

STATEMENTS | 20.03.2023 *

Pressekonferenz anlässlich der Veröffentlichung der Stellungnahme „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“

- Es gilt das gesprochene Wort. -

Einführung – Alena Buyx

Sehr geehrte Damen und Herren,

herzlich willkommen! Ich danke Ihnen, auch im Namen meiner Kolleginnen und Kollegen im Deutschen Ethikrat, dafür, dass Sie heute gekommen sind und wir Ihnen unsere Stellungnahme „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“ vorstellen dürfen.

Spätestens seit ChatGPT im November ins öffentliche Bewusstsein drang, ist die breite Anwendung von Künstlicher Intelligenz in unserem Alltag und in der öffentlichen Debatte angekommen. Dabei leben wir bereits viel länger in einer Welt, in der Algorithmen des maschinellen Lernens viele Funktionen übernehmen. Sie können eingesetzt werden, um Krebs zu diagnostizieren oder mit Schüler*Innen Englischvokabeln zu lernen, aber auch um zu bestimmen, wer bestimmte Sozialleistungen bekommen soll und um unser Verhalten in den sozialen Medien zu beeinflussen.

Gerade weil KI uns inzwischen nahezu alle erreicht – ob uns das nun klar ist, oder nicht – und gerade weil sie immer besser wird, haben wir vom Deutschen Ethikrat uns dieses Themas angenommen. Aufbauend auf unsere Arbeit in der Stellungnahme zu Big Data und Gesundheit von 2017 und zu Robotik und Pflege von 2019 reagieren wir damit auch auf eine Bitte des damaligen Präsidenten des Deutschen Bundestages Schäuble im Oktober 2020.

Wir stellen in der Stellungnahme die technischen, methodischen und philosophischen Grundlagen von KI vor; zu letzteren wird mein Kollege Julian Nida-Rümelin, stellvertretender Vorsitzender und AG-Sprecher, noch ein wenig sagen. Wir entwickeln eine Beurteilungsmatrix, die die AG-Sprecherin Judith Simon einführen wird und wenden diese dann konkret auf vier Anwendungsfelder an: die Medizin, die schulische Bildung, die öffentliche Kommunikation und Meinungsbildung sowie die öffentliche Verwaltung. Zum Schluss führen wir eine Reihe von Querschnittsthemen ein, die für alle Anwendungsfelder relevant sind.

Unser Thema ist groß und tiefgreifend. Welche Auswirkungen hat es, wenn Tätigkeiten, die vorher Menschen vorbehalten waren, an Maschinen delegiert werden, und wie werden durch diese Übertragung die Möglichkeiten anderer Menschen beeinflusst, vor allem von jenen, über die entschieden wird?

Pressekontakt

Jonas Huggins
Telefon: +49/30/20370-524
E-Mail: presse@ethikrat.org

Geschäftsstelle

Jägerstraße 22/23
D-10117 Berlin
Telefon: +49/30/20370-242
Telefax: +49/30/20370-252
E-Mail: kontakt@ethikrat.org
Internet: www.ethikrat.org

* Korrigierte Fassung vom 27.03.2023

Das übergreifende Ergebnis unserer Analyse lautet: Der Einsatz von KI muss menschliche Entfaltung, Autorschaft und Handlungsmöglichkeiten erweitern und darf sie nicht vermindern. KI darf den Menschen nicht ersetzen. Das klingt nach einfachen ethischen Regeln – ist aber alles andere als trivial. Man muss nämlich für die vielen Bereiche der Gesellschaft, in denen KI Einfluss hat und zunehmend mehr Einfluss haben wird – vom Arztzimmer bis zur Amtsstube, und von der Schule bis in die sozialen Medien – sehr genau hinschauen: was sind die Grade der Ersetzung, wer ist denn jetzt wie und in welchem Maße betroffen? Wir haben uns die Mühe gemacht, das an vielen konkreten Beispielen auf 300 Seiten durchzugehen.

In der Anwendung besorgt uns beispielsweise, dass die algorithmisch vermittelten Informations- und Kommunikationsangebote digitaler Plattformen und Sozialer Medien Konsequenzen für das demokratische Legitimationsgefüge haben; mehr dazu später noch. Und bei den philosophisch-anthropologischen Grundlagen fragen wir: kann es eine starke, eigenständige maschinelle Intelligenz geben, die der menschlichen gleichkommt? Können (zukünftige) Maschinen im vollen Sinne intentional handeln? Die kurze Antwort auf diese Fragen lautet: Nein; die längere wird Julian Nida-Rümelin andeuten, der als nächstes zu Ihnen sprechen wird. Vielen Dank!

^ Alena Buyx

Zur philosophischen Grundlegung der Stellungnahme – Julian Nida-Rümelin

Ludwig Wittgenstein (Philosophische Grammatik §97):

„Die Intention von außen gesehen, das hat mit der Frage zu tun, ob eine Maschine denken könnte. Welches Phänomen immer man sähe, es könnte nie die Intention sein. Denn die muss ja das intendierte enthalten. Und jedes Phänomen wäre ein in sich Vollständiges, sich um nichts Bekümmernendes, das tot daläge, wenn man es für sich betrachtet.“

Der Ethikrat hat sich für diese Stellungnahme „Mensch und Maschine. Herausforderungen durch Künstliche Intelligenz“ intensiv mit den philosophischen Grundlagen des Verhältnisses Mensch - Maschine auseinandergesetzt, was sich vor allem im dritten Kapitel niederschlägt. Hier werden die Merkmale menschlicher Intelligenz und menschlicher Vernunft herausgearbeitet, die sich auf Maschinen nicht übertragen lassen, daher ist von „Künstlicher Intelligenz“ im Text durchgängig nur als Eigenname also mit großem „K“ geschrieben die Rede.

Deutlicher wird dies noch in der Analyse der Begriffe Handlung und Verantwortung Menschen interagieren mit Maschinen und ihre Handlungsoptionen können sich durch den Einsatz digitaler Tools und speziell Künstlicher Intelligenz erweitern, aber sie können ihre Verantwortung nicht an Maschinen delegieren oder mit Maschinen teilen. Auch wenn Maschinen in Gestalt Künstlicher Intelligenz hochentwickelt sind, können sie den Menschen ihre Verantwortung nicht abnehmen. Damit ergibt sich eine doppelte Frontstellung, gegen das Verständnis des Menschen als einer besonders hochentwickelten Maschine und gegen die anthropomorphe Interpretation Künstlicher Intelligenz. Die Vernunft des Menschen ist mit seiner Leiblichkeit, die Teil seiner Identität ist, unauflösbar verbunden. Damit übergebe ich an Judith Simon, vielen Dank!

^ Julian Nida-Rümelin

Mensch-Technik-Relationen – Judith Simon

Neben den soeben dargelegten Unterschieden zwischen Mensch und Maschine sind auch die Wechselverhältnisse zwischen beiden wichtig. Wir entwickeln Technologien zwar zu bestimmten Zwecken, doch diese Technologien wirken auch auf uns zurück. So können Technologien neue Handlungsmöglichkeiten eröffnen, sie aber – ggf. an anderer Stelle oder für andere Personen – auch beschränken.

Kern unserer Analyse ist die Betrachtung dieses Wechselspiels: zwischen dem Delegieren vormals menschlicher Tätigkeiten an Technologie einerseits – bis hin zu einer möglichen Ersetzung des Menschen durch Technologie – und den Rückwirkungen dieses Delegierens auf menschliche Handlungsmöglichkeiten andererseits.

Ziel der Delegation von Handlungssegmenten an Technologie im Allgemeinen und KI im speziellen sollte prinzipiell die Erweiterung menschlicher Handlungsfähigkeit und Autorschaft sein. Umgekehrt gilt es, die Verminderung menschlicher Handlungsfähigkeit und Autorschaft sowie eine Diffusion von Verantwortung zu verhindern.

Im Folgenden werden wir die zentralen Einsichten aus den vier Sektoren darlegen. Hier zeigt sich, dass für die ethische Bewertung von KI der Teufel oft im Detail steckt. Genauer: in den Details der Technik, der Einsatzkontexte sowie der institutionellen Rahmenbedingungen. Darüber hinaus gilt es die unterschiedlichen Beteiligten und Betroffenen in den Blick zu nehmen – die Erweiterung der Handlungsmöglichkeiten für die einen kann nämlich durchaus mit deren Verminderung für andere einhergehen.

^ Judith Simon

KI in der Medizin – Alena Buyx

KI-gestützte digitale Produkte kommen zunehmend im Gesundheitssystem zum Einsatz. Insbesondere in der medizinischen Forschung kann der Einsatz von KI vorteilhaft sein, so bei der Durchsichtung großer Datenbanken, um treffsichere Vorhersagen über individuelle Krankheiten, die Ausbreitung eines Virus oder zur Struktur komplexer Moleküle zu machen. In der medizinischen Versorgung werden KI-Instrumente zunehmend auch für die Diagnostik und Therapie entwickelt, beispielsweise bei Brust- und Prostatakrebs. Einer der wenigen medizinischen Handlungsbereiche, in denen KI-basierte Systeme die weitgehende oder sogar vollständige Ersetzung von ärztlichem bzw. anderem Gesundheitspersonal zumindest suggerieren, ist die Psychotherapie.

Um sicher zu stellen, dass menschliche Entfaltung und Handlungsmöglichkeiten erweitert und nicht vermindert werden, formulieren wir eine Reihe von Empfehlungen, die sich unter anderem auf die Qualitätssicherung bei der Entwicklung und Nutzung von KI-Produkten beziehen. Wir unterstreichen die Vermeidung ärztlicher Kompetenzverluste – das sogenannte Deskillung – und das Ziel, die Privatsphäre von Patientinnen und Patienten mit gleichzeitig intensiver Datennutzung in der medizinischen Forschung in Einklang zu bringen. KI-Anwendungen, die nachweislich herkömmlichen Behandlungsmethoden überlegen sind, sollten allen einschlägigen Patientengruppen zur Verfügung stehen. Es gilt, das Vertrauensverhältnis zwischen allen beteiligten

Personen zu schützen. Eine vollständige Ersetzung von Ärzt*Innen durch ein KI-System gefährdet das Patientenwohl und ist auch nicht dadurch zu rechtfertigen, dass schon heute in bestimmten Versorgungsbereichen ein akuter Personalmangel besteht.

^ Alena Buyx

KI in der schulischen Bildung – Julian Nida-Rümelin

Auf der Grundlage eines anspruchsvollen Bildungsbegriffs, der die Befähigung zur Lebensautorschaft in den Mittelpunkt stellt, werden die Einsatzmöglichkeiten von KI in der Bildungspraxis erörtert, dabei stehen anthropologische und ethische Fragen im Mittelpunkt. Wie verändert die maschinelle Ersetzung bestimmte Handlungssegmente den Unterricht, wie ändert sich dadurch das Selbstverständnis der Beteiligten und deren Kooperationsbereitschaft? Welcher Einsatz von KI fördert die Fähigkeit zu freiem und vernunftgeleitetem Handeln?

Digitalisierung ist kein Selbstzweck, die technologische Gestaltung des Bildungsprozesses sollte die Persönlichkeitsbildung, die Urteilskraft und die Entscheidungsstärke, sowie das erforderliche Orientierungswissen fördern. Auch ChatGPT zwingt uns zur Auseinandersetzung damit, was Bildung ist und soll, welches Wissen relevant ist und welche Fähigkeiten und Fertigkeiten die Lernenden benötigen.

Der Einsatz von KI kann das Eingehen auf individuelle Bedingungen und Neigungen der Lernenden fördern und zudem den Unterrichtenden wichtige Informationen zum Lerngeschehen bereitstellen. Die Datenerfassung und -bereitstellung muss dabei dem Lernprozess dienen, ohne zur Überwachung und zur Stigmatisierung von Lernenden missbraucht zu werden. Daher wird die vor allem in China verbreitete Methode der Classroom-Analytics kritisch gesehen. Eine Unterstützung der Lehrkräfte durch kontinuierliches Feedback kann die Unterrichtsergebnisse zwar erheblich verbessern, ein umfassendes Kontrollregime im Unterricht ist jedoch nicht wünschenswert.

Auch die Frage, wie sehr die Technik menschliche Interaktion im Unterricht ersetzt, ist relevant. KI-Tools können Lehrkräfte in bestimmten Bereichen, etwa Vokabellernen oder Arithmetik ersetzen, die personale Beziehung zwischen Unterrichtenden und Lernenden darf durch ihren Einsatz im Unterricht jedoch nicht beeinträchtigt werden.

Diese Punkte greifen wir in den Empfehlungen des Bildungsabschnitts ebenso auf wie die Förderung der Nutzungskompetenz der Lehrkräfte, ohne die der Einsatz von KI im Unterricht zu einer Verengung des pädagogischen Ansatzes und einem Deskillings der Unterrichtenden führen kann.

^ Julian Nida-Rümelin

KI in der öffentlichen Kommunikation – Julian Nida-Rümelin

Auch in der öffentlichen Kommunikation und Meinungsbildung verändert sich durch KI einiges – mit potenziellen Auswirkungen auf die Demokratie. Einerseits erleichtert digitale Kommunikation den Zugang zu Informationen; andererseits haben Polarisierung und Diskursverrohung besonders in den Sozialen Medien zugenommen, und die demokratischen Institutionen sind

verletzlicher geworden. Damit ist die politische Meinungsbildung im digitalen Zeitalter herausgefordert.

Das kommerzielle Interesse an einem möglichst langen Verbleib der einzelnen Nutzerin auf der Plattform, um ihr Werbung zu zeigen, führt zu personalisierten Angeboten, die den Eindruck objektiver Auswahl vermitteln, tatsächlich aber das vorausgegangene Nutzerverhalten widerspiegeln. Eine Verengung des angebotenen Informations- und Meinungsspektrums ist die Folge.

Wenn Inhalte auf Plattformen negativ auffallen, weil sie beispielsweise Falsch-aussagen enthalten oder Hass schüren, versuchen Firmen darauf mit „Content Moderation“ zu reagieren. Sie setzen dafür menschliches Personal ein, nutzen aber auch Softwaresysteme. Das soll sicherstellen, dass rechtlich unzulässige oder als problematisch eingestufte Inhalte gelöscht werden.

Damit entsteht ein Spannungsverhältnis zwischen Meinungsfreiheit und Zivildkultur der Kommunikation, das in den Händen privater Digital-Unternehmen nach eigenen Regeln aufgelöst wird. Overblocking kann die Folge sein. Die grundsätzlichere Frage aber ist, ob wir privaten kommerziellen Konzernen derart wichtige Entscheidungen überlassen wollen.

Digitale Kommunikation verändert die Informations- und Diskursqualität positiv wie negativ. Sie ermöglicht vielen Menschen Zugänge zu Daten und Prozessen, die im vor-digitalen Zeitalter nur wenigen vorbehalten waren. Auch der Meinungsaustausch ist nun in Formaten möglich, die Interessierte aus ganz unterschiedlichen Milieus, Disziplinen und Orten zusammenführen. Zugleich führen die Konkurrenz um Aufmerksamkeit und die kommerziellen Interessen der Plattformbetreiber zu Polarisierung und Diskursverrohung; populistische Tendenzen und die Verbreitung von Verschwörungstheorien werden begünstigt.

In welchem Ausmaß digitale Kommunikation zur Gefährdung der Demokratie beiträgt, ist zwar wissenschaftlich umstritten. Dennoch stellt sich die Frage, ob eine alternative digitale Kommunikationsinfrastruktur in öffentlicher Verantwortung nicht zum Schutz von Vielfalt, kommunikativem Ethos und Demokratie erforderlich ist.

^ Julian Nida-Rümelin

Öffentliche Verwaltung – Judith Simon

Der vierte Sektor, der in dieser Stellungnahme exemplarisch behandelt wurde, ist die öffentliche Verwaltung. Hier lässt sich in den vergangenen Jahren ein zunehmender Einsatz von Software zur Entscheidungsunterstützung in einer Vielzahl unterschiedlicher Bereiche beobachten, bspw. im Sozial- oder Polizeiwesen.

Häufig wird der Einsatz solcher Systeme einerseits mit einer Steigerung der Effizienz von Verwaltungsvorgängen begründet und andererseits mit einer möglichen Verbesserung der Qualität von – teils sehr tiefgreifenden – Entscheidungen.

Demgegenüber stehen zahlreiche ethische Fragen und Probleme. So ist nicht erwiesen, dass die Verwendung von Entscheidungssystemen zwangsläufig zu besseren Entscheidungen führt. Durch die verwendeten Daten werden zudem

oftmals existierende gesellschaftliche Ungleichheiten und Ungerechtigkeiten aufgegriffen und durch den Einbau in scheinbar neutrale Entscheidungssysteme zementiert.

Gerade in staatlichen Kontexten sowie in Bezug auf Entscheidungen, die eine hohe Tragweite haben oder besonders vulnerable Gruppen betreffen, müssen hier hohe und verbindliche Anforderungen an Genauigkeit, Diskriminierungsvermeidung und Nachvollziehbarkeit gestellt werden und die Einhaltung dieser Kriterien der externen Überprüfung zugänglich gemacht werden.

Darüber stellen sich zahlreiche weitere Fragen, bspw. wie die Autonomie der unterschiedlichen Beteiligten und Betroffenen geschützt werden kann oder wie sichergestellt werden kann, dass es zu keiner schleichenden Aushöhlung der sozialrechtlich gebotenen Identifizierung individueller Hilfebedarfe zugunsten einseitiger externer Interessen an Gefahrenminimierung oder Kostenersparnis kommt.

^ Judith Simon

Querschnittsthemen – Judith Simon

In unseren Analysen wurde deutlich, dass es Themen gibt, welche beim Einsatz von KI in den verschiedenen Sektoren immer wieder auftauchen - aber nicht immer auf gleiche Weise. Diese Querschnittsthemen stellen wir im letzten Teil unserer Stellungnahme vor.

Die ersten vier Querschnittsthemen beschäftigen sich mit der Frage, wie genau der Einsatz von KI zu einer Erweiterung oder Verminderung von Handlungsmöglichkeiten für unterschiedliche Akteure führt. So geht es darum, die Möglichkeiten von KI zur Generierung von Wissen, zur Verbesserung von Diagnosen und Prognosen zu nutzen. Gleichzeitig gilt es jedoch auch langfristige Verschiebung in den Blick zu nehmen. Diese betreffen unerwünschte individuelle oder kollektive Kompetenzverluste, welche durch gezielte Gegenmaßnahmen minimiert bzw. kompensiert werden müssen. Aber es geht auch um die grundlegende Frage, welche Rolle Prognosen für Entscheidung insbesondere über andere Menschen haben sollen - in Anbetracht des unüberbrückbaren Spaltes zwischen Wahrscheinlichkeitstheoretischer Prognose und individuellem Fall. Denn auch wenn eine Software vorhersagt, dass eine Person zu 99% straffällig wird, so können wir eben niemals wissen, ob die Person vor uns nicht genau das 1% ist.

Die Querschnittsthemen 5 und 6 wiederum beschäftigen sich mit der Nutzung von Daten und der Frage, wie wir einerseits die Privatsphäre und Autonomie von Nutzerinnen und Nutzern schützen können, aber andererseits auch eine gemeinwohlorientierte Datennutzung ermöglichen können. Hier erscheint eine stärker beziehungsorientierte Perspektive auf Datennutzung notwendig. Dabei ist auf asymmetrische Machtverhältnisse zu achten und Sorge zu tragen, dass die Erweiterung der Handlungsmöglichkeiten einiger nicht zulasten der Verminderung der Handlungsmöglichkeiten anderer, insbesondere benachteiligter Gruppen stattfindet.

Die Querschnittsthemen 7 und 8 beschäftigen sich mit der Frage, inwiefern digitale Technologien als kritische Infrastrukturen betrachtet werden müssen: wie können wir Resilienz herstellen und Abhängigkeiten und Missbrauch vermeiden? Hier betonen wir die Notwendigkeit von Interoperabilität, aber

auch des Auf- und Ausbaus alternativer Infrastrukturen. Zudem können hohe Standards für die Sicherheit und den Schutz der Privatsphäre dazu beitragen, spätere missbräuchliche Anwendungen einzuhegen bzw. möglichst zu verhindern.

Die letzten beiden Querschnittsthemen betreffen die Vermeidung von Bias und Diskriminierung einerseits, sowie die Beförderung von Transparenz und Nachvollziehbarkeit andererseits. Hier gilt: je größer die Eingriffstiefe und je unumgänglicher die Systeme, desto höher und verbindlicher müssen die Anforderungen an Diskriminierungsminimierung, Transparenz und Nachvollziehbarkeit sein.

^ Judith Simon