Open Educational Resources

DATEN SCHUTZ

KOLLEKTIV & MACHTKRITISCH

Workshopkonzept für die Erwachsenenbildung

FORSCHUNGSDOSSIER

DECOL
Data Ethics Outreach Lab
Universität Osnabrück

Autor:in
Annemarie Witschas

Herausgeber
Prof. Dr. Rainer Mühlhoff
Universität Osnabrück
Ethics and Critical Theories of AI
Wachsbleiche 27
D - 49090 Osnabrück

Redaktion
Nora Freya Lindemann
Anna Kraher

Design
Anna Kraher

Coverbild
Anna Kraher

Kontakt
outreachlab@ethikderki.de

5. Februar 2024

DECOL Data Ethics Outreach Lab
Ethics and Critical Theories of AI Universität Osnabrück
UNIVERSITÄT OSNABRÜCK

INHALT

Der folgende Text gibt einen Überblick über die wichtigsten Konzepte, die im Rahmen des Workshops „Datenschutz: kollektiv und machtkritisch“ verwendet werden.

TEIL 1: DATEN

WAS SIND DATEN?

„Daten“ ist ein Oberbegriff für jegliche Informationen, wie z.B. numerische Werte, aber auch Text, Kontakt- oder Kalendereinträge. Daten können, aber müssen aber nicht, digital sein, also zum Beispiel auf einem Computer abgespeichert sein. In diesem Workshop liegt der Fokus auf digital erhobenen und verarbeiteten Daten liegen. Da der Workshopschwerpunkt des Weiteren auf gesellschaftlichen Auswirkungen von Daten liegt, werden primär „personenbezogene Daten“ thematisiert, also Informationen, die über Personen gesammelt und verarbeitet werden (DSGVO, Art. 4).

WAS PASSIERT MIT UNSEREN DATEN?

SAMMLUNG

Daten sind allgegenwärtig in unserem digitalen Alltag. Wenn wir Websites besuchen, Apps nutzen oder smarte Geräte verwenden, werden nahezu immer Informationen über uns gesammelt. Ein Teil dieser Daten ist unerlässlich für die Nutzung des jeweiligen Dienstes, wie unsere Accountdaten, die es uns ermöglichen, uns einzuloggen. Doch darüber hinaus werden auch umfangreichere Informationen über uns, unser Verhalten und unsere Vorlieben erfasst. Dies kann die Erfassung unserer IP-Adresse beinhalten, welche Rückschlüsse auf unseren Standort ermöglicht, sowie die Aufzeichnung von Nutzungsdaten von Apps und Websites, etwa wie häufig wir eine App nutzen, wann wir sie nutzen und wie lange wir sie nutzen. Zusätzlich werden aufmerksamkeitsbasierte Daten erfasst, wie etwa die Dauer des Betrachtens von Inhalten oder die Geschwindigkeit des Tippens, sowie das Verhalten beim Wechseln zwischen verschiedenen Tabs.

AUSWERTUNG

Selbst wenn diese Daten auf den ersten Blick unscheinbar oder unwichtig erscheinen mögen, können durch die Analyse großer Datenmengen Muster erkannt werden, die wiederum Rückschlüsse auf sensiblere Informationen ermöglichen. Dazu werden verschiedene statistische, algebraische oder Machine Learning Verfahren genutzt, um Muster und Regelmäßigkeiten in den Daten aufzuspüren. Diese Verfahren werden unter dem Überbegriff der *prädiktiven Analytik* zusammengefasst. Sie dienen dem Ziel, sensible Eigenschaften über Personen zu inferieren, oder auch zukünftige Ereignisse oder Verhaltensweisen vorherzusagen.

So hat eine Studie aus den USA bereits 2013 ergeben, dass durch eine Handvoll *Facebook Likes* sensible biografische Informationen vorhergesagt werden können, die die Nutzer:innen selbst nicht angegeben hatten (Kosinski et al., 2013). Sehr gute Vorhersagbarkeit hatten dabei die Kategorien *weiß* oder „African American“, sowie Geschlecht. Diese können laut Studie in 95% bzw. 93% der Fälle korrekt vorhergesagt werden. Auch die Kategorien „schwul“, demokratisch oder republikanisch erzielten gute Vorhersageergebnisse. Zigaretten- und Alkoholkonsum können ebenfalls einigermaßen zuverlässig anhand der Like-Daten vorhergesagt werden (73% bzw. 70%).

Um noch ein Beispiel aus Deutschland zu Rate zu ziehen: David Kriesel vom Chaos Computer Club (CCC) analysierte über zweieinhalb Jahre 100.000 Artikel der Zeitschrift *Spiegel Online*. Allein mit den Daten, wann die Artikel veröffentlicht wurden und von wem sie geschrieben wurden, konnte er rekonstruieren, welche Autor:innen wiederholt zeitgleich in den Urlaub gefahren sind. Daraus leitet er die Vermutung ab, welche Autor:innen miteinander liiert sind (Kriesel, 2016).

Auch wenn es sich dabei um ein illustratives und relativ harmloses Beispiel handelt, können daran doch einige Merkmale der prädiktiven Daten-Logik gezeigt werden. Zum einen wird hier verdeutlicht, wie sich aus frei verfügbaren Daten (Datum der Veröffentlichung und Verfasser:in) eine potenziell intime Information ermitteln lässt. Diese wiederum ließe sich für diverse Zwecke von personalisierter Werbung nutzbar machen (z.B. Urlaubsreisen für Pärchen). Des Weiteren muss die vorhergesagte Eigenschaft der partnerschaftlichen Beziehung gar nicht stimmen. Es könnte sich auch um sogenannte Scheinkorrelationen handeln, bei denen zwei Autor:innen aus anderen Gründen wiederholt zeitgleich Urlaub nehmen. Auch datenauswertende Unternehmen können nicht wissen, ob wir tatsächlich schwanger sind, eine neue Waschmaschine suchen oder empfänglich für Glücksspiel sind. Trotzdem können Unternehmen davon profitieren, wenn sie uns einfach entsprechend der *wahrscheinlichsten Kategorie* behandeln (Joque, 2022, Kap. 7).

HANDEL

Die Daten, die verschiedene Anbieter über uns sammeln, bleiben oft nicht nur in deren Händen. Ihre wirtschaftliche Relevanz für die Betreiber steigt häufig erst, wenn sie weiterverkauft werden. Hier kommen *Datenhändler (Data Broker)* ins Spiel. Diese Unternehmen sammeln riesige Datensätze aus verschiedenen Quellen und durchforsten sie nach relevanten Informationen, um diese dann weiterzuverkaufen. Dabei agieren sie größtenteils im Verborgenen und sind daher der breiten Öffentlichkeit weitgehend unbekannt. Dennoch verfügen sie über unzählige Profile von praktisch jeder:m Einzelnen.

Eine Recherche von Netzpolitik.org zu Datenhändler:innen von 2023 dokumentierte *650.000 Kategorien*, die Datenhändler:innen nutzen, um Nutzer:innen zu segmentieren. Darunter gab es zahlreiche Kategorien basierend auf *sensiblen* Informationen wie „Brustkrebs“, „Grindr Nutzer“ oder „moms who shop like crazy“ (Dachwitz, 2023).

WAS IST DIE GRÖßERE PERSPEKTIVE?

Einige Expert:innen beobachten, wie die kommerzielle Datensammlung, die darauf abzielt, möglichst viele Informationen über Nutzer:innen zu sammeln und in umfassenden Profilen zusammenzuführen, zum vorherrschenden Paradigma des heutigen Internets geworden ist. Die Wirtschaftswissenschaftlerin Shoshana Zuboff prägte hierfür den Begriff des *"Überwachungskapitalismus" (*Zuboff, 2020). Durch diese Praktiken werden wir für Unternehmen vorhersehbar und berechenbar, was wiederum für zielgerichtete Werbung und personalisierte Dienste genutzt wird.

Erwähnenswert ist dabei auch, dass durch die groß angelegte Speicherung von Daten nicht nur jene Analyseverfahren eine Rolle spielen, die bereits jetzt entwickelt sind und genutzt werden. Auch zukünftige, noch präzisere Technologien könnten rückwirkend auf heutige Daten angewendet werden.

TEIL 2: DATENSCHUTZ

MACHTKRITISCHE DIMENSION

Ab den 1980er Jahren begannen die Möglichkeiten zur elektronischen Verarbeitung und Speicherung von Informationen rasant zuzunehmen. Allerdings wurden bereits zu dieser Zeit Bedenken laut. So bestand die Befürchtung, dass diejenigen Akteure, die Zugang zu umfangreichen Datenmengen hatten und die technischen Mittel besaßen, diese auszuwerten, einen erheblichen Vorteil hätten.

Wie die Redewendung "Wissen ist Macht" bekanntermaßen zusammenfasst, formieren sich entlang solcher Informationsanhäufungen *Machtpotentiale*. Das Wissen über die Lebensumstände, das Verhalten und die Routinen der Menschen kann dazu dienen, soziale Kontrolle auszuüben. Der Philosoph und Historiker Michel Foucault beschreibt, wie das das Gefühl, unter ständiger potenzieller Beobachtung zu stehen, dazu führt, dass Subjekte die Erwartungen, die an sie gestellt werden, internalisieren und nach ihnen handeln, weil sie ständig potenziell beobachtet werden (Foucault, 1995).

Dies lässt sich auch im Volkszählungsurteil 1983 des Bundesverfassungsgerichts finden, welches urteilte, dass eine erzwungene Datenerhebung einen derartigen Anpassungsdruck erzeugen würde, dass er einer freien und ungezwungenen Entfaltung der Persönlichkeit schade. Dadurch würde ebenfalls die funktionierende Demokratie gefährdet, welche auf der Selbstbestimmung ihrer Bürger:innen beruht.

Der Datenschützer Wilhelm Steinmüller brachte diese Problematik prägnant auf den Punkt: "Es geht nicht um Privatsphäre, sondern es geht darum, eine Technik sozial beherrschbar zu machen" (Steinmüller, 2009).

Aus dieser machtkritischen Perspektive lässt sich folgern, dass Datenschutz nicht primär darauf abzielt, Daten zu schützen, sondern vielmehr darauf, *informationelle Machtgefälle* zu verhindern, die durch undurchsichtige Datensammlungen, langfristige Datenspeicherung und automatisierte Datenverarbeitung entstehen können. Dadurch schützt Datenschutz letztlich Individuen und Gesellschaft vor Grundrechtseingriffen.

KOLLEKTIVE DIMENSION

Datenschutz ist also nicht einfach gleichzusetzen mit dem Schutz der „Privatsphäre“ eines Individuums. Statt Daten als Attribut oder Eigentum einer bestimmten Person anzusehen, ist es fruchtbar, Daten als Verhältnis zwischen den beteiligten Akteuren zu betrachten: Eine Nutzerin wäre dementsprechend über jene Daten mit einer Social Media Plattform verbunden, welche die Plattform über sie verarbeitet. Gleichzeitig sind auch andere Nutzer:innen der Plattform nicht einfach voneinander unabhängige Besucher:innen, sondern sie sind über ihre Daten miteinander verwoben und die Daten Anderer wirken zurück auf die jeweilige Nutzer:in (Mühlhoff, 2023, s. 62).

Um diese kollektive Dimension von Datenschutz zu beleuchten, demonstriert Rainer Mühlhoff, inwiefern auf großen Plattformen sensible Informationen mithilfe der Daten anderer Nutzer:innen ermittelt werden (Abb. Präsentationsfolie S. 8–10).

Dafür nutzen die Plattformen – etwa eine Social Media Plattform wie Facebook oder Instagram – die Daten, welche ihnen von den Nutzer:innen zuhauf überlassen werden: Daten zur Aktivität auf der Plattform, wie Likes, geteilte Inhalte, Reaktionen, sowie Daten aus anderen Aktivitäten aus dem Browser, welche beispielsweise durch Drittanbieter-Cookies gewonnen wurden (Grüne Punkte in der Abbildung).

Des Weiteren geben manche Nutzer:innen auch sensiblere Daten über sich preis, etwa ihre sexuelle Identität oder eine Neurodivergenz. Dies können Nutzer:innen direkt in ihrem Facebook-Profil angeben. Gleichzeitig lassen sich aber auch Hinweise über sensible Daten von Nutzer:innen aber auch etwa über den Beitritt in einschlägige Gruppen oder Likes inferieren (Sensible Daten werden in der Abbildung als rote Punkte dargestellt.). Dies sind sogenannte Hilfsdaten. Folglich haben die Plattform-Betreibenden von manchen Nutzer:innen sowohl Hilfs-/ Nutzungsdaten als auch sensible Daten. Selbst wenn nur eine kleine Minderheit der Nutzer:innen diese Daten preisgegeben hat (etwa 5%), wären dies immer noch ausreichend Informationen, um darauf ein Modell zu trainieren, welches die Korrelationen zwischen den Hilfsdaten und den sensiblen Daten lernt. Diese trainierten Modelle können dann verwendet werden, um die sensiblen Kategorien ebenfalls über die restlichen Nutzer:innen vorherzusagen. (Rosa Punkte in der Abbildung.)

Dieser Mechanismus erlaubt es Plattformen wie Facebook oder Instagram gezielte Werbung auf Basis spezifischer biographischer Informationen anzubieten (z.B. Beziehungsstatus, medizinischer Zustand, politische Einstellungen), obwohl die meisten Nutzer:innen diese Informationen gar nicht über sich preisgeben.

DATENVERSCHMUTZUNG

Um diese kollektive Abhängigkeit voneinander greifbarer zu machen, entwickelt Omri Ben-Shahar (2019) das Konzept der „Datenverschmutzung“. Analog zu Umweltverschmutzungen sind die Konsequenzen unserer Handlungen nicht nur auf uns selbst beschränkt, sondern prägt das Ökosystem, in dem wir uns befinden und beeinflusst dadurch auch andere Menschen und andere Lebewesen darin. Ob wir also viel Auto fahren und fliegen oder Produkte von „Datenkraken“ nutzen – für uns persönlich können wir dabei wahrscheinlich viele Vorteile herausholen. Doch die Auswirkungen davon beeinflussen auch andere Menschen und bürgen Risiken für sie, hervorgerufen etwa durch dem Ausgesetztsein von Luftverschmutzung, Klimaerhitzung oder ein engmaschiges Datennetz an personenbezogenen Daten.

Das bedeutet nicht, dass es unsere individuelle Schuld ist, dass wir an dieser jeweiligen „Verschmutzung“ beteiligt sind. Oft sind es infrastrukturelle Abhängigkeiten, die uns keine weitere Wahl lassen, als bei der kollektiven Praktik mitzumachen.

Daraus ergibt sich auch, dass Veränderung strukturell gedacht werden müssen: So wie eine Person, die auf ihr Auto verzichtet, noch nicht die Klimakrise abwenden wird, wird auch ein einzelner Mensch, der sein Smartphone von Google befreit, auch das Problem der massiven Datensammlung nicht ändern.

TEIL 3: HANDLUNGSMÖGLICHKEITEN

Ansätze zur Bewältigung der besprochenen Herausforderungen können auf verschiedenen Ebenen verortet werden. Eine ideale Strategie kombiniert Ansätze der unterschiedlichen Ebenen miteinander.

Auf *politischer Ebene* stellen *Regulierungen* ein zentrales Werkzeug zur Sicherung von Datenschutzstandards und sozialer Kontrolle über prädiktive Technologien dar. Die Datenschutz-Grundverordnung (DSGVO) kann dabei als ein erfolgreiches Beispiel betrachtet werden, welche auch jenseits der EU neue Standards setzte. Jedoch gibt es auch innerhalb der bestehenden Gesetzgebung noch zahlreiche Schlupflöcher und Grauzonen, welche juristisch aufgearbeitet werden sollten. Gleichzeitig bedarf es weiterer Regulierungen wie beispielsweise die Zweckbindung von Vorhersagemodellen (Mühlhoff & Ruschemeier, 2024).

Obgleich politische Regulierung auf effektivste und großflächigste Weise Veränderungen anstoßen kann, sind die damit einhergehenden institutionellen Prozesse oft langwierig, voller Kompromisse und umkämpft von unterschiedlichen Interessensgruppen. Um die Teilnehmenden aufgrund dieser Hürden nicht einem Ohnmachtsgefühl zu überlassen, werden im Workshop verschiedene *individuell umsetzbare Handlungsmöglichkeiten* thematisiert. Die Liste der „Tipps für kollektiven & machtkritischen Datenschutz“ ist auf einem Flyer zusammengefasst. Um nur beispielhaft einige Ansätze zu nennen, können mit Einstellungen im technischen Set-Up wie Wahl des Browsers (Firefox statt Chrome) oder datenschutzfreundliche Add-Ons (z.B. uBlock Origin) bereits viel ungewollte Datensammlung unterbunden werden (Digitalcourage, 2022).

Dabei ist jedoch zu beachten, dass solche individualisierten Ansätze nur begrenzt Veränderungen bewirken. Ebenfalls besteht die Gefahr, dass ein zu starker Fokus auf individuelle Nutzungsumstellungen zu einem responsibilisierenden Diskurs beiträgt, also dass die Verantwortung für den Schutz digitaler Grundrechte zu sehr auf die Einzelnen abgewälzt wird (vgl. Lisker, 2023).

Daher empfehlen wir, die bereitgestellten Tipps nicht als verpflichtende „Hausaufgabe“ zu betrachten, sondern als Anregung, mehr über die Zusammenhänge und Hintergründe unserer digitalen Welt zu lernen. Dadurch werden auch die *Auseinandersetzung* mit Büchern, Filmen, Spielen etc. zu einem fruchtbaren Ansatzpunkt für Interventionen auf individueller Ebene. Eine Liste mit Leseempfehlungen sowie Filmen und Online-Tools befindet sich in den Präsentationsfolien.

Ebenfalls lehrreich kann das Einholen eigener *Datenauskünfte* sein. Laut DSGVO haben Nutzer:innen das „Recht auf Auskunft“ über Daten, die über sie gespeicherte wurden. Dazu reicht meist eine formlose E-Mail an die Betreibenden. Es gibt aber auch online Tools, die Anfragen in Ihrem Namen herausschicken können[1].

Erwähnenswert ist auch, dass zwischen individueller und politisch-regulatorischer Ebene ebenfalls Spielraum für Handlungsmöglichkeiten besteht. Auf dieser *kollektiven Ebene* können sich Menschen gemeinsam organisieren und die kollektive Relevanz von Datenschutz in ihren Kreisen bekannt machen, sei es am Arbeitsplatz, in der Universität, in Schulen oder in Vereinen.

[1] https://selbstauskunft.net/

REFERENZEN

Ben-Shahar, O. (2019). Data Pollution. *Journal of Legal Analysis*, *11*, 104–159. https://doi.org/10.1093/jla/laz005

Digitalcourage. (2022, Juli 25). *Sicher surfen mit Laptop oder PC*. https://digitalcourage.de/digitale-selbstverteidigung/sicher-surfen-pc

Joque, J. (2022). *Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism*. Verso Books.

Mühlhoff, R. (2023). *Die Macht der Daten* (Bd. 10). Universitätsverlag Osnabrück V&R unipress.

DSGVO, Art. 4 Datenschutz-Grundverordnung § Kapitel I - Allgemeine Bestimmungen (Art. 1 - 4).

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kriesel, D. (2016, Dezember 28). *SpiegelMining – Reverse Engineering von Spiegel-Online*. https://media.ccc.de/v/33c3-7912-spiegelmining_reverse_engineering_von_spiegel-online

Lisker, M. (2023). *Von der (Un-)Möglichkeit, digital mündig zu sein* [Masterarbeit]. Technische Universität Berlin.

Mühlhoff, R. (2023). *Die Macht der Daten: Warum künstliche Intelligenz eine Frage der Ethik ist*. V&R unipress. https://doi.org/10.14220/9783737015523

Mühlhoff, R. and Ruschemeier, H. (2024). *Updating Purpose Limitation for AI: A normative approach from law and philosophy*. http://dx.doi.org/10.2139/ssrn.4711621

Steinmüller, W. (2009). *Interview mit Prof. Wilhelm Steinmüller—ULD*. https://www.datenschutzzentrum.de/artikel/939-Interview-mit-Prof.-Wilhelm-Steinmueller.html