Datenschutzrecht für künstliche Intelligenz in der öffentlichen Verwaltung

Eine Auslegeordnung am Beispiel des Kantons Zürich

Philip Glass

Inhaltsübersicht

I. Künstliche Intelligenz und damit verbundene Risiken[1]

A. Eine grosse Diversität an Definitionen von KI

1. Klassische begriffliche Unterscheidungen

Die Erfassung des Phänomens der künstlichen Intelligenz («KI») bereitet gewisse Schwierigkeiten, weil bisher keine schlüssige Definition existiert. Dies ist einerseits historisch begründet, da das Ziel einer intelligent handelnden Maschine auf verschiedene Arten verfolgt wurde.[2] Andererseits liegt es auch daran, dass trotz Jahrhunderten der Nachforschung und des Überlegens nach wie vor keine anerkannte allgemeine Theorie von «Intelligenz» existiert und insgesamt unklar ist, was man unter dem Begriff der Intelligenz verstehen soll.[3] In Bezug auf «künstliche Intelligenz» besteht immerhin ein gewisser Konsens, dass es (vorerst) lediglich um die Simulation der von aussen beobachtbaren Attribute von Intelligenz geht, sogenannte *weak AI*,[4] insbesondere

---

[1] An dieser Stelle möchte ich Dr. Sc. (ETH Zürich) Colin W. Glass für die kritische Durchsicht des Manuskripts und Besprechung der technischen Aspekte der Untersuchung danken.

[2] Stuart Russel/Peter Norvig, Artificial Intelligence – A Modern Approach, Global Edition, 4. Ed., Pearson Education 2022, Introduction; vgl. dazu die Zusammenfassung der gängigsten Konzepte und Methoden bei Alfred Früh/Dario Haux, Foundations of Artificial Intelligence and Machine Learning, Weizenbaum Series 29, Berlin 2022, https://www.weizenbaum-institut.de/media/Publikationen/Weizenbaum_Series/Weizenbaum_Series_29.pdf (Abruf 06.09.2022), *passim*.

[3] Mark Coeckelbergh, Ai Ethics, Cambridge (MA)/London 2020, Kapitel 3; Yann Le Cun, Quand la machine apprend – La révolution des neurones artificiels et de l'apprentissage profond, Paris 2019, 374; Melanie Mitchell, Artificial Intelligence – A Guide for Thinking Humans, Pelican Books 2019, 6 f.; Stuart Russell, Human Compatible – AI and the Problem of Control, Penguin Books 2019, 13. f.; Stan Franklin, History, motivations, and core themes, in: Keith Frankish/William R. Ramsey (Eds.), The Cambridge Handbook of Artificial Intelligence, Cambridge University Press 2014, 15; für ein Beispiel der technischen Quantifizierung der universellen Intelligenz eines Systems siehe Jörg Zimmermann/Armin B. Cremers, Foundations of Artificial Intelligence and Effective Universal Induction, in: Joachim von Braun/Margaret S. Archer/Gregory M. Reichberg/Marcelo Sánchez Sorondo (Eds.), Robotics, AI, And Humanity – Science Ethics, and Policy, Springer Open Access 2021, 37.

[4] Zur Unterscheidung zwischen *weak AI* und *strong AI* siehe Russel/Norvig (FN 2), 1032 f.

Sprachverständnis, Wissensrepräsentation, automatisierte Schlussfolgerung – erweitert durch kognitive Werkzeuge der Simulation von Sinneswahrnehmung durch Sensoren wie beispielsweise die Bilderkennung.[5] Konkret einsatzfähige KI-Systeme müssen im Einzelnen jeweils für die Aufgabe programmiert bzw. trainiert werden, für die sie vorgesehen sind. Ob und gegebenenfalls wann eine allgemeine KI, *artificial general intelligence (AGI),* verschiedentlich auch *strong AI* oder *human level intelligence* genannt, für welche dies nicht mehr zwingend notwendig wäre, verfügbar sein wird, ist unklar.[6]

Im Kern geht es bei KI um konkrete Problemlösung durch Algorithmen als anpassbares, zunehmend intelligentes Verhalten.[7] Künstliche Intelligenz ist somit ein technologisches Querschnittskonzept, dessen methodische Konkretisierung mit dem Fortschritt der technologischen Entwicklung ändert. Entsprechend werden KI-Probleme, die als «gelöst» gelten, «in die Schublade der klassischen Werkzeuge versorgt» und mit der Zeit als *good old fashioned AI* (GOFAI) bezeichnet.[8] Aktuell bezeichnet der Begriff jene KI-Systeme, die auf der Grundlage von logischer Verknüpfung von formalen Regeln arbeiten.[9] Solche Methoden werden auch *symbolic AI* genannt, da ihre Wissensbasis aus symbolischen Repräsentationen für Aspekte der Maschinenumwelt be-

---

[5] Russel/Norvig (FN 2), 20; aus rechtlicher Sicht zuletzt Wolfgang Hoffmann-Riem, Recht im Sog der Digitalisierung – Herausforderungen, Tübingen 2022, 39 ff. m.w.H.

[6] Micheal Wooldridge, The Road to Conscious Machines – The Story of AI, Pelican Books 2021, 303 ff.; Le Cun (FN 3), 372 f.; Mitchell (FN 3), 363; Margaret A. Boden, Artificial Intelligence – A Very Short Introduction, Oxford University Press 2018, 47 f.; skeptisch gegenüber der Möglichkeit einer hierzu notwendigen «Intelligenzexplosion» Erik J. Larson, The Myth of Artificial Intelligence – Why Computers Can't Think the Way We Do, Cambridge MA/London 2021, 37 f.

[7] Wooldridge (FN 6), 22.

[8] Le Cun (FN 3), 17 f. (Übersetzung durch den Verfasser); Nadja Braun Binder/Matthias Spielkamp/Catherine Egli/Laurent Freiburghaus/Eliane Kunz/Nina Laukenmann/Michele Loi/Anna Mätzener/Liliane Obrecht/Jessica Wulf, Einsatz Künstlicher Intelligenz in der Verwaltung: rechtliche und ethische Fragen – Schlussbericht vom 28. Februar 2021 zum Vorprojekt IP6.4, Staatskanzlei des Kantons Zürich 2021, 10, hier: «Basistechnologie».

[9] Le Cun (FN 3), 384; Russel/Norvig (FN 2), 1033; Margaret A. Boden, GOFAI, in: Keith Frankish/William R. Ramsey (Eds.), The Cambridge Handbook of Artificial Intelligence, Cambridge University Press 2014, 89 ff.; zu den Kognitiven Grenzen des Ansatzes zusammenfassend Manuela Lenzen, Natürliche und künstliche Intelligenz – Einführung in die Kognitionswissenschaft, Frankfurt a.M. 2002, 63 ff.

steht.[10] Demgegenüber sind neue Methoden des *machine learning* in der Lage, «lernende» Algorithmen zu erzeugen, die auf ihr Lernziel hin optimiert werden können. Heutige KI-Systeme sind vielfach Mischungen verschiedener Technologien, oftmals GOFAI in Kombination mit maschinellem Lernen.[11] Aufgrund der Vielfalt der methodischen Zugänge enthalten wissenschaftliche, politische und rechtliche Stellungnahmen zum Thema je eigene Umschreibungen des Phänomens.

2. Definitionen der OECD und des Europarates

Auf einen offenen, technologieneutralen Begriff von KI legte sich die AI Expert Group der OECD fest. Die Definition besteht aus drei Elementen. Demnach handelt es sich bei KI-Systemen zunächst um *Maschinen-basierte Systeme*, die in der Lage sind, im Rahmen von Zielen, die durch Menschen definiert werden, Voraussagen zu treffen, Empfehlungen zu generieren oder auch ihre reale bzw. virtuelle Umgebung zu beeinflussen. Sodann nutzen solche Systeme begriffsnotwendig durch Menschen oder Maschinen generierte Inputdaten, um reale oder virtuelle Umgebungen wahrzunehmen und diese in statistische Modelle umzuwandeln. Bei der Anwendung auf neue Daten sind sie schliesslich in der Lage, auf der Grundlage dieser Modelle mittels *model inference*[12] Optionen und Informationen für Handlungen zu formulieren. Dabei verfügen solche Systeme über unterschiedliche Grade an Autonomie.[13]

---

[10] Boden (FN 6), 5, 19 ff.; Mitchell (FN 3), 9 ff.; Wooldridge (FN 6), 42 ff.: «This approach is called symbolic AI, because it makes use of symbols that stand for things that the system is reasoning about.»; vgl. auch die Erläuterung anhand von *natural language processing* bei Markus Christen/Clemens Mader/Johann Čas/Tarik Abou-Chadi/Abraham Bernstein/ Nadja Braun Binder/Daniele Dell'Aglio/ Luca Fábián/Damian George/Anita Gohdes/Lorenz Hilty/ Markus Kneer/Jaro Krieger-Lamina/Hauke Licht/Anne Scherer/Claudia Som/Pascal Sutter/Florent Thouvenin, Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz, TA-Swiss 72/2020, 84.

[11] Le Cun (FN 3), 2: «tricotage d'apprentissage-machine, de GOFAI et d'informatique classique».

[12] Vgl. dazu Paul DeBeasi, Training versus Inference, blogs.gartner.com, 14.02.2019, https://blogs.gartner.com/paul-debeasi/2019/02/14/training-versus-inference/ (Abruf 06.09.2022).

[13] Siehe die Definition bei OECD, Artificial Intelligence in Society, Paris 2019, https://doi.org/10.1787/eedfee77-en (Abruf 02.04.2022), 22.

Etwas abstrakter formuliert dies der Europarat im Rahmen seiner Studie über die Machbarkeit sowie mögliche Elemente eines rechtlichen Rahmens für die Entwicklung, das Design und den Einsatz von KI-Systemen im Hinblick auf den Schutz der Menschenrechte, der Demokratie und der Verwirklichung von Rechtsstaatlichkeit.[14] Als künstliche Intelligenz wird hier die Entdeckung von Mustern und Trends in grossen Datensets durch statistische Methoden bezeichnet. Hierdurch ermöglichten intelligente Algorithmen die Erkennung von Bildern und Tönen, das *streamlining* von Produkten und Dienstleistungen sowie grosse Effizienzgewinne bei der Verwirklichung komplexer Aufgaben.[15] Es handelt sich somit um einen Begriff von künstlicher Intelligenz, der mit *machine learning* gleichzusetzen ist und eine enge Verwandtschaft zum Konzept *big data* aufweist. Es war denn auch der Aufstieg von Big Data-Systemen, bestehend aus grossen Speichern, grossen Datenmengen und schnellen Prozessoren, welche im Nachgang zum Aufkommen des *world wide web* die Entwicklung neuer Methoden zur Entwicklung von Algorithmen beschleunigte, auf deren Grundlage probabilistische Modelle erstellt und trainiert werden können.[16]

3. «Big Algo» und «algorithmische Systeme»

In der Literatur wird deshalb auch der Begriff «Big Algo» vorgeschlagen, was den Fokus von den Daten zu Art und Kontext der Bearbeitung verschieben soll.[17] Soweit ersichtlich, hat sich der Begriff noch nicht durchgesetzt. Derweil scheint sich der Fokus der KI-Entwicklung aus verschiedenen Gründen eher weg von grossen Mengen hin zu qualitativ hochwertigen Datensets (*good data*) zu verlagern,[18] sowie zu optimierten simplen Regelmodellen.[19] Mit an-

---

[14] Ad Hoc Committee on Artificial Intelligence, Feasibility Study, CAHAI (2020)23, N 4.

[15] Feasibility Study 2020 (FN 14), N 4.

[16] Siehe dazu Russel/Norvig (FN 2), 43 ff.

[17] Siehe den Hinweis bei Hoffmann-Riem (FN 5), 38.

[18] Siehe dazu das Interview von Eliza Strickland, Andrew Ng: Unbiggen AI, IEEE Spectrum, 09.02.2022, abrufbar unter https://spectrum.ieee.org/andrew-ng-data-centric-ai (Abruf 06.09.2022); bzw. «smart data», Rolf H. Weber, Big Data: Herausforderung für das Datenschutzrecht, in: Astrid Epiney/Daniela Nüesch (Hrsg.), Big Data und Datenschutzrecht, Zürich/Basel/Genf 2016, 18 f.

[19] Zu dieser Entwicklung siehe Brian Christian, The Alignment Problem: Machine Learning and Human Values, New York City 2020, 98 ff.

deren Worten wird versucht, die Auswahl der Parameter für KI-Modelle besser empirisch abzustützen.

In eine ähnliche Richtung wie «Big Algo» geht die *Digital Society Initiative* der Universität Zürich, die in ihrem Positionspapier zur Regulierung von künstlicher Intelligenz anstelle von KI von «algorithmischen Systemen» spricht. Dies soll nicht bestimmte Technologien bezeichnen, sondern «auf die Anwendung dieser Technologien in einem sozialen Kontext [verweisen]».[20] Auch hier wird der Kontext der Anwendung in den Vordergrund gestellt und als primäre Grundlage für eine mögliche Regulierung gesehen. Der Begriff ist indes in mehrfacher Hinsicht problematisch, denn er erscheint sehr weit gefasst und enthält keine inhärenten Hinweise auf die (zusätzlichen) Bedeutungen, die ihm beigemessen werden.

Erstens bezeichnet der Begriff «Algorithmus» klar strukturierte Handlungsanweisungen, die zunächst einmal nicht als künstlich intelligent bezeichnet werden, beispielsweise Kochrezepte und andere Abläufe logischer Deduktion.[21] Mithin wird als Algorithmus «jede eindeutige Handlungsanweisung gekennzeichnet, die dafür eingesetzt wird, bestimmte Probleme in definierten Einzelschritten zu lösen».[22] Das Moment der künstlichen Intelligenz besteht denn auch nicht im Algorithmus, sondern im Modell, das durch den Algorithmus umgesetzt wird.[23] Mithin muss ein algorithmisches System nicht begriffsnotwendig künstlich intelligente Funktionen aufweisen.

Zweitens ist mit der Bezeichnung als System in der klassischen Bedeutung des Begriffs nicht ein sozio-technisches (Gesamt-)System gemeint, sondern ein Informationsverarbeitungssystem. Entsprechend bezeichnet der Begriff des algorithmischen Systems – analog zum verwandten, aber nicht identischen

---

[20] Florent Thouvenin/Markus Christen/Abraham Bernstein/Nadja Braun Binder/Thomas Burri/Karsten Donnay/Lena Jäger/Mariela Jaffé/Michael Krauthammer/Melinda Lohmann/Anna Mätzener/Sophie Mützel/Liliane Obrecht/Nicole Ritter/Matthias Spielkamp/Stephanie Volz, Positionspapier: Ein Rechtsrahmen für künstliche Intelligenz, DSI 2021, https://www.zora.uzh.ch/id/eprint/211386/ (Abruf 01.03.2022), 1.

[21] Russel/Norvig (FN 2), 27.; Vgl. beispielsweise den Herzstillstand-Algorithmus bei Hans Rickli (Hrsg.), Kardiovaskuläres Manual Kantonsspital St. Gallen, https://www.kssg.ch/sites/default/files/2016-05/kv-manual2011.pdf (Abruf 01.03.2022), 12.

[22] Hoffmann-Riem (FN 5), 36.

[23] Siehe III.A.1.

Begriff des KI-Systems[24] – ein Informationsverarbeitungssystem, das auf Basis von Algorithmen arbeitet.[25]

Aus rechtsmethodischer Sicht lassen die Überlegungen hinter den Begriffen «Big Algo» und «algorithmisches System» bzw. «algorithmisches/automatisiertes Entscheidungssystem» insofern aufhorchen, als sie an eine ähnliche Entwicklung im Datenschutzrecht erinnern. Hier hat sich mittlerweile durchgesetzt, dass sich der rechtliche Schutz nicht primär nach dem Inhalt der Daten richtet, sondern nach dem Kontext der Datenbearbeitung und den damit verbundenen Risiken für die Betroffenen.[26] Der Grund liegt zunächst darin, dass der Informationsgehalt von Daten, aufgrund dessen Entscheidungen letztlich getroffen werden, stets eine kontextbezogene Interpretativleistung darstellt.[27] Parallel dazu zeigt sich zunehmend deutlich, dass der Wert der erweiterten Persönlichkeitssphäre als Gegenstand der informationellen Selbstbestimmung – das primäre Schutzgut des Datenschutzrechts – ebenfalls kontextsensibel ist.[28] Indes blieb der Rechtsbegriff des Datums unverändert technisch.

Der Verweis auf die sozialen Auswirkungen ist aus dieser Perspektive von Bedeutung, als KI-Instanzen ein komplexes Zusammenspiel von Algorithmen

---

[24] Dazu Martin Ebers, Regulierung von KI und Robotik, in: Martin Ebers/Christian Heinze/Tina Krügel/Björn Steinrötter (Hrsg.), Künstliche Intelligenz und Robotik – Rechtshandbuch, München 2020, § 3 N 4.

[25] Hoffmann-Riem (FN 5), 195 ff.

[26] Zuletzt Joel Drittenbass, Regulierung von autonomen Robotern – Angewendet auf den Einsatz von autonomen Medizinrobotern: Eine datenschutzrechtliche und medizinproduktrechtliche Untersuchung, Diss. Univ. St. Gallen, Zürich/St. Gallen 2021, N 140; Philip Glass, Die rechtstaatliche Bearbeitung von Personendaten in der Schweiz – Regelungs- und Begründungsstrategien des Datenschutzrechts mit Hinweisen zu den Bereichen Polizei, Staatsschutz, Sozialhilfe und elektronische Informationsverarbeitung, zugl. Diss Univ. Basel 2016, Zürich/St. Gallen 2017, 125 ff. m.w.H.; Thomas Gächter/Gregori Werder, Einbettung ausgewählter Konzepte in das schweizerische Datenschutzrecht, in: Astrid Epiney/Tobias Fasnacht/Gaëtan Blaser (Hrsg.), Instrumente zur Umsetzung des Rechts auf informationelle Selbstbestimmung, Zürich/Basel/Genf 2013, 88; Eva Maria Belser, in: Eva Maria Belser/Astrid Epiney/Bernhard Waldmann, Datenschutzrecht – Grundlagen und öffentliches Recht, Bern 2011 (zit. VerfasserIn, Datenschutzrecht Grundlagen), 27 f. N 50 u. 53.

[27] Marion Albers, Informationelle Selbstbestimmung, Baden-Baden 2005, 95.

[28] Grundlegend Helen Nissenbaum, Privacy in Context – Technology, Policy, and the Integrity of Social Life, Stanford 2010, 129 ff.; vgl. zum Zweckbindungsgrundsatz IV.A.1.b.

bilden, und solche technischen Systeme über einzelne Nutzer(gruppen) zunehmend auf die Gesellschaft insgesamt wirken.[29] Dennoch sollte begrifflich weiterhin zwischen Algorithmen als Regulierungsgegenstand auf der einen und deren Einbettung in den sozialen Kontext als Zielgrösse der Regulierung auf der anderen Seite getrennt werden. Der Grund liegt darin, dass sowohl der Regulierungsgegenstand als auch die Regulierung desselben je eigene, individuelle wie gesellschaftliche Risiken erzeugen, deren Auswirkungen auch je separat zu beachten sind.

Besonders problematisch erscheint damit die gleichsetzende Umdeutung des Begriffs des algorithmischen Systems als sozio-technisches KI-System schliesslich deshalb, weil die Automatisierung von Prozessen durch einen in Software codierten Algorithmus und die Lernfunktion von künstlich intelligenten Modellen je eigene intrinsische Risiken bergen.[30] Der Begriff «algorithmisches System» verweist in seiner üblichen Bedeutung auf die Risiken von Automatisierung, während der Begriff «KI-System» darüber hinaus die spezifischen Risiken der Umsetzung von Ergebnissen der maschinellen Modellbildung mitumfasst. Daher erscheint der Begriff des algorithmischen Systems nicht als geeigneter Ersatz für den Begriff des KI-Systems. Dies gilt auch für verwandte Begriffe, wie «algorithmisches Entscheidungssystem»[31] oder «automatisiertes Entscheidungssystem»[32]. Immerhin wird hier durch den Fokus auf Entscheidungen eine gewisse Verbindung zu gesellschaftlichen Systemen hergestellt. Eine diesbezügliche Abgrenzung kann an dieser Stelle indes nicht geleistet werden.

---

[29] Hoffmann-Riem (FN 5), 35 ff.; Mario Martini, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, Berlin 2019, 64 f.

[30] Siehe I.D.

[31] Vgl. dazu Julia Krüger/Konrad Lischka, Damit Maschinen den Menschen dienen – Lösungsansätze, um algorithmische Entscheidungen in den Dienst der Gesellschaft zu stellen, Arbeitspapier im Auftrag der Bertelsmannstiftung, Mai 2018, abrufbar unter https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Algorithmenethik-Loesungspanorama.pdf (Abruf 01.03.2022).

[32] Vgl. dazu das Positionspapier der Digitalen Gesellschaft zur Regulierung von automatisierten Entscheidungssystemen vom 21. Februar 2022, abrufbar unter https://www.digitale-gesellschaft.ch/uploads/2022/02/Position-der-Digitalen-Gesellschaft-zur-Regulierung-von-automatisierten-Entscheidungssystemen-1.0.pdf (Abruf 01.03.2022).

4. Definition der EU

Einen konkreteren Begriff verwendet schliesslich die EU in ihrem Entwurf zur KI-Regulierung, indem sie neben einer ähnlich abstrakten Definition verschiedene Kategorien von Technologien bezeichnet, die ein KI-System konstituieren können. Dies steht in einem gewissen Gegensatz zur Entwicklung in der Schweiz. Hier wird eher eine technologieneutrale Regulierung angestrebt bzw. empfohlen.[33] Indes ist der vorgeschlagene Begriff trotz der enumerativen Elemente sehr weit gefasst. Als KI-System nach der Konzeption des EU-Entwurfs kommt demgemäss Software in Frage, die unter Verwendung gewisser, im Anhang der Regulierung beschriebenen methodischen bzw. technologischen Ansätzen entwickelt wurde. Solche Software soll indes nur dann als künstlich intelligent gelten, wenn sie dazu geeignet sind, zur Verwirklichung vorgegebener menschlicher Ziele sinnvollen Output zu generieren, insbesondere Inhalte, Prognosen, Empfehlungen oder Entscheidungen mit Auswirkungen auf die Umgebung, mit der sie interagieren.

B. Ein einfaches Modell einer *learner*-KI

1. Elemente und zyklische Phasen des Lernens

Moderne künstlich intelligente Systeme bestehen aus einem mehr oder weniger komplexen Algorithmus (*KI-Funktion*), der ein mathematisches Modell (*Transformationsregel*) an eingegebenen Daten (*Input*) abarbeitet, unter Umständen mittels Resultat-Feedback verbessert (*Lernfunktion*), und dessen Resultate (*Output*) Aussagen über statistisch belastbare Zusammenhänge in den Inputdaten (*Korrelationen*) erlauben. Hieraus lassen sich Tatsachenbehauptungen generieren, die mit einer gewissen Wahrscheinlichkeit zutreffend sind, beispielsweise in der Form von Prognosen, aufgrund derer geplant oder Empfehlungen, aufgrund derer gehandelt werden kann. Soweit diese als Daten erfasst und gespeichert werden, handelt es sich um *probabilistische*

[33] Herausforderungen der künstlichen Intelligenz – Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, SBFI Forschung und Innovation, Dezember 2019, 10; Christen et al. (FN 10), 291 f.; Thouvenin et al. (FN 20), 2.

Daten, also nicht um empirisch erhobene Daten.[34] Darüber hinaus kann die Auswahl der nachfolgenden Handlung ebenfalls automatisiert werden, etwa zur Begründung, Änderung oder Aufhebung von Rechten oder Pflichten einer Person oder zur Steuerung des Verkehrsflusses mittels Verkehrsanalyse mit automatisierten Geschwindigkeitsbeschränkungen. Je nachdem, ob schlussendlich Mensch oder Maschine entscheidet, spricht man von *Teilautomation* oder *Vollautomation* eines Entscheidungsprozesses.[35]

KI-Systeme sind demnach Informationsverarbeitungsmaschinen, die stets weiter optimiert werden können. Die Möglichkeit der fortlaufenden Verbesserung führt zu einer zyklischen Betrachtungsweise von solchen Systemen, dem sogenannten *Lebenszyklus* einer KI, der in diskrete Phasen unterteilt wird. Die Unterteilung erfolgt üblicherweise in die Phasen der Erstellung (Auswahl und Zusammenstellung der KI-Technologien), des Trainings, der Anwendung oder des Einsatzes sowie des Feedbacks, das den Lernprozess auslöst.[36] Während des Trainings, in der Anwendung sowie in der Feedbackphase können Personendaten bearbeitet werden, was je eigene Datenschutzfragen aufwirft. Bei der Auswahl und Zusammenstellung der KI-Technologien für eine geplante KI-Instanz ergeben sich die Datenschutzrisiken indes aus der gewählten Architektur, die opak oder transparent – also interpretier und erklärbar – ausgestaltet sein kann.[37]

2. Überwachtes versus unüberwachtes Lernen

Ziel einer KI ist die Berechnung von sinnvollen, prädiktiven Outputdaten. Der Sinn von Outputdaten ergibt sich wiederum aus dem Einsatzzweck der KI. Der Lernprozess, aufgrund dessen solche Programme als «intelligent» bezeichnet werden, besteht in der Optimierung einer sog. Transformationsregel, welche

---

[34] Dies birgt Fragen in Bezug auf die Datenrichtigkeit, siehe dazu IV.A.3.

[35] Braun Binder et al. (FN 8), 11.

[36] Vgl. dazu die übersichtliche Darstellung für neuronale Netze bei Samuel Klaus, KI trifft Datenschutz: Risiken und Lösungsansätze, in: Astrid Epiney/Sophia Rovelli (Hrsg.), Künstliche Intelligenz und Datenschutz – L'intelligence artificielle et protection des données, Zürich/Basel/Genf 2021, 83 f.

[37] Vgl. dazu die Liste möglicher Risiken und Lösungsansätze pro Phase bei Klaus (FN 36), 92 ff.; Zur Frage der Transparenz und Erkennbarkeit siehe IV.A.1.c. u. d.

auf Inputdaten nützliche Outputdaten abbildet.[38] Diese wird zunächst anhand von repräsentativen Trainingsdaten in Hinblick auf den Einsatzzweck trainiert. Beispielsweise können dies Bilder von Katzen und Hunden sein,[39] die der Algorithmus voneinander unterscheiden bzw. jeweils korrekt bezeichnen soll. Im Rahmen des Trainings werden diese Daten als Bilddateien in das System eingelesen. Je nach Automatisierungsgrad der KI sind die Inputdaten von Menschen vorsortiert, d.h. als Hundebilder und Katzenbilder beschriftet oder nicht. Man spricht hier von überwachtem Lernen (*supervised learning)* bzw. unüberwachtem Lernen (*unsupervised learning)*.[40]

Von den Daten, die für das Training zur Verfügung stehen, sollte ein zufällig ausgewählter Teil vor dem Training heraussortiert, als getrenntes Datenset gespeichert und vorerst zurückbehalten, d.h. nicht für das Training der Transformationsregel verwendet werden. Es handelt sich hierbei um jene Daten, die dazu benutzt werden, den fertigen Algorithmus anhand von «neuen», d.h. nicht im Trainingsset enthaltenen Daten zu testen oder auch zu validieren. Die Daten in diesem Set werden daher Test- oder Validierungsdaten genannt.[41]

Soweit mit *supervised learning* gearbeitet wird, besteht das Ziel der Transformationsregel darin, die Daten des Trainingsdatensets den jeweils korrekten Labels zuzuweisen (die Anwendung auf die Testdaten folgt später). Im Rahmen von *unsupervised learning* besteht das Ziel darin, dass die KI unterschiedliche diskrete Kategorien von Daten in den Trainingsdaten bildet und neue Daten nach diesen Kriterien sinnvoll unterscheiden kann, z.B. dahinge-

---

[38] Andreas Kaminski/Colin W. Glass, Das Lernen der Maschinen, in: Kevin Liggieri/Oliver Müller (Hrsg.), Mensch-Maschinen-Interaktion – Handbuch zu Geschichte – Kultur – Ethik, Berlin 2019, 130, 132.

[39] Offenbar ein beliebtes Beispiel; vgl. dazu die Grafik bei Klaus (FN 36), 83.

[40] Kaminski/C.W. Glass (FN 38), 130 f.; mittlerweile wird an sehr grossen und unspezifizierten Modellen geforscht, sog. *foundation models*; siehe dazu die Übersicht bei https://research.ibm.com/blog/what-are-foundation-models/ (Abruf 11.06.2022).

[41] Andriy Burkov, The Hundred-Page Machine Learning Book, Eigenpublikation 2019, 49, der hier zusätzlich den Begriff *holdout set* verwendet, weil die Daten *zurückbehalten* werden; Kaminski/C.W. Glass (FN 38), 131.

hend, ob sie Bilder von Katzen und Hunden enthalten. Es existieren Ansätze, die diesbezügliche Qualitätsprüfung ebenfalls zu automatisieren.[42]

Die Ergebnisse des Trainings sind nicht notwendigerweise empirisch belastbare Kausalzusammenhänge. Vielmehr handelt es sich beim Output des KI-Trainings um stochastische Daten, also um Daten, die in einem gegebenen Datensatz eine Wahrscheinlichkeitsverteilung in Bezug auf bestimmte Merkmale in den Daten beschreiben. Wird festgestellt, dass der Algorithmus die Trainingsdaten gut voraussagt, kann ein Testlauf mit den Testdaten starten. Hier werden klassische Fehlerquellen, insbesondere eine mögliche Überanpassung an die Trainingsdaten geprüft.

3. Klassische Trainings-«Fehler»

a. Underfitting

«Fehler» im Sinne von unerwünschten Ergebnissen können entstehen, wenn Abkürzungen genommen werden, indem beispielsweise das Modell einfacher angelegt wird, als der gewünschte Output erfordert oder die benutzten Features für die gesuchte Funktion nicht informativ genug sind, um sinnvolle, regelhafte Unterscheidungen zu ermöglichen. Es handelt sich um eine Form von hohem Bias.[43]

b. Overfitting

Umgekehrt besteht im Rahmen des überwachten Lernens immer die Gefahr, dass ein Modell zu komplex ist, d.h. über zu viele Freiheitsgrade verfügt, und in der Folge zu präzise für die spezifische Anwendung auf die Trainingsdaten optimiert wird.[44] Dies kann darin resultieren, dass der Algorithmus idiosynkratische Muster des Trainingsdatensets als verallgemeinerbare Unterscheidungsmerkmale übernimmt, wodurch das Modell in der allgemeinen Anwendung

---

[42] Beispiel für eine automatisierte Überprüfung sind *generative adversarial networks* (GAN). Hier generiert ein neuronales Netzwerk Bilder, beispielsweise von Katzen, während ein zweites Netzwerk diese Bilder mit echten Bildern von Katzen vergleicht und Feedback gibt; Vgl. dazu Martin Giles, The GANfather: The man who's given machines the gift of imagination, Technology Review 21.02.2018.

[43] Burkov (FN 41), 51.

[44] Kaminski/C.W. Glass (FN 38), 130.

schlechte Resultate zeigen wird.[45] Es liegt mit anderen Worten eine *Überanpassung* des Modells an die Trainingsdaten vor.[46]

c. Benachteiligender Bias (und Diskriminierung)

Als Bias im technischen Sinn werden gemeinhin Verzerrungen des KI-Modells relativ zu den realen Begebenheiten der KI-Umwelt bezeichnet. Es handelt sich um architektonisch bedingte Effekte von Lernalgorithmen.[47] Diese können eine moralische Bedeutung[48] sowie eine aus rechtlicher Sicht signifikante Form annehmen, indem sie Unterscheidungen treffen oder nicht treffen, und daraus eine unsachliche bzw. ungerechtfertigte Ungleichbehandlung oder Diskriminierung resultiert. Obwohl der Grund für das Entstehen von Bias in der logischen Architektur von Lernalgorithmen zu finden ist, liegt die Quelle von rechtlich signifikanten Verzerrungen oftmals in den Trainingsdaten, indem diese unsorgfältig ausgesucht wurden. Da sich technischer Bias kaum vermeiden lässt, müssen KI-Systeme in der Anwendung stets mit Blick auf mögliche rechtsungleiche oder gar diskriminierende Auswirkungen in der realen Welt im Auge behalten werden.

Je nach Algorithmus kann es indes schwierig bis unmöglich sein, die diskriminierende Wirkung eines Bias *ex post* rechtsgenüglich zu belegen. Lässt sich nämlich eine strukturelle Benachteiligung in den Ergebnissen statistisch nachweisen, wird dadurch zunächst nur erkennbar, dass der KI eine tendenziell diskriminierende Entscheidungsstruktur antrainiert wurde. Damit ist nicht belegt, dass diese diskriminierende Verzerrung in einem Einzelfall für das Ergebnis ausschlaggebend war, da andere Faktoren «KI-intern» höher gewichtet worden sein könnten. Vor allem lässt sich möglicherweise nicht erkennen, ob die Vermutung einer Diskriminierung, welche aufgrund des Nachweises des Bias entstanden ist, innerhalb der internen Logik des Algorithmus durch eine

---

[45] Burkov (FN 41), 52.

[46] Kaminski/C.W. Glass (FN 38), 130 f.; Burkov (FN 41), 23 f.

[47] Stefan Bauberger/Birgit Beck/Aljoscha Burchardt/Petter Remmers, Ethische Fragen der künstlichen Intelligenz, in: Günther Görz, Ute Schmid, Tanya Braun (Hrsg.), Handbuch der künstlichen Intelligenz, 6. A. Berlin Boston 2021, 918; Batya Friedman/Helen Nissenbaum, Bias in Computer Systems, ACM Transactions on Information Systems, Vol. 14, No. 03.07.1996, abgedruckt in: John Weckert (Hrsg.), Computer Ethics, London New York 2007 (2018), 220.

[48] Friedman/Nissenbaum (FN 47), 217.

besonders qualifizierte sachliche Begründung entkräftet und die Entscheidung dadurch gerechtfertigt wurde.[49] In solchen Fällen müsste die qualifizierte Rechtfertigung scheitern.

Das klassische Beispiel für einen technischen Bias, der zu diskriminierenden Entscheidungen führen kann, stammt aus der Gesichtserkennung.[50] Idealerweise sucht man hierfür ein KI-System, das grundsätzlich in der Lage wäre, jedes Gesicht der Welt zu erkennen, egal welchen Alters, Geschlechts oder Hautfarbe. Aufgrund der Tatsache, dass Erkennungsalgorithmen oft mit Bildern von Gesichtern trainiert werden, die überwiegend männlich und hellhäutig sind, zeigen KI-Systeme immer wieder Schwächen bei der Erkennung von Gesichtern von Frauen bzw. von dunkler Hautfarbe, was in einer schlechteren Trefferquote für die Erkennung von dunkelhäutigen weiblichen Gesichtern resultieren kann. Je nach Verwendung des KI-Systems entstehen hierdurch Nachteile für die Betroffenen.[51] Der Schlüssel zur Vermeidung von benachteiligendem Bias liegt in diesen Fällen in der sorgfältigen Auswahl von repräsentativen Trainingsdaten.

---

[49] Zur Diskriminierungsprüfung siehe René Rhinow/Markus Schefer/Peter Uebersax, Schweizerisches Verfassungsrecht, 3. erw. u. akt. Aufl., Basel 2016, N 1891; Giovanni Biaggini, in: Giovanni Biaggini (Hrsg.), BV Kommentar, 2. A., Zürich 2017 (zit. OFK BV-Biaggini), Art. 8 N 22; Bernhard Waldmann, in: Bernhard Waldmann/Eva Maria Belser/Astrid Epiney (Hrsg.), Schweizerische Bundesverfassung, Basler Kommentar, Basel 2015, Art. 8 BV N 87; Rainer Schweizer, St. Galler Kommentar zu Art. 8 BV, in: Bernhard Ehrenzeller/Benjamin Schindler/Rainer J. Schweizer/Klaus A. Vallender (Hrsg.), Die Schweizerische Bundesverfassung, St. Galler Kommentar, 3. A. St. Gallen 2014, (zit. SGK BV-VerfasserIn) N 54; Vincent Martenet, Commentaire sur article 8 Cst., in: Vincent Martenet/Jacques Dubey (Hrsg.), Constitution fédérale, Commentaire Romand, Basel 2021 (zit. CR Cst.-VerfasserIn), N 98.

[50] Zu den weiteren Herausforderungen der Gesichtserkennung siehe VII.A.3.

[51] Notorisch ist das «Gorilla»-Debakel von Google Photos: Mitchell (FN 3), 123 ff.; Tom Simonite, When It Comes to Gorillas, Google Photos Remains Blind, Wired.com 11.01.2018; Marco Metzler, Wie Computer lernen, uns zu diskriminieren, NZZ vom 04.03.2017; Richard Nieva, Google apologizes for algorithm mistakenly calling black people ‹gorillas›, cnet.com 01.07.2015; Maggie Zhang, Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software, Forbes.com 01.07.2015.

Probleme können aber dennoch (oder erst recht) auftauchen, wenn sorgfältig ausgewählte Daten eine in der Gesellschaft vorhandene, gruppenspezifische Benachteiligungen korrekt transportieren.[52] In diesem Zusammenhang sorgt es für Verwirrung, dass der Begriff «Bias» mit einer anderen Bedeutung verwendet wird, die jener des technischen Bias aber sehr ähnlich ist. Von einem vorbestehenden Bias (*pre-existing bias*[53] oder *societal bias*[54]) wird gesprochen, wenn die Ergebnisse der KI tatsächliche, unerwünschte Ungleichbehandlung in der Gesellschaft widerspiegeln.[55] Hier bezieht sich die Verzerrung nicht auf die realen Tatsachen, wie sie sind, sondern auf die realen Tatsachen, wie sie aus rechtlicher oder moralischer Sicht sein sollten. Aus technischer Sicht ist anzumerken, dass der nicht-technische, vorbestehende Bias sich auf die Ergebnisse der KI als Abbild gesellschaftlicher Zustände bezieht und nicht auf die korrekte Funktionsweise der KI.

Die Vermeidung von Bias mit rechtlich relevanten negativen Folgen kann demnach auf mindestens zwei konzeptionell nachvollziehbare Arten angegangen werden. Zum einen kann durch sorgfältige Auswahl der Trainingsdaten darauf geachtet werden, dass das zu analysierende Phänomen in repräsentativer Weise wiedergeben wird.[56] Zum anderen ist es denkbar, dass mittels Bias in KI-Systemen tatsächliche Benachteiligungen in der Gesellschaft identifiziert und der politischen Diskussion zugeführt werden.[57] Die erste

---

[52] Bauberger et al. (FN 47), 919; siehe die Beispiele bei Coeckelbergh (FN 3), 127 ff.; vgl. dazu den «Referenzfall COMPAS» bezüglich Strafaussetzung auf Bewährung in den USA bei Martini (FN 29), 55 f.; sowie Julia Angwin/Jeff Larson/Surya Mattu/Lauren Kirchner, Machine Bias, ProPublica.org, 23.05.2016.

[53] Friedman/Nissenbaum (FN 47), 218.

[54] Russel/Norvig (FN 2), 1043 f.

[55] Bauberger et al. (FN 47), 918 f.; a.M. Coeckelbergh (FN 3), AI Ethics, Cambridge (MA)/London 2020, 126, der mit *bias* offenbar Bias mit rechtlich relevanten Unterscheidungsfehlern (z.B. Diskriminierung) meint; ebenso Friedman/Nissenbaum (FN 47), 217.

[56] Zu den diesbezüglichen Herausforderungen Ben Schneiderman, Human-Centered AI, Oxford University Press 2022, 161 ff.

[57] Vgl. dazu Aria Khademi/David Foley/Sanghack Lee/Vasant Honavar, Fairness in algorithmic decision making: An excursion through the lens of causality, Proceedings of the World Wide Web Conference, ACM 2019, 2907-2914, arXiv:1903.11719; Christian (FN 19), 47 ff.

Vorgehensweise betrifft den Datenschutz direkt, indem über die Qualität der Daten grundrechtliche Risiken der Betroffenen vermindert werden. Die zweite Methode ist dagegen primär eine Frage der politischen und verfassungsrechtlichen Entwicklung im Umgang mit KI. Sie betrifft Datenschutzfragen insofern, als die zur entsprechenden KI-Analyse notwendigen Daten personenbezogene Daten sein werden.

C. Überprüfbarkeit in der Anwendung

Um herauszufinden, wie gut ein KI-System die gestellte Aufgabe in der Anwendung meistert, müssen dessen Resultate in irgendeiner Form überprüfbar sein.[58] Dabei ist zu unterscheiden, ob sich der zu prüfende Algorithmus im Trainingsstadium oder in jenem der Anwendung befindet. Im Training dient die Überprüfung dazu, den *Lernerfolg* zu messen, während in der Anwendung die *Nützlichkeit* des Algorithmus in Bezug auf die ihm gestellte Aufgabe im Vordergrund steht. Der Lernerfolg kann auf strukturell benachteiligenden oder diskriminierenden Bias untersucht,[59] die Nützlichkeit der jeweiligen Prognose dagegen im Einzelfall angefochten werden.

Die Überprüfung erfolgt grundsätzlich mithilfe der durch Menschen in einem Datenset feststellbaren Tatsachen, bzw. festgelegten Ein- und Ausgabewerte, der sog. *ground truth*.[60] In dem vorhin benutzten Beispiel sind dies die Tatsachen darüber, welche Bilder tatsächlich Hunde zeigen, welche Bilder tatsächlich Katzen zeigen. So kann festgestellt werden, ob und gegebenenfalls inwieweit die Transformationsregel, welche durch den Algorithmus gebildet wurde, das Verhältnis zwischen Ein- und Ausgabedaten korrekt abbildet.

---

[58] Siehe zum Transparenzprinzip IV.A.1.c.

[59] Siehe I.B.3.c.

[60] Kaminski/C.W. Glass (FN 38), 128.

Schwierig wird der Vergleich dort, wo die tatsächlichen Gegebenheiten nur schwer feststellbar sind. In solchen Fällen kann möglicherweise mittels statistischer Evidenz ermittelt werden, ob die individuellen, wie gesellschaftlichen Auswirkungen der Verwendung von Ergebnissen einer KI-Instanz den damit verfolgten Zweck erfüllen oder nicht.[61] Auf diese Weise könnte in manchen Fällen auch ein benachteiligender bzw. diskriminierender Bias nachgewiesen werden.[62] Ein statistischer Prüfungsansatz bedingt indes, dass sinnvolle Messwerte sowie eine Bandbreite von akzeptablen Resultaten festgelegt werden können.

D. Nicht-lernende KI-Systeme

Als künstlich intelligent gelten neben Lernalgorithmen nach wie vor Systeme aus den ersten Jahrzehnten der KI-Forschung, die auf unveränderlichen, deterministischen Entscheidungsalgorithmen basieren. Im Gegensatz zu Lern-KI arbeiten einfache logische Agenten mit vorprogrammierten Regeln und nicht mit erlernten Wahrscheinlichkeiten. Sie sind daher grundsätzlich berechenbar und transparent.[63] Ein gutes Beispiel sind einfache Expertensysteme, deren Programm aus einer Sammlung von logisch verknüpften Aussagen bestehen, die jeweils Expertenwissen in einer Domäne wiedergeben.

Die intrinsischen Datenschutzprobleme solcher Algorithmen erscheinen aufgrund ihrer Architektur als weit weniger weitreichend wie jene von Lernalgorithmen. Soweit KI-Systeme indes nicht-lernende Elemente enthalten, werden diese – wie andere Software auch – in die datenschutzrechtliche Beurteilung des Gesamtsystems einbezogen werden müssen.

---

[61] Joanna J. Bryson, The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation, in: Markus D. Dubber/Frank Pasquale/Sunit Das (Hrsg.), The Oxford Handbook of Ethics of AI, Oxford University Press 2020, 9.

[62] OECD (FN 13), 92.

[63] Siehe IV.A.1.d.

II. Vorüberlegungen zu KI und Datenschutz

A. KI-Wertung versus Selbstwertung

Die Möglichkeit von künstlicher Intelligenz, Entscheidungsprozesse zu automatisieren, steht in einem gewissen Widerspruch zum Kerngedanken des Datenschutzes, wonach jedem Menschen ein verfassungsmässiges Recht auf informationelle Selbstbestimmung zukommt. Der Grund liegt darin, dass die Automatisierung eine Einflussnahme der Betroffenen auf das Ergebnis im Einzelfall verunmöglichen oder zumindest in unzumutbarer Weise erschweren und zudem zu einer unübersichtlichen Datenlage hinsichtlich der «eigenen» Personendaten führen kann.[64]

Weitere Risiken für die Grundrechte der Betroffenen entstehen, wenn KI-Systeme ihnen gewisse Persönlichkeitsaspekte wie etwa «erhöhte Gefährlichkeit»[65] oder Emotionen[66] zuweisen, insb. wenn diese Zuweisung die Grundlage für eine (hier: staatliche) Entscheidung bildet. Die (implizite) Verknüpfung mit besonders geschützten Merkmalen von Art. 8 BV kann zudem zu einer (mittelbaren) Diskriminierung durch einen benachteiligenden Bias im verwendeten Algorithmus führen.[67] Schliesslich ist ebenso bedenklich, dass die Ergebnisse der Ausforschung von Personen durch KI als Hebel für Manipulation verwendet werden können, und zwar auf einer individuellen wie auch auf einer gesellschaftlichen bzw. demokratischen Ebene.[68]

---

[64] Thomas Wischmeyer, Regierungs- und Verwaltungshandeln durch KI, in: Martin Ebers/Christian Heinze/Tina Krügel/Björn Steinrötter (Hrsg.), Künstliche Intelligenz und Robotik – Rechtshandbuch, München 2020, § 20 N 51 ff.

[65] Beispielsweise im Rahmen von *predictive policing,* siehe VII.A.6.

[66] Catrin Misselhorn, Künstliche Intelligenz und Empathie – Vom Leben mit Emotionserkennung, Sexrobotern & Co., Reclam Ditzingen 2021, 20 ff.

[67] Siehe I.B.3.c.

[68] Zum Ganzen BVerfG, 06.11.2019 – 1 BvR 16/13 – Recht auf Vergessen I, N 85; Nadja Braun Binder/Thomas Burri/Melinda Florina Lohmann/ Monika Simmler/ Florent Thouvenin/Kerstin Noëlle Vokinger, Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, in: Jusletter vom 28.06.2021, N 31 ff.; Dirk Helbling, Societal, Economic, Ethical and Legal Challenges of the Digital Revolution – From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies, in: Jusletter IT vom 21.05.2015, N 40; Sami Coll, Big Data, Big Problem?, in: Astrid Epiney/Daniela Nüesch (Hrsg.), Big Data und Datenschutzrecht, Zürich/Basel/Genf 2016, 26 f.

Datenschutz ist einer jener Bereiche, in denen das sogenannte *value alignment problem* der KI-Technologien sehr deutlich zum Vorschein kommt. Der Begriff bezeichnet die Schwierigkeit, die Wertvorstellungen der menschlichen Nutzerinnen – inklusive jenen Werten, die durch die Normen des Rechts transportiert werden – mit den durch die Automatisierungsfunktion der KI tatsächlich beförderten Werten in Einklang miteinander zu bringen.[69] Aus technischer Sicht zeigt sich das Problem dort, wo die *utility function,* d.h. die internalisierte Werteskala bzw. -gewichtung der KI in ihrer Lösungsstrategie gesellschaftliche und individuelle Wertvorstellungen bezüglich des zu lösenden Problems nicht miterfasst.[70] Gerade weil KI-Systeme laufend verfeinert und vergrössert werden und in der Form von sog. *foundation models* mittlerweile mit Millionen von automatisch generierten Parametern operieren können,[71] ist die Auflösung dieses Zielkonflikts alles andere als trivial, zumal die Erarbeitung der Parameter für die internalisierte Werteskala oftmals als Geschäftsgeheimnis aufgefasst und entsprechend intransparent und entkoppelt von Nutzern bzw. gesellschaftlichen Entscheidungsstrukturen erfolgt.[72] Soweit öffentliche Organe solche KI-Produkte nutzen, kann dies die demokratische Legitimation von Normgebungs- oder anderen staatlichen Entscheidungsprozessen unterwandern, so etwa, wenn Beurtei-

---

[69] Le Cun (FN 3), 370 ff.; zur Problemstellung in Bezug auf Moral Lukas Brand, Künstliche Tugend – Roboter als moralische Akteure, Regensburg 2018, 80 ff.

[70] Russel/Norvig (FN 2), 1054 f.; dies kann sich in einem diskriminierenden Bias zeigen, siehe I.B.3.c.

[71] Siehe beispielsweise Suman Bhattacharyya, Meta Unveils New AI Supercomputer, WSJ vom 24.01.2022; zum Begriff siehe FN 40; zur Problematik siehe auch Briefing, Huge «foundation models» are turbo-charging AI progress, The Economist, 11.07.2022, https://www.economist.com/interactive/briefing/2022/06/11/huge-foundation-models-are-turbo-charging-ai-progress (Abruf 12.06.2022); Rishi Bommasani/Percy Liang, Reflections on Foundation Models, Stanford HAI, 18.10.2021, https://hai.stanford.edu/news/reflections-foundation-models (Abruf 01.06.2022).

[72] Martini (FN 29), 33 ff.; Glass (FN 26), 214 f.

lungsspielräume der Verwaltung durch Entscheidungen eines KI-Systems konkretisiert werden.[73]

Schliesslich ist zu bedenken, dass die internalisierte Werteskala im Hinblick auf ihre Nützlichkeit in Bezug auf das Zielergebnis der KI optimiert wird. Letzteres kann moralisch signifikante Auswirkungen in der KI-Umwelt zeitigen, wird aber auf absehbare Zeit nicht von einer KI-Funktion moralisch bewertet werden können.[74] Damit bleibt auch die Möglichkeit einer rechtlichen Selbstkontrolle durch die betreffende KI vorerst auf die Anwendung von entscheidbaren Regelnormen bzw. die Erkennung von klar definierten Sachverhaltselementen (beispielsweise die Identifikation eines Nummernschildes) beschränkt.

B. Selbstbestimmung in Bezug auf personenbezogene Daten

In Zusammenhang mit der Digitalisierung im Allgemeinen und künstlicher Intelligenz im Besonderen wird in der Literatur erneut darauf hingewiesen, dass das Recht auf informationelle Selbstbestimmung durch die technische Entwicklung obsolet geworden oder gar von Beginn an eine Fehlkonstruktion gewesen sei.[75] Diese Kritik wurde bereits nach der Entwicklung des Rechts durch das Bundesverfassungsgericht im Volkszählungsurteil vorgebracht,[76]

---

[73] Danielle Keats Citron, Technological Due Process, 85 WASH. U. L. REV. 1249 (2008), https://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2 (Abruf 07.02.2022), 1294 ff.; zur Opazität von COMPAS bezüglich Rückfallwahrscheinlichkeit Monika Simmler/Giulia Canova, Smart Government in der Strafrechtspflege: Wann ist Smart Criminal Justice smart?, in: Monika Simmler (Hrsg.), Smart Criminal Justice – Der Einsatz von Algorithmen in der Polizeiarbeit und Strafrechtspflege, Basel 2021, 50.

[74] Catrin Misselhorn, Grundfragen der Maschinenethik, Reclam Ditzingen 2018, 70 ff.; Brand (FN 69), 89 f.

[75] Hinweise bei Braun Binder et al. (FN 68), N 17.; Vgl. auch die vorgeschlagene Neukonzeption für das deutsche Recht bei Marion Albers, Informationelle Selbstbestimmung als vielschichtiges Bündel von Rechtsbindungen und Rechtspositionen, in: Michael Friedewald/Jörn Lamla/Alexander Rossnagel (Hrsg.), Informationelle Selbstbestimmung im digitalen Wandel, Wiesbaden 2017, 21 ff.

[76] Siehe die Hinweise bei Glass (FN 26), 154 f.; Hanspeter Bull, Informationelle Selbstbestimmung – Vision oder Illusion? – Datenschutz im Spannungsverhältnis von Freiheit und Sicherheit, Tübingen 2009, 45 ff.; Florent Thouvenin, Informational Self-Determination: A Convincing Rationale for Data Protection Law?, JIPITEC 4/2021, N 4 ff.

sowie nach der Übernahme des Grundrechts als Teilgehalt der persönlichen Freiheit[77] in das schweizerische Recht durch das Bundesgericht.[78] Aus verschiedenen Gründen ist hier Vorsicht geboten.

Zunächst ist die Idee, ein Grundrecht aufgrund mangelnder Wirksamkeit infrage zu stellen, mit Gefahren für den Bestand des Grundrechts und damit auch für die geschützten Personen verbunden. Die Relativierung eines Grundrechts in seinem materiellen Gehalt aufgrund einer festgestellten mangelhaften Umsetzung käme einer Kapitulation gegenüber den betreffenden Verletzungskonstellationen gleich. Als legitim erscheint dagegen, die Wirkungsweise eines Grundrechts in Bezug auf die Verwirklichung von dessen Schutzgehalt zu hinterfragen und gegebenenfalls anzupassen.[79]

Das Konzept eines Rechts auf informationelle Selbstbestimmung wurde im Volkszählungsurteil des deutschen Bundesverfassungsgerichts begründet und war eine Reaktion auf einen drohenden Kontrollverlust,[80] den das Gericht in Bezug auf persönliche Informationen sowie das Bild, welches Dritte sich von der eigenen Person machen, diagnostiziert hatte.[81] Als Ausdruck eines informationellen Persönlichkeitsschutzes war es nie als «absolute Verfügungsbefugnis»[82] über «eigene» Personendaten konzipiert.[83]

Die aktuelle Rechtsprechung des Bundesverfassungsgerichts bekräftigt dies: Die ursprüngliche Formel aus dem Volkszählungsurteil, entwickelt als Abwehrrecht gegen den Staat,[84] wonach das Grundrecht eine Befugnis des Einzelnen gewährleiste, «grundsätzlich selbst über die Preisgabe und Ver-

---

77 Beat Rudin, in: Beat Rudin/Bruno Baeriswyl (Hrsg.), Praxiskommentar zum Informations- und Datenschutzgesetz des Kantons Basel-Stadt, Zürich/Basel/Genf 2014 (zit. VerfasserIn, PraKom IDG BS), Grundlagen, N 2.

78 Dazu Rudin, PraKom IDG BS (FN 77), Grundlagen, N 5 ff.

79 Vgl. dazu Eva Maria Belser, Zur rechtlichen Tragweite des Grundrechts auf Datenschutz: Missbrauchsschutz oder Schutz der informationellen Selbstbestimmung?, in: Astrid Epiney/Tobias Fasnacht/Gaëtan Blaser (Hrsg.), Instrumente zur Umsetzung des Rechts auf informationelle Selbstbestimmung, Bern 2013, 27.

80 Zu aktuellen Formen des drohenden Kontrollverlustes Albers (FN 75), 27.

81 BVerfGE, 65,1 (42 f.).

82 Belser (FN 79) 25; Albers (FN 75), 16.

83 Für das schweizerische Recht Beat Rudin, Kollektives Gedächtnis und informationelle Integrität, AJP 1998, 248 f.; Glass (FN 26), 155.

84 Albers (FN 75), 19.

wendung seiner persönlichen Daten zu bestimmen»[85] wurde im Verhältnis zwischen Privaten präzisierend ergänzt. Hier entfalte sie eine mittelbare Drittwirkung im Sinne einer «Gewährleistung, über der eigenen Person geltende Zuschreibungen selbst substanziell mitzuentscheiden».[86] Diese neue Formel eines «Rechts auf substanzielle Mitentscheidung» entspricht nach wie vor der ursprünglichen Funktion der informationellen Selbstbestimmung als ergänzendem Schutz der Privatheit in Bezug auf Daten über die eigene Person, welche sich ausserhalb des Herrschaftsbereichs des Individuums bei Dritten befinden.[87]

Es erscheint daher nach wie vor als sachgerecht, im schweizerischen Recht die informationelle Selbstbestimmung als Teilgehalt bzw. informationelle Dimension des in Art. 10 Abs. 2 BV garantierten Rechts auf persönliche Freiheit aufzufassen,[88] und als eigenständigen Aspekt des in Art. 13 Abs. 2 BV garantierten Missbrauchsschutz gegen grundrechtsverletzende Datenbearbeitungen beizubehalten.[89] Als solcher tritt es gleichsam als Recht auf «informationelle Integrität»[90] neben den körperlichen und geistigen Integritätsschutz. Der Schutzbereich erstreckt sich hierbei zunächst auf Personendaten, welche Informationen über elementare Erscheinungsformen der Persönlichkeitsentfaltung abbilden – unabhängig davon, wo sich diese befinden.[91] Da nun aber das allgemeine Persönlichkeitsrecht in enger Beziehung zu den übrigen Grundrechten steht, indem es als «verfassungsrechtliche Grundgarantie zum Schutz der Persönlichkeit» gilt, welche «hinter die speziellen Garantien zurücktritt»,[92] können die aus den speziellen Garantien fliessenden informationsspezifischen Teilgehalte umgekehrt als Ausbildungen eines so verstandenen informationel-

---

85 BVerfGE 65, 1 (43).

86 BVerfG, 06.11.2019 – 1 BvR 16/13 – Recht auf Vergessen I, N 86 f.; siehe auch Braun Binder et al. (FN 68), FN 31.

87 Glass (FN 26), 178 f.

88 Belser (FN 79), 27 ff..; Zur historischen Entwicklung dieser Verbindung und der diesbezüglichen bundesgerichtlichen Rechtsprechung siehe Belser (FN 26), Datenschutzrecht Grundlagen, 322 f.; Vgl. die Hinweise bei Rudin (FN 83), 248; SGK BV-Breitenmoser (FN 49), Art. 13 BV, N 4.

89 Im Ergebnis ähnlich Belser (FN 26), Datenschutzrecht Grundlagen, 378 N 121.

90 Rudin, PraKom IDG BS (FN 77), Grundlagen, N 4.

91 Glass (FN 26), 164 f.

92 OFK BV-Biaggini (FN 49), Art. 10 N 17.

len Persönlichkeitsschutzes verstanden werden. Dies bringt die ursprüngliche, vom Bundesgericht bestätigte Funktion der persönlichen Freiheit als Garantin der übrigen Rechte zum Ausdruck.[93]

Ein auf diese Schutzfunktion der persönlichen Freiheit bezogenes Recht auf informationelle Selbstbestimmung ist Garantin der übrigen Verfassungsrechte in Bezug auf die damit zusammenhängenden Informationen bzw. Personendaten. Als solche erstreckt sie sich über den Schutzbereich der Grundrechte insgesamt und kann auch als Grundlage und Massstab[94] dienen, um über das Verwirklichungsgebot in Art. 35 BV entsprechende Schutzpflichten der Datenbearbeiter festzulegen. Damit wird auch deutlich, dass die informationelle Selbstbestimmung primär gegenüber dem Staat wirksam und für Informationsvorgänge zwischen Privaten in der Regel konkretisierungsbedürftig ist. Mithin ist die «informationelle Integrität» als Persönlichkeitsgut im Sinne des zivilrechtlichen Persönlichkeitsschutzes zu betrachten.[95] Die in der Literatur beklagte faktische Wirkungslosigkeit oder auch Inhaltsleere der informationellen Selbstbestimmung[96] wird damit primär zum Thema für den Gesetzgeber,[97] die Gerichte, den eidgenössischen Datenschutzbeauftragten sowie die Lehre und die Anbieter von Rechtsdienstleistungen.[98]

Eine so verstandene informationelle Selbstbestimmung korrespondiert mit der Gesetzgebung in der Schweiz, soweit Datenschutzgesetze ausdrücklich den Schutz der Grundrechte der betroffenen Personen als Gesetzeszweck nennen,

---

93 BGE 90 I 29 E. 3a: «En d'autres termes, elle (die persönliche Freiheit; Anm. d. Verfassers) vise à garantir l'existence des conditions de fait indispensables pour que l'homme puisse effectivement exercer ces autres libertés»; siehe auch bei Belser (FN 26), Datenschutzrecht Grundlagen, 322 N 9.

94 Glass (FN 26), 161.

95 Dazu Philip Glass, Die Schutzparameter des zivilrechtlichen und des verfassungsrechtlichen Persönlichkeitsrechts, datalaw.ch, 28.05.2018, N 4 ff. m.w.H.

96 Belser (FN 79), 27 m.w.H.; Gächter/Werder (FN 26), 88.

97 Vgl. dazu Drittenbass (FN 26), N 144.

98 Ähnlich Gächter/Werder (FN 26), 95, allerdings nicht als Ausdruck der informationellen Selbstbestimmung, sondern als konkretisierungsbedürftige «justiziable Minimalgarantie» des verfassungsrechtlichen Datenschutzes.

insbesondere Art. 1 DSG bzw. Art. 1 nDSG[99], sowie für den Kanton Zürich § 1 Abs. 2 Bst. b IDG ZH[100].[101]

Schliesslich sollte nicht vergessenen gehen, dass die Datenschutzidee aus der Beobachtung entstand, dass die Verletzung von Grundrechten durch die moderne Bearbeitung von Personendaten einfacher, leichter skalierbar und zugleich für den Einzelnen weniger durchschaubar geworden war. Mittlerweile droht das Ausmass an Ausforschung der Persönlichkeitsstruktur des Einzelnen in ein Übergewicht der Fremd- gegenüber der Eigenwahrnehmung zu münden. Damit verbunden besteht das Risiko der Zurechnung bzw. Übernahme eines datafizierten, d.h. aus quantifizierten Datenmustern zusammengesetzten,[102] unterkomplexen und aus der Drittperspektive entwickelten Selbst.[103] Die informationelle Selbstbestimmung im Sinne des datenschutzrechtlichen Grundrechtsschutzes betrifft daher stets nur die Bearbeitung von personenbezogenen Daten, d.h. insbesondere Erhebung, Aufbewahrung, Nutzung, Veränderung,

---

[99] Bundesgesetz vom 19. Juni 1992 über den Datenschutz (Datenschutzgesetz, DSG; SR 235.1); zum neuen Datenschutzgesetz des Bundes (nDSG) siehe Botschaft vom 25. September 2020 zum Bundesgesetz über den Datenschutz (Datenschutzgesetz, DSG), BBl 2020 7639.

[100] Gesetz vom 12. Februar 2007 über die Information und den Datenschutz des Kanton Zürich (Informations- und Datenschutzgesetz, IDG ZH; ON 170.4).

[101] BSK DSG-Urs Maurer-Lambrou/Simon Kunz, in: Urs Maurer-Lambrou/Gabor Blechta, Datenschutzgesetz – Öffentlichkeitsgesetz, Basler Kommentar, Basel 2014 (zit. BSK DSG-VerfasserIn), Art. 1 N 26; Rudin, PraKom IDG BS (FN 77), § 1 N 12; Bruno Baeriswyl, in: Bruno Baeriswyl/Beat Rudin (Hrsg.), Praxiskommentar zum Informations- und Datenschutzgesetz des Kantons Zürich, Zürich/Basel/Genf 2012 (zit. VerfasserIn, PraKom IDG ZH), § 1 N 10 f.; eher restriktiv interpretiert bei David Rosenthal, in: David Rosenthal/Yvonne Jöhri (Hrsg.), Handkommentar zum Datenschutzgesetz sowie weiteren, ausgewählten Bestimmungen, Zürich 2008 (zit. VerfasserIn, in: Handkommentar DSG), Art. 1 N 3; Vgl. auch die qualifizierte Form der gesetzlichen Verbindung von Datenschutz und Grundrechten in § 3 Abs. 4 Bst. a IDG BS, der die besonderen Personendaten definiert als «Personendaten, bei deren Bearbeitung eine besondere Gefahr der Grundrechtsverletzung besteht»; zum neuen Datenschutzgesetz des Bundes (nDSG) siehe BBl 2020 7639.

[102] Zum Begriff *datafication* siehe Viktor Mayer-Schönberger/Kenneth Cukier, Big Data – The Essential Guide to Work, Life and Learning in the Age of Insight, üb. u. erw. A., London 2017, 78 ff.

[103] Mireille Hildebrandt, Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning, Theoretical Inquiries in Law, Vol. 20.1 83 (2019), 92 f.

Weitergabe und Löschung.[104] Die tatsächliche Verwirklichung eines datenbasierten grundrechtlichen Risikos fällt nach wie vor in den «traditionellen» Schutzbereich des betreffenden Grundrechts. Denn das Datenschutzrecht schützt nicht direkt vor Grundrechtsverletzungen, sondern vor Datenbearbeitungen, welche diese als Risiko oder Gefahr für die Betroffenen begünstigen. Es ist mithin ein mit den Grundrechten eng verknüpfter, diesen vorgelagerter, präventiver und vor allem *eigenständiger* Schutzmechanismus.[105]

Im Ergebnis setzt das Datenschutzrecht angesichts der neuen Herausforderungen durch vernetzte künstlich intelligente Systeme und Agenten weiterhin auf die Autonomie des Einzelnen und die rechtliche Absicherung der hierzu notwendigen Bedingungen. In dieser Betonung der Selbstbestimmung ist die Forderung zu erblicken, auch in einer von Automation durchdrungenen Gesellschaft ein autonomes Subjekt bleiben zu dürfen – zumindest in einem gewissen Umfang. Schlussendlich ist für die Betroffenen unerheblich, ob die Bedrohung für ihre persönliche Entwicklung von empirischen oder probabilistischen Daten ausgeht. Letztere sind indes schwieriger zu verifizieren und stellen damit in der Tendenz die komplexere Bedrohung dar.

III. KI und Personendaten

A. Die künstlich intelligente Bearbeitung von Personendaten

Künstlich intelligente Systeme werden datenschutzrelevant, wenn mit ihnen Personendaten bearbeitet werden. Viele KI-Systeme, wie etwa solche zur vorausschauenden Instandhaltung von Maschinen, bearbeiten nur Sachdaten. Die Ergebnisse von KI-Bearbeitungen können aber potenziell auf dieselbe Weise wie empirisch erhobene Sachdaten mit Personen verknüpft werden, wodurch sie den Status von Personendaten erlangen.

---

[104] Albers (FN 75), 17.

[105] Vgl. aus der Perspektive des Privatrechts Alfred Früh, Roboter und Privacy: Informationsrechtliche Herausforderungen datenbasierter Systeme, AJP 2017 141–151, 145; Gächter/Werder (FN 26), 91.

Auch hier gilt, dass unter der Bearbeitung jede Form des Umgangs mit Personendaten durch das Programm gemeint ist. Dies ist grundsätzlich während sämtlichen Phasen des Lebenszyklus von Daten innerhalb einer KI-Anwendung[106] möglich. Personendaten können zunächst als Trainings-, Test- (bzw. Validierungs-) oder Inputdaten verwendet werden. Auch kann ein System darauf ausgerichtet sein, im Rahmen seines Outputs neue Personendaten zu generieren.

Schlussendlich birgt jedes Stadium des Lebenszyklus eines KI-Systems, also die Entwicklung bzw. das Trainieren von Transformationsregeln, die Erzeugung von Outputs aus Inputs, sowie die Einbettung in Entscheidungsprozesse eigene Risiken. Aus Sicht des Datenschutzrechts ist dies von Bedeutung, da sämtliche dieser Phasen je eigenständige Formen der Datenbearbeitung darstellen, die durch Verknüpfung mit Personen in den Geltungsbereich der Datenschutzgesetze fallen können. Die nachfolgende skizzenhafte Darstellung bezieht sich vornehmlich auf den Geltungsbereich des IDG ZH, indes sind die darin enthaltenen Überlegungen auf andere Datenschutzgesetze übertragbar.

1. Bearbeitung von personenbezogenen Daten als Trainings- oder Validierungsdaten

KI-Systeme, die nützliche künstlich intelligente Funktionen bereitstellen, können deterministischen Vorgaben folgen oder trainiert werden. Die initiale Programmierung eines Lernalgorithmus bleibt insofern statisch, als zunächst eine Grundformel entwickelt werden muss, auf deren Basis das KI-System trainiert wird.

Für die Beurteilung des grundrechtlichen Risikos ist von besonderer Bedeutung, dass das Training einer künstlich intelligenten Funktion stets darauf abzielt, eine Transformationsregel zwischen Input- und Outputdaten zu bilden; Ziel ist die Bildung einer Regel als Resultat des Lernprozesses und nicht eine Einzelfallentscheidung; in diesem Lernvorgang liegt das Moment der künstlichen Intelligenz.[107] Das KI-System als Ergebnis des Trainings

[106] Zum Lebenszyklus von KI-Systemen vgl. OECD (FN 13), 26.

[107] Kaminski/C.W. Glass (FN 38), 130.

prägt dadurch die spätere Entscheidungsfindung in der Anwendung.[108] Auch wenn das Modell anhand von personenbezogenen Daten trainiert wird, bleibt das Training damit auf eine Kategorienbildung beschränkt. Die Outputdaten des Trainings dienen der Überprüfung der Transformationsregel und bilden weder eine Unterstützung im Hinblick auf die Entscheidung eines Einzelfalls, noch beziehen sie sich auf eine bestimmte oder bestimmbare Person.[109]

2. Bekanntgabe von Personendaten zu Trainingszwecken

Aufgrund dessen ist davon auszugehen, dass für die Nutzung von Personendaten als Trainingsdaten die Bestimmungen über die Nutzung von Personendaten zu nicht personenbezogenen Zwecken zur Anwendung gelangen. Für den Kanton Zürich bestimmt das Gesetz zunächst in § 9 Abs. 2 IDG ZH, dass öffentliche Organe Personendaten zu nicht personenbezogenen Zwecken bearbeiten dürfen, «wenn sie anonymisiert werden und aus den Auswertungen keine Rückschlüsse auf betroffene Personen möglich sind». Sie dürfen überdies gemäss § 18 IDG ZH Personendaten zum Zweck der nicht personenbezogenen Bearbeitung an Dritte bekanntgeben, soweit diese nachweisen, «dass die Personendaten anonymisiert werden, aus den Auswertungen keine Rückschlüsse auf betroffene Personen möglich sind und die ursprünglichen Personendaten nach der Auswertung vernichtet werden».

Entscheidend erscheint hier, dass nur eine (relative) Unmöglichkeit von Rückschlüssen auf *betroffene* Personen nachzuweisen ist – also jene Personen, deren Daten zum Zweck des Trainings bekannt gegeben wurden. Erwiesen und damit bedenkenswert erscheint in diesem Zusammenhang, dass es unter Umständen möglich sein kann, durch *re-engineering* des Modells einzelne Trainingsdaten zu de-anonymisieren und der richtigen Person zuzuordnen.[110] In diesen Fällen bzw. wenn keine Prognose bezüglich der Anonymität des Mo-

[108] Martini (FN 29), 50.

[109] Siehe I.B.

[110] Lena Leffer/Maximilian Leicht, Datenschutzrechtliche Herausforderungen beim Einsatz von Trainingsdaten für KI-Systeme, in: Jusletter IT vom 24.02.2022, N 21, schliessen daraus, dass gewisse KI-Modelle wie pseudonymisierte Personendaten zu behandeln sind, d.h. nach wie vor dem Datenschutzrecht unterstehen.

dells möglich ist, müsste ein KI-Modell im Zweifelsfall als «nicht anonym» und damit als Personendatum eingestuft werden.[111]

Die Zulässigkeit, zu einem späteren Zeitpunkt durch Anwendung der trainierten KI-Funktion Schlüsse auf *weitere* Personen zu ziehen, stellt insofern eine separate Datenschutzfrage dar, die bei Einsatz des KI-Systems zu beantworten ist.

Soweit schliesslich ein Teil der verfügbaren Daten von den Trainingsdaten getrennt wird, um die Leistungsfähigkeit des KI-Systems testen bzw. dessen Funktion validieren zu können,[112] sind diese Validierungsdaten wohl gleich zu behandeln wie die Trainingsdaten.

3. Bearbeitung von personenbezogenen Daten als Inputdaten

Soweit nun personenbezogene Daten als Inputdaten für ein KI-System verwendet werden, das System also im Rahmen der Tätigkeit eines öffentlichen Organs eingesetzt wird, sind verschiedene Szenarien denkbar. Erstens können personenbezogene Daten als Personendaten verwendet werden, beispielsweise für eine Profilbildung einer bestimmten bzw. bestimmbaren Person bzw. Gruppe von Personen oder als Grundlage für eine automatisierte Einzelentscheidung, beispielsweise für eine Parkbusse. In diesem Fall gelten die üblichen Bestimmungen des IDG ZH. Zweitens können die personenbezogenen Daten von den jeweiligen Personen entkoppelt werden, indem man die Sachdaten von den identifizierenden Daten trennt. Je nachdem, wie endgültig die Entkoppelung eingeschätzt wird, gelten andere gesetzliche Regelungen. Soweit eine Wiederherstellung des Personenbezugs ausgeschlossen werden kann, spricht man von *anonymen* Personendaten. Für sie gelten die besonderen Grundsätze der Bearbeitung von Personendaten in Abschnitt 2 des IDG ZH grundsätzlich nicht. Dies ergibt sich im Umkehrschluss aus der in § 3 Abs. 3 IDG ZH enthaltenen Definition von Personendaten als auf eine bestimmte oder bestimmbare Person bezogene Angaben.

---

[111] Dies entspricht offenbar gängiger Praxis, vgl. dazu David Rosenthal, Datenschutz und KI: Worauf in der Praxis zu achten ist, in: Sandra Husi-Stämpfli (Hrsg.), Jusletter IT vom 26.04.2022, N 52 ohne weitere Hinweise.

[112] Siehe I.B.2.

4. Bearbeitung von Personendaten als Outputdaten

Die Transformationsregel einer KI gilt gemeinhin als triviale Maschine, da die Inputdaten stets ein Paar mit den korrespondierenden Outputdaten bilden, d.h. die Regel deterministisch und unveränderbar funktioniert. Für Lernmaschinen ist dies insofern nicht uneingeschränkt der Fall, da diese darauf ausgelegt sind, einzelne Variablen der angewendeten Regel zu verändern, um den zu einem Input korrespondierenden Output zu optimieren. Ihre triviale Natur gilt demnach nur insoweit, als die Transformationsregel nicht verändert wird. Im Rahmen des Lernprozesses gelten sie daher als *nicht-trivial*.[113]

Die Verbindung zwischen Input und Output kann indes derart komplex sein, dass sie von Experten kaum oder gar nicht mehr ermittelt bzw. erklärt werden kann.[114] Aus Sicht des Datenschutzrechts stellt die korrelative Natur von berechneten Personendaten in solchen Fällen die Richtigkeit der Daten[115] sowie die Transparenz der Datenbearbeitung in Frage.[116] Zudem können (probabilistische) Personendaten generiert werden, ohne dass dies den Betroffenen bewusst ist. In diesen Fällen ist auch die Erkennbarkeit der Erhebung von Personendaten nicht ohne Weiteres gegeben.

B. Einbettung in den Verwaltungsprozess: KI-Technologie als qualifizierendes Merkmal für Datenbearbeitungen

Neben den gewöhnlichen Personendaten qualifiziert das Datenschutzrecht gewisse Personendaten bzw. gewisse Bearbeitungszusammenhänge von Daten, als «besondere»[117] bzw. «besonders schützenswerte»[118] Personendaten. Als besondere Personendaten gelten gem. § 3 Abs. 4 Bst. a IDG ZH jegliche «Informationen, bei denen wegen ihrer Bedeutung, der Art ihrer Bearbeitung oder der Möglichkeit ihrer Verknüpfung mit anderen Informationen die besondere Gefahr einer Persönlichkeitsverletzung besteht» sowie gem. § 3 Abs. 4 Bst. a

---

113 Kaminski/C.W. Glass (FN 38), 132.

114 OECD (FN 13), 82.

115 Zur Datenrichtigkeit IV.A.3.

116 Zur Transparenz bzw. Erkennbarkeit siehe IV.A.1.c u. d.

117 In manchen Kantonen, z.B. § 3 Abs. 4 IDG ZH.

118 Bund: Art. 3 Bst. c DSG (Art. 5 Bst. c nDSG).

Ziff. 1–4 IDG ZH Personendaten aus gewissen besonderen Lebensbereichen. Weiter fallen hierunter gemäss § 3 Abs. 4 Bst. b IDG ZH auch Persönlichkeitsprofile.

Der einfachste Fall einer qualifizierten Bearbeitung von Personendaten durch KI ist jener, dass Personendaten aus einem gesetzlich geschützten Lebensbereich von § 3 Abs. 4 IDG ZH im Rahmen eines KI-unterstützten Entscheidungsprozesses der Verwaltung bearbeitet werden. Hier stellt die KI-Funktion ein Werkzeug bereit, mit dessen Hilfe eine administrative Bearbeitung von Personendaten vorgenommen wird, die für sich bereits den Tatbestand der Bearbeitung von besonderen Personendaten erfüllt. Ein weiterer einfacher Fall ist jener, dass KI-Funktionen dazu genutzt werden, ein Profiling i.S.v. § 3 Abs. 4 Bst. c IDG ZH vorzunehmen. In anderen Fällen wird die Qualifikation der Daten nach den üblichen Gesichtspunkten beurteilt, namentlich anhand der Bestimmbarkeit von Personen aus den generierten Daten sowie des hierzu benötigten Aufwandes für die Datenbearbeiterin.[119]

IV. Rechtliche Regelungen im Kanton Zürich

A. Regelungsansätze im IDG ZH und die damit zusammenhängenden Fragen

Das IDG ZH enthält mit den Bestimmungen zu der automatisierten Auswertung von personenbezogenen Informationen bzw. Profiling[120] nur einen einzigen Normenkomplex, der ausdrücklich den Einsatz von KI-Technologie regelt. Indes sind die übrigen Bestimmungen selbstverständlich auf KI-Technologien anwendbar. Die folgende Aufstellung zeigt die wichtigsten Grundsätze und inwiefern deren Umsetzung im Rahmen der Anwendung von KI eine Herausforderung darstellen kann.

[119] Siehe dazu Braun Binder et al. (FN 8), 42.

[120] Siehe § 3 Abs. 4 lit. c IDG ZH.

1. Die klassischen Grundsätze der Datenbearbeitung

a. Gesetzmässigkeit

Bezüglich der Gesetzmässigkeit wird in der Lehre die Position vertreten, dass der Einsatz von KI-Technologien im Hinblick auf das Legalitätsprinzip «keine besondere Herausforderung» darstellt.[121] Dem ist insofern zuzustimmen, als die Gültigkeitsvoraussetzungen der genügenden Normstufe und Normdichte auch für den Einsatz von KI-Technologien gelten und grundsätzlich auf diese angewendet werden können.[122] Eine separate gesetzlich Befugnis für den Einsatz von KI-Systemen zur Bearbeitung von Personendaten ist erforderlich, soweit der Einsatz von KI als zusätzlicher Eingriff in die Rechte der Betroffenen, bzw. als eine Bearbeitung von besonderen Personendaten i.S.v. § 3 Abs. 4 IDG ZH zu werten ist.[123]

Sofern aber der Einsatz von KI zur Erfüllung einer gesetzlichen Aufgabe als denknotwendig erscheint, muss das Gesetz nicht die Erlaubnis regeln, sondern den Einsatz in berechtigten Ausnahmefällen gegebenenfalls einschränken. Der Grund liegt darin, dass bei einer genügend engen funktionalen Verknüpfung einer Bearbeitungsmethode mit einer ausdrücklich im Gesetz vorgesehenen Aufgabe erstere als mitgeregelt gilt, also eine implizite Grundlage für die betreffende Bearbeitung vorliegt. Dies gilt grundsätzlich auch für besondere Datenbearbeitungen, wenn auch die Anforderungen an die Rechtsicherheit sehr viel höher sind.[124]

b. Zweckbindung und Verhältnismässigkeit

Die Bearbeitung von Personendaten durch öffentliche Organe darf nur zweckgebunden erfolgen. Öffentliche Organe dürfen sich nicht ohne Weiteres für Private interessieren, sondern nur in Zusammenhang mit der Erfüllung einer ihnen zugewiesenen gesetzlichen Aufgabe. Der Bearbeitungszweck bindet somit die Datenbearbeitung über die staatlichen Aufgaben an das Recht und bildet somit einen Teilgehalt der Gesetzmässigkeit von Datenbearbeitungen.[125]

---

[121] Braun Binder et al. (FN 8), 35.

[122] Siehe dazu Braun Binder et al. (FN 8), 34.

[123] Baeriswyl, PraKom IDG ZH (FN 101), § 8 N 13 ff.

[124] Glass (FN 26), 225 f.

[125] Glass (FN 26), 191 f.

Mithin wird durch die Festlegung des Zwecks einer Datenbearbeitung deren gesetzliche Grundlage im Einzelfall vervollständigt und ist damit ein «unverzichtbarer Bestandteil der Gesetzmässigkeit».[126] Es handelt sich um einen genuin datenschutzrechtlichen Grundsatz, der zudem auch Parallelen zur Verhältnismässigkeit aufweist.[127] Das Zusammenspiel dieser drei Grundsätze (Gesetzmässigkeit, Zweckbindung und Verhältnismässigkeit) läuft darauf hinaus, dass die Erhebung und Speicherung von Daten ohne konkreten Bearbeitungszweck bzw. «auf Vorrat» unverhältnismässig und somit unzulässig ist.[128] Ausnahmen in Bundesgesetzen sind indes gemäss Art. 191 BV «massgeblich» und damit unabhängig von dieser Beurteilung anzuwenden.[129]

Für den Kanton Zürich definiert das Gesetz diese Zweckbindung in § 9 Abs. 1 IDG ZH dahingehend, dass die öffentlichen Organe Personendaten nur zu dem Zweck bearbeiten dürfen, zu dem sie erhoben wurden. Jede Erhebung von Daten wiederum setzt gemäss § 8 IDG ZH einen plausiblen Zusammenhang mit der Erfüllung einer öffentlichen Aufgabe voraus. Die Zweckbindung an den Erhebungsgrund der Daten gilt neben den im Gesetz genannten Arten – Beschaffen, Aufbewahren, Verwenden, Umarbeiten, Vernichten – für sämtliche Bearbeitungsarten über den gesamten Lebenszyklus eines Personendatums.[130]

Nach dem Gesagten liegt der rechtmässige Zweck einer Datenbearbeitung stets in der Erfüllung einer gesetzlichen Aufgabe. Die beiden Voraussetzungen sind untrennbar miteinander verbunden. Somit muss stets klar sein, welche gesetzlich begründete Aufgabe durch die Datenbearbeitung befördert wird. Entsprechend muss der Bearbeitungszweck jeweils vor der Bearbeitung fest-

---

[126] Harb, PraKom IDG ZH (FN 101), § 9 N 1.

[127] Epiney (FN 26), Datenschutzrecht Grundlagen, 539.

[128] Florent Touvenin, Forschung im Spannungsfeld von Big Data und Datenschutzrecht: eine Problemskizze, in: Volker Boehme-Nessler/Manfred Rehbinder (Hrsg.), Big Data: Ende des Datenschutzes? – Gedächtnisschrift für Martin Usteri, Bern 2017, 36 m.w.H.; Bruno Baeriswyl, in: Bruno Bäriswyl/Kurt Pärli (Hrsg.), Datenschutzgesetz, Stämpflis Handkommentar, 1. A. Bern 2015 (zit. SHK DSG-VerfasserIn), Art. 4 N 34; Rosenthal, in: Handkommentar DSG (FN 101), Art. 4 N 20; Gerrit Hornung, Erosion traditioneller Prinzipien des Datenschutzrechts durch Big Data, in: Wolfgang Hoffmann-Riem (Hrsg.), Big Data – Regulative Herausforderungen, Baden-Baden 2018, 85.

[129] So beispielsweise die Vorratsspeicherung auf Grundlage des BÜPF (SR 780.1).

[130] Rudin, PraKom IDG ZH (FN 101), § 3 N 32 ff.

gelegt[131] und auf eine rechtlich begründete Bearbeitungsbefugnis des Organs gestützt werden.[132] Bearbeitungen, die zu einem anderen Zweck erfolgen, sind demzufolge als separate Datenbearbeitungen zu betrachten und erfordern demnach grundsätzlich eine neue rechtliche Grundlage.

Für besondere Personendaten i.S.v. § 3 Abs. 4 IDG ZH müssen gemäss § 8 Abs. 2 IDG ZH nebst dem Zweck auch die Art und Weise der Bearbeitung selbst gesetzlich vorgesehen sein, wobei diese sich aus der Umschreibung der Aufgabe notwendigerweise ergeben kann.[133] Die Anforderungen sowohl an die Erlassstufe (Gesetz, Verordnung, interne Richtlinie etc.), als auch an die Bestimmtheit der rechtlichen Bearbeitungsgrundlage im Einzelfall, variieren dabei je nach Risiko für die Betroffenen.[134]

Eine Zweckänderung für bereits vorhandene Personendaten ist gemäss § 9 Abs. 1 IDG ZH nur möglich, wenn das Gesetz dies ausdrücklich ermöglicht oder die betroffene Person einwilligt. Zudem kann das öffentliche Organ die bei ihm vorhandenen Personendaten im Rahmen der rechtmässigen Bekanntgabe an Dritte einem weiteren Zweck zuführen. Ebenso ist eine Bearbeitung zu nicht personenbezogenen Zwecken möglich. Einer drohenden «Aushöhlung»[135] des Zweckbindungsprinzips kann über die Informationspflicht begegnet werden, indem von den Datenempfängern bzw. Umsetzungsbefugten des neuen Zwecks als Beschaffer der Daten eine erneute Information gemäss § 12 IDG ZH verlangt wird, in der die Betroffenen auf die Zweckänderung hingewiesen werden.

c. Transparenz, Erkennbarkeit sowie das Handeln nach Treu und Glauben

Die Transparenz des Handelns öffentlicher Organe bildet neben dem Schutz der Grundrechte den zweiten in § 1 Abs. 2 IDG ZH ausdrücklich genannten Zweck des Gesetzes. Sie birgt verschiedene Aspekte, namentlich eine «aktive» und eine «passive» Transparenz der Verwaltung gegenüber der Allgemeinheit,

---

131 Harb, PraKom IDG ZH (FN 101), § 9 N 1.

132 Glass (FN 26), 92.

133 Harb, PraKom IDG ZH (FN 101), § 9 N 3 f.

134 Baeriswyl, PraKom IDG ZH (FN 101), § 8 N 14.

135 Harb, PraKom IDG ZH (FN 101), § 9 N 12.

die in § 4 IDG ZH verankert ist,[136] sowie Transparenz gegenüber den von einer Datenbearbeitung betroffenen Personen.[137] Aus letzterer folgt der in § 12 IDG ZH konkretisierte Grundsatz der *Erkennbarkeit* der Datenbeschaffung, welche die Betroffenen in die Lage versetzen soll, die Rechtmässigkeit der Bearbeitung zu beurteilen und gegebenenfalls dagegen vorzugehen.[138] Es handelt sich mithin um eine Ausprägung des Grundsatzes von Treu und Glauben.[139]

Obwohl § 1 Abs. 2 Bst. a IDG ZH primär das in Art. 17 und 49 KV verankerte Öffentlichkeitsprinzip umschreibt,[140] gilt dieses als Ordnungsprinzip[141] der Verwaltung ebenso bei der Umsetzung des Datenschutzrechts. Dies zeigt sich am deutlichsten in den Zugangsrechten zu Informationen und «eigenen» Personendaten gemäss § 20 ff. IDG ZH[142] sowie in den Informationspflichten der öffentlichen Organe in Bezug auf die Bearbeitung von Personendaten gemäss § 12 IDG ZH und in der in § 12a IDG ZH statuierten Meldepflicht für gewisse Datenschutzverletzungen. Da diese Pflichten ebenfalls spezifisch datenrechtliche Ausprägungen des Grundsatzes von Treu und Glauben sind, gehen sie diesem vor, wobei letzterer ergänzend anzuwenden ist.[143]

Die Folgen von intransparenten Datenbearbeitungen sind im Gesetz angedeutet. Aus § 1 Abs. 1 IDG ZH kann die Vermutung oder auch Befürchtung des Gesetzgebers herausgelesen werden, dass dadurch die freie Meinungsbildung, die Wahrnehmung der demokratischen Rechte sowie die Kontrolle staatlichen Handelns beeinträchtigt werden können. Entsprechend ist das Transparenzprinzip auch kein Selbstzweck, sondern stets in Hinblick auf diese Vermutung

---

136 Baeriswyl, PraKom IDG ZH (FN 101), § 4 N 2.

137 Siehe den Hinweis bei Rudin, PraKom IDG BS (FN 77), § 2 N 17.

138 Harb, PraKom IDG ZH (FN 101), § 12 N 2.

139 Epiney (FN 26), Datenschutzrecht Grundlagen, 544 f.; BSK DSG-Maurer-Lambrou/Steiner (FN 101), Art. 4 DSG N. 16a f.; Rosenthal, in: Handkommentar DSG (FN 101), Art. 4 N 51.

140 Baeriswyl, PraKom IDG ZH (FN 101), § 1 N 4.

141 Baeriswyl, PraKom IDG ZH (FN 101), § 1 N 4.

142 Baeriswyl, PraKom IDG ZH (FN 101), § 1 N 7.

143 Epiney (FN 26), Datenschutzrecht Grundlagen, 545.

zu lesen.[144] Mit anderen Worten kann eine Datenbearbeitung in dem Masse intransparent erfolgen, als der Nachweis gelingt, die gesetzliche Vermutung in Bezug auf die genannten Gefahrenmomente sei unzutreffend.

d. Transparenz als Explainability von KI-Systemen

Die Transparenz von KI-Systemen kann sich sowohl auf den Output des Systems beziehen als auch auf dessen Wirkungsweise oder Hintergrund (Design, Entwicklung, Einbettung in Entscheidungsprozesse).[145] Die Transparenz einer KI-Instanz misst sich an der Möglichkeit, Entscheidungsprozesse sowie einzelne Entscheidungen zu interpretieren und erklären. Verlangt wird also eine Form von *actionable transparency*, welche die Opazität[146] von KI-Systemen sowie das jeweils damit einhergehende Informationsgefälle genügend auszugleichen vermag.[147]

Zusammengefasst wird dies alles unter dem Stichwort der *explainability*.[148] Dieses Konzept einer «nutzbaren Transparenz» umfasst die Interpretierbarkeit sowie die Erklärbarkeit, die sich als allgemein anerkannte Grundsätze der «ethischen» KI-Nutzung herausgebildet haben.[149] Sie kann grundsätzlich *by design* in jede KI als Funktion des Systems integriert werden.[150]

Aus rechtlicher Sicht entscheidend ist die Nachvollziehbarkeit des Outputs eines KI-Systems im Hinblick auf Nachvollziehbarkeit staatlichen Handelns

---

[144] Baeriswyl, PraKom IDG ZH (FN 101), § 1 N 5 ff.; Robert van den Hoven van Genderen, Transparancy Requirements for Algorithms and AI, Wishful Thinking?, in: Jusletter IT vom 27.05.2021, N 32.

[145] Siehe dazu Nicholas Diakopoulos, Transparency, in: Markus D. Dubber/Frank Pasquale/Sunit Das (Hrsg.), The Oxford Handbook of Ethics of AI, Oxford University Press 2020, 199 f.

[146] Zur Begriffsbildung Hoffmann-Riem (FN 5), 41.

[147] Emre Bayamlioglu, Contesting Automated Decisions, in: EDPL 4/2018, 438 f.; oder auch *usable transparency*, Diakopoulos (FN 145), 204.

[148] Ella Hafermalz/Marlen Huysman, Please Explain: Key Questions for Explainable AI Research from an Organizational Perspective, Morals + Machines 2/2021, 15.

[149] Bernahrd Waltl/Roland Vogl, Explainable Artificial Intelligence – the New Frontier in Legal Informatics, in: Erich Schweighofer/Franz Kummer/Ahti Saarenpää/Burkhard Schafer (Hrsg.), Datenschutz/Legal Tech, Tagungsband des 21. Internationalen Rechtsinformatik Symposions IRIS 2018, 117 f.

[150] Bryson (FN 61), 8.

sowie die Möglichkeit der Anfechtung der rechtlichen und sachlichen Begründung eines darauf gestützten Entscheids. Das Prinzip der Transparenz sowie die damit einhergehenden Erfordernisse der Interpretier- und Erklärbarkeit, sind somit eng mit der Begründungspflicht für rechtliche Entscheidungen verknüpft.[151]

Technisch betrachtet hängen Transparenz, Interpretier- und Erklärbarkeit von der Nachvollziehbarkeit der internen mathematischen Struktur des Algorithmus ab. Die technisch bedingte Veranlagung zu Intransparenz unterscheidet sich je nach Modell.[152] Lineare Modelle oder Entscheidungsbäume (*decision trees*) funktionieren beispielsweise nach zusammenhängenden Regeln und sind daher sowohl im Hinblick auf ihre Wirkungsweise als auch auf die Nachvollziehbarkeit einer Entscheidung im Einzelfall nachvollziehbar.[153]

Als zweites Problem entpuppt sich die Bandbreite an unterschiedlichen denkbaren Adressaten der Transparenz und deren Verständnishorizont, so etwa Laien, zur Entscheidung befugte Fachpersonen oder KI-Ingenieure.[154] Denn je komplexer die zugrundeliegenden mathematischen Formeln einer KI-Instanz sind, desto weniger ist diese für Laien nachvollziehbar und wird entsprechend für sie intransparent.[155] Soweit indes eine Überprüfung sowie eine darauf aufbauende nützliche Erklärung durch Fachpersonen möglich bleibt, kann die Transparenz grundsätzlich als gewahrt gelten. Bei steigender Komplexität wird indes möglicherweise ein Punkt erreicht, an dem Entscheidungen auch von Fachleuten nicht mehr in rechtsgenüglicher Weise nachvollzogen werden können. Hier ist die Rechtmässigkeit einer auf KI-Output basierten

---

[151] Zum Ganzen Paul Vogel, Künstliche Intelligenz und Datenschutz, – Vereinbarkeit intransparenter Systeme mit geltendem Datenschutzrecht und potentielle Regulierungsansätze, zugl. Diss. Univ. Würzburg 2021, Baden-Baden 2022, 199 f.

[152] Für eine Übersicht verschiedener Methoden siehe Waltl/Vogl (FN 149), 119, Tabelle 1.

[153] Waltl/Vogl (FN 149), 119; für automatisch generierte *trees* siehe Badr Hssina/Abdelkarim Merbouha/Hanane Ezzikouri/Mohammed Erritali, A comparative study of decision tree ID3 and C4.5, International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Advances in Vehicular Ad Hoc Networking and Applications 2014, http://dx.doi.org/10.14569/SpecialIssue.2014.040203 (Abruf 01.06.2022), 13 ff.

[154] Rolf H. Weber, Künstliche Intelligenz: Regulatorische Überlegungen zum «Wie» und «Was», in: EuZ 1/2022, B12.

[155] van den Hoven van Genderen (FN 144), N 8 f.; Hafermalz/Huysman (FN 148), 17 ff.

Entscheidung wohl nur anzunehmen, wenn diese auf andere Weise plausibilisiert werden kann. Dies kann beispielsweise durch entsprechende statistische Studien erfolgen.[156] Je nach Qualität der behaupteten Rechtsverletzung ist es zudem denkbar, dass der Nachweis genügt, welches Gewicht im Rahmen der Entscheidung welchem Input zugemessen wurde bzw. welche Faktoren für das Ergebnis von Bedeutung waren.[157]

Soweit ein Modell eine konstruktionsbedingte Intransparenz aufweist, muss diese im Rahmen der Festlegung der Parameter eines geplanten KI-Modells berücksichtigt und im Hinblick auf die Vorgaben des Datenschutzes mit technischen Mitteln dahingehend umgeformt werden, dass die daraus generierten Erklärungen für die jeweiligen Adressaten nützlich sind, um ihre Rechte wahrzunehmen.

2. Von der Qualität zur Qualitätssicherung der Datenbearbeitung

Insgesamt ist den klassischen Grundsätzen der Datenbearbeitung anzumerken, dass sie aus einer Zeit stammen, in der künstliche Intelligenz eine Wissenschaftsdisziplin ohne merkliche Auswirkungen auf die Datenbearbeitungspraxis der Verwaltung war, und Datenbanken noch eher in der Form von Karteikästen als von Computerservern daherkamen. Erkennbar ist dies an dem Umstand, dass keiner der Grundsätze ausdrücklich auf die Minimierung der mit den spezifischen Eigenschaften von KI-Technologien verbundenen Risiken gerichtet ist. Dies erstaunt nicht, wenn man bedenkt, dass das weltweit erste Datenschutzgesetz aus dem Jahr 1970 stammt,[158] und in der Schweiz das Datenschutzgesetz des Bundes am 1. Juli 1993 in Kraft trat.[159] Seither entstanden neue Datenschutzgesetze in den Kantonen, von denen die meisten zwi-

---

156 Siehe I.C.

157 Siehe dazu Christen et al. (FN 10), 136.

158 Das Hessische Datenschutzgesetz von 1970; eine historische Kurzdarstellung findet man auf der Seite des Hessischen Beauftragten für Datenschutz und Informationsfreiheit unter https://datenschutz.hessen.de/ueber-uns/geschichte-des-datenschutzes (Abruf 07.02.2022).

159 AS (1993) 1945, 1958.

schenzeitlich revidiert wurden.[160] Auch das DSG des Bundes wurde mehrfach revidiert, die neuste Version tritt voraussichtlich auf den 1. September 2023 in Kraft.[161] Die neuen Gesetzesbestimmungen tragen den KI-Technologien vermehrt Rechnung.

Gemeinsam ist den klassischen Bearbeitungsgrundsätzen, dass sie sich auf die Art und Weise bzw. auf die *Qualität der Datenbearbeitung* beziehen. Dies im Gegensatz zu den weiter unten thematisierten «neuen» Grundsätzen der Datenbearbeitung, die mit Einzug von Computern und nun auch der Möglichkeit des Einsatzes von KI-Systemen in die Datenschutzgesetze aufgenommen wurden. Diese betreffen nicht die Qualität der Datenbearbeitung als solche, sondern schreiben den öffentlichen Organen vielmehr gewisse Massnahmen des Risikomanagements vor,[162] mithin der *Qualitätssicherung der Datenbearbeitung*. Darunter fallen die Pflicht zur Durchführung von Datenschutzfolgeabschätzungen bzw. zur Einrichtung eines laufenden Risikomonitorings,[163] eine damit verbundene Pflicht der Vorlage zur Vorabkontrolle von riskanten Datenbearbeitungen an das Aufsichtsorgan sowie eine Meldepflicht in Bezug auf festgestellte Datenschutzverletzungen.

Ergänzt werden diese Massnahmen durch die Schutzziele der Informationssicherheit, die auch für Informationen aus Sachdaten gelten. Diesbezüglich enthält das Gesetz in § 7 Abs. 2 IDG ZH die Pflicht der öffentlichen Organe, Massnahmen zu ergreifen, welche gewisse Schutzziele im Hinblick auf die durch sie bearbeiteten Informationen sicherstellen. Als Schutzziele nennen § 7 Abs. 2 Bst. a-e IDG ZH die Verhinderung unrechtmässiger Kenntnisnahme, die Richtigkeit, Vollständigkeit und Verfügbarkeit der Information, die Zurechenbarkeit der Bearbeitung zu bestimmten Personen sowie die Erkennbarkeit und Nachvollziehbarkeit von Änderungen. Diese Pflichten gelten denknotwendig auch für die den Informationen zugrundeliegenden Daten, d.h. auch

---

[160] So auch das IDG des Kantons Zürich, dessen jüngste Teilrevision im Juni 2020 in Kraft trat.

[161] Hinweis auf https://www.bj.admin.ch/bj/de/home/staat/gesetzgebung/datenschutzsta erkung.html (Abruf 06.09.2022).

[162] Vgl. Baeriswyl, PraKom IDG BS (FN 77), § 8 N 3 ff.

[163] Dazu Philip Glass, Gedanken zur Revision des DSG, datalaw.ch, 23.01.2018, https://www.datalaw.ch/gedanken-zur-revision-des-dsg/ (Abruf 01.06.2022) N 13.

für Personendaten. Entsprechend gelten die Schutzziele der Informationssicherheit als Bearbeitungsgrundsätze für Personendaten.[164]

3. Insbesondere Datenrichtigkeit

a. Richtigkeit als Voraussetzung der rechtmässigen Bearbeitung

Die Richtigkeit der von den öffentlichen Organen als Entscheidungsgrundlagen herangezogenen Daten ist in § 7 Abs. 2 Bst. b IDG ZH vorgeschrieben und bildet ein zentrales Schutzziel für den Umgang mit Informationen und Daten, insbesondere auch im Hinblick auf künftige KI-Anwendungen.[165] Die Bearbeitung von unrichtigen Daten im Sinne von § 7 Abs. 2 Bst. b IDG ZH bzw. die «unrichtige Datenbearbeitung» stellt eine Persönlichkeitsverletzung dar.[166]

Richtigkeit von Personendaten bedeutet zunächst, dass diese keine Falschaussagen über die Betroffenen enthalten, d.h. keine Informationen abbilden dürfen, die nicht den objektiv feststellbaren Tatsachen entsprechen. In dieser Hinsicht ist der gesetzliche Anspruch an die Richtigkeit der bearbeiteten Daten absolut. Dabei ist stets im konkreten Zusammenhang zu ermitteln, worauf sich der Anspruch der Richtigkeit bezieht, was in der Lehre als relative Richtigkeit bezeichnet wird.[167] So muss die Voraussetzung beispielsweise im Zeitpunkt der Bearbeitung erfüllt sein,[168] wovon sich das öffentliche Organ zu vergewissern hat[169] und die Daten gegebenenfalls nachführen muss.

Soweit Daten subjektiv festgestellte Tatsachen oder Werturteile enthalten, etwa in polizeilichen Protokollen, ist die Richtigkeit der subjektiven Bewertung der in den Daten abgebildeten Umstände nicht objektiv feststellbar. Auch ist dies nicht erwünscht, da sich der Richtigkeitsanspruch auf die Dokumentation der amtlichen Beobachtungen bezieht. Mit anderen Worten geht es nicht

---

[164] So auch Baeriswyl, PraKom IDG ZH (FN 101), § 7 N 2.

[165] Braun Binder et al. (FN 8), 46.

[166] SHK DSG-Baeriswyl/Blonski (FN 128), Art. 5 N 5.

[167] Rudin, PraKom IDG BS (FN 77), § 11 N 3; SHK DSG-Baeriswyl/Blonski (FN 128), Art. 5 N 5, welche die Richtigkeit der Daten insgesamt als relativen Begriff bezeichnen; ebenso Rosenthal/Jöhri, in: Handkommentar DSG (FN 101), Art. 5 N 2.

[168] Botschaft vom 23. März 1988 zum Bundesgesetz über den Datenschutz (DSG), BBl 1988 II 413, 450.

[169] Sog. «Vergewisserungspflicht»; Waldmann/Oeschger (FN 26), Datenschutzrecht Grundlagen, 815 f. m.w.H.

um die Richtigkeit von objektiv feststellbaren Angaben zu den betroffenen Personen, sondern um die Authentizität des Protokolls. Ausschlaggebend ist, ob die vorhandenen Daten die subjektive Beobachtung der ermächtigten Person zum Zeitpunkt der Vornahme der Bewertung wiedergeben.[170]

Um zu vermeiden, dass rechtmässig bearbeitete «subjektive Personendaten» nachträglich geändert und so verfälscht werden, sieht das Gesetz im Rahmen der Rechtsbehelfe in § 21 IDG ZH vor, dass in diesen Fällen ein Bestreitungsvermerk angebracht und die Bearbeitung der Daten gegebenenfalls eingeschränkt wird.

b. Richtigkeit der Outputdaten

Aufgrund der Funktionsweise von KI-Technologien kann es Schwierigkeiten bereiten oder gar unmöglich sein, eine mangelhafte Qualität der Outputdaten im Einzelfall nachzuweisen. Dies betrifft insbesondere auch Personendaten, die durch elektronisches Profiling generiert werden. Ergebnisse von wahrscheinlichkeitsbasierten Algorithmen können unter Umständen ebenso wenig objektiv überprüft werden, wie subjektive Tatsachen und Werturteile. Das Problem besteht darin, dass nicht ohne Weiteres auf eine intrinsische kausale Begründung zugegriffen werden kann, durch welche das Ergebnis objektiv nachprüfbar würde. Insofern besteht auch hier eine gewisse «Subjektivität der Bearbeitungsperspektive».

Im Gegensatz zu subjektiven Tatsachen und Werturteilen von Menschen, basieren die Ergebnisse eines KI-Systems indes auf mathematischen Formeln; aufgrund der Ergebnisse der Validierung besteht zumindest die Möglichkeit, eine statistische Wahrscheinlichkeit dafür anzugeben, dass die im Ergebnis abgebildeten Zusammenhänge gültig auf gewisse Tatsachen schliessen lassen. Ab welchem Grad an Wahrscheinlichkeit ein Datum als «richtig» im Sinne des Gesetzes gilt, ist indes unklar.[171]

---

[170] Waldmann/Oeschger (FN 26), Datenschutzrecht Grundlagen, 814 f.; Rosenthal/Jöhri, in: Handkommentar DSG (FN 101), Art. 5 N 2.

[171] Vgl. dazu David Vasella, Zur Freiwilligkeit und zur Ausdrücklichkeit der Einwilligung im Datenschutzrecht, in: Jusletter vom 16.11.2015, N 10, der im Hinblick auf das Ergebnis im Schlussbericht des EDÖB i.S. PostFinance darauf hinweist, dass «Wahrscheinlichkeitsaussagen wie beispielsweise ein Rating i.d.R. als Werturteil beurteilt [werden], das nicht falsch, sondern höchstens unvertretbar sein kann»; zur Validierung siehe I.B.2.

Da die Richtigkeit eine gesetzliche Voraussetzung für die Bearbeitung von Personendaten durch öffentliche Organe ist, muss die Frage der Rechtmässigkeit des durch die negative und positive Fehlerquote des KI-Systems begründeten Risikos jeweils für die betreffende Bearbeitung im Rahmen einer Folgenabschätzung geklärt werden.

c. Spezialfall: Richtigkeit der Trainingsdaten

Die Verwendung von künstlicher Intelligenz in der Form von trainierten Modellen setzt die Nutzung Trainingsdaten voraus, die ihrerseits Personendaten sein können. Obwohl diese Daten dem Zweck dienen, statistische Modelle zu berechnen – und daher nicht der Output von Personendaten bezweckt wird – spielt die Datenrichtigkeit dennoch eine wichtige Rolle im Hinblick auf die Schutzziele des Datenschutzrechts. Dies ist insbesondere der Fall, wenn falsche Erhebung oder falsche Auswahl der Trainingsdaten zu einem rechtlich relevanten Bias im trainierten Modell führen.[172]

4. Die neuen Grundsätze der Datenbearbeitung

a. Vorabkontrolle und Datenschutz-Folgenabschätzung

Das Datenschutzgesetz des Kantons Zürich enthält seit geraumer Zeit eine Pflicht der staatlichen Organe, gewisse Datenbearbeitungen der Datenschutzbeauftragten zur Vorabkontrolle vorzulegen. Mit Inkrafttreten einer Änderung des IDG ZH am 1. Juni 2020 kommt die Pflicht zur Durchführung einer Datenschutz-Folgenabschätzung hinzu.[173]

Die Datenschutz-Folgenabschätzung verpflichtet die öffentlichen Organe von Kanton und Gemeinden gemäss § 10 Abs. 1 IDG ZH dazu «bei einer beabsichtigten Bearbeitung von Personendaten deren Risiken für die Grundrechte der betroffenen Personen [zu bewerten]». Dies bedeutet zunächst, dass öffentlichen Organe grundsätzlich bei der Bearbeitung von Personendaten die damit verbundenen Persönlichkeitsrisiken im Auge behalten müssen.

---

[172] Zur Bearbeitung von Personendaten als Trainingsdaten siehe III.A.1.; Zur Problematik von Bias siehe I.B.3.c.

[173] Gesetz über die Information und den Datenschutz (IDG) (Änderung) vom 25. November 2019 (OS 75, 263; ABl 2018-07-13).

Je nach Ergebnis, bzw. ermitteltem Risikoprofil einer Datenbearbeitung besteht sodann die Pflicht, die betreffende Datenbearbeitung der kantonalen Datenschutzbehörde zur Vorabkontrolle vorzulegen. Dies ist gemäss § 10 Abs. 2 IDG ZH zwingend vorgeschrieben, wenn eine beabsichtigte Datenbearbeitung «mit besonderen Risiken für die Grundrechte der betroffenen Personen» verbunden ist. Gemäss § 24 Abs. 1 Bst. a-e IDV ZH[174] erfüllt eine Datenbearbeitung diese Voraussetzung «insbesondere» dann, wenn sie ein Abrufverfahren vorsieht, die Sammlung einer Vielzahl besonderer Personendaten betrifft, mit dem Einsatz neuer Technologien verbunden ist, wenn sie vorsieht, dass mindestens drei verschiedene öffentliche Organe gemeinsam Personendaten bearbeiten, oder wenn sie eine grosse Anzahl von Personen betrifft. Als typische Fälle eines besonderen Risikos gelten zudem auch die automatisierte Einzelentscheidung, die systematische Überwachung von Personen, die Bearbeitung von besonderen Personendaten im Allgemeinen, die Bearbeitung von Personendaten in grossem Umfang (insb. hohe Anzahl an Betroffenen bzw. grosse Datenmengen), das Zusammenführen/Kombinieren von Personendaten aus unterschiedlichen Prozessen, der Einsatz neuer Technologien oder biometrischer Verfahren, sowie Scoring bzw. Profiling.[175]

Von diesen gesetzlich typisierten Fällen der besonderen Risiken basieren die automatisierte Einzelentscheidung sowie das automatisierte Profiling i.S.v. § 3 Abs. 4 Bst. c IDG ZH notwendigerweise auf Technologien der künstlichen Intelligenz, während in anderen Fällen die Nutzung solcher Technologien naheliegend oder zumindest denkbar erscheint. Sowohl die systematische Überwachung, die Bearbeitung von besonderen Personendaten, die Bearbeitung in grossem Umfang (durch Excel-Tabellen und klassische Datenbanken aber auch Big Data-Anwendungen), die Zusammenführung aus verschiedenen Bereichen (z.B. durch elektronische Zugriffsberechtigungen auf verschiedene Datenbanken) und die biometrischen Verfahren[176] (z.B. Auswertung von DNA-Spuren) müssen nicht notwendigerweise, können aber durch KI-Technologien unterstützt werden. Zudem wird die Verwendung von KI-Technologien auf

[174] Verordnung vom 28. Mai 2008) über die Information und den Datenschutz des Kanton Zürich (IDV ZH; ON 170.41)

[175] Datenschutzbeauftragte des Kantons Zürich, Merkblatt Datenschutz-Folgenabschätzung DSFA, V.1.1. November 2020, Abschnitt 2

[176] Siehe VII.A.2.

absehbare Zeit als «Einsatz neuer Technologien» i.S.v. § 24 Abs. 1 Bst. c IDV ZH zu qualifizieren sein.[177]

Hier stellt sich die Frage, ob und gegebenenfalls wann solche Technologien zu einem späteren Zeitpunkt nicht mehr als «neu» gelten, und ihre Verwendung somit nicht mehr einer automatischen Vorlagepflicht unterliegt. In diesem Zusammenhang gilt es zum einen, zu bestimmen, was «neu» bedeutetet. Zum anderen, worauf sich die Qualität des «neu-seins» bezieht.

b. Der Einsatz von «neuen Technologien» i.S.v. § 24 Abs. 1 Bst. c IDV ZH

Dem Wortlaut des Gesetzes nach bezieht sich die Qualität des neu-seins auf eine Technologie, also eine Art und Weise oder auch Methode, wie Personendaten bearbeitet werden. Die Neuheit tritt in gewissen Situationen als zusätzlicher, nicht durch die Technologie selbst erzeugter Risikofaktor hinzu. Aus diesem Blickwinkel bezieht sich die Qualifizierung als «neu» nicht darauf, ob die Technologie erst kürzlich entwickelt wurde, sondern ob der Kontext ihres Einsatzes zur Datenbearbeitung – und damit auch deren Risikoprofil – neu sind. Ein neuer Kontext der Bearbeitung ist in diesem Zusammenhang anzunehmen, wenn weder das einsetzende öffentliche Organ noch die zuständige Datenschutzbehörde Erfahrungen mit dieser Technologie haben, und daher auch nicht *prima vista* über das damit zusammenhängende Risiko befinden können. Mithin muss nicht die Technologie, sondern deren Risikoprofil im konkreten Datenbearbeitungskontext für die betreffenden Stellen «neu» sein, um eine Vorlagepflicht auszulösen.

Im Ergebnis ist somit – ähnlich wie im Falle von besonderen Personendaten – die Tatsache, dass eine Technologie neu ist, lediglich ein beispielhaftes Indiz für ein ebenso neues Risikoprofil. Dies bedeutet, dass bereits bekannte Technologien, die in einem neuen Kontext eingesetzt werden, und dort ein qualitativ neues Risiko für die Grundrechte der Betroffenen begründen, grundsätzlich als «neu» gelten und daher vorlagepflichtig werden können. Ausschlaggebend kann nur sein, dass weder die betreffende Behörde noch die zuständige Datenschutzstelle sich zuvor mit der neuen Risikokonstellation auseinandergesetzt

[177] Datenschutzbeauftragte des Kantons Zürich, In der Krise ist nicht alles anders – Tätigkeitsbericht 2020, 33.

haben. Handkehrum gilt eine Technologie nicht mehr als neu im Sinne der Vorlagepflicht, wenn für ihre Verwendung in einem spezifischen Kontext genügend Erfahrungswerte in Bezug auf das eigentliche Risikoprofil vorliegen, um einschätzen zu können, ob eine Vorlagepflicht besteht oder nicht.

Auf der anderen Seite kann dies in seltenen Fällen dazu führen, dass Datenbearbeitungen, die unter Verwendung einer allgemein als «neu» empfundenen Technologie vorgenommen werden, nicht notwendigerweise als «neu» i.S. von § 24 Abs. 1 Bst. c IDV ZH gelten. In Bezug auf KI-Technologien dürfte die Vorlagepflicht regelmässig trotzdem gegeben sein, da solche Technologien oftmals auch ohne «neu» zu sein eine der Voraussetzungen erfüllen dürften.[178]

c. Ähnliche Risikostruktur bei voll- und teilautomatisierten Einzelentscheidung

Die Frage, ob ein KI-unterstützter Prozess voll- oder teilautomatisiert erfolgen soll, ist zunächst eine Frage des *designs*[179] eines Mensch-Maschine-Entscheidungsprozesses. Die möglichen Herausforderungen und Risiken für öffentliche Organe ergeben sich somit aus der Natur von KI-Systemen einerseits, sowie andererseits aus den Eigenheiten der Menschen im Allgemeinen sowie der Verwaltung im Speziellen.[180] Zunächst erscheint die Unterscheidung einfach, da im ersten Fall der Output der KI die Entscheidung und im zweiten Fall einen Vorschlag für eine Entscheidung durch den zuständigen Menschen darstellt. Allerdings ist zu bedenken, dass auch vollautomatisierte Einzelfallentscheidungen in einen Verwaltungsprozess eingebunden sein müssen, nicht zuletzt, um die Wahrung der Rechte der Betroffenen sicherzustellen bzw. den Rechtsweg zu öffnen. Für den Verwaltungsprozess läuft die Entscheidung für Voll- oder Teilautomatisierung – neben der technischen Machbarkeit – auf die Frage hinaus, wer eine solche Entscheidung als erstes beurteilen wird: eine Fachperson, welche direkt für die fragliche Materie zuständig ist, oder eine Instanz des Rechtsmittelwegs.

Ungeachtet dessen, wer entscheidet, wird sich das Problem des *automation bias* stellen. Es handelt sich hierbei um eine mentale Verzerrung in der Kritik-

---

[178] Siehe IV.A.4.a.

[179] Zu den *by design*-Prinzipien siehe V.D.

[180] Braun Binder et al. (FN 8), 6.

fähigkeit von Menschen zugunsten von automatisierten Vorgängen, insbesondere von Computern und deren Output. Als Hauptursachen gelten die durch vermeintliche Neutralität und vermittelte Autorität von Computern sowie die Tatsache, dass es kognitiv einfacher ist, eine Aufgabe an die Automation zu delegieren.[181] Aufgrund der Beobachtung, dass auch Fachpersonen wie beispielsweise Piloten, Nuklearingenieure oder Fachpersonen in der Intensivpflege durchaus für solche Fehlurteile anfällig sind,[182] wird eine pauschale Aussage darüber, wer bei einer ersten Plausibilitätsprüfung durch den Menschen (Fachstelle oder Rechtsmittelinstanz) bessere Chancen hat, nicht einer solchen Voreingenommenheit zu erliegen, schwierig zu treffen sein. Aus allgemeinen verwaltungsrechtlichen Überlegungen erscheint es indes sinnvoll, für automatisierte Entscheidungen nicht devolutive Rechtsmittel vorzusehen, etwa durch Einsprachemöglichkeit an die verfügende Instanz oder durch Einwendungsverfahren.[183]

5. Die Meldepflicht gemäss § 12a IDG ZH

Das IDG ZH enthält seit Sommer 2020 eine Meldepflicht der öffentlichen Organe für qualifizierte Datenschutzverletzungen. Das Gesetz schreibt neu in § 12a IDG ZH vor, dass öffentliche Organe «unverzüglich die unbefugte Bearbeitung oder den Verlust von Personendaten» dem oder der Datenschutzbeauftragten melden, «wenn die Grundrechte der betroffenen Person gefährdet sind».[184] Dies dürfte der Fall sein, wenn Bearbeitungszusammenhänge betroffen sind, die besondere Personendaten i.S.v. § 3 Abs. 4 IDG ZH darstellen, da diese definitionsgemäss eine «besondere Gefahr einer Persönlichkeitsverletzung» bergen. Im Übrigen ist unklar, wie diese beiden gesetzlichen Massstäbe

---

[181] Kathleen L. Mosier/Linda J. Skitka, Human Decision Makers and Automated Decision Aids: Made for Each Other?, in: Raja Parasuraman/Mustapha Mouloua (Hrsg.), Automation and Human Performance: Theory and Applications. NJ: Erlbaum 1996/ CRC Press 2009, 201–220, 206; Siehe auch Matthias Van Der Haegen, Quantative Legal Prediction: the Future of Dispute Resolution, in: Jan De Bruyne/Cedric Vanleenhove (Hrsg.), Artificial Intelligence and the Law, Cambridge Antwerp Chicago 2021, 85 ff.

[182] Mosier/Skitka (FN 181), 201.

[183] Dazu allgemein Ulrich Häfelin/Georg Müller/Felix Uhlmann, Allgemeines Verwaltungsrecht, 8. Auflage, Zürich 2020, N 1194 ff.

[184] Eingefügt durch das Gesetz vom 25. November 2019 (OS 75, 263; ABl 2018-07-13). In Kraft seit 1. Juni 2020.

systematisch zueinanderstehen, da die Unterscheidung zwischen einer Gefährdung von Grundrechten und einer besonderen Gefahr für die Persönlichkeit in der Praxis kaum sinnvoll vorzunehmen sein wird.

Schliesslich ist daran zu erinnern, dass die Bearbeitung von Personendaten durch KI-Technologien auf absehbare Zeit *prima vista* als «neue Technologien» im Sinne des IDG ZH gelten und daher aufgrund eines immanenten besonderen Risikos für die Grundrechte zur Vorabkontrolle vorgelegt werden müssen.[185] Parallel dazu dürften der Verlust oder die unrechtmässige Bearbeitung von Personendaten im Rahmen der Nutzung von KI-Systemen regelmässig auch meldepflichtig i.S.v. § 12a IDG ZH sein.

B. Schweiz

Im schweizerischen Recht sind erst wenige Normen auszumachen, die ausdrücklich auf die Regulierung von KI-Technologien ausgerichtet sind.[186] Auch wenn noch eine gewisse Unklarheit darüber auszumachen ist, wie eine allfällige Regulierung angegangen werden soll, so wurden von verschiedenen Seiten bereits einige Prinzipien ausgearbeitet, welche die Erschaffung eines rechtlichen Rahmens anleiten sollen.[187]

Eine Ausnahme bildet das Datenschutzrecht. Mit der Revision des neuen DSG werden Bestimmungen in Kraft treten, welche den Einsatz von KI-Technologien betreffen, insbesondere den Einsatz von automatisierten Einzelentscheiden (Art. 21 nDSG) und von automatisiertem Profiling (Art. 5 Bst. f u. g nDSG). Letzterer Begriff wurde ausdrücklich aus dem europäischen Recht übernommen.[188] Eine weitere Angleichung besteht darin, dass neu der Begriff

---

[185] Siehe IV.A.4.b.

[186] Vgl. die Bestandsaufnahme in Braun Binder et al. (FN 68).

[187] Vgl. Herausforderungen der künstlichen Intelligenz – Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat, SBFI Forschung und Innovation, Dezember 2019, Kapitel 4, https://www.sbfi.admin.ch/dam/sbfi/de/dokumente/2019/12/bericht_idag_ki.pdf.download.pdf/bericht_idag_ki_d.pdf (Abruf wann Januar 2022); Thouvenin et al. (FN 20), 2 f.

[188] Botschaft vom 15. September 2017 zum Bundesgesetz über die Totalrevision des Bundesgesetzes über den Datenschutz und die Änderung weiterer Erlasse zum Datenschutz, BBl 2017 6941 ff. (zit. Botschaft E-DSG), 6971 u. 7021 f.

des automatisierten Profilings mit hohem Risiko geschaffen wurde. Der Tatbestand ist gemäss Art. 5 Bst. g nDSG erfüllt, wenn ein automatisiertes Profiling ein besonderes Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person mit sich bringt. In diesem Bereich geht das Gesetz somit ausdrücklich von einem risikobasierten Regelungsmodell aus.

C. Europa

Der Entwurf für eine Regulierung der künstlichen Intelligenz in der EU sieht primär einen risikobasierten Ansatz vor: Neben der Aufzählung der regulierten Klassen von Technologien schlägt die EU-Kommission vor, die Risiken von KI-Systemen nach Graden einzuteilen und Risikoklassenanalysen vorzuschreiben. Angedacht ist eine Klassifizierung nach minimalem, geringem oder auch begrenztem sowie hohem Risiko, wobei gewisse Praktiken aufgrund eines unzulässigen Risikos verboten wären.[189]

Interessant ist, dass die Kategorien, die als typische Bereiche mit hohem Risiko aufgeführt sind, zum Teil konkreter beschrieben sind als die gesetzlich geschützten Lebensbereiche des Datenschutzrechts. Als Beispiele werden aufgeführt: kritische Infrastrukturen, Schul- und Berufsbildung, Sicherheitskomponenten von Produkten, Beschäftigung, Personalmanagement und Zugang zu selbstständiger Tätigkeit, wichtige private und öffentliche Dienstleistungen, Strafverfolgung, Migration, Asyl und Grenzkontrolle, Rechtspflege und demokratische Prozesse.[190] Die Liste in Anhang III des Entwurfs ist nicht abschliessend und soll von der Kommission bei Bedarf ergänzt werden können.[191]

---

[189] Siehe die Zusammenfassung bei Angela Müller, Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz – mit Auswirkungen auf die Schweiz, EuZ 1/2022, A7 f.; spannend wird nun die parlamentarische Debatte, siehe dazu Luca Bertuzzi, AI regulation filled with thousands of amendments in the European Parliament, https://www.euractiv.com/section/digital/news/ai-regulation-filled-with-thousands-of-amendments-in-the-european-parliament/ (Abruf 15.06.2022).

[190] Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über die künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union vom 22. April 2021, COM(2021) 206 final, Anhang III.

[191] COM(2021) 206 final (FN 190), Art. 7.

Die vorgesehene Risikoklassifizierung für KI ist von besonderem Interesse für das Datenschutzrecht, weil sie dieselbe Funktion erfüllt, wie der Begriff der besonderen Personendaten im IDG ZH: den Schutz der Grundrechte der betroffenen Personen.[192] Aufgrund der mit der Risikoklasse verbundenen Annahme einer hohen Gefährdung von Grundrechten, sollte die Klassifizierung auch bei Risikoanalysen berücksichtigt werden, welche die öffentlichen Organe im Kanton Zürich gemäss § 10 Abs. 1 IDG ZH (Datenschutz-Folgeabschätzung) vornehmen müssen. Aufgrund der gemeinsamen europäischen Grundrechts- und Datenschutzkultur, wie sie insbesondere in der EMRK bzw. im Übereinkommen SEV NR. 108 des Europarates zum Ausdruck kommt, erscheint es naheliegend, dass aus datenschutzrechtlicher Sicht die künstlich-intelligente Bearbeitung von Personendaten mit hohem Risiko im Sinne des Verordnungsentwurfs der EU oftmals zugleich als Bearbeitung von besonderen Personendaten i.S.v. § 3 Abs. 4 IDG ZH zu qualifizieren sein wird.

Schliesslich sieht der Entwurf für KI-Systeme, deren Einsatz nur mit einem geringen oder minimalen Risiko verbunden ist, grundsätzlich keine besonderen Regeln vor. Hierunter werden nach Ansicht des Parlaments und des Rates die meisten KI-Anwendungen fallen, so beispielsweise Videospiele oder Spamfilter. Ausnahmsweise kann für solche Systeme eine Transparenzpflicht gelten, d.h. die Betroffenen müssen darüber informiert werden, dass ein KI-System im Einsatz ist. Gemäss Art. 52 des Entwurfs gilt dies grundsätzlich für alle Systeme, die auf Interaktion mit Nutzern ausgelegt sind, wie beispielsweise Chatbots.[193] Die Transparenzpflicht gilt überdies auch für Systeme, die «zur Erkennung von Emotionen oder zur Assoziierung (gesellschaftlicher) Kategorien anhand biometrischer Daten eingesetzt werden oder Inhalte erzeugen oder manipulieren («Deepfakes»)».[194]

---

[192] COM(2021) 206 final (FN 190), Begründung Ziff. 3.5.

[193] Vgl. die Pressemitteilung der Europäischen Kommission vom 21. April 2021, abrufbar unter https://ec.europa.eu/commission/presscorner/detail/de/ip_21_1682.

[194] COM(2021) 206 final (FN 190), Begründung Ziff. 5.2.4 sowie Titel VI.

V. Herausbildung von «ethischen» Grundsätzen des Einsatzes von KI

A. Metaprinzipien für den Einsatz von KI-Technologien

In den letzten Jahren haben internationale Organisationen, Regierungen sowie private Organisationen verschiedentlich dazu Stellung genommen, wie mit dem Phänomen der künstlichen Intelligenz und insbesondere den hiervon ausgehenden Risiken umzugehen sei. Solche Erklärungen beziehen sich regelmässig auf «ethische Standards» für den Umgang mit KI.[195]

Die Vielzahl von Erklärungen zur KI-Ethik hat zu ersten Metastudien geführt, die Gemeinsamkeiten untersucht und gewisse Metaprinzipien ausgearbeitet haben – und sich in den Ergebnissen (nur) zum Teil überschneiden.[196] Beispielsweise identifizierte eine Studie der ETH Zürich gewisse Metaprinzipien, namentlich Autonomie, Freiheit, Nachhaltigkeit, Abwendung von Schaden, Privatheit, Transparenz, Verantwortung und Würde,[197] während eine spätere Studie der Universität Harvard die Metathemen von Privatheit, Sicherheit, Fairness und Nichtdiskriminierung, Erklärbarkeit und Transparenz, Verantwortung, menschliche Kontrolle, Professionalität und die Förderung von menschlichen Werten herausarbeitet.[198]

---

[195] Eine aktuelle Zusammenstellung findet sich im «AI Ethics Guidelines Global Inventory» von AlgorithmWatch, abrufbar unter https://inventory.algorithmwatch.org (Abruf 13.06.2022).

[196] Anna Jobin/Marcello Ienca/Effy Vayena, Artificial Intelligence: the global landscape of ethics guidelines, Nat. Mach. Intell. (2019); Vgl. auch die Hinweise bei Michal Cichocki, Guidelines für Künstliche Intelligenz (KI): Besteht aus rechtlicher Sicht Handlungsbedarf?, in: Jusletter IT vom 25.02.2021, N 3 f.

[197] Jobin et al. (FN 196), Tabelle 3.

[198] Jessica Fjeld/Nele Achten/Hannah Hilligoss/Adam Nagy/Madhulika Srikumar, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, Berkman Klein Center for Internet & Society, 2020, open access, https://dash.harvard.edu/handle/1/42160420, 66 f. (Abruf 01.06.2022).

B. Einbindung in das Recht durch Verweise

Von Interesse ist an dieser Stelle, dass beide Metastudien die Themenkomplexe Privatheit, Transparenz und Verantwortung als überschneidende Prinzipien in den verschiedenen Erklärungen identifizieren. Hierbei handelt es sich um rechtlich gefestigte Prinzipien, die im Datenschutzrecht von Bedeutung sind. Der Umstand, dass sie Gegenstand eines grossen Teils der globalen Diskussion zu «KI-Ethik» zwischen privaten und staatlichen Akteuren sind, zeugt von einem hohen Stellenwert in der Gesellschaft. Den Ausgleich zwischen verschiedenen, gegenläufigen gesellschaftlichen Werten bzw. wertungsbedürftigen Interessen kann Ethik als Befragungsmethode der Moral indes kaum leisten, da die jeweiligen konkreten sozialen Moralvorstellungen der verschiedenen Kulturen subjektiv und oftmals bezüglich Inhalt sowie Verbindlichkeit bzw. Kompromissbereitschaft sehr unterschiedlich konzipiert sind.[199] Das Recht kann hier seine Funktion wahrnehmen, widerstrebende moralische Ansprüche voreinander zu schützen,[200] indem es entsprechende Wertungsgesichtspunkte in die Rechtsfindung einbindet.

Das Bundesgericht anerkennt beispielsweise die rechtliche Verbindlichkeit «ethischer Leitlinien» von Fachverbänden, wenn und soweit dies rechtlich vorgesehen ist, typischerweise in einer Verweisnorm durch Gesetz oder Verordnung.[201] Klassische Fälle sind überdies ausdrückliche Verweise auf Normenkomplexe sozial-moralischer Wertung, wie etwa in Art. 2 Abs. 1 ZGB (Treu und Glauben), in Art. 19 Abs. 2 OR, Art. 20 Abs. 1 OR und Art. 230 Abs. 1 OR (gute Sitten) oder die bereits erwähnten Schutzziele der Informationssicherheit (Stand der Technik),[202] aber auch implizite Verweise, etwa die

---

[199] Bernd Rüthers/Christian Fischer/Axel Birk, Rechtstheorie und Juristische Methodenlehre, 11. üb. Aufl. München 2020, N 401 ff.; Vgl. dazu auch Karen Hao, Should a self-driving car kill the baby or the grandma? Depends on where you're from, Technology Review 24.10.2018.

[200] Kurt Seelmann/Daniela Demko, Rechtsphilosophie, 7. Auflage München 2019, § 3 N 13; zu Staaten mit «Einheit von Recht und Moral» Bernd Rüthers/Christian Fischer/Axel Birk, Rechtstheorie und Juristische Methodenlehre, N 406.

[201] BGE 136 VI 97 E. 6.2.2, bezüglich ethisch-medizinischer Richtlinien der Schweizerischen Akademie der Medizinischen Wissenschaften SAMW.

[202] Siehe IV.A.2.

Kerngehaltsgarantie in Art. 36 Abs. 4 BV (Menschenwürde)[203]. Insgesamt gilt es zu beachten, dass für die öffentlichen Organe vorrangig die Wertungen des Rechts verbindlich sind; Das Recht gibt grundsätzlich den Rahmen und den Platz für die Anwendung von «ethischen» Wertungen vor. Vorbehalten bleiben nach allgemeinem Verständnis lediglich offensichtliche Fälle von untragbaren Verstössen gegen den «Kern der Gerechtigkeit».[204]

Für öffentliche Organe können demnach «ethische» Anleitungen eine rechtliche Verbindlichkeit erlangen, wenn sie beispielsweise die *best practice* in einer Fachdomäne im Umgang mit moralischen Dilemmata im Berufsalltag wiedergeben, und das Recht auf eine solche verweist. Allerdings erfüllen die eingangs erwähnten Erklärungen zur ethischen Nutzung von KI-Technologie diese Anforderungen regelmässig nicht, da sie zu abstrakt und generisch formuliert sind und kaum Anleitungen zur Auflösung praktischer Probleme enthalten.[205]

C. Indizien für öffentliche Interessen und Auslegungshilfen

Ein weiterer Aspekt der Thematik besteht darin, dass in den entsprechenden Erklärungen jeweils von «ethischen» Prinzipien für die Entwicklung und Nutzung von KI die Rede ist. Damit wird zunächst mitgeteilt, dass es sich um eine moralische Position handeln soll, beispielsweise zur Verwirklichung der Grund- und Menschenrechte in ihrer Funktion als moralische Werte im Recht. Zweitens wird signalisiert, dass es sich nicht um rechtliche Normen handelt, dass also keine rechtlich durchsetzbare Verbindlichkeit erwartet wird. Schliesslich ermöglicht die Berufung auf diese Erklärungen die Teilnahme an

---

[203] Markus Schefer, Die Kerngehalte von Grundrechten – Geltung, Dogmatik, inhaltliche Ausgestaltung, Bern 2001, 83 f. «Ziel grundrechtlicher Kerngehalte bleibt ein absoluter Schutz der Menschenwürde»; OFK BV-Biaggini (FN 49), Art. 36 N 24 ff.; SGK BV-Rainer Schweizer (FN 49), Art. 36 BV, N 44 m.w.H., Menschenwürde als «Auffangkerngehalt»; CR Cst.-Dubey (FN 49), Art. 36 N 126.

[204] Radbruch'sche Formel, vgl. Bernd Rüthers/Christian Fischer/Axel Birk, Rechtstheorie und Juristische Methodenlehre, N 970 f.; Kurt Seelmann/Daniela Demko, Rechtsphilosophie, § 2 N 27.

[205] Hoffmann-Riem (FN 5), 291; Bezüglich privater Compliance David Rosenthal, Datenethik – ein praktischer Zugang aus Sicht der Compliance, in: Recht relevant, Ausgabe 1/2022, 2–4, 3.

der entsprechenden (globalen) Debatte. Die tatsächliche Verbindlichkeit der «ethischen» KI-Prinzipien ist entsprechend unklar.[206] Es handelt sich mithin um eine laufende politische Diskussion,[207] die das Recht unter Umständen sinnvoll ergänzen kann.[208]

Im Anwendungsbereich des öffentlichen Rechts ist es zudem naheliegend, Erklärungen zur «ethischen KI» von Staaten und öffentlichen Organen i.S.v. § 3 IDG ZH als Absichtserklärungen in Bezug auf die Sicherung der Grundrechte und insofern als (mehr oder weniger deutliche) Indizien für den Bestand von entsprechenden öffentlichen Interessen bzw. Schutzpositionen für die Grundrechte Dritter i.S.v. Art. 36 Abs. 2 BV zu werten.

Gewisse «ethische» Grundsätze überschneiden sich denn auch mit den Grundsätzen der Datenbearbeitung, indem sie den Schutz derselben Güter oder Interessen bezwecken.[209] Diese Überschneidung der Schutzfunktion kann für die Beurteilung der betreffenden rechtlichen Grundsätze von Bedeutung sein, etwa in Bezug auf deren Auslegung.[210] Zudem kann die öffentliche Erklärung von «ethische Grundsätzen» als Indiz für das Vorhandensein eines entsprechenden öffentlichen Interesses gewertet werden, wodurch punktuell das öffentliche Interesse an der Durchsetzung der datenschutzrechtlichen Schutzpflichten öffentlicher Organe verstärkt und damit den jeweiligen grundrechtlichen Schutzpositionen ein höheres Gewicht verliehen würde. Aufgrund der bereits genannten subjektiven Qualität von Moral muss indes beachtet werden, dass auch jene Interessen zu berücksichtigen sind, die in den beigezogenen «ethischen» Prinzipien weniger zum Ausdruck kommen.

---

[206] Zum Ganzen Philip Glass, Eine Skizze zur rechtlichen Verbindlichkeit «ethischer» KI-Prinzipien, in: Jusletter IT vom 28.02.2020.; zur Verankerung im deutschen Recht durch das Bundesverfassungsgericht siehe Robert Alexy, Begriff und Geltung des Rechts, erw. Neuausgabe, Freiburg/München 2020, 52 f.

[207] Fjeld et al. (FN 198), 200 ff.

[208] Hoffmann-Riem (FN 5), 291; OECD (FN 13), 85.

[209] Glass (FN 206), N 18 ff.; vgl. auch die Gegenüberstellung (aus der Perspektive von privaten Akteuren) bei Cichocki (FN 196), N 5 ff.

[210] Hoffmann-Riem (FN 5), 291; Glass (FN 206), N 13 ff.; im Ergebnis auch Cichocki (FN 196), N 26.

D. Insbesondere die «Förderung menschlicher Werte»

Einen besonderen Platz nimmt das Prinzip der Förderung menschlicher Werte ein, da es einerseits auf das bereits angesprochene Problem des *value alignment* verweist,[211] und nicht auf klassische verfassungsrechtliche Schutzgüter wie im Falle des Verbots der Diskriminierung oder andere Verletzungen der Menschenwürde «beschränkt» ist. Andererseits entstammt es einer Denktradition und -methode aus den Ursprüngen der Computerwissenschaften, und bildet dort in der Form von *value sensitive design* eine zentrale Komponente guten Maschinendesigns. Wie jede Technologie ist auch KI immer ein Ergebnis von Design.[212] Computertechnologie bzw. computerbasierte Algorithmen als Ergebnis von Design transportieren immer Werte. Sie verleihen dadurch Macht und sind politisch.[213] Wertebezogene Designprinzipien, wie *value sensitive design* bzw. *design for values*[214] oder auch *X by design*[215], folgen der entsprechenden Erkenntnis, dass Computersysteme und andere technologische Artefakte auf die Beachtung von erwünschten moralischen Wertungsgesichtspunkten ausgerichtet werde müssen.[216] Hierzu ist es unabdingbar, im Vorfeld und während sämtlichen Phasen der Entwicklung diese Werte zu identifizieren und gegebenenfalls anzupassen.[217] Werteorientierte Designansätze fordern somit eine bewusste Technikgestaltung zur Beförderung von bestimmten Werten. Für lernende Algorithmen bedeutet dies die Identifikation von möglichen Fehlerquellen im Vorfeld und während der Entwicklung sowie

---

[211] Siehe II.A.

[212] Bryson (FN 61), 6.

[213] Woodrow Hartzog, Privacy's Blueprint – The Battle to Control the Design of New Technologies, Harvard University Press 2018, 51; Martini (FN 29), 48 f.

[214] Für einen Überblick siehe Janet Davis/Lisa P. Nathan, Value Sensitive Design: Applications, Adaptations, and Critiques, in: Jeroen van den Hoven/Pieter E. Vermaas/Ibo van de Poel (Eds.), Handbook of Ethics, Values, and Technological Design – Sources, Theories, Values and Application Domains, Dordrecht 2015, 11–35.

[215] AI High Level Expert Group, Draft Ethics Guidelines for Trustworthy AI, European Commission, April 2019, 21.

[216] Zur historischen Entwicklung Batya Friedman/David G. Hendry, Value Sensitive Design – Shaping Technology with Moral Imagination, The MIT Press, 2019, 11 f.

[217] Friedman/Hendry (FN 216), 29 f.; Lawrence Lessig, Code Version 2.0, New York 2006, 6.

des Einsatzes.[218] In dieser Hinsicht stellen *by design*-Methoden einen Aspekt des Qualitätsmanagements dar.

Das Datenschutzrecht kennt spezifische Formen von Vorgaben für *value sensitive design*, namentlich *privacy by design* sowie das spezifischere *privacy by default*, wobei letzteres als «Prinzip der datenschutzfreundlichen Voreinstellung»[219] für öffentliche Organe nur eine geringe Rolle spielt, da diese in der Regel aufgrund von gesetzlichen Ermächtigungen Daten bearbeiten und nicht aufgrund von Einwilligungen der Betroffenen.[220] Beide sind Ausprägungen des Prinzips der Rechtsdurchsetzung durch Technikgestaltung.[221]

Mit Inkrafttreten des neuen Datenschutzgesetzes des Bundes wird *privacy by design* unter dem Begriff «Datenschutz durch Technik» in Art. 6 Abs. 1 nDSG im schweizerischen Recht auf Bundesebene kodifiziert. Der Bundesrat verspricht sich hiervon eine Symbiose der gegenseitigen Ergänzung zwischen Recht und Technik sowie eine durch das Ethos der technischen Risikominimierung bedingte verringerte Notwendigkeit rechtlicher Technikregulierung; Die Bearbeitungsgrundsätze des Datenschutzes sollen so weit wie möglich auf technischem Weg verwirklicht werden.[222]

Das Informations- und Datenschutzgesetz des Kantons Zürich kennt bereits die eine oder andere ausdrückliche *by design*-Norm. So hält § 4 IDG ZH die

---

[218] Lucia M. Sommer, Personenbezogenes Predictive Policing – Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose, Diss. Univ. Göttingen, Baden-Baden 2020, 105 ff.; in der Praxis beispielsweise durch die Definition von *failure modes*; vgl. dazu eingehend Sophie Stalla-Bourdillon/Alfred Rossi/Gabriela Zanfir-Fortuna, Data Protection by Process – How to Operationalize Data Protection by Design for Machine Learning, Future of Privacy Forum, White Paper V. 1.0, Dezember 2019, abrufbar unter https://iapp.org/resources/article/fpf-how-to-operationalize-data-protection-by-design-for-machine-learning/ (Abruf 24.10.2022), 11 ff.

[219] Vgl. Art. 7 Abs. 3 nDSG, BBl 2020 7639, 7642 f.

[220] Botschaft E-DSG (FN 188), 7030; Glass (FN 26), 235 ff.; Thomas Gächter/Philipp Egli, Informationsaustausch im Umfeld der Sozialhilfe, in: Jusletter vom 06.09.2010, N 55.

[221] Walter Hötzendorfer, Zum Verhältnis von Recht und Technik: Rechtsdurchsetzung durch Technikgestaltung, in: Walter Hötzendorfer/Christof Tschohl/Franz Kummer, International Trends in Legal Informatics – Festschrift for Erich Schweighofer, Bern 2020, 424 f.

[222] Botschaft E-DSG (FN 188), 7029.

öffentlichen Organe an, ihre Informationsverwaltung auf die Ermöglichung von Transparenz und Erkennbarkeit auszurichten, während der mehr oder weniger analoge § 5 IDG ZH auf die Gewährleistung der Nachvollziehbarkeit ausgerichtet ist. Dadurch werden insbesondere das Akteneinsichtsrecht, die Informations- und Datenzugangsrechte sowie das Auskunftsrecht durch technische Zielvorgaben gesichert.[223] Auf ähnliche Weise wirken die Schutzziele der «Informationssicherheit» bzw. Datensicherheit in § 7 Abs. 2 IDG ZH als *design choices* für Informationssysteme förderlich für die Ziele des Datenschutzes.[224]

VI. Spezifische Datenschutzfragen

A. Geltungsbereich des Datenschutzrechts

Die Bestimmung des Geltungsbereichs des IDG ZH durch die Qualität des Personenbezugs der bearbeiteten Daten steht auch ohne den Einsatz von KI-Systemen in der Kritik.[225] Durch den Einsatz von KI-Systemen wird die Rechtslage noch etwas komplizierter.[226] Dies zeigte sich bereits in Zusammenhang mit Diskussionen um *big data*, eine Sammelbezeichnung für die Bearbeitung von sehr grossen Mengen von Daten mittels sehr leistungsfähiger Computersysteme, deren Funktionsweise auf KI-Technologie, genauer: *machine learning*, zurückzuführen ist. Insofern ist die datenschutzrechtliche Diskussion um KI schon seit einigen Jahren im Gange und kann auf die entsprechende Literatur verwiesen werden.[227] Nun aber beginnt sich die Diskussion vermehrt auf die Technologie hinter diesen Anwendungen zu verlagern.

---

[223] Baeriswyl, PraKom IDG ZH (FN 101), § 5 N 6.

[224] Zur Datensicherheit siehe IV.A.2

[225] Hoffmann-Riem (FN 5), 164; David Rosenthal, Personendaten ohne Identifizierbarkeit?, in: digma 2017, 199 f.

[226] Hoffmann-Riem (FN 5), 175.

[227] Eine aktualisierte Problemübersicht findet man bei Astrid Epiney, Big Data und Datenschutzrecht – Gibt es einen gesetzgeberischen Handlungsbedarf?, Jusletter vom 27.04.2020; siehe auch Michal Cichocki, Big Data und Datenschutz: Ausgewählte Aspekte, in: Jusletter IT vom 21.05.2015; René Huber, «Big Data», das kantonale Recht und der Datenschutz, in: Jusletter IT vom 21.05.2015; Rolf H. Weber, Big Data: Sprengkörper des Datenschutzrechts?, in: Jusletter IT vom 11.12.2013; Bruno Baeriswyl, «Big Data» ohne Datenschutz-Leitplanken, in: digma 2013, 14–17.

Im Vordergrund steht damit die Frage, ob Anonymisierung als Begrenzung des Geltungsbereichs des Datenschutzrechts nach wie vor sinnvoll ist.

In Zusammenhang mit Datenschutzfragen beruht die Wirkung von Anonymisierung primär auf der Tatsache, dass die verwendeten Daten personenbezogene Informationen einer grossen Zahl unbekannter Personen abbilden. Damit wird die Bestimmbarkeit erschwert oder gar verunmöglicht. Indes sind die Tatsachen über einzelne Personen nach wie vor als Muster in den Daten vorhanden. Es besteht somit grundsätzlich die Möglichkeit, durch künstlich intelligente Mustererkennung Gruppen von Personen zu bilden oder gar einzelne Personen zu *singularisieren*. Dadurch kann eine Re-identifikation möglich werden. [228] Aufgrund dessen ist eine erfolgreiche Anonymisierung nicht ohne Weiteres anzunehmen.[229]

Die aktuelle Lehre geht denn auch von einem relativen Anonymisierungsbegriff aus, der sich nach dem Grad der Bestimmbarkeit der betroffenen Person und dem hierzu benötigten Aufwand richtet – ähnlich wie im Falle der Qualifikation von Daten als Personendaten. Insofern können auch «anonyme» Personendaten als Personendaten im Sinne des Datenschutzrechts gelten.[230] Verfügt eine Behörde beispielsweise über einen Schlüssel zur Wiederherstellung des Personenbezugs oder kann sie mit vertretbarem Aufwand einen solchen beschaffen, gelten anonyme Datensätze lediglich als *pseudonymisiert* und damit als Personendaten.[231]

---

[228] Vgl. dazu David Vasella, DSB Österreich: Einsatz von Google Analytics untersagt; Standard bei der Drittstaatsprüfung; Singularisierung statt Identifizierung, https://www.datenrecht.ch 26.01.2022, https://datenrecht.ch/dsb-oesterreich-einsatz-von-google-analytics-untersagt-standard-bei-der-drittstaatspruefung-singularisierung-statt-identifizierung/ (Abruf 01.06.2022); Philip Glass, Singularisierung und Identifizierung, https://www.datalaw.ch, 24.02.2018, https://www.datalaw.ch/singularisierung-und-identifizierung/ (Abruf 01.06.2022), N 6 ff.; Rosenthal (FN 225), 198 ff.; Philippe Meier, Le défi de Big Data dans les relations entre privés, in: Astrid Epiney/Daniela Nüesch (Hrsg.), Big Data und Datenschutzrecht, Zürich/Basel/Genf 2016, 56 ff.

[229] Früh (FN 105), AJP 2017, 144 m.w.H.; OECD (FN 13), 87.

[230] Epiney (FN 227), N 12.

[231] Früh (FN 105), AJP 2017, 144; Glass (FN 26), 114; vgl. Botschaft E-DSG (FN 188), 7076, wonach pseudonymisierte Personendaten bei fehlendem Schlüssel als «faktisch anonymisiert» bezeichnet werden.

In Zusammenhang mit KI-Technologien stellt sich neu die Frage, ob ein solcher Schlüssel vorbestehen muss oder ob es für die Annahme von Pseudonymisierung ausreichend ist, wenn ein Schlüssel durch maschinelles Lernen modelliert, also nachgebildet und auf diese Weise ermittelt werden kann. Aus der Perspektive des laufenden technologischen Fortschritts auf diesem Gebiet wird die Grenze zwischen anonym und pseudonym zunehmend unklar, und es stellt sich die weitergehende Frage, ob «anonyme» Personendaten nicht grundsätzlich im Geltungsbereich des Datenschutzrechts verbleiben sollten.[232]

Bedenkenswert ist hier der Einwand, dass in diesem Fall eine «Löschung durch Anonymisierung»[233] nicht mehr oder nur in einem sehr eingeschränkten Umfang möglich wäre. Dagegen lässt sich wiederum einwenden, dass die Aufnahme von anonymisierten Personendaten in den Geltungsbereich des Datenschutzrechts am Status von derart «gelöschten Personendaten» zunächst nichts ändern würde. Indes würde es bedeuten, dass die öffentlichen Organe für die weitere Bearbeitung der anonymen Daten weiterhin nach Datenschutzrecht verantwortlich wären.

Die Diskussion erscheint allerdings – zumindest in Bezug auf das IDG ZH – nicht dringlich, als bereits heute nach § 23 Abs. 1 IDG ZH von der Bekanntgabe von anonymen Personendaten – quasi als personenbezogenen Sachdaten – abgesehen werden müsste, wenn ein überwiegendes privates Interesse dagegen spricht. Abgesehen von den typischen Fällen der Gefährdung der Privatsphäre von Dritten gemäss § 23 Abs. 2 IDG ZH muss dies umso mehr dann gelten, wenn eine plausible Gefahr der De-Anonymisierung nachgewiesen werden kann.

B. Durchsetzung von Datenschutzrechten gegenüber KI-Bearbeitungen

Das Datenschutzrecht sieht gewisse Rechte vor, die den Betroffenen gegenüber Datenbearbeitern zustehen. Es handelt sich hierbei um Informationsrechte auf der einen und Rechte zur Beseitigung von Verletzungen auf der

---

[232] Meier (FN 228), 57; Roland Mathys, Big Data in der Rechtspraxis, in: Astrid Epiney/Daniela Nüesch (Hrsg.), Big Data und Datenschutzrecht, Zürich/Basel/Genf 2016, 99; Hoffmann-Riem (FN 5), 164; Weber (FN 227), 457 f.

[233] Baeriswyl, PraKom IDG ZH (FN 101), § 11 N 12; dazu eingehend David Rosenthal, Löschen und doch nicht löschen, in: digma 2019/4, 190 ff.

anderen Seite. Die Informationsrechte beinhalten das Recht, über gewisse Datenbearbeitungen informiert zu werden sowie das Recht, beim öffentlichen Organ Einsicht in die «eigenen» Personendaten zu nehmen. Während die Informationspflicht im Hinblick auf die Bearbeitung von Personendaten mittels künstlicher Intelligenz nicht offensichtlich problematisch ist, kann die Einsichtnahme in KI-Personendaten gewisse Probleme bergen.

1. Recht auf Information über die Erhebung von Personendaten

Das Gesetz verlangt in § 12 Abs. 1 IDG ZH, dass öffentliche Organe jene Personen informieren, über die sie Personendaten erheben. Dieser Grundsatz wird in § 12 Abs. 3 IDG ZH relativiert, indem weitreichende Ausnahmen von dieser Informationspflicht vorgesehen sind, namentlich wenn die betroffene Person bereits genügend Kenntnis von der Datenerhebung hat, die Beschaffung der Personendaten gesetzlich vorgesehen ist, die Mitteilung nicht möglich ist oder einen unverhältnismässigen Aufwand erfordern würde sowie in den Ausnahmefällen von § 23 IDG ZH, in denen eine Datenbekanntgabe verweigert werden kann.

Grundsätzlich stellt die Erhebung von Personendaten zum Zweck der oder durch die Bearbeitung von KI keinen signifikanten Unterschied dar zur «empirischen» Erhebung von Personendaten. In beiden Fällen muss klar sein, welche Kategorien von Daten erhoben werden und wozu. Dass für Personendaten, die als Output einer KI generiert wurde, ein spezifisches Risiko besteht, dass die Daten falsch sind, stellt nicht ein Problem der Information über die Erhebung der Daten dar, sondern ein Problem der Richtigkeit der Daten bzw. der rechtmässigen Bearbeitung.

2. Recht auf Einsichtnahme in die vorhandenen Personendaten

Das Recht auf Zugang zu den «eigenen» Personendaten gemäss § 20 Abs. 2 IDG ZH gilt insoweit, als Daten bei einem öffentlichen Organ vorhanden und mit der betreffenden Person verknüpft sind oder zumindest ohne grossen Aufwand verknüpft werden können. Dies ist der Fall für Personendaten, die das öffentliche Organ im Rahmen seiner normalen Aufgabenerfüllung bearbeitet

bzw. auf die es zugreifen kann. Vorhanden sind die Daten demnach, wenn die Behörde diese ohne weiteres bearbeiten kann, insbesondere ohne Rückgriff auf eine amtshilfeweise Bekanntgabe oder die Erfüllung einer gesetzlichen Auskunftspflicht durch ein anderes öffentliches Organ.[234] Hinsichtlich des Bestands eines Einsichtsrechts ist nicht von Belang, auf welche Weise die Daten erhoben wurden, ob durch empirische Tatsachenfeststellung oder mathematische Berechnung mittels KI.

In Bezug auf KI-Systeme dürfte das Zugangsrecht insbesondere gegenüber Outputdaten wirksam werden, da diese in der Regel als Ergebnis einer Datenbearbeitung gespeichert und weiterbearbeitet werden. Inputdaten werden hingegen nicht notwendigerweise gespeichert, zumindest nicht durch das Organ, welches den Input veranlasst. Nutzt eine Mitarbeiterin eines öffentlichen Organs beispielsweise eine Such- oder Übersetzungsmaschine, werden die Suchbegriffe in der Regel nicht durch das öffentliche Organ gespeichert. Unter Umständen besteht aber dennoch eine aus der Begründungs- und Dokumentationspflicht fliessende Pflicht die Inputdaten zu speichern oder eine entsprechende Aktennotiz zu erstellen. Dies erscheint denkbar, wenn die KI-Funktion einen Output generieren soll, der als Begründung für eine rechtlich relevante Entscheidung dient, und dessen Erklärungszusammenhang sich allein aus der KI-Funktion ergibt. Soweit der für eine Begründung relevante Erklärungszusammenhang allein in der KI-Funktion abgebildet ist bzw. durch diese verborgen bleibt, ist davon auszugehen, dass die Inputdaten gespeichert werden müssen, damit im Rahmen einer rechtlichen Überprüfung des Entscheids gegebenenfalls die Funktionsweise der KI nachvollzogen werden kann.[235]

[234] Vgl. Rudin, PraKom IDG ZH (FN 101), § 20 N 13.

[235] Siehe zur Explainability IV.A.1.d

VII. Ausgewählte Use Cases

A. KI-Bearbeitungen in gesetzlich besonders geschützten Lebensbereichen

1. Klassisch sensitive Bereiche: Religiöse, weltanschauliche, politische oder gewerkschaftliche Ansichten oder Tätigkeiten, Gesundheit, Intimsphäre, ethnische Herkunft

Diese in § 3 Abs. 4 IDG ZH genannten Bereiche können aus zwei Gründen als «klassisch sensitiv» bezeichnet werden. Erstens stehen sie durch thematische Überschneidung in einem gewissen Zusammenhang mit den Diskriminierungsmerkmalen in Art. 8 Abs. 2 BV.[236] Zweitens unterscheiden sie sich von den nachfolgenden besonderen Bereichen durch die Art und Weise, in der das grundrechtliche Risiko der entsprechenden Datenbearbeitung wirkt. Sie bezeichnen Themen, die üblicherweise als «sensibel» bzw. «privat» angesehen werden und den klassischen «Schutzbereich» des Datenschutzrechts im Sinne einer erweiterten Privatsphäre mitumschreiben. Ihre Bearbeitung ist denn auch nicht notwendigerweise mit einer Bedrohung für die Grundrechte verbunden, diese ist vielmehr kontextabhängig.[237]

Die Bearbeitung von Daten aus diesen Bereichen ist rechtlich ambivalent. Zum einen können durch die daraus erkennbaren Informationen wesentliche Aspekte der Persönlichkeit rekonstruiert werden, was der Gesetzgeber grundsätzlich als eine Bedrohung für die Grundrechte der Betroffenen wertet. Zum anderen kann eine in der Gesellschaft vorhandene diskriminierende Benachteiligung durch den Verzicht auf die Bearbeitung von Diskriminierungsmerkmalen i.S.v. Art. 8 Abs. 2 BV aus diesen Bereichen nicht ausgeglichen werden. Es besteht vielmehr die Gefahr, dass bereits vorhandene Probleme verschlimmert bzw. die Identifikation von diskriminierenden Praktiken verhindert wird, indem korrelative Daten zu einem Diskriminierungsmerkmal dieses redundant in den Datensatz hinein codieren.[238]

---

[236] Dazu Glass (FN 26), 141.

[237] Rudin, PraKom IDG ZH (FN 101), § 3 N 25.

[238] Dazu ausführlich Christian (FN 19), 64 ff.

Den klassischen sensitiven Merkmalen ist schliesslich gemeinsam, dass sie persönliche Eigenschaften einer Person erfassen, und dass diese Eigenschaften für sich allein keine eineindeutige Identifikation der betroffenen Person ermöglichen, diese aber oftmals einer definierbaren Gruppe von Personen zuweisen. Demgegenüber handelt es sich bei den nachfolgend dargestellten gesetzlichen Kategorien von besonderen Personendaten einerseits um Fälle der eineindeutigen Identifikationsmöglichkeit (biometrische Daten), andererseits um Daten von Behörden, deren Verbindung zu einer Person eine Stigmatisierung bewirken können (Sozialhilfe, Verfolgung und Sanktionen). Auch hier sind unproblematische Bearbeitungskonstellationen grundsätzlich denkbar.[239]

2. Genetische und biometrische Daten

Das Datenschutzrecht betrachtet biometrische Daten aufgrund ihrer Universalität, Einzigartigkeit und Beständigkeit als besonders risikoreich für die Grundrechte der betreffenden Person. Unter den Begriff der biometrischen Daten fallen gemäss der Botschaft des Bundesrates zum neuen DSG Personendaten, «die durch ein spezifisches technisches Verfahren zu den physischen, physiologischen oder verhaltenstypischen Merkmalen eines Individuums gewonnen werden und die eine eindeutige Identifizierung der betreffenden Person ermöglichen oder bestätigen», wobei ausdrücklich betont wird, dass ein technisches Verfahren beteiligt sein muss, welches «die eindeutige Identifizierung oder Authentifizierung einer Person erlaubt».[240] Dies entspricht wohl zugleich der Qualifizierung von biometrischen Daten als Personendaten.

Die ausdrückliche Nennung der genetischen Daten im Gesetz – die sich unbestrittenermassen zur Verwendung als biometrische Daten eignen – weist darauf hin, dass diese grundsätzlich auch dann besonders geschützt sind, wenn

---

[239] Rudin, PraKom IDG ZH (FN 101), § 3 N 25.

[240] Botschaft E-DSG (FN 188), 7020; dazu eingehend Dominika Blonski, Biometrische Daten als Gegenstand des informationellen Selbstbestimmungsrechts, Diss. Univ. Bern 2015, 6 ff.; siehe auch EDÖB – Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter, Leitfaden zu biometrischen Erkennungssystemen, Version 1.0 2009, https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/dokumentation/leitfaeden/leitfaden-zu-biometrischen-erkennungssystemen.html (Abruf 01.06.2022); *privatim*, Leitfaden zur datenschutzrechtlichen Beurteilung von biometrischen Verfahren, Version 1.0 Oktober 2006, https://www.privatim.ch/wp-content/uploads/2017/06/privatim_Leitfaden_Biometrie_2006_d-1.pdf (Abruf 01.06.2022).

sie nicht zur eindeutigen Identifizierung oder Authentifizierung einer Person mittels eines technischen Verfahrens genutzt werden. Allerdings fallen nach wie vor nur genetische Personendaten in den Geltungsbereich des DSG,[241] d.h. genetische Daten, die geeignet sind, die betreffende Person mit einem zumutbaren Aufwand zu ermitteln.

Aufgrund der gesetzlichen Qualifikation von biometrischen Daten als besondere Personendaten müssen Bearbeitungen gemäss § 8 Abs. 2 IDG ZH in einer hinreichend bestimmten Regelung in einem formellen Gesetz geregelt sein, d.h. insbesondere Bestimmung des verantwortlichen Organs, Ziel und Zweck der Bearbeitung, Kategorien der bearbeiteten Daten sowie die Art und Weise der Bearbeitung.[242]

Das Gesetz muss somit die Befugnis erteilen, eine Identifikation oder Authentifizierung mittels biometrischer Daten vorzunehmen. Von einer genügenden Regelung gedeckt wären sämtliche begleitenden Datenbearbeitungen gedeckt, die zur Durchführung eines biometrischen Vergleichs notwendig sind, und die zugleich kein zusätzliches persönliches Risiko für die Betroffenen schaffen, insbesondere die Erhebung und Speicherung sowie der Vorgang des Vergleichens. Spezifisch geregelt werden müsste indes die Frage, was nach Abschluss eines Identifizierungs- bzw. Authentifizierungsvorgangs mit den biometrischen Profilen geschieht, die für den Vergleich benutzt wurden.[243] Dies gilt vor allem auch dann, wenn es sich um Identifizierungs- oder Authentifizierungsvorgänge handelt, die üblicherweise mehrmals durchgeführt werden. Da die Vorratsdatenspeicherung mangels Bearbeitungszweck grundsätzlich als unrechtmässig gilt,[244] muss die Bearbeitung in irgendeiner Dimension begrenzt werden, beispielsweise durch Ablauf des Arbeitsverhältnisses (zeitlich) bzw. die Ausübung der Funktion, welche für die betreffende Bearbeitung zuständig ist (sachlich/personell).

Schliesslich sind die weiteren Umstände der Datenbearbeitung in die Risikobewertung mitaufzunehmen. Zu unterscheiden wäre insbesondere zwischen

---

[241] David Rosenthal, Das neue Datenschutzgesetz, in: Jusletter vom 16.11.2020, N 20, 22.

[242] Baeriswyl, PraKom IDG ZH (FN 101), § 8 N 14.

[243] Für das Beispiel der Gesichtserkennung siehe VII.A.3.

[244] Siehe IV.A.1.b.

offenen, von den Betroffenen steuerbaren Bearbeitungsvorgängen (etwa der Nutzung eines Fingerabdrucks zum Entsperren eines Arbeitscomputers) und verdeckten Bearbeitungen bzw. einer biometrischen Fernidentifizierung. Eine solche liegt beispielsweise vor, wenn Personen ohne ihr Wissen bzw. auf Distanz mittels Sensoren aufgrund ihres Gesichts, ihrer Gangart, Pulsrythmen und ähnlichen biometrischen Eigenheiten identifiziert werden. Dies erzeugt ein zusätzliches Risiko für die Grundrechte der Betroffenen und ist daher rechtlich separat zu bewerten.[245]

3. Insbesondere biometrische Gesichtserkennung

a. Automatisierte biometrische Erkennung

Biometrische Erkennung bedeutet die Erkennung einer Person anhand von biometrischen Merkmalen. Im Falle einer automatisierten Erkennung erfolgt diese über Sensoren, welche mit einem entsprechenden KI-System verbunden sind. Als Sensoren werden beispielsweise Kameras bzw. Scanner oder Mikrofone eingesetzt. Die Outputdaten der KI stellen eine maschinelle Erkennung dar, soweit sie die Identität einer Person bestätigen. Neben Finger- bzw. Handabdruck-, Iris- oder Venenmustern ist vor allem die Gesichtserkennung im Einsatz. Sie hat den Vorteil, dass sie auf Distanz bzw. anhand von alltäglichen Fotos bzw. Videoaufnahmen der Betroffenen vorgenommen werden kann.[246]

Die spezifischen Risiken von biometrischer Erkennung sind mittlerweile anerkannt und führen dazu, dass Datenschutzgesetze in der Schweiz die Bearbeitung biometrischer Daten als neue Kategorie von besonders schützenswerten bzw. besonderen Personendaten eingeführt haben oder einführen werden, so beispielsweise im Kanton Zürich.[247] Noch nicht in Kraft ist die entsprechende Anpassung im neuen Datenschutzgesetz des Bundes. Unter den Begriff der besonders schützenswerten Personendaten fallen künftig gemäss Art. 5 Bst. c

---

[245] EDÖB (FN 240), 3.

[246] Gerrit Hornung/Stefan Schindler, Datenschutz bei der biometrischen Gesichtserkennung – Künstliche Intelligenz und Mustererkennung als Herausforderung für das Recht, DuD 8/2021, 515–521, 515.

[247] In § 3 Abs. 4 lit. a Ziff. 2 IDG ZH mit der Änderung vom 25. November 2019, in Kraft seit dem 1. Juni 2020 (OS 75, 63).

Ziff. 4 nDSG ausdrücklich «biometrische Daten, welche eine Person eindeutig identifizieren».[248]

Derweil sieht ein Regulierungsentwurf der Europäischen Union vor, die biometrischen Fernidentifizierung in Echtzeit zu Strafverfolgungszwecken im öffentlichen Raum in der EU je nach dem damit verbundenen Risiko für die Betroffenen stark einzuschränken bzw. grundsätzlich zu verbieten.[249]

b. Gesichtserkennung als stellvertretendes Beispiel

Die biometrische Erkennung ist vermutlich eine der in der Öffentlichkeit bekanntesten KI-Technologien mit einer sehr grossen Bandbreite an möglichen Einsatzfeldern, wie beispielsweise der Authentifizierung von Telefon- und Computernutzern. Konzeptionell bedeutet automatisierte Gesichtserkennung, dass ein mit Personendaten verbundenes Gesichtsmuster durch die Auswertung von physiognomischen Eigenheiten einer Person zugeordnet wird. Ihre grundlegende Funktion ist demnach jene des biometrischen Erkennens von Personen. Durch sie kann beispielsweise eine Zugangsberechtigung bestätigt oder Bilder und Videomaterial in Bezug auf die darin erfassten Personen ausgewertet werden. Im ersten Fall spricht man von einem *one-to-one matching*: das Muster, das von der zu erkennenden Person erstellt wird, wird mit einem Muster in der Datenbank verglichen, um die betreffende Person zu verifizieren. Diese Person ist dem System also bekannt, muss aber ihre Identität beweisen. Im zweiten Fall spricht man von einem *one-to-many matching*: die erfassten Personen werden mit vielen Mustern in einer Datenbank verglichen, um sie zu identifizieren. Die sensorisch erfassten Personen sind dem System nicht bekannt und sollen zwecks Identifikation mit den in der Datenbank vorhandenen Personen verglichen werden. [250]

---

[248] BBl 2020 7639, 7641.

[249] Siehe den Vorschlag für ein Gesetz über die künstliche Intelligenz vom 22. April 2021, COM(2021) 206 final, Art. 5 Abs. 1 Bst. d des Entwurfs.

[250] Vgl. EDÖB (FN 240), 5; FRA – Agentur der Europäischen Union für Grundrechte, Gesichtserkennungstechnologien: grundrechtsrelevante Erwägungen im Rahmen der Strafverfolgung, Januar 2020, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper_de.pdf (Abruf 01.06.2022), 4; Nadja Braun Binder/Eliane Kunz/Liliane Obrecht, Maschinelle Gesichtserkennung im öffentlichen Raum, in: sui generis 2022, N 7.

Die automatisierte Gesichtserkennung ist als allgemeines technologisches Konzept zu verstehen, dessen Hauptfunktion, die Erkennung, sowie die hierdurch ermöglichten Funktionen der Verifizierung und Identifizierung bzw. Suche in sämtlichen Lebensbereichen einsetzbar sind.[251] Das Konzept lässt sich in der Regel datenschutzkonform ausgestalten, indem die mit dem konkreten Einsatz von automatisierter Gesichtserkennung verbundenen Risiken identifiziert und durch geeignete Massnahmen gemindert werden. Diese Risiken ergeben sich einerseits aus der verwendeten Technologie und andererseits aus den Umständen ihrer Anwendung bzw. aus der Automatisierung von Verwaltungs- und Entscheidungsprozessen in der Gesellschaft.[252] Dabei betreffen die technologischen Risiken die Qualität der durch sie generierten Daten, die Risiken der Automatisierung und gesellschaftlichen Einbettung hingegen die Verwendung dieser Daten. Hierbei stellen sich je unterschiedliche Rechtsfragen.[253]

c. Grundlegende Risikostruktur

Im Gegensatz zu anderen eineindeutigen biometrischen Merkmalen, beispielsweise Fingerabdruck- oder DNA-Muster, verändern sich die Gesichter über Zeit merklich. Es handelt sich demnach um eine permanente, nicht ohne weiteres änderbare aber zugleich veränderliche Eigenschaft einer Person. Mit der zunehmenden Leistungsfähigkeit der Gesichtserkennungstechnologie, scheint diese Veränderung mittel- bis langfristig eine überwindbare Hürde darzustellen: Personen werden auch dann erkannt, wenn zwischen Bild und Gesicht (derselben Person) mehrere Jahre liegen, wobei die Fehlerrate leicht

---

[251] Zu den Begriffen der Verifizierung (oder auch Authentifizierung) und Identifizierung siehe Blonski (FN 240), 12 ff.; siehe auch die illustrativen Beispiele bei Ramona Keist, Gesichtserkennung im zivilrechtlichen Persönlichkeitsschutz, in: Jusletter vom 20.05.2019, N 7.

[252] Vgl. dazu I.A.3.

[253] Vgl. dazu IV.A.2. und IV.A.3.; Unklar hier Braun Binder et al. (FN 250), 53 ff., die betonen, dass die mit der automatisierten Gesichtserkennung zusammenhängenden Rechtsfragen sich «grundsätzlich unabhängig» von der eingesetzten Technologie stellen (N 7) und an anderer Stelle auf die Gefahr einer Diskriminierung durch *false positives* hinweisen (N 32) – letztere wird aber durch das Training des verwendeten Modells begründet und stellt ein intrinsisches Problem trainierter KI-Modelle dar; siehe dazu I.B.3.c.

zunimmt.[254] Theoretisch können alte Fotos (von genügender Qualität) und möglicherweise Bilder aus der Kindheit zur Erkennung von erwachsenen Gesichtern ausreichen.[255] Neuerdings soll das Gesicht gar aus der DNA einer Person berechnet werden können.[256]

Neben dem zeitlichen Faktor kann als weiterer Risikofaktor der Umstand genannt werden, dass Gesichter in der Regel für Dritte mit verhältnismässig wenig Aufwand zugänglich sind. Sie können per Foto- oder Videoaufnahme auf Distanz ermittelt werden. Dies kann in unmittelbarer Weise mittels Livebilder geschehen oder mittelbar[257] aufgrund von Fotos und Videos.

d. Risiken durch Datenbearbeitung

Neben den intrinsischen Risiken der Gesichtserkennung bestehen weitere, durch die Art und Weise sowie den Zweck der Datenbearbeitung begründete Risiken. Dabei ist zu unterscheiden zwischen Risiken, welche die Betroffenen einer Bearbeitung von Personendaten tragen sowie Risiken, die eine unbekannte Anzahl von Personen betreffen. Denn mit der vereinfachten Skalierbarkeit moderner KI- und Big-Data-Technologien weiten sich die hier besprochenen Risiken der automatisierten Gesichtserkennung auf die gesellschaftliche Ebene aus. Es wird befürchtet, dass ein unreflektierter Einsatz von Gesichtserkennungstechnologien die demokratische Meinungsbildung erschweren und einzelne Gruppen in diskriminierender Weise benachteiligen könnte.[258]

Betroffene einer Bearbeitung von Personendaten im Rahmen einer Gesichtserkennung sind zunächst jene Personen, deren biometrisches Gesichtsmuster

---

[254] Lacey Best-Rowden/Anil K. Jain, Longitudinal Study of Automatic Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 40/1 2018, 148.

[255] Alexis C. Madrigal, Computers See Your Face as a Child: Will They Recognize You as an Adult?, The Atlantic, 13.05.2014, legt die (damalige) untere Altersgrenze bei 7 Jahren fest.

[256] Tate Ryan-Mosley, This company says it's developing a system that can recognize your face from just your DNA, Technology Review 31.01.2022; unklar, ob dies tatsächlich realisierbar ist.

[257] Auch als »nachträgliche Gesichtserkennung« bezeichnet; vgl. Braun Binder et al. (FN 250), N 14.

[258] FRA (FN 250), 4; Braun Binder et al. (FN 250), N 32 m.w.H.

als Vergleichsmuster vom System für den Vergleichsvorgang herangezogen wird. Erstens werden biometrischen Personendaten verwendet und zweitens besteht der Zweck der Bearbeitung darin, diese Personen zu erkennen, also zu identifizieren. Dagegen gelten Personen, die sich beispielsweise im Sichtfeld einer Videokamera mit Gesichtserkennungsfunktion aufhalten, *datenschutzrechtlich* nicht als Betroffene der Gesichtserkennung, solange sie nicht einem bestimmten Gesichtsprofil zugeordnet wurden, da sie diesbezüglich weder bestimmt noch bestimmbar sind, bzw. nicht identifiziert wurden.[259] Soweit aber ein KI-System sie zum Zweck des Vergleichs sensorisch erfasst, sind sie – wie auch jene Personen, die tatsächlich identifiziert werden – von der Überwachung betroffen, die gegebenenfalls durch die Gesichtserkennung bewirkt werden soll. Eine solche Überwachung birgt eigene Grundrechtsrisiken, namentlich jenes des Einschüchterungseffekts oder auch *chilling effect*.[260]

Die Unterscheidung der beiden Gruppen von Betroffenen ist somit von Bedeutung, weil die Risikostruktur für die datenschutzrechtlich von der Gesichtserkennung betroffene Personen sich von jener der übrigen erfassten Personen unterscheidet. Indes sind die Übergänge fliessend, da für Personen, die an und für sich datenschutzrechtlich nicht von der Gesichtserkennung betroffen wären, die Möglichkeit besteht, dass sie fälschlicherweise das Lager wechseln, wenn die KI «einen Fehler macht»[261] und für sie ein *false positive* ausgibt, d.h. sie falsch, d.h. als eine andere Person identifiziert.[262] Durch den fehlerhaften Personenbezug entstehen unrichtige Personendaten,[263] die je nach Verwendungszusammenhang ein beträchtliches persönliches Risiko für die Betroffenen bedeuten, und die im Sinne der Datenrichtigkeit mittels geeigneter Qualitätskontrollen möglichst schnell zu berichtigen bzw. zu löschen sind.[264]

---

[259] Dazu Rosenthal (FN 225), 200; Glass (FN 228), N 5; hier zeigt sich erneut die Problematik, dass der Geltungsbereich der Datenschutzgesetze durch den relativen Begriff des Personendatums mitbestimmt wird.

[260] Glass (FN 26), 154; Braun Binder et al. (FN 250), N 30.

[261] Technisch handelt es sich nicht um einen Fehler, sondern um die Manifestation der statistischen Fehlerquote des Systems; vgl. FRA (FN 250), 9.

[262] FRA (FN 250), 9.; zur Terminologie zuletzt Braun Binder et al. (FN 250), N 31.

[263] Tina Krügel/Julia Pfeiffenbring, Datenschutzrechtliche Herausforderungen von KI, in: Martin Ebers /Christian Heinze/Tina Krügel/Björn Steinrötter (Hrsg.), Künstliche Intelligenz und Robotik – Rechtshandbuch, München 2020, § 11 N 26.

[264] Zur Datenrichtigkeit siehe IV.A.3.

e. Rechtliche Vorgaben

Im kantonalen Recht existieren soweit ersichtlich keine ausdrücklichen Vorgaben für den Einsatz von Gesichtserkennungstechnologie durch öffentliche Organe, wohingegen dies im Bundesrecht vereinzelt vorgesehen ist.[265] Dies ist insofern bemerkenswert, als die betreffenden Vergleichs- und Outputdaten aufgrund ihrer Eigenschaft als biometrische Daten als besondere Personendaten im Sinne von § 3 Abs. 4 Bst. a Ziff. 2 IDG ZH gelten. Der Einsatz von Gesichtserkennungstechnologien muss daher gemäss § 8 Abs. 2 IDG ZH in einem formellen Gesetz hinreichend geregelt sein.[266]

Was in diesem Zusammenhang eine hinreichende Regelung im Gesetz bedeutet, ist unklar. Je nach Einsatzzweck der Gesichtserkennung wird die betreffende gesetzliche Grundlage mehr oder weniger bestimmt ausfallen müssen. So erscheint es beispielsweise als naheliegend, dass die Verwendung von Laptops mit kamerabasiertem Login durch das Personal eines öffentlichen Organs auf Gesetzes- bzw. Verordnungsebene weniger klar vorgespurt sein muss – tatsächlich reicht hierzu wohl die Pflicht zur Sicherstellung der Datensicherheit aus – als beispielsweise der Einsatz zur Identifikation von Fahrzeuglenkern bei Geschwindigkeitskontrollen.

Für die Normdichte einer Befugnis zum Einsatz von Gesichtserkennungstechnologie gelten grundsätzlich die allgemeinen Kriterien für gesetzliche Bearbeitungsgrundlagen für besondere Personendaten. Zumindest müssen das verantwortliche Organ, Ziel und Zweck der damit verbundenen Datenbearbeitungen, die zulässigen Datenkategorien sowie die Art und Weise der Datenbearbeitungen geregelt werden.[267] Hierbei muss das Gesetz deutlich zum Ausdruck bringen, wie die Gesichtserkennungsfunktion in den betreffenden Arbeitsprozess eingebunden ist und welche Aufgabe durch sie in diesem Rahmen erfüllt werden soll. Weiter muss in den Materialien plausibel begründet werden, weshalb kein milderes Mittel das verfolgte Ziel ausreichend

---

[265] So etwa in Art. 103 Abs. 1 und 5 AIG, die es den zuständigen kantonalen Behörden ermöglichen, ankommende Flugpassagiere mittels elektronischer Erkennung zu identifizieren; dazu Braun Binder et al. (FN 250), N 29.

[266] Gesichtsdaten werden im Rahmen eines Trainings nicht personenbezogen bearbeitet und stellen insofern keine Personendaten dar; Siehe dazu III.A.1.

[267] Baeriswyl, PraKom IDG ZH (FN 101), § 8 N 14.

verwirklichen würde.[268] Hierbei muss die Risikobeurteilung für jede mit dem Einsatz der Gesichtserkennungstechnologie verbundene Bearbeitung von personenbezogenen Daten (inkl. Beschaffen, Aufbewahren, Verwenden, Umarbeiten, Vernichten)[269] je separat sowie kumulativ vorgenommen und rechtlich bewertet werden.

4. Massnahmen der sozialen Hilfe

Die Sozialhilfe in der Schweiz wird seit Jahren von Debatten über die Verhinderung unrechtmässiger Leistungen geprägt, welche dazu geführt hat, dass verschiedene Kantone entsprechende Massnahmen zur Bekämpfung von unrechtmässig bezogenen Leistungen im Gesetz verankert haben. Im Kanton Zürich ist dies insbesondere die in § 48a SHG geregelte Observation, welche den Einsatz von Observationsspezialisten ermöglicht sowie die in § 48 SHG geregelten, weitreichenden Auskunftsbefugnisse von Behörden und Privatpersonen gegenüber der Sozialhilfe. Flankiert werden diese Massnahmen durch die in § 47b SHG[270] enthaltenen, umfassenden gesetzlichen Anzeigepflichten von Verwaltungsbehörden des Kantons, der Gemeinden sowie der mit der Erfüllung von öffentlichen Aufgaben betrauten Organisationen und Personen gegenüber der Sozialhilfe.

Die Sozialhilfebehörden verfügen auch inhaltlich über sehr weitreichende Befugnisse zur Bearbeitung von Personendaten ihrer Klientinnen und Klienten. Tatsächlich enthält § 18 SHG, der die zulässigen Datenkategorien bezeichnet, welche durch die Sozialhilfebehörde in Erledigung ihrer Aufgaben bearbeiten dürfen, keine nennenswerten Einschränkungen. Entsprechend sind die Behörden grundsätzlich berechtigt, vorbehalten der genügenden Plausibilisierung eines Zusammenhangs mit ihren weit gefassten Aufgaben, Personendaten aus beliebigen Lebensbereichen der Betroffenen zu bearbeiten. Aufgrund dieser weitreichenden Bearbeitungsbefugnisse zur Erfüllung ihrer Aufgaben, insbesondere die Bedürftigkeit von Gesuchstellerinnen zu beurteilen, über entsprechende Hilfe zu entscheiden und unrechtmässig bezogene Unterstützung zu

[268] Vgl. dazu Braun Binder et al. (FN 250), N 17.

[269] Siehe FN 130.

[270] Sozialhilfegesetz vom 14. Juni 1981 des Kantons Zürich (SHG; ON 851.1).

entdecken, verfügen Sozialhilfebehörden in der Regel über sehr detaillierte Daten zu den Lebensumständen ihrer Klientinnen und Klienten.

An und für sich wäre damit eine Fülle von Trainingsdaten für den Einsatz von KI-Technologien gegeben, mit deren Hilfe möglicherweise Fälle von unrechtmässiger Unterstützung oder gar Missbrauch identifiziert werden könnten. Systeme zur Aufdeckung von unrechtmässig ausbezahlten Sozialhilfeleistungen bzw. der diesbezüglichen Risikoanalyse sind denn auch vereinzelt bereits im Einsatz, so beispielsweise in Dänemark und den Niederlanden.[271] Solche Analysen sind funktional eng mit der personenbezogenen prädiktiven Polizeiarbeit verwandt. Aufgrund der noch jungen Technologie, die zum Einsatz kommt, sind Ergebnisse mit Vorbehalt zu nutzen bzw. nur zurückhaltend als Indizien für das Vorliegen eines tatsächlichen unrechtmässigen Bezugs zu werten. Auch ist die Transparenz der Bearbeitung stets zu wahren,[272] d.h. verdeckte Missbrauchsanalysen durch intelligente Mustererkennung in den Klientinnendaten müssten gesetzlich vorgesehen und geregelt sein und die Betroffenen in irgendeiner Form informiert werden.[273] Ersteres ergibt sich nicht zuletzt aus dem Umstand, dass Ergebnisse einer Mustererkennungsanalyse von Sozialhilfedaten einer Person in der Regel als eine «Zusammenstellungen von Informationen, die eine Beurteilung wesentlicher Aspekte der Persönlichkeit natürlicher Personen erlauben» i.S.v. § 3 Abs. 4 Bst. b IDG ZH zu werten sein werden, und daher eine Bearbeitung von besonderen Personendaten i.S.v. § 8 Abs. 2 IDG ZH vorliegt.

---

[271] Braun Binder et al. (FN 8), 29 (Dänemark), 31 (Niederlande); sowie Hinweis bei OECD (FN 13), 70. m.w.H (GB); siehe auch Melissa Heikkilä, Dutch scandal serves as a warning for Europe over risks of using algorithms, Politico.eu, 29.03.2022.

[272] Zur Transparenz siehe IV.A.1.c.

[273] Zur Rechtsgrundlage für KI-Bearbeitungen siehe IV.A.1.a.

5. Administrative oder strafrechtliche Verfolgungen oder Sanktionen

Neben *predictive policing*[274] und künstlich intelligenten Onlinetools zur Bekämpfung komplexer krimineller Strukturen[275] sind hier insbesondere auch niederschwellige Verfolgungs- und Sanktionsmassnahmen der Massenverwaltung zu nennen, so beispielsweise automatische Geschwindigkeitskontrollen oder die Videoüberwachung und -analyse von Menschenmengen zwecks *crowd management, hotspot policing* oder die anlässlich von Sportveranstaltungen durchgeführte Erkennung von Personen, die im «Hooligan-Register» des Bundes[276] verzeichnet sind. Weiter fallen auch die im Rahmen der Darstellung von Bias genannten Systeme, wie COMPAS, in diese Kategorie.[277]

6. Insbesondere Predictive Policing

a. Breiter Abwehr- und Präventionsauftrag der Polizei

Polizeigesetze sind von offenen Normen geprägt, welche die Aufgaben und Mittel der Polizeien beschreiben. Ergänzt werden diese zunehmend durch verfassungskonkretisierende Grundsätze der Polizeiarbeit, welche zur verfassungskonformen Auslegung der Gesetze anmahnen. Abgesehen davon besteht für die Polizeien ein weiter gesetzlicher Spielraum mit nur wenigen Vorentscheidungen des Gesetzgebers in Bezug auf Einzelfälle. Eine Konkretisierung besteht insofern, als die zulässigen Massnahmen der Aufgabenerfüllung gewisse Konturen verleihen, indem sie die zulässigen Datenbearbeitungen genauer beschreiben.[278] Diese werden ihrerseits durch die Regelungen der zulässigen Datenbanken

---

[274] Zu *predictive policing* siehe sogleich VII.A.6.

[275] Beispielsweise das Projekt Memex der DARPA (Entdecken von Menschenhandel), https://www.defense.gov/News/News-Stories/Article/Article/1041509/darpa-program-helps-to-fight-human-trafficking/ (Abruf 01.06.2022); Projekt zur Bekämpfung von Kinderpornografie des Bundeslandes NRW in Zusammenarbeit mit Microsoft Deutschland vgl. https://www.sueddeutsche.de/digital/software-maschinenlernen-ki-kuenstiche-intelligenz-kinderpornografie-polizei-nrw-1.4553870 (Abruf 01.06.2022).

[276] Zum Informationssystem Hoogan und den in den letzten Jahren erfassten Personengruppen und Straftaten siehe den Bericht des Fedpol vom 1. Juli 2021, https://www.fedpol.admin.ch/fedpol/de/home/sicherheit/hooliganismus/zahlen/hoogan.html (Abruf 01.06.2022).

[277] Siehe I.B.3.c.

[278] Glass (FN 26), 245 ff.

sowie den darin zu bearbeitenden Datenkategorien ergänzt, insbesondere auch bezüglich der Zugriffsrechte, woraus in der Regel die wichtigsten Eckdaten der polizeilichen Informationssysteme ersichtlich werden.[279]

Die hier beschriebene offene Normierung findet ihre Begründung in der Offenheit der Gefahrenabwehr als zentrale Aufgabe der Polizei, insbesondere in der Aufklärung und dem Vorbeugen von Straftaten.[280] Da die Gefahrenabwehr regelmässig im Schutzbereich der Grundrechte erfolgt, muss sie sich in besonderem Masse an den Bestimmungen über die Umsetzung der Grundrechte in Art. 35 BV orientieren. Dies gilt insbesondere auch für die Frage, welche Gefahren im Einzelfall prioritär bekämpft werden.[281]

b. Das Konzept der automatisierten polizeilichen Gefahrenprognose

In der Schweiz werden diverse polizeiliche Prognoseprogramme eingesetzt.[282] Diese Formen der Polizeiarbeit nutzen algorithmische Prognosen, um «lagebezogene Wahrscheinlichkeitsaussagen» in Bezug auf mögliche künftige Straftaten zu generieren.[283] Ziele sind die Verbesserung der Präventionsarbeit sowie die effizientere Nutzung von Ressourcen.[284] Es handelt sich um modellbasierte Vorfeldermittlung.[285]

Dabei wird versucht, die Aufgabe der präventiven Gefahrenabwehr durch künstlich intelligente Algorithmen anzugehen und automatisch voraussagen zu lassen, wo (raumbezogen) oder durch wen (personenbezogen) künftig die

---

[279] Siehe die Hinweise bei Belser/Noureddine (FN 26), Datenschutzrecht Grundlagen, 448.

[280] Belser/Noureddine (FN 26), Datenschutzrecht Grundlagen, 448.

[281] Glass (FN 26), 251.

[282] Monika Simmler/Simone Brunner, Die Kantone im Bann der Algorithmen?, in: Monika Simmler (Hrsg.), Smart Criminal Justice – Der Einsatz von Algorithmen in der Polizeiarbeit und Strafrechtspflege, Basel 2021, 15 f.; Braun Binder et al. (FN 8), 25.

[283] Johanna Sprenger, Verbrechensbekämpfung: Predictive Policing, in: Künstliche Intelligenz und Robotik – Rechtshandbuch, München 2020, § 31 N 5 f.; Sommer (FN 218), 36.

[284] Sprenger (FN 283), § 31 N 38.

[285] Glass (FN 26), 258 f.

Verwirklichung einer polizeilichen Gefahr droht.[286] Gewisse Systeme verbinden beide Formen, indem sie per Videoanalysen in Echtzeit Gefahren erkennen sollen. Dies geschieht anhand verschiedener biometrischer Merkmale, wie etwa Gesichtserkennung (bekannte Person) oder Erkennung von gefährlichen Bewegungsabläufen (unbekannte Personen).[287]

Raumbezogene Gefahrenprognosen können datenschutzrechtlich relevant werden, indem sie Personengruppen erzeugen, die unter einem Generalverdacht stehen, gefährlicher zu sein als die allgemeine Bevölkerung, indem sie deren Basisrate[288] im Risikoscore übertreffen. Weiter darf nicht vergessen gehen, dass die verwendeten Analyseparameter (beispielsweise Tatzeitraum, Tatort, Tatobjekt, *modus operandi* und Beute vergangener Delikte)[289] personenbezogene Daten sind, die (vorerst) nicht bestimmbare Personen betreffen. Ziel der Prognose ist es schlussendlich, jene Personen zu identifizieren und überführen, welche durch ihre Tätigkeit die Daten liefern. Schliesslich kann mit dem Gedanken gespielt werden, Erkenntnisse aus der raumbezogenen Gefahrenprognose Personen «aus der Gegend», für die ein Risikoscore berechnet wird, als erschwerenden Faktor zuzuweisen.

B. Profiling

Neu definiert Art. 5 Bst. f nDSG den Begriff des Profiling als «jede Art der automatisierten Bearbeitung von Personendaten, die darin besteht, dass diese Daten verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten, insbesondere um Aspekte bezüglich Arbeitsleistung, wirtschaftlicher Lage, Gesundheit, persönlicher Vorlieben, Interessen, Zuverlässigkeit, Verhalten, Aufenthaltsort oder Ortswechsel dieser natürlichen Person zu analysieren oder vorherzusagen». Der Begriff des Persönlichkeitsprofils wird aus dem Gesetz gestrichen, ausschlaggebend soll

---

[286] Jennifer Pullen/Patricia Schefer, Predictive Policing – Grundlagen, Funktionsweise und Wirkung, in: Monika Simmler (Hrsg.), Smart Criminal Justice – Der Einsatz von Algorithmen in der Polizeiarbeit und Strafrechtspflege, Basel 2021, 103–122, 105 f.; Christen et al. (FN 10), 211 f.; Sommer (FN 218), 36 f.

[287] Wischmeyer (FN 64), § 20 N 17.

[288] Zum Begriff Sommer (FN 218), 52.

[289] Vgl. die Darstellung zu PRECOBS bei Simmler/Brunner (FN 282), 17 ff.

neu der Vorgang sein und nicht mehr das Ergebnis.[290] Das nDSG unterscheidet zwischen einfachem Profiling und Profiling mit hohem Risiko. Letzteres wurde vom Ständerat vorgeschlagen und die endgültige Fassung erst auf Antrag der Einigungskonferenz in beiden Räten genehmigt.[291]

Ein hohes Risiko liegt gemäss Art. 5 Bst. g nDSG dann vor, wenn das Profiling «ein hohes Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person mit sich bringt, indem es zu einer Verknüpfung von Daten führt, die eine Beurteilung wesentlicher Aspekte der Persönlichkeit einer natürlichen Person erlaubt». Hier klingt die Legaldefinition des Persönlichkeitsprofils von Art. 3 Bst. d DSG nach: eine solches besteht in einer «Zusammenstellung von Daten, die eine Beurteilung wesentlicher Aspekte der Persönlichkeit einer natürlichen Person erlaubt». Ob und wie sich diese Unterscheidung in der Praxis bewähren wird, ist noch unklar. Es kann indes davon ausgegangen werden, dass Datenauswertungen mittels KI-Technologien, welche sich auf bestimmte oder bestimmbare Personen beziehen, zumindest als «einfaches» Profiling gelten werden.[292]

In jedem Fall soll gemäss Vorschlag des Bundesrates ein Profiling nur dann vorliegen, wenn die *Bewertung* einer Person *vollautomatisiert* vorgenommen wird. Hierunter fällt «jede Auswertung mit Hilfe von computergestützten Analysetechniken».[293] Damit wird die Betonung auf die Automatisierung der Auswertung gelegt und nicht auf den Grad der Automatisierung des gesamten Entscheidungsprozesses, in den eine Auswertung eingebettet ist. Die hieraus ersichtliche Relativierung der Unterscheidung zwischen voll- und teilautomatisierten Entscheiden ist zu begrüssen, da die Problemlage jeweils eine ähnliche ist.[294]

Unklar bleibt schliesslich, weshalb die Botschaft betont, dass die Definition des Profiling nicht bedeute, dass Algorithmen verwendet werden müssten,[295]

---

[290] Botschaft E-DSG (FN 188), 7021.

[291] Geschäft des Bundesrates 17.059, Datenschutzgesetz. Totalrevision und Änderung weiterer Erlasse zum Datenschutz, Beschlüsse gemäss Antrag der Einigungskonferenz des National-, bzw. des Ständerates vom 24.09.2020.

[292] Siehe auch den Hinweis bei Klaus (FN 36), 86.

[293] Botschaft E-DSG (FN 188), 7022.

[294] Siehe IV.A.4.c.

[295] Botschaft E-DSG (FN 188), 7022.

da *prima vista* nicht ersichtlich ist, wie eine vollautomatisierte computergestützte Auswertung ohne die Verwendung von Algorithmen stattfinden soll.

C. KI-Bearbeitungen in allgemeinen Verwaltungsprozessen

Unabhängig davon, ob Daten aus einem gesetzlich geschützten Bereich gemäss § 3 IDG ZH bearbeitet werden, ist absehbar, dass Verwaltungsprozesse künftig durch neue KI-Technologien unterstützt werden.[296] Man verspricht sich davon beispielsweise eine Entlastung der Arbeitsprozesse durch Steigerung der Effizienz, indem etwa Routinearbeit durch einen KI-Agenten erledigt wird, bessere statistische Prognosen für komplexe wirtschafts-, sozial-, umwelt-, oder sicherheitspolitische Zusammenhänge, besseren Service und mehr Kundennähe durch Chatbots und andere e-Government-Dienstleistungen mit KI-Funktionen, die Überprüfung und Vereinfachung von Prozessabläufen oder eine bessere Verwendung der bei der Verwaltung vorhandenen grossen Datenmengen.[297] Im Folgenden werden zwei aktuelle Anwendungen mit KI-Einschlag thematisiert: *chatbots* und online-Übersetzungen.

1. Chatbots

Verschiedene Verwaltungen in der Schweiz setzten für die Kommunikation mit ihren Kundinnen und Kunden Chatbots ein. Im Vordergrund steht zurzeit die Bereitstellung von interaktiv aufbereiteter Information in Dialogform, doch sollen künftig auch rechtswirksame Handlungen gegenüber öffentlichen Organen möglich werden (etwa die Einreichung von Gesuchen).[298] Als Chatbot werden Userinterfaces bezeichnet, die einen Text- oder auch Sprachdialog in Alltagssprache simulieren (to chat = plaudern). Es handelt sich demnach um eine «konversationsbasierte Schnittstelle»[299] zwischen Nutzerinnen und Nutzern und einem Informationssystem. Dagegen gelten Schnittstellen, die auf der Grundlage eines vorprogrammierten Frage-/Antwort-Systems (FAQs)

---

[296] Gewisse KI-Anwendungen sind schon länger in Betrieb. Es handelt sich um mittlerweile alltägliche, «niederschwellige» Anwendungen wie Rechtschreibeprüfungen, Spamfilter oder Antimalware.

[297] Zum Ganzen siehe Braun Binder et al. (FN 8), 15.

[298] Übersicht bei Braun Binder et al. (FN 8), 27.

[299] Braun Binder et al. (FN 8), 27.

aufgebaut sind, nicht als Chatbots – können aber durch erläuternde Chatbots ergänzt werden.[300]

Im Rahmen des Einsatzes als Auskunftssystem muss ein Bot zusätzlich auf entsprechende Inhalte zugreifen können, welche die Grundlage für eine korrekte Antwort bilden. Hierbei kommen verschiedene KI-Funktionen zusammen: das Programm muss zunächst die Eingabe der Nutzerinnen und Nutzern sprachlich korrekt interpretieren, diese erfolgreich mit den zur Verfügung stehenden Inhalten *vergleichen*, d.h. ein *intent matching* vornehmen, und eine sinnvolle und inhaltlich richtige Antwort formulieren. Während die Spracherkennung und -aufbereitung für den Dialog auf *machine learning* bzw. *natural language processing* (NLP) basiert, wird die Auswahl der Antwort typischerweise von einem Expertensystem übernommen.[301]

Die Bearbeitung von Personendaten durch Chatbot-Applikationen, welche durch ein öffentliches Organ bereitgestellt werden, stellt eine Bearbeitung von Personendaten i.S.v. § 3 Abs. 5 IDG ZH dar und unterliegt damit – innerhalb des Geltungsbereichs von § 2 IDG ZH – den Bestimmungen dieses Gesetzes. Als (vorerst) neue Technologie muss vor der Verwendung eines KI-basierten Chatbots, durch den Personendaten bearbeitet werden sollen, stets geprüft werden, ob die Applikation gemäss § 10 Abs. 2 IDG ZH bei der kantonalen Datenschutzbeauftragten zur Vorabkontrolle angemeldet werden muss.[302] In der Anwendung ergibt sich schliesslich aus dem Transparenzprinzip eine Pflicht, die KI-Funktionalität eines Chatbots gegenüber den Nutzerinnen und Nutzern auszuweisen.

Ob eine Bearbeitung von Personendaten durch den Chatbot stattfindet, ergibt sich aus dem Zweck sowie der angeschlossenen Systemarchitektur. Ein Chatbot, der anonym genutzt wird, bearbeitet keine Personendaten. Eine anonyme Nutzung liegt vermutungsweise vor, wenn kein Login in ein Onlineportal notwendig ist, um den Chatbot zu nutzen, und wenn im Rahmen des Chats keine Personendaten, wie etwa Kontaktdaten, erhoben werden.

---

[300] Für den Kanton Zürich vgl. Braun Binder et al. (FN 8), 27.

[301] Siehe zum Ganzen Michael Krähenbühl, Ein Chatbot als Rechtsberater – ein haftungsrechtlicher Albtraum für den Betreiber?, in: Jusletter IT vom 30.09.2021, N 10; dies scheint sich zu ändern, vgl. https://openai.com/blog/chatgpt/ (Aufruf 23.12.2022).

[302] Siehe IV.A.4.b.

Die Anonymität ist insofern relativ, als gewisse Randdaten der technischen Verbindung zwischen Nutzerin und Nutzer und Chatinterface durch den ISP gemäss Art. 2 BÜPF i.V.m. Art. 21 BÜPF während der gesetzlichen Frist von 6 Monaten gespeichert werden und insofern eine theoretische Möglichkeit der De-anonymisierung bzw. Identifikation der betreffenden Person über ihre IP-Adresse besteht.[303] Soweit aber der Chatbot sich in einem durch Login geschützten Bereich befindet, besteht zumindest gegenüber dem Portalsystem, in das er eingebettet ist, keine Anonymität. Konzeptionell könnten sämtliche im Portal vorhandenen Daten der betreffenden Person mit dem Chat verknüpft werden, wodurch diese grundsätzlich Personendaten darstellen. Eine solche Verknüpfung kann typischerweise dazu dienen, die Antworten für die betreffende Person zu verbessern.[304] In solchen Fällen ist der Chatbot als Teil des Portals zu sehen und sind die datenschutzrechtlichen Vorgaben für elektronische Portale[305] sinngemäss anzuwenden.

Schliesslich bearbeiten Chatbots die Personendaten der Nutzerinnen und Nutzern immer dann, wenn sich der Chat auf ein spezifisches, diese Person betreffendes Geschäft bezieht, etwa indem er das Stellen eines Gesuchs unterstützt oder eine medizinische Erstberatung anbietet. Inwiefern hier ein zusätzliches Risiko für die Persönlichkeit der Nutzerinnen und Nutzern besteht, hängt von der Architektur des Gesamtsystems ab, beispielsweise davon, ob die Chatinhalte gespeichert und anderen Applikationen zur Verfügung gehalten werden. Wäre dies der Fall, müssten solche Zugriffe die Voraussetzungen der Bekanntgabe gemäss § 16 bzw. 17 IDG ZH erfüllen.

---

[303] Bundesgesetz vom 18. März 2016 betreffend die Überwachung des Post- und Fernmeldeverkehrs (BÜPF; SR 780.1).

[304] Jörn von Lucke/Jan Etscheid, Wie Ansätze künstlicher Intelligenz die öffentliche Verwaltung und die Justiz verändern könnten, in: Walter Hötzendorfer/Christof Tschohl/Franz Kummer, International Trends in Legal Informatics – Festschrift für Erich Schweighofer, Bern 2020, 253 f.

[305] Vgl. dazu *privatim* – Konferenz der schweizerischen Datenschutzbeauftragten, Merkblatt für Online-Portale der öffentlichen Verwaltung, https://www.privatim.ch/wp-content/uploads/2018/10/031018_privatim_Merkblatt_Online-Portale.pdf (Abruf 28.01.2022).

2. Online-Übersetzung

Moderne Übersetzungsprogramme nutzen neuronale Netze, sog. *neural machine translation*, um anhand von Sprachmustern und deren Kontext die korrekte Übersetzung von einer Sprache in eine andere Sprache zu erstellen. Dazu werden sie mit Satzpaaren trainiert und anhand der Resultate optimiert.[306] Der Wechsel der grossen Anbieter wie Google oder Microsoft von regelbasierten Systemen auf Lernalgorithmen erfolgte soweit ersichtlich innerhalb der letzten fünf bis zehn Jahre.[307]

Wenn öffentliche Organe Dokumente, wie beispielsweise Verfügungen, amtliche Schreiben oder fremdsprachige Originaldokumente, die Personendaten enthalten, mit Hilfe von Übersetzern übersetzen lassen, so gilt dies je nach Inhalt des Textes als eine Bearbeitung von (besonderen) Personendaten. Wird die Onlineversion eines Übersetzungstools benutzt, gilt die Übermittlung des zu übersetzenden Originaltextes an den Server, der die KI-Funktion bereitstellt, als Bekanntgabe von (besonderen) Personendaten an Private und sind die Bestimmungen in § 16 Abs. 1 bzw. § 17 Abs. 1 IDG ZH anwendbar. Da es sich um eine Form von Cloud-Computing handelt, sind die entsprechenden Vorschriften zu beachten.[308]

---

[306] Zur Beschreibung der Funktionsweise von DeepL siehe Ruth Fulterer, Warum automatische Übersetzer so gut funktionieren, NZZ vom 29.01.2022.

[307] Barak Turovsky, Found in translation: More accurate, fluent sentences in Google Translate, Google Blog, 15.11.2016; Microsoft Translator, Microsoft brings AI-powered translation to end users and developers, whether you're online or offline, Microsoft Translator Blog, 18.04.2018.

[308] Vgl. dazu DSB Kanton Zürich, Merkblatt Cloud Computing, V 1.5/April 2021, https://docs.datenschutz.ch/u/d/publikationen/formulare-merkblaetter/merkblatt_cloud_computing.pdf (Abruf 01.06.2022).