Teil I
Künstliche Intelligenz und Selbstbestimmung

Prädiktive Privatheit: Kollektiver Datenschutz im Kontext von Big Data und KI

Rainer Mühlhoff

Zusammenfassung

Big Data und künstliche Intelligenz (KI) stellen eine neue Herausforderung für den Datenschutz dar. Denn diese Techniken werden dazu verwendet, anhand der anonymen Daten vieler Menschen Vorhersagen über Dritte zu treffen – etwa über Kaufkraft, Geschlecht, Alter, sexuelle Orientierung, ethnische Zugehörigkeit, den Verlauf einer Krankheit etc. Die Grundlage für solche Anwendungen „prädiktiver Analytik" ist ein Vergleich von Verhaltensdaten (z.B. Nutzungs-, Tracking- oder Aktivitätsdaten) des betreffenden Individuums mit den potenziell anonymisiert verarbeiteten Daten vieler Anderer anhand von Machine Learning Modellen oder einfacherer statistischer Verfahren. Der Artikel weist zunächst darauf hin, dass mit prädiktiver Analytik erhebliche Missbrauchspotenziale verbunden sind, welche sich als soziale Ungleichheit, Diskriminierung und Ausgrenzung manifestieren. Diese Missbrauchspotenziale werden vom geltenden Datenschutzrecht (EU DSGVO) nicht reguliert; tatsächlich findet die Verwendung anonymisierter Massendaten in einem weitestgehend rechtsfreien Raum statt. Unter dem Begriff "prädiktive Privatheit" wird ein datenschützerischer Ansatz vorgestellt, der den Missbrauchsgefahren prädiktiver Analytik begegnet. Die prädiktive Privatsphäre einer Person oder Gruppe wird verletzt, wenn anhand der Daten vieler anderer Individuen ohne ihr Wissen und gegen ihren Willen sensible Informationen über sie vorausgesagt werden. Prädiktive Privatheit wird sodann als Schutzgut eines kollektiven Ansatzes im Datenschutz formuliert und verschiedene Verbesserungen der DSGVO im Hinblick auf die Regulierung prädiktiver Analytik werden vorgeschlagen.

1. Einleitung

Eine der aktuell wichtigsten Anwendungen von KI-Technologie ist die sogenannte prädiktive Analytik. Unter diesen Begriff fasse ich datenbasierte Vorhersagemodelle, die über beliebige Individuen anhand verfügbarer

Daten Prognosen stellen. Diese Prognosen können sich auf zukünftiges Verhalten beziehen (z.B., was wird jemand wahrscheinlich kaufen?), auf unbekannte persönliche Attribute (z.B. sexuelle Identität, ethnische Zugehörigkeit, Wohlstand, Bildungsgrad) oder auf persönliche Risikofaktoren (z.B. psychische oder körperliche Krankheitsdispositionen, Suchtverhalten oder Kreditrisiko). Prädiktive Analytik ist brisant, denn neben den gesellschaftlich nutzbringenden Anwendungsmöglichkeiten besitzt die Technologie ein enormes Missbrauchspotenzial und ist aktuell gesetzlich kaum reguliert. Prädiktive Analytik ermöglicht die automatisierte und daher großflächige Ungleichbehandlung von Individuen und Gruppen beim Zugriff auf ökonomische und gesellschaftliche Ressourcen wie Arbeit, Bildung, Wissen, Gesundheitsversorgung und Rechtsdurchsetzung. Speziell im Kontext von Datenschutz und Antidiskriminierung muss die Anwendung prädiktiver KI-Modelle als eine neue Form von Datenmacht großer IT-Unternehmen analysiert werden, die im Zusammenhang mit der Stabilisierung und Hervorbringung von Strukturen der Diskriminierung, der sozialen Klassifizierung und der datenbasierten sozialen Ungleichheit steht.

Vor dem Hintergrund der enormen gesellschaftlichen Auswirkungen prädiktiver Analytik werde ich in diesem Kapitel argumentieren, dass wir im Kontext von Big Data und KI neue Ansätze im Datenschutz benötigen. Mit dem Begriff *prädiktive Privatheit* werde ich den Schutz der Privatheit einer Person oder Gruppe gegen ihre neuartige Form der Verletzbarkeit durch *abgeleitete* oder *vorhergesagte* Informationen fassen und normativ verankern. Die Anwendung prädiktiver Modelle auf Einzelindividuen, um damit Entscheidungen zu stützen, stellt eine Verletzung der Privatheit dar – die jedoch neuartigerweise weder durch „Datenklau" noch durch einen Bruch von Anonymisierung zustande kommt. Die Verletzung der prädiktiven Privatheit erfolgt mittels eines Abgleichs der über das Zielindividuum bekannten Hilfsdaten (z.B. Nutzungsdaten auf Social Media, Browserverlauf, Geo-Location-Daten) mit den Daten vieler tausend *anderer* Nutzer:innen. Prädiktive Analytik verfährt nach dem Prinzip des „pattern matching" und ist immer dort möglich, wo es eine hinreichend große Gruppe von Nutzer:innen gibt, welche die sensiblen Zielattribute über sich preisgibt, weil sie sich der Big Data-basierten Verwertungsweisen nicht bewusst sind oder denkt, „nichts zu verbergen zu haben". Deshalb markiert das Problem der prädiktiven Privatheit eine Grenze des im Datenschutz weit verbreiteten Individualismus und gibt dazu Anlass, kollektivistische Schutzgüter und kollektivistische Abwehrrechte im Datenschutz zu verankern.

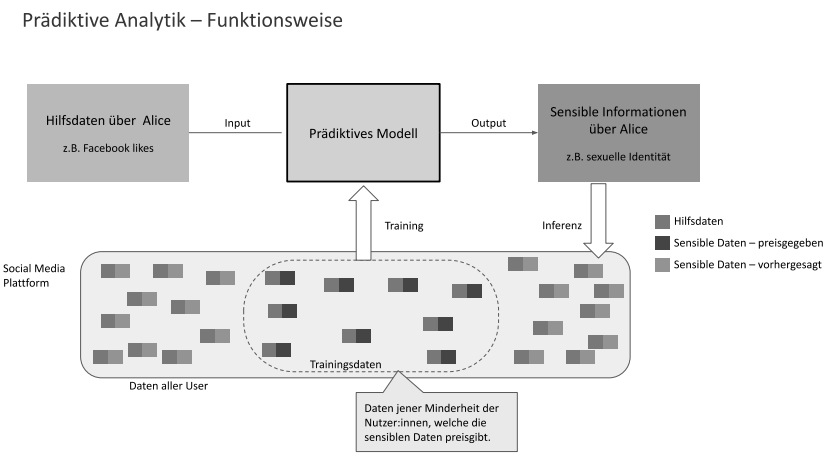
Eine solche kollektivistische Perspektive im Datenschutz berücksichtigt erstens, dass Individuen *nicht* in jeder Hinsicht frei entscheiden können sollten, welche Daten sie über sich gegenüber modernen Datenunternehmen preisgeben, denn die eigenen Daten können potenziell negative Auswirkungen auch auf andere Individuen haben. Zweitens bringt diese kollektivistische Perspektive zur Geltung, dass große Ansammlungen anonymisierter Daten vielen Individuen aufgrund der darin „lernbaren" Korrelationen zwischen sensiblen und weniger sensiblen Datenfeldern von Datenverarbeitenden *nicht* frei verarbeitet werden können sollten, wie es die aktuelle Rechtslage nach DSGVO bei anonymisierten Daten erlaubt. Drittens schließlich werde ich fordern, dass die Betroffenenrechte des Datenschutzes (Recht auf Auskunft, Rektifizierung, Löschung, etc.) kollektivistisch neu formuliert werden sollten, so dass betroffene Kollektive und das Gemeinwesen im Ganzen befugt wären, solche Rechte im Sinne des Gemeinwohls gegenüber datenverarbeitenden Organisationen auszuüben.

2. *Prädiktive Analytik*

Für den Gegenstand dieses Artikels ist es unerheblich, auf welchen Algorithmen und Verfahren ein prädiktives Modell konkret beruht. Es handelt sich bei prädiktiver Analytik um einen Container-Begriff, der sowohl Verfahren des maschinellen Lernens als auch einfachere statistische Auswertungen umfasst. Während prädiktive Analytik die technologische Disziplin bezeichnet, verweist „prädiktives Modell auf eine konkrete Manifestation dieser Technologie. Jedoch ist für ein adäquates Verständnis des Datenschutzproblems eine funktionale Charakterisierung prädiktiver Modelle hilfreich. Es handelt sich dabei um Datenverarbeitungssysteme, die als Input eine Reihe verfügbarer Daten über ein Individuum (oder einen „Fall") erhalten und als Ausgabe die Schätzung einer unbekannten Information, eine Klassifikation oder eine Entscheidung in Bezug auf das Individuum angeben (im Folgenden kurz „Zielvariable" genannt).

Die Inputdaten sind dabei typischerweise leicht verfügbare Hilfsdaten, zum Beispiel Trackingdaten, der Browser- oder Standort-Verlauf, oder Social Media Daten (Likes, Postings, Freund:innen, Gruppenmitgliedschaften). Bei der Zielvariablen handelt es sich typischerweise um schwer zugängliche oder besonders sensible Informationen über das Individuum, oder um eine Entscheidung über das Individuum in Bezug auf die Geschäftsvorgänge des Betreibers des prädiktiven Modells (zum Beispiel: zu welchem Preis dem Individuum eine Versicherung oder ein Kredit angeboten wird).

In der prädiktiven Analytik möchte man also anhand leicht zugänglicher Daten schwer zugängliche Informationen über Individuen abschätzen. Dazu vergleichen prädiktive Modelle den durch die Inputdaten gegebenen Fall nach Prinzipien der Mustererkennung mit Tausenden oder Millionen anderen Fällen, die das Modell zuvor während einer Lernphase (oder mittels anderer, statistischer Verfahren) ausgewertet hat. Häufig werden solche Modelle mit Verfahren des überwachten Lernens trainiert. Dazu wird eine große Menge Trainingsdaten benötigt, also ein Datensatz, in dem für eine Kohorte von Individuen beide Datenfelder, die Hilfsdaten und die Zieldaten, erfasst sind. Solche Datensätze fallen regelmäßig im Kontext sozialer Alltagsmedien an, zum Beispiel produziert die Teilmenge aller Facebook-Nutzer:innen, die in ihrem Profil explizit Angaben über ihre sexuelle Orientierung machen, einen Trainingsdatensatz für prädiktive Modelle zur Abschätzung der sexuellen Orientierung *beliebiger* Facebook-Nutzer:innen anhand der auf Facebook anfallenden Nutzungsdaten, wie zum Beispiel Facebook-Likes (s. Abb. 1).



Schematische Darstellung der Vorgehensweise prädiktiver Analytik

Wenn nur wenige Prozent der mehr als zwei Mrd. Facebook-Nutzer:innen Angaben über ihre sexuelle Orientierung machen, dann sind das einige Millionen Nutzer:innen. Das damit trainierbare prädiktive Modell kann die Plattform im nächsten Schritt dazu verwenden, die sexuelle Orientierung für alle anderen Facebook-Nutzer:innen abzuschätzen – auch für Nutzer:innen, die der Verarbeitung dieser Information nicht zustimmen

würden, diese Angabe bewusst nicht getätigt haben oder möglicherweise nicht wissen, dass das Unternehmen in der Lage ist, diese Informationen über sie abzuschätzen (vgl. auch Skeba und Baumer 2020).

Mediziner:innen von der University of Pennsylvania haben gezeigt, dass sich mit dieser Vorgehensweise anhand von Facebook-Daten beispielsweise vorhersagen lässt, ob eine Nutzer:in an Krankheiten wie Depression, Psychosen, Diabetes oder Bluthochdruck leidet (Merchant u.a. 2019). Facebook selbst hat bekannt gegeben, suizidale Nutzer:innen anhand ihrer Postings erkennen zu können (Goggin 2019). Eine viel beachtete Studie von Kosinski et al. zeigt, dass die Daten über Facebook-Likes dazu verwendet werden können, „eine Reihe höchst sensibler persönlicher Attribute vorherzusagen, darunter sexuelle Orientierung, Ethnie, religiöse und politische Ansichten, Persönlichkeitseigenschaften, Intelligenz, happiness, Suchtverhalten, Trennung der Eltern, Alter und Geschlecht" (Kosinski u.a. 2013).

Solche prädiktiven Analysen stoßen bei Versicherungs- und Finanzkonzernen auf großes Interesse, weil sie eine individuelle Risikobemessung jenseits der klassischen Credit Scores erlauben.[1] Auch im Personalmanagement werden solche prädiktiven Modelle verwendet, um zum Beispiel eine automatisierte Vorauswahl von Bewerber:innen bei Einstellungsvorgängen durchzuführen (O'Neil 2016, S. 108, 148). Zu den ersten und häufigsten Anwendungen prädiktiver Analytik gehört außerdem die gezielte Werbung (targeted advertising). So ist es einer US-amerikanischen Supermarktkette im Jahr 2011 gelungen, anhand der Einkaufsdaten, die über Rabattprogramme (customer loyalty cards) gesammelt werden, schwangere Kundinnen zu identifizieren (Duhigg 2012).

3. *Prädiktive Privatheit*

Prädiktive Analytik erlaubt es, unbekannte oder potenziell sensible Informationen über Individuen oder Gruppen anhand vermeintlich weniger sensibler und leicht verfügbarer Daten (Hilfsdaten) abzuschätzen. Dies ist mit modernen maschinellen Lernverfahren möglich, wenn viele Nutzer:innen einer digitalen Plattform die Datengrundlage schaffen, um Korrelationen zwischen den Hilfsdaten und den Zielinformationen zu ermitteln. Wir stehen hier also vor einer Situation, in der die *Datenfreigiebigkeit*

1 Siehe Lippert 2014 zum Beispiele der Firma ZestFinance sowie O'Neil 2016, Kap. 8 zu sogenannten "e-scores" als alternative credit scoring-Verfahren.

einer Minderheit von Nutzer:innen (zum Beispiel die prozentual wenigen Facebook-Nutzer:innen, die Angaben über ihre sexuelle Orientierung machen) den Standard der über *alle* Gesellschaftsmitglieder ableitbaren Informationen setzt. In der industriellen Verwendung prädiktiver Analytik im Kontext digital vernetzter Medien hat sich eine Praxis etabliert, in der die von Vorhersagen betroffenen Individuen in den meisten Fällen nicht informiert oder gefragt werden. Auch auf regulatorischer Ebene ist das Problem bisher im EU-Kontext weitestgehend unbeleuchtet: Insbesondere die DSGVO verfehlt es, die Herstellung oder Verwendung prädiktiver Modelle an geeignete Voraussetzungen zu knüpfen oder verantwortungsvoll einzuschränken.[2]

Vorhergesagte Informationen über Individuen oder Gruppen ermöglichen neben vorstellbar nutzbringenden Anwendungen zahlreiche schädliche und missbräuchliche Verwendungsweisen, welche mit Diskriminierung, Ungleichbehandlung und weiteren Grundrechtseingriffen der Betroffenen verbunden sein können. Um einen Schutz vor der missbräuchlichen Verwendung abgeschätzter Informationen normativ zu verankern – zunächst ethisch, sodann politisch und rechtlich –, möchte ich deshalb ein neues Schutzgut konstruieren. In direkter Antwort auf die Gefahrenlage der prädiktiven Analytik schlage ich dazu den Begriff der *prädiktiven Privatheit* vor (vgl. Mühlhoff 2020b, Mühlhoff 2021).[3] Prädiktive Privatheit lässt sich am besten negativ definieren, indem fixiert wird, wann sie *verletzt* ist:

> Die prädiktive Privatheit einer Person oder Gruppe wird verletzt, wenn sensible Informationen ohne ihr Wissen oder gegen ihren Willen über sie vorhergesagt werden, und zwar in solcher Weise, dass daraus die Ungleichbehandlung eines Individuums oder einer Gruppe resultieren könnte. (vgl. Mühlhoff 2021)

2 In diesem Sinne argumentiert auch Roßnagel (2018, S. 365–367) für eine Modernisierung der DSGVO angesichts der Gefahr durch prognostizierte Informationen.

3 Es gibt verwandte Begriffsvorschläge, die in eine ähnliche Richtung zielen. Darunter ist insbesondere „categorical privacy" von Vedder 1999 zu erwähnen, sowie die jüngere Debatte zu „group privacy" im Kontext von Big Data (Floridi 2014; Taylor u.a. 2016; Mittelstadt 2017) und „inferential privacy" (Loi und Christen 2020). Auch die Arbeiten zu einem „right to reasonable inferences" von Sandra Wachter und Brent Mittelstadt (Wachter und Mittelstadt 2018) schlagen eine ähnliche Richtung ein. Eine Auseinandersetzung mit den Gemeinsamkeiten und Unterschieden dieser Begriffe zu dem hier konstruierten Konzept der prädiktiven Privatheit findet sich in Mühlhoff 2021.

Hinter dieser sehr allgemeinen Begriffsbildung steht zunächst das Anliegen, angesichts der durch KI veränderten technologischen Situation auch die gesellschaftliche und kulturelle Auffassung von Privatheit anzupassen und zu erweitern. Denn bisher hat man sich unter Verletzungen von (informationeller) Privatheit meist einen nicht-autorisierten Zugriff auf die private „Informationssphäre" oder Eingriffe in die informationelle Selbstbestimmung des Einzelnen vorgestellt, durch die dem Datensubjekt Informationen „entwendet" werden, die es nicht über sich preisgeben wollte.[4] Zwar werden bei einer Verletzung prädiktiver Privatheit ebenfalls Informationen gewonnen, die das betroffene Subjekt mutmaßlich nicht preisgeben möchte, jedoch geschieht dies nicht auf dem Weg der „Entwendung" oder des Eindringens in eine private Sphäre (diese Metapher ist in der neuen technologischen Situation längst nicht mehr adäquat, siehe dazu auch Ruschemeier in diesem Band). Vielmehr werden die Informationen über das Datensubjekt abgeschätzt, und zwar anhand eines Vergleichs mit den Daten, die viele *andere* Datensubjekte über sich preisgeben. Hierbei kommt es darauf an, dass diese Verletzungen prädiktiver Privatheit *nicht* von der Genauigkeit oder Korrektheit der geschätzten Informationen abhängen, sondern allein davon, dass diese Informationen das Potenzial einer Ungleichbehandlung der betroffenen Individuen oder Gruppen bergen. Das heißt, es wäre nach der ethischen und datenschützerischen Norm der prädiktiven Privatheit nicht automatisch legitim, Menschen anhand von über sie vorhergesagten Informationen unterschiedlich zu behandeln, bloß weil die Vorhersagen bestimmte Anforderungen der Genauigkeit erfüllt.[5]

---

4 Zum Begriff Privatheit werden häufig zwei oder mehr Haupttraditionen unterschieden, die im anglophonen Raum als „nonintrusion theory" und als „control theory" of privacy in Erscheinung treten (vgl. Tavani 2007, der insgesamt vier Kategorien unterscheidet). Das Verständnis von Privatheit als Nicht-Intrusion betont dabei eine (oder sogar mehrere geschachtelte) private Sphäre(n) jedes Individuums, die vor Einblicken und Eingriffen zu schützen sei(en); Kontrolltheorien setzen dagegen weniger auf die Abgeschiedenheit für sich, sondern auf das Vermögen des Individuums, effektiv und potenziell differenziert darüber zu verfügen, wer welchen „Zugang" zu den eigenen persönlichen Informationen hat.

5 In diesem Punkt weicht die ethische und datenschützerische Norm der prädiktiven Privatheit von der zu kurz greifenden Forderung eines „right to reasonable inferences" von Sandra Wachter und Brent Mittelstadt (vgl. Wachter und Mittelstadt 2018) ab.

4. Ein neues Datenschutzproblem: Drei Angriffstypen

Die Abschätzung persönlicher und potenziell sogar sensibler Informationen über Individuen anhand von Massendaten stellt ein neues dominantes Angriffsszenario im Datenschutz unter den Bedingungen unzureichend regulierter KI- und Big Data-Technologie dar. Dies ist ein Angriffsszenario, das erst seit etwa zehn Jahren virulent ist. Um die neue Qualität dieser Herausforderung und die entsprechend neuen Schutzbedarfe herauszuarbeiten, lohnt sich eine vergleichende Gegenüberstellung des neuartigen mit zwei älteren Angriffsszenarien, die in den Diskursen über Datenschutz und Privatheit der letzten Jahrzehnte jeweils zu ihrer Zeit eine prominente Rolle gespielt haben (siehe zur Übersicht Tab. 1).

Tab. 1: Qualitativer Vergleich von Angriffsszenarien, die im öffentlichen Diskurs um Datenschutz zu verschiedenen Zeiten eine dominante Bedrohung darstellen. Die jeweils anderen Angriffsszenarien waren zu jeder Zeit ebenfalls denkbar, aufgrund der technologischen Entwicklung besitzt die Relevanz der Szenarien jedoch unterschiedliche zeitliche Schwerpunkte.

	Typ 1: Intrusion	Typ 2: Re-Identifikation	Typ 3: Vorhersage
Virulent seit	1960 ff.	1990 ff.	2010 ff.
Mittel	Hacking, Datenlecks, Bruch von Verschlüsselung etc.	De-Anonymisierung mittels statistischer Attacken oder Hintergrundwissen	Abschätzung unbekannter Informationen anhand des Abgleichs mit kollektiven Datenbeständen
Angriffsziel	persönliche Daten	Anonymität in Datensätzen	Gleichheit der Behandlung, Fairness
Schutz	Datensicherheit	Differential Privacy, Federated ML	Predictive Privacy

Typ 1: Intrusion

Den Urtypus eines Gefahrenszenarios im Datenschutz kann man als Intrusion bezeichnen. Damit eng zusammen hängt die zielgerichtete, auf konkrete Individuen oder Gruppen begrenzte Überwachung. Die Gefahr der gewaltsamen Entwendung von Daten aus mehr oder weniger gesicherten, jedenfalls nicht-öffentlichen Zonen ist tragend für Debatten zum Datenschutz spätestens seit dem Verbreiten der elektronischen Datenverarbeitung in den 1960er Jahren. Das Mittel dieser Form der Verletzungen von Privatheit ist der klassische „Datenklau" als gezielter Akt der Entwendung

von Daten über technische oder organisatorische Schutzbarrieren hinweg. Obwohl die wichtigste potenzielle Angreiferin immer die datenverarbeitende Organisation selbst ist, wird dieser Angriffstypus in der populären Imagination oft mit *hacking* und Cyberattacken durch Kriminelle oder Geheimdienste in Verbindung gebracht. Das Angriffsziel der intrusiven Verletzung von Privatsphäre sind *konkrete* sensible Datenbestände (über Einzelpersonen, Kohorten, Firmen, staatliche Prozesse, …), die den Angreifenden eigentlich nicht zugänglich sein sollten.

Typ 2: Re-Identifikation

Eine zweiter Angriffstyp wird als Re-Identifikation bezeichnet. Dieser Typus wurde erst in den 1990er Jahren virulent, nachdem durch die Digitalisierung des Gesundheitswesens – zum Beispiel der Abrechnungsvorgänge mit Versicherungen oder der Patientenverwaltung in Krankenhäusern – umfassende digitale Datenbestände über die Prozesse der medizinischen Versorgung verfügbar wurden. Es kam dann die Idee auf, diese Daten für statistische Auswertungen im Rahmen wissenschaftlicher Forschung verwenden zu wollen. Dazu stellte sich die Frage, wie man die Einträge in solchen Datenbanken anonymisieren könnte, um sie dann zu veröffentlichen.

In einem mittlerweile legendären Fall hat der US-Bundesstaat Massachusetts Ende der 1990er Jahre die Krankenhaus-Behandlungsdaten seiner ca. 135.000 staatlichen Bediensteten und ihrer Angehörigen in vermeintlich anonymisierter Form der Forschung zugänglich gemacht. Die Anonymisierung der Datensätze erfolgte, indem Name und Anschrift, sowie die Sozialversicherungsnummer aus den Datensätzen herausgelöscht wurden. Latanya Sweeney, damals Informatik-Studentin am MIT, konnte mit einer linkage-Attacke den Datensatz des damaligen Gouverneurs von Massachusetts, William Weld, in den anonymisierten Daten identifizieren und seine Krankenakte rekonstruieren (Sweeney 2002; Ohm 2010). Dieser Fall hat in Wissenschaft und Politik eine intensive Diskussion über Grenzen und Machbarkeit von Anonymisierung ausgelöst. Die Frage der „sicheren" Anonymisierungsverfahren wird davon ausgehend bis heute diskutiert; jeweils aktuelle Vorschläge für Anonymisierungsverfahren in der Informatik werden immer wieder einige Zeit später durch spektakuläre Angriffe

gebrochen[6]; es ist klar geworden, dass „Anonymität" ein komplexer, nicht absolut definierbarer Begriff ist, der stets von Annahmen in Bezug auf das Hintergrundwissen der Angreifer:in und der statistischen Verteilung der Daten im zu anonymisierenden Datensatz abhängt. Auf Verfahren der Anonymisierung lastet die Anforderung, dass ein heute verwendetes Verfahren alle zukünftigen Angriffstechniken antizipieren und alle möglichen Konfigurationen von Hintergrundwissen zukünftiger Angreifer:innen abdecken muss.[7]

Die Gefahr der Re-Identifizierbarkeit von Individuen in anonymisierten Datensätzen wurde seit den 1990er Jahren zu einem zweiten, viel diskutierten Gefahrenszenario im Datenschutz. Die Diskussion hatte insbesondere spürbaren Einfluss auf die Datenschutzgesetzgebung im Kontext medizinischer Daten, in den USA zum Beispiel auf den *Health Information Portability and Accountability Act* (HIPPA) von 1996. Für die Zwecke des vorliegenden Kapitels kommt es darauf an, auf die qualitative Differenz zum Angriffstyp der Intrusion (und der Prädiktion) hinzuweisen. Im Unterschied zum Datenklau ist das Ziel von Re-Identifikationsattacken ein Bruch der Anonymität. Auch wenn hier ebenfalls sensible Daten über Einzelne oder definierte Kohorten ermittelt werden, ist das etwas anderes als intrusiver Datenklau, da die zugrundeliegenden Daten zuvor bewusst veröffentlicht wurden, jedoch mit dem Versprechen, dabei nichts über Einzelindividuen, sondern nur über statistische Zusammenhänge preiszugeben.

Typ 3: Prädiktion

Mein Argument ist nun, dass auch Re-Identifikation heute schon nicht mehr als der wichtigste und dominante Angriffstypus im Datenschutz gelten kann. Das Prinzip der Vorhersage von unbekannten Daten mittels Big Data und KI-Technologie löst die Gefahr der Re-Identifizierung freilich nicht auf (genauso wenig wie die Gefahr der Intrusion). Die Gefährdung

6 Vgl. Ohm 2010 und besonders spektakulär: Die Re-Identifikation von Netflix-Usern in einer pseudonymisiert publizierten Datenbank aus Film-Bewertungen (Narayanan und Shmatikov 2008) oder die Rekonstruktion des Familiennamens anhand anonym vorliegender Genom-Daten (Gymrek u.a. 2013).

7 Die Bundesrepublik Deutschland hat im Dezember 2019 im Rahmen des „Digitale-Versorgung-Gesetz" erst die Zusammenführung der Behandlungsdaten aller ca. 70 Millionen gesetzlich Krankenversicherten zu einer zentralen Forschungsdatenbank beschlossen, vgl. *Bundesgesetzblatt Teil I*, Nr. 49 vom 18.12.2019, S. 2562. Vgl. dazu auch Mühlhoff 2020a.

durch unregulierte prädiktive Analytik übertrifft jedoch beide klassische Angriffsszenarien bei Weitem hinsichtlich Reichweite und Skalierbarkeit. Ist ein prädiktives Modell einmal erstellt – und hierfür gibt es zur Zeit keine wirksamen rechtlichen Beschränkungen –, kann es auf Millionen Nutzer:innen automatisiert und nahezu ohne Grenzkosten angewandt werden. Die Datenfreigiebigkeit der oft privilegierten Gruppe von Nutzer:innen, die vorbehaltlos die Trainingsdaten für prädiktive Analysen bereitstellen (z.B. Gruppe der Facebook-Nutzer:innen, die explizite Angaben über ein sensibles Attribut machen, siehe oben), setzen den Standard des über beinahe *alle* Menschen ermittelbaren Wissens, solange prädiktive Analytik-Technologie nicht reguliert wird.

Dies stellt eine qualitativ neue Gefahrenlage im Datenschutz dar, denn das Mittel der Verletzung prädiktiver Privatheit ist weder der Datenklau noch der Bruch eines Anonymisierungsversprechens. Die Gefährdung durch prädiktive Analytik unterscheidet sich von den älteren Angriffsszenarien in drei Hinsichten: hinsichtlich ihres Ursprungs beruht sie auf der Verfügbarkeit *kollektiver* Datenbestände; hinsichtlich der verübenden Instanz ist sie genau jenen Akteuren vorbehalten, die über aggregierte kollektive Datenbestände verfügen; und hinsichtlich ihrer Effekte zeigt sie nicht allein individuelle sondern vielmehr *gesamtgesellschaftliche* Auswirkungen. Dies bedeutet erstens eine *kommerzielle Zentralisierung* der von prädiktiver Analytik ausgehenden Datenmacht bei wenigen großen Unternehmen. Zweitens liegt der potenzielle Schaden von Verletzungen prädiktiver Privatheit nicht nur in der Abschätzung von Informationen über gezielt ausgewählte Einzelindividuen, sondern in der automatischen und synchronen Abschätzung dieser Informationen über sehr große Nutzer:innen-Kohorten, die eine breite Mehrheit unserer Gesellschaften betreffen. Im Zentrum der Verletzung prädiktiver Privatheit steht also nicht Spionage, die sich auf Einzelne richtet, sondern automatisierte und serienmäßige Ungleichbehandlung von Menschen in der Breite der Gesellschaft. Diese Ungleichbehandlung ist ein *struktureller Faktor* insofern sie sich nicht nur auf Einzelindividuen richtet, sondern auf uns alle in der Interaktion mit automatisierten Systemen, zum Beispiel, wenn uns unterschiedliche Preise für Versicherungen angeboten werden, automatisiert entschieden wird, wer für ein Jobinterview eingeladen wird usf. Das Angriffsziel der Verletzung prädiktiver Privatheit ist somit die Gleichheit und Fairness der gesellschaftlichen Behandlung. Das Schutzgut, das hier verletzt wird, ist im Unterschied zu den anderen Angriffstypen erst in einer kollektivistischen Perspektive erkennbar.

5. Prädiktive Privatheit als neues Schutzgut

Der Problemkomplex der prädiktiven Privatheit stellt eine neuartige und aktuell wohl die bedeutsamste Herausforderung für den Datenschutz dar. Um den Schutz prädiktiver Privatheit im vollen Sinne als ein Problem des Datenschutzes zu erkennen, ist es erforderlich, das Denken des Datenschutzes von der Perspektive individueller Schutzansprüche zu lösen, die er qua Konstruktion aus der Bindung an den Schutz der Grundrechte erbt, die stets Individualrechte sind. Das Schutzgut prädiktiver Privatheit ist jenseits der Perspektive individueller Rechte in einer *kollektivistischen ethischen Blickweise* zu konstruieren, die auf der Wertsetzung beruht, das Kollektiv gegenüber den Individuen zu priorisieren. Zwar ist es auch eine Gefahr für das Individuum, aufgrund abgeschätzter Informationen nachteilig behandelt zu werden. Doch diese Gefahr allein ist nichts Neues. Schon lang bevor es KI-basierte prädiktive Analytik gab, haben Bankberater:innen anhand von Bauchgefühl, Erfahrung und Vorurteilen über Kreditwürdigkeit entschieden, Ärzt:innen anhand persönlicher Einschätzungen Behandlungsprogramme priorisiert oder Human Resource Manager:innen bei Einstellungsvorgängen die Performanz von Bewerber:innen prognostiziert.

Die neue Qualität der Gefährdung durch prädiktive Analytik liegt weniger darin, dass Informationen über eine konkrete Person X gegen ihren Willen oder ohne ihr Wissen prognostiziert werden, sondern darin, dass der Platzhalter „X" *jede beliebige Person* zugleich repräsentieren kann. Die für prädiktive Analytik verwendeten Technologien können Prognosen zeitgleich und auf großer Skala über *beliebige* Personen X stellen. Prädiktive Analytik-Technologien werden dort entwickelt, wo es um die algorithmische Verwaltung von Nutzer:innen-Kohorten und Populationen im Ganzen geht (Mühlhoff 2020c), also um die Sortierung großer Menschenmengen, nicht um die Überwachung oder das Ausspionieren von Einzelpersonen. Das Wesen der Verletzung prädiktiver Privatheit liegt somit nicht darin, in *eine* private „Sphäre" einzudringen, sondern eine *strukturelle* Neukonfiguration von Privatheit in der digitalen Gesellschaft zu bewirken – auf die mit der Konstruktion eines Schutzgutes der prädiktiven Privatheit reagiert werden muss. Diese Neukonfiguration betrifft die technologisch realistischen Erwartungen an Privatheit, die Skalierungsfähigkeit der Methoden zur Unterwanderung von Privatheit und die politischen Werte, die mit Privatheit auf dem Spiel stehen: Im Kontext von KI

und Big Data betreffen diese verstärkt die Fragen von Gleichheit, Fairness und Anti-Diskriminierung.[8]

Die Missbrauchsgefahren prädiktiver Analytik sind also noch nicht erkannt, wenn man nur auf die Belange eines Einzelnen schaut – etwa auf Selbstbestimmungs- und Kontrollrechte in Bezug auf die eigenen Daten, inklusive des Rechts auf informationelle Selbstbestimmung. Der Blick ist auf die strukturelle Machtasymmetrie zwischen Individuen und datenverarbeitenden Organisationen zu richten. Eine *positive* Bestimmung des Schutzgutes „prädiktive Privatheit" geht somit auch über den Gehalt der oben zunächst eingeführten *negativen* Bestimmung der „Verletzung prädiktiver Privatheit eines Individuums oder einer Gruppe" hinaus. Es geht bei prädiktiver Privatheit darum, eine Technologie zu regulieren, die es ermöglicht, jedes *beliebige* Individuum – also potenziell jede:n von uns in seiner prädiktiven Privatheit, und somit unsere Gesellschaft in ihren Werten der Gleichheit, Fairness und Menschenwürde – zu verletzen. Indem der Blick von den Schutz- und Abwehransprüchen des Einzelnen auf positiv bestimmbare Werte der Gleichbehandlung verschoben wird, gelangt der Schutz des Gemeinwohls in den Mittelpunkt und sodann geht es bei prädiktiver Privatheit tatsächlich um den Ausgleich einer Machtasymmetrie, die daraus resultiert, dass Technologie auf neue Weise an der Stabilisierung und Produktion sozialer Unterschiede und Diskriminierung mitwirkt.

Das Datenschutzanliegen der prädiktiven Privatheit betrifft in besonderer Weise eine kollektive Dimension des Datenschutzes, die durch neue Technologie auf neue Weise virulent wird. Diese Dimension kann gestärkt werden, indem prädiktive Privatheit – also der Schutz *der Gesellschaft* vor der Macht einzelner Akteure, Prädiktionen über beliebige Individuen zu stellen – als Schutzgut im Sinne des Gemeinwohls konstruiert wird. Diese kollektivistische Bestimmung eines erweiterten Schutzguts des Datenschutzes ist eine direkte Antwort auf das Missbrauchspotenzial prädiktiver Analytik, welches eben Kollektive und nicht allein Individuen betrifft. Dieses Missbrauchspotenzial der anhand vieler Datenpunkte erstellten prädiktiven Modelle ist von struktureller Qualität, es betrifft synchron und potenziell *alle*. Zwar ist ein individueller Schaden durch *prädiktive Verletzungen von Privatheit* spürbar. Eine *Verletzung prädiktiver Privatheit*

8 Um Antidiskriminierung geht es hierbei nicht nur in Bezug auf die Fragen möglicher Verzerrungen (bias) in Prognosesystemen, sondern insofern solche Systeme, auch wenn sie „unverzerrt" sind (falls das möglich sein sollte), erhebliche gesellschaftliche Folgen haben.

– man beachte den subtilen Unterschied beider Begriffe – jedoch bedeutet in gesamtgesellschaftlicher Perspektive eine Zementierung oder Neuproduktion sozialer Ungleichheit und datenbasierter sozioökonomischer Selektion, die einen Gemeinschaftsschaden darstellt. Prädiktive Privatheit benennt folglich einen Schutzanspruch des Gemeinwesens; das Schutzgut der prädiktiven Privatheit stützt die Grundwerte freier, egalitärer und demokratischer Gesellschaften.

Neben der kollektivistischen Konstruktion des Schutzgutes prädiktiver Privatheit gibt es einen weiteren ethischen Punkt zu bedenken: Die *Verletzung* prädiktiver Privatheit ist durch eine kollektive „Täterschaft" bzw. Verursachungsstruktur gekennzeichnet. Denn prädiktive Analysen sind nur dort möglich, wo eine hinreichend große Menge Nutzer:innen bei der Benutzung digitaler Dienste sensible Daten im Zusammenhang mit den Hilfsdaten zur Verfügung stellt und wo es Plattformunternehmen und anderen wirtschaftlichen Akteuren rechtlich gestattet ist, diese Daten (potenziell auch in anonymisierter Form) zu aggregieren und damit prädiktive Modelle zu trainieren. Der Schutz prädiktiver Privatheit erfordert sodann nichts weniger als den Bruch mit einem tief verankerten liberalistischen Denken westlicher Bevölkerungen in Bezug auf die Ethik der täglichen Datenproduktion bei der Verwendung digitaler Dienste. Er fordert ein breit geteiltes Bewusstsein dafür, dass die eigenen Daten potenziell anderen schaden – und dass deshalb ein moderner Datenschutz noch nicht gewährleistet ist, wenn jede einzelne Nutzer:in die Kontrolle darüber erhält, welche personenbezogenen Daten bei der Benutzung eines Service von ihr erfasst werden. Das erforderliche kollektive Bewusstseins über diesen Umstand könnte durch die umgekehrte Einsicht vermittelt werden, dass persönliche Informationen über *einen selbst* mittels prädiktiver Analytik anhand der Daten abgeschätzt werden, die viele andere Menschen mehr oder weniger wissentlich und freiwillig über sich preisgeben (und die von Plattformunternehmen völlig legal gesammelt werden).[9]

9 In diesem Vorschlag für eine Rhetorik zur öffentlichen Vermittlung des Anliegens prädiktiver Privatheit findet jetzt wieder eine pragmatische Rückübersetzung des kollektivistischen Anliegens eines Schutzes vor Verletzungen prädiktiver Privatheit in die Terminologie der drohenden *prädiktiven Verletzung (individueller) Privatheit* statt. Dieses Changieren zwischen Gemeinwohl- und Individualwohlterminologie sehe ich ganz pragmatisch im Sinne der Überzeugungskraft des Arguments auch bei Menschen, die in ihrem politischen Empfinden weniger kollektivistisch gestimmt sind.

Diese elementaren Überlegungen zu den gesellschaftlichen Externalitäten der eigenen Datenpraxis,[10] die sich aus der technische Grundstruktur prädiktiver Analytik ergeben, zeigen eine gewichtige Grenze der Rechtsgrundlage der Einwilligung auf, die im Kontext sozialer Medien eines der relevantesten, von der DSGVO bereitgestellten Vehikel darstellt, um große Datenmengen bei Plattformunternehmen zu aggregieren. Hier wird nämlich klar: wenn eine Nutzer:in nach einer Einwilligung gefragt wird, dann trifft sie eine Entscheidung auch für viele andere Menschen, die anhand dieser Daten diskriminiert werden können, sofern auch noch einige weitere Nutzer:innen solche Daten über sich preisgeben, was, wie die zahlreichen Beispiele der sozialen Medien zeigen, überwiegend der Fall ist. In unserer aktuellen rechtlichen und regulatorischen Situation, in der Bau und Verwendung prädiktiver Modelle *nicht* reguliert sind, sind *individuelle* Einwilligungsentscheidungen von *über-individueller*, nicht auf das Datensubjekt selbst beschränkter Tragweite.

Hierbei ist verschärfend zu beachten, dass für das Training prädiktiver Analysen *anonyme Daten ausreichen.* Man benötigt dazu nur die Korrespondenz von Hilfsdaten und Zielinformationen, zum Beispiel von Facebook-Likes und Informationen über Krankheitsdispositionen; die Trainingsdaten für prädiktive Analysen müssen jedoch keine *identifizierenden* Datenfelder enthalten. Anonymisierungsversprechen werden deshalb routinemäßig in Stellung gebracht, um die Einwilligungsbereitschaft von Nutzer:innen zu befördern; für Big Data Geschäftsmodelle, die auf prädiktiver Analytik beruhen, ist das unschädlich.[11] In Situationen, in denen Nutzer:innen bei der Benutzung eines digitalen Service nicht anonym auftreten, ist davon auszugehen, dass Plattformunternehmen es vermeiden können, das Training prädiktiver Analysen als Datenverarbeitungszweck aufzuführen. Denn die Unternehmen können die Daten ihrer Nutzer:innen nach der Erfassung direkt anonymisieren und dann der weiteren Verwertung für das Training prädiktiver Modelle zuführen. Die anonymisierten Daten fallen nicht in den Schutzbereich der DSGVO und können – insbesondere in aggregierter Form – frei verwendet werden. Sie können auch auf unbestimmte Zeit gespeichert und erst später für prädiktive Analytik ver-

10 In diesem Sinne auch das Konzept „data pollution", siehe Ben-Shahar 2019.

11 Siehe dazu insbesondere die Forschung zu differential privacy in machine learning, vgl. Abadi u.a. 2016; Dwork 2006. Warum moderne Anonymisierungsverfahren wie differentially private machine learning prädiktive Privatheit eher gefährden als schützen, habe ich auch hier argumentiert: https://rainermuehlhoff.de/differential-privacy-in-machine-learning-is-a-data-protection-challenge/

wendet werden.[12] Schließlich ist zu bedenken, dass die trainierten prädiktiven Modelle selbst abgeleitete, höchst aggregierte, anonymisierte Daten darstellen,[13] die somit nicht in den Schutzbereich der DSGVO fallen und insbesondere ohne effektive Datenschutzhürden verkauft und zirkuliert werden.

6. Regulierungsvorschläge

Durch prädiktive Analytik und KI-Technologie ist das Missbrauchspotenzial anonymisierter Massendaten in den vergangenen 15 Jahren erheblich gestiegen (vgl. auch Tabelle 1). In der aktuellen rechtlichen Situation sind Herstellung und Verwendung prädiktiver Modelle weitestgehend unreguliert, so dass das Missbrauchspotenzial eine potenziell gravierende gesellschaftliche Kraft darstellt, die sozio-ökonomische Ungleichheit und Diskriminierungsmuster stabilisieren und produzieren kann.

Ich möchte nun einige Vorschläge in die Diskussion einbringen, wie man dem Missbrauchspotenzial prädiktiver Analytik vorbeugen und das kollektivistische Schutzgut der prädiktiver Privatheit im Kontext der EU DSGVO stärken könnte. Nach dem Grundsatz des Datenschutzes als eines „Vorfeldschutzes" (Britz 2010; Lewinski 2009; Lewinski 2014) kommt es hierbei darauf an, die Schutzwirkung als *präventive* Absicherung von Gleichheit und Fairness in der Behandlung durch privatwirtschaftliche und öffentliche datenverarbeitende Organisationen zu betrachten. Es geht um den Ausgleich einer Machtasymmetrie zwischen Gesellschaft und Organisationen; diese besteht bereits in der *potenziellen* und *drohenden* Verletzung prädiktiver Privatheit, sowie in der unterschiedlich verteilten Vulnerabilität verschiedener Gruppen und Akteure in Bezug auf das Missbrauchspotenzial anonymisierter Massendaten und prädiktiver Modelle.

Die Schutzwirkung einer Datenschutzregulierung, die wirksam die Missbrauchsgefahren prädiktiver Analytik einschränkt, kann somit nicht allein auf die Schultern der Abwehrrechte betroffener Einzelindividuen gelegt werden. Denn ein solcher Ansatz läuft dem tatsächlichen Verlet-

12 Das liegt daran, dass anonyme Daten nicht in den Gegenstandsbereich der DSGVO fallen und zum Beispiel das Recht auf Löschung im Kontext der DSGVO auch durch Anonymisierung der Daten erfüllt werden kann, vgl. dazu Abschn. 6 unten.

13 Dies setzt voraus, dass etablierte Anonymisierungsverfahren dabei eingesetzt werden, die unter Stichworten wie differential privacy und differentially private machine learning seit fünfzehn Jahren dafür entwickelt werden.

zungsvorfall stets hinterher. Die Wirksamkeit solcher Instrumente wird im vorliegenden Kontext noch dadurch abgeschwächt, dass der Verletzungstatbestand aus der individuellen Opferperspektive oft schwer identifizierbar oder gar nachweisbar ist. Aus Sicht des betroffenen Einzelindividuums ist außerdem der nachweisbare Schaden durch prädiktive Verletzungen von Privatheit häufig geringwertig, so dass der individuelle Rechtsweg wenig Erfolg verspricht; durch einen Streueffekt, der dadurch zustande kommt, dass die entsprechenden Techniken automatisiert auf tausende Individuen parallel angewandt werden, kann der gesamtgesellschaftliche Schaden jedoch erheblich sein (vgl. Ruschemeier 2021).

Statt einer Betonung individueller Schutzrechte müssen deshalb auf Ebene (a) des Schutzgutes, (b) der Abwehrrechte und (c) des Prozessrechts jeweils Strukturen geschaffen werden, die *kollektives Handeln* – sowohl von Gruppen als auch des Gemeinwesens insgesamt – gegenüber Datenunternehmen ermöglichen.

6.1 Aktuell fehlende Regulierung

6.1.1 Herstellung prädiktiver Modelle

Zunächst stellt sich die Frage, warum die DSGVO nicht effektiv die *Herstellung* prädiktiver Modelle reguliert. Ein Grund liegt in der individualistischen normativen Konzeption der DSGVO, die letztlich in der Konzeption der Grundrechte als Individualrechte gründet. Darüber hinaus ist es ein Kennzeichen des öffentlichen Diskurses, der Rechtsprechung und der Geschäftspraktiken, die sich rund um die DSGVO entspinnen, dass sowohl Schutzgut als auch Abwehrrechte des Datenschutzes stets auf die Relation des Individuums zu seinen eigenen Daten zugespitzt werden. Die Auslegung lautet meist: Die Souveränität der Einzelnen in Bezug auf die Verwendung ihrer (persönlichen) Daten muss gewahrt bleiben; jede:r wird in Bezug auf seine eigenen Daten um Zustimmung gebeten oder bekommt eine andere Rechtsgrundlage erklärt. Verletzungstatbestände beziehen sich folglich darauf, dass eine Einzelperson geltend macht, dass personenbezogene Daten *über sie* auf eine Weise verarbeitet wurden, die durch die beanspruchte Rechtsgrundlage nicht gedeckt war. Insbesondere die Betroffenenrechte wie das Recht auf Auskunft (Art. 15), Rektifizierung (Art. 16), Löschung (Art. 17), Einschränkung der Verarbeitung (Art. 18), und Portabilität (Art. 20), sind in der DSGVO als individuelle Rechte gefasst, die stets nur das Individuum in Bezug auf seine eigenen Daten ausüben kann.

Ein weiterer, hiermit zusammenhängender Grund für die schwache Regulierung prädiktiver Analytik durch die DSGVO liegt darin, dass sich die DSGVO auf „personenbezogene Daten" (Art. 4 (1)) bezieht und anonyme Daten nicht betrifft. Die Abgrenzung personenbezogener und anonymer Daten ist im Kontext von KI und Massendaten jedoch veraltet. Dies jedoch *nicht* bloß, weil Anonymisierung gebrochen werden kann,[14] sondern weil mittels prädiktiver Analytik die anonymisierten Daten *vieler* Individuen dazu verwendet werden können, sensible und „persönliche" Daten über wiederum *andere* Individuen abzuschätzen. In der juristischen und unternehmerischen Praxis wird die Unterscheidung zwischen personenbezogenen und anonymen Daten oft nur im „Input Stadium" der Datenverarbeitung evaluiert und berücksichtigt, etwa um zu beurteilen, ob bestimmte Daten rechtmäßig erfasst wurden (Wachter-Mittelstadt 2018, S. 122, 125f.; Wachter 2019), obwohl nach Art. 4(1) DSGVO alle Stadien der Datenverarbeitung zu betrachten sind. Hinzu kommt, dass bei der Evaluation hinsichtlich Personenbezug vs. Anonymität in der Praxis ausschließlich die Relation der fraglichen Daten zu dem *einen* Datensubjekt betrachtet wird, von dem die Daten erhoben werden.[15] Dass die anonymisierten Daten *vieler* Datensubjekte einen neuartigen Bruch von Privatheit *beliebiger Anderer* ermöglichen, bleibt in diesem Schema unerkannt. Im Laufe der Verarbeitung *abgeleitete* Informationen können die Unterscheidung von anonymen vs. personenbezogenen somit unterlaufen, und zwar nicht nur insofern vermeintlich anonyme Daten durch Schlussfolgerung wieder dem Datensubjekt zugeordnet werden könnten, auf das sie sich vor Anonymisierung bezogen haben, sondern vielmehr weil durch Verknüpfungen anonymer Daten neue Erkenntnisse in Bezug auf beliebige Dritte gewonnen werden können. Der Personenbezug würde hier also variable Individuen und insbesondere Dritte treffen und ist als Konzept damit überholt.

Die rechtliche und theoretische Würdigung der Gefahr durch abgeleitete Daten ist umstritten und uneinheitlich. Das Bundesverfassungsgericht argumentierte bereits im Volkszählungsurteil 1983, dass es keine „belanglosen" Daten gebe (BVerfGE 1983, S. 34) – doch der Fokus liegt hier nicht auf dem Massendaten-Szenario, das es damals noch nicht in der heutigen Form gab, sondern auf der Ableitung sensibler Informationen über ein In-

14 Das ist natürlich *auch* ein Problem, es entspräche Typ 2 der Angriffsszenarien; dieses steht jedoch hier nicht im Vordergrund, da ja argumentiert werden soll, dass es darüber hinaus noch eine neue Gefahrenlage gibt (Typ 3).

15 So willigt zum Beispiel beim Zugriff einer Social Media App auf das Telefonbuch eines Smartphones nur die Smartphonebesitzer:in in die Verarbeitung dieser Daten ein, nicht all die Personen, die in dem Telefonbuch eingetragen sind.

dividuum X aus scheinbar weniger sensiblen oder anonymisierten Daten über dasselbe Individuum X. Die ehemalige *Artikel 29 Arbeitsgruppe* hat in verschiedenen Stellungsnahmen empfohlen, abgeleitete Informationen unter die personenbezogenen Daten nach Art. 4 DSGVO zu fassen (Article-29-Datenschutzgruppe, 2018); in ihren Richtlinien und Stellungnahmen ist jedoch ebenfalls keine trennscharfe Adressierung des Phänomens anonymer Massendaten im Unterschied zur Gefahr der Re-Identifizierung zu erkennen. In Bezug auf die Kategorisierung von Daten (wie oben diskutiert zum Beispiel als anonym vs. personenbezogen) befürwortet die *Artikel 29 Arbeitsgruppe* in progressiver Weise, auf Verarbeitungszwecke und -folgen zu schauen anstatt auf Personenbezug im Input-Stadium (Article-29-Datenschutzgruppe, 2007; Wachter und Mittelstadt 2018, S. 126). Der Europäische Gerichtshof hingegen hat in mehreren Urteilen klargestellt, dass sich der Anwendungsbereich der DSGVO auf das „Input-Stadium" der Datenverarbeitung beschränke (Wachter und Mittelstadt 2018, S. 6) und die Abwehr gegenüber Verarbeitungsfolgen, auch hinsichtlich automatisierter Entscheidungen, auf sektorspezifische Regulierungen gestützt werden müsse (Wachter und Mittelstadt 2018, S. 7). Mit dem Instrument der Datenschutzfolgenabschätzung wiederum sieht die DSGVO einen Mechanismus vor, der die Folgen der Datenverarbeitung auch jenseits des „Input Stadiums" und somit insbesondere auch hinsichtlich der Effekte von anonymisierten Massendaten explizit einbeziehen kann. Doch auch diesem vergleichsweise sperrigen Instrument dürfte durch die Unterscheidung anonymisierter und personenbezogener Daten Grenzen gesetzt sein. Denn insbesondere gilt nach der aktuellen Auslegung des Rechts auf Löschung, dass diesem Recht auch durch Anonymisierung von Datensätzen Genüge getan werden kann.[16] Hier öffnet sich ein Schlupfloch für die unbefristete und unregulierte Verarbeitung von ehemals personenbezogenen Daten über die Zweckbindung hinaus, zum Beispiel für das Training prädiktiver Modelle, insofern dafür anonymisierte Daten ausreichen.

16 Ich folge hier meiner Auslegung der Entscheidung der Österreichischen Datenschutzbehörde DSB 2018. Vgl. in diesem Sinne außerdem direkt auf den Seiten der Europäischen Kommission: „Data can also be kept if it has undergone an appropriate process of anonymisation." https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/dealing-citizens/do-we-always-have-delete-personal-data-if-person-asks_en (letzter Besuch: 2022-03-10). In hiervon abweichender Auffassung vertritt Roßnagel 2021, dass Datenlöschung und Anonymisierung nach DSGVO nicht gleichgestellt sind.

6.1.2 Verwendung prädiktiver Modelle

Die zweite Frage ist, warum die DSGVO nicht effektiv die *Verwendung* prädiktiver Modelle – das heißt die Anwendung bereits vorhandener und trainierter Modelle auf Einzelpersonen – reguliert. Dies hängt eng mit der Frage zusammen, ob die DSGVO in ausreichendem Maße vor der Herstellung und Verwendung personenbezogener Vorhersagen schützt. Sandra Wachter und Brent Mittelstadt sehen hier Unterschiede des Schutzniveaus im Vergleich zu explizit erhobenen personenbezogenen Daten:

> "Compared to other types of personal data, inferences are effectively "economy class" personal data in the General Data Protection Regulation ("GDPR"). Data subjects' rights to know about (Art. 13–15), rectify (Art. 16), delete (Art. 17), object to (Art. 21), or port (Art. 20) personal data are significantly curtailed when it comes to inferences, often requiring a greater balance with the controller's interests (e.g., trade secrets or intellectual property) than would otherwise be the case. Similarly, the GDPR provides insufficient protection against sensitive inferences (Art. 9) or remedies to challenge inferences or important decisions based on them (Art. 22(3))." (Wachter und Mittelstadt 2018, S. 6).

Insbesondere sehen Wachter und Mittelstadt in der Rechtsprechung des EuGH der letzten Jahre Anhaltspunkte, dass abgeleitete Informationen über Individuen in Bezug auf die Rechtsfolgen der Datenverarbeitung nicht vollständig als „personenbezogene Daten" gemäß DSGVO behandelt werden müssen (Wachter und Mittelstadt 2018, S. 5 ff., 105 ff.). Das ist ein Punkt, in dem der 2018 beschlossene und 2020 in Kraft gesetzt *California Consumer Privacy Act* (CCPA) eine eindeutigere Regelung trifft: Im Unterschied zur DSGVO bietet der CCPA eine Definition von „personal information", die abgeleitete Daten explizit umfasst (Blanke 2020). Im Wortlaut des CCPA fallen unter den Begriff der „personal information" neben verschiedenen unmittelbar personenbezogenen Daten auch:

> "Inferences drawn […] to create a profile about a consumer reflecting the consumer's preferences, characteristics, psychological trends, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes." (CCPA § 1798.140 (o), CCPA 2018)

In diesem Zusammenhang ist zweitens zu nennen, dass die Regulierung von Profiling und automatisierten Entscheidungen durch die DSGVO (siehe Art. 22) zu schwach ausfällt, weil sie explizit auf voll-automatisierte Verfahren beschränkt ist. Verfahren, die mittels prädiktiver Modelle Men-

schen unterschiedlich behandeln, können vergleichsweise leicht als halbautomatische Routinen implementiert werden, indem menschliche Aufsicht und Eingriffsmöglichkeiten (z.B. durch Klickarbeiter:innen) in den Verarbeitungsprozess integriert werden, um die Bestimmungen des Art. 22 zu umgehen.

Ein dritter Grund für die effektiv schwache Regulierung der Verwendung prädiktiver Modelle liegt darin, dass bei der Erhebung der Hilfsdaten über Zielsubjekte, also derjenigen Daten, die als Input für die inferenzielle Verwendung eines prädiktiven Modells benötigt werden, die Einwilligungshürde psychologisch niedrig liegt. Die meisten User willigen in die Verarbeitung solcher Daten vorbehaltlos ein, weil ihnen Verhaltensdaten wie Facebook Likes als wenig sensibel erscheinen. Außerdem werden diese Daten häufig routinemäßig, ohne jeweils zweckbezogene Einwilligung oder auf der Rechtsgrundlage eines „berechtigten Interesses" (Art. 6(1)(f) DSGVO) bei der Verwendung sozialer Alltagsmedien erfasst.

6.2 Die DSGVO fit machen für KI & Big Data

6.2.1 Abgeleitete Informationen

Entlang dieser Bestandsaufnahme ergeben sich Vorschläge, wie die Regulierung prädiktiver Analytik im Kontext der DSGVO verbessert werden kann. Die Vorschläge zielen in die Richtung, abgeleitete personenbezogene Informationen mit Blick auf die Rechtsfolgen der Datenverarbeitung vollumfänglich mit explizit erhobenen personenbezogenen Informationen gleichzustellen – analog dem Kalifornischen CCPA. Das würde insbesondere bedeuten, dass die Legitimität einer Datenverarbeitung nicht allein in Bezug auf das Moment der Datenerhebung festzustellen ist, sondern im Hinblick auf die Zwecke und Auswirkungen der Verarbeitung beliebiger, auch etwa im Zuge der Verarbeitung anonymisierter Daten und Daten anderer Personen. Im Unterschied zu dem Vorschlag eines „Right to reasonable inferences" von Wachter und Mittelstadt (Wachter und Mittelstadt 2018) schlage ich hier jedoch *nicht* den Weg ein, den Datenschutz mit Instrumenten auszustatten, welche Individuen vor falschen oder wenig akkuraten Vorhersagen schützen möchten. Dieser Vorschlag greift zu kurz, um die Machtasymmetrie zwischen globalen Unternehmen, die über „prediction power" verfügen, und Gesellschaften auszugleichen. Wie argumentiert, können auch zutreffende Prädiktionen missbräuchlich und in für Gesellschaft und Individuum schädlicher Weise verwendet werden – eine

Schutzwirkung dagegen zu entfalten, wäre das Anliegen der rechtlichen Umsetzung prädiktiver Privatheit.

6.2.2 *Anonymisierte Daten*

Um die Gesellschaft vor dem Vermögen großer Firmen, anhand aggregierter Daten Vorhersagen zu stellen, zu schützen, sollten zweitens auch anonymisierte Daten unter die DSGVO-Prinzipien gefasst werden.[17] Vor dem Hintergrund der Missbrauchspotenziale sollte es *nicht* selbstverständlich sein, dass die Verarbeitung anonymisierter Daten umstandslos erlaubt ist und sich in einem weitestgehend unregulierten Geschäftsfeld außerhalb des Zugriffs der DSGVO abspielt. Anonymisierung von Datensätzen sollte überdies nicht der Löschung gleichgestellt werden.[18] Bei einer entsprechenden Neuregelung ist darauf zu achten, sich nicht auf die Gefahr der Re-Identifikation von Einzelpersonen in anonymisierten Datensätzen (Angriffsszenario von Typ 2) zu beschränken. Bei der Regulierung der Verarbeitung anonymer Daten zur Milderung der Risiken von Typ 3 ist erstens zu bedenken, dass es hierbei um die Missbrauchspotenziale *großer Sammlungen* anonymisierter Daten geht. Zweitens geht es um solche Sammlungen, in denen verschiedene mehr und weniger sensible Datenfelder auf Korrelationen untersucht werden können. Ein steigendes gesellschaftliches Bewusstsein um den Informationsreichtum anonymisierter Massendaten wäre für dieses Regulierungsanliegen förderlich, um es in der öffentlichen und politischen Debatte nicht auf die Gefahr der Re-Identifikation zu reduzieren. Aufzuklären ist auch darüber, wie der Informationsreichtum anonymisierter Massendaten durch Datenanalyse- und KI-Verfahren kommerziell abgeschöpft wird, mit potenziell großen gesellschaftliche Auswirkungen, die auch Menschen betreffen, deren anonymisierte Daten dem Modell *nicht* zugrunde liegen. Die DSGVO ist hiergegen bisher zahnlos, da Big Data-Geschäftsmodelle genau die Verwendungsmöglichkeiten von Da-

17 Dies bedeutet nicht, wie der Vorschlag häufig missverstanden wird, die Verarbeitung anonymisierter Daten kategorial zu verbieten, sondern, analog den personenbezogenen Daten, sie unter ein grundsätzliches Verarbeitungsverbot zu stellen, dessen Ausnahmen durch Rechtsgrundlagen geregelt werden müssen. Die Rechtsgrundlage der Einwilligung scheidet hierbei aus, wenn die Folgen der Datenverarbeitung potenziell Dritte betreffen, siehe unten. Eine politische Debatte ist darüber zu führen, welche Verwendungsweisen anonymisierter Massendaten als gesellschaftlich förderlich vs. schädlich gelten.

18 Vgl. oben, Fußnote 16.

ten kapitalisieren, die trotz Anonymisierung und DSGVO-Regulierungen möglich sind.

Auch die Beschränkung der Verarbeitung anonymisierter Daten darf sich nicht auf das Input-Stadium der Datenverarbeitung beschränken. Besonders ist zu bedenken, dass trainierte prädiktive Modelle *selbst aggregierte, anonymisierte Daten darstellen.*[19] Die Regulierung der Verarbeitung anonymisierter Daten muss daher die Zirkulation und Verwendung trainierter Machine Learning-Modelle umfassen. Prädiktive Modelle, die aus Kundendatensätzen erzeugt werden, können aktuell frei und insbesondere zweckbindungsfrei zirkulieren oder verkauft werden, weil sie nicht in den Schutzbereich der DSGVO fallen. Im Rahmen einer Neuregelung wäre eine erweiterte Form des Zweckbindungsgebots und der Aufsicht durch Aufsichtsbehörden oder unabhängiger Stellen einzubeziehen: Im Rahmen der Genehmigung der Herstellung eines prädiktiven Modells wäre im Vorhinein der Zweck anzugeben und zu genehmigen, für den dieses Modell durch benannte Akteure verwendet werden soll, so dass anderweitige Verwendungsweisen oder die Weitergabe verboten werden können.

6.2.3 Einwilligung beschränken

Eine dritte Säule der Modernisierung des Datenschutzes betrifft die Rechtsgrundlage der Einwilligung. Da im Kontext von Big Data und KI-Technologie die Verarbeitung der eigenen Daten grundsätzlich Auswirkungen auf andere hat, steht die Validität der Rechtsgrundlage der Einwilligung grundsätzlich in Frage. Die Einwilligung sollte nur verfügbar sein, wenn die Konsequenzen der Einwilligungsentscheidung allein das einwilligende Individuum betreffen (vgl. dazu auch Ruschemeier in diesem Band).

Bei einem durchschnittlichen Gebrauch von Internet und Smartphone-Apps ist die Einwilligung heute eine der dominantesten Manifestationen von Datenschutzregulierungen im täglichen Mediengebrauch. Die bewusstseinsbildende Funktion der Einwilligung ist nicht zu unterschätzen;

19 Solche Modelle werden durch Millionen Einträge in einer großen Matrix repräsentiert, die im Trainingsverfahren simulierter neuronaler Netze kalibriert werden. Diese Parameter sind selbst abgeleitete Daten und wenn das Trainingsverfahren bestimmte, technisch wohldefinierte Anforderungen erfüllt, lassen sich daraus keine individuellen Einträge der Trainingsdaten rekonstruieren, so dass es sich dabei formal um anonyme Daten handelt. Siehe hierzu den Diskurs zu differential privacy in machine learning; Fußnote 11 oben.

doch sie bestätigt das liberalistische Missverständnis von Datenschutz, das von den Gefahren prädiktiver Analytik ablenkt (Kröger u.a. 2021): Jeder neue Einwilligungsdialog, mit dem die Nutzer:in konfrontiert wird, affirmiert das gesellschaftlich schädliche Verständnis, dass es dem Datenschutz um die individuelle Wahlmöglichkeit jedes Einzelnen in Bezug auf die Preisgabe seiner Daten gehe. Darüber hinaus ist bekannt und wurde viel thematisiert, dass Einwilligungsdialoge über Dark Patterns, Design-Tricks, Nudges, längliches Kleingedrucktes und weil sie in den unpassendsten Momenten erscheinen, die Nutzer:innen nicht richtig informieren, sondern nicht selten zur Einwilligung überlisten oder nötigen (vgl. Baruh und Popescu 2017; Mühlhoff 2018).

Weiterhin könnte dem Instrument der Einwilligung bei der *Anwendung* prädiktiver Modelle auf Einzelpersonen Bedeutung zukommen. Ein von prädiktiver Wissensproduktion betroffenes Individuum sollte in die Ermittlung von Informationen oder Entscheidungen über es einwilligen müssen, bevor die dafür herangezogenen und meist weniger sensibel erscheinenden Hilfsdaten erfasst werden. In diesem Zusammenhang sei auf die erste Forderung zurückverwiesen, dass abgeleitete Daten hinsichtlich der Rechtsfolgen in vollem Umfang wie personenbezogene Daten behandelt werden sollten. In welchen Anwendungsbereichen die Einwilligung in diesem Fall als Rechtsgrundlage zur Verfügung gestellt und in welchen Domänen sie verboten werden sollte, wäre ausführlicher zu erwägen (vgl. Mühlhoff 2021).

6.2.4 Kollektive Schutzrechte

Ein weiterer zentraler Vorschlag betrifft die Einrichtung kollektivistischer Pendants zu den Betroffenenrechten der DSGVO. Das heißt, die Rechte zu Auskunft, Rektifizierung, Löschung, Portierbarkeit usw. sind kollektivistisch zu erweitern, so dass zum Beispiel von Diskriminierung betroffene Gruppen, aber auch das Gemeinwesen im Ganzen, in die Lage versetzt werden, von Plattformbetreibern über prädiktive Modelle und die Verarbeitung anonymisierter Daten Auskunft zu verlangen.[20] Eine solche Regelung sollte Interessenverbänden und der demokratischen Gesellschaft im Ganze mehr Kontrolle darüber ermöglichen, welche Informationen kommerzielle Organisationen über beliebige Individuen aus Hilfsdaten ableiten können und welche prädiktiven Modelle eine Organisation an-

20 Vgl. auch ähnliche Vorschläge bei Mantelero 2016; Pohle 2016.

hand der Daten vieler Nutzer:innen trainiert. Dieses kollektive Recht auf Einsichtnahme soll dazu dienen, dass aufgedeckt werden kann, welche Diskriminierungsmuster den prädiktiven Modellen eingeschrieben sind. Ein kollektives Recht auf Berichtigung oder Löschung solcher Modelle sollte aktivierbar sein, wenn Muster der Ausgrenzung und Diskriminierung, oder stabilisierende und verstärkende Effekte in Bezug auf soziale Ungleichheit beobachtbar sind. Für die Ausübung dieser kollektiven Abwehrrechte sollten Aufsichtsorgane sowie geeignete Instrumente der kollektiven Rechtsdurchsetzung wie Verbandsklagen oder Musterfeststellungsklagen vorgesehen werden (vgl. ausführlich Ruschemeier 2021).

6.3 Antidiskriminierung

Angesichts der Missbrauchspotenziale von Big Data und KI wird sich ein wirkungsvoller Datenschutz im aktuellen Jahrzehnt daran messen lassen müssen, zu welchem Grade er in eine tragfähige Allianz mit Antidiskriminierungsgesetzgebung tritt. Denn es geht im Kontext dieser Technologien nicht um das Ausspionieren Einzelner, sondern um massenweise parallel durchgeführte Abschätzungsoperationen, die uns alle betreffen und auf individualisierte – und das heißt, unterschiedliche – Behandlung von Individuen und Gruppen, mithin auf soziale Ungleichheit, Diskriminierung und Ausschlussmechanismen hinauslaufen können. Das Feld der prädiktiven Wissensgewinnung anhand von anonymisieren Massendaten, die wir alle täglich kostenfrei für große Datenunternehmen produzieren, ist aktuell weitestgehend unreguliert. Um den Regulierungsbedarf zu erkennen, muss sich der Datenschutz (und insbesondere der anglophone Diskurs um *privacy*) von seinem Lieblingsbezugspunkt, dem Schutz der informationellen Sphäre des Einzelnen, lösen, und die soziale Strukturierungswirkung moderner Datenverarbeitung in den Blick nehmen.

Acknowledgements

Ich danke Prof. Dr. Hannah Ruschemeier für die intensiven Diskussionen sowie den drei Reviewern für ihre Kommentare und Verbesserungsvorschläge.

Diese Arbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 16SV8480 unterstützt. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

Literatur

Abadi, Martín, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, und Li Zhang (2016). „Deep Learning with Differential Privacy". *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*, 308–18. https://doi.org/10.1145/2976749.2978318.

Article-29-Datenschutzgruppe (2007). „Stellungnahme 4/2007 zum Begriff ‚personenbezogene Daten'". 01248/07/DE WP 136. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp140_de.pdf.

Article-29-Datenschutzgruppe (2018). „Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679". 17/DE WP251rev.01. https://ec.europa.eu/newsroom/article29/items/612053/en.

Baruh, Lemi, and Mihaela Popescu (2017). "Big Data Analytics and the Limits of Privacy Self-Management." *New Media & Society* 19, no. 4: 579–96. https://doi.org/10.1177/1461444815614001.

Ben-Shahar, Omri (2019). „Data Pollution". *Journal of Legal Analysis* 11 (Januar): 104–59. https://doi.org/10.1093/jla/laz005.

Blanke, Jordan M. (2020). „Protection for 'Inferences Drawn': A Comparison Between the General Data Protection Regulation and the California Consumer Privacy Act". *Global Privacy Law Review* 1 (2).

Bundesverfassungsgericht (1983). BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983 – Zur Verfassungsmäßigkeit des Volkszählungsgesetzes 1983, 209/83 1 BvR 1–215. Bundesverfassungsgericht.

California Consumer Privacy Act, Cal. Legis. Serv. Ch. 55 (A.B. 375) (west) § (2018). https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

Duhigg, Charles (2012). „How Companies Learn Your Secrets". *The New York Times*, 16. Februar 2012, Abschn. Magazine. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html.

Dwork, Cynthia (2006). „Differential Privacy". In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10–14, 2006, Proceedings, Part II*, herausgegeben von Michele Bugliesi, Bart Preneel, Vladimiro Sassone, und Ingo Wegener, 2:1–12. Lecture Notes in Computer Science 4052. Berlin and Heidelberg: Springer.

Goggin, Benjamin (2019). „Inside Facebook's suicide algorithm: Here's how the company uses artificial intelligence to predict your mental state from your posts". *Business Insider*, 6. Januar 2019. https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12.

Gymrek, M., A. L. McGuire, D. Golan, E. Halperin, und Y. Erlich (2013). „Identifying Personal Genomes by Surname Inference". *Science* 339 (6117): 321–24. https://doi.org/10.1126/science.1229566.

Kosinski, Michal, David Stillwell, und Thore Graepel (2013). „Private Traits and Attributes Are Predictable from Digital Records of Human Behavior". *Proceedings of the National Academy of Sciences* 110 (15): 5802–5. https://doi.org/10.1073/pnas.1218772110.

Kröger, Jacob Leon, Otto Hans-Martin Lutz, und Stefan Ullrich (2021). „The Myth of Individual Control: Mapping the Limitations of Privacy Self-Management". In *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3881776.

Lewinski, Kai von (2009). „Geschichte des Datenschutzrechts von 1600 bis 1977". In *Freiheit – Sicherheit – Öffentlichkeit: 48. Assistententagung Öffentliches Recht, Heidelberg 2008*, 196–220. Nomos. https://doi.org/10.5771/9783845215532-196.

von Lewinski, Kai (2014). *Die Matrix des Datenschutzes Besichtigung und Ordnung eines Begriffsfeldes.* Tübingen: Mohr Siebeck. http://public.eblib.com/choice/PublicFullRecord.aspx?p=6624481.

Lippert, John (2014). „ZestFinance Issues Small, High-Rate Loans, Uses Big Data to Weed out Deadbeats". *Washington Post*, 11. Oktober 2014, Abschn. Business. https://www.washingtonpost.com/business/zestfinance-issues-small-high-rate-loans-uses-big-data-to-weed-out-deadbeats/2014/10/10/e34986b6-4d71-11e4-aa5e-7153e466a02d_story.html.

Loi, Michele, und Markus Christen (2020). "Two Concepts of Group Privacy." *Philosophy & Technology* 33: 207–24. https://doi.org/10.1007/s13347-019-00351-0.

Mantelero, Alessandro (2016). „Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection". *Computer Law & Security Review* 32 (2): 238–55. https://doi.org/10.1016/j.clsr.2016.01.014.

Merchant, Raina M., David A. Asch, Patrick Crutchley, Lyle H. Ungar, Sharath C. Guntuku, Johannes C. Eichstaedt, Shawndra Hill, Kevin Padrez, Robert J. Smith, und H. Andrew Schwartz (2019). „Evaluating the Predictability of Medical Conditions from Social Media Posts". *PLOS ONE* 14 (6): e0215476. https://doi.org/10.1371/journal.pone.0215476.

Mittelstadt, Brent (2017). "From Individual to Group Privacy in Big Data Analytics." *Philosophy & Technology* 30, no. 4: 475–94. https://doi.org/10.1007/s13347-017-0253-7.

Mühlhoff, Rainer (2018). „Digitale Entmündigung und User Experience Design: Wie digitale Geräte uns nudgen, tracken und zur Unwissenheit erziehen". *Leviathan – Journal of Social Sciences* 46 (4): 551–74. https://doi.org/10.5771/0340-0425-2018-4-551.

Mühlhoff, Rainer (2020a). „Die Illusion der Anonymität: Big Data im Gesundheitssystem". *Blätter für Deutsche und Internationale Politik* 8: 13–16.

Mühlhoff, Rainer (2020b). „Prädiktive Privatheit: Warum wir alle »etwas zu verbergen haben«". In *#VerantwortungKI – künstliche intelligenz und gesellschaftliche folgen*, herausgegeben von Christoph Markschies und Isabella Hermann. Bd. 3/2020. Berlin-Brandenburgische Akademie der Wissenschaften.

Mühlhoff, Rainer (2020c). „Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus". *Deutsche Zeitschrift für Philosophie* 68 (6): 867–90. https://doi.org/10.1515/dzph-2020-0059.

Mühlhoff, Rainer (2021). „Predictive Privacy: Towards an Applied Ethics of Data Analytics". *Ethics and Information Technology*, Juli. https://doi.org/10.1007/s10676-021-09606-x.

Narayanan, Arvind, und Vitaly Shmatikov (2008). „Robust De-anonymization of Large Sparse Datasets". In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, 111–25. Oakland, CA, USA: IEEE. https://doi.org/10.1109/SP.2008.33.

Ohm, Paul (2010). „Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization". *UCLA Law Review*, 77.

O'Neil, Cathy (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Österreichische Datenschutzbehörde (2018). Datenschutzbeschwerde von Dr. Xaver X.

Pohle, Jörg (2016). „PERSONAL DATA NOT FOUND: Personenbezogene Entscheidungen als überfällige Neuausrichtung im Datenschutz". *Datenschutz Nachrichten*, 2016.

Roßnagel, Alexander (2018). "Notwendige Schritte zu einem modernen Datenschutzrecht." In *Die Fortentwicklung des Datenschutzes: zwischen Systemgestaltung und Selbstregulierung*, herausgegeben von Alexander Roßnagel, Michael Friedewald und Marit Hansen, 361–84. Wiesbaden: Springer Vieweg. https://doi.org/10.1007/978-3-658-23727-1_20.

Ruschemeier, Hannah (2021). „Kollektiver Rechtsschutz und strategische Prozessführung gegen Digitalkonzerne". *MMR* 24 (12): 942–46.

Skeba, Patrick, und Eric PS Baumer (2020). „Informational Friction as a Lens for Studying Algorithmic Aspects of Privacy." *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2: 1–22.

Sweeney, Latanya (2002). „K-Anonymity: A Model for Protecting Privacy". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10 (05): 557–70. https://doi.org/10.1142/S0218488502001648.

Tavani, Herman T. (2007). „Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy". *Metaphilosophy* 38 (1): 1–22. https://doi.org/10.1111/j.1467-9973.2006.00474.x.

Taylor, Linnet, Luciano Floridi, und Bart van der Sloot (2016). *Group Privacy: New Challenges of Data Technologies*. New York: Springer Berlin Heidelberg.

Vedder, Anton (1999). "KDD: The Challenge to Individualism." *Ethics and Information Technology* 1, no. 4: 275–81.

Wachter, Sandra (2019). „Data Protection in the Age of Big Data". *Nature Electronics* 2 (1): 6–7. https://doi.org/10.1038/s41928-018-0193-y.

Wachter, Sandra, und Brent Mittelstadt (2018). „A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI". Preprint. LawArXiv. https://doi.org/10.31228/osf.io/mu2kf.