



STELLUNGNAHME

Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz

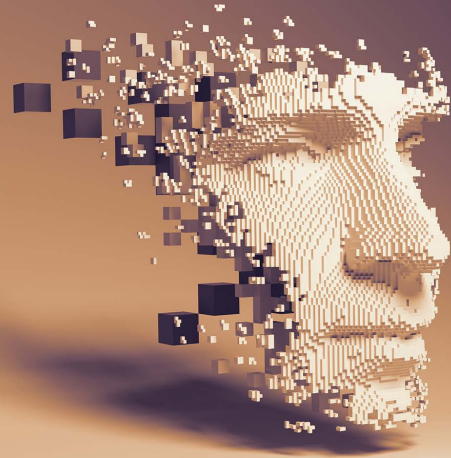


Illustration: pinkeyes/Shutterstock.com

In seiner Stellungnahme untersucht der Deutsche Ethikrat umfassend die Auswirkungen digitaler Technologien auf das menschliche Selbstverständnis und Miteinander. Die philosophisch grundierten Überlegungen und Empfehlungen geben Orientierung zu den rasanten Innovationen im Bereich der Künstlichen Intelligenz (KI).

Veröffentlicht:
20. März 2023

ChatGPT, bitte schreib ein Gedicht über die Stellungnahme des Deutschen Ethikrates zur KI!



Menschliche Entfaltung muss KI erweitern, nicht etwa vermindern oder gar ersetzen. Der Ethikrat warnt vor diesen Fehlern, die uns als Menschheit ins Verderben hetzen.

KI ist inzwischen allgegenwärtig, sie ist in vielen Bereichen zu finden, von der Medizin bis zur Verwaltung, in der Schule und auch im Internetzinden.

Doch Ethik ist hier von hoher Bedeutung, denn es geht nicht nur um Technologie, sondern um den Menschen und seine Entwicklung, sowie um eine faire Gesellschaftspolitie.

KI kann helfen und unterstützen, doch muss der Mensch stets im Mittelpunkt bleiben, um eine Zukunft zu gestalten, in der Mensch und KI in Harmonie gedeihen.

Was ist der Anlass für die Stellungnahme?

Nicht erst seit OpenAIs Chatbot ChatGPT auf Zuruf teils beindruckende, teils absurde Textschöpfungen zu jedem vorgeschlagenen Thema anbietet, hat der Begriff der Künstlichen Intelligenz in der öffentlichen Debatte zunehmend an Aufmerksamkeit gewonnen. Die dichterischen Fähigkeiten von ChatGPT sind sicherlich noch ausbaufähig, aber in vielen gesellschaftlichen Bereichen werden mit KI ernsthafte Hoffnungen wie auch Befürchtungen verknüpft. Vor diesem Hintergrund hat der Deutsche Ethikrat eine Stellungnahme zu den ethischen Fragen des Verhältnisses von Mensch und Maschine erarbeitet.

Was sind die normativen Überlegungen des Deutschen Ethikrates zum Thema KI?

Der Deutsche Ethikrat geht von einem normativ grundlegenden Unterschied zwischen Mensch und Maschine aus. Softwaresysteme verfügen weder über theoretische noch über praktische Vernunft. Sie handeln oder entscheiden nicht selbst und können keine Verantwortung übernehmen. Sie sind kein personales Gegenüber, auch dann nicht, wenn sie Anteilnahme, Kooperationsbereitschaft oder Einsichtsfähigkeit simulieren.

Menschliche Vernunft ist immer zugleich eingebunden in die konkrete soziale Mit- und Umwelt. Nur so ist zu erklären, dass sie handlungswirksam wird. Vernünftig handelt der einzelne Mensch als Teil einer sozialen Mitwelt und einer kulturellen Umgebung.

„Auch wenn Maschinen in Gestalt Künstlicher Intelligenz hochentwickelt sind, können sie den Menschen ihre Verantwortung nicht abnehmen.“

Julian Nida-Rümelin

Menschen entwickeln digitale Technik und nutzen sie als Mittel zu menschlichen Zwecken. Jedoch wirken diese Technologien zurück auf menschliche Handlungsmöglichkeiten. Auch wenn Maschinen also nicht selbst handeln, so verändern sie die Handlungsfähigkeit von Menschen tiefgreifend und können Handlungsmöglichkeiten erheblich erweitern oder vermindern.



MENSCH \neq MASCHINE
MASCHINE \neq MENSCH

„Wenn menschliche Tätigkeiten an Maschinen delegiert werden, kann dies für verschiedene Personengruppen, Akteure und Betroffene ganz unterschiedliche Auswirkungen haben.“

Judith Simon

Für den Ethikrat lautet die zentrale Schlüsselfrage für die ethische Beurteilung: Werden menschliche Autorschaft und die Bedingungen für verantwortliches Handeln durch den Einsatz von KI erweitert oder vermindert? Ziel der Delegation menschlicher Tätigkeiten an Maschinen sollte prinzipiell die Erweiterung menschlicher Handlungsfähigkeit sein. Ihre Verminderung sowie eine Diffusion oder Evasion von Verantwortung gilt es hingegen zu verhindern.

„KI darf den Menschen nicht ersetzen.“

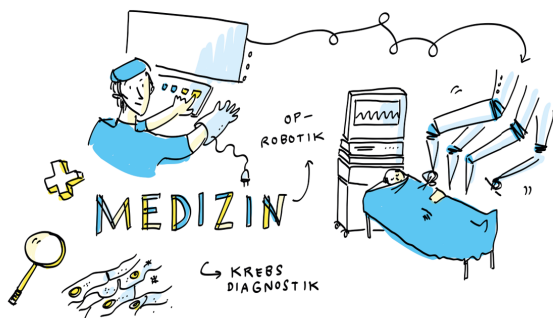
Alena Buyx

Um über Wert und Nutzen der Delegation vormals menschlichen Handelns an Maschinen ethisch zu befinden, bedarf es immer eines kontextspezifischen Blicks, der die Perspektiven unterschiedlicher Beteiligter und Betroffener ebenso berücksichtigt wie die langfristigen Auswirkungen solcher Übertragungen. Die Herausforderungen stecken also wie so oft im Detail, genauer: in den Details der Technik, der Einsatzkontexte sowie der institutionellen und sozio-technischen Umgebung.

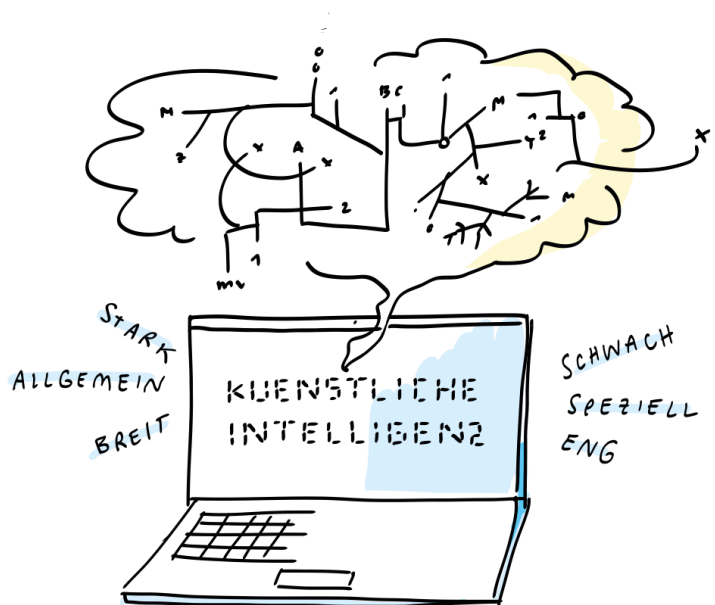
Welche Anwendungsbereiche von KI betrachtet der Deutsche Ethikrat in der Stellungnahme?

Um einen kontextspezifischen Blick zu ermöglichen, hat sich der Deutsche Ethikrat in seiner Stellungnahme exemplarisch mit Anwendungen in vier Bereichen beschäftigt: der Medizin, der schulischen Bildung, der öffentlichen Kommunikation und Meinungsbildung sowie der öffentlichen Verwaltung.

Medizin



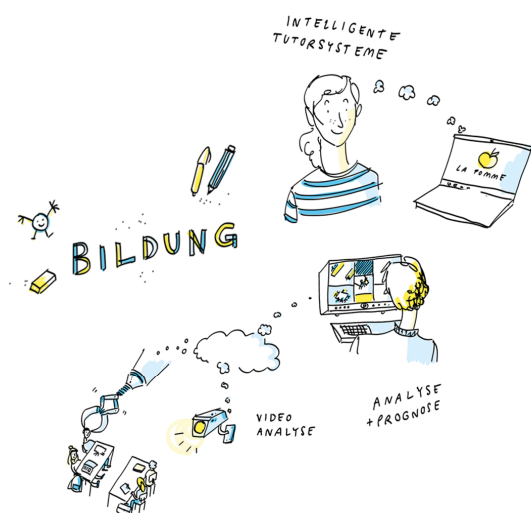
Im besonders sensiblen Bereich der Medizin kommen digitale Produkte mit KI-Komponenten in rasch wachsendem Umfang zum Einsatz. KI-Systeme können große Datenmengen und viele relevante Parameter auswerten und in Gestalt präziser Mustererkennung großes diagnostisches und therapeutisches Potenzial entwickeln. Das Spektrum der Delegation an KI-Systeme reicht hier von der punktuellen Unterstützung ärztlichen Handelns über dessen ständige Begleitung bis hin zur weitgehenden oder gar vollständigen Ersetzung ärztlicher Fachkräfte durch ein KI-System.



„Wir unterstreichen die Vermeidung ärztlicher Kompetenzverluste – das sogenannte *Deskilling* – 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 Ärztinnen und Ärzten 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

Bildung



In der schulischen Bildung kommen KI-gestützte Technologien zum Einsatz, um personalisiertes Lernen und Lehren zu unterstützen oder Informationen zum Lerngeschehen bereitzustellen. Nach Einschätzung des Ethikrates darf dies aber nicht auf Kosten von sozialer Interaktion und auch nicht um den Preis von Überwachung oder Stigmatisierung von Lernenden geschehen.

„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. [...] KI-Tools können Lehrkräfte in bestimmten Bereichen, etwa Vokabellernen oder Arithmetik ersetzen, die personale Beziehung zwischen Unterrichtenden und Lernenden darf durch den Einsatz von KI im Unterricht jedoch nicht beeinträchtigt werden.“

Julian Nida-Rümelin

Öffentliche Kommunikation und Meinungsbildung

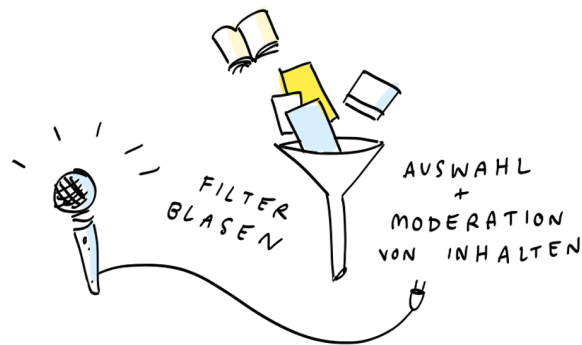
Ein großer Teil der öffentlichen Kommunikation und Meinungsbildung findet inzwischen online statt. Dabei kommt zunehmend auch KI zum Einsatz. Suchmaschinen, Newsfeeds und Empfehlungssoftware steuern beabsichtigt oder unbeabsichtigt menschliches Verhalten. Sie beeinflussen unseren Zugang zu Information und Kommunikation und somit zur Realität.

„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 Meinungs austausch 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.“

Julian Nida-Rümelin

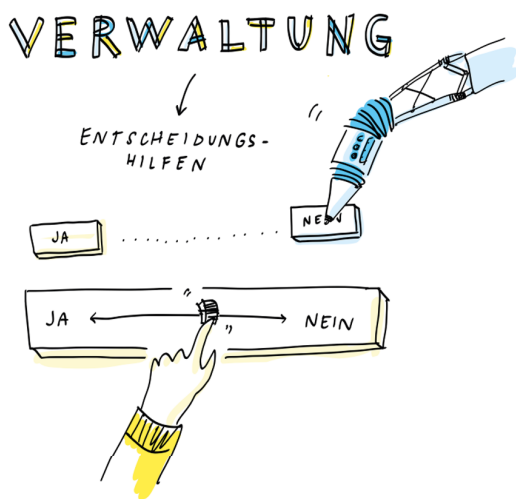


Der Ethikrat empfiehlt, die Regeln für Onlineplattformen und private Social-Media-Angebote weiterzuentwickeln und diesen im europäischen Rahmen eine digitale Kommunikationsinfrastruktur in öffentlich-rechtlicher Verantwortung zur Seite zu stellen, deren Betrieb sich nicht am Unternehmensinteresse eines möglichst langen Verweilens von Menschen auf der Plattform oder an anderen kommerziellen Interessen orientiert.



KAPITEL 7

Öffentliche Verwaltung

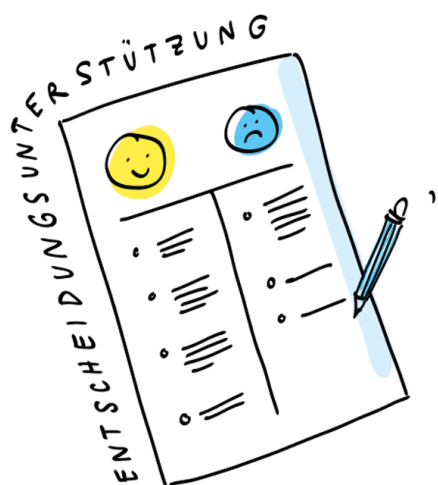


Der Ethikrat rät daher zu Ansätzen, die vor Diskriminierung schützen, Transparenz erhöhen, die Rechte der Betroffenen stärken und dem blinden Befolgen maschineller Empfehlungen vorbeugen.

„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“

Judith Simon

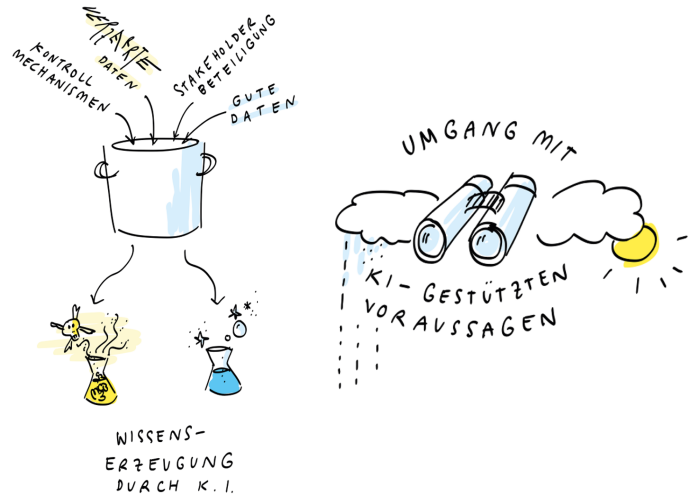
In der öffentlichen Verwaltung kommt zunehmend KI zum Einsatz, um Entscheidungen zu unterstützen, beispielsweise bei der Bewertung der Arbeitsmarktchancen Jobsuchender, der Prüfung und Vergabe von Sozialleistungen, oder auch der Vorhersage von Straftaten. Häufig wird der Einsatz solcher Software mit Effizienzsteigerungen und Qualitätsgewinnen begründet. Es ist jedoch nicht erwiesen, dass ihre Verwendung zwangsläufig zu besseren Entscheidungen führt. Zudem können insbesondere datenbasierte Systeme existierende gesellschaftliche Ungleichheiten übernehmen und Diskriminierung fortführen.



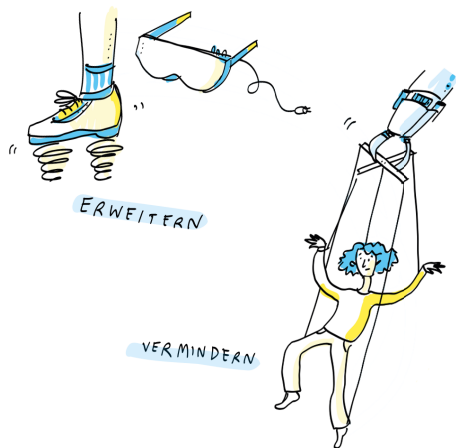
KAPITEL 8

Welche ethischen Querschnittsthemen spricht der Ethikrat in der Stellungnahme an?

In seiner Stellungnahme identifiziert der Deutsche Ethikrat zehn Querschnittsthemen und -herausforderungen, die sich durch alle Anwendungsbereiche von KI ziehen und formuliert zu jedem eine Empfehlung. Die Querschnittsthemen sollen als Anregung für eine breitere Debatte zur zukünftigen Politik- und Technikgestaltung dienen.



1. Erweiterung und Verminderung von Handlungsmöglichkeiten



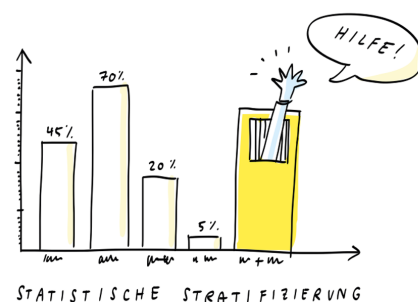
Eine vollständige Ersetzung von Menschen durch KI verbietet sich überall dort, wo die konkrete zwischenmenschliche Begegnung eine notwendige Voraussetzung für das Erreichen von Handlungszielen darstellt. Wie stark KI menschliche Handlungsmöglichkeiten erweitert oder vermindert, hängt jedoch erheblich von den konkreten Einsatzszenarien und ihrer Wirkung auf verschiedene Personengruppen ab. Daher muss der Einsatz von KI jeweils kontextspezifisch geplant und müssen seine Folgen für unterschiedliche Betroffene zeitnah evaluiert werden.

2. Wissenserzeugung durch KI und Umgang mit KI-gestützten Voraussagen

KI-gestützte Vorhersagen werden immer genauer. Trotzdem bleiben immer Restunsicherheiten und darf man Korrelationen und Datenmuster

nicht mit Erklärungen und Begründungen von Ereignisursachen gleichsetzen. Um Verantwortungsdiffusion zu vermeiden, sollte KI nur eingesetzt werden, um Entscheidungen zu unterstützen, nicht jedoch, um sie zu ersetzen. Den Betroffenen algorithmisch gestützter Entscheidungen muss Zugang zu den jeweiligen Entscheidungsgrundlagen gewährt werden.

3. Gefährdung des Individuums durch statistische Stratifizierung



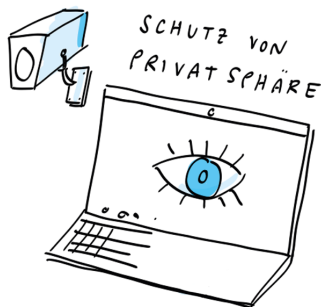
KI stellt auf Grundlage großer Datenmengen Zusammenhänge zwischen Merkmalen fest, und ordnet dabei auch Menschen unterschiedlichen Gruppen zu. Hier können Probleme entstehen, wenn für eine Person im konkreten Fall bestimmte Merkmale einer Gruppe und statistisch getroffene Diagnosen oder Prognosen nicht zutreffen. Das Individuum entzieht sich einer vollständigen statistischen Annäherung. Einzelfallbeurteilungen bleiben daher grundsätzlich wichtig. KI-basierte Urteile und Voraussagen können als Hilfsmittel verwendet werden, nicht jedoch für definitive Entscheidungen.

4. Auswirkungen von KI auf menschliche Kompetenzen und Fertigkeiten



Der Einsatz von KI-Anwendungen beeinflusst menschliche Handlungsfähigkeiten. Wenn KI gute Ergebnisse erzielt, können Menschen sich angewöhnen, sich weniger auf das eigene als auf das technisch erzeugte Urteil zu verlassen – mit der Folge, dass eigene Fähigkeiten nicht eingeübt werden und so mit der Zeit verloren gehen. Einem möglichen Verlust menschlicher Kompetenzen durch den Einsatz von KI muss anhand einer sinnvollen Gestaltung des Zusammenspiels von Mensch und Technik entgegengewirkt werden.

5. Schutz von Privatsphäre und Autonomie versus Gefahren durch Überwachung und Chilling-Effekte



Viele KI-Anwendungen basieren auf der Sammlung und Auswertung großer Mengen an personenbezogenen Daten. Das bringt Risiken für die Privatsphäre mit sich. Schon die Sorge vor Überwachung und Nachteilen durch Datensammlung kann Handlungsmöglichkeiten einschränken. Um solchen Gefahren und Chilling-Effekten entgegenzuwirken, muss übermäßiges Tracking von Onlineverhalten und der Handel mit personenbeziehenden Daten rechtlich und technisch angemessen und effektiv eingedämmt werden.

6. Datensouveränität und gemeinwohlorientierte Datennutzung



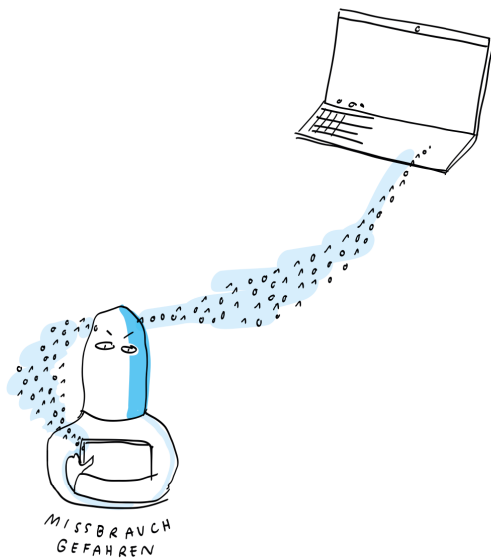
Um die großen Chancen von KI zu nutzen, müssen Lösungen entwickelt werden, wie Daten sinnvoll für verschiedene wichtige Zwecke genutzt werden können, ohne zugleich den Schutz der Privatsphäre unzulässig zu beeinträchtigen. Gemeinwohlorientierte Datennutzung muss unter Berücksichtigung der jeweiligen spezifischen Herausforderungen und Nutzenpotenziale vereinfacht werden. Dies beinhaltet unter anderem einen gesellschaftlichen Bewusstseinswandel von einer individualistisch geprägten zu einer stärker gemeinwohlbasierten Perspektive.

7. Kritische Infrastrukturen, Abhängigkeiten und Resilienz



Kritische Infrastrukturen wie beispielsweise Stromnetze werden zunehmend digital und mithilfe von KI überwacht und gesteuert. Gleichzeitig werden digitale Technologien wie soziale Netzwerke selbst zu wichtigen Infrastrukturen. Wo solche Infrastrukturen verwundbar sind oder abhängig von kommerziellen Akteuren, kann menschliche Autonomie gefährdet werden. Daher braucht der Schutz und die Resilienz kritischer digitaler Infrastrukturen mehr Aufmerksamkeit und müssen einseitige Abhängigkeiten vermieden werden. Alternative Infrastrukturen wie etwa unabhängige öffentliche Plattformen sollten auf- und ausgebaut werden.

8. Pfadabhängigkeiten, Zweitverwertung und Missbrauchsgefahren



Entscheidungen, die zu Beginn einer Entwicklung getroffen werden, können noch lange nachwirken und einmal gesetzte Normen und Standards sind teils schwer wieder aufzuheben. Wurde eine Technologie erst einmal mit womöglich hohen Investitionen etabliert, gehen damit zudem oft Bestrebungen einher, sie möglichst umfänglich zu nutzen – auch über das ursprüngliche Anwendungsfeld hinaus. Daher müssen mögliche Langzeitfolgen wie Pfadabhängigkeiten und Missbrauchspotenziale von Anfang an mitbedacht werden. Negativen Entwicklungen sollte mit hohen Schutzstandards vorgebeugt werden.

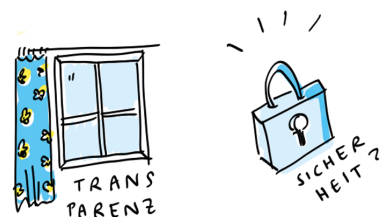
9. Bias und Diskriminierung



Zur Vermeidung von Diskriminierungen müssen KI-Systeme angemessen beaufsichtigt und kontrolliert werden. Je größer die Eingriffstiefe und je unumgänglicher die

Systeme, desto höher die Anforderungen. Schon bei der Technologieentwicklung gilt es, Diskriminierungen zu minimieren und Fairness, Transparenz und Nachvollziehbarkeit zu gewährleisten. Technische und regulatorische Maßnahmen stoßen hier an Grenzen, unter anderem weil unterschiedliche Fairnessziele technisch nicht gleichzeitig erfüllt werden können. Es müssen daher zugleich ethische und politische Entscheidungen getroffen werden, welche Kriterien für Gerechtigkeit in welchem Kontext gelten sollen. Diese Entscheidungen dürfen nicht den Personen, die Software entwickeln, und anderen direkt Beteiligten überlassen werden. Stattdessen bedarf es der Entwicklung geeigneter Verfahren und Institutionen, um diese Kriterien kontextspezifisch und demokratisch, gegebenenfalls immer wieder neu auszuhandeln.

10. Transparenz und Nachvollziehbarkeit – Kontrolle und Verantwortung



KI-Systeme sind aus unterschiedlichen Gründen oft wenig transparent und nachvollziehbar. Es sollten ausgewogene kontextspezifische Standards für Transparenz, Erklärbarkeit und Nachvollziehbarkeit entwickelt werden, deren Umsetzung durch verbindliche technische und organisatorische Vorgaben gewährleistet wird. Dabei muss Anforderungen an Sicherheit und Schutz vor Missbrauch, Datenschutz sowie dem Schutz von intellektuellem Eigentum und Geschäftsgeheimnissen in angemessener Weise Rechnung getragen werden.

