Trainingsprozesses, beispielsweise im Falle des unüberwachten Trainings.
- Verwendung von Verifizierungstechniken, die weniger Daten erfordern, wie z. B. die Kreuzvalidierung.
- Analyse und Konfiguration von algorithmischen Hyperparametern, die die Menge oder den Umfang der verarbeiteten Daten beeinflussen könnten, um diese zu minimieren.
- Verwendung von föderierten statt zentralisierten Lernmodellen.
- Anwendung von differenziellen Datenschutzstrategien.
- Training mit verschlüsselten Daten unter Verwendung homomorpher Techniken.
- Datenaggregation.
- Anonymisierung und Pseudonymisierung, nicht nur bei der Datenkommunikation, sondern auch in Bezug auf Trainingsdaten, möglichen personenbezogenen Daten im Modell und bei der Verarbeitung der Herleitung.

4.2.2 Erkennen und Eliminieren von Verzerrungen

Selbst wenn die Mechanismen zur Vermeidung von Verzerrungen bereits in den vorangegangenen Phasen angewandt wurden (siehe vorheriger Abschnitt über das Training), muss dennoch sichergestellt werden, dass die Ergebnisse der Trainingsphase die Verzerrungen auf ein Mindestmaß senken. Dies kann schwierig sein, da einige Arten von Verzerrungen und Diskriminierung oft besonders schwer zu erkennen sind. Die Teammitglieder, die die Eingabedaten kuratieren, sind sich dessen manchmal nicht bewusst, ebenso wenig wie die jeweiligen Nutzer. Daher sind die vom KI-Entwickler in der Validierungsphase eingesetzten Überwachungssysteme äußerst wichtige Faktoren, um Verzerrungen zu vermeiden.

Es gibt zahlreiche technische Hilfsmittel, die sich hervorragend zum Erkennen von Verzerrungen eignen, darunter die algorithmische Folgenabschätzung.[553] KI-Entwickler sollten daher deren wirksame Umsetzung berücksichtigen.[554] Wie die Literatur zeigt[555], kann es jedoch vorkommen, dass sich ein Algorithmus nicht vollständig von allen verschiedenen Arten von Verzerrungen befreien lässt. Der Entwickler sollte jedoch zumindest versuchen, sich ihrer Existenz und der daraus resultierenden Folgen bewusst zu sein (siehe Abschnitte Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz und Richtigkeit, Kapitel Grundsätze).

---

[553] D. Reisman, K. Crawford und M. Whittaker (2018), Algorithmic impact assessments: a practical framework for public agency accountability. AI Now Institute, New York, NY. Verfügbar unter: https://ainowinstitute.org/aiareport2018.pdf (abgerufen am 15. Mai 2020).

[554] ICO (2020) AI auditing framework – draft guidance for consultation. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (abgerufen am 15. Mai 2020).

[555] Chouldechova, A. (2017) „Fair prediction with disparate impact: a study of bias in recidivism prediction instruments", *Big Data* 5(2): 153–163, http://doi.org/10.1089/big.2016.0047

4.2.3 Ausübung der Rechte der betroffenen Person

Selbstverständlich müssen die Verantwortlichen alle Rechte der betroffenen Personen während des gesamten Lebenszyklus ermöglichen. In dieser speziellen Phase sind jedoch das Auskunftsrecht, das Recht auf Berichtigung und das Recht auf Löschung besonders sensibel. Sie weisen bestimmte Merkmale auf, derer sich die Verantwortlichen bewusst sein müssen.

a) Recht auf Auskunft

Im Allgemeinen können Trainingsdaten kaum mit einer einzelnen betroffenen Person in Verbindung gebracht werden, da sie in der Regel nur Informationen enthalten, die für Vorhersagen relevant sind, beispielsweise frühere Transaktionen, demografische Daten oder den Standort, jedoch keine Kontaktangaben oder eindeutige Kundenkennungen. Darüber hinaus werden sie häufig vorverarbeitet, um sie für die Algorithmen besser nutzbar zu machen. Dies bedeutet jedoch nicht, dass diese Daten als vollständig pseudonymisiert oder anonymisiert betrachtet werden können. Es handelt sich somit weiterhin um personenbezogene Daten. Im Falle eines Modells zur Kaufvorhersage könnte das Training beispielsweise ein für einen Kunden spezifisches Kaufmuster enthalten. Wenn in diesem Beispiel ein Kunde eine Liste seiner letzten Einkäufe im Rahmen seiner Anfrage zur Verfügung stellt, könnte das Unternehmen möglicherweise den Teil der Trainingsdaten identifizieren, der sich auf diese Person bezieht.

Unter diesen Umständen sollten KI-Entwickler auf Antrag der betroffenen Personen in Bezug auf Auskunft über sie betreffende personenbezogenen Daten reagieren, vorausgesetzt, sie haben angemessene Maßnahmen ergriffen, um die Identität der betroffenen Person zu überprüfen, und es gelten keine weiteren Ausnahmen. Wie das ICO zudem feststellt, „sollten Anträge auf Auskunft, Berichtigung oder Löschung von Trainingsdaten nicht als offensichtlich unbegründet oder übertrieben angesehen werden, nur weil sie möglicherweise schwieriger zu erfüllen sind oder die Motivation der Anfrage im Vergleich zu anderen Auskunftsanfragen, die ein Unternehmen in der Regel erhält, unklar sein könnte.“ [556]

Natürlich sind Unternehmen nicht verpflichtet, zusätzliche personenbezogene Daten zwecks Identifizierung der betroffenen Personen in den Trainingsdaten zu erheben oder zu speichern, nur um die Verordnung einzuhalten. Wenn die KI-Entwickler eine betroffene Person in den Trainingsdaten nicht identifizieren können und die betroffene Person keine zusätzlichen Informationen bereitstellen kann, die ihre Identifizierung ermöglichen würden, sind KI-Entwickler nicht verpflichtet, einem Antrag nachzukommen, der nicht erfüllt werden kann.

b) Recht auf Berichtigung

Im Falle des Rechts auf Berichtigung muss der Verantwortliche das Recht auf Berichtigung der Daten gewährleisten, insbesondere der Daten, die durch die von der KI-Entwicklung erstellten Herleitungen und Profile erzeugt wurden.

Auch wenn der Zweck von Trainingsdaten darin besteht, Modelle zu trainieren, die auf allgemeinen Mustern in großen Datensätzen beruhen, und einzelne Ungenauigkeiten somit kaum direkte Auswirkungen auf eine betroffene Person haben werden, darf das Recht auf Berichtigung nicht eingeschränkt werden. Der Verantwortliche könnte höchstens eine längere Frist (zwei zusätzliche Monate) für die Berichtigung beantragen, wenn das technische Verfahren besonders komplex ist (Artikel 11 Absatz 3).

---

[556] ICO (2019) Enabling access, erasure, and rectification rights in AI tools. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (abgerufen am 15. Mai 2020).

Kasten 19: Berichtigung

Beispielsweise ist es unter Umständen wichtiger, eine falsch erfasste Kundenlieferadresse zu berichtigen als die gleiche falsche Adresse in den Trainingsdaten zu korrigieren. Dies liegt daran, dass Ersteres zu einer fehlgeschlagenen Lieferung führen könnte, Letzteres aber kaum Auswirkungen auf die Gesamtgenauigkeit des Modells hätte.[557]

c) Recht auf Löschung

Die betroffenen Personen haben das Recht, zu verlangen, dass sie betreffende personenbezogene Daten unverzüglich gelöscht werden. Dieses Recht kann jedoch eingeschränkt sein, wenn bestimmte konkrete Umstände vorliegen. Laut dem ICO „können Unternehmen Anträge auf Löschung von Trainingsdaten erhalten. Unternehmen müssen auf Löschungsanträge reagieren, wenn die betroffenen Personen angemessene Gründe vorgebracht haben, es sei denn, es gilt eine einschlägige gesetzliche Ausnahmeregelung. Wenn beispielsweise die Trainingsdaten nicht mehr benötigt werden, weil das ML-Modell bereits trainiert wurde, muss das Unternehmen dem Antrag nachkommen. Falls die Entwicklung des Systems jedoch noch nicht abgeschlossen ist, kann es jedoch notwendig sein, die Trainingsdaten für die Zwecke des erneuten Trainings, der Verfeinerung und der Bewertung eines KI-Tools zu speichern. In diesem Fall sollte das Unternehmen von Fall zu Fall entscheiden, ob es dem Antrag nachkommen kann. Einem Antrag auf Löschung von Trainingsdaten nachzukommen, bedeutet jedoch nciht, dass auf diesen Daten beruhenden ML-Modelle gelöscht werden müssen, es sei denn, die Modelle selbst enthalten diese Daten oder können zu deren Herleitung verwendet werden.“[558]

5 Bewertung (Validierung)

„Vor der endgültigen Einführung des vom Datenanalysten erstellten Modells sollte das Modell unbedingt gründlicher bewertet und die Konstruktion des Modells überprüft werden, um sicher zu sein, dass es die Geschäftsziele ordnungsgemäß erreicht. In dieser Phase ist die Prüfung von entscheidender Bedeutung, ob ein wichtiger Geschäftsaspekt eventuell nicht ausreichend berücksichtigt wurde. Am Ende dieser Phase sollte der Projektleiter dann entscheiden, wie die Data-Mining-Ergebnisse zu verwenden sind. Die wichtigsten Schritte sind hier die Bewertung der Ergebnisse, die Überprüfung des Prozesses und die Festlegung der nächsten Schritte.“[559]

---

[557] ICO (2019) Enabling access, erasure, and rectification rights in AI systems. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (abgerufen am 15. Mai 2020).

[558] ICO (2019) Enabling access, erasure, and rectification rights in AI systems. Information Commissioner’s Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (abgerufen am 15. Mai 2020).

[559] C. Shearer (2000), The CRISP-DM model: the new blueprint for data mining, *Journal of Data Warehousing* 5(4): 13–23, Seite 17. Verfügbar unter: https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf (abgerufen am 15. Mai 2020).

5.1 Beschreibung

Diese Phase umfasst mehrere Aufgaben, die wichtige Datenschutzfragen aufwerfen. Im Allgemeinen muss der Entwickler:

- die Ergebnisse des Modells bewerten, z. B. ob es genau ist oder nicht; wobei es der KI-Entwickler zu diesem Zweck auch in der realen Welt testen könnte
- den Prozess überprüfen: Der Entwickler könnte den Data-Mining-Auftrag überprüfen, um festzustellen, ob ein wichtiger Faktor oder eine wichtige Aufgabe übersehen wurde. Dazu gehören auch Fragen der Qualitätssicherung.

5.2 Wichtigste zu ergreifende Maßnahmen

5.2.1 Prozesse der dynamischen Validierung

Die Validierung der eine KI-Komponente umfassenden Verarbeitung muss unter Bedingungen erfolgen, die die tatsächliche Umgebung widerspiegeln, in der die Verarbeitung stattfinden soll. Darüber hinaus muss der Validierungsprozess regelmäßig überprüft werden, wenn sich die Bedingungen ändern oder wenn der Verdacht besteht, dass die Lösung selbst verändert werden könnte. KI-Entwickler müssen gewährleisten, dass die Validierung die Bedingungen, unter denen der Algorithmus validiert wurde, genau widerspiegelt.

Dazu sollte die Validierung alle Komponenten eines KI-Tools einbeziehen, einschließlich der Daten, im Voraus geschulter Modelle, Umgebungen und des Verhaltens des Systems als Ganzes, und so früh wie möglich einsetzen. Schließlich muss sichergestellt sein, dass die Ergebnisse bzw. Handlungen mit den Ergebnissen der vorangegangenen Prozesse übereinstimmen, indem sie zur Gewährleistung ihrer Konformität mit den zuvor definierten Zielsetzungen verglichen werden.[560] Zur Validierung müssen gelegentlich neue personenbezogene Daten erhoben werden. In einigen anderen Fällen verwenden die Verantwortlichen die Daten für andere Zwecke als die ursprünglich vorgesehenen. In all diesen Fällen sollten die Verantwortlichen die Einhaltung der DSGVO sicherstellen (siehe Abschnitt „Zweckbindung“, Kapitel „Grundsätze“ sowie „Datenschutz und wissenschaftliche Forschung“ in Abschnitt „Konzepte“).

5.2.2 Löschen unnötiger Datensätze

Häufig sind der Validierungs- und Trainingsprozess auf gewisse Weise miteinander verbunden. Wenn im Rahmen der Validierung erforderliche Verbesserungen des Modells erkannt wurden, sollte das Training erneut durchgeführt werden.

Sobald die KI-Entwicklung abgeschlossen ist, ist im Prinzip auch die Trainingsphase des KI-Tools beendet. Zu diesem Zeitpunkt sollten Sie den zu diesem Zweck verwendeten Datensatz

---

[560] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI. Europäische Kommission, Brüssel, Seite 22. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

löschen, es sei denn, es besteht eine rechtliche Notwendigkeit, die Daten für die Weiterentwicklung oder Bewertung des Systems oder für andere Zwecke zu speichern, die gemäß den Bestimmungen von Artikel 6 Absatz 4 DSGVO mit den Zwecken vereinbar sind, für die sie erhoben wurden (siehe Abschnitt „Definition angemessener Speicherungsrichtlinien“ in diesem Dokument). KI-Entwickler sollten jedoch stets bedenken, dass die Löschung personenbezogener Daten im Widerspruch zur Notwendigkeit stehen kann, die Genauigkeit von Tools zu aktualisieren, die auf dem Echtzeit-Selbstlernen von Algorithmen beruhen: Wird ein Fehler gefunden, müssen wahrscheinlich die zuvor in der Trainingsphase verwendeten Daten abgerufen werden.

Beantragen betroffene Personen die Löschung der Daten, muss der Verantwortliche einen Einzelfallansatz wählen und dabei etwaige in der Verordnung vorgesehene Einschränkungen dieses Rechts berücksichtigen (siehe Artikel 17 Absatz 3).[561]

5.2.3 Durchführung eines externen Audits der Datenverarbeitung

In Fällen, in denen die Risiken der Verarbeitung personenbezogener Daten im Rahmen des KI-Tools hoch sind, muss ein Audit des Systems durch einen unabhängigen Dritten in Betracht gezogen werden. Es können verschiedene Audits durchgeführt werden. Sie können intern oder extern erfolgen, sie können sich nur auf das Endprodukt beziehen oder in Bezug auf weniger weit entwickelte Prototypen durchgeführt werden. Sie können als eine Form der Überwachung oder als ein Transparenzinstrument betrachtet werden. Anhang I am Ende dieses Dokuments enthält einige Empfehlungen der spanischen Datenschutzbehörde, die als Modell dienen könnten.

Im Hinblick auf die rechtliche Richtigkeit müssen KI-Tools überprüft werden, um festzustellen, ob die Verarbeitung personenbezogener Daten in ihrem System die in der DSGVO festgelegten Verpflichtungen erfüllt, wobei eine Vielzahl von Fragen zu berücksichtigen ist. Die Hochrangige Expertengruppe für künstliche Intelligenz stellte in Bezug auf das Erproben und Prüfen fest: „Es sollte von einer möglichst vielfältig zusammengesetzten Personengruppe entworfen und entwickelt werden. Für die Kategorien, die aus verschiedenen Perspektiven getestet werden, sollten vielfältige Metriken entwickelt werden. Kontradiktorische Tests durch vertrauenswürdige, gemischte „rote Teams“, die bewusst versuchen, das System zu „knacken“, um Schwachstellen zu finden, und „Bug-Bounty“-Programme, die Außenstehende dazu anregen, Systemfehler und -schwächen zu erkennen und verantwortungsbewusst zu melden, sollten in Betracht gezogen werden.“[562] Es gibt jedoch gute Gründe, skeptisch zu sein, ob ein Prüfer in der Lage ist, die Funktionsweise eines maschinellen Lernsystems zu überprüfen.

Deshalb ist es sinnvoll, sich auf die Punkte zu konzentrieren, die die spanische Datenschutzbehörde (AEPD) in ihrer empfohlenen Checkliste aufführt: Es wäre einfacher, sich auf die Maßnahmen zur Vermeidung von Verzerrungen, Unklarheiten, versteckter Profilerstellung usw. zu konzentrieren, indem beispielsweise die Umsetzung von Grundsätzen

---

[561] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 26. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

[562] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI. Europäische Kommission, Brüssel, Seite 22. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020).

wie Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen oder Datenminimierung und der angemessene Einsatz von Instrumenten wie der DSFA oder das Eingreifen eines qualifizierten DSB geprüft werden, als zu versuchen, die Funktionsweise eines komplexen Algorithmus in allen Einzelheiten zu verstehen (das Blackbox-Problem ist zu diesem Zweck natürlich sehr wichtig). Die Umsetzung angemessener Datenschutzmaßnahmen in den ersten Phasen des Lebenszyklus des Tools ist die beste Möglichkeit, um Datenschutzprobleme zu vermeiden.

Kasten 20: Die Schwierigkeit, ein System des maschinellen Lernens zu überprüfen: Die Watson-Plattform von IBM

Die Richtlinie von IBM betont, dass Watson durch „überwachtes Lernen" trainiert wird. Anders ausgedrückt: Das System wird Schritt für Schritt beim Lernen angeleitet. Der Prozess sollte somit überwacht werden können, im Gegensatz zum unüberwachten Lernen, bei dem die Maschine ihre Betriebskriterien völlig selbständig bestimmt. Laut Aussage von IBM wird auch überprüft, was die Systeme getan haben, bevor eine Entscheidung zur Beibehaltung eines bestimmten Lerntyps getroffen wird. Die Experten, die sich im Rahmen der verschiedenen organisierten Debatten (darunter auch durch den Forschungsausschuss für Ethik von Allistene, CERNA) zu diesem Thema geäußert haben, haben jedoch immer wieder darauf hingewiesen, dass solche Aussagen falsch sind. Nach dem derzeitigen Stand der Forschung ist der „Output" der neuesten Algorithmen des maschinellen Lernens nicht erklärbar, wobei erklärbare KI ein Konzept ist, an dem noch geforscht wird. Laut den Experten ist es zudem sehr schwierig ist, ein System des maschinellen Lernens in der Praxis zu überprüfen.[563]

6 Bereitstellung

„Die Bereitstellung ist der Prozess, bei dem ein IT-System in seiner Umgebung betriebsbereit gemacht wird, einschließlich Installation, Konfiguration, Betrieb, Tests und Vornahme der erforderlichen Änderungen. Die Bereitstellung erfolgt in der Regel nicht durch die Entwickler eines Systems, sondern durch das IT-Team des Kunden. Aber selbst in diesem Fall sind die Entwickler dafür verantwortlich, dem Kunden ausreichende Informationen für eine erfolgreiche Bereitstellung des Modells zu liefern. Dazu gehört in der Regel ein (allgemeiner) Bereitstellungsplan mit den notwendigen Schritten für eine erfolgreiche Bereitstellung und deren Durchführung sowie ein (allgemeiner) Überwachungs- und Wartungsplan für die Wartung des Systems, die Überwachung der Bereitstellung und die korrekte Nutzung der Data-Mining-Ergebnisse."[564]

---

[563] CNIL (2017), How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence. Commission Nationale de l'Informatique et des Libertés, Paris, Seite 28. Verfügbar unter: www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf (abgerufen am 15. Mai 2020).

[564] SHERPA-Projekt (2020) Guidelines for the ethical development of AI and big data systems: an ethics by design approach. SHERPA, Seite 13. Verfügbar unter: www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf (abgerufen am 15. Mai 2020).

6.1 Wichtigste zu ergreifende Maßnahmen

6.1.1 Allgemeine Bemerkungen

Wenn das KI-Tool personenbezogene Daten enthält, ist zum Zeitpunkt der Verteiliung Folgendes erforderlich (siehe auch Abschnitt „Erwerb oder Förderung des Zugangs zu einer Datenbank“, Kapitel „Maßnahmen und Tools“):

- Löschen oder ggf. Begründung, warum dies aufgrund der Beeinträchtigung des Modells ganz oder teilweise nicht möglich ist (siehe Abschnitt „Speicherbegrenzung“ , Kapitel „Grundsätze“).
- Bestimmen der Rechtsgrundlage für die Übermittlung personenbezogener Daten an Dritte, insbesondere wenn besondere Datenkategorien betroffen sind (siehe Unterabschnitt „Rechtmäßigkeit“, Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“).
- Inkenntnissetzung der betroffenen Personen über die oben beschriebene Verarbeitung.
- Nachweis, dass die Richtlinien zum Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen umgesetzt worden sind (insbesondere Datenminimierung).
- Je nach den Risiken, die sich für die Betroffenen ergeben könnten, und unter Berücksichtigung des Umfangs oder der Kategorien der zu verwendenden personenbezogenen Daten sollte die Durchführung einer Datenschutz-Folgenabschätzung (DSFA) in Betracht gezogen werden.[565]

Sobald das Modell zum Einsatz kommt, werden die Trainingsdaten grundsätzlich aus dem Algorithmus gelöscht und das Modell verarbeitet nur noch die personenbezogenen Daten, auf die es angewendet wird. Der Verantwortliche kann ggf. die Daten der betroffenen Person zur Anpassung der von dem KI-Tool angebotenen Dienstleistung speichern. Sobald diese Dienstleistung jedoch beendet ist, müssen diese Daten gelöscht werden, es sei denn, überzeugende Gründe sprechen dafür, sie zu behalten. Das bedeutet natürlich nicht, dass die Datenspeicherung ewig dauern sollte.

Der KI-Entwickler muss sicherstellen, dass der Algorithmus keine versteckten personenbezogenen Daten enthält (oder die notwendigen Maßnahmen ergreifen, wenn dies unvermeidbar ist). In jedem Fall muss der Entwickler eine formale Bewertung durchführen, um festzustellen, welche personenbezogenen Daten der betroffenen Personen identifizierbar sein könnten.[566] Dies kann mitunter kompliziert sein. Beispielsweise könnten einige KI-Tools wie Vector Support Machines (VSM) in der Logik des Modells von Trainingsdaten durch Technikgestaltung enthalten. In anderen Fällen können im Modell Muster gefunden werden, die eine bestimmte Person identifizieren.[567] In all diesen Fällen könnten Unbefugte in der

---

[565] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 26. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

[566] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 41. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

[567] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 13. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

Lage sein, Elemente der Trainingsdaten wiederherzustellen oder durch Analyse des Verhaltens des Modells auf die Person zu schließen, die darin enthalten war.

Unter solchen Bedingungen könnte die Gewährleistung schwierig sein, dass die betroffenen Personen ihr Recht auf Auskunft, Berichtigung und Löschung ausüben können. Denn „sofern die betroffene Person nicht nachweist, dass sie betreffende personenbezogene Daten aus dem Modell hergeleitet werden können, kann der Verantwortliche möglicherweise nicht feststellen, ob personenbezogene Daten hergeleitet werden können und ob der Antrag daher begründet ist".[568] Die Verantwortlichen sollten jedoch regelmäßig Prüfungen durchführen, um die Wahrscheinlichkeit beruhend auf dem Stand der Technik zu bewerten, dass personenbezogene Daten aus Modellen hergeleitet werden können, sodass das Risiko einer versehentlichen Offenlegung minimiert wird. Lassen diese Prüfungen auf eine erhebliche Möglichkeit der Offenlegung von Daten schließen, sollten die erforderlichen Maßnahmen zur Verhinderung ergriffen werden (siehe Abschnitt „Integrität und Vertraulichkeit", Kapitel „Grundsätze").

6.1.2 Aktualisierung der Informationen

Wenn der Algorithmus von einem Dritten implementiert wird, sollten die KI-Entwickler die Ergebnisse des eingesetzten Validierungs- und Überwachungssystems mitteilen und ihre Mitarbeit bei der weiteren Überwachung der Validierung der Ergebnisse anbieten. Es wäre auch ratsam, diese Art der Koordinierung mit allen Dritten vorzusehen, von denen Datenbanken oder andere relevante Komponenten im Lebenszyklus des Systems erworben werden. Wenn es sich um eine Datenverarbeitung durch einen Dritten handelt, muss der Verantwortliche sicherstellen, dass der Zugang im Rahmen einer Rechtsgrundlage erfolgt.

Darüber hinaus müssen dem Endnutzer in jeder Phase Echtzeitinformationen über die Werte in Bezug auf die Richtigkeit und/oder Qualität der hergeleiteten Informationen angeboten werden (siehe Abschnitt „Richtigkeit", Kapitel „Grundsätze"). Wenn die hergeleiteten Informationen die Mindestschwellenwert in Bezug auf die Qualität nicht erreichen, muss ausdrücklich darauf hingewiesen werden, dass diese Informationen keinen Wert besitzen.[569] Diese Anforderung bedeutet häufig, dass die Entwickler detaillierte Informationen über die Trainings- und Validierungsphasen bereitstellen sollten. Besonders wichtig sind Informationen über die zu diesem Zweck verwendeten Datensätze. Andernfalls könnte die verwendete Lösung zu enttäuschenden Ergebnissen für die Endnutzer führen, die dann über die Ursache spekulieren müssen.

---

[568] ICO (2019) Enabling access, erasure, and rectification rights in AI tools. Information Commissioner's Office, Wilmslow. Verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/ (abgerufen am 15. Mai 2020).

[569] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 35. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 15. Mai 2020).

Anhang I: Auditierung von AI-Tools[570]

Laut der spanischen Datenschutzbehörde sollten die Audits eine lange Liste von Punkten abdecken, darunter:

- Vorhandensein oder Nichtvorhandensein von personenbezogenen Daten, Profiling oder automatischen Entscheidungen über betroffene Personen ohne das Eingreifen von Personen
- Wirksamkeit von Anonymisierungs- und Pseudonymisierungsmethoden
- Vorhandensein und Legitimität der Verarbeitung besonderer Kategorien von Daten, insbesondere der hergeleiteten Informationen
- Rechtsgrundlage für die Verarbeitung und Bestimmung der Verantwortlichkeiten
- Bewertung des Gleichgewichts zwischen den verschiedenen Interessen und der Auswirkungen auf die Rechte und Freiheiten der betroffenen Personen unter Berücksichtigung der angenommenen Garantien, insbesondere, wenn die Rechtsgrundlage das berechtigte Interesse ist
- Angabe und Wirksamkeit der implementierten Transparenzmechanismen
- Anwendung des Grundsatzes der proaktiven Rechenschaftspflicht und des Risikomanagements in Bezug auf die Rechte und Freiheiten der betroffenen Personen und insbesondere die Frage, ob die Verpflichtung oder die Notwendigkeit zur Durchführung einer DSFA besteht sowie ggf. deren Ergebnisse
- Anwendung von Maßnahmen zum Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellung, darunter:
    - Analyse der Notwendigkeit der Menge und des Umfangs der Verarbeitung personenbezogener Daten in den verschiedenen Phasen der KI-Entwicklung
    - Analyse der Richtigkeit, Zuverlässigkeit, Qualität und Verzerrungen der für die Entwicklung oder den Betrieb der KI-Komponente verwendeten oder erfassten Daten sowie der verwendeten Datenbereinigungsmethoden
    - Überwachung und Durchführung von Test- und Validierungsprozessen in Bezug auf die Präzision, Richtigkeit, Konvergenz, Einheitlichkeit, Vorhersagbarkeit und andere Messgrößen für die Güte der verwendeten Algorithmen, Profile und Herleitungen. Darüber hinaus ist zu prüfen, ob diese Parameter den Verarbeitungsanforderungen entsprechen.
- Die Angemessenheit der Sicherheitsmaßnahmen zur Vermeidung von Datenschutzrisiken
- Schulung und Weiterbildung des Personals des Verantwortlichen im Zusammenhang mit der Entwicklung oder Implementierung der KI-Komponente, gegebenenfalls mit besonderem Augenmerk auf die korrekte Auslegung der Herleitungen

Anhang II: Maschinelles Lernen und Forschung im Bereich der künstlichen

[570] AEPD (2020) Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción. Agencia Espanola Proteccion Datos, Madrid, Seite 45–47. Verfügbar unter: www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf (abgerufen am 3. Juni 2020).

Intelligenz zum Nutzen von Patienten: 20 kritische Fragen zu Transparenz, Replizierbarkeit, Ethik und Wirksamkeit[571]

Einführung

1. Wie lautet die Gesundheitsfrage in Bezug auf den Patientennutzen?

2. Welche Belege gibt es dafür, dass bei der Entwicklung des Algorithmus bewährte Verfahren der klinischen Forschung und des epidemiologischen Studiendesigns berücksichtigt wurden?

Studie

1. Wann und wie sollten Patienten in die Erhebung, Analyse, Bereitstellung und Nutzung von Daten einbezogen werden?

2. Sind die Daten zur Beantwortung der klinischen Frage geeignet, d. h. erfassen sie die relevante Heterogenität in der realen Welt, und weisen sie eine ausreichende Detailliertheit und Qualität auf?

3. Spiegelt die Validierungsmethodik die mit der Datenerhebung und -speicherung verbundenen Sachzwänge und Arbeitsabläufe in der Praxis wider?

4. Welche Rechen- und Softwareressourcen sind für die Aufgabe erforderlich, und reichen die verfügbaren Ressourcen aus, um dieses Problem zu lösen?

Statistische Methoden

1. Sind die berichteten Leistungskennzahlen für den klinischen Kontext, in dem das Modell verwendet werden soll, relevant?

2. Wird der ML/KI-Algorithmus mit der derzeit besten Technologie und mit anderen geeigneten Baselines verglichen?

3. Ist der angegebene Gewinn an statistischer Leistung durch den ML/KI-Algorithmus in Bezug auf etwaigen Kompromissen gerechtfertigt?

Reproduzierbarkeit

1. Auf welcher Grundlage sind die Daten für andere Forscher zugänglich?

2. Sind der Code, die Software und alle anderen relevanten Teile der Vorhersagemodellierungspipeline für andere verfügbar, um die Reproduzierbarkeit zu erleichtern?

3. Liegt eine organisatorische Transparenz über den Daten- und Ergebnisfluss vor?

Folgenabschätzung

1. Lassen sich die Ergebnisse auf andere Bereiche als den, in dem das System entwickelt wurde, verallgemeinern (d. h. Reproduzierbarkeit der Ergebnisse/externe Gültigkeit)?

2. Führt das Modell zu Ungleichheiten in der Gesundheitsversorgung nach Alter, Geschlecht, ethnischer Zugehörigkeit oder anderen geschützten Merkmalen oder verschärft es diese?

3. Welche Belege gibt es dafür, dass Mediziner und Patienten das Modell und dessen Ergebnisse als (einigermaßen) interpretierbar erachten?

---

[571] S. Vollmer et al. (2020), Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness, *BMJ* 2020;368:l6927, http://dx.doi.org/10.1136/bmj.l6927

4. Wie wird die Wirksamkeit des Modells in der Praxis im vorgeschlagenen klinischen Umfeld nachgewiesen und wie werden unbeabsichtigte Folgen verhindert?

Umsetzung

1. Wie wird das Modell regelmäßig neu bewertet und aktualisiert, wenn sich die Datenqualität und die klinische Praxis ändern (d. h. Überwachung nach der Bereitstellung)?

2. Ist das ML/KI-Modell in Bezug auf Entwicklung, Umsetzung und Wartung kosteneffizient?

3. Wie werden die potenziellen finanziellen Vorteile im Falle einer Vermarktung des ML/KI-Modells verteilt?

4. Wie wurden die rechtlichen und behördlichen Anforderungen in Bezug auf die Akkreditierung/Genehmigung berücksichtigt?

Anhang III: Checklisten

Checkliste: Geschäftsverständnis

 Die Verantwortlichen haben die für die Entwicklung der KI-Lösung benötigte Datenmenge bzw. die Art dieser Daten bewertet und sichergestellt, dass sie mit dem Grundsatz der Minimierung übereinstimmen.

 Die Verantwortlichen haben je nach Anwendungsfall akzeptable Schwellenwerte für falsch positive/negative Ergebnisse oder Bereiche festgelegt und dann eine Nutzenbilanz erstellt.

 Die Verantwortlichen haben ein angemessenes Gleichgewicht zwischen dem erforderlichen Genauigkeitsgrad und dem Umfang der dafür erforderlichen personenbezogenen Daten hergestellt.

 Die Verantwortlichen haben, wann immer möglich, die Entwicklung verständlicherer Algorithmen gegenüber weniger verständlichen Algorithmen vorgesehen.

 Die Verantwortlichen haben für eine optimale Schulung aller am Projekt beteiligten Personen bzw. eine angemessene interne oder externe Bewertung ethischer und rechtlicher Fragen gesorgt.

 Die Verantwortlichen haben die Tools zur Legitimierung der Datenverarbeitung sorgfältig entwickelt. Zu diesem Zweck haben sie geprüft, ob die Einschaltung einer Ethik-Kommission erforderlich oder irgendeine Art von weicher Regelung anwendbar ist.

 Die Verantwortlichen haben einen risikobasierten Ansatz gewählt (einschließlich technischer und organisatorischer Sicherheitsmaßnahmen), der die Risiken für die Rechte, Interessen und Freiheiten der betroffenen Personen minimiert.

Die Verantwortlichen haben Tools und Richtlinien implementiert, um die Wirksamkeit der technischen und organisatorischen Maßnahmen regelmäßig zu bewerten und zu evaluieren.

 Die Verantwortlichen haben geprüft, ob der Rechtsrahmen in Bezug auf die wissenschaftliche Forschung anwendbar ist.

Die Speicherrichtlinien sehen vor, dass personenbezogene Daten in einer Form

gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es für die Zwecke, für die sie verarbeitet werden, erforderlich ist.

 Die Verantwortlichen haben die Benennung eines Datenschutzbeauftragten in Betracht gezogen.

Checkliste: Datenverständnis

 Die Verantwortlichen haben geeignete technische und organisatorische Maßnahmen getroffen, die sicherstellen, dass durch Voreinstellung nur personenbezogene Daten, deren Verarbeitung für den jeweiligen bestimmten Verarbeitungszweck erforderlich ist, verarbeitet werden.

 Die Verantwortlichen haben Richtlinien eingeführt, die die Menge der erhobenen personenbezogenen Daten, den Umfang ihrer Verarbeitung, ihre Speicherfrist und ihre Zugänglichkeit auf ein Minimum beschränken. Durch diese Maßnahmen wird sichergestellt, dass personenbezogene Daten durch Voreinstellungen nicht ohne Eingreifen der Person einer unbestimmten Zahl von natürlichen Personen zugänglich gemacht werden.

 Die Verantwortlichen erheben keine unnötigen Daten. Wenn Daten bereits gespeichert sind, haben sie Maßnahmen ergriffen, um unnötige Datenelemente zu löschen.

 Die Verantwortlichen haben die Auflösung der Daten auf das für die mit der Verarbeitung verfolgten Zwecke erforderliche Mindestmaß beschränkt.

 Die Verantwortlichen haben die Rechtsgrundlage gewählt, die der wahren Natur ihrer Beziehung zu der betroffenen Person und dem Zweck der Verarbeitung am ehesten entspricht.

 Die Verantwortlichen haben sorgfältig analysiert, ob die Verarbeitung die Deanonymisierung anonymisierter Daten und die Schaffung neuer personenbezogener Informationen beinhaltet, die im ursprünglichen Datensatz nicht enthalten waren, und haben angemessene Maßnahmen getroffen, um diese Herausforderungen zu bewältigen.

 Die Verantwortlichen haben sichergestellt, dass die Zusammenführung von Datensätzen keine ethischen oder rechtlichen Probleme in Bezug auf die Rechte und Freiheiten der betroffenen Personen aufwirft.

Checkliste: Datenaufbereitung

 Die Verantwortlichen haben sichergestellt, dass die Daten richtig sind, d. h. dass sie richtig und auf dem neuesten Stand sind.

 Wenn Profiling oder automatisierte Entscheidungsfindung vorgesehen ist:

- Die Verantwortlichen haben den Personen einen Link zu ihrer Datenschutzerklärung gesendet, wenn sie sie betreffende personenbezogene Daten indirekt erhalten haben.
- Die Verantwortlichen haben erläutert, wie Personen auf Details der Informationen zugreifen können, die sie zur Erstellung ihres Profils verwendet haben.
- Die Verantwortlichen haben den betroffenen Personen, die ihnen sie betreffende personenbezogene Daten zur Verfügung stellen, mitgeteilt, wie sie dem Profiling widersprechen können.
- Die Verantwortlichen haben Verfahren eingeführt, anhand der Kunden auf die in ihre Profile eingegebenen personenbezogenen Daten zuzugreifen und diese im Hinblick auf etwaige Fehler überprüfen und bearbeiten können.

- Die Verantwortlichen haben zusätzliche Kontrollen für ihre Systeme zur Profilerstellung/automatisierten Entscheidungsfindung eingeführt, um schutzbedürftige Gruppen (einschließlich Kinder) zu schützen.
- Die Verantwortlichen haben sichergestellt, dass sie nur die erforderliche Mindestmenge an Daten erheben und über eine klare Speicherungsrichtlinie für die von ihnen erstellten Profile verfügen.
- Die Verantwortlichen haben eine DSFA durchgeführt, um die Risiken bei Einführung einer neuen automatisierten Entscheidungsfindung oder eines neuen Profilings zu berücksichtigen und zu bewältigen.
- Die Verantwortlichen haben den zuständigen DSB in diese Aktivitäten einbezogen.
- Die Verantwortlichen haben die Systemanforderungen, die zur Unterstützung einer aussagekräftigen Überprüfung durch Personen erforderlich sind, **bereits in der Designphase berücksichtigt.** Insbesondere die Anforderungen an die Interpretierbarkeit und die effektive Gestaltung der Benutzeroberfläche zur Unterstützung von Überprüfungen und Eingriffen durch Personen.
- Die Verantwortlichen haben eine angemessene Schulung und Unterstützung für menschliche Prüfer gestaltet und umgesetzt.
- Die Verantwortlichen haben dem an der Verarbeitung beteiligten Personal die entsprechenden Befugnisse, Anreize und Unterstützung gegeben, um die Bedenken einzelner Personen anzusprechen oder zu eskalieren und gegebenenfalls die Entscheidung des KI-Systems außer Kraft zu setzen.

- Die Verantwortlichen haben sichergestellt, dass die mit der Auswahl der in die Datensätze zu integrierenden Daten betrauten Teams aus Personen bestehen, die die Diversität gewährleisten, die von der KI-Entwicklung erwartet wird.
- Die Verantwortlichen haben sichergestellt, dass Faktoren, die zu Ungenauigkeiten in personenbezogenen Daten führen, korrigiert werden und das Risiko von Fehlern minimiert wird.
- Die Verantwortlichen haben Tools mit dem Ziel implementiert, diskriminierende Auswirkungen auf natürliche Personen aufgrund der Rasse oder der ethnischen Herkunft, der politischen Meinung, der Religion oder der Weltanschauung, der Gewerkschaftszugehörigkeit, des genetischen oder gesundheitlichen Status oder der sexuellen Orientierung zu verhindern oder die zu Maßnahmen führen, die eine solche Wirkung haben.

Checkliste: Modellierung (Training)

 Die Verantwortlichen haben den Verwendungszweck des KI-Systems zu Beginn des Trainings oder Einsatzes festgelegt und eine Neubewertung dieser Festlegung vorgenommen, falls die Verarbeitung des Systems unerwartete Ergebnisse hervorgebracht hat.

 Die Verantwortlichen haben die in der Trainingsphase verwendeten Daten von allen Informationen bereinigt, die für das Training des Modells nicht unbedingt erforderlich sind.

 Die Verantwortlichen haben den Einsatz technischer Hilfsmittel in Erwägung gezogen, die sich zur Erkennung von Verzerrungen eignen könnten, beispielsweise die algorithmische Folgenabschätzung.

 Die Verantwortlichen haben die Durchführung einer DSFA in dieser Phase in Betracht gezogen.

 Die Verantwortlichen haben sichergestellt, dass sie in der Lage sind, auf Anträge betroffener Personen auf Ausnahmen vom Auskunftsrecht zu reagieren.

 Die Verantwortlichen können das Recht auf Berichtigung der Daten gewährleisten, insbesondere derjenigen, die durch die von der KI-Entwicklung erstellten Herleitungen und Profile erzeugt wurden.

 Die Verantwortlichen sind in der Lage, Anträgen auf Löschung nachzukommen, es sei denn, es gilt eine einschlägige Ausnahmeregelung, und sofern die betroffene Person angemessene Gründe vorbringen kann.

Checkliste: Bewertung (Validierung)

 Die Verantwortlichen haben sichergestellt, dass die Validierung die Bedingungen, unter denen der Algorithmus validiert wurde, genau widerspiegelt.

 Die Verantwortlichen haben die betroffenen Personen über die zusätzliche Verarbeitung in dieser Phase informiert.

 Die Verantwortlichen haben sichergestellt, dass der zu Validierungszwecken verwendete Datensatz entfernt wird, es sei denn, es besteht eine rechtliche Notwendigkeit, ihn für die Weiterentwicklung oder Bewertung des Systems oder für andere Zwecke zu speichern, die mit denen, für die er erhoben wurde, vereinbar sind.

 Die Verantwortlichen haben die Durchführung einer DSFA in dieser Phase in Betracht gezogen.

 Sollten die betroffenen Personen die Löschung ihrer Daten beantragen, haben die Verantwortlichen einen Einzelfallansatz gewählt, der etwaige in der Verordnung vorgesehene Einschränkungen dieses Rechts berücksichtigt.

 Die Verantwortlichen haben ein Audit des Systems durch einen unabhängigen Dritten erwogen.

Checkliste: Bereitstellung

 Die Verantwortlichen haben alle unnötigen personenbezogenen Daten gelöscht bzw. die Unmöglichkeit der Löschung begründet.

 Die Verantwortlichen haben die betroffenen Personen über die zusätzliche Verarbeitung in dieser Phase informiert.

 Die Verantwortlichen haben die angemessene Rechtsgrundlage für die Übermittlung

personenbezogener Daten an Dritte bestimmt, insbesondere wenn besondere Datenkategorien betroffen sind.

 Die Verantwortlichen haben die Durchführung einer DSFA in Betracht gezogen.

 Die Verantwortlichen haben sichergestellt, dass der Algorithmus keine versteckten personenbezogenen Daten enthält (oder die notwendigen Maßnahmen ergriffen, wenn dies unvermeidbar ist).

 Die KI-Entwickler haben Tools zu dem Zweck implementiert, die Ergebnisse des eingesetzten Validierungs- und Überwachungssystems zu mitzuteilen, und haben ihre weitere Mitarbeit angeboten.

 Die KI-Entwickler haben sich verpflichtet, dem Endnutzer in jeder Phase Echtzeitinformationen über die Werte in Bezug auf die Richtigkeit und/oder Qualität der hergeleiteten Informationen anzubieten.

Erstes Szenario: Entwicklung eines KI-Tools zur Diagnose der COVID-19-Krankheit

Iñigo de Miguel Beriain (UPV/EHU)

Dieser Teil des Leitfadens wurde von Marko Sijan, Senior Advisor Specialist, (HR DPA) überprüft und bestätigt.

Beschreibung

Die Pandemie führte zu Situationen, in denen viele Patienten eine medizinische Versorgung benötigten, die jedoch aufgrund der hohen Zahl der erkrankten Mitarbeiter im Gesundheitswesen nur schwer zu gewährleisten war. In dieser Situation konnte zum Beispiel ein Radiologe die große Zahl der zu analysierenden Röntgenbilder nicht bewältigen, weil seine Kollegen krankgeschrieben waren. Der Einsatz von KI für solche Zwecke könnte eine große Hilfe für die Zukunft sein. Es gibt jedoch zahlreiche ethische und rechtliche Fragen, die dabei berücksichtigt werden müssen. In diesem Szenario werden die verschiedenen, von einem Forscherteam zu erfüllenden Schritte analysiert, um einen Algorithmus zu trainieren, der bei der Diagnose von Lungenkrankheiten helfen kann.

Vorbemerkungen

Die Forschung mit Gesundheitsdaten ist mit besonders großen ethischen Herausforderungen verbunden. Wenn es sich zudem um Patienten handelt, die an einer Krankheit wie COVID leiden, ist das Dilemma besonders gravierend. Im Zusammenhang mit der Gesundheitsfürsorge kann die mit der klinischen Praxis verbundene Einwilligung in Kenntnis der Sachlage leicht mit der Einwilligung in die biomedizinische Forschung vermischt werden. Dies ist immer problematisch. Die beiden Dinge sind grundverschieden. Bei der Planung einer Aktivität wie der Entwicklung eines KI-Tools für die Diagnose sollte dies berücksichtigt werden. Dies gilt insbesondere für Patienten, die sich in einer schutzbedürftigeren Situation befinden als gewöhnlich. Es sollte nie vergessen werden, dass sich die Ziele der biomedizinischen Forschung nicht mit den Interessen und dem Wohlergehen der Menschen überschneiden dürfen.

Es gibt mehrere wichtige Instrumente, die Forscher im Rahmen eines Plans zur Entwicklung eines KI-Tools immer im Auge behalten sollten. Die Checkliste für ethische Fragen im Leitfaden des Programms Horizont 2020 „How to complete your ethics self-assessment", Seite 6[572], ist sehr zu empfehlen. Die wichtigsten zu konsultierenden Dokumente umfassen:

- Hochrangige Expertengruppe für künstliche Intelligenz: Ethik-Leitlinien für eine vertrauenswürdige KI.[573]
- Weißbuch der Europäische Kommission, Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen.[574]
- **T**raining and **R**esources in **R**esearch **E**thics **E**valuation (TRREE)[575] ist ein Online-Tool, das kostenlosen Zugang zu folgenden Themen bietet:
    - **E-Learning**: ein internetgestützter Fernlehrgang über Forschungsethik mit Zertifikatsabschluss
    - **Online-Ressourcen:** eine interaktive Website mit Zugang zu Dokumenten legislativer und politischer Art von internationaler, regionaler und nationaler Bedeutung
- Weitere Online-Schulungstools sind auf der EUREC-Webseite zu finden[576].

SCHRITTWEISE ANALYSE

1 Geschäftsverständnis (Business Understanding)

1.1 Beschreibung

„Im Mittelpunkt der ersten Phase steht ein präzises Verständnis der Projektziele aus geschäftlicher Perspektive, der Abbildung dieses Wissen in einer Data Mining-

[572] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf

[573] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[574] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf

[575] https://elearning.trree.org/mod/page/view.php?id=70&lang=de

[576] http://www.eurecnet.org/materials/index.html

Problemdefinition und der anschließenden Entwicklung eines vorläufigen Plans, um die Ziele zu erreichen. Um zu verstehen, welche Daten später wie analysiert werden sollen, benötigen Data Mining-Fachleute ein umfassendes Verständnis des Unternehmens, für das sie eine Lösung finden wollen. Die Phase des Geschäftsverständnisses umfasst mehrere wichtige Schritte, darunter die Bestimmung der Geschäftsziele, die Bewertung der Situation, die Bestimmung der Data Mining-Ziele und die Erstellung des Projektplans."[577]

Dieses allgemeine Ziel umfasst vier Hauptaufgaben:

5. Bestimmung der Geschäftsziele: Dies bedeutet:
   - **a.** Ermittlung des primären Geschäftsziels sowie der damit verbundenen Fragen, die das Unternehmen angehen möchte.
   - **b.** Bestimmung des Erfolgsmaßstabs.

6. Bewertung der Situation
   - a. Ermittlung der für das Projekt verfügbaren materiellen und personellen Ressourcen.
   - b. Ermittlung der verfügbaren Daten verfügbar, um das primäre Geschäftsziel zu erreichen.
   - c. Auflistung der dem Projekt zugrunde liegenden Annahmen.
   - d. Auflistung der Projektrisiken, Auflistung möglicher Lösungen für diese Risiken, Erstellung eines Glossars der Geschäfts- und Data Mining-Begriffe und Erstellung einer Kosten-Nutzen-Analyse für das Projekt.

7. Bestimmung der Data Mining-Ziele: Festlegung, welcher Grad an Vorhersagegenauigkeit erwartet wird, um das Projekt als erfolgreich zu betrachten.

8. Erstellung eines Projektplans: Beschreibung des beabsichtigten Plans zur Erreichung der Data Mining-Ziele, einschließlich der Beschreibung spezifischer Schritte und eines vorgeschlagenen Zeitplans, einer Bewertung potenzieller Risiken und einer ersten Bewertung der zur Unterstützung des Projekts erforderlichen Tools und Techniken.

1.2 Wichtigste zu ergreifende Maßnahmen

1.2.1 Bestimmung der Geschäftsziele

Bei der Entwicklung eines KI-Tools müssen Sie zunächst klären, was Sie erreichen wollen. Ein Tool zur Erkennung einer Pathologie anhand eines Röntgenbildes kann beispielsweise zu dem Zweck entwickelt werden:

1) die Arbeit des Radiologen zu unterstützen;
2) die Arbeit eines Hausarztes zu unterstützen, d. h. zu entscheiden, ob der Patient an einen Spezialisten zu überweisen ist;
3) den Arzt zu ersetzen und selbständig eine Diagnose zu erstellen, z. B. COVID;
4) eine erste Triage durchzuführen (d. h. die Empfehlung, ob ein Hausarzt oder ein Spezialist eingreifen sollte).

Jedes dieser Szenarien weist sehr unterschiedliche Merkmale auf. Einige von ihnen erfordern ein höheres Maß an Genauigkeit als andere. Wenn Sie beispielsweise den Arzt ersetzen wollen, muss die KI ein beeindruckend hohes Maß an Präzision erreichen.

---

[577] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 14.

Die ethischen und rechtlichen Implikationen der verschiedenen Zwecke sind ebenfalls sehr unterschiedlich. Wenn das Instrument zur automatisierten Entscheidungsfindung eingesetzt werden soll, wie in den Fällen 3) oder 4), unterliegt die Verarbeitung der Daten einer wesentlich strengeren rechtlichen Regelung. In vielen Ländern kann eine solche Verwendung sogar direkt rechtswidrig sein.

All diese Überlegungen müssen von Anfang an berücksichtigt werden. Der Entwicklungsprozess sollte erst dann eingeleitet werden, nachdem Sie als Verantwortlicher geklärt haben, welche Ergebnisse erzielt werden sollen, denn diese Frage ist entscheidend dafür, ob die geplante Datenverarbeitung im Einklang mit der DSGVO steht oder nicht. Die Festlegung des erwarteten Grads an Vorhersagegenauigkeit, damit das Projekts als Erfolg angesehen werden kann, ist von wesentlicher Bedeutung, um die für die Entwicklung des KI-Tools benötigte Datenmenge oder die Art dieser Daten bewerten zu können. Der Grad der Vorhersagbarkeit oder Präzision des Algorithmus, die Validierungskriterien der Tests, die maximale Menge oder Mindestqualität der personenbezogenen Daten, die für den Einsatz in der realen Welt erforderlich sind, sind nur einige der grundlegenden Merkmale einer KI-Entwicklung.

Diese Schlüsselelemente der Entwicklung sollten bereits in der ersten Phase des Lebenszyklus der Lösung berücksichtigt werden. Dies ist äußerst hilfreich für die Umsetzung einer Richtlinie des „Datenschutzes durch Technikgestaltung“. Wenn ein akzeptables Genauigkeitsniveau bei Verwendung einer deutlich geringeren Menge personenbezogener Daten erreicht werden, als es ein höheres Niveau erfordert, sollte dies unbedingt in Betracht gezogen werden. Je ungenauer Ihre Bewertungen ausfallen, desto schwieriger wird es, die genauen Zwecke zu bestimmen, die mit der Verarbeitung verfolgt werden (siehe Unterabschnitt „Voraussetzungen für die Rechtmäßigkeit – festgelegte, eindeutige Zwecke“, Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“, Kapitel „Grundsätze“). Wenn wir uns vor Augen halten, dass die Verantwortlichen die Zwecke der Verarbeitung eindeutig angeben müssen, d. h. „in verständlicher Weise offenlegen, erläutern oder ausdrücken müssen“, sind genaue Erwartungen sehr empfehlenswert.

1.2.2 Die Wahl der technischen Lösungen

Im Allgemeinen sollten Sie stets verständlichere Algorithmen gegenüber weniger verständlichen Algorithmen bevorzugen. Zielkonflikte zwischen der Erklärbarkeit/Transparenz und der besten Leistung des Systems müssen je nach Nutzungskontext angemessen abgewogen werden. Auch wenn im Gesundheitswesen die Genauigkeit und Leistung des Systems wichtiger sein können als seine Erklärbarkeit, sollte man immer im Hinterkopf behalten, dass die Erläuterung einer Empfehlung beispielsweise der Schulung von Ärzten, der Bereitstellung angemessener Informationen für Patienten, die eine Wahl zwischen verschiedenen möglichen Behandlungen treffen müssen, oder der Begründung einer Triage-Entscheidung dienen kann. Wenn also eine recht ähnliche Dienstleistung sowohl durch einen leicht verständlichen als auch durch einen undurchsichtigen Algorithmus angeboten werden kann, d. h. wenn es keinen Zielkonflikt zwischen Erklärbarkeit und Leistung gibt, sollten Sie sich für den besser interpretierbaren Algorithmus entscheiden (siehe Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“, Kapitel „Grundsätze“).

1.2.3 Einführung einer Schulung zu ethischen und rechtlichen Fragen

Diese Maßnahme ist eine der wichtigsten Ratschläge, der vom ersten Moment einer geschäftlichen KI-Entwicklung an berücksichtigt werden sollte. Designer von Algorithmen (Entwickler, Programmierer, Datenwissenschaftler, Ingenieure), die das erste Glied in der algorithmischen Kette darstellen, sind sich der ethischen und rechtlichen Auswirkungen ihres Handelns wahrscheinlich nicht bewusst. Wenn alle beteiligten Mitarbeiter in engem Kontakt mit den betroffenen Personen stehen, lassen sich ethische Überlegungen leichter umsetzen. Dies wird jedoch wahrscheinlich nicht der Fall sein. Eines der Hauptprobleme eines KI-Tools, das sich mit Fragen des Gesundheitswesens befasst, besteht nämlich darin, dass es im Allgemeinen personenbezogene Daten verwendet, die in großen Datensätzen enthalten sind. Dadurch wird die Beziehung zwischen den Daten und der betroffenen Person in gewisser Weise verwischt, sodass es zu Verletzungen von Vorschriften kommt, die bei einer direkten Beziehung zwischen dem Verantwortlichen und der betroffenen Person nur selten auftreten.

Dies könnte schwerwiegende Folgen für die angemessene Einhaltung der Datenschutzstandards haben, vor allem weil es um Daten besonderer Kategorien geht. Diese Schlüsselpersonen müssen sich somit unbedingt der ethischen und sozialen Auswirkungen ihrer Arbeit und der Tatsache umfassend bewusst sein, dass diese sich sogar auf gesellschaftliche Entscheidungen erstrecken können, die sie von Rechts wegen nicht allein beurteilen sollten. Die Silo-Mentalität muss sorgfältig bekämpft werden.

Um zu vermeiden, dass die falsche Darstellung ethischer und rechtlicher Fragen unerwünschte Folgen nach sich zieht, gibt es zwei Hauptmaßnahmen, die ergriffen werden können. Erstens könnten die Entwickler sicherstellen, dass Designer von Algorithmen die Auswirkungen ihres Handelns sowohl für den Einzelnen als auch für die Gesellschaft verstehen und sich ihrer Verantwortung bewusst sind, indem sie lernen, kontinuierlich aufmerksam und wachsam zu sein.[578] In diesem Zusammenhang könnte eine optimale Schulung aller am Projekt beteiligten Personen (Entwickler, Programmierer, Datenwissenschaftler, Ingenieure, Forscher) noch vor Projektbeginn eines der effizientesten Instrumente sein, um Zeit und Ressourcen im Hinblick auf die Einhaltung der Datenschutzvorschriften zu sparen. So sollten grundlegende Schulungsprogramme implementiert werden, die zumindest die Grundlagen der Charta der Grundrechte, die in Artikel 5 DSGVO dargelegten Grundsätze, die Notwendigkeit einer Rechtsgrundlage für die Verarbeitung (einschließlich Verträge zwischen den Parteien) usw. umfassen.

Die Schulung von Personen, die noch nie mit Datenschutzfragen in Berührung gekommen sind, könnte jedoch schwierig sein. Eine alternative Maßnahme ist die Einbeziehung eines Experten für Datenschutz sowie ethische und rechtliche Fragen in das Entwicklungsteam, um ein disziplinübergreifendes Team zu bilden. Zu diesem Zweck könnte ein Experte (ein interner Mitarbeiter oder ein externer Berater) eingestellt werden, der die Strategie und die nachfolgenden Entscheidungen über personenbezogene Daten, die für die Entwicklung der Toole erforderlich sind, in enger Zusammenarbeit mit dem Datenschutzbeauftragten ausarbeitet.

Die Umsetzung angemessener Maßnahmen zur Gewährleistung der Vertraulichkeit ist ebenfalls sehr zu empfehlen (siehe Unterabschnitt „Unterabschnitt „Maßnahmen zur Unterstützung der Vertraulichkeit““, Abschnitt „Integrität und Vertraulichkeit““, Kapitel „Grundsätze“).

---

[578] Ibid., Seite 55.

1.2.4 Entwurf legitimer Datenverarbeitungstools

Gemäß Artikel 5 Absatz 1 Buchstabe b DSGVO müssen personenbezogene Daten „für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden“. Das Konzept der Legitimität ist in der DSGVO nicht genau definiert. Laut der Artikel 29-Datenschutzgruppe beinhaltet die Legitimität jedoch, dass die Daten „im Einklang mit dem Gesetz“ verarbeitet werden müssen, wobei „Gesetz“ als weit gefasster Begriff verstanden werden sollte, der „alle Formen des geschriebenen Rechts und des Gewohnheitsrechts, primäre und sekundäre Rechtsvorschriften, kommunale Verordnungen, gerichtliche Präzedenzfälle, Verfassungsgrundsätze, Grundrechte, andere Rechtsgrundsätze sowie die Rechtsprechung umfasst, d. h. wie ein solches „Gesetz“ von einem zuständigen Gericht ausgelegt und berücksichtigt würde“.

Es handelt sich also um ein umfassenderes Konzept als das der Rechtmäßigkeit. Es umfasst die Einhaltung der wichtigsten Werte der geltenden Verordnung und der wichtigsten ethischen Grundsätze. So werden beispielsweise einige konkrete KI-Entwicklungen das Eingreifen einer Ethik-Kommission erfordern. In einigen anderen Fällen könnten Richtlinien oder andere weiche Regelungen anwendbar sein. Sie sollten eine angemessene Erfüllung dieser Anforderungen gewährleisten, indem sie bereits während dieser Vorstufe des Tool-Lebenszyklus einen entsprechenden Plan erstellen (siehe „Legitimität und Rechtmäßigkeit“ in „Rechtmäßigkeit, Fairness und Transparenz“, Kapitel „Grundsätze“). Zu diesem Zweck sollten Sie sich insbesondere über die Anforderungen im Klaren sein, die sich aus den auf nationaler Ebene geltenden Vorschriften ergeben. In vielen Mitgliedstaaten wird die Entwicklung eines Algorithmus im Zusammenhang mit der Gesundheitsfürsorge mit Sicherheit die Einschaltung von Ethikkommissionen nach sich ziehen, und zwar höchstwahrscheinlich in einem ersten Stadium. Stellen Sie sicher, dass Ihr Forschungsplan mit diesen Anforderungen vereinbar ist.

1.2.5 Annahme eines risikobasierten Denkansatzes

Da bei der Erstellung Ihres Algorithmus mit Sicherheit eine große Menge besonderer Kategorien personenbezogener Daten, vor allem Gesundheitsdaten, verwendet wird, müssen Sie sicherstellen, dass Sie geeignete Maßnahmen ergreifen, um die Risiken für die Rechte und Freiheiten der betroffenen Personen zu minimieren (siehe „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“). Zu diesem Zweck müssen Sie die Risiken für die Rechte und Freiheiten der am Forschungs- und Entwicklungsprozess beteiligten Personen bewerten und beurteilen, welche Maßnahmen zu deren Schutz angemessen sind. In jedem Fall müssen Sie sicherstellen, dass sie die Datenschutzanforderungen erfüllen.

Risikobasiertes Denken in Bezug auf die Vertraulichkeit von Daten oder ein risikobasierter Ansatz in Bezug auf die Frage, welcher Schaden den Menschen/betroffenen Personen zugefügt werden könnte, muss von den ersten Schritten des Prozesses an berücksichtigt werden. Es könnte für den Verantwortlichen rechtliche Konsequenzen in Bezug auf die in der DSGVO festgelegten Verpflichtungen haben, wenn dies erst später in Betracht gezogen wird. Daher müssen Sie die impliziten Bedrohungen für die geplante Datenverarbeitung ermitteln und das damit verbundene Risiko bewerten. Wenn Sie planen, Software zu Verarbeitungszwecken zu verwenden, sollten Sie sicherstellen, dass angemessene Maßnahmen zur Wahrung der Vertraulichkeit getroffen werden. Wenn Ihre KI Software von Drittanbietern oder Standardsoftware verwendet, ist es von entscheidender Bedeutung, Funktionen auszuschließen, die personenbezogene Daten verarbeiten, für die es keine Rechtsgrundlage gibt oder die mit den beabsichtigten Zwecken nicht vereinbar sind.

Vermeiden Sie nach Möglichkeit die Nutzung von Datenspeichern oder Softwarediensten, die in einem Drittland angesiedelt sind. Sollte dies unvermeidlich sein, ist sicherzustellen, dass Ihre Datenverarbeitungsverträge mit diesen Dritten einen angemessenen, der DSGVO entsprechenden Schutz bieten. Anderenfalls sollten Sie dafür sorgen, dass die Forschungsteilnehmer über die Datenschutz-/Sicherheitsrisiken für ihre Daten vollständig aufgeklärt sind. *Sie sollten sich auch der angemessenen Sicherheitsmaßnahmen bewusst sein, die von den Anbietern von Datenspeichern und* Softwarediensten umgesetzt werden, und darüber informiert sein, dass die Unterlassung von Sicherheitsmaßnahmen zu einem Verstoß gegen die Sicherheitsverarbeitung führen kann.

Darüber hinaus müssen Sie sicherstellen, dass geeignete technische und organisatorische Maßnahmen getroffen werden, um das Risiko zu eliminieren oder zumindest abzuschwächen, sodass die Eintrittswahrscheinlichkeit der festgestellten Bedrohungen verringert wird oder deren Auswirkungen gemindert werden. Die Sicherheitsmaßnahmen müssen Teil Ihres Verzeichnisses von Verarbeitungstätigkeiten werden, und alle durchgeführten Maßnahmen sind Teil der Datenschutz-Folgenabschätzung.

Sobald die ausgewählten Maßnahmen umgesetzt sind, sollte das verbleibende Restrisiko bewertet und weiterhin kontrolliert werden. Sowohl die Risikoanalyse als auch die DSFA sind die Instrumente, die dabei zum Einsatz kommen. In Ihrem konkreten Fall müssen Sie eine DSFA durchführen, da die Entwicklung des KI-Tools die umfangreiche Verarbeitung besonderer Kategorien von Daten umfassen wird.

Schließlich darf nicht vergessen werden, dass sich bei der Nutzung von Big Data und KI die künftigen Risiken nur schwer vorhersehen lassen, sodass eine Bewertung der ethischen Auswirkungen nicht ausreicht, um alle möglichen Risiken zu berücksichtigen. Daher ist es wichtig, eine Neubewertung der Risiken in Erwägung zu ziehen, und es ist auch sehr empfehlenswert, eine dynamischere Methode zur Bewertung von Forschungsrisiken zu integrieren. Zögern Sie nicht, in anderen Phasen des Prozesses bei Bedarf zusätzliche Datenschutz-Folgenabschätzungen durchzuführen.

1.2.6 Vorbereiten der Dokumentation der Verarbeitung

Wer auch immer personenbezogene Daten verarbeitet (einschließlich Verantwortlicher und Auftragsverarbeiter), muss seine Tätigkeiten in erster Linie zur Verwendung durch qualifizierte/zuständige Aufsichtsbehörden dokumentieren. Dies muss durch Verzeichnisse über die Verarbeitung geschehen, die von Ihrem Unternehmen zentral für alle Verarbeitungstätigkeiten geführt werden, sowie durch zusätzliche Dokumentation, die sich auf eine einzelne Datenverarbeitungstätigkeit bezieht (siehe Abschnitt „Dokumentation der Verarbeitung", Kapitel „Maßnahmen und Tools"). Diese Vorbereitungsphase ist der ideale Zeitpunkt, um eine systematische Erfassung der erforderlichen Dokumentation festzulegen, da Sie zu diesem Zeitpunkt die Verarbeitungstätigkeit konzipieren und planen.

In der Tat sollten Sie eine Datenschutzrichtlinie erstellen (siehe Unterabschnitt „Größenvorteile bei der Einhaltung der Vorschriften und deren Nachweis", Abschnitt „Rechenschaftspflicht", Kapitel „Grundsätze"), die die Rückverfolgbarkeit von Informationen ermöglicht (falls genehmigte Verhaltenskodizes existieren, sollten diese umgesetzt werden, siehe wiederum Unterabschnitt „Größenvorteile bei der Einhaltung der Vorschriften und deren Nachweis", Abschnitt „Rechenschaftspflicht", Kapitel „Grundsätze"). Diese Richtlinie sollte auch die den Auftragsverarbeitern zugewiesenen Verantwortlichkeiten deutlich machen, wenn Sie sie in Ihr Projekt einbeziehen wollen, und die Aufgaben der Verarbeitungsvereinbarung enthalten, die ihnen im Zusammenhang mit der Ausübung der Rechte der betroffenen Personen übertragen werden. Sie sollten immer daran denken, dass Artikel 32 Absatz 4 DSGVO festlegt, dass ein wichtiges Sicherheitselement in der

Gewährleistung besteht, dass Mitarbeiter nur auf Anweisung und gemäß den Weisungen des Verantwortlichen handeln (siehe Abschnitt „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“).

Die Entwicklung Ihres KI-Tools kann die Verwendung verschiedener Datensätze beinhalten. Die Rückverfolgbarkeit der Verarbeitung, Informationen über die mögliche Wiederverwendung von Daten und die Verwendung von Daten, die zu verschiedenen Datensätzen gehören, in verschiedenen oder denselben Phasen des Lebenszyklus müssen durch die Aufzeichnungen sichergestellt werden.

Wie in den Anforderungen und Akzeptanztests für den Kauf und/oder die Entwicklung der verwendeten Software, Hardware und Infrastruktur (Unterabschnitt des Abschnitts „Dokumentation der Verarbeitung“) dargelegt, müssen die Risikobewertung und die getroffenen Entscheidungen dokumentiert werden, um die Anforderung „Datenschutz durch Technikgestaltung“ (Artikel 25 DSGVO) zu erfüllen. In der Praxis kann dies in der folgenden Form erfolgen:

- **Datenschutzanforderungen**, die für den Kauf (z. B. eine Ausschreibung) oder die Entwicklung von Software, Hardware und Infrastruktur festgelegt werden;
- **Abnahmetests**, mit denen überprüft wird, ob die gewählte Software, die Systeme und die Infrastruktur für den Zweck geeignet sind und einen angemessenen Schutz sowie angemessene Garantien bieten.

Eine solche Dokumentation kann ein integraler Bestandteil der Datenschutz-Folgenabschätzung sein.

Schließlich sollten Sie sich stets bewusst sein, dass gemäß Artikel 32 Absatz 1 Buchstabe d DSGVO der Datenschutz ein Verfahren ist. Daher **sollten Sie die Wirksamkeit der technischen und organisatorischen Maßnahmen regelmäßig überprüfen, bewerten und evaluieren.** Dies ist ein hervorragender Zeitpunkt, um eine Strategie zur Bewältigung dieser Herausforderungen zu entwickeln.

1.2.7 Nutzung des Rechtsrahmens

Die DSGVO enthält einen spezifischen Rechtsrahmen für die Verarbeitung zu wissenschaftlichen Forschungszwecken (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung“, Kapitel „Konzepte“).[579] Ihre KI-Entwicklung stellt eine wissenschaftliche Forschung dar, unabhängig davon, ob sie mit einer Gewinnerzielungsabsicht erstellt wird oder nicht. Daher „können vorbehaltlich der Bedingungen und Garantien gemäß Absatz 1 des vorliegenden Artikels im Unionsrecht oder im Recht der Mitgliedstaaten insoweit Ausnahmen von den Rechten gemäß der Artikel 15, 16, 18 und 21 vorgesehen werden, als diese Rechte voraussichtlich die Verwirklichung der spezifischen Zwecke unmöglich machen oder ernsthaft beeinträchtigen und solche Ausnahmen für die Erfüllung dieser Zwecke notwendig sind“ (Artikel 89 Absatz 2). Darüber hinaus gilt gemäß Artikel 5 Buchstabe b „eine Weiterverarbeitung der erfassten Daten gemäß Artikel 89 Absatz 1 nicht als unvereinbar mit den ursprünglichen Zwecken („Zweckbindung“). Einige andere besondere Ausnahmen von dem allgemein gültigen Rahmenwerk für die Verarbeitung zu Forschungszwecken (z. B. die Beschränkung der Speicherung) sollten ebenfalls berücksichtigt werden."

[579] Dieses besondere Rahmenwerk umfasst auch historische Forschungszwecke oder statistische Zwecke. Die IKT-Forschung ist jedoch in der Regel nicht auf diese Zwecke ausgerichtet. Daher werden wir sie hier nicht analysieren.

Sie können sicherlich von diesem günstigen Rahmenwerk profitieren. Dennoch müssen Sie sich über den konkreten Rechtsrahmen, der für diese Forschung gilt, im Klaren sein (vor allem über die umzusetzenden Garantien). Dieser kann je nach den jeweiliten nationalen Vorschriften wichtige Änderungen enthalten. Zu diesem Zweck wird eine Konsultation mit Ihrem Datenschutzbeauftragten dringend empfohlen.

1.2.8 Definition angemessener Speicherungsrichtlinien

Gemäß Artikel 5 Absatz 1 Buchstabe e DSGVO sollten personenbezogene Daten „in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es für die Zwecke, für die sie verarbeitet werden, erforderlich ist“ (siehe Abschnitt „Speicherbegrenzung“, Kapitel „Grundsätze“). Dieses Erfordernis umfasst zwei Aspekte. Auf der einen Seite geht es um die Identifizierung: Personenbezogene Daten sollten in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es erforderlich ist. Folglich sollten Sie Richtlinien festlegen, um eine Identifizierung zu vermeiden, sobald sie für die Verarbeitung nicht mehr erforderlich ist. Dies umfasst die Ergreifung geeigneter Maßnahmen, um sicherzustellen, dass zu jedem Zeitpunkt nur **das Mindestmaß an Identifizierung verwendet wird, das für die Erfüllung der Zwecke erforderlich ist** (siehe Unterabschnitt „Zeitlicher Aspekt“, Abschnitt „Speicherbegrenzung“, Kapitel „Grundsätze“).

Andererseits bedeutet Datenspeicherung, dass Daten nur für einen **begrenzten Zeitraum** gespeichert werden dürfen: den Zeitraum, der für die Zwecke, für die die Daten verarbeitet werden, unbedingt erforderlich ist. Die Datenschutz-Grundverordnung erlaubt jedoch eine „längere Speicherung“, wenn der einzige Zweck die wissenschaftliche Forschung ist (wie in Ihrem konkreten Fall).

Diese Ausnahme birgt somit das Risiko, dass Sie beschließen, die Daten länger als unbedingt erforderlich zu speichern und so sicherzustellen, dass sie für andere Zwecke als die ursprünglichen, für die sie erhoben wurden, zur Verfügung stehen. Tun Sie dies nur dann, wenn es guten Gründe dafür gibt (wenn beispielsweise Röntgenaufnahmen aus einer Krankenakte stammen, müssen Sie sie in der Krankenakte des Patienten speichern). Sie sollten sich darüber im Klaren sein, dass Sie, auch wenn die DSGVO eine längere Speicherung zulässt, **einen guten Grund für einen solchen längeren Zeitraum haben müssen.** Wenn Sie die Daten somit nicht benötigen und es keine zwingenden rechtlichen Gründe gibt, die Sie zur Speicherung derselben verpflichten, sollten Sie diese besser anonymisieren oder löschen. Dies könnte auch ein hervorragender Zeitpunkt sein, **um Fristen für die Löschung der verschiedenen Datenkategorien vorzusehen und diese Entscheidungen zu dokumentieren** (siehe „Rechenschaftspflicht“ im Kapitel „Grundsätze“).

1.2.9 Benennung eines Datenschutzbeauftragten

Laut Artikel 37 DSGVO müssen Sie einen Datenschutzbeauftragten benennen, da Sie eine umfangreiche Verarbeitung besonderer Kategorien von Daten gemäß Artikel 9 durchführen werden. In jedem Fall sollte das Schlüsselpersonal von Verantwortlichen die Rolle des DSB in Bezug auf das Gesamtmanagement des Projekts ausarbeiten und sicherstellen, dass die Rolle des DSB nicht am Rande steht, sondern in die Entscheidungsprozesse des Unternehmens/Projekts eingebunden ist. Es sollte auch deutlich machen, wie diese Rolle in Bezug auf Aufsicht, Entscheidungsfindung und Ähnliches aussehen könnte.

1.2.10 Sicherstellung der Einhaltung des Rechtsrahmens für Medizinprodukte

Auch wenn sich dieser Leitfaden hauptsächlich mit Datenschutzfragen befasst, können wir nicht umhin zu erwähnen, dass Sie sich bereits in dieser ersten Phase darüber im Klaren sein sollten, dass Sie für eine angemessene Einhaltung des Rechtsrahmens für Medizinprodukte sorgen müssen. Wir beziehen uns hauptsächlich auf die Verordnung (EU) 2017/745 über Medizinprodukte (MDR) und die Verordnung (EU) 2017/746 über In-vitro-Diagnostika (IVDR). Höchstwahrscheinlich wird es nationale, auf diese Themen anwendbare Regelungen geben. Bitte ergreifen Sie Maßnahmen, die auf die Einhaltung dieser Vorschriften abzielen. Hilfreiche Leitlinien zu diesem Zweck finden Sie hier: https://ec.europa.eu/docsroom/documents/40323

Im Hinblick auf die Regulierung von Gesundheitsdaten auf Ebene der Mitgliedstaaten könnte diese Ressource von besonderer Bedeutung sein:

https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf

2 Datenverständnis

2.1 Beschreibung

„Die Phase des Datenverständnisses beginnt mit einer ersten Datenerhebung. Der Analytiker fährt dann fort, sich mit den Daten vertraut zu machen, Probleme in Bezug auf die Datenqualität zu erkennen, erste Erkenntnisse über die Daten zu gewinnen oder interessante Teilmengen zu entdecken, um Hypothesen über verborgene Informationen zu bilden. Die Phase des Datenverständnisses umfasst vier Schritte, darunter die Erhebung von Ausgangsdaten, die Beschreibung der Daten, die Untersuchung der Daten und die Verifizierung der Datenqualität.“[580]

In dieser Phase werden die ersten Daten erhoben und eine erste Untersuchung derselben durchgeführt. Sie umfasst vier aufeinanderfolgende Aufgaben:

- Erhebung von Ausgangsdaten
- Beschreibung der Daten
- Untersuchung der Daten
- Verifizierung der Datenqualität

Alle diese Aufgaben zielen darauf ab, die verfügbaren Daten zu ermitteln. In dieser Phase müssen Sie sich über die Daten, mit denen Sie arbeiten müssen, im Klaren sein und Entscheidungen über die Art und Weise treffen, in der die wichtigsten Grundsätze des Datenschutzes umgesetzt werden sollen.

2.2 Wichtigste zu ergreifende Maßnahmen

In diesem Stadium gibt zahlreiche grundlegende Fragen im Zusammenhang mit dem Schutz personenbezogener Daten, die behandelt werden müssen. Je nach den getroffenen Entscheidungen müssen Grundsätze wie Datenminimierung, Datenschutz durch Technikgestaltung oder durch datenschutzfreundliche Voreinstellungen, Rechtmäßigkeit,

[580] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 15

Verarbeitung nach Treu und Glauben und Transparenz usw. in angemessener Weise geregelt werden.

2.2.1 Art der erhobenen Daten

Laut DSGVO müssen Sie „geeignete technische und organisatorische Maßnahmen treffen, die sicherstellen, dass durch Voreinstellung nur personenbezogene Daten, deren Verarbeitung für den jeweiligen bestimmten Verarbeitungszweck erforderlich ist, verarbeitet werden. Diese Verpflichtung gilt für die Menge der erhobenen personenbezogenen Daten, den Umfang ihrer Verarbeitung, ihre Speicherfrist und ihre Zugänglichkeit. Solche Maßnahmen müssen insbesondere sicherstellen, dass personenbezogene Daten durch Voreinstellungen nicht ohne Eingreifen der Person einer unbestimmten Zahl von natürlichen Personen zugänglich gemacht werden.“[581] (Siehe „Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen“ im Kapitel „Konzepte“). Dies ist spezifisch während dieser Phase zu beachten, da zu diesem Zeitpunkt häufig Entscheidungen über die Art der zu verwendenden Daten getroffen werden. Im Allgemeinen würde Ihre KI in Bezug auf Datenschutzfragen am besten so entwickelt, dass sie ausschließlich Röntgenbilder verwendet. Es könnte jedoch auch interessant sein, Daten über frühere Pathologien wie Alter oder Geschlecht einzubeziehen. Außerdem könnte man über die Verwendung von Daten wie Ernährungsgewohnheiten, Postleitzahl, Sportgewohnheiten usw. nachdenken. Unter Umständen würde sich die Genauigkeit des Modells durch das Hinzufügen zahlreicher neuer Merkmale deutlich erhöhen. Dies muss jedoch nicht unbeding der Fall sein**. Sie sollten abwägen, ob die Einführung zusätzlicher Daten (neben den Röntgenbildern) Diagnosen mit einem ausreichenden Grad an erhöhter Genauigkeit liefert, um deren Verwendung zu rechtfertigen.** Dies lässt sich im Voraus nicht immer eindeutig bewerten, aber zumindest die Trainingsphase sollte diese Frage klären. Wenn die erhöhte Genauigkeit eine unverhältnismäßige Verwendung personenbezogener Daten nicht rechtfertigt, sollte sie vermieden werden.

Prüfen Sie also unbedingt, ob Sie wirklich große Datenmengen benötigen. Intelligente Daten könnten viel nützlicher sein als Big Data. Natürlich ist die Verwendung intelligenter, gut ausgearbeiteter Daten unter Umständen mit einem enormen Aufwand für die Vereinheitlichung, Homogenisierung usw. verbunden, aber es wird die Umsetzung des Grundsatzes der Datenminimierung wesentlich effizienter gestalten. Zu diesem Zweck **könnte ein Expert extrem hilfreich sein, der in der Lage ist, relevante Merkmale auszuwählen.**

Darüber hinaus sollten Sie versuchen, die **Auflösung der Daten auf das für die mit der Verarbeitung verfolgten Zwecke erforderliche Mindestmaß zu beschränken**. Sie sollten zudem **ein optimales Niveau der Datenaggregation bestimmen,** bevor Sie mit der Verarbeitung beginnen (siehe „Angemessen, erheblich und beschränkt“ in Abschnitt „Datenminimierung“, Kapitel „Grundsätze“).

Die Datenminimierung kann im Falle des Deep Learning komplex sein, da eine Unterscheidung nach Merkmalen unter Umständen nicht möglich ist. Es gibt jedoch eine effiziente Möglichkeit, die Menge der erhobenen Daten zu regulieren und sie nur dann zu erhöhen, wenn es notwendig erscheint: die Lernkurve. Sie sollten damit beginnen, eine begrenzte Menge an Trainingsdaten zu erheben und zu verwenden, und dann die Genauigkeit des Modells überwachen, während es mit neuen Daten gefüttert wird.

---

[581] Artikel 25.

2.2.2 Überprüfung der rechtmäßigen Verwendung von Datensätzen

Datensätze können auf unterschiedliche Weise beschafft werden. Als erste Option kann der Entwickler entscheiden, auf eine Datenbank zuzugreifen, die bereits von jemand anderem erstellt wurde. In diesem Fall sollten Sie besonders vorsichtig sein, da der Erwerb des Zugangs zu einer Datenbank mit zahlreichen rechtlichen Fragen verbunden ist (siehe „Zugang zu einer Datenbank", Kapitel „Maßnahmen und Tools").[582]

Eine zweite und sehr gängige Alternative darin, eine Datenbank zu erstellen. Natürlich müssen Sie in diesem Fall sicherstellen, dass Sie bei der Erstellung dieser Datenbank alle rechtlichen Anforderungen der DSGVO erfüllen (siehe „Erstellen einer Datenbank", Kapitel „Maßnahmen und Tools").

Drittens können Sie auch einen alternativen Weg wählen. Sie können **verschiedene Datensätze mischen, um einen großen Trainingsdatensatz und einen weiteren für Validierungszwecke zu erstellen.** Dies könnte einige Probleme mit sich bringen, darunter die Möglichkeit, dass die Kombination dieser personenbezogenen Daten zusätzliche Informationen über die betroffenen Personen liefert. So könnten Sie beispielsweise die betroffenen Personen identifizieren, was bisher nicht möglich war. Dies könnte bedeuten, dass anonymisierte Daten deanonymisiert und neue personenbezogene Informationen geschaffen werden, die im ursprünglichen Datensatz nicht enthalten waren, ein Umstand, der dramatische ethische und rechtliche Fragen aufwerfen würde. Wenn beispielsweise „die betroffenen Personen in Kenntnis der Sachlage in die Verarbeitung personenbezogener Daten in den ursprünglichen Datensätzen für bestimmte Zwecke eingewilligt haben, haben sie damit nicht notwendigerweise auch in die Zusammenführung von Datensätzen und das Data Mining eingewilligt, das neue Informationen offenbart. Neue Informationen, die auf diese Weise gewonnen werden, können auch auf Wahrscheinlichkeiten oder Vermutungen beruhen und daher falsch sein oder Verzerrungen bei der Darstellung von Personen enthalten." [583] Daher sollten Sie solche Folgen nach Möglichkeit vermeiden, indem Sie sicherstellen, dass die Zusammenführung von Datensätzen nicht gegen die Rechte und Interessen der betroffenen Personen verstößt.

Wenn Sie mehrere Datensätze verwenden, die unterschiedlichen Zwecken dienen, sollten Sie geeignete Maßnahmen ergreifen, um die verschiedenen Verarbeitungsvorgänge zu trennen. Andernfalls könnten Sie leicht Daten, die zu einem bestimmten Zweck erhoben wurden, für andere Vorgänge verwenden, was zu Problemen im Zusammenhang mit dem Grundsatz der Zweckbindung führen könnte.

2.2.3 Auswahl der geeigneten Rechtsgrundlage

Sie sollten sich vor Beginn der Verarbeitung für eine Rechtsgrundlage entscheiden, Ihre Entscheidung in einer Datenschutzerklärung (zusammen mit den Zwecken) dokumentieren und die Gründe für Ihre Entscheidung angeben (siehe Abschnitt „Rechenschaftspflicht", Kapitel „Grundsätze").

Sie sollten die **Rechtsgrundlage wählen, die der wahren Natur Ihrer Beziehung zu der betroffenen Person und dem Zweck der Verarbeitung am ehesten entspricht.** Diese

---

[582] Yeong Zee Kin, Legal Issues in AI Deployment, verfügbar unter: https://lawgazette.com.sg/feature/legal-issues-in-ai-deployment/, abgerufen am 15. Mai 2020

[583] SHERPA, Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design approach, 2020, Seite 38. Verfügbar unter: https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf, abgerufen am 15. Mai 2020.

Entscheidung ist von grundlegender Bedeutung, da die Rechtsgrundlage für die Verarbeitung nur dann geändert werden darf, wenn es stichhaltige Gründe gibt, die dies rechtfertigen (siehe Abschnitt „Zweckbindung“, Kapitel „Grundsätze“).

Im Falle eines KI-Tools, das Patientendaten verarbeitet, fühlen sich die Entwickler in der Regel versucht, die Einwilligung als Rechtsgrundlage für die Verarbeitung zu verwenden (siehe Abschnitt „Einwilligung“ im Kapitel „Hauptkonzepte“). Dies könnte sinnvoll sein, wenn Sie Daten wiederverwenden, die bereits für einen anderen Zweck erhoben wurden, und die Einwilligung die Grundlage war, die die primäre Verwendung der Daten erlaubte. In der Tat erlaubt die DSGVO die Wiederverwendung von Daten für wissenschaftliche Zwecke. In Artikel 5 Absatz 1 Buchstabe b heißt es, dass die Weiterverarbeitung für wissenschaftliche Forschungszwecke nicht als unvereinbar mit den ursprünglichen Zwecken („Zweckbindung“) gilt. Im Prinzip könnten Sie also diese Daten auf der Grundlage der ursprünglichen Einwilligung weiterverwenden. Sie müssen jedoch bedenken, dass gemäß Artikel 9 Absatz 4 DSGVO „die Mitgliedstaaten zusätzliche Bedingungen, einschließlich Beschränkungen, einführen oder aufrechterhalten können, soweit die Verarbeitung von genetischen, biometrischen oder Gesundheitsdaten betroffen ist.“ Es kann also durchaus sein, dass Ihre jeweilige nationale Regelung Ausnahmen oder besondere Bedingungen für die Weiterverwendung personenbezogener Daten vorsieht. In jedem Fall sollten Sie immer daran denken, dass Ihre Informationspflichten bestehen bleiben. Sie sollten der betroffenen Person vor jeder weiteren Verarbeitung ihrer Daten Informationen über diesen anderen Zweck und alle weiteren relevanten Informationen gemäß Artikel 13 Absatz 2 DSGVO zur Verfügung stellen.

Die Diskussion über die Wiederverwendung von Daten

Gegenwärtig wird eine lebhafte Diskussion über die Wiederverwendung von Daten zu Forschungszwecken geführt. Laut Artikel 5 Absatz 1 Buchstabe b DSGVO gilt die Weiterverarbeitung für wissenschaftliche Forschungszwecke nicht als unvereinbar mit dem ursprünglichen Zweck. Sofern Ihre nationale Regelung also nichts anderes vorsieht, können Sie die verfügbaren Daten für Forschungszwecke weiterverwenden, da diese mit dem ursprünglichen Zweck, für den sie erhoben wurden, vereinbar sind.

Der EDSB führte jedoch an: „Um die Achtung der Rechte der betroffenen Person zu gewährleisten, sollte die Vereinbarkeitsprüfung gemäß Artikel 6 Absatz 4 vor der Weiterverwendung von Daten für wissenschaftliche Forschungszwecke weiterhin in Betracht gezogen werden, insbesondere wenn die Daten ursprünglich für ganz andere Zwecke oder außerhalb des Bereichs der wissenschaftlichen Forschung erhoben wurden. Laut einer Analyse aus Sicht der medizinischen Forschung sollte die Anwendung dieses Tests in der Tat einfach sein“[584]. Nach dieser Auslegung sollten Sie personenbezogene Daten nur dann weiterverwenden, wenn die Bedingungen von Artikel 6.4 zutreffen.

Diese Auslegung steht in gewisser Weise im Widerspruch zur Auslegung dieser Frage durch den EDSB, der feststellte, dass Artikel 5 Absatz 1 Buchstabe b DSGVO vorsieht, „dass eine Weiterverarbeitung für im öffentlichen Interesse liegende Archivzwecke, für wissenschaftliche oder historische Forschungszwecke oder für statistische Zwecke von vornherein nicht als unvereinbar mit dem ursprünglichen Zweck gelten soll, sofern sie gemäß den Bestimmungen von Artikel 89 erfolgt, der für diese Fälle spezielle geeignete Garantien und Ausnahmen enthält. Wenn dies der Fall ist, könnte der für die Datenverarbeitung Verantwortliche unter bestimmten Voraussetzungen zur Weiterverarbeitung der Daten in der Lage sein, ohne eine neue Rechtsgrundlage zu benötigen. Diese Voraussetzungen werden

[584] EDSB, A preliminary opinion on data protection and scientific research, 6. Januar 2020, Seite 23.

aufgrund ihrer horizontalen Anwendbarkeit und Komplexität künftig besondere Aufmerksamkeit und Anleitung seitens des EDSA erfordern. Zum gegenwärtigen Zeitpunkt sollte die Möglichkeit der Anwendbarkeit der Vermutung der Vereinbarkeit mit den ursprünglichen Zwecken vorbehaltlich der Bedingungen in Artikel 89 nicht stets für die Sekundärnutzung von Daten aus klinischen Prüfungen außerhalb des Prüfplans für andere wissenschaftliche Zwecke ausgeschlossen werden.“[585]

Daher ist die Situation derzeit noch unklar, auch wenn wir der Meinung sind, dass die Auslegung des EDSA sinnvoller ist und sich in Zukunft wahrscheinlich durchsetzen wird.

Wenn Sie neue Daten für Ihre Forschung erheben können, empfehlen wir Ihnen, die Einwilligung als Rechtsgrundlage zu vermeiden. Dies gilt insbesondere dann, wenn die Daten in einer Situation erhoben werden, in der die Patienten eine dringende medizinische Versorgung benötigen, z. B. wenn sie an Symptomen im Zusammenhang mit COVID leiden. Im Zusammenhang mit klinischen Prüfungen hat der EDSA[586] erklärt, dass „auch wenn die Bedingungen für eine Einwilligung nach Aufklärung gemäß der Verordnung über klinische Prüfungen vorliegen, jedes klare Ungleichgewicht in der Machtverteilung zwischen dem Teilnehmer und dem Sponsor bzw. Prüfer bedeutet, dass die Einwilligung nicht im Sinne der DSGVO „freiwillig erteilt“ wurde. Dies wird nach Ansicht des EDSA zum Beispiel dann der Fall sein, wenn ein Teilnehmer einen schlechten Gesundheitszustand aufweist, Teilnehmer einer wirtschaftlich oder sozial benachteiligten Gruppe angehören oder sich in einer institutionellen oder hierarchischen Abhängigkeit befinden. Daher und wie in den Leitlinien der Artikel-29-Datenschutzgruppe in Bezug auf die Einwilligung erläutert, wird die Einwilligung in den meisten Fällen nicht die geeignete Rechtsgrundlage darstellen und daher auf andere Rechtsgrundlagen als die Einwilligung zurückzugreifen sein (siehe alternative Rechtsgrundlagen weiter unten). Der EDSA vertritt daher die Ansicht, dass für die Verarbeitung Verantwortliche die Umstände der jeweiligen klinischen Prüfung besonders gründlich prüfen sollten, bevor sie die Einwilligung der betreffenden Person als Rechtsgrundlage für die Verarbeitung personenbezogener Daten zum Zwecke der Forschungstätigkeiten der jeweiligen Prüfung heranziehen.“

Unseres Erachtens kann diese Stellungnahme auch auf andere Szenarien ausgedehnt werden, in denen das Machtgleichgewicht verzerrt ist. Es kann jedoch vorkommen, dass der entsprechende Ethikausschuss unsere Ansicht nicht teilt. Bitte seien Sie sich solcher Umstände bewusst und versuchen Sie, mögliche Probleme im Voraus zu vermeiden, indem Sie gegebenenfalls den Ausschuss und/oder Ihren DSB und die Aufsichtsbehörden konsultieren.

3 Datenaufbereitung

[585] EDSA, Stellungnahme 3/2019 zu den Fragen und Antworten zum Zusammenspiel der Verordnung über klinische Prüfungen und der Datenschutz-Grundverordnung (DSGVO) (Artikel 70 Absatz 1 Buchstabe b) angenommen am 23. Januar 2019, Seite 9.

[586] STELLUNGNAHME 3/2019 ZU DEN FRAGEN UND ANTWORTEN ZUM ZUSAMMENSPIEL DER VERORDNUNG ÜBER KLINISCHE PRÜFUNGEN UND DER DATENSCHUTZ-GRUNDVERORDNUNG (DSGVO), VERFÜGBAR UNTER: HTTPS://EDPB.EUROPA.EU/SITES/DEFAULT/FILES/FILES/FILE1/EDPB_OPINIONCTRQ_A_FINAL_DE.PDF

3.1 Beschreibung

„Die Phase der Datenaufbereitung umfasst alle Aktivitäten zur Erstellung des endgültigen Datensatzes bzw. der Daten, die aus den anfänglichen Rohdaten in das/die Modellierungswerkzeug(e) eingespeist werden. Zu den Aufgaben gehören die Auswahl von Tabellen, Datensätzen und Attributen sowie die Transformation und Bereinigung der Daten für die Modellierungswerkzeuge. Die fünf Schritte der Datenaufbereitung umfassen die Auswahl der Daten, die Bereinigung der Daten, den Aufbau der Daten, die Integration der Daten und die Formatierung der Daten.“[587]

Diese Phase umfasst alle erforderlichen Aktivitäten, um aus den anfänglichen Rohdaten den endgültigen Datensatz zu erstellen, der dann in das Modell eingespeist wird. Sie umfasst die folgenden fünf Aufgaben, die nicht unbedingt nacheinander ausgeführt werden müssen:

1. Auswahl der Daten. Festlegung der für die Analyse zu verwendenden Daten, auf der Grundlage der Relevanz für die Data-Mining-Ziele, der Qualität sowie der technischen Beschränkungen, wie z. B. Einschränkungen des Datenvolumens oder der Datentypen.

2. Bereinigung der Daten. Anhebung der Datenqualität auf das erforderliche Niveau, beispielsweise durch Auswahl sauberer Teilmengen der Daten, Einfügen von Standardwerten und Schätzung fehlender Daten durch Modellierung.

3. Aufbau der Daten. Der Aufbau neuer Daten durch die Erstellung abgeleiteter Attribute, neuer Datensätze oder transformierter Werte für vorhandene Attribute.

4. Integration der Daten. Kombinieren von Daten aus mehreren Tabellen oder Datensätzen, um neue Datensätze oder Werte zu erstellen.

5. Formatierung der Daten. Syntaktische Änderungen an Daten, die vom Modellierungswerkzeug benötigt werden.

3.2 Wichtigste zu ergreifende Maßnahmen

3.2.1 Einführung der in Artikel 89 DSGVO vorgesehenen Garantien

Da Sie die Daten für wissenschaftliche Zwecke verwenden, müssen Sie sie gemäß den in Artikel 89 DSGVO vorgesehenen Garantien aufbereiten. Wenn die Zwecke Ihrer Forschung durch die Weiterverarbeitung, bei der die Identifizierung von betroffenen Personen nicht oder nicht mehr möglich ist, erfüllt werden können, d. h. durch Pseudonymisierung, sollten diese Zwecke auf diese Weise erfüllt werden. Wenn dies nicht möglich ist, müssen Sie Garantien einführen, die sicherstellen, dass technische und organisatorische Maßnahmen ergriffen werden, mit denen insbesondere die Achtung des Grundsatzes der Datenminimierung gewährleistet wird. Bitte beachten Sie die konkreten Vorschriften, die in Ihrer nationalen Regelung für Garantien festgelegt sind. Konsultieren Sie Ihren Datenschutzbeauftragten.

[587] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 16.

3.2.2 Gewährleistung der sachlichen Richtigkeit in Bezug auf die Verarbeitung personenbezogener Daten

Laut der DSGVO müssen die Daten sachlich richtig sein (siehe „Richtigkeit“, Kapitel „Grundsätze“).

Dies bedeutet, dass die Daten korrekt und auf dem neuesten Stand sind, bezieht sich aber auch auf die Genauigkeit der durchgeführten Analysen. Der EDSB hat die Bedeutung der Richtigkeit der Profilerstellung oder des (nicht ausschließlich) automatisierten Entscheidungsfindungsprozesses in allen Phasen (von der Erhebung der Daten bis zur Anwendung des Profils auf die Person) hervorgehoben.[588]

Die Verantwortlichen müssen die Richtigkeit gewährleisten. Sobald Sie die Datenerhebung abgeschlossen haben, sollten Sie daher geeignete Instrumente einsetzen, um die Richtigkeit der Daten zu gewährleisten. Dies bedeutet in der Regel, dass Sie einige grundlegende Entscheidungen über die technischen und organisatorischen Maßnahmen treffen müssen, damit dieser Grundsatz erfüllt wird (siehe Unterabschnitt „Verbundene technische und organisatorische Maßnahmen“, Abschnitt „Richtigkeit“, Kapitel „Grundsätze“). Da die meisten Daten von Patienten stammen und quantitativ sind, können Sie davon ausgehen, dass sie richtig sind. In jedem Fall erfordert die Richtigkeit der Daten eine angemessene Umsetzung von Maßnahmen, damit betroffene Personen ihr Recht auf Berichtigung in Anspruch nehmen können (siehe Abschnitt „Recht auf Berichtigung“, Kapitel „Rechte der betroffenen Person“).

3.2.3 Konzentration auf Profiling-Probleme

Im Falle einer Datenbank, die ein KI-Tool trainieren oder validieren soll, besteht eine besonders wichtige Verpflichtung, die betroffenen Personen darüber zu informieren, **dass sie betreffende Daten zu einer automatisierten Entscheidungsfindung oder einem Profiling führen könnten, es sei denn, Sie können gewährleisten, dass das Tool in keiner Weise zu diesen Folgen führen wird.** Auch wenn eine automatisierte Entscheidungsfindung im Rahmen der Forschung kaum vorkommen kann, sollten Sie diesem Punkt Aufmerksamkeit schenken. Profiling hingegen kann bei der KI-Entwicklung einige Probleme mit sich bringen.

Laut Artikel 22 Absatz 3 sind automatisierte Entscheidungen, die besondere Kategorien personenbezogener Daten wie beispielsweise Gesundheitsdaten betreffen, nur dann zulässig sind, wenn die betroffene Person eingewilligt hat oder wenn sie auf einer gesetzlichen Grundlage erfolgen. Diese Ausnahme gilt nicht nur, wenn die entsprechenden Daten in diese Kategorie fallen, sondern auch dann, **wenn der Abgleich verschiedener Arten personenbezogener Daten sensible Informationen über Personen offenbaren kann oder wenn abgeleitete Daten in diese Kategorie fallen.**

Einige zusätzliche Maßnahmen können ebenfalls äußerst nützlich sein, um ein unnötiges Profiling zu vermeiden:

- Berücksichtigen Sie bereits **in der Designphase die Systemanforderungen, die zur Unterstützung einer aussagekräftigen Überprüfung durch Personen erforderlich**

---

[588] *Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679* (wp251rev.01). 22.08.*2018, Seite* 13; Ross Ducato, Private Ordering of Online Platforms in Smart Urban Mobility The Case of Uber’s Rating System, CRIDES Working Paper Series no. 3/20202 Februar 2020 Aktualisiert am 26 Juli 2020, Seite 20–21, verfügbar unter: https://poseidon01.ssrn.com/delivery.php?ID=247104118003073117118086021112071111102048023015008020118084071112086000027097102088036101006014057116105116119119026079007006118044033055000114023106007076115096073024007094081002078064098028091093003078095099082108113086098120001079015123027083125024&EXT=pdf&INDEX=TRUE

sind. Insbesondere die Anforderungen an die Interpretierbarkeit und die effektive Gestaltung der Benutzeroberfläche zur Unterstützung von Überprüfungen und Eingriffen durch Personen;

- Gestalten und setzen Sie eine geeignete Schulung und Unterstützung für menschliche Prüfer um.
- Geben Sie dem Personal die entsprechenden Befugnisse, Anreize und Unterstützung, damit es die Bedenken einzelner Personen ansprechen oder eskalieren und gegebenenfalls die Entscheidung des KI-Tools außer Kraft setzen kann.

Wenn Sie ein Profiling oder eine automatisierte Entscheidungsfindung durchführen, müssen Sie die betroffenen Personen über Ihre Entscheidung informieren und alle erforderlichen Informationen gemäß der DSGVO und ggf. nationalen Vorschriften bereitstellen.

3.2.4 Auswahl unverzerrter Daten

Verzerrungen sind eines der Hauptprobleme bei der Entwicklung von KI, ein Problem, das dem Grundsatz der Fairness zuwiderläuft. Verzerrungen können durch viele verschiedene Faktoren verursacht werden. Bei der Erfassung der Daten können sozial bedingte Verzerrungen, Ungenauigkeiten, Fehler und andere Mängel auftreten. Manchmal kann es auch vorkommen, dass Datensätze durch böswillige Handlungen verzerrt werden. Die Eingabe schädlicher Daten in ein KI-Tool kann sein Verhalten verändern, insbesondere bei selbstlernenden Systemen.[589] Daher werfen Fragen im Zusammenhang mit der Zusammensetzung der für das Training verwendeten Datenbanken entscheidende ethische und rechtliche Fragen auf, nicht nur Fragen der Effizienz oder technischer Natur.

Sie müssen diese Fragen somit vor dem Training des Algorithmus klären. Erkennbare und diskriminierende Verzerrungen sollten nach Möglichkeit bereits in der Phase der Datensatzerstellung eliminiert werden. Im Falle von COVID könnte Patienten z. B. nach Alter, Geschlecht oder ethnischer Zugehörigkeit unterschieden werden. Sie müssen sicherstellen, dass der Algorithmus diesen Faktor bei der Auswahl der Daten berücksichtigt. **Die mit der Auswahl der in die Datensätze zu integrierenden Daten betrauten Teams sollten somit aus Personen bestehen, die die Diversität gewährleisten, die von der KI-Entwicklung erwartet wird.** Wenn sich Ihre Daten hauptsächlich auf eine bestimmte Gruppe beziehen, z. B. die kaukasische Bevölkerung im Alter von über vierzig Jahren, dürfen Sie schließlich niemals die Erklärung vergessen, dass der Algorithmus auf dieser Grundlage trainiert wurde und daher bei anderen Bevölkerungsgruppen möglicherweise nicht so gut funktioniert.

4 Modellierung (Training)

4.1 Beschreibung

„In dieser Phase werden verschiedene Modellierungstechniken ausgewählt und angewandt sowie deren Parameter auf optimale Werte kalibriert. In der Regel gibt es für ein und dieselbe Art von Data-Mining-Problem mehrere Techniken. Einige Techniken haben spezifische Anforderungen an die Form der Daten. Daher kann es notwendig sein, zur Phase der

[589] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI, 2019, Seite 17. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020)

Datenaufbereitung zurückzukehren. Die Modellierungsschritte umfassen die Auswahl der Modellierungstechnik, die Erstellung des Testdesigns, die Erstellung der Modelle und die Bewertung der Modelle.“[590]

Diese Phase umfasst mehrere wichtige Aufgaben. Alles in allem müssen Sie:

- Die zu verwendende Modellierungstechnik auswählen. Je nach Art der Technik sind Folgen wie Herleitung, Unklarheit oder Verzerrungen von Daten mehr oder weniger wahrscheinlich.
- Das zu verwendende Trainingstool festlegen. Mit diesem Tool kann der Entwickler messen, wie gut das Modell die Vergangenheit vorhersagen kann, bevor er es für die Vorhersage der Zukunft einsetzt. Zum Training gehören immer empirische Tests mit Daten. Manchmal testen die Entwickler das Modell mit Daten, die sich von den zur Erstellung des Modells verwendeten Daten unterscheiden. Daher kann man in dieser Phase von verschiedenen Arten von Datensätzen sprechen. Unter Umständen kann es schwierig sein, die Personen zu identifizieren, auf die sich die Trainingsdaten beziehen. Dies führt zu Problemen bei der Wahrung der Rechte dieser Personen, was geeigneter Weise angegangen werden sollte.

4.2 Wichtigste zu ergreifende Maßnahmen

4.2.1 Umsetzung des Grundsatzes der Datenminimierung

Nach dem Grundsatz der Datenminimierung müssen Sie die Menge der bereitgestellten Daten und/oder den Umfang der Informationen über die betroffene Person so schnell wie möglich auf das notwendige Maß beschränken. Folglich müssen Sie die in der Trainingsphase verwendeten Daten von allen Informationen bereinigen, die für das Training des Modells nicht unbedingt erforderlich sind. (Siehe Unterabschnitt „Zeitlicher Aspekt“, Abschnitt „Datenminimierung“, Kapitel „Grundsätze“). Es gibt mehrere Strategien, um die Datenminimierung in der Trainingsphase zu gewährleisten. Natürlich sollten Sie damit beginnen, alle personenbezogenen Daten im Zusammenhang mit dem von Ihnen verwendeten Röntgengerät zu löschen, aber dies wäre nur ein erster Schritt zur Einhaltung des Grundsatzes der Minimierung. Zu diesem Zweck sollten sorgfältig strengere Maßnahmen implementiert werden. Die Techniken entwickeln sich ständig weiter. Zu den gängigsten zählen[591] (siehe auch Abschnitt „Integrität und Vertraulichkeit” im Kapitel „Grundsätze”):

- Analyse der von den Daten zu erfüllenden Bedingungen, um als qualitativ hochwertig und mit einer hohen Vorhersagekraft für die spezifische Anwendung zu gelten.

- Kritische Analyse des Umfangs der in jeder Phase des KI-Tools verwendeten Datentypologie.

- Löschung von unstrukturierten Daten und unnötigen Informationen, die bei der Vorverarbeitung der Informationen gesammelt wurden.

- Identifizierung und Unterdrückung der Datenkategorien, die keinen signifikanten Einfluss auf das Lernen oder das Ergebnis der Herleitung haben.

---

[590] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 17.

[591] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, Seite 40. Verfügbar unter: https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf, abgerufen am 15. Mai 2020.

- Unterdrückung irrelevanter Schlussfolgerungen im Zusammenhang mit personenbezogenen Informationen während des Trainingsprozesses, beispielsweise im Falle des unüberwachten Trainings.

- Verwendung von Verifizierungstechniken, die weniger Daten erfordern, wie z. B. die Kreuzvalidierung.

- Analyse und Konfiguration von algorithmischen Hyperparametern, die die Menge oder den Umfang der verarbeiteten Daten beeinflussen könnten, um diese zu minimieren.

- Verwendung von föderierten statt zentralisierten Lernmodellen.

- Anwendung von differenziellen Datenschutzstrategien.

- Training mit verschlüsselten Daten unter Verwendung homomorpher Techniken.

- Datenaggregation.

- Anonymisierung und Pseudonymisierung, nicht nur bei der Datenkommunikation, sondern auch in Bezug auf Trainingsdaten, möglichen personenbezogenen Daten im Modell und bei der Verarbeitung der Herleitung.

4.2.2 Erkennen und Eliminieren von Verzerrungen

Selbst wenn die Mechanismen zur Vermeidung von Verzerrungen bereits in den vorangegangenen Phasen angewandt wurden (siehe vorheriger Abschnitt über das Training), muss dennoch sichergestellt werden, dass die Ergebnisse der Trainingsphase die Verzerrungen auf ein Mindestmaß senken. Dies kann schwierig sein, da einige Arten von Verzerrungen und Diskriminierung oft besonders schwer zu erkennen sind. Die Teammitglieder, die die Eingabedaten kuratieren, sind sich dessen manchmal nicht bewusst, ebenso wenig wie die jeweiligen Nutzer. Daher sind die vom KI-Entwickler in der Validierungsphase eingesetzten Überwachungssysteme äußerst wichtige Faktoren, um Verzerrungen zu vermeiden.

Es gibt zahlreiche technische Hilfsmittel, die sich hervorragend zum Erkennen von Verzerrungen eignen, darunter die algorithmische Folgenabschätzung.[592] Sie sollten daher deren wirksame Umsetzung berücksichtigen.[593] Wie die Literatur zeigt[594], kann es jedoch vorkommen, dass sich ein Algorithmus nicht vollständig von allen verschiedenen Arten von Verzerrungen befreien lässt. Sie sollten jedoch zumindest versuchen, sich ihrer Existenz und der daraus resultierenden Folgen bewusst zu sein (siehe Abschnitte Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz und Richtigkeit, Kapitel Grundsätze).

4.2.3 Ausübung der Rechte der betroffenen Person

Manchmal vervollständigen die Entwickler die verfügbaren Daten durch Herleitungen. Wenn man zum Beispiel nicht über die konkreten Daten zum arteriellen Druck eines Patienten

---

[592] D. Reisman, K. Crawford und M. Whittaker, Algorithmic impact assessments: A practical framework for public agency accountability, 2018, verfügbar unter: https://ainowinstitute.org/aiareport2018.pdf, abgerufen am 15. Mai 2020

[593] https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf, abgerufen am 15. Mai 2020

[594] Alexandra Chouldechova, Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments, Big Data. Band: 5, Ausgabe 2: 1. Juni 2017. 153–163. http://doi.org/10.1089/big.2016.0047

verfügt, kann man einen Algorithmus verwenden, um ihn aus den übrigen Daten herzuleiten. Dies bedeutet jedoch nicht, dass diese Daten als vollständig pseudonymisiert oder anonymisiert betrachtet werden können. Dies gilt insbesondere für genomische Daten, da deren Anonymisierung nahezu unmöglich ist. Es handelt sich somit weiterhin um personenbezogene Daten. Darüber hinaus müssen auch hergeleitete Daten als personenbezogene Daten betrachtet werden. Daher haben die betroffenen Personen einige grundlegende Rechte in Bezug auf diese Daten, die Sie respektieren müssen.

Tatsächlich müssen Sie alle Rechte der betroffenen Personen während des gesamten Lebenszyklus ermöglichen. In dieser speziellen Phase sind das Auskunftsrecht, das Recht auf Berichtigung und das Recht auf Löschung besonders sensibel. Sie weisen bestimmte Merkmale auf, derer sich die Verantwortlichen bewusst sein müssen. Im Falle einer Forschung zu wissenschaftlichen Zwecken wie die von Ihnen entwickelte sieht die DSGVO jedoch einige Garantien und Ausnahmen in Bezug auf die Verarbeitung vor (Artikel 89). Sie müssen sich über die konkrete Regelung in Ihrem Mitgliedstaat im Klaren sein. Laut der DSGVO können im Unionsrecht oder im Recht der Mitgliedstaaten insoweit Ausnahmen von den Rechten gemäß der Artikel 15, 16, 18 und 21 vorgesehen werden, als diese Rechte voraussichtlich die Verwirklichung der spezifischen Zwecke unmöglich machen oder ernsthaft beeinträchtigen und solche Ausnahmen für die Erfüllung dieser Zwecke notwendig sind.

- Auskunftsrecht (siehe „Auskunftsrecht“, Kapitel „Rechte der betroffenen Person“)

Grundsätzlich sollten Sie auf Anträge betroffener Personen in Bezug auf Auskunft über sie betreffende personenbezogenen Daten reagieren, vorausgesetzt, Sie haben angemessene Maßnahmen ergriffen, um die Identität der betroffenen Person zu überprüfen, und es gelten keine weiteren Ausnahmen. Allerdings sind Sie nicht verpflichtet, zusätzliche personenbezogene Daten zwecks Identifizierung der betroffenen Personen in den Trainingsdaten zu erheben oder zu speichern, nur um die Verordnung einzuhalten. Wenn Sie eine betroffene Person in den Trainingsdaten nicht identifizieren können und die betroffene Person keine zusätzlichen Informationen bereitstellen kann, die ihre Identifizierung ermöglichen würden, sind sie nicht verpflichtet, einem Antrag nachzukommen, der nicht erfüllt werden kann.

- Recht auf Berichtigung (siehe Abschnitt „Recht auf Berichtigung“, Kapitel „Rechte der betroffenen Person“)

Im Falle des Rechts auf Berichtigung müssen Sie das Recht auf Berichtigung der Daten gewährleisten, insbesondere der Daten, die durch die vom KI-Tool erstellten Herleitungen und Profile erzeugt wurden. Auch wenn der Zweck von Trainingsdaten darin besteht, Modelle zu trainieren, die auf allgemeinen Mustern in großen Datensätzen beruhen, und einzelne Ungenauigkeiten somit kaum direkte Auswirkungen auf eine betroffene Person haben werden, darf das Recht auf Berichtigung nicht eingeschränkt werden. Sie könnten höchstens eine längere Frist (zwei zusätzliche Monate) für die Berichtigung beantragen, wenn das technische Verfahren besonders komplex ist (Artikel 11 Absatz 3).

- Recht auf Löschung (siehe „Recht auf Löschung“, Kapitel „Rechte der betroffenen Person“)

Die betroffenen Personen haben das Recht, zu verlangen, dass sie betreffende personenbezogene Daten unverzüglich gelöscht werden. Dieses Recht kann jedoch eingeschränkt sein, wenn bestimmte konkrete Umstände vorliegen. Laut dem ICO

„können Unternehmen Anträge auf Löschung von Trainingsdaten erhalten. Unternehmens müssen auf Löschungsanträge reagieren, es sei denn, es gilt eine einschlägige Ausnahmeregelung, und sofern die betroffene Person angemessene Gründe vorbringen kann. Wenn beispielsweise die Trainingsdaten nicht mehr benötigt werden, weil das ML-Modell bereits trainiert wurde, muss das Unternehmen dem Antrag nachkommen. Falls die Entwicklung des Systems jedoch noch nicht abgeschlossen ist, kann es jedoch notwendig sein, die Trainingsdaten für die Zwecke des erneuten Trainings, der Verfeinerung und der Bewertung eines KI-Tools zu speichern. In diesem Fall sollte das Unternehmen von Fall zu Fall entscheiden, ob es dem Antrag nachkommen kann. Einem Antrag auf Löschung von Trainingsdaten nachzukommen, bedeutet jedoch nciht, dass auf diesen Daten beruhenden ML-Modelle gelöscht werden müssen, es sei denn, die Modelle selbst enthalten diese Daten oder können zu deren Herleitung verwendet werden.“[595]

5 Bewertung (Validierung)

5.1 Beschreibung

„Vor der endgültigen Einführung des vom Datenanalysten erstellten Modells sollte das Modell unbedingt gründlicher bewertet und die Konstruktion des Modells überprüft werden, um sicher zu sein, dass es die Geschäftsziele ordnungsgemäß erreicht. In dieser Phase ist die Prüfung von entscheidender Bedeutung, ob ein wichtiger Geschäftsaspekt eventuell nicht ausreichend berücksichtigt wurde. Am Ende dieser Phase sollte der Projektleiter dann entscheiden, wie die Data-Mining-Ergebnisse zu verwenden sind. Die wichtigsten Schritte sind hier die Bewertung der Ergebnisse, die Überprüfung des Prozesses und die Festlegung der nächsten Schritte.“[596]

Diese Phase umfasst mehrere Aufgaben, die wichtige Datenschutzfragen aufwerfen. Alles in allem müssen Sie:

- Die Ergebnisse des Modells bewerten, zum Beispiel, ob es genau ist oder nicht. Zu diesem Zweck könnte der KI-Entwickler sie in der realen Welt testen.
- Den Prozess überprüfen. Sie sollten den Data-Mining-Auftrag überprüfen, um festzustellen, ob ein wichtiger Faktor oder eine wichtige Aufgabe übersehen wurde. Dazu gehören auch Fragen der Qualitätssicherung.

5.2 Wichtigste zu ergreifende Maßnahmen

5.2.1 Prozesse der dynamischen Validierung

Die Validierung der eine KI-Komponente umfassenden Verarbeitung muss unter Bedingungen erfolgen, die die tatsächliche Umgebung widerspiegeln, in der die Verarbeitung

---

[595] ICO, Enabling access, erasure, and rectification rights in AI tools, verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/, abgerufen am 15. Mai 2020

[596] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 17

stattfinden soll. Wenn Sie also im Voraus wissen, wo das KI-Tool eingesetzt werden soll, sollten Sie den Validierungsprozess an diese Umgebung anpassen. Wenn das Tool beispielsweise in Italien eingesetzt werden soll, sollten Sie es mit Daten aus der italienischen Bevölkerung oder, falls dies nicht möglich ist, einer ähnlichen Bevölkerung validieren. Andernfalls könnten die Ergebnisse völlig falsch sein. In jedem Fall sollten Sie jeden möglichen Nutzer über die Bedingungen der Validierung informieren.

Darüber hinaus muss der Validierungsprozess regelmäßig überprüft werden, wenn sich die Bedingungen ändern oder wenn der Verdacht besteht, dass die Lösung selbst verändert werden könnte. Wenn der Algorithmus beispielsweise mit Daten älterer Menschen gefüttert wird, sollten Sie prüfen, ob sich dadurch seine Genauigkeit bei einer jungen Bevölkerung ändert oder nicht. Sie müssen gewährleisten, dass die Validierung die Bedingungen, unter denen der Algorithmus validiert wurde, genau widerspiegelt.

Dazu sollte die Validierung alle Komponenten eines KI-Tools einbeziehen, einschließlich der Daten, im Voraus geschulter Modelle, Umgebungen und des Verhaltens des Systems als Ganzes. Darüber hinaus sollte die Validierung so bald wie möglich erfolgen. Schließlich muss sichergestellt sein, dass die Ergebnisse bzw. Handlungen mit den Ergebnissen der vorangegangenen Prozesse übereinstimmen, indem sie zur Gewährleistung ihrer Konformität mit den zuvor definierten Zielsetzungen verglichen werden.[597] Zur Validierung müssen gelegentlich neue personenbezogene Daten erhoben werden. In einigen anderen Fällen verwenden die Verantwortlichen die Daten für andere Zwecke als die ursprünglich vorgesehenen. In all diesen Fällen sollten die Verantwortlichen die Einhaltung der DSGVO sicherstellen (siehe Abschnitt „Zweckbindung“, Kapitel „Grundsätze“ sowie „Datenschutz und wissenschaftliche Forschung“ in Abschnitt „Konzepte“).

5.2.2 Löschen unnötiger Datensätze

Häufig sind der Validierungs- und Trainingsprozess auf gewisse Weise miteinander verbunden. Wenn im Rahmen der Validierung erforderliche Verbesserungen des Modells erkannt wurden, sollte das Training erneut durchgeführt werden. Sobald die KI-Entwicklung abgeschlossen ist, ist im Prinzip auch die Trainingsphase des KI-Tools beendet. Zu diesem Zeitpunkt sollten Sie den zu diesem Zweck verwendeten Datensatz löschen, es sei denn, es besteht eine rechtliche Notwendigkeit, die Daten für die Weiterentwicklung oder Bewertung des Systems oder für andere Zwecke zu speichern, die gemäß den Bestimmungen von Artikel 6 Absatz 4 DSGVO mit den Zwecken vereinbar sind, für die sie erhoben wurden (siehe Abschnitt „Definition angemessener Speicherungsrichtlinien“ in diesem Dokument). Sie sollten jedoch stets bedenken, dass die Löschung personenbezogener Daten im Widerspruch zur Notwendigkeit stehen kann, die Genauigkeit von Tools zu aktualisieren, die auf dem Echtzeit-Selbstlernen von Algorithmen beruhen: Wird ein Fehler gefunden, müssen Sie wahrscheinlich die zuvor in der Trainingsphase verwendeten Daten abrufen. Beantragen betroffene Personen die Löschung der Daten, müssen Sie einen Einzelfallansatz wählen und dabei etwaige in der Verordnung vorgesehene Einschränkungen dieses Rechts berücksichtigen (siehe Artikel 17 Absatz 3).[598]

---

[597] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI, 2019, Seite 22. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai v

[598] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, Seite 26. Verfügbar unter: https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

5.2.3 Durchführung eines externen Audits der Datenverarbeitung

Da die Risiken des von Ihnen entwickelten Systems hoch sind, **muss ein Audit des Systems durch einen unabhängigen Dritten erwogen werden**. Es können verschiedene Audits durchgeführt werden. Sie können intern oder extern erfolgen, sie können sich nur auf das Endprodukt beziehen oder in Bezug auf weniger weit entwickelte Prototypen durchgeführt werden. Sie können als eine Form der Überwachung oder als ein Transparenzinstrument betrachtet werden.

Im Hinblick auf die rechtliche Richtigkeit müssen KI-Tools daraufhin überprüft werden, ob sie personenbezogene Daten im Einklang mit den Bestimmungen der DSGVO verarbeiten, wobei ein breites Spektrum von Fragen zu berücksichtigen ist, die mit dieser Verarbeitung zusammenhängen könnten. Die Hochrangige Expertengruppe für künstliche Intelligenz stellte in Bezug auf das Erproben und Prüfen fest: „Es sollte von einer möglichst vielfältig zusammengesetzten Personengruppe entworfen und entwickelt werden. Für die Kategorien, die aus verschiedenen Perspektiven getestet werden, sollten vielfältige Metriken entwickelt werden. Kontradiktorische Tests durch vertrauenswürdige, gemischte „rote Teams“, die bewusst versuchen, das System zu „knacken“, um Schwachstellen zu finden, und „Bug-Bounty“-Programme, die Außenstehende dazu anregen, Systemfehler und -schwächen zu erkennen und verantwortungsbewusst zu melden, sollten in Betracht gezogen werden.“[599] Es gibt jedoch gute Gründe, skeptisch zu sein, ob ein Prüfer in der Lage ist, die Funktionsweise eines maschinellen Lernsystems zu überprüfen.

Deshalb ist es sinnvoll, sich auf die Punkte zu konzentrieren, die die spanische Datenschutzbehörde (AEPD) in ihrer empfohlenen Checkliste aufführt: Es wäre einfacher, sich auf die Maßnahmen zur Vermeidung von Verzerrungen, Unklarheiten, verstecktem Profiling usw. und auf den angemessenen Einsatz von Instrumenten wie der Datenschutz-Folgenabschätzung zu konzentrieren, die mehrfach durchgeführt werden kann, als zu versuchen, die Funktionsweise eines komplexen Algorithmus in allen Einzelheiten zu verstehen. Die Umsetzung angemessener Datenschutzmaßnahmen in den ersten Phasen des Lebenszyklus des Tools ist die beste Möglichkeit, um Datenschutzprobleme zu vermeiden.

5.2.4 Sicherstellung der Einhaltung des Rechtsrahmens für Medizinprodukte

Bevor Sie Ihr Gerät in Betrieb nehmen, sollten Sie sich vergewissern, dass Sie die Vorschriften über die Entwicklung von Medizinprodukten ordnungsgemäß befolgt haben. Bitte stellen Sie sicher, dass dies der Fall ist. Eine klinische Bewertung und eine Leistungsbeurteilung sollten ebenfalls erstellt werden. Der Leitfaden „Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software“ (https://ec.europa.eu/docsroom/documents/40323) könnte ein hervorragendes Hilfsmittel zu diesem Zweck sein.

---

[599] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI, 2019, Seite 22. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, abgerufen am 15. Mai 2020

5.2.5 Information des an der Entwicklung beteiligten Gesundheitspersonals über mögliche Probleme

Häufig werden KI-Instrumente validiert, indem ihre Leistung mit der von menschlichen Elementen, in diesem Fall von Angehörigen der Gesundheitsberufe, verglichen wird. Dies kann ungewollt dazu führen, dass ihre Teilnahme eine Bewertung ihrer eigenen beruflichen Fähigkeiten hervorruft. Wenn wir die Erfolgsquote einiger Fachleute mit der anderer vergleichen, könnten einige von ihnen das Gefühl haben, ungewollt getestet zu werden. Es ist sehr wichtig, diesen Effekt zu vermeiden. Wenn er dennoch auftritt, sollten die Teilnehmer gewarnt werden und dies akzeptieren.

6 Bereitstellung

6.1 Beschreibung

„Die Bereitstellung ist der Prozess, bei dem ein IT-System in seiner Umgebung betriebsbereit gemacht wird, einschließlich Installation, Konfiguration, Betrieb, Tests und Vornahme der erforderlichen Änderungen. Die Bereitstellung erfolgt in der Regel nicht durch die Entwickler eines Systems, sondern durch das IT-Team des Kunden. Aber selbst in diesem Fall sind die Entwickler dafür verantwortlich, dem Kunden ausreichende Informationen für eine erfolgreiche Bereitstellung des Modells zu liefern. Dazu gehört in der Regel ein (allgemeiner) Bereitstellungsplan mit den notwendigen Schritten für eine erfolgreiche Bereitstellung und deren Durchführung sowie ein (allgemeiner) Überwachungs- und Wartungsplan für die Wartung des Systems, die Überwachung der Bereitstellung und die korrekte Nutzung der Data-Mining-Ergebnisse.“[600]

6.2 Wichtigste zu ergreifende Maßnahmen

6.2.1 Allgemeine Bemerkungen

Sobald Sie Ihren Algorithmus erstellt haben, stehen Sie vor einem wichtigen Problem. Es könnte passieren, dass er offen oder verdeckt personenbezogene Daten enthält. In jedem Fall müssen Sie eine formale Bewertung durchführen, um festzustellen, welche personenbezogenen Daten der betroffenen Personen identifizierbar sein könnten. Dies kann mitunter kompliziert sein. Beispielsweise könnten einige KI-Tools wie Vector Support Machines (VSM) in der Logik des Modells von Trainingsdaten durch Technikgestaltung enthalten. In anderen Fällen können im Modell Muster gefunden werden, die eine bestimmte Person identifizieren. In all diesen Fällen könnten Unbefugte in der Lage sein, Elemente der Trainingsdaten wiederherzustellen oder durch Analyse des Verhaltens des Modells auf die Person zu schließen, die darin enthalten war. Wenn Sie wissen oder vermuten, dass das KI-Tool personenbezogene Daten enthält (siehe auch Abschnitt „Erwerb oder Förderung des Zugangs zu einer Datenbank“, Kapitel „Maßnahmen und Tools“), sollten Sie:

---

[600] SHERPA, Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design approach, 2020, Seite 13. Verfügbar unter: https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf, abgerufen am 15. Mai 2020.

- Die Daten löschen oder ggf. begründen, warum dies aufgrund der Beeinträchtigung des Modells ganz oder teilweise nicht möglich ist (siehe Abschnitt „Speicherbegrenzung“, Kapitel „Grundsätze“).

- Die Rechtsgrundlage für die Übermittlung personenbezogener Daten an Dritte bestimmen, insbesondere wenn besondere Datenkategorien betroffen sind (siehe Unterabschnitt „Rechtmäßigkeit“, Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“).

- Die betroffene Person über die oben beschriebene Verarbeitung inkenntnissetzen.

- Nachweisen, dass die Richtlinien zum Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen umgesetzt worden sind (insbesondere die Datenminimierung).

- Eine Datenschutz-Folgenabschätzung (DSFA) durchführen.

Schließlich sollten Sie jedoch regelmäßig Prüfungen durchführen, um die Wahrscheinlichkeit beruhend auf dem Stand der Technik zu bewerten, dass personenbezogene Daten aus Modellen hergeleitet werden können, sodass das Risiko einer versehentlichen Offenlegung minimiert wird. Lassen diese Prüfungen auf eine erhebliche Möglichkeit der Offenlegung von Daten schließen, sollten die erforderlichen Maßnahmen zur Verhinderung ergriffen werden (siehe Abschnitt „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“).

6.2.2 Aktualisierung der Informationen

Wird der Algorithmus von einem Dritten implementiert, müssen Sie die Ergebnisse des in den Entwicklungsphasen eingesetzten Validierungs- und Überwachungssystems mitteilen und Ihre Mitarbeit bei der weiteren Überwachung der Ergebnisvalidierung anbieten. Es wäre auch ratsam, diese Art der Koordinierung mit allen Dritten vorzusehen, von denen Sie Datenbanken oder andere relevante Komponenten im Lebenszyklus des Systems erwerben. Wenn es sich um eine Datenverarbeitung durch einen Dritten handelt, müssen Sie sicherstellen, dass der Zugang im Rahmen einer Rechtsgrundlage erfolgt.

Darüber hinaus müssen dem Endnutzer in jeder Phase Echtzeitinformationen über die Werte in Bezug auf die Richtigkeit und/oder Qualität der hergeleiteten Informationen angeboten werden (siehe Abschnitt „Richtigkeit“, Kapitel „Grundsätze“). Wenn die hergeleiteten Informationen die Mindestschwellenwert in Bezug auf die Qualität nicht erreichen, müssen Sie hervorheben, dass diese Informationen keinen Wert besitzen. Diese Anforderung bedeutet häufig, dass Sie detaillierte Informationen über die Trainings- und Validierungsphasen bereitstellen sollten. Besonders wichtig sind Informationen über die zu diesem Zweck verwendeten Datensätze. Andernfalls könnte die verwendete Lösung zu enttäuschenden Ergebnissen für die Endnutzer führen, die dann über die Ursache spekulieren müssen.

Zweites Szenario: KI zur Verhütung, Ermittlung und Aufdeckung von Straftaten

Johann Čas (ITA/OEAW)

Dieser Teil des Leitfadens wurde von Marko Sijan, Senior Advisor Specialist, (HR DPA) überprüft und bestätigt.

Einführung und Vorbemerkungen

Der Einsatz fortschrittlicher IKT spielt – als wesentliche Technologie für alle wirtschaftlichen, staatlichen oder gesellschaftlichen Aktivitäten – eine immer wichtigere Rolle bei der Vorhersage, Verhütung, Untersuchung und Verfolgung krimineller oder terroristischer Aktivitäten. Dementsprechend bildet die Forschung zur Entwicklung und Verbesserung der technischen Fähigkeiten von Strafverfolgungsbehörden einen Schwerpunktbereich vergangener, aktueller und zukünftiger EG-Förderprogramme. Fortgeschrittene und neu entstehende IKT-Services verfügen über beispiellose Fähigkeiten zur Überwachung und Analyse großer und vielfältiger Datensätze, insbesondere in Verbindung mit KI[601]-Technologien. Die Forschung im Bereich dieser Technologien sowie der Einsatz fortgeschrittener IKT im Zusammenhang mit der Sicherheit werfen ernsthafte Bedenken hinsichtlich der Ethik und der Einhaltung von Rechtsvorschriften auf. EU-finanzierte Sicherheitsforschungsprogramme verlangen ausdrücklich die vollständige Einhaltung der Bestimmungen der Charta der Grundrechte der Europäischen Union,[602] die Berücksichtigung des Datenschutzes durch Technikgestaltung sowie durch datenschutzfreundliche Voreinstellungen [603]und zusätzlich zur Selbstbewertungstabelle für ethische Fragen[604] auch die Ausfüllung einer „Societal Impact Table", d. h. einer Tabelle der gesellschaftlichen Auswirkungen. „Eine „Societal Impact Table" ist eine Besonderheit dieses

[601] KI ist ein (zu) häufig verwendeter Begriff, dem eine eindeutige Definition fehlt. Hier beziehen wir uns auf die breite Definition von KI, die von der Hochrangigen Expertengruppe für künstliche Intelligenz entwickelt wurde: „Künstliche-Intelligenz-(KI)-Systeme sind vom Menschen entwickelte Software- (und möglicherweise auch Hardware-) Systeme, die in Bezug auf ein komplexes Ziel auf physischer oder digitaler Ebene agieren, indem sie ihre Umgebung durch Datenerfassung wahrnehmen, die gesammelten strukturierten oder unstrukturierten Daten interpretieren, Schlussfolgerungen daraus ziehen oder die aus diesen Daten abgeleiteten Informationen verarbeiten und über die geeignete(n) Maßnahme(n) zur Erreichung des vorgegebenen Ziels entscheiden. KI-Systeme können entweder symbolische Regeln verwenden oder ein numerisches Modell erlernen, und sie können auch ihr Verhalten anpassen, indem sie analysieren, wie die Umgebung von ihren vorherigen Aktionen beeinflusst wird.
Als wissenschaftliche Disziplin umfasst die KI mehrere Ansätze und Techniken wie z. B. maschinelles Lernen (Beispiele dafür sind „Deep Learning“ und bestärkendes Lernen), maschinelles Denken (es umfasst Planung, Terminierung, Wissensrepräsentation und Schlussfolgerung, Suche und Optimierung) und die Robotik (sie umfasst Steuerung, Wahrnehmung, Sensoren und Aktoren sowie die Einbeziehung aller anderen Techniken in cyber-physische Systeme).“ https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

[602] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2010:083:0389:0403:en:PDF

[603] Weitere Informationen finden Sie unter EDSA. (2019). Leitlinien 4/2019 zu Artikel 25 Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen. Angenommen am 20. Oktober 2020. https://edpb.europa.eu/system/files/2021-04/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_de.pdf.

[604] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf

Teils des Arbeitsprogramms. Diese Tabelle legt den Schwerpunkt auf die gesellschaftlichen Aspekte der Sicherheitsforschung. Es wird geprüft, ob die vorgeschlagene Sicherheitsforschung den Bedürfnissen der Gesellschaft entspricht, ihr nützt und keine negativen Auswirkungen auf die Gesellschaft hat. Die Antragsteller müssen die „Societal Impact Table" als Teil des Einreichungsverfahrens ausfüllen.“ [605] Ähnliche Verfahren sollten auch auf der Ebene der Gestaltung der Arbeitsprogramme eingeführt werden. Es sollten zusätzliche Sicherheitsvorkehrungen vorgesehen werden, damit die Programme keine Forderungen enthalten, die nur schwer oder gar nicht erfüllt werden können, ohne schwerwiegende ethische Fragen aufzuwerfen oder unverhältnismäßige Menschenrechtsverletzungen zu verursachen. Dies könnte durch eine obligatorische Einbeziehung von Vertretern der Zivilgesellschaft sowie von Ethik- und Rechtsexperten in die Expertengruppen erreicht werden, die EU-finanzierte Forschungsprogramme entwerfen.

Diese Garantien sind unerlässlich, um die Sicherheitsforschung mit Grundsätzen wie Menschenrechten und Demokratie in Einklang zu bringen. Dennoch bleiben Bedenken, dass sie die Legitimität von Sicherheitsforschungsprojekten erhöhen könnten, ohne die Einhaltung ethischer und rechtlicher Vorschriften in der Praxis zu garantieren.[606] Der Einsatz von KI im Zusammenhang mit der Vorhersage oder Verhütung von Straftaten stellt eine ernsthafte Bedrohung für die bürgerlichen Freiheiten dar. Eine einfache Abwägung zwischen Sicherheit und Freiheit ist weder angemessen noch ausreichend. Die komplexe Beziehung sollte als eine Art feindselige Symbiose betrachtet werden,[607] was bedeutet, dass beide für das Überleben des anderen notwendig sind.

Um diese Bedenken gebührend zu berücksichtigen, bezieht dieses Szenario auch Informationen aus bestehenden Ausschreibungen zur Sicherheitsforschung im Rahmen des Projekts H2020 ein, insbesondere aus der Ausschreibung H2020-SEC-2016-2017 und aus derzeit laufenden oder kürzlich abgeschlossenen Projekten. MAGNETO[608] (Multimedia Analysis and Correlation Engine for Organised Crime Prevention and Investigation), CONNEXIONs[609] (InterCONnected NEXt-Generation Immersive IoT Platform of Crime and Terrorism DetectiON, PredictiON, InvestigatiON, and PreventiON Services) oder RED-Alert[610] (Real-time Early Detection and Alert System for Online Terrorist Content based on Natural Language Processing, Social Network Analysis, Artificial Intelligence and Complex Event Processing) sind Beispiele für Projekte, die für diese Fallstudie relevant sind. Sie werden im Rahmen der Ausschreibungen zur Einreichung von Vorschlägen für Technologien zur Verhütung, Ermittlung und Eindämmung von Straftaten im Zusammenhang mit der Bekämpfung von Kriminalität und Terrorismus 2016-2017 finanziert.[611] Der ursprüngliche Plan, eines dieser Projekte als konkrete Grundlage für dieses Szenario zu verwenden, wurde aufgegeben, da die meisten bzw. fast alle Ergebnisse der genannten Projekte gemäß den H2020-Vorschriften[612] als geheim eingestuft und nicht öffentlich zugänglich sind. Obwohl die Klassifizierung spezifischer Ergebnisse von Sicherheitsforschungsprojekten notwendig und verständlich sein mag, schränkt sie sicherlich auch die Möglichkeit einer öffentlichen Prüfung

[605] Siehe Seite 5 https://ec.europa.eu/research/participants/data/ref/h2020/wp/2018-2020/main/h2020-wp1820-security_en.pdf

[606] M. Leese, K. Lidén und B. Nikolova, 2019, Putting critique to work: Ethics in EU security research, Security Dialogue 50(1), 59-76 < https://journals.sagepub.com/doi/abs/10.1177/0967010618809554 >.

[607] B. Wittes, (2011). Against a Crude Balance: Platform Security and the Hostile Symbiosis Between Liberty and Security. Project on Law and Security, Harvard Law School and Brookings, <https://www.brookings.edu/wp-content/uploads/2016/06/0921_platform_security_wittes.pdf>

[608] http://www.magneto-h2020.eu/

[609] https://www.connexions-project.eu/

[610] https://redalertproject.eu/

[611] https://cordis.europa.eu/programme/id/H2020_SEC-12-FCT-2016-2017

[612] http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/secur/h2020-hi-guide-classif_en.pdf

und Diskussion dieser Technologien ein, die angesichts der potenziellen Verletzungen von Menschenrechten und europäischen Werten obligatorisch sein sollte.

Die Komplexität dieses Anwendungsfalls wird noch dadurch erhöht, dass für die Forschungs- und Entwicklungsphase einerseits und für die Umsetzungs- und Nutzungsphase andererseits unterschiedliche Vorschriften gelten. Forschungstätigkeiten unterliegen der Datenschutz-Grundverordnung, künftige Anwendungen der Forschungsergebnisse dagegen der (Richtlinie 2016/680,[613] die eine spezifische Umsetzung und Gesetzgebung in einzelnen Mitgliedsstaaten ermöglicht.

Die Entwicklung künstlicher Intelligenz für Sicherheitszwecke erfordert eine besonders sorgfältige und strenge Prüfung und Einhaltung der ethischen Anforderungen im Allgemeinen, d. h. des bereits erwähnte Leitfadens des Programms Horizont 2020 „How to complete your ethics self-assessment“, die entsprechenden Schlüsseldokumente im Zusammenhang mit künstlicher Intelligenz, sowie der Leitlinien der Hochrangigen Expertengruppe für künstliche Intelligenz: „Ethik-Leitlinien für eine vertrauenswürdige KI“[614] und des Weißbuches der der Europäische Kommission „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“[615] sowie zusätzlicher sicherheitsspezifischer Überlegungen und Dokumente, wie sie in der „Societal Impact Table“, der Stellungnahme Nr. 28 der Europäischen Gruppe für Ethik und Naturwissenschaften und neuen Technologien „Ethik von Sicherheits- und Überwachungstechnologien“[616] oder in den vom Europäischen Datenschutzbeauftragten (EDSB) veröffentlichten einschlägigen Dokumenten behandelt werden.[617] Der Vorschlag für eine Verordnung über künstliche Intelligenz befasst sich speziell mit der Nutzung von KI-Technologien zum Zweck der Strafverfolgung und „... enthält eine solide Risiko-Methodik zur Einstufung von Hochrisiko-KI-Systemen, d. h. solchen Systemen, die erhebliche Risiken für die Gesundheit und Sicherheit oder die Grundrechte von Personen bergen. Solche KI-Systeme müssen horizontalen Auflagen für vertrauenswürdige KI genügen und Konformitätsbewertungsverfahren unterzogen werden, bevor sie in der Union in Verkehr gebracht werden dürfen.“[618] In Anhang III ist eine Reihe von KI-Anwendungen für die Strafverfolgung als Hochrisiko-KI-Werkzeuge aufgeführt, für die Konformitätsbewertungsverfahren vorgeschrieben sind.

Die folgende schrittweise Analyse folgt der Struktur und Terminologie des CRISP-DM-Modells[619], wie in der nachstehenden Beschreibung dargelegt. Um die Vergleichbarkeit der Ansätze und Ergebnisse zu erhöhen, wird diese Struktur allgemein auf alle Fallstudien angewandt, die im Rahmen der von PANELFIT durchgeführten MLEs (Mutual Learning

---

[613] Richtlinie (EU) 2016/680 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung sowie zum freien Datenverkehr und zur Aufhebung des Rahmenbeschlusses 2008/977/JI des Rates, Amtsblatt <http://eur-lex.europa.eu/legal-content/EL/TXT/?uri=OJ:L:2016:119:TOC >.

[614] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[615] https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf

[616] Europäische Gruppe für Ethik und Naturwissenschaften und neuen Technologien. (2014). Stellungnahme Nr. 28: Ethik von Sicherheits- und Überwachungstechnologien (10.2796/22379). Abgerufen von Luxemburg: Brüssel: https://publications.europa.eu/en/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-54fc3225c969/language-en/format-PDF/source-77404258

[617] https://edps.europa.eu/data-protection/our-work/subjects_en

[618] Europäische Kommission. (2021). COM(2021) 206 final. Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, Seite 3 <https://ec.europa.eu/newsroom/dae/redirection/document/75788>

[619] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 14.

Encounters) vorgestellt und diskutiert werden. Die Annahme einer gemeinsamen Struktur impliziert, dass die einzelnen Begriffe nicht wörtlich verstanden werden dürfen. Geschäftsverständnis kann beispielsweise bedeuten, eine ganzheitliche Sicht auf die Projektziele und die Mittel und Schritte zu ihrer Erreichung zu entwickeln, wenn das geplante Projekt nicht (primär) kommerzielle Absichten verfolgt. Es bedeutet auch, dass einige der im gemeinsamen Rahmen enthaltenen Schritte oder Aufgaben für verschiedene Kontexte der Fallstudien nicht anwendbar oder weniger relevant sind. So zeichnet sich beispielsweise die erste der vier Hauptaufgaben, die das allgemeine Ziel umfassen, d. h. die Festlegung der Unternehmensziele, durch wenig oder weniger Wahlfreiheit aus, wenn die Ziele wie im vorliegenden Fall in einer Ausschreibung zur Einreichung von Forschungsvorschlägen definiert und beschrieben werden. Aus dieser Aussage kann jedoch nicht abgeleitet werden, dass Wahlfreiheiten überhaupt nicht bestehen oder nicht berücksichtigt werden sollten, sondern dass die Wahlmöglichkeiten für Projektantragsteller im Vergleich zu denen bei der Entscheidung über die Themen von Forschungsaufrufen eingeschränkt sind.

In der Diskussion der Entwurfsfassung mit externen Experten wurden auch Empfehlungen ausgesprochen, die über dieses konkrete Szenario hinausgehen. Dazu zählt beispielsweise die Entwicklung von Lehrplänen für Ethik und deren verpflichtende Integration in das technische Studium oder Angebote von Datenschutz- und Ethikschulungen für Ingenieure. Entsprechende Schulungsprogramme sollten auch für Polizeikräfte (die KI einsetzen) als allgemeine Sensibilisierungsmaßnahme angeboten werden.

SCHRITTWEISE ANALYSE

1 Geschäftsverständnis (Business Understanding)

1.1 Beschreibung

„Im Mittelpunkt der ersten Phase steht ein präzises Verständnis der Projektziele aus geschäftlicher Perspektive, der Abbildung dieses Wissen in einer Data Mining-Problemdefinition und der anschließenden Entwicklung eines vorläufigen Plans, um die Ziele zu erreichen. Um zu verstehen, welche Daten später wie analysiert werden sollen, benötigen Data Mining-Fachleute ein umfassendes Verständnis des Unternehmens, für das sie eine Lösung finden wollen. Die Phase des Geschäftsverständnisses umfasst mehrere wichtige Schritte, darunter die Bestimmung der Geschäftsziele, die Bewertung der Situation, die Bestimmung der Data Mining-Ziele und die Erstellung des Projektplans.“[620]

Im Zusammenhang mit FuE im Bereich der Technologien zur Verhütung, Ermittlung und Aufdeckung von Straftaten, die im Rahmen von H2020 durchgeführt werden, muss die allgemeine Beschreibung und Struktur der Aufgaben entsprechend angepasst werden. Dies kann bedeuten, dass sowohl die Terminologie als auch der konkrete Inhalt der Aufgaben interpretiert und geändert werden müssen, um den besonderen Zielen zu entsprechen.

[620] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 14.

Die oben erwähnten allgemeinen Ziele umfassen vier Hauptaufgaben:

1. Bestimmung der Projektziele: Dies bedeutet:
    a. Ermittlung des primären Ziels sowie der damit verbundenen Fragen, die das Projekt (die geplante Lösung) angehen soll.
    b. Bestimmung des Erfolgsmaßstabs.
2. Bewertung der Situation
    a. Ermittlung der für das Projekt verfügbaren materiellen und personellen Ressourcen.
    b. Ermittlung der verfügbaren Daten verfügbar, um das primäre Ziel zu erreichen.
    c. Auflistung der dem Projekt zugrunde liegenden Annahmen.
    d. Auflistung der Projektrisiken, Auflistung möglicher Lösungen für diese Risiken, Erstellung eines Glossars der Projekt- und Datenverarbeitungsbegriffe und Erstellung einer Kosten-Nutzen-Analyse für das Projekt.
3. Bestimmung der Datenverarbeitungsziele: Festlegung, welcher Grad an Vorhersagegenauigkeit erwartet wird, um das Projekt als erfolgreich zu betrachten.
4. Erstellung eines Projektplans: Beschreiben des vorgesehenen Plans zur Erreichung der Datenverarbeitungsziele, einschließlich der Beschreibung spezifischer Schritte und eines vorgeschlagenen Zeitplans. Bereitstellung einer Bewertung der potenziellen Risiken und einer ersten Einschätzung der für das Projekt erforderlichen Instrumente und Techniken.

1.2 Wichtigste zu ergreifende Maßnahmen

1.2.1 Bestimmung der Projektziele

Für unser Szenario werden die allgemeinen Ziele durch die jeweilige Ausschreibung definiert. Die oben genannten Projekte beziehen sich auf die Ausschreibung SEC-12-FCT-2016-2017: Technologies for prevention, investigation, and mitigation in the context of the fight against crime and terrorism.[621] Die spezifische Herausforderung wird wie folgt beschrieben: „Organisationen des organisierten Verbrechens und des Terrorismus sind bei der Planung, Durchführung und Verschleierung ihrer kriminellen Aktivitäten und der daraus resultierenden Einnahmen oft führend in der technologischen Innovation. Die Strafverfolgungsbehörden hinken oft hinterher, wenn es darum geht, kriminelle Aktivitäten zu bekämpfen, die durch „fortschrittliche“ Technologien unterstützt werden.“

Der Rahmen dieser Ausschreibung umfasst:

- Neue Erkenntnisse und gezielte Technologien zur Bekämpfung alter und neuer Formen der Kriminalität und terroristischer Verhaltensweisen, die durch fortgeschrittene Technologien unterstützt werden;
- Test und Demonstration neu entwickelter Technologien durch die an den Vorschlägen beteiligten Strafverfolgungsbehörden;

[621] https://ec.europa.eu/research/participants/data/ref/h2020/wp/2016_2017/main/h2020-wp1617-security_en.pdf

- Innovative Lehrpläne, Schulungen und (gemeinsame) Übungen zur Erleichterung der EU-weiten Einführung dieser neuen Technologien, insbesondere in den Bereichen der folgenden Unterthemen:

    1. Cyber-Kriminalität: Deanonymisierung/Verfolgung/Verhinderung von virtuellen/Kryptowährungen, wenn sie Untergrundmärkte im Darknet unterstützen;

    2. Erkennung und Neutralisierung von unseriösen/verdächtigen leichten Drohnen/UAV, die Sperrgebiete überfliegen, und gegebenenfalls Einbeziehung der Betreiber von Infrastrukturen als Empfänger.

    3. Videoanalyse im Zusammenhang mit juristischen Ermittlungen

    und ein viertes offenes Unterthema.

Die in dieser Ausschreibung festgelegten Bedingungen lassen einen gewissen, wenn auch begrenzten Ermessensspielraum bei der Gestaltung des Projekts zu. Die Antragsteller können die Art der Technologien frei wählen; nichttechnische Lösungsstrategien scheinen jedoch nicht förderfähig zu sein. Auch wenn das Spektrum der Technologien offenbleibt, fordert die Ausschreibung eindeutig technische Lösungen und schließt damit Ansätze zur Lösung der angesprochenen spezifischen Sicherheitsprobleme ohne die Einbeziehung potenziell stark die Privatsphäre einschränkender Technologien aus. Der Begriff „fortgeschrittene Technologien“ legt zumindest nahe, die Entwicklung und den Einsatz von Technologien der künstlichen Intelligenz und des maschinellen Lernens zu untersuchen. Begrenzte Wahlmöglichkeiten bestehen auch in Bezug auf die Zielsetzung, beispielsweise auf welche Formen von Kriminalität oder terroristischen Verhaltensweisen das Projekt abzielt. Daher ist es wichtig, die Endnutzer (d. h. die Strafverfolgungsbehörden) bereits in der Entscheidungsphase über die Ziele und Mittel zu deren Erreichung einzubeziehen.

Die Auswahl spezifischer Technologien oder, in einem allgemeineren Kontext, spezifischer Methoden beeinflusst auch das Spektrum der mit dem Projekt verbundenen ethischen oder rechtlichen Fragen. Im Falle der Sicherheitsforschung können speziell ausgewählte Technologien – in unserem Fall KI-Tools oder Ansätze des maschinellen Lernens – neben den üblichen ethischen Fragen wie der Verarbeitung personenbezogener Daten auch ethische Bedenken im Zusammenhang mit einem doppelten Verwendungszweck, der ausschließlichen Ausrichtung der Forschung auf zivile Anwendungen oder dem Missbrauch aufwerfen, sodass die entsprechenden besonderen Vorschriften zu berücksichtigen sind.

1.2.2 Entscheidung für technische Lösungen, die Erklärbarkeit und Transparenz gewährleisten

Während Erklärbarkeit und Transparenz allgemeine Anforderungen an KI-Tools darstellen, sind sie im Falle von KI-Technologien, die auf Menschen angewendet werden oder Auswirkungen auf Menschen haben, zwingend vorgeschrieben (siehe auch Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“, Kapitel „Grundsätze“). Im Falle von KI, die zur Profilerstellung oder Entscheidungsfindung in einem Sicherheitskontext eingesetzt wird, sind diese Grundsätze von grundlegender Bedeutung. KI-Werkzeuge sind anfällig für Verzerrungen. Erklärbarkeit und Transparenz können jedoch dazu beitragen, Verzerrungen der mit solchen Methoden erstellten Algorithmen aufzudecken und zu beseitigen. Technologien zur Unterstützung der Verbrechensverhütung, -aufdeckung und -verfolgung müssen nachweisbare und belegbare Ergebnisse als gültige Beweise liefern, auch vor Gericht. Ungenaue Ergebnisse können schwerwiegende Folgen für den Einzelnen haben, insbesondere in Form von falsch positiven Ergebnissen oder verheerenden Folgen im

Falle von falsch negativen Ergebnissen. Daher kann es notwendig sein, das KI-Tool zur Unterstützung menschlicher Entscheidungen einzusetzen, zusammen mit verbindlichen Maßnahmen für den Einsatz. So wird sichergestellt, dass die Verantwortlichen die Vorschläge des Systems nicht einfach zu ihrer eigenen Entscheidung machen, sondern verstehen, dass sie ihre Entscheidung begründen müssen – sowohl bei Befolgung als auch Ablehnung des Vorschlags. Damit die Menschen die Vorschläge eines KI-Tools verstehen können, müssen diese Systeme sehr transparent hinsichtlich der Faktoren sein, die das Ergebnis einer Berechnung beeinflussen. Letztendlich müssen Menschen die Verantwortung und die Haftung für eine Entscheidung übernehmen. Transparenz ist auch wichtig, um ein ausreichendes Verständnis des Modells, der verwendeten Daten und der erzielten Ergebnisse zu gewährleisten, insbesondere im Falle von Beschwerden oder der Notwendigkeit von Beweisen.

Die Entwickler von in diesem Zusammenhang eingesetzten KI-Tools könnten die Umsetzung erleichtern, indem sie unterstützende Anwendungen für den gesamten Entscheidungsprozess programmieren, z. B. ein Pflichtfeld, das ausgefüllt werden muss, wenn eine Entscheidung auf Vorschlag des Systems getroffen wird, bevor das Ergebnis weiterverarbeitet werden kann.

1.2.3 Implementierung eines Schulungsprogramms

In unserem Fall sind „Schulungen und (gemeinsame) Übungen zur Erleichterung der EU-weiten Einführung dieser neuen Technologien“ bereits in der Beschreibung der Ausschreibung enthalten. Solche Schulungen dürfen sich nicht auf die Nutzung der entwickelten Technologien beschränken, sondern müssen bereits zu Beginn der Forschungstätigkeiten beginnen und insbesondere alle Personen einbeziehen, die an der Gestaltung von KI-Technologien beteiligt sind (z. B. Designer von Algorithmen, Entwickler, Programmierer, Datenwissenschaftler, Ingenieure). Diese Maßnahme ist einer der wichtigsten Ratschläge, die vom ersten Moment eines Projekts zur Verhütung, Ermittlung und Aufdeckung von Straftaten an berücksichtigt werden sollten. Designer von Algorithmen, die das erste Glied in der algorithmischen Kette darstellen, sind sich der ethischen und rechtlichen Auswirkungen ihres Handelns wahrscheinlich nicht bewusst. Eines der Hauptprobleme von KI-Tools zur Bekämpfung von Kriminalität und Terrorismus besteht darin, dass sie häufig in großen Datensätzen enthaltene personenbezogene Daten verwenden, die große Teile der Bürger umfassen, z. B. die Nutzer bestimmter sozialer Netzwerke. Auch wenn die Analyse von Massenüberwachungsdaten durch KI-Tools im Rahmen bestimmter nationaler Rechtsordnungen oder der Umsetzung der *Richtlinie über den Datenschutz im Bereich der Strafverfolgung* (Richtlinie 2016/680) zulässig sein mag, ist sie aus mehreren Gründen sehr problematisch. Erstens mag die Erfüllung von Rechtsvorschriften eine notwendige Bedingung für die Einhaltung ethischer Grundsätze sein, kann aber keinesfalls als ausreichende Bedingung angesehen werden. In einem Informationsdokument der Europäischen Kommission über „Ethik und Datenschutz“[622] heißt es ganz klar: „Die Tatsache, dass einige Daten öffentlich zugänglich sind, bedeutet nicht, dass es keine Grenzen für ihre Verwendung gibt“ (siehe Kasten 4 auf Seite 13). Zweitens bedeutet die Erfüllung der nationalen oder EU-Rechtsvorschriften nicht zwangsläufig, dass die Grundrechte rechtlich eingehalten werden. Die Richtlinie über die Vorratsspeicherung von Daten ist ein prominentes Beispiel dafür. Sie wurde vom Gerichtshof der Europäischen Union (EuGH) in einem Urteil vom 8. April 2014 für nichtig erklärt, da der Gerichtshof der Ansicht war, dass die Richtlinie „einen Eingriff von großem Ausmaß und besonderer Schwere in die Grundrechte auf Achtung des Privatlebens und auf den Schutz personenbezogener Daten beinhaltet, der sich nicht auf das absolut Notwendige beschränkt.“ Drittens müssen die öffentliche Meinung und die Akzeptanz der

[622] https://ec.europa.eu/info/sites/default/files/5._h2020_ethics_and_data_protection_0.pdf

Bürger respektiert werden. Umfassende Bürgerbefragungen zu Überwachungstechnologien haben gezeigt, dass die Bürger im Allgemeinen schwerwiegende Eingriffe in ihre Privatsphäre akzeptieren, wenn sie auf einem konkreten und plausiblen Verdacht beruhen, ungezielte Massenüberwachungsmaßnahmen dagegen ablehnen.[623] Die Anwendung von Data Mining zur Aufdeckung krimineller oder terroristischer Aktivitäten kann mit der Suche nach der Stecknadel im Heuhaufen verglichen werden[624]. Zudem wird die Verarbeitung personenbezogene Daten betroffener Personen umfassen, die weder gegenwärtig noch in der Vergangenheit in kriminelle oder terroristische Aktivitäten verwickelt waren. Je nach Ausrichtung der analysierten Daten können die verarbeiteten Daten überwiegend oder fast ausschließlich unschuldige Personen betreffen. Eine solche Datenverarbeitung verstößt gegen die Unschuldsvermutung, verändert das Verhältnis zwischen Bürger und Staat und kann schwerwiegende gesellschaftliche und individuelle (im Falle falsch positiver Ergebnisse) Folgen haben.[625][626]

Als Entwickler von Algorithmen müssen Sie daher die Folgen Ihres Handelns sowohl für den Einzelnen als auch für die Gesellschaft verstehen und sich Ihrer Verantwortung bewusst sein, indem Sie lernen, ständig aufmerksam und wachsam zu sein. Die Befolgung dieses Ratschlags kann Ihnen helfen, viele ethische und rechtliche Probleme zu vermeiden oder zu mildern. In diesem Zusammenhang könnte eine optimale Schulung aller am Projekt beteiligten Personen noch vor Projektbeginn eines der effizientesten Instrumente sein, um Zeit und Ressourcen im Hinblick auf die Einhaltung von Datenschutz, Ethik, EU- und nationalem Recht oder gesellschaftlicher Akzeptanz zu sparen. Dazu gehört auch die Beteiligung von Ethik- und Rechtsexperten sowohl an Schulungsmaßnahmen als auch an der Durchführung des Projekts. Angemessene Maßnahmen zur Gewährleistung der Vertraulichkeit sollten ebenfalls besonders beachtet werden (siehe Unterabschnitt „Maßnahmen zur Unterstützung der Vertraulichkeit“, Abschnitt „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“). Die Sicherheit und Vertraulichkeit der verarbeiteten Daten ist einerseits unerlässlich, andererseits sind allgemeine Kenntnisse über die Art der verarbeiteten Daten, die betroffenen Personen oder die verwendeten Algorithmen zwingend erforderlich, um die Einhaltung der Menschenrechte und der europäischen Werte zu gewährleisten. Die Einhaltung der Vorschriften des restriktivsten Mitgliedstaates unterstützt zudem die Geschäftsziele, da sie die Implementierung und Nutzung entwickelter Systeme ermöglicht, ohne dass individuelle Anpassungen erforderlich sind.

1.2.4 Nutzung des für die Datenverarbeitung geltenden Rechtsrahmens

Bei sicherheitsrelevanten FuE-Projekten ist dieser Schritt besonders komplex und anspruchsvoll. Für das Forschungsprojekt als solches gelten die Bestimmungen der DSGVO; für spätere Umsetzungen müssen jedoch die Regeln und Bestimmungen der *Richtlinie über*

---

[623] S. Strauß (2015). *D 6.10–Citizen Summits on Privacy, Security and Surveillance: Synthesis Report*. <http://surprise-project.eu/wp-content/uploads/2015/02/SurPRISE-D6.10-Synthesis-report.pdf>

[624] Das bedeutet auch, dass die Suche nach mehr Daten nur den Heuhaufen vergrößert, nicht unbedingt die Anzahl der Stecknadeln.

[625] Richtlinie 2006/24/EG des Europäischen Parlaments und des Rates vom 15. März 2006 über die Vorratsspeicherung von Daten, die bei der Bereitstellung öffentlich zugänglicher elektronischer Kommunikationsdienste oder öffentlicher Kommunikationsnetze erzeugt oder verarbeitet werden, und zur Änderung der Richtlinie 2002/58/EG, Amtsblatt der Europäischen Union <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0024&from=en>

[626] https://www.europarl.europa.eu/legislative-train/theme-area-of-justice-and-fundamental-rights/file-data-retention-directive

den Datenschutz im Bereich der Strafverfolgung (Richtlinie 2016/680) befolgt werden. Darüber hinaus sind möglicherweise abweichende Gesetzgebungen der beteiligten (Mitglied-)Staaten zu berücksichtigen. Daher müssen die entwickelten Technologien und Systeme zumindest anpassungsfähig und flexibel sein, um mit unterschiedlichen Regelungen umgehen zu können. Aus menschenrechtlicher und ethischer Sicht sollte die Einhaltung der restriktivsten Vorschriften in die entwickelten Technologien integriert werden, um die größtmögliche Achtung der Grundrechte und der damit verbundenen Werte zu gewährleisten und gleichzeitig, wie bereits erwähnt, die Notwendigkeit von Änderungen zu verringern oder zu beseitigen, wenn sie in Ländern mit abweichenden Vorschriften angewendet werden.

Gemäß Artikel 5 Absatz 1 Buchstabe b DSGVO müssen personenbezogene Daten „für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden". Das Konzept der Legitimität ist in der DSGVO nicht genau definiert. Laut der Artikel 29-Datenschutzgruppe beinhaltet die Legitimität jedoch, dass die Daten „im Einklang mit dem Gesetz" verarbeitet werden müssen, wobei „Gesetz" als weit gefasster Begriff verstanden werden sollte, der „alle Formen des geschriebenen Rechts und des Gewohnheitsrechts, primäre und sekundäre Rechtsvorschriften, kommunale Verordnungen, gerichtliche Präzedenzfälle, Verfassungsgrundsätze, Grundrechte, andere Rechtsgrundsätze sowie die Rechtsprechung umfasst, d. h. wie ein solches „Gesetz" von einem zuständigen Gericht ausgelegt und berücksichtigt würde".[627]

Es handelt sich also um ein umfassenderes Konzept als das der Rechtmäßigkeit. Es umfasst die Einhaltung der wichtigsten Werte der geltenden Verordnung und der wichtigsten ethischen Grundsätze. So werden beispielsweise einige konkrete KI-Tools das Eingreifen einer Ethik-Kommission erfordern. In einigen anderen Fällen könnten Richtlinien oder andere weiche Regelungen anwendbar sein. Sie sollten eine angemessene Erfüllung dieser Anforderungen gewährleisten, indem sie bereits für diese Vorstufe des Tool-Lebenszyklus einen entsprechenden Plan erstellen (siehe „Legitimität und Rechtmäßigkeit" in „Rechtmäßigkeit, Fairness und Transparenz", Kapitel „Grundsätze"). Zu diesem Zweck sollten Sie sich insbesondere über die Anforderungen im Klaren sein, die sich aus den auf nationaler Ebene geltenden Vorschriften ergeben. Die Entwicklung von Algorithmen zur Verhütung, Ermittlung und Aufdeckung von Straftaten erfordert eindeutig die frühzeitige Einbeziehung von Ethikausschüssen und gemäß Art. 35 DS-GVO eine Datenschutz-Folgenabschätzung durchzuführen. Wie bereits erwähnt, ist gemäß Art. 10 DSGVO zu prüfen, ob die Verarbeitung personenbezogener Daten über strafrechtliche Verurteilungen und Straftaten oder damit zusammenhängende Sicherungsmaßregeln nach dem Unionsrecht oder dem Recht der Mitgliedstaaten zulässig ist. Vergewissern Sie sich, dass Ihr Forschungsplan all diesen Anforderungen in beiden Phasen, d. h. bei der Durchführung des Forschungsprojekts und der künftigen Umsetzung der entwickelten Systeme, gerecht wird.

Der Ethik-Leitfaden für EU-finanzierte Forschung (siehe Fußnote 604) stellt einen umfassenden Rahmen für die Überprüfung der Einhaltung der Ethikvorschriften dar, der zusätzlich zu den institutionellen Ethikvorschriften oder Verhaltenskodizes konsultiert werden sollte, unabhängig davon, ob Ihre Forschung tatsächlich von der EU finanziert wird. Bitte beachten Sie, dass die ethische Bewertung keine Checkliste ist, sondern immer auch eine Abwägung potenziell widersprüchlicher Normen beinhaltet. Insbesondere die Anwendung neuer IKT und die Berücksichtigung des „Datenschutzes durch Technikgestaltung" in einem

[627] Artikel-29-Datenschutzgruppe (2013) Stellungnahme 03/2013 zur Zweckbindung, angenommen am 2. April 2013, WP203. Europäische Kommission, Brüssel, Seite 20. Verfügbar unter: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

so sensiblen Bereich erfordern vorausschauendes Denken auf beiden Seiten, d. h. bei den beteiligten Forschern und den Ethik-Bewertern.

Selbst wenn Ihr Projekt oder Ihre Forschungseinrichtung keinen spezifischen ethischen Vorschriften unterliegt, ist die Beachtung und Einhaltung der einschlägigen nationalen oder EU-Vorschriften unerlässlich. Sobald Sie die entwickelten Technologien und Systeme auf den Markt bringen, ist die Einhaltung der Vorschriften sowohl für die Bereitstellung innerhalb der EU als auch für die Erlangung von Exportlizenzen für die kommerzielle Nutzung außerhalb der EU unerlässlich.

1.2.5 Annahme eines risikobasierten Denkansatzes

Die Erstellung Ihres Algorithmus wird wahrscheinlich die Verwendung mehrerer besonderer Kategorien personenbezogener Daten beinhalten, darunter politische Meinungen, religiöse oder philosophische Überzeugungen oder Daten über das Sexualleben oder die sexuelle Ausrichtung einer natürlichen Person im Falle des Data Mining in sozialen Netzwerken. Daher müssen Sie sicherstellen, dass Sie geeignete Maßnahmen ergreifen, um die Risiken für die Rechte, Interessen und Freiheiten der betroffenen Personen zu minimieren (siehe „Integrität und Vertraulichkeit“, Kapitel „Grundsätze“). Zu diesem Zweck müssen Sie die Risiken für die Rechte und Freiheiten der am Forschungs- und Entwicklungsprozess beteiligten Personen bewerten und beurteilen, welche Maßnahmen zu deren Schutz angemessen sind. In jedem Fall müssen Sie die Einhaltung der Datenschutzanforderungen sicherstellen.

Im Zusammenhang mit Technologien zur Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten macht ein risikobasierter Ansatz eine Datenschutz-Folgenabschätzung obligatorisch, da zwangsläufig mindestens eine der drei spezifischen Bedingungen von Art. 35 Absatz 3 DSGVO zutreffen wird:

„*3. Eine Datenschutz-Folgenabschätzung gemäß Absatz 1 ist insbesondere in folgenden Fällen erforderlich:*

a) systematische und umfassende Bewertung persönlicher Aspekte natürlicher Personen, die sich auf automatisierte Verarbeitung einschließlich Profiling gründet und die ihrerseits als Grundlage für Entscheidungen dient, die Rechtswirkung gegenüber natürlichen Personen entfalten oder diese in ähnlich erheblicher Weise beeinträchtigen;

b) umfangreiche Verarbeitung besonderer Kategorien von personenbezogenen Daten gemäß Artikel 9 Absatz 1 oder von personenbezogenen Daten über strafrechtliche Verurteilungen und Straftaten gemäß Artikel 10 oder

c) systematische umfangreiche Überwachung öffentlich zugänglicher Bereiche.“

Die risikobasierte Analyse sollte auch potenzielle ethische Fragen im Zusammenhang mit dem Missbrauch[628] der entwickelten Technologien und den Ausfuhrbeschränkungen im Zusammenhang mit doppeltem Verwendungszweck,[629] die für die entwickelten Systeme gelten können, berücksichtigen.

Zu bedenken ist auch, dass sich die Risiken nicht auf den Datenschutz und die Verletzung der Privatsphäre durch die entwickelten Systeme beschränken. Auch verfassungsmäßige Rechte und andere Menschenrechte wie die Unschuldsvermutung, der gleichberechtigte Zugang zur Justiz, die Nichtdiskriminierung oder die Meinungsfreiheit können verletzt oder beeinträchtigt

[628] Siehe https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-misuse_en.pdf
[629] Siehe https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-dual-use_en.pdf

werden. Darüber hinaus beschränken sich diese Auswirkungen nicht auf potenzielle Verdächtige, sondern betreffen die Gesellschaft als Ganzes. Sie werden durch den Mangel an Transparenz und menschlicher Kontrollierbarkeit vieler KI-Tools noch verschärft.

1.2.6 Vorbereiten der Dokumentation der Verarbeitung

Wer auch immer personenbezogene Daten verarbeitet (einschließlich Verantwortlicher und Auftragsverarbeiter) verarbeitet, muss seine Tätigkeiten in erster Linie zur Verwendung durch qualifizierte/zuständige Aufsichtsbehörden dokumentieren. Dies muss durch Verzeichnisse über die Verarbeitung geschehen, die von Ihrem Unternehmen zentral für alle Verarbeitungstätigkeiten geführt werden, sowie durch zusätzliche Dokumentation, die sich auf eine einzelne Datenverarbeitungstätigkeit bezieht (siehe Abschnitt „Dokumentation der Verarbeitung", Kapitel „Maßnahmen und Tools"). Diese Vorbereitungsphase ist der ideale Zeitpunkt, um eine systematische Erfassung der erforderlichen Dokumentation festzulegen, da Sie zu diesem Zeitpunkt die Verarbeitungstätigkeit konzipieren und planen.

Die Entwicklung Ihres KI-Tools kann die Verwendung verschiedener Datensätze beinhalten. Die Verzeichnisse müssen die Rückverfolgbarkeit der Verarbeitung, Informationen über die mögliche Wiederverwendung von Daten und die Verwendung von Daten, die zu verschiedenen Datensätzen gehören, in verschiedenen oder denselben Phasen des Lebenszyklus sicherstellen.

Bei Systemen, die zu Strafverfolgungszwecken eingesetzt werden, muss die Dokumentation der Verarbeitung auch die Dokumentation des Zugriffs auf das System nach seiner Implementierung umfassen, um einen möglichen Missbrauch, z. B. den unbefugten Zugriff auf generierte Ergebnisse, zu verhindern und zu erkennen.

Wie in den Anforderungen und Akzeptanztests für den Kauf und/oder die Entwicklung der verwendeten Software, Hardware und Infrastruktur (Unterabschnitt des Abschnitts „Dokumentation der Verarbeitung") dargelegt, müssen die Risikobewertung und die getroffenen Entscheidungen *dokumentiert werden, um die Anforderung „Datenschutz durch Technikgestaltung" (Artikel 25 DSGVO) zu erfüllen. In der Praxis kann dies in der folgenden Form erfolgen:*

> ***Datenschutzanforderungen**, die für den Kauf (z. B. eine Ausschreibung) oder die Entwicklung von Software, Hardware und Infrastruktur festgelegt werden;*
>
> ***Abnahmetests**, mit denen überprüft wird, ob die gewählte Software, die Systeme und die Infrastruktur für den Zweck geeignet sind und einen angemessenen Schutz sowie angemessene Garantien bieten.*
>
> *Eine solche Dokumentation sollte ein integraler Bestandteil der Datenschutz-Folgenabschätzung sein.*

Schließlich sollten Sie sich stets bewusst sein, dass gemäß Artikel 32 Absatz 1 Buchstabe d DSGVO der Datenschutz ein Verfahren ist. Daher **sollten Sie die Wirksamkeit der technischen und organisatorischen Maßnahmen regelmäßig überprüfen, bewerten und evaluieren.** Dies ist ein perfekter Zeitpunkt, um eine Strategie zur Bewältigung dieser Herausforderungen zu entwickeln.

1.2.7 Überprüfung des Rechtsrahmens

Die DSGVO enthält spezifische Vorschriften für die Verarbeitung zu wissenschaftlichen Forschungszwecken (siehe Abschnitt „Datenschutz und wissenschaftliche Forschung",

Kapitel „Konzepte“).[630] Ihr KI-Tool könnte als wissenschaftliche Forschung eingestuft werden, unabhängig davon, ob es mit einer Gewinnerzielungsabsicht erstellt wird oder nicht. *Ist dies der Fall, „können vorbehaltlich der Bedingungen und Garantien gemäß Absatz 1 des vorliegenden Artikels im Unionsrecht oder im Recht der Mitgliedstaaten insoweit Ausnahmen von den Rechten gemäß der Artikel 15, 16, 18 und 21 vorgesehen werden, als diese Rechte voraussichtlich die Verwirklichung der spezifischen Zwecke unmöglich machen oder ernsthaft beeinträchtigen und solche Ausnahmen für die Erfüllung dieser Zwecke notwendig sind“ (Artikel 89 Absatz 2 DSGVO). Darüber hinaus gilt gemäß Artikel 5 Buchstabe b „eine Weiterverarbeitung der erfassten Daten gemäß Artikel 89 Absatz 1 nicht als unvereinbar mit den ursprünglichen Zwecken („Zweckbindung“). Einige andere besondere Ausnahmen von dem allgemein gültigen Rahmenwerk für die Verarbeitung zu Forschungszwecken (z. B. die Beschränkung der Speicherung) sollten ebenfalls berücksichtigt werden."*

Je nach Land, in dem die Forschung durchgeführt wird, und je nach Rechtsform der beteiligten Partner, z. B. ob es sich um akademische oder kommerzielle Einrichtungen handelt, können Sie möglicherweise von diesem günstigen Rahmen profitieren. Dennoch müssen Sie sich über die konkreten (nationalen) Vorschriften, die für diese Forschung gelten, im Klaren sein (vor allem über die umzusetzenden Garantien). Diese können je nach den jeweiligen nationalen Gesetzen spezifische Anforderungen enthalten.

Vorsicht bedeutet auch, dass Sie sowohl die rechtlichen als auch die ethischen Grenzen der geplanten Forschung berücksichtigen müssen. Nur weil bestimmte (nationale) Vorschriften die geplante Datenverarbeitung zulassen, heißt das nicht, dass sie auch aus ethischer Sicht akzeptabel oder konform ist. Analog dazu darf die Einhaltung ethischer Grundsätze nicht als Flucht[631] vor Vorschriften missbraucht werden.

1.2.8 Definition der Speicherungsrichtlinien

Gemäß Artikel 5 Absatz 1 Buchstabe e DSGVO sollten personenbezogene Daten *„in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es für die Zwecke, für die sie verarbeitet werden, erforderlich ist“* (siehe Abschnitt „Speicherbegrenzung“, Kapitel „Grundsätze“). Dieses Erfordernis umfasst zwei Aspekte. Auf der einen Seite geht es um die Identifizierung: Personenbezogene Daten sollten in einer Form gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es erforderlich ist. Folglich sollten Sie Richtlinien festlegen, um eine Identifizierung zu vermeiden, sobald sie für die Verarbeitung nicht mehr erforderlich ist. Solch Richtlinien umfassen die Ergreifung geeigneter Maßnahmen, um sicherzustellen, dass zu jedem Zeitpunkt nur **das Mindestmaß an Identifizierung verwendet wird, das für die Erfüllung der Zwecke erforderlich ist** (siehe Unterabschnitt „Zeitlicher Aspekt“, Abschnitt „Speicherbegrenzung“, Kapitel „Grundsätze“).

Andererseits bedeutet Datenspeicherung, dass Daten nur für einen **begrenzten Zeitraum** gespeichert werden dürfen: den Zeitraum, der für die Zwecke, für die die Daten verarbeitet werden, unbedingt erforderlich ist. Die Datenschutz-Grundverordnung erlaubt jedoch eine

---

[630] Dieses besondere Rahmenwerk umfasst auch historische Forschungszwecke oder statistische Zwecke. Die IKT-Forschung ist jedoch in der Regel nicht auf diese Zwecke ausgerichtet. Daher werden wir sie hier nicht analysieren.

[631] B. Wagner (2018). Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping? In E. BayamlioĞLu, I. Baraliuc, L. Janssens und M. Hildebrandt (Eds.), Being Profiled (Seite 84–89): Amsterdam University Press.

„längere Speicherung“, wenn der einzige Zweck die wissenschaftliche Forschung ist (was in der FuE-Phase der Fall sein könnte).

Die Ausnahme für die wissenschaftliche Forschung birgt das Risiko, dass Sie beschließen, die Daten länger zu speichern als unbedingt erforderlich. Sie sollten sich darüber im Klaren sein, dass Sie, auch wenn die DSGVO eine längere Speicherung zulässt, **einen berechtigten Grund für einen solchen längeren Zeitraum haben müssen.** Bei den von Ihnen entwickelten Systemen müssen Sie organisatorische und technische Vorkehrungen treffen, um die verschiedenen nationalen Rechtsvorschriften über die maximalen Datenspeicherfristen einzuhalten. Dies könnte auch ein hervorragender Zeitpunkt sein, **um Fristen für die (automatische) Löschung der verschiedenen Datenkategorien vorzusehen und diese Entscheidungen zu dokumentieren** (siehe „Rechenschaftspflicht“ im Kapitel „Grundsätze“).

1.2.9 Benennung eines Datenschutzbeauftragten

Gemäß Art. 37 Absatz 1 DSGVO müssen Sie einen Datenschutzbeauftragten benennen:

„1. Der Verantwortliche und der Auftragsverarbeiter benennen auf jeden Fall einen Datenschutzbeauftragten, wenn

> *a) die Verarbeitung von einer Behörde oder öffentlichen Stelle durchgeführt wird, mit Ausnahme von Gerichten, soweit sie im Rahmen ihrer justiziellen Tätigkeit handeln,*
>
> *b) die Kerntätigkeit des Verantwortlichen oder des Auftragsverarbeiters in der Durchführung von Verarbeitungsvorgängen besteht, welche aufgrund ihrer Art, ihres Umfangs und/oder ihrer Zwecke eine umfangreiche regelmäßige und systematische Überwachung von betroffenen Personen erforderlich machen, oder*
>
> *c) die Kerntätigkeit des Verantwortlichen oder des Auftragsverarbeiters in der umfangreichen Verarbeitung besonderer Kategorien von Daten gemäß Artikel 9 oder von personenbezogenen Daten über strafrechtliche Verurteilungen und Straftaten gemäß Artikel 10 besteht.“*

2 Datenverständnis

2.1 Beschreibung

„Die Phase des Datenverständnisses beginnt mit einer ersten Datenerhebung. Der Analytiker fährt dann fort, sich mit den Daten vertraut zu machen, Probleme in Bezug auf die Datenqualität zu erkennen, erste Erkenntnisse über die Daten zu gewinnen oder interessante Teilmengen zu entdecken, um Hypothesen über verborgene Informationen zu bilden. Die Phase des Datenverständnisses umfasst vier Schritte, darunter die Erhebung von Ausgangsdaten, die Beschreibung der Daten, die Untersuchung der Daten und die Verifizierung der Datenqualität.“[632]

Alle diese Schritte zielen darauf ab, die verfügbaren Daten zu ermitteln. In dieser Phase müssen Sie sich über die Daten, mit denen Sie arbeiten müssen, im Klaren sein und Entscheidungen über die Art und Weise treffen, in der die wichtigsten Grundsätze des Datenschutzes umgesetzt werden sollen. Sie sollten das Dokument „Ethics and data

[632] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 15

protection“ vom 14. November 2018[633] konsultieren, um die rechtlichen und ethischen Anforderungen zu erfüllen. Im Falle der Verwendung von Daten aus sozialen Netzwerken sind die Informationen in Kasten 4 „Verwendung von „Open-Source“-Daten“ auf Seite 13 besonders relevant.

Sie sollten sich auch darüber im Klaren sein, dass Datenbanken, die personenbezogene Daten über strafrechtliche Verurteilungen und Straftaten enthalten, sensibel sind und Sie als Entwickler in der Regel keinen Zugang zu ihnen haben.

2.2 Wichtigste zu ergreifende Maßnahmen

In diesem Stadium müssen zahlreiche grundlegende Fragen im Zusammenhang mit dem Schutz personenbezogener Daten behandelt werden. Je nach den getroffenen Entscheidungen müssen Grundsätze wie Datenminimierung, Datenschutz durch Technikgestaltung oder durch datenschutzfreundliche Voreinstellungen, Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz usw. in angemessener Weise geregelt werden. Es muss eine Kommunikation zwischen Ethik- und Rechtsexperten einerseits und Projektentwicklern andererseits hergestellt werden, um die Grundsätze des „Datenschutzes durch Technikgestaltung“ oder „durch datenschutzfreundliche Voreinstellungen“ verwirklichen zu können.

2.2.1 Entscheidung über die Art der zu verarbeitenden Daten

Laut DSGVO trifft der Verantwortliche *„geeignete technische und organisatorische Maßnahmen, die sicherstellen, dass durch Voreinstellung nur personenbezogene Daten, deren Verarbeitung für den jeweiligen bestimmten Verarbeitungszweck erforderlich ist, verarbeitet werden. Diese Verpflichtung gilt für die Menge der erhobenen personenbezogenen Daten, den Umfang ihrer Verarbeitung, ihre Speicherfrist und ihre Zugänglichkeit. Solche Maßnahmen müssen insbesondere sicherstellen, dass personenbezogene Daten durch Voreinstellungen nicht ohne Eingreifen der Person einer unbestimmten Zahl von natürlichen Personen zugänglich gemacht werden.“*[634] (Siehe „Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen“ im Kapitel „Konzepte“). Diese Forderung ist spezifisch während dieser Phase zu beachten, da zu diesem Zeitpunkt häufig Entscheidungen über die Art der zu verwendenden Daten getroffen werden.

Vergewissern Sie sich also, ob Sie wirklich große Datenmengen benötigen. Zielgerichtete, „intelligente Daten“ könnten viel nützlicher sein als Big Data. Natürlich ist die Verwendung intelligenter, gut ausgearbeiteter Daten unter Umständen mit einem enormen Aufwand für die Vereinheitlichung, Homogenisierung usw. verbunden, aber es wird die Umsetzung des Grundsatzes der Datenminimierung wesentlich effizienter gestalten. Zu diesem Zweck ist es wichtig, **über Fachwissen zu verfügen, um die relevanten Merkmale auszuwählen.** Dieser Schritt beinhaltet auch die Überprüfung der Erforderlichkeit der Verarbeitung für jede Datenkategorie. Es muss somit nachgewiesen werden, dass keine alternativen Maßnahmen oder Methoden, die aus Sicht des Datenschutzes und der Menschenrechte weniger verletzend sind, angewandt werden könnten, um das gleiche Ergebnis zu erzielen.

Darüber hinaus sollten Sie versuchen, die **Auflösung der Daten auf das für die mit der Verarbeitung verfolgten Zwecke erforderliche Mindestmaß zu beschränken**. Sie sollten zudem **ein optimales Niveau der Datenaggregation bestimmen,** bevor Sie mit der

[633] https://ec.europa.eu/info/sites/info/files/5._h2020_ethics_and_data_protection_0.pdf
[634] Artikel 25 Absatz 2.

Verarbeitung beginnen (siehe „Angemessen, erheblich und beschränkt“ in Abschnitt „Datenminimierung“, Kapitel „Grundsätze“). Im Falle von KI, die zur Verhütung, Ermittlung oder Aufdeckung von Straftaten eingesetzt wird, ist der mögliche Grad der Datenaggregation, d. h. der Anonymisierung von Daten, zweifellos begrenzt, zumindest bei späteren Implementierungen und Verwendungen der entwickelten Systeme. Da ein Hauptziel in der Identifizierung (potenzieller) Täter besteht, muss es zumindest möglich sein, Daten über potenzielle Bedrohungen zu (re-)personalisieren.

Die Datenminimierung kann im Falle des Deep Learning kompliziert sein, da eine Unterscheidung nach Merkmalen unter Umständen nicht möglich ist. Es gibt jedoch eine effiziente Möglichkeit, die Menge der erhobenen Daten zu regulieren und sie nur dann zu erhöhen, wenn es notwendig erscheint: die Lernkurve. Sie sollten damit beginnen, eine begrenzte Menge an Trainingsdaten zu erheben und zu verwenden, und dann die Genauigkeit des Modells überwachen, während es mit neuen Daten gefüttert wird.

2.2.2 Überprüfung der rechtmäßigen Verwendung von Datensätzen

Datensätze können auf unterschiedliche Weise beschafft werden. Als erste Option kann sich der Entwickler dafür entscheiden, eine bereits von jemand anderen erstellte Datenbank oder den Zugriff darauf zu erwerben. In diesem Fall sollten Sie besonders vorsichtig sein, da der Erwerb des Zugangs zu einer Datenbank mit zahlreichen rechtlichen Fragen verbunden ist (siehe „Erwerb des Zugangs zu einer Datenbank”, Kapitel „Maßnahmen und Tools”).[635]

Eine zweite und sehr gängige Alternative darin, eine Datenbank zu erstellen. Natürlich müssen Sie in diesem Fall sicherstellen, dass Sie bei der Erstellung dieser Datenbank alle rechtlichen Anforderungen der DSGVO erfüllen (siehe „Erstellen einer Datenbank“, Kapitel „Maßnahmen und Tools“).

Drittens können Sie auch einen alternativen Weg wählen. Sie können lizenzierte Daten von Dritten miteinander oder mit Ihrem eigenen Datensatz mischen, um einen großen Trainingsdatensatz und einen weiteren für Validierungszwecke zu erstellen. Dies könnte einige Probleme mit sich bringen, darunter die Möglichkeit, dass die Kombination dieser unterschiedlichen Datensätze zusätzliche Informationen über die betroffenen Personen liefert. Unter Umständen könnten Sie dadurch die betroffenen Personen identifizieren, was bei Verwendung nur eines der Datensätze bisher nicht möglich war. Dies könnte bedeuten, dass anonymisierte Daten deanonymisiert und neue personenbezogene Informationen geschaffen werden, die im ursprünglichen Datensatz nicht enthalten waren. Diese Situation würde erhebliche ethische und rechtliche Fragen aufwerfen. Wenn beispielsweise *„die betroffenen Personen in Kenntnis der Sachlage in die Verarbeitung personenbezogener Daten in den ursprünglichen Datensätzen für bestimmte Zwecke eingewilligt haben, haben sie damit nicht notwendigerweise auch in die Zusammenführung von Datensätzen und das Data Mining eingewilligt, das neue Informationen offenbart. Neue Informationen, die auf diese Weise gewonnen werden, können auch auf Wahrscheinlichkeiten oder Vermutungen beruhen und daher falsch sein oder Verzerrungen bei der Darstellung von Personen enthalten.“* [636] Daher sollten Sie solche Folgen nach Möglichkeit vermeiden, indem Sie sicherstellen, dass die Zusammenführung von Datensätzen nicht gegen die Rechte und Interessen der betroffenen Personen verstößt.

---

[635] Yeong Zee Kin, Legal Issues in AI Deployment, verfügbar unter: https://lawgazette.com.sg/feature/legal-issues-in-ai-deployment/, abgerufen am 15. Mai 2020

[636] SHERPA, Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design approach, 2020, Seite 38. Verfügbar unter: https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf, abgerufen am 15. Mai 2020.

Wenn Sie mehrere Datensätze verwenden, die unterschiedlichen Zwecken dienen, sollten Sie geeignete Maßnahmen ergreifen, um die verschiedenen Verarbeitungsvorgänge zu trennen. Andernfalls könnten Sie leicht Daten für einen Zweck verwenden, für den sie nicht erhoben wurden, was zu Problemen im Zusammenhang mit dem Grundsatz der Zweckbindung führen könnte.

Beachten Sie, dass die oben genannten Maßnahmen nur für die Durchführungsphase des Forschungsprojekts ausreichend sind. Eine Einwilligung in Kenntnis der Sachlage ist im Rahmen von Strafverfolgungsmaßnahmen in der Regel von sehr begrenztem Nutzen. Das Gleiche gilt für die Erstellung und Verwendung von Dummy- oder synthetischen Daten. Auch die Verwendung synthetischer Daten kann trotzdem Probleme in Bezug auf die potenzielle Re-Identifikation sowie die Frage aufwerfen, ob man solchen Daten beim Training von KI-Algorithmen vertrauen kann. All diese Maßnahmen können wirksam dazu beitragen, ethische oder rechtliche Probleme in der Forschungsphase zu entschärfen oder zu beseitigen. Es muss unbedingt sichergestellt werden, dass die für reale Anwendungen benötigten Datensätze auch den ethischen und rechtlichen Anforderungen der EU und der Mitgliedstaaten entsprechen; dies gilt auch für die Verwendung von Datensätzen, die sich im Besitz der Polizei oder von Behörden befinden. Seien Sie sich auch bewusst, dass es schwierig oder sogar unmöglich sein kann, Zugang zu ausreichend großen realen Datensätzen zu erhalten, die für das praktische Training des KI-Tools erforderlich sind.

2.2.3 Auswahl der geeigneten Rechtsgrundlage für die Verarbeitung

Sie müssen entscheiden, welche Rechtsgrundlage Sie für die Verarbeitung verwenden, bevor Sie mit der Verarbeitung beginnen, Ihre Entscheidung (zusammen mit den Zwecken) dokumentieren und die Gründe für Ihre Wahl angeben (siehe Abschnitt „Rechenschaftspflicht“, Kapitel „Grundsätze“).

Sie sollten die Rechtsgrundlage wählen, die der wahren Natur Ihrer Verarbeitung personenbezogener Daten am ehesten entspricht. Wenn es sich um menschliche Teilnehmer handelt, müssen auch die Beziehung zu den Teilnehmern und der Zweck der Verarbeitung berücksichtigt werden. Diese Entscheidung ist von grundlegender Bedeutung, da die Rechtsgrundlage für die Verarbeitung nur dann geändert werden darf, wenn es stichhaltige Gründe gibt, die dies rechtfertigen (siehe Abschnitt „Zweckbindung“, Kapitel „Grundsätze“).

Bei KI-Tools, die zum Zweck der Verhütung, Ermittlung, Aufdeckung usw. von Straftaten entwickelt werden, müssen Sie wiederum zwischen der Forschungsphase und späteren Implementierungen unterscheiden. In der Forschungsphase können Sie möglicherweise die Einwilligung als Rechtsgrundlage für die Verarbeitung heranziehen (siehe Abschnitt „Einwilligung“ im Kapitel „Hauptkonzepte“), je nachdem, ob menschliche Teilnehmer konkret beteiligt sind. Beispiele hierfür könnten KI-Tools sein, die eine biometrische Identifizierung oder die Interpretation von Videodaten verwenden und die Einbeziehung menschlicher Teilnehmer für Tests erfordern. Die Einwilligung könnte auch eine gültige Rechtsgrundlage sein, wenn Sie Daten wiederverwenden, die bereits für einen anderen Zweck erhoben wurden, und die Einwilligung die Grundlage war, die die primäre Verwendung der Daten erlaubte. Die DSGVO erlaubt die Wiederverwendung von Daten für wissenschaftliche Zwecke. In Artikel 5 Absatz 1 Buchstabe b heißt es, dass die Weiterverarbeitung für wissenschaftliche Forschungszwecke nicht als unvereinbar mit den ursprünglichen Zwecken („Zweckbindung“) gilt. Im Prinzip könnten Sie also diese Daten auf der Grundlage der ursprünglichen Einwilligung weiterverwenden. Sie müssen jedoch bedenken, dass gemäß Artikel 9 Absatz 4 DSGVO *„die Mitgliedstaaten zusätzliche Bedingungen, einschließlich Beschränkungen, einführen oder aufrechterhalten können, soweit die Verarbeitung von genetischen, biometrischen oder Gesundheitsdaten betroffen ist.“* Es kann also durchaus sein, dass Ihre jeweilige nationale Regelung Ausnahmen oder besondere Bedingungen für die

Weiterverwendung personenbezogener Daten vorsieht. In jedem Fall sollten Sie immer daran denken, dass Ihre Informationspflichten bestehen bleiben. Sie sollten der betroffenen Person vor jeder weiteren Verarbeitung ihrer Daten Informationen über diesen anderen Zweck und alle weiteren relevanten Informationen gemäß Artikel 13 Absatz 2 DSGVO zur Verfügung stellen.

Bitte bedenken Sie, dass die oben genannten Bestimmungen nur für die Durchführung der Forschung als solche gelten. Künftige Verwendungen der entwickelten Systeme müssen den geltenden Rechtsvorschriften der EU und der Mitgliedstaaten über Strafverfolgungsmaßnahmen entsprechen. Seien Sie sich auch bewusst, dass die Entwicklung von Technologien, die nicht mit den geltenden Vorschriften, den ethischen Grundsätzen oder den europäischen Werten übereinstimmen, eine Verschwendung von Aufwand und Ressourcen bedeuten würde.

2.2.4 Wiederverwendung von Daten

Gegenwärtig wird eine lebhafte Diskussion über die Wiederverwendung von Daten für Forschungszwecke geführt. Laut Artikel 5 Absatz 1 Buchstabe b DSGVO gilt die Weiterverarbeitung für wissenschaftliche Forschungszwecke nicht als unvereinbar mit dem ursprünglichen Zweck. Sofern Ihre nationale Regelung also nichts anderes vorsieht, können Sie die verfügbaren Daten für Forschungszwecke weiterverwenden, da diese mit dem ursprünglichen Zweck, für den sie erhoben wurden, vereinbar sind.

Der EDSB führte jedoch an: „*Um die Achtung der Rechte der betroffenen Person zu gewährleisten, sollte die Vereinbarkeitsprüfung gemäß Artikel 6 Absatz 4 vor der Weiterverwendung von Daten für wissenschaftliche Forschungszwecke weiterhin in Betracht gezogen werden, insbesondere wenn die Daten ursprünglich für ganz andere Zwecke oder außerhalb des Bereichs der wissenschaftlichen Forschung erhoben wurden. Laut einer Analyse aus Sicht der medizinischen Forschung sollte die Anwendung dieses Tests in der Tat einfach sein*“.[637] Nach dieser Auslegung sollten Sie personenbezogene Daten nur dann weiterverwenden, wenn die Bedingungen von Artikel 6.4 zutreffen. Bitte prüfen Sie in diesem Zusammenhang auch die Anwendbarkeit von Artikel 10 „*Die Verarbeitung personenbezogener Daten über strafrechtliche Verurteilungen und Straftaten oder damit zusammenhängende Sicherungsmaßregeln aufgrund von Artikel 6 Absatz 1 darf nur unter behördlicher Aufsicht vorgenommen werden oder wenn dies nach dem Unionsrecht oder dem Recht der Mitgliedstaaten, das geeignete Garantien für die Rechte und Freiheiten der betroffenen Personen vorsieht, zulässig ist.*“

3 Datenaufbereitung

3.1 Beschreibung

„*Die Phase der Datenaufbereitung umfasst alle Aktivitäten zur Erstellung des endgültigen Datensatzes bzw. der Daten, die aus den anfänglichen Rohdaten in das/die Modellierungswerkzeug(e) eingespeist werden. Zu den Aufgaben gehören die Auswahl von Tabellen, Datensätzen und Attributen sowie die Transformation und Bereinigung der Daten für die Modellierungswerkzeuge. Die fünf Schritte der Datenaufbereitung umfassen die Auswahl der Daten, die Bereinigung der Daten, den Aufbau der Daten, die Integration der Daten und die Formatierung der Daten.*“[638]

[637] EDSB, A preliminary opinion on data protection and scientific research, 6. Januar 2020, Seite 23.
[638] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 16.

Diese Phase umfasst alle erforderlichen Aktivitäten, um aus den anfänglichen Rohdaten den endgültigen Datensatz zu erstellen, der dann in das Modell eingespeist wird. Sie umfasst die folgenden fünf Aufgaben, die nicht unbedingt nacheinander ausgeführt werden müssen:

1. Auswahl der Daten: Festlegung der für die Analyse zu verwendenden Daten, auf der Grundlage der Relevanz für die Data-Mining-Ziele, der Qualität sowie der technischen Beschränkungen, wie z. B. Einschränkungen des Datenvolumens oder der Datentypen.
2. Bereinigung der Daten: Anhebung der Datenqualität auf das erforderliche Niveau, beispielsweise durch Auswahl sauberer Teilmengen der Daten, Einfügen von Standardwerten und Schätzung fehlender Daten durch Modellierung.
3. Aufbau der Daten: Der Aufbau neuer Daten durch die Erstellung abgeleiteter Attribute, neuer Datensätze oder transformierter Werte für vorhandene Attribute.
4. Integration der Daten: Kombinieren von Daten aus mehreren Tabellen oder Datensätzen, um neue Datensätze oder Werte zu erstellen.
5. Formatierung der Daten: Syntaktische Änderungen an Daten, die vom Modellierungswerkzeug benötigt werden.

3.2 Wichtigste zu ergreifende Maßnahmen

3.2.1 Einführung der in Artikel 89 DSGVO vorgesehenen Garantien

Da Sie die Daten für wissenschaftliche Zwecke verwenden, müssen Sie sie gemäß den in Artikel 89 DSGVO vorgesehenen Garantien aufbereiten. Wenn die Zwecke Ihrer Forschung durch die Weiterverarbeitung, bei der die Identifizierung von betroffenen Personen nicht oder nicht mehr möglich ist, erfüllt werden können, d. h. durch Pseudonymisierung, sollten diese Zwecke auf diese Weise erfüllt werden. Wenn dies nicht möglich ist, müssen Sie Garantien einführen, die sicherstellen, dass die technischen und organisatorischen Maßnahmen eine angemessene Umsetzung des Grundsatzes der Datenminimierung ermöglichen. Bitte beachten Sie die konkreten Vorschriften, die in Ihrer nationalen Regelung für Garantien festgelegt sind. Konsultieren Sie Ihren Datenschutzbeauftragten.

3.2.2 Gewährleistung der sachlichen Richtigkeit in Bezug auf die Verarbeitung personenbezogener Daten

Laut der DSGVO müssen die Daten sachlich richtig sein (siehe „Richtigkeit“, Kapitel „Grundsätze“). Dies bedeutet, dass die verarbeiteten Daten richtig und auf dem neuesten Stand sind. Die Verantwortlichen müssen die Richtigkeit gewährleisten. Sobald Sie die Datenerhebung abgeschlossen haben, sollten Sie daher geeignete Instrumente einsetzen, um die Richtigkeit der Daten zu gewährleisten. Dies bedeutet in der Regel, dass Sie einige grundlegende Entscheidungen über die technischen und organisatorischen Maßnahmen treffen müssen, damit dieser Grundsatz erfüllt wird (siehe Unterabschnitt „Verbundene technische und organisatorische Maßnahmen“, Abschnitt „Richtigkeit“, Kapitel „Grundsätze“). Da die meisten Daten wahrscheinlich aus ganz unterschiedlichen Quellen ohne standardisierte Qualitätsanforderungen stammen und zudem im Falle der Vorhersage von Straftagen wahrscheinlich qualitativ sind, kann man nicht davon ausgehen, dass sie per se sachlich richtig sind. Dies liegt vor allem daran, dass diese Daten auf individuellen Einschätzungen verschiedener Personen beruhen können, während die betroffenen Personen möglicherweise nicht einmal wissen, dass diese Art von Daten über sie gespeichert wird.

In jedem Fall erfordert die Richtigkeit der Daten eine angemessene Umsetzung von Maßnahmen, damit betroffene Personen ihr Recht auf Berichtigung in Anspruch nehmen können (siehe Abschnitt „Recht auf Berichtigung“, Kapitel „Rechte der betroffenen Person“).

Es ist auch sicherzustellen, dass die Ergebnisse möglichst sachlich richtig sind. Die Arten von falschen positiven und falschen negativen Ergebnissen sollten in der Phase der Datenaufbereitung im Voraus festgelegt werden. Falsche Ergebnisse sind einer der wesentlichen Punkte, die sich auf die Grundrechte der Betroffenen auswirken.

3.2.3 Konzentration auf Profiling-Probleme

Im Falle einer Datenbank, die ein KI-Tool trainieren oder validieren soll, besteht im Allgmeinen eine besonders wichtige Verpflichtung, die betroffenen Personen darüber zu informieren, **dass sie betreffende Daten zu einer automatisierten Entscheidungsfindung oder einem Profiling führen könnten.** Profiling ist in der KI-Entwicklung besonders problematisch, dies gilt auch für KI-Tools, die für die Zwecke von Strafverfolgungsbehörden entwickelt werden.

Gemäß Artikel 22 Absatz 2 Buchstabe c sind automatisierte Entscheidungen, die besondere Kategorien personenbezogener Daten betreffen, wie Daten, aus denen die ethnische Herkunft, politische Meinung, religiöse oder philosophische Überzeugung bzw. die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder zur sexuellen Orientierung einer natürlichen Person (Artikel 9 Absatz 1) nur zulässig, wenn die betroffene Person eingewilligt hat oder wenn sie auf einer gesetzlichen Grundlage erfolgen. Diese Ausnahme gilt nicht nur, wenn die entsprechenden Daten in diese Kategorie fallen, sondern auch dann, **wenn der Abgleich verschiedener Arten personenbezogener Daten sensible Informationen über Personen offenbaren kann oder wenn abgeleitete Daten in diese Kategorie fallen.** Im Falle der Verhütung, Ermittlung und Aufdeckung von Straftaten gilt die ausdrückliche Einwilligung der betroffenen Personen in der Regel nur für freiwillige menschliche Teilnehmer in der FuE-Phase. Die Verarbeitung besonderer Kategorien personenbezogener Daten, z. B. in Bezug auf politische Meinungen oder religiöse Überzeugungen, kann zum Kerndatenbestand von KI-Tools gehören, die im Bereich der Terrorismusprävention eingesetzt werden.

Folgende zusätzliche Maßnahmen können ebenfalls äußerst nützlich sein, um eine nicht erforderliche, automatisierte Entscheidungsfindung zu vermeiden:

- Berücksichtigen Sie bereits **in der Designphase die Systemanforderungen, die zur Unterstützung einer aussagekräftigen Überprüfung durch Personen erforderlich sind.** Insbesondere die Anforderungen an die Interpretierbarkeit und die effektive Gestaltung der Benutzeroberfläche zur Unterstützung von Überprüfungen und Eingriffen durch Personen;
- Gestalten und setzen Sie eine geeignete Schulung und Unterstützung für menschliche Prüfer um.
- Geben Sie dem Personal die entsprechenden Befugnisse, Anreize und Unterstützung, damit es die Bedenken einzelner Personen ansprechen oder

eskalieren und gegebenenfalls die Entscheidung des KI-Tools außer Kraft setzen kann.[639]

Wenn Sie ein Profiling oder eine automatisierte Entscheidungsfindung durchführen, müssen Sie die betroffenen Personen über Ihre Entscheidung informieren und alle erforderlichen Informationen gemäß der DSGVO und ggf. nationalen Vorschriften bereitstellen.

3.2.4 Auswahl unverzerrter Daten

Verzerrungen sind eines der Hauptprobleme bei der Entwicklung von KI, ein Problem, das dem Grundsatz der Fairness zuwiderläuft. Verzerrungen können durch viele verschiedene Faktoren verursacht werden. Bei der Erfassung der Daten können sozial bedingte Verzerrungen, Ungenauigkeiten, Fehler und andere Mängel auftreten. Manchmal kann es auch vorkommen, dass Datensätze durch böswillige Handlungen verzerrt werden. Die Eingabe schädlicher Daten in ein KI-Tool kann sein Verhalten verändern, insbesondere bei selbstlernenden Systemen.[640] Daher werfen Fragen im Zusammenhang mit der Zusammensetzung der für das Training verwendeten Datenbanken entscheidende ethische und rechtliche Fragen auf, nicht nur Fragen der Effizienz oder technischer Natur.

Sie müssen diese Fragen somit vor dem Training des Algorithmus klären. Erkennbare und diskriminierende Verzerrungen sollten nach Möglichkeit bereits in der Phase der Datensatzerstellung eliminiert werden. Wie wir in der Vergangenheit gesehen haben, ist die Vorstellung, dass bestimmte Personengruppen (Schwarze, Araber oder Ausländer im Allgemeinen, Muslime …) häufiger verurteilt werden, weil sie häufiger gegen das Gesetz verstoßen, in den meisten Fällen nicht zutreffend. Sie werden häufiger durchsucht, häufiger von der Polizei diskriminiert, sind häufiger mit exzessiver Gewalt, Willkür oder Anfeindungen durch die Polizei konfrontiert und kommen daher häufiger in problematische Situationen. Diese Beobachtung würde bei Gleichbehandlung höchstwahrscheinlich auch für jede andere Bevölkerungsgruppe gelten. Daraus eine höhere Kriminalitätsrate in Gebieten abzuleiten, in denen viele Ausländer leben, könnte zu einer sich selbst erfüllenden Prophezeiung werden.

Ein weiteres Beispiel wäre die Annahme, dass ein KI-Tool die richtigen Ergebnisse liefert, wenn diese mit den Ergebnissen von Menschen übereinstimmen. Auch menschliche Entscheidungen sind oft voreingenommen, und das KI-Tool würde höchstwahrscheinlich solche diskriminierenden Praktiken fortsetzen, anstatt objektivere Ergebnisse zu liefern.

Wenn der Algorithmus voreingenommen ist, kann er auch die Zahl der falschen positiven oder falschen negativen Ergebnisse erhöhen. Falsche positive Ergebnisse können schwerwiegende negative Auswirkungen auf die betroffenen Personen haben, falsche negative Ergebnisse dagegen auf die Gesellschaft und natürlich auch auf die Opfer krimineller oder terroristischer Aktivitäten, die möglicherweise hätten vermieden werden können.

Sie müssen sicherstellen, dass der Algorithmus diese Faktoren bei der Auswahl der Daten entsprechend bewertet. **Die mit der Auswahl der in die Datensätze zu integrierenden Daten betrauten Teams sollten somit aus Personen bestehen, die die Diversität gewährleisten, die von KI-Tools erwartet wird.** Wenn sich Ihre Daten hauptsächlich auf eine bestimmte Gruppe beziehen, dürfen Sie schließlich niemals die Erklärung vergessen,

---

[639] https://ico.org.uk/about-the-ico/news-and-events/ai-blog-fully-automated-decision-making-ai-systems-the-right-to-human-intervention-and-other-safeguards/

[640] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI, 2019, Seite 17. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020)

dass der Algorithmus auf dieser Grundlage trainiert wurde und daher bei anderen Bevölkerungsgruppen möglicherweise nicht so gut funktioniert.

4 Modellierung (Training)

4.1 Beschreibung

„In dieser Phase werden verschiedene Modellierungstechniken ausgewählt und angewandt sowie deren Parameter auf optimale Werte kalibriert. In der Regel gibt es für ein und dieselbe Art von Data-Mining-Problem mehrere Techniken. Einige Techniken haben spezifische Anforderungen an die Form der Daten. Daher kann es notwendig sein, zur Phase der Datenaufbereitung zurückzukehren. Die Modellierungsschritte umfassen die Auswahl der Modellierungstechnik, die Erstellung des Testdesigns, die Erstellung der Modelle und die Bewertung der Modelle.“[641]

Diese Phase umfasst mehrere wichtige Aufgaben. Alles in allem müssen Sie:

- Die zu verwendende Modellierungstechnik auswählen. Je nach Art der Technik sind Folgen wie Herleitung, Unklarheit oder Verzerrungen von Daten mehr oder weniger wahrscheinlich.
- Das zu verwendende Trainingstool festlegen. Mit diesem Tool kann der Entwickler messen, wie gut das Modell die Vergangenheit vorhersagen kann, bevor er es für die Vorhersage der Zukunft einsetzt. Im Falle der Vorhersage von Straftaten könnte dies selbst ein Problem darstellen. Dies ist anders als die Vorhersage, dass jemand mit einer Vorliebe für Joghurt diesen wieder kaufen wird. Wir sprechen hier von Menschen und ihren Chancen im Leben. Die Annahme, dass jemand wieder straffällig wird, weil er/sie in der Vergangenheit gegen das Gesetz verstoßen hat, vernachlässigt im Prinzip die Tatsache, dass wir Bürger als Menschen mit einem freien Willen und der Chance betrachten, beim nächsten Mal eine bessere Entscheidung zu treffen. Es ist von Natur aus problematisch, davon auszugehen, dass die Zukunft eine Extrapolation der Vergangenheit sein wird. Je nach den individuellen und gesellschaftlichen Folgen ist dies in einigen Fällen weniger problematisch und in anderen Fällen unvertretbar. Zum Training gehören immer empirische Tests mit Daten. Manchmal testen die Entwickler das Modell mit Daten, die sich von den zur Erstellung des Modells verwendeten Daten unterscheiden. Daher kann man in dieser Phase von verschiedenen Arten von Datensätzen sprechen.

4.2 Wichtigste zu ergreifende Maßnahmen

4.2.1 Umsetzung des Grundsatzes der Datenminimierung

Nach dem Grundsatz der Datenminimierung müssen Sie die Menge der bereitgestellten Daten und/oder den Umfang der Informationen über die betroffene Person so schnell wie möglich auf das notwendige Maß beschränken. Folglich müssen Sie die in der Trainingsphase verwendeten Daten von allen Informationen bereinigen, die für das Training des Modells

[641] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 17.

nicht unbedingt erforderlich sind. (Siehe Unterabschnitt „Zeitlicher Aspekt“, Abschnitt „Datenminimierung“, Kapitel „Grundsätze“). Es gibt mehrere Strategien, um die Datenminimierung in der Trainingsphase zu gewährleisten. Die Techniken entwickeln sich ständig weiter. Zu den gängigsten zählen[642] (siehe auch Abschnitt „Integrität und Vertraulichkeit” im Kapitel „Grundsätze”):

- Analyse der von den Daten zu erfüllenden Bedingungen, um als qualitativ hochwertig und mit einer hohen Vorhersagekraft für die spezifische Anwendung zu gelten.
- Kritische Analyse des Umfangs der in jeder Phase der KI-Lösung verwendeten Datentypologie.
- Löschung von unstrukturierten Daten und unnötigen Informationen, die bei der Vorverarbeitung der Informationen gesammelt wurden.
- Identifizierung und Unterdrückung der Datenkategorien, die keinen signifikanten Einfluss auf das Lernen oder das Ergebnis der Herleitung haben.
- Unterdrückung irrelevanter Schlussfolgerungen im Zusammenhang mit personenbezogenen Informationen während des Trainingsprozesses, beispielsweise im Falle des unüberwachten Trainings.
- Verwendung von Verifizierungstechniken, die weniger Daten erfordern, wie z. B. die Kreuzvalidierung.
- Analyse und Konfiguration von algorithmischen Hyperparametern, die die Menge oder den Umfang der verarbeiteten Daten beeinflussen könnten, um diese zu minimieren.
- Verwendung von föderierten statt zentralisierten Lernmodellen.
- Anwendung von differenziellen Datenschutzstrategien.
- Training mit verschlüsselten Daten unter Verwendung homomorpher Techniken.
- Datenaggregation.
- Anonymisierung und Pseudonymisierung, nicht nur bei der Datenkommunikation, sondern auch in Bezug auf Trainingsdaten, möglichen personenbezogenen Daten im Modell und bei der Verarbeitung der Herleitung.

4.2.2 Erkennen und Eliminieren von Verzerrungen

Selbst wenn die Mechanismen zur Vermeidung von Verzerrungen bereits in den vorangegangenen Phasen angewandt wurden (siehe den obigen Abschnitt über das Training), muss dennoch sichergestellt werden, dass die Ergebnisse der Trainingsphase die Verzerrungen auf ein Mindestmaß senken. Dies kann schwierig sein, da einige Arten von Verzerrungen und Diskriminierung oft besonders schwer zu erkennen sind. Die Teammitglieder, die die Eingabedaten kuratieren, sind sich dessen manchmal nicht bewusst, ebenso wenig wie die jeweiligen Nutzer. Daher sind die vom KI-Entwickler in der Validierungsphase eingesetzten Überwachungssysteme äußerst wichtige Faktoren, um Verzerrungen zu vermeiden.

Es gibt zahlreiche technische Hilfsmittel, die sich hervorragend zum Erkennen von Verzerrungen eignen, darunter die algorithmische Folgenabschätzung.[643] Sie sollten daher

[642] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, Seite 40. Verfügbar unter: https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf, abgerufen am 15. Mai 2020.

[643] D. Reisman, K. Crawford und M. Whittaker, Algorithmic impact assessments: A practical framework for public agency accountability, 2018, verfügbar unter: https://ainowinstitute.org/aiareport2018.pdf, abgerufen am 15. Mai 2020

deren wirksame Umsetzung berücksichtigen.[644] Wie die Literatur zeigt[645], kann es jedoch vorkommen, dass sich ein Algorithmus nicht vollständig von allen verschiedenen Arten von Verzerrungen befreien lässt. Sie sollten jedoch zumindest versuchen, sich ihrer Existenz und der daraus resultierenden Folgen bewusst zu sein (siehe Abschnitte Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz und Richtigkeit, Kapitel Grundsätze).

4.2.3 Ausübung der Rechte der betroffenen Person

Manchmal vervollständigen die Entwickler die verfügbaren Daten durch Herleitungen. Wenn Sie beispielsweise nicht über die faktischen Daten verfügen, die den politischen Ansichten eines Straftäters entsprechen, können Sie einen anderen Algorithmus verwenden, um diese aus den übrigen Daten herzuleiten, z. B. aus der beobachteten Teilnahme an Demonstrationen. Dies bedeutet jedoch keinesfalls, dass diese Daten als pseudonymisiert oder anonymisiert betrachtet werden können. Es handelt sich somit weiterhin um personenbezogene Daten. Dementsprechend müssen auch hergeleitete Daten als personenbezogene Daten betrachtet werden. Daher haben die betroffenen Personen einige grundlegende Rechte in Bezug auf diese Daten, die Sie respektieren müssen.

Tatsächlich müssen Sie alle Rechte der betroffenen Personen während des gesamten Lebenszyklus respektieren. In dieser speziellen Phase sind das Auskunftsrecht, das Recht auf Berichtigung und das Recht auf Löschung besonders sensibel. Sie weisen bestimmte Merkmale auf, derer sich die Verantwortlichen bewusst sein müssen. Im Falle einer Forschung zu wissenschaftlichen Zwecken wie die von Ihnen entwickelte sieht die DSGVO jedoch einige Garantien und Ausnahmen in Bezug auf die Verarbeitung vor (Art. 89). Sie müssen sich über die konkrete Regelung in Ihrem Mitgliedstaat im Klaren sein. Laut der DSGVO können im Unionsrecht oder im Recht der Mitgliedstaaten insoweit Ausnahmen von den Rechten gemäß der Artikel 15, 16, 18 und 21 vorgesehen werden, als diese Rechte voraussichtlich die Verwirklichung der spezifischen Zwecke unmöglich machen oder ernsthaft beeinträchtigen und solche Ausnahmen für die Erfüllung dieser Zwecke notwendig sind.

- Auskunftsrecht (siehe „Auskunftsrecht“, Kapitel „Rechte der betroffenen Person“)

Grundsätzlich sollten Sie auf Anträge betroffener Personen in Bezug auf Auskunft über sie betreffende personenbezogenen Daten reagieren, vorausgesetzt, Sie haben angemessene Maßnahmen ergriffen, um die Identität der betroffenen Person zu überprüfen, und es gelten keine weiteren Ausnahmen. Allerdings sind Sie nicht verpflichtet, zusätzliche personenbezogene Daten zur Identifizierung der betroffenen Personen in den Trainingsdaten zu erheben oder zu speichern, nur um die Verordnung einzuhalten. Wenn Sie eine betroffene Person in den Trainingsdaten nicht identifizieren können und die betroffene Person keine zusätzlichen Informationen bereitstellen kann, die ihre Identifizierung ermöglichen würden, sind sie nicht verpflichtet, einem Antrag nachzukommen, der nicht erfüllt werden kann.

- Recht auf Berichtigung (siehe Abschnitt „Recht auf Berichtigung“, Kapitel „Data subject rights“)

[644] https://ico.org.uk/media/about-the-ico/consultations/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf, abgerufen am 15. Mai 2020

[645] Alexandra Chouldechova, Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments, Big Data. Band: 5, Ausgabe 2: 1. Juni 2017. 153–163. http://doi.org/10.1089/big.2016.0047

Im Falle des Rechts auf Berichtigung müssen Sie das Recht auf Berichtigung der Daten gewährleisten, insbesondere der Daten, die durch die vom KI-Tool erstellten Herleitungen und Profile erzeugt wurden. Auch wenn der Zweck von Trainingsdaten darin besteht, Modelle zu trainieren, die auf allgemeinen Mustern in großen Datensätzen beruhen, und einzelne Ungenauigkeiten somit kaum direkte Auswirkungen auf eine betroffene Person haben werden, darf das Recht auf Berichtigung nicht eingeschränkt werden. Sie könnten höchstens eine längere Frist (zwei zusätzliche Monate) für die Berichtigung beantragen, wenn das technische Verfahren besonders komplex ist (Artikel 11 Absatz 3).

- Recht auf Löschung (siehe „Recht auf Löschung", Kapitel „Rechte der betroffenen Person")

Die betroffenen Personen haben ein Recht auf Löschung sie betreffender personenbezogener Daten. Dieses Recht kann jedoch eingeschränkt sein, wenn bestimmte konkrete Umstände vorliegen. Laut dem britischen ICO „können Unternehmen Anträge auf Löschung von Trainingsdaten erhalten. Unternehmens müssen auf Löschungsanträge reagieren, es sei denn, es gilt eine einschlägige Ausnahmeregelung, und sofern die betroffene Person angemessene Gründe vorbringen kann. Wenn beispielsweise die Trainingsdaten nicht mehr benötigt werden, weil das ML-Modell bereits trainiert wurde, muss das Unternehmen dem Antrag nachkommen. Falls die Entwicklung des Systems jedoch noch nicht abgeschlossen ist, kann es jedoch notwendig sein, die Trainingsdaten für die Zwecke des erneuten Trainings, der Verfeinerung und der Bewertung eines KI-Tools zu speichern. In diesem Fall sollte das Unternehmen von Fall zu Fall entscheiden, ob es dem Antrag nachkommen kann. Einem Antrag auf Löschung von Trainingsdaten nachzukommen, bedeutet jedoch nicht, dass auf diesen Daten beruhenden ML-Modelle gelöscht werden müssen, es sei denn, die Modelle selbst enthalten diese Daten oder können zu deren Herleitung verwendet werden."[646]

5 Bewertung (Validierung)

5.1 Beschreibung

„Vor der endgültigen Einführung des vom Datenanalysten erstellten Modells sollte das Modell unbedingt gründlicher bewertet und die Konstruktion des Modells überprüft werden, um sicher zu sein, dass es die Geschäftsziele ordnungsgemäß erreicht. In dieser Phase ist die Prüfung von entscheidender Bedeutung, ob ein wichtiger Geschäftsaspekt eventuell nicht ausreichend berücksichtigt wurde. Am Ende dieser Phase sollte der Projektleiter dann entscheiden, wie die Data-Mining-Ergebnisse zu verwenden sind. Die wichtigsten Schritte sind hier die Bewertung der Ergebnisse, die Überprüfung des Prozesses und die Festlegung der nächsten Schritte."[647]

[646] ICO, Enabling access, erasure, and rectification rights in AI tools, verfügbar unter: https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/, abgerufen am 15. Mai 2020

[647] Colin Shearer, The CRISP-DM Model: The New Blueprint for Data Mining, Seite 17

Diese Phase umfasst mehrere Aufgaben, die wichtige Datenschutzfragen aufwerfen. Alles in allem müssen Sie:

- Die Ergebnisse Ihres Modells bewerten, zum Beispiel, ob es genau ist oder nicht. Zu diesem Zweck könnte der KI-Entwickler sie in der realen Welt testen. Dieser Test kann oft in Koordination mit einem projektbezogenen Partner aus dem Bereich, in dem das System eingeführt werden soll, durchgeführt werden (beispielsweise Strafverfolgungsbehörden).
- Den Prozess überprüfen. Sie sollten das Datenverarbeitungssystem überprüfen, um festzustellen, ob ein kritischer Faktor oder eine wichtige Aufgabe übersehen wurde. Dazu gehören auch Fragen der Qualitätssicherung. Dies ist tatsächlich die allerletzte Phase, um potenzielle Endnutzer in den Entwicklungsprozess einzubeziehen. Allerdings sollten Sie die Bedürfnisse der Endnutzer bereits in einer sehr frühen Phase Ihres Projekts einbeziehen und ermitteln (Geschäftsverständnis). In dieser Phase können die Stakeholder und Endnutzer Einblicke in die Stärken und Schwächen des Systems in der Praxis geben.

5.2 Wichtigste zu ergreifende Maßnahmen

5.2.1 Prozesse der dynamischen Validierung

Die Validierung der eine KI-Komponente umfassenden Verarbeitung muss unter Bedingungen erfolgen, die die reelle Umgebung widerspiegeln, in der die Verarbeitung stattfinden soll. Wenn Sie also im Voraus wissen, wo das KI-Tool eingesetzt werden soll, sollten Sie den Validierungsprozess an diese Umgebung anpassen. Dies geschieht am besten unter Einbeziehung der jeweiligen Partner aus dem betreffenden Bereich. Wenn das Tool in Land X eingesetzt werden soll, sollten Sie es mit Daten aus der entsprechenden Bevölkerung oder, falls dies nicht möglich ist, einer ähnlichen Bevölkerung validieren. Andernfalls könnten die Ergebnisse völlig falsch sein. In jedem Fall sollten Sie jeden möglichen Nutzer über die Bedingungen der Validierung informieren.

Darüber hinaus muss der Validierungsprozess regelmäßig überprüft werden, wenn sich die Bedingungen ändern oder wenn der Verdacht besteht, dass die Lösung selbst verändert werden könnte. Wenn der Algorithmus beispielsweise mit Daten einer spezifischen Bevölkerungsgruppe gefüttert wird, sollten Sie prüfen, ob sich dadurch seine Genauigkeit bei einem anderen Teil der Bevölkerung ändert oder nicht. Sie müssen gewährleisten, dass die Validierung die Bedingungen, unter denen der Algorithmus validiert wurde, genau widerspiegelt.

Dazu sollte die Validierung alle Komponenten eines KI-Tools einbeziehen, einschließlich der Daten, im Voraus geschulter Modelle, Umgebungen und des Verhaltens des Systems als Ganzes. Darüber hinaus sollte die Validierung so bald wie möglich erfolgen. Schließlich muss sichergestellt sein, dass die Ergebnisse bzw. Handlungen mit den Ergebnissen der vorangegangenen Prozesse übereinstimmen, indem sie zur Gewährleistung ihrer Konformität mit den zuvor definierten Zielsetzungen verglichen werden.[648] Zur Validierung müssen gelegentlich neue personenbezogene Daten erhoben werden. In einigen anderen Fällen verwenden die Verantwortlichen die Daten für andere Zwecke als die ursprünglich vorgesehenen. In all diesen Fällen sollten die Verantwortlichen die Einhaltung der DSGVO

[648] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI, 2019, Seite 22. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020)

sicherstellen (siehe Abschnitt „Zweckbindung“, Kapitel „Grundsätze“ sowie „Datenschutz und wissenschaftliche Forschung“ in Abschnitt „Konzepte“).

5.2.2 Löschen unnötiger Datensätze

Häufig sind der Validierungs- und Trainingsprozess auf gewisse Weise miteinander verbunden. Wenn im Rahmen der Validierung erforderliche Verbesserungen des Modells erkannt wurden, sollte das Training erneut durchgeführt werden. Nach Abschluss des KI-Tools ist auch seine Trainingsphase beendet. Zu diesem Zeitpunkt sollten Sie den zu diesem Zweck verwendeten Datensatz löschen, es sei denn, es besteht eine rechtliche Notwendigkeit, ihn für die Weiterentwicklung oder Bewertung des Systems oder für andere Zwecke zu speichern, die gemäß den Bestimmungen von Artikel 6 Absatz 4 DSGVO mit den Zwecken vereinbar sind, für die sie erhoben wurden (siehe Abschnitt „Definition angemessener Speicherungsrichtlinien“ in diesem Dokument).

Beantragen betroffene Personen die Löschung der Daten, müssen Sie einen Einzelfallansatz wählen und dabei etwaige in der Verordnung vorgesehene Einschränkungen dieses Rechts berücksichtigen (siehe Artikel 17 Absatz 3).[649]

5.2.3 Durchführung eines externen Audits der Datenverarbeitung

Da die Risiken des von Ihnen entwickelten Systems hoch sind, **muss ein Audit des Systems durch einen unabhängigen Dritten einbezogen werden**. Es können verschiedene Audits durchgeführt werden. Sie können intern oder extern erfolgen, sie können sich nur auf das Endprodukt beziehen oder in Bezug auf weniger weit entwickelte Prototypen durchgeführt werden. Man könnte sie als eine Form von Überwachungs- und Transparenzinstrument betrachten, das ebenfalls eine Qualitätsfunktion erfüllen soll.

KI-Lösungen müssen im Hinblick auf die rechtliche Richtigkeit geprüft werden, um festzustellen, ob sie unter Berücksichtigung einer Vielzahl von Aspekten ordnungsgemäß mit der DSGVO übereinstimmen. Die Hochrangige Expertengruppe für künstliche Intelligenz stellte in Bezug auf das Erproben und Prüfen fest: *„Es sollte von einer möglichst vielfältig zusammengesetzten Personengruppe entworfen und entwickelt werden. Für die Kategorien, die aus verschiedenen Perspektiven getestet werden, sollten vielfältige Metriken entwickelt werden. Kontradiktorische Tests durch vertrauenswürdige, gemischte „rote Teams“, die bewusst versuchen, das System zu „knacken“, um Schwachstellen zu finden, und „Bug-Bounty“-Programme, die Außenstehende dazu anregen, Systemfehler und -schwächen zu erkennen und verantwortungsbewusst zu melden, sollten in Betracht gezogen werden.“*[650] Die Prüfung muss auch die Erfüllung des Grundsatzes der Erklärbarkeit umfassen. *„Bis zu welchem Grad Erklärbarkeit notwendig ist, hängt sehr stark vom Kontext und der Tragweite der Konsequenzen eines fehlerhaften oder anderweitig unzutreffenden Ergebnisses ab.“*[651] In Anbetracht der sehr schwerwiegenden Folgen für Personen, die krimineller Aktivitäten verdächtigt oder überführt werden, müssen die angewandten Technologien des maschinellen Lernens neben anderen erforderlichen Maßnahmen die Erklärbarkeit ermöglichen, damit die entwickelten Systeme die Grundrechte respektieren. Die Prüfung sollte sich auch auf die Maßnahmen zur Vermeidung von Verzerrungen, Unklarheiten, verstecktem Profiling usw.

---

[649] AEPD, Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, 2020, Seite 26. Verfügbar unter: https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

[650] Unabhängige hochrangige Expertengruppe für künstliche Intelligenz (2019) Ethik-Leitlinien für eine vertrauenswürdige KI, 2019, Seite 22. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 15. Mai 2020)

[651] Ibid., Seite 15.

sowie auf die korrekte Anwendung von Instrumenten wie der mehrfach durchführbaren Datenschutz-Folgenabschätzung konzentrieren. Die Umsetzung angemessener Datenschutzmaßnahmen ab den ersten Phasen des Lebenszyklus des Tools ist die beste Möglichkeit, um Datenschutzprobleme zu vermeiden.

6 Bereitstellung

6.1 Beschreibung

„Die Bereitstellung ist der Prozess, bei dem ein IT-System in seiner Umgebung betriebsbereit gemacht wird, einschließlich Installation, Konfiguration, Betrieb, Tests und Vornahme der erforderlichen Änderungen. Die Bereitstellung erfolgt in der Regel nicht durch die Entwickler eines Systems, sondern durch das IT-Team des Kunden. Aber selbst in diesem Fall sind die Entwickler dafür verantwortlich, dem Kunden ausreichende Informationen für eine erfolgreiche Bereitstellung des Modells zu liefern. Dazu gehört in der Regel ein (allgemeiner) Bereitstellungsplan mit den notwendigen Schritten für eine erfolgreiche Bereitstellung und deren Durchführung sowie ein (allgemeiner) Überwachungs- und Wartungsplan für die Wartung des Systems, die Überwachung der Bereitstellung und die korrekte Nutzung der Data-Mining-Ergebnisse.“[652]

6.2 Wichtigste zu ergreifende Maßnahmen

6.2.1 Allgemeine Bemerkungen

Sobald Sie Ihren Algorithmus erstellt haben, stehen Sie vor einem wichtigen Problem. Es könnte passieren, dass er offen oder verdeckt personenbezogene Daten enthält. In jedem Fall müssen Sie eine formale Bewertung durchführen, um festzustellen, welche personenbezogenen Daten der betroffenen Personen identifizierbar sein könnten. Dies kann mitunter kompliziert sein. Beispielsweise könnten einige KI-Lösungen wie Vector Support Machines (VSM) in der Logik des Modells von Trainingsdaten durch Technikgestaltung enthalten. In anderen Fällen können im Modell Muster gefunden werden, die eine bestimmte Person identifizieren. In all diesen Fällen könnten Unbefugte in der Lage sein, Elemente der Trainingsdaten wiederherzustellen oder durch Analyse des Verhaltens des Modells auf die Person zu schließen, die darin enthalten war. Wenn Sie wissen oder vermuten, dass das KI-Tool personenbezogene Daten enthält (siehe auch Abschnitt „Erwerb oder Förderung des Zugangs zu einer Datenbank“, Kapitel „Maßnahmen und Tools“), sollten Sie:

- Die Daten löschen oder ggf. begründen, warum dies aufgrund der Beeinträchtigung des Modells ganz oder teilweise nicht möglich ist (siehe Abschnitt „Speicherbegrenzung“, Kapitel „Grundsätze“).
- Die Rechtsgrundlage für die Übermittlung personenbezogener Daten an Dritte bestimmen, insbesondere wenn besondere Datenkategorien betroffen sind (siehe Unterabschnitt „Rechtmäßigkeit“, Abschnitt „Rechtmäßigkeit, Verarbeitung nach Treu und Glauben und Transparenz“).

[652] SHERPA, Guidelines for the Ethical Development of AI and Big Data Systems: An Ethics by Design approach, 2020, Seite 13. Verfügbar unter: https://www.project-sherpa.eu/wp-content/uploads/2019/12/development-final.pdf, abgerufen am 15. Mai 2020.

- Inkenntnissetzung der betroffenen Personen über die oben beschriebene Verarbeitung.
- Nachweisen, dass die Richtlinien zum Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen umgesetzt worden sind (insbesondere die Datenminimierung).
- Eine Datenschutz-Folgenabschätzung (DSFA) durchführen.

Schließlich sollten Sie jedoch regelmäßig Prüfungen durchführen, um die Wahrscheinlichkeit beruhend auf dem Stand der Technik zu bewerten, dass personenbezogene Daten aus Modellen hergeleitet werden können, sodass das Risiko einer versehentlichen Offenlegung minimiert wird. Lassen diese Prüfungen auf eine erhebliche Möglichkeit der Offenlegung von Daten schließen, sollten die erforderlichen Maßnahmen zur Verhinderung ergriffen werden (siehe Abschnitt „Integrität und Vertraulichkeit", Kapitel „Grundsätze").

6.2.2 Aktualisierung der Informationen

Wird der Algorithmus von einem Dritten implementiert, müssen Sie die Ergebnisse des in den Entwicklungsphasen eingesetzten Validierungs- und Überwachungssystems mitteilen und Ihre Mitarbeit bei der weiteren Überwachung der Ergebnisvalidierung anbieten. Es wäre auch ratsam, diese Art der Koordinierung mit allen Dritten vorzusehen, von denen Sie Datenbanken oder andere relevante Komponenten im Lebenszyklus des Systems erwerben. Wenn es sich um eine Datenverarbeitung durch einen Dritten handelt, müssen Sie sicherstellen, dass der Zugang auf einer Rechtsgrundlage erfolgt.

Darüber hinaus müssen dem Endnutzer in jeder Phase Echtzeitinformationen über die Werte in Bezug auf die Richtigkeit und/oder Qualität der hergeleiteten Informationen angeboten werden (siehe Abschnitt „Richtigkeit", Kapitel „Grundsätze"). Wenn die hergeleiteten Informationen die Mindestschwellenwert in Bezug auf die Qualität nicht erreichen, müssen Sie hervorheben, dass diese Informationen keinen Wert besitzen. Diese Anforderung bedeutet häufig, dass Sie detaillierte Informationen über die Trainings- und Validierungsphasen bereitstellen sollten. Besonders wichtig sind Informationen über die zu diesem Zweck verwendeten Datensätze. Andernfalls könnte die verwendete Lösung zu enttäuschenden Ergebnissen für die Endnutzer führen, die dann über die Ursache spekulieren müssen.

Sie müssen auch sicherstellen, dass jede reale Implementierung zudem die *Richtlinie über den Datenschutz im Bereich der Strafverfolgung* (Richtlinie 2016/680)[653] und deren spezifischer Umsetzung in den einzelnen Mitgliedsstaaten erfüllt. Bitte beachten Sie, dass dies in der Regel weniger restriktive Vorschriften für die Verwendung personenbezogener Daten für die Strafverfolgungsbehörden bedeutet. In der Strafjustiz ist die Bereitstellung von Beweisen oft eine mit einem erheblichen Aufwand verbunden. Es ist daher eine natürliche Tendenz, so viele Daten wie möglich zu sammeln und zu verarbeiten, die sich letztendlich als nützlich erweisen könnten. Diese Tendenz wird durch die zunehmenden technischen Möglichkeiten zur automatischen Analyse großer Datenmengen durch KI-Tools noch verstärkt. Eine Datenminimierung ist jedoch notwendig, und wirksame Gegenmaßnahmen gegen eine umfangreiche Datenerfassung und -verarbeitung müssen daher bereits in das Design der KI-Tools integriert werden.

[653] Richtlinie (EU) 2016/680 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung sowie zum freien Datenverkehr und zur Aufhebung des Rahmenbeschlusses 2008/977/JI des Rates, Amtsbaltt <http://eur-lex.europa.eu/legal-content/EL/TXT/?uri=OJ:L:2016:119:TOC >.

Die Einhaltung von Menschenrechten und ethischen Grundsätzen erfordert die Erfüllung weiterer wesentlicher Bedingungen:

„In Bezug auf Überwachungstechnologien muss die Beweislast bei den Staaten und/oder Unternehmen liegen, die öffentlich und transparent Nachweise erbringen müssen, bevor sie Überwachungsaktionen durchführen,

- dass diese notwendig sind,

- dass diese wirksam sind,

- dass diese verhältnismäßig sind (z. B. Zweckbindung),

- dass es keine besseren Alternativen gibt, die diese Überwachungstechnologien ersetzen könnten.

Die Einhaltung dieser Kriterien ist einer nachträglichen Beurteilung zu unterziehen. Dies muss entweder auf der Ebene der normalen politischen Analysen oder durch die diesbezüglichen Regelungen der Mitgliedstaaten geschehen.“

654

[654] Europäische Gruppe für Ethik und Naturwissenschaften und neuen Technologien. (2014). Stellungnahme Nr. 28: Ethik von Sicherheitsund Überwachungstechnologien (10.2796/22379). Abgerufen von Luxemburg: Brüssel: https://publications.europa.eu/en/publication-detail/-/publication/6f1b3ce0-2810-4926-b185-54fc3225c969/language-en/format-PDF/source-77404258