forum <privatheit>
selbstbestimmtes_leben_in_der_digitalen_welt

FORUM PRIVATHEIT UND SELBSTBESTIMMTES LEBEN IN DER DIGITALEN WELT

Policy Paper

RISIKEN KÜNSTLICHER INTELLIGENZ FÜR DIE MENSCHLICHE SELBSTBESTIMMUNG

IMPRESSUM

Autor_innen:
Thilo Hagendorff, Christian Geminn, Jörn Lamla, Murat Karaboga, Nicole Krämer, Maxi Nebel, Markus Uhlmann

Kontakt:
Michael Friedewald

Telefon +49 721 6809-146
Fax +49 721 6809-315
E-Mail info@forum-privatheit.de

Fraunhofer-Institut für System- und Innovationsforschung ISI
Breslauer Straße 48
76139 Karlsruhe

www.isi.fraunhofer.de
www.forum-privatheit.de

Schriftenreihe:
Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt
ISSN-Print 2199-8906
ISSN-Internet 2199-8914

1. Auflage, September 2020

Einleitung

In den letzten Jahren konnten bei Technologien des maschinellen Lernens durch Sprunginnovationen rasante Fortschritte in der Entwicklung zunehmend leistungsstarker Systeme gemacht werden. Diese zumeist unter dem Oberbegriff „künstliche Intelligenz" (KI) gefassten Anwendungen umfassen insbesondere Methoden, die in ihrer Funktionsweise vom menschlichen Gehirn inspiriert sind, dabei aber nicht notwendigerweise das Vorgehen menschlicher Lern- und Denkprozesse kopieren. KI-Anwendungen werden in unterschiedlichen Ausprägungen in verschiedensten technischen Systemen eingesetzt, die zu Veränderungen unterschiedlicher Tiefe und Geschwindigkeit in Wirtschaft, Industrie, Recht, Medizin, der Medienbranche oder der Wissenschaft führen. Zudem werden KI-Technologien zunehmend für Endverbraucher_innen oder Endnutzer_innen digitaler Medien verfügbar, sei es in Form von Sprachassistent_innen, Spamfiltern, Kaufempfehlungen, Übersetzungssoftware, Suchmaschinen und vielem mehr.

Viele der Anwendungen stehen in der Diskussion, die Autonomie oder Selbstbestimmung von Menschen einzuschränken. Inhaltsempfehlungssysteme auf Social-Media-Plattformen werden hier ebenso thematisiert wie personalisierte Werbeanzeigen oder Techniken des „Big Nudging", der subtilen Manipulation auf Grundlage der Auswertung großer Datenmengen (Big Data). Aber auch das Problem der technischen und organisationalen Intransparenz ist zu nennen, welche es Nutzer_innen erschwert, das Zustandekommen von algorithmischen Entscheidungen auch nur ansatzweise nachzuvollziehen. Demnach besteht in vielen Diskursen eine eher kritische Haltung gegenüber dem Einfluss von KI-Technologien auf die informationelle Privatheit sowie die damit verbundene Idee der persönlichen Selbstbestimmung. Problematisiert wird, dass in vielen Bereichen des sozialen Lebens Handlungsweisen im Umgang mit digitalen Medien und Technologien erst durch ein umfassendes Tracking ausgeforscht und vermessen werden, das Nutzungsverhalten anschließend per maschinellem Lernen analysiert und statistisch modelliert wird, um es dann in einem dritten Schritt vorhersagen und durch affektive Stimuli beeinflussen zu können. Dieser Dreischritt vollzieht sich insbesondere – aber nicht nur – im Rahmen von Social-Media-Plattformen und geschieht vor allem im Interesse einzelner Unternehmen, die durch derartige Methoden größere Umsätze erzielen können. Methoden der Mustererkennung, Vorhersagen und die Beeinflussung des Nutzungsverhaltens haben potenziell einen kritischen Einfluss auf die Selbstbestimmungsansprüche und -möglichkeiten der Betroffenen.

Das folgende Policy Paper analysiert den beschriebenen Themenkomplex vorrangig aus ethischer, rechtlicher und gesellschaftswissenschaftlicher Perspektive und skizziert gesellschaftlichen Handlungsbedarf. Es beginnt mit einer Einführung in die unterschiedlichen Perspektiven auf Selbstbestimmung. Im Anschluss geht es der Frage nach, inwiefern sich KI-Systeme fördernd oder einschränkend auf die Selbstbestimmung von Menschen auswirken können, nennt Beispiele, zeigt Zukunftsperspektiven auf und gibt Empfehlungen. Insgesamt soll das Policy Paper einen Beitrag zur derzeitigen breit geführten Diskussion über die Chancen und Risiken von KI-basierten Technologien leisten. Es baut dazu auf die Zusammenführung unterschiedlicher disziplinärer Perspektiven, die Ethik und Philosophie sowie Rechtswissenschaften, Soziologie, Medienpsychologie und Informatik einschließen.

Was ist Selbstbestimmung?

Klassische Verständnisse von Selbstbestimmung fokussieren auf die Autonomie der und des Einzelnen und begreifen Gesellschaft als die Summe einzelner autonomer Individuen. Unter Selbstbestimmung wird ein idealer Zustand begriffen, in dem eine Person unabhängig von anderen Menschen Entscheidungs- und Handlungsfreiheit genießt. Der Philosoph Immanuel Kant etwa sah Autonomie als den „Grund der Würde der menschlichen Natur", demgemäß als etwas dem Menschen Vorbehaltenes und an sich Gegebenes, das die Bedingung für moralisches Handeln ist. Das Ziel vieler klassischer Ansätze ist es, die individuelle Selbstbestimmung zu maximieren, indem externe Einschränkungen dieser Selbstbestimmung reduziert werden (negative Freiheit, Abwesenheit von Zwang). Historisch betrachtet spielte die individuelle Selbstbestimmung eine wichtige Rolle im Hinblick auf die Befreiung des Menschen aus der Herrschaft autoritärer Glaubenssysteme, aus feudalen Herrschaftsstrukturen und sonstigen Formen der Fremdbestimmung im Kontext westlicher Gesellschaften.

Im Laufe der Jahre geriet dieses am Individuum orientierte Verständnis der Selbstbestimmung jedoch vielfach in die Kritik. Eingewendet wurde etwa, dass Selbstbestimmung als vorsoziale Eigenschaft postuliert wurde, die nicht erst im Ergebnis von Sozialisierung erworben wird oder dass Gesellschaft nicht als eigene Kategorie begriffen, sondern auf die Summe der Individuen reduziert wurde. Kritisiert wurde auch, dass klassische Verständnisse zwar als eine zentrale notwendige Bedingung zur Erreichung von Selbstbestimmung die Abwesenheit von Zwang bestimmten, zugleich aber die Verwirklichungsbedingungen dieser Selbstbestimmung unberücksichtigt ließen. Denn wenn eine Person zwar rein formal alles wählen kann, aber keinerlei Möglichkeiten hat, diese Chancen auch tatsächlich zu ergreifen, könne von Freiheit nicht die Rede sein. Andere, insbesondere sozialpsychologisch motivierte Kritiken verwiesen schließlich auf die empirische Unhaltbarkeit des Konzepts der individuellen Selbstbestimmung, da jedes Individuum Teil eines dichten zwischenmenschlichen Netzwerks sei, in dessen Rahmen allenfalls von einer miteinander verschränkten Selbstbestimmung in Abhängigkeit von den jeweiligen Netzwerken bzw. sozialen Kontexten die Rede sein könne.

Im Folgenden möchten wir unser Verständnis von Selbstbestimmung darlegen. Wir verstehen das Individuum als soziales Wesen, das zwar mit gewissen prä-sozialen (biologischen/genetischen) Eigenschaften ausgestattet (Hunger- und Schlaftrieb, ein gewisses Maß an Autonomieverlangen, usw.) ist, aber insbesondere in Bezug auf seine Identität und sein Bewusstsein weitgehend von seiner sozialen Umwelt abhängig ist. Das bedeutet zum Beispiel, dass Individuierung bzw. Selbstbestimmung kein abgeschlossener Zustand, sondern ein Prozess ist, der, abhängig vom jeweiligen sozialen Kontext, ganz unterschiedlich gelebt werden kann.

Ein entscheidendes Stichwort in diesem Zusammenhang lautet Vertrauen. Individuen sind in ein Netz aus unterschiedlichen gesellschaftlichen Kontexten und Beziehungen zu anderen Menschen und Gegenständen eingebettet, in denen sie stets einen Teil ihrer Autonomie notwendigerweise abgeben müssen. Dies zeigt sich besonders deutlich im Kontext hochkomplexer informationstechnischer Systeme, zu denen auch KI-basierte Technologien hinzuzuzählen sind. Selbst technisch versierte und interessierte Individuen sind in aller Regel außer Stande, die Funktionsweise aller von ihnen genutzten Technologien und Objekte vollends zu durchdringen. Die Nutzung informationstechnischer Systeme basiert stattdessen zu einem Großteil auf dem Vertrauen in Programmierer_innen, Firmen, die Hardware und Infrastruktur bereitstellen, andere Nutzer_innen, Expert_innen, die Code und Funktion überprüfen und viele mehr. Die Entscheidung, zu vertrauen, ist teilweise eine individuelle, weil jede Person entsprechende Fragen für sich selbst beantworten muss. Diese individuelle Entscheidung ist jedoch stets auch Produkt der Umwelteinwirkungen, Normen, die im Laufe einer Sozialisierung vermittelt werden,

Ge- und Verbote, die den individuellen Handlungsrahmen vorstrukturieren, usw. Wir verstehen Selbstbestimmung somit als eine gesellschaftliche Praxis, die im Wechselverhältnis zwischen Individuum und Gesellschaft stattfindet.

Selbstbestimmung steigern

Es können verschiedene Bereiche identifiziert werden, in denen sich KI-Technologien förderlich auf die menschliche Selbstbestimmung auswirken. Prominent ist in diesem Kontext der Einsatz von KI-Technologien im Gesundheitswesen, wo sie oft der Verbesserung von Diagnosemethoden sowie der Individualisierung von Therapien dienen. Menschen mit Seh- oder Hörbehinderungen können zudem Text-to-Speech- oder Speech-to-Text-Technologien einsetzen, um Wahrnehmungseinschränkungen zu kompensieren. Digitale Assistenzsysteme aller Art, sowohl innerhalb als auch außerhalb des Pflege- und Gesundheitssystems, wirken darüber hinaus selbstbestimmungsfördernd.

Menschen können zudem mehr Freiheit erlangen, indem sie lästige, gefährliche oder anderweitig unerwünschte Aufgaben an KI-Systeme delegieren. Hierzu gehört klassischerweise das Feld der Robotik, durch das prekäre oder gefährliche Aufgaben, etwa solche in lebensfeindlichen Umgebungen, durch automatisierte technische Systeme übernommen werden können. Fortschritte vor allem im Bereich des maschinellen Sehens, aber auch beim verstärkenden Lernen, einem Unterbereich des maschinellen Lernens, ermöglichen die Konstruktion immer komplexerer und eigenständiger agierender Systeme.

Gerade in der Arbeitswelt sind mit dieser Entwicklung neben Hoffnungen auch Ängste verbunden. Die Hoffnungen beziehen sich vor allem auf eine für die Zukunft antizipierte Reduktion durchschnittlicher Arbeitszeiten sowie einen Freizeitgewinn durch die Delegation von Hausarbeit auf maschinelle Helfer. Wenn, einfach gesagt, KI und Robotik für eine Automatisierung der Produktion lebensnotwendiger Güter sorgen, kann nicht nur eine bessere Balance aus Arbeit und Erholung gefunden werden, sondern es werden gleichsam Handlungs- und Selbstverwirklichungspotenziale freigesetzt, die in kulturelle, sportliche, künstlerische oder anderweitige soziale Tätigkeiten münden können, welche bislang durch zeitintensive Erwerbsarbeit vernachlässigt worden sind. Diese Vision eines Gewinns von Selbstverwirklichungspotentialen in der Arbeitswelt bestimmt das philosophische Denken bereits seit der Antike. Aus heutiger Sicht gesprochen sowie unter Beachtung wirtschaftswissenschaftlicher Erkenntnisse wird es durch KI zwar zu einem starken Strukturwandel der Arbeitswelt kommen – insbesondere im Hinblick auf routinemäßige und durch geringe Komplexität oder durch soziale Intelligenz ausgezeichnete Arbeitsfelder –, mit sehr hoher Wahrscheinlichkeit aber nicht zu einer utopischen Freisetzung menschlicher Schaffenskraft außerhalb tradierter Wirtschaftsprozesse.

Selbstbestimmung einschränken

Neben positiven Einflüssen auf Selbstbestimmungspotenziale, wie sie im vorherigen Kapitel angesprochen wurden, können KI-Technologien gleichzeitig einschränkend wirken. In diesem Zusammenhang werden Beispiele genannt und kritisch diskutiert. Diese reichen von synthetischen Medien – das sind generierte oder manipulierte textuelle, auditive oder (audio-)visuelle Medienbeiträge, z. B. so genannte Deepfakes - über Technologien der KI-gestützten Verhaltensmanipulation bis hin zu besonders privatheitsverletzenden Technologien der probabilistischen Erkennung „interner Zustände".

Medienproduktion

Vor allem, aber nicht nur dank sogenannter „Generative Adversarial Networks" (GAN) sind in den letzten Jahren massive Fortschritte bei der automatisierten Generierung synthetischer Medien erreicht worden. GANs bestehen aus zwei gegeneinander antretenden neuronalen Netzen, von denen das erste „Generative Network" beispielsweise ein artifizielles Bild erstellt, während das zweite „Discriminating Network" („Discriminating" im Sinne von deutsch „eine Unterscheidung treffen") auf der Basis eines Trainingsdatensatzes mit „echten" Bildern den „Realitätsgrad" des artifiziellen Bildes beurteilt und so das generierende Netz optimiert. So können schlussendlich fotorealistische, aber gänzlich fiktive Bilder erstellt werden. Aber nicht nur Bilder, auch realistische Videos, Audio-Dateien oder Texte können durch KI-Systeme generiert werden. Da die Kosten für derartige Fälschungen weit unter denjenigen liegen, die für die Erstellung herkömmlicher Fälschungen nötig waren, können massenhaft synthetische Medien erzeugt und beispielsweise durch soziale Medien in Umlauf gebracht werden mit dem Ziel, Publika und Meinungsbilder zu beeinflussen, politische Gegner zu diskreditieren, zu polarisieren, Ängste und Konflikte zu schüren oder Sachverhalte zu verzerren. In diesem Kontext haben unter anderem Desinformationskampagnen in den letzten Jahren einen gewissen Strukturwandel der Öffentlichkeit herbeigeführt, der in Teilen mit einer Erosion demokratischer Werte wie beispielsweise der sinnvollen Nutzung des Rechts auf Meinungsfreiheit einherging. Diese Entwicklungen stellen ein Problem sowohl für den Wert der individuellen Selbstbestimmung als auch der demokratischen Selbstorganisation dar. Schließlich können demokratisch verfasste Gesellschaften nur dann gelingend selbstorganisiert werden, wenn Wahlentscheidungen auf überprüf- und kritisierbaren Informationen basieren und dementsprechend durch mündige Personen auf Basis einer freien und qualitativ hochwertig informierten Meinungs- und Willensbildung getroffen werden.

Synthetische Medien wirken sich noch in einem weiteren Sinne einschränkend auf die Selbstbestimmung einzelner Personen aus: Durch sie ist es möglich geworden, gänzlich neue Dimensionen des Identitätsdiebstahls zu eröffnen. Bei den dabei verwendeten Technologien handelt es sich um die KI-gestützte Generierung von fremden Stimmaufzeichnungen, deren genauer Wortlaut durch beliebig geschriebenen Text vorgegeben werden kann. Oder es handelt sich um „Deepfakes", also künstlich generierte Videos zumeist prominenter Personen. Derlei Fälschungen eröffnen neue Möglichkeiten der Cyberkriminalität und des Social Engineerings, etwa beim Telefonbetrug oder der Stimmauthentifizierung und sind mit aktuellen Mitteln schwer zu erkennen.

Verhaltensmanipulation

Dass KI-gestützte Massenbeeinflussung funktioniert, ist durch verschiedene Studien deutlich geworden. In einer Studie wurden beispielsweise mehr als drei Millionen Facebook-Nutzer_innen mit personalisierter Werbung angesprochen, die in Abhängigkeit von psychometrisch ermittelten Persönlichkeitseigenschaften individualisiert wurde. Bei

Werbung, die auf das psychologische Profil, also etwa die gemessene Extra- oder Introvertiertheit einer Person abgestimmt war, ergab sich eine signifikant höhere Klick- und Kaufrate als bei nicht-individualisierter Werbung. Dass ein solches auf Affekte oder subtile Instinkte abzielendes „Micro-Targeting" vermutlich im großen Stil funktioniert, zeigen möglicherweise auch die Wahlbeeinflussungen der letzten Jahre, sei es in der Leave.EU-Kampagne in Großbritannien oder dem US-Wahlkampf der Republikanischen Partei 2016. Unternehmen wie Cambridge Analytica haben sich auf diese Wahlbeeinflussung spezialisiert. Maschinelles Lernen hilft bei der digitalen Psychografie, also der genauen Auslesung der Persönlichkeit eines Menschen. Wer diese kennt, besitzt gleichzeitig Manipulationsmöglichkeiten. Ängste und psychologische Schwächen oder Erwartungen können ausgenutzt werden. KI-gestütztes psychometrisches Vermessen von Persönlichkeiten und das daraufhin erfolgende Anpassen und Individualisieren von Marketingbotschaften oder Wahlkampfmotiven steht in fundamentalem Widerspruch zur Idee menschlicher Selbstbestimmtheit. Waren persuasive Botschaften bei der Vermittlung durch vor-digitale Technologien klar als solche erkennbar, helfen KI-gestützte Auswertungsverfahren und inhaltlich individualisierte Onlineplattformen bei der unauffälligen oder verdeckten Verhaltensbeeinflussung.

Scoring

Bereits Anfang der 2000er-Jahre wurde mit dem Begriff des „Social Sortings" eine neue Qualität der digitalen Massenüberwachung beschrieben, bei der Menschen durch technische Verfahren in Reputations- oder Risikoklassen eingeteilt wurden. Auf Basis dieser Einteilungen lassen sich verschiedene „Zukünfte" für die betroffenen Personen ableiten, also verschiedene Möglichkeiten der Selbstentfaltung in unterschiedlichen sozialen Feldern. Während die technische Grundlage für das „Social Sorting" anfangs einfaches Data-Mining war, werden heute Verfahren des maschinellen Lernens eingesetzt. Das „Social Sorting" wurde zum „Social Scoring" ausgebaut, bei dem algorithmische Entscheidungssysteme automatisiert Punktwerte mit bestimmten Verhaltensweisen, die digital erfasst werden, verbinden. Als problematisch ist dabei neben vielen anderen Faktoren zu erachten, dass die eingesetzten technischen Verfahren Ergebnisse produzieren, die lediglich probabilistische Projektionen sind. Dennoch haben diese Projektionen reale Folgen, da sie Realität mehr konstruieren, als abbilden. Bonitäts-Scores, Gefährdungsindexe, Branchenscores, akademische Scores, Rückfälligkeits-Scores, personalisierte Preise und derlei mehr beeinflussen - egal ob faktisch korrekt oder nicht - den Grad sowohl möglicher individueller als auch kollektiver Selbstbestimmung, sei dies im Hinblick auf finanzielle Freiheiten, auf Mobilitätseinschränkungen, auf Bildungschancen etc.

Datenschutzrechtlich gefasst und einem grundsätzlichen Verbot unterworfen sind dabei nur Entscheidungen, die ausschließlich auf einer automatisierten Verarbeitung personenbezogener Daten basieren. Nicht erfasst sind diejenigen Risiken, die durch eine teilautomatisierte Entscheidung bedingt sind, bei der zwar ein Mensch die finale Entscheidung trifft, dabei jedoch das Ergebnis der automatisierten Entscheidungsvorbereitung faktisch übernimmt und damit nur formal zum_zur Entscheider_in wird. Hier bestehen rechtliche Schutzlücken.

Zudem erfasst das Datenschutzrecht bezüglich der automatisierten Entscheidung im Einzelfall nur jene Entscheidungen, die Rechtswirkung entfalten oder die betroffene Person auf ähnliche Weise erheblich beeinträchtigen. Nicht erfasst ist damit beispielsweise die automatisierte Beschränkung von Zahlungsmöglichkeiten im E-Commerce.

Qualitative Anforderungen an Scoring enthält § 31 BDSG. Danach müssen die zur Berechnung des Wahrscheinlichkeitswerts genutzten Daten unter Zugrundelegung eines wissenschaftlich anerkannten mathematisch-statistischen Verfahrens nachweisbar für die Berechnung der Wahrscheinlichkeit des bestimmten Verhaltens erheblich sein. Die

Vorschrift umfasst allerdings nur die Entscheidung über die Begründung, Durchführung oder Beendigung eines Vertragsverhältnisses mit einer Person.

Wesentliches Problem des Scorings ist seit jeher die Intransparenz des Verfahrens. Hier besteht ein Spannungsfeld zwischen Transparenzpflichten und dem Geheimhaltungsinteresse des Scoring-Anbieters oder der Scoring-Anbieterin bezogen auf die verwendete Score-Formel. Diese Formel ist als Geschäftsgeheimnis geschützt. Beim Scoring auf Basis von künstlicher Intelligenz verschärft sich diese Problematik noch. Das Datenschutzrecht fordert gegenüber der betroffenen Person die Offenlegung „aussagekräftiger Informationen über die involvierte Logik". Wie dies gerade bei selbstlernenden Systemen, bei denen vielleicht nicht einmal die Programmierer_innen die Entstehung der Logik nachvollziehen können, realisiert werden kann, bleibt eine offene Frage.

Interne Zustände

KI hilft, um aus „unverdächtigen" Datenspuren wie etwa den Beschleunigungsdaten des Smartphones, der Nachverfolgung von Klicks (Clickstreams), Gesichtsabbildungen etc. intime Informationen mit hoher Genauigkeit abzuleiten. Möglich wird dies durch computergestützte Mustererkennung, bei der aus Datensätzen aus der Kombination von Merkmalen Strukturen erkannt werden, die wiederum bei Datensätzen mit eingeschränktem Merkmalsumfang helfen, auf unbekannte Faktoren zu schließen. Beispielsweise soll beim Vorliegen von Gesichtsabbildungen und Informationen über die jeweilige sexuelle Orientierung von Personen durch Mustererkennung probabilistisch auf die sexuelle Orientierung von Personen geschlossen werden können, von denen nur Gesichtsabbildungen vorliegen. Diese mächtige Technologie, mit der auf Aspekte wie Drogenkonsum, Beziehungsstatus, Intelligenzquotient, Persönlichkeitseigenschaften und vieles mehr probabilistisch geschlossen werden kann, stellt eine Gefahr für die Selbstbestimmung von Menschen dar, welche in bestimmten Fällen gerade auf das Nichtwissen von privaten Informationen bei Dritten angewiesen sind, etwa zum Schutz vor Diskriminierung. Moderne KI-Technologien bieten jedoch „superhumane" Analysemethoden, bei der nicht nur über tagtägliche Datenspuren, sondern gleichsam über Eye-Tracking, über Gang- oder Körperspracheanalysen, über Gesichtsausdrücke oder andere physiologische Parameter auf eigentlich „geheime" interne Zustände oder intime Persönlichkeitseigenschaften geschlossen werden kann. Basierend auf diesen Technologien ist ein neuer Industriezweig erschlossen worden, welcher insbesondere durch Sicherheitstechnik Milliardenumsätze macht.

Der Erfolg gerade von Gesichtserkennungstechnologien muss jedoch kritisch betrachtet werden, und zwar auf zwei Ebenen: Zum einen hinsichtlich der Beschneidung des Werts der Selbstbestimmung, die sich bereits in Bezug auf Beobachtungs- oder Selbstzensur-Effekte („chilling-effects") beim Einsatz von Gesichtserkennungstechnologien im öffentlichen Raum manifestiert. Zum anderen müssen die Technologien methodisch hinterfragt werden. Denn die immer wieder getroffene Behauptung, man könne Emotionen technisch aus Gesichtsausdrücken oder -bewegungen auslesen, ist falsch. Der Link zwischen Gesichtsausdruck und Emotion ist weder spezifisch – derselbe Gesichtsausdruck verweist nicht zuverlässig auf dieselbe Emotion – noch zuverlässig – dieselben Emotionen werden nicht immer durch denselben Gesichtsausdruck angedeutet –, oder generalisierbar – es gibt je nach Kontext und Kultur Unterschiede in bestimmten Gesichtsmimiken. In der Emotionspsychologie wird darüber hinaus bereits seit Jahrzehnten auf Basis empirischer Studien debattiert, ob Emotionen und Mimik überhaupt „hartverdrahtet" verbunden sind oder ob Mimik sich nicht vielmehr nach der sozialen Situation richtet. Diese Feststellung sollte den Einsatzbereich und die Leistungsversprechungen von Gesichtserkennungssoftware stark einschränken bzw. relativieren, verhindert aber natürlich nicht ihre Anwendung in diesem Feld. Vor allem in Bezug auf die Selbstbestimmung ist dabei allerdings relevant, dass die Gefahr falscher Schlüsse wächst, wenn von falschen Voraussetzungen und wissenschaftlich nicht abgesicherten psychologischen Vorstellungen ausgegangen wird.

Zukunftsperspektiven

Unabhängig von aktuell bestehenden KI-basierten Anwendungen ist der wissenschaftliche Diskurs reich an Visionen, Utopien sowie Dystopien über den Entwicklungsverlauf zukünftiger KI-Technologien. Insbesondere steht in diesem Zusammenhang die Schließung der Differenz zwischen schwacher und starker KI zugunsten letzterer im Fokus: Während schwache KI unflexibel auf einen bestimmten Bereich der Mustererkennung spezialisiert ist und nicht auf andere Bereiche übertragen werden kann, entfällt diese Limitierung bei starker KI, welche als allgemeine Problemlösungsmaschine eingesetzt werden kann. Die Befürchtung oder Hoffnung in diesem Kontext ist, dass in nicht allzu ferner Zukunft eine „superintelligente" Maschine menschliche Informationsverarbeitungskapazitäten übersteigt. Während ein solches Zukunftsszenario in der engeren Forschung zum maschinellen Lernen tendenziell als unrealistisch betrachtet wird, ist in der breiteren KI-Forschungsgemeinschaft, wie vor einigen Jahren in einer Umfrage herausgefunden wurde, durchaus ein signifikanter Anteil an Wissenschaftler_innen überzeugt, dass Technologien mit starker KI zukünftig Realität werden.

Unter dieser Voraussetzung könnte starke KI nicht nur zur Intelligenzverstärkung, dem kognitiven „Enhancement", zum Hacking oder zur Steigerung ökonomischer Produktivität eingesetzt werden, sondern auch zur sozialen Manipulation. Welche genauen Gefahren sich daraus ergeben, ist noch nicht absehbar. Zentral sind hier die Diskussionen über das Kontrollproblem. Mit verschiedenen technischen Ansätzen sollen - mit dem Zweck der Bewahrung menschlicher Selbstbestimmung - Systeme starker KI so unter Kontrolle gehalten werden, dass das Moment der Erlangung vollständiger technischer Autonomie verhindert wird. „Boxing"-Ansätze, in denen sowohl physische als auch informationelle Einhegungen von KI-Artefakten überlegt werden, werden neben Ansätzen der zur Lenkung von KI-Systemen getätigten Anreizsetzung oder der Vordefinition von bestimmten technischen Motivationsstrukturen diskutiert. Alle diese Bemühungen um KI-Sicherheit haben langfristig das Ziel der Erhaltung menschlicher und der Verhinderung technischer Selbstbestimmung. Dennoch muss nochmals der rein spekulative Charakter dieser Diskussionen betont werden.

In nicht weniger spekulativen post- oder transhumanistischen Diskursen, in denen die „conditio humana" als solche aufgelöst wird und die Definition dessen, was ein Mensch ist, einer fundamentalen Umdeutung unterliegt, gilt Ähnliches für das traditionelle Verständnis menschlicher Selbstbestimmung. Wenn Menschen nur noch als „Anhängsel" einer globalen, superintelligenten Maschine verstanden werden, oder wenn sie mit maschinellen Strukturen verschmelzen, sodass evolutionär ausgehärtete Eigenschaften und Veranlagungen durch technische Erweiterungen disponibel werden, stellt sich die Frage sowohl individueller als auch kollektiver Selbstbestimmung auf andere Weise. Schließlich geht es dann weniger um den Erhalt menschlicher Würde inklusive des Wertes menschlicher Autonomie, als vielmehr um die Entfaltung eines technischen „Willens", der sich durch die Selbstreparatur, Selbstverteidigung, den Selbsterhalt, die Selbsterweiterung und Selbstkontrolle technischer Strukturen auszeichnet. Inwiefern derlei Spekulationen aus der Zukunftsforschung, dem Post- oder Transhumanismus bereits handlungsleitend für aktuelle politische Entscheidungen sein sollten, kann durchaus skeptisch gesehen werden. Schließlich gibt es, wie oben beschrieben, bereits heute handfeste Probleme bei der Unterminierung individueller und kollektiver Selbstbestimmung durch KI-Systeme, deren Lösung große Herausforderungen der politischen Steuerung mit sich bringen.

Fazit und Empfehlungen

Bei allen Technologien, welche zu verschiedenen Einsatzzwecken geeignet sind, liegen Chancen und Risiken dicht beieinander. Erstere gilt es zu fördern, letztere zu minimieren; so auch bei KI-Anwendungen. An dieser Stelle sollen einige Empfehlungen formuliert werden, welche sich auf den Teilbereich der Erhaltung des Werts der menschlichen Selbstbestimmung fokussieren. Grundsätzlich unterschieden werden kann zwischen Empfehlungen, die auf die Selbstbestimmung des Individuums fokussieren und solchen, die eine Steigerung der gesellschaftlichen Selbstbestimmungsfähigkeit zum Ziel haben. Zu beachten ist, dass beide Ziele je nach Kontext regelmäßig in Konflikt miteinander geraten können.

Technologien der Verhaltensmanipulation beispielsweise können demokratisch ausgehandelt und dazu eingesetzt werden, Menschen zu einer bestimmten gesellschaftlich erwünschten Handlungsweise anzutreiben („Musst Du heute trotz guten Wetters mit dem Auto zur Arbeit fahren?“), sie können dabei aber auch individuelle Wünsche übergehen und bei den Betroffenen Kritik und Gefühle von Kontrollverlust etc. hervorrufen. Derartige Konfliktfälle bleiben auch in einer auf KI-Anwendungen bauenden Gesellschaft Gegenstand öffentlicher Debatten. Zudem sollte trotz aller Typisierungen und argumentativ notwendigen Verkürzungen nicht vergessen werden, dass in einem hochkomplexen technologischen Feld wie der KI auch weiterhin zumindest zwischen verschiedenen Anwendungsfeldern und noch besser zwischen einzelnen Anwendungen und Einzelfällen unterschieden werden sollte. Was in einem Kontext als unerfreuliche Einmischung in individuelle Angelegenheiten oder kollektive Handlungspraktiken gelten mag, kann in einem anderen Kontext durchaus erwünscht sein. Nichtsdestotrotz sollen die folgenden Empfehlungen einen Handlungsrahmen umreißen, der zur Gewährleistung eines anhaltenden gesellschaftlichen Diskurses zum Thema KI und Selbstbestimmung beitragen kann.

Zunächst ist zu beachten, dass es eine starke Skalierbarkeit von KI-Technologien gibt. Das heißt, dass der Anwendungsbereich derselben potenziell deutlich größer ist als derjenige von nicht-digitalen Technologien. Die hohe Skalierbarkeit bedeutet aber auch eine starke Skalierbarkeit von Chancen oder Risiken.

Ausreichende Vielfalt von algorithmischen Entscheidungssystemen gewährleisten

In diesem Sinne ist grundsätzlich eine ausreichende Vielfalt von algorithmischen Entscheidungsfindungssystemen ein wichtiger Beitrag zur Gewährleistung von Entscheidungsfreiheit in der Wahl eines Systems. Diese Vielfaltssicherung hat auch Einfluss auf die Verwirklichung von Fairness in KI-Systemen. Wenn etwa bei der automatisierten Vorauswahl von Bewerber_innen bei Unternehmen immer dieselbe, potenziell diskriminierende Software zum Einsatz kommt, schränkt dies die freie Entfaltung unzähliger Personen auf unlautere Weise ein. Andererseits kann eine solche Software aber auch einen Entwicklungsstand erreichen, auf dem der Grad der Diskriminierung seitens der jeweiligen Technologie niedriger ist, als die Diskriminierung seitens eines Menschen. Entscheidend werden aller Voraussicht nach die ethischen, rechtlichen und technischen Anforderungen sein, die an derartige Systeme gestellt werden. Diese gilt es zwar demokratisch festzulegen, jedoch zugleich ausreichend viel Raum für die individuelle Entfaltung eines jeden Menschen übrig zu lassen.

Algorithmische Entscheidungssysteme sollten demokratisch festgelegten Fairnesskriterien entsprechen

Dies kann etwa dadurch erreicht werden, dass algorithmische Entscheidungssysteme demokratisch festgelegten Fairnesskriterien entsprechen, die zum Ziel haben, Verzerrungen in der Datenbasis der Systeme zu verhindern, Widerspruchsmöglichkeiten gegen KI-basierte Entscheidungen zu etablieren, sowie das Zustandekommen algorithmischer Entscheidungen so transparent und nachvollziehbar wie möglich zu erklären.

Zwei entscheidende Fragekomplexe in diesem Zusammenhang sind: Welche Anwendungen oder Anwendungsfelder müssen überhaupt berücksichtigt werden und wer

legt diese fest? Und welcher Akteur ist für die Umsetzung der Fairnesskriterien zuständig?

Risikoadaptiven Regulierungsansatz wählen

Wir schlagen diesbezüglich grundsätzlich eine Orientierung an den Überlegungen zum Thema Kritikalität der Datenethikkommission (DEK) vor. Diese sieht einen risikoadaptierten Regulierungsansatz für algorithmische Systeme vor, der nicht nur Fragen der Fairness umfasst, sondern auch Gefährdungen der Selbstbestimmung oder der Gesundheit. In diesem und ähnlichen Ansätzen der Technikfolgenbewertung gehen mit einem steigenden Schädigungspotenzial wachsende Anforderungen und Eingriffstiefen der regulatorischen Methoden einher. Regulierungen kommen nicht zur Anwendung, wenn es sich um Anwendungen ohne oder mit sehr geringem Schädigungspotential handelt. Andernfalls reichen die regulatorischen Methoden von Transparenzpflichten über Risikoabschätzungen und Ex-post-Kontrollen im Falle von Anwendungen mit einem gewissen Schädigungspotenzial bis hin zu Ex-ante-Zulassungsverfahren bzw. zu Verboten in Fällen mit deutlichem Schädigungspotenzial oder mit unvertretbarem Schädigungspotenzial. Die DEK empfiehlt, dass der Gesetzgeber ein Prüfschema festlegt, nach welchem die Kritikalität eines algorithmischen Systems gemessen werden kann. Nicht adressiert bleibt indes, welche Instanz über die Zuordnung von Anwendungsfeldern und einzelnen Anwendungen in die jeweiligen Kritikalitätsstufen entscheidet. Für die Aufsicht der Einhaltung der Maßnahmen, die im Rahmen der verschiedenen Stufen zu treffen sind, sieht die DEK insbesondere sektorale Aufsichtsbehörden in der Pflicht, deren Arbeit von einem zu gründenden bundesweiten Kompetenzzentrum Algorithmische Systeme unterstützt werden soll. Darüber hinaus wird aber auch zivilgesellschaftlichen Initiativen sowie Formen der Ko- und Selbstregulierung Raum gegeben.

Identifikation der Risiken seitens der Produkt- bzw. Technologiehersteller …

Die ungeklärte Frage über die Zuständigkeit der Zuordnung von Anwendungsfeldern und Anwendungen zu den Kritikalitätsstufen könnte beispielsweise durch eine Verlagerung der Verantwortlichkeit hin zu den Produkt- bzw. Technologieherstellen beantwortet werden. Ähnlich wie im Aufsichtssystem der DSGVO wären dann die einzelnen Unternehmen und andere Organisationen in der Pflicht, eine erste interne Folgenabschätzung ihrer Systeme vorzunehmen und darauf aufbauend die Einstufung selbst vorzunehmen. Falls Unklarheit über die Zuordnung zu einer Kritikalitätsstufe besteht, müssten die Unternehmen einen Anspruch auf Beratung seitens der jeweils zuständigen Aufsichtsstelle haben. Dieses Vorgehen würde die Aufsichtsinstanzen entlasten, da sie nicht jeden Fall selbst prüfen müssten. Zugleich erhielten die Unternehmen mehr Freiheiten, da keine externe Instanz, sondern sie selbst für die Zuordnung verantwortlich wären. Im Falle von Unsicherheiten würde der Beratungsanspruch wiederum garantieren, dass ein Unternehmen im Zweifelsfall nicht ohne Unterstützung agieren muss. Schließlich müsste für offensichtlich beabsichtigte Verstöße, Fehleinstufungen usw. ein Sanktionssystem eingeführt werden, das abschreckend wirkt und dem Missbrauch der Verlagerung der Verantwortung hin zu den Unternehmen wirksam entgegenwirkt.

… flankiert durch die Beratung seitens der Aufsichtsstellen …

… sowie die Beteiligung nichtstaatlicher Initiativen …

Zudem sollten nichtstaatliche Initiativen nicht nur – wie von der DEK vorgeschlagen – bei der Festlegung von technisch-statistischen Standards für die Qualität von Testverfahren und Audits beteiligt werden, sondern je nach Kontext durchaus auch generell die Rolle eines Interessenvertreters ausfüllen. In für Verbraucherthemen relevanten Fällen könnte diese Rolle von Verbraucherschutzorganisationen ausgefüllt werden, bei arbeitnehmerrelevanten Themen könnte diese Rolle den Gewerkschaften zukommen. Selbiges gilt für Unternehmen: (Dach-)Verbände und Berufsvereinigungen sollten die Möglichkeit erhalten, für die jeweilige Gruppe oder ihren Sektor Prüfkriterien, Verhaltensregeln und anderes festzulegen.

Diskussion um die Weiterentwicklung der rechtlichen Rahmenbedingungen fortführen

Ferner muss gerade das Recht die technische Entwicklung in diesem Bereich kritisch begleiten. Dass mit dem Recht der betroffenen Person aus Art. 22 DSGVO, nicht einer ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher

Weise erheblich beeinträchtigt, der alte Art. 15 der 1995 in Kraft getretenen Vorgängerrichtlinie 95/46/EG fast wörtlich übernommen wurde, zeigt, dass die Diskussion um die Weiterentwicklung der rechtlichen Rahmenbedingungen nicht stehenbleiben darf. Zentrale Frage ist dabei, wie grundrechtlich geschützte Positionen eine konkrete technische Umsetzung erfahren können. Dies ist umso wichtiger, wenn künstliche Intelligenz durch staatliche Stellen zum Einsatz kommen sollte. Spätestens dann muss ein diskriminierungsfreier und die Selbstbestimmung der Bürger_innen wahrender Einsatz sicher gewährleistet werden. Eine besondere Rolle spielt dabei das Verbot, dass staatliche Stellen keinesfalls teilweise oder weitgehend vollständige Persönlichkeitsbilder der Bürger_innen erzeugen dürfen, denn die gänzliche oder teilweise Registrierung und Katalogisierung der Persönlichkeit stellt einen nicht zu rechtfertigenden Eingriff in die Würde des Menschen dar. Aber auch auf einfachgesetzlicher Ebene stellt künstliche Intelligenz das Recht vor enorme Herausforderungen, die abseits des Datenschutzrechts von Fragen der Haftung über Schöpfung durch KI bis hin zur Zulässigkeit von Legal Tech reichen und ebenso vielfältig sind, wie die potenziellen Einsatzbereiche von künstlicher Intelligenz.

Hieraus ergibt sich ein umfassender Regelungsbedarf, um die Selbstbestimmung der betroffenen Personen zu erhalten und zu fördern. Um den Gestaltungs- und Regelungsbedarf von KI zu eruieren, muss zunächst für jeden Anwendungsbereich bestimmt werden, wie KI die Selbstbestimmung verändert, beeinflusst, einschränkt oder aber erweitert. Gleichzeitig müssen alle anderen verfassungsrechtlich verbrieften Rechte gewährleistet werden. Vor allem muss sichergestellt sein, dass die betroffenen Personen durch den Einsatz von KI nicht diskriminiert werden, was bereits die Verwendung entsprechend möglichst diskriminierungsfreier Trainingsdaten erfordert.

Selbstbestimmung muss auch beim Einsatz von KI gewährleistet sein, indem …

KI muss letztlich so gestaltet sein, dass die betroffene Person auch beim Einsatz der KI selbstbestimmt entscheiden kann, ob und welche personenbezogenen Daten verarbeitet werden und ihre Betroffenenrechte geltend machen kann. Ist dies nicht gewährleistet, würde die Person zum bloßen Datenobjekt degradiert. Hierfür müssen verschiedene Grundsätze Beachtung finden:

… der Zweckbindungsgrundsatz …

Der Zweckbindungsgrundsatz nach Art. 5 Abs. 1 lit. b DSGVO muss auch beim Einsatz von KI gewährleistet sein. Die Anwendung muss sicherstellen, dass eine über den legitimen Zweck hinausgehende Datenverarbeitung nicht möglich ist. Dies sollte auch für die Verarbeitung von Verhaltensdaten ohne unmittelbaren Personenbezug gelten, die über probabilistische Musterbildung aber auf eine selbstbestimmte Lebensführung einschränkend zurückwirken können. Der Zweck sollte möglichst konkret und durch eine entsprechende technologiespezifische Rechtsgrundlage festgelegt werden und nicht allein der Wahl des Verantwortlichen überlassen bleiben.

… der Grundsatz der Datensparsamkeit …

Auch der Grundsatz der Datenminimierung nach Art. 5 Abs. 1 lit. c DSGVO – oder besser noch der Grundsatz der Datensparsamkeit – muss Beachtung finden. Gerade für Anwendungen, die in erheblichem Maß auf die dauerhafte Ansammlung und Verwendung großer Mengen von Daten ausgerichtet sind, müssen entsprechende Grenzen definiert und kontrolliert werden können.

… und Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen Beachtung finden

Die Pflicht zum Ergreifen technischer und organisatorischer Maßnahmen, Datenschutz durch Technikgestaltung sowie datenschutzfreundlicher Voreinstellungen nach 25 Abs. 1 und 2 DSGVO sowie deren regelmäßige Kontrolle und Anpassung nach Art. 24 Abs. 1 Satz 2 DSGVO gelten auch für die Gestaltung von KI und konkretisieren gleichzeitig den Grundsatz des Systemdatenschutzes nach Art. 5 Abs. 1 lit. f DSGVO. Neben der Sicherstellung der oben genannten datenschutzrechtlichen Grundsätze muss durch technische und organisatorische Maßnahmen auch verhindert werden, dass sich KI selbst weiterentwickelt und so Datenschutzgrundsätze umgeht oder andere Schutzmechanismen aushebelt. Auch bedarf es technischer Interventionsmöglichkeiten, die es den be-

Weiterentwicklung von KI, die zur Umgehung der Datenschutzgrundsätze führen kann, wirksam vermeiden

troffenen Personen ermöglichen, die Grundlagen der Datenverarbeitung nachzuvollziehen, Voreinstellungen zu ändern und Betroffenenrechte, vor allem das Recht auf Löschen, auszuüben.

Transparenz und Nachvollziehbarkeit algorithmischer Entscheidungen gewährleisten

Die KI muss im Sinne des Art. 5 Abs. 1 lit. a DSGVO zudem transparent, nachvollziehbar und erklärbar sein. Die betroffene Person muss darüber informiert sein, dass und unter welchen Voraussetzungen KI zum Einsatz kommt und die Grundlagen des Entscheidungsvorgangs und der Logik des KI-Systems nachvollziehen können. Hier muss geklärt werden, auf welchem Wege algorithmische Entscheidungen für nicht vorgebildete Nutzer_innen so transparent und nachvollziehbar wie möglich erklärt werden können. Dabei reicht es nicht aus, dass Personen gemäß Art. 13 Abs. 2 lit. f über die allgemeine Logik von KI-basierten Technologien informiert werden; auch müssen die konkreten normativen Implikationen transparent gemacht werden, die etwa den Entscheidungen von Algorithmen zugrunde liegen. Nur so können KI-basierte Verfahren kritisiert werden.

Ansätze zur kollektiven Beteiligung an der Gestaltung von KI entwickeln …

Schließlich ist es unerlässlich, dass Selbstbestimmung auch im Sinne einer kollektiven Beteiligung an der Gestaltung von KI verstanden wird. So läuft die KI-Gestaltung meist in den von der Internetökonomie festgelegten Bahnen ab und wird nur selten mit normativ reichhaltigeren Alternativen konfrontiert. Gegenwärtig wird die Beteiligung von Verbraucher_innen an der KI Gestaltung auf die Bereitstellung von Trainingsdaten für KI reduziert. Um eine demokratisch gehaltvolle Selbstbestimmung zu ermöglichen, sind indes Ansätze notwendig, welche die kritischen Bewertungskompetenzen von Verbraucher_innen bei der Entwicklung von KI und ihrer Nutzung nicht nur erhalten, sondern auch zielgerichtet fördern. Weiterhin gibt es einen Bedarf an institutionellen Rahmenbedingungen für öffentliche Verfahren, die eine gesellschaftsweite Problematisierung über die normativen Grundlagen von KI ermöglichen und die Inklusion sonst marginalisierter Gruppen in der KI-Gestaltung gewährleisten. Hierzu ist eine Stärkung von Intermediären wie Datenschutzbehörden oder Verbraucherschutzorganisationen notwendig, die unabhängige Kontrollen durchführen, Zertifikate für vertrauenswürdige KI vergeben und die Öffentlichkeit über Normverstöße in Kenntnis setzen.

… und Intermediäre stärken

Professionsethische Selbstverpflichtungen für KI-Entwickler_Innen gewährleisten

Zudem sind langfristig professionsethische Selbstverpflichtungen für KI-Entwickler_innen zu gewährleisten, welche sicherstellen, dass Interessen von Verbraucher_innen bereits im Herstellungsprozess berücksichtigt werden. Diese Gestaltungsvorschläge zielen entsprechend auf die Etablierung von Rahmenbedingungen für kollektive Selbstbestimmung ab, bei der sich Selbstbestimmung nicht nur durch individuelles Tun, sondern durch Prozesse der kollektiven Interessendelegation verwirklicht.

Vertraulichkeit sicherstellen mittels technischer und organisatorischer Maßnahmen

Schließlich muss vor dem Hintergrund der Globalisierung der Datenverarbeitung und der Verlagerung von Rechenleistungen auf ein Backend auch der Zugriffsschutz gewährleistet werden. Dies umfasst technische und organisatorische Maßnahmen zur Sicherstellung der Vertraulichkeit, der lokalen Speicherung personenbezogener Daten im Endgerät oder zumindest auf europäischen Servern, aber auch rechtliche Restriktionen im Umgang mit KI. Mit Blick auf die Sensitivität der verarbeiteten Daten müssen auch klare Regeln und Grenzen für die Nutzung der Daten durch Sicherheitsbehörden erarbeitet werden.

Die Vielfalt der hier angerissenen Empfehlungen zeigt den breiten Handlungsbedarf innerhalb des diskutierten Themenkomplexes. Letztlich bleibt nur zu wiederholen, dass KI-Technologien für verschiedene Einsatzzwecke geeignet sind, sodass sowohl Chancen als auch Risiken entstehen, die sich wiederum auf Aspekte der individuellen als auch der kollektiven Selbstbestimmungsfähigkeit beziehen können. In aller Grundsätzlichkeit bleibt schließlich das Ziel, eine sinnvolle Aushandlung von Wert- oder Zielkonflikten zu erreichen, sodass gemeinsame Chancen identifiziert und gefördert werden können.

forum <privatheit>
selbstbestimmtes_leben_in_der_digitalen_welt

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

PROJEKTPARTNER

Fraunhofer

Natur Technik Kultur Gesellschaft
UNIKASSEL VERSITÄT

provet
Projektgruppe verfassungsverträgliche Technikgestaltung

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

EBERHARD KARLS UNIVERSITÄT TÜBINGEN
INTERNATIONALES ZENTRUM FÜR ETHIK IN DEN WISSENSCHAFTEN

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

ULD
Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein