



Brüssel, den 8.4.2019
COM(2019) 168 final

**MITTEILUNG DER KOMMISSION AN DAS EUROPÄISCHE PARLAMENT, DEN
RAT, DEN EUROPÄISCHEN WIRTSCHAFTS- UND SOZIALAUSSCHUSS UND
DEN AUSSCHUSS DER REGIONEN**

Schaffung von Vertrauen in eine auf den Menschen ausgerichtete künstliche Intelligenz

1. EINLEITUNG – DIE EUROPÄISCHE KI-STRATEGIE

Die künstliche Intelligenz (KI) hat das Potenzial, unsere Welt zum Besseren zu verändern: Sie kann die Gesundheitsversorgung verbessern, den Energieverbrauch senken, Autos sicherer machen und ermöglicht Landwirten eine effizientere Nutzung von Wasser und Naturressourcen. KI kann eingesetzt werden, um Umwelt- und Klimaveränderungen vorherzusagen, das Management finanzieller Risiken zu verbessern und die Werkzeuge zu schaffen, die wir brauchen, um genau auf unsere Bedürfnisse zugeschnittene Produkte mit weniger Abfällen herzustellen. Sie kann auch helfen, Betrug und Bedrohungen der Cybersicherheit zu erkennen, und versetzt die Strafverfolgungsbehörden in die Lage, Kriminalität wirksamer zu bekämpfen.

KI kann für die Gesellschaft und Wirtschaft als Ganzes von großem Nutzen sein. Es handelt sich um eine strategische Technologie, die sich derzeit in der ganzen Welt rasch weiterentwickelt und verbreitet. Allerdings bringt die KI auch neue Herausforderungen für die Zukunft der Arbeit mit sich und wirft rechtliche und ethische Fragen auf.

Um diese Herausforderungen zu meistern und die von der KI eröffneten Chancen bestmöglich zu nutzen, hat die Kommission im April 2018 eine europäische Strategie¹ veröffentlicht. Diese Strategie rückt die Menschen in den Mittelpunkt der KI-Entwicklung – es geht um eine auf den Menschen ausgerichtete, **menschenzentrierte KI**. Dabei wird ein dreiteiliger Ansatz verfolgt: Förderung der technologischen und industriellen Leistungsfähigkeit der EU sowie der weiteren KI-Verbreitung in der gesamten Wirtschaft, Vorbereitung auf die mit KI verbundenen sozioökonomischen Veränderungen und Gewährleistung eines geeigneten ethischen und rechtlichen Rahmens.

Zur Umsetzung der KI-Strategie hat **die Kommission gemeinsam mit den Mitgliedstaaten einen koordinierten Plan für die KI² ausgearbeitet**, den sie im Dezember 2018 vorgestellt hat, um Synergien zu schaffen, Daten – den Rohstoff für zahlreiche KI-Anwendungen – zusammenzuführen und verstärkt gemeinsame Investitionen zu fördern. Ziel ist es, die grenzüberschreitende Zusammenarbeit zu fördern und alle Akteure zu mobilisieren, damit die öffentlichen und privaten Investitionen in den nächsten zehn Jahren **auf mindestens 20 Mrd. EUR** jährlich steigen³. Die Kommission hat ihre im Rahmenprogramm Horizont 2020 vorgesehenen Investitionsmittel für KI verdoppelt und plant zudem, 1 Mrd. EUR jährlich aus den Programmen „Horizont Europa“ und „Digitales Europa“ zu investieren, um insbesondere gemeinsame Datenräume auf Gebieten wie Gesundheit, Verkehr und Fertigung, große Versuchseinrichtungen wie intelligente Krankenhäuser und Infrastrukturen für automatisierte Fahrzeuge sowie eine strategische Forschungsagenda zu unterstützen.

Zur Umsetzung einer solchen gemeinsamen strategischen Forschungs-, Innovations- und Einführungsagenda hat die Kommission ihren **Dialog mit allen relevanten Akteuren** aus Industrie, Forschungsinstituten und Behörden intensiviert. Das neue Programm „Digitales Europa“ wird ebenfalls maßgeblich dazu beitragen, dass gerade kleinen und mittleren Unternehmen in allen Mitgliedstaaten – mithilfe digitaler Innovationszentren –

¹ COM(2018) 237.

² COM(2018) 795.

³ Um dieses Ziel zu erreichen, hat die Kommission für den nächsten Programmplanungszeitraum 2021–2027 vorgeschlagen, dass die Union jährlich mindestens 1 Mrd. EUR an Fördermitteln aus den Programmen „Horizont Europa“ und „Digitales Europa“ für Investitionen in KI bereitstellt.

verbesserte Test- und Versuchseinrichtungen, Datenräume und Schulungsprogramme zur Verfügung gestellt werden können.

Das ethische Herangehen Europas an die KI wird – aufbauend auf dem guten Ruf seiner sicheren und hochwertigen Produkte – einerseits das Vertrauen der Bürger in die digitale Entwicklung stärken und zielt andererseits darauf ab, einen Wettbewerbsvorteil für europäische KI-Unternehmen zu erringen. Mit dieser Mitteilung soll eine umfassende Pilotphase unter breitestmöglicher Beteiligung der Interessenträger eingeleitet werden, um die praktische Umsetzung der Ethik-Leitlinien bei der Entwicklung und Nutzung der KI zu erproben.

2. SCHAFFUNG VON VERTRAUEN IN EINE MENSCHENZENTRIERTE KI

Die europäische KI-Strategie und der koordinierte Plan machen deutlich, dass **Vertrauen eine Grundvoraussetzung für die Verfolgung eines auf den Menschen ausgerichteten Ansatzes für die KI** ist: KI ist kein Selbstzweck, sondern ein Instrument, das den Menschen dienen muss und letztlich das Wohlergehen der Menschen steigern soll. Dabei kommt es entscheidend auf die **Vertrauenswürdigkeit der KI** an. Die unseren Gesellschaften zugrunde liegenden Wertvorstellungen müssen bei der Entwicklung der KI vollständig zum Tragen kommen.

Die **Werte**, auf denen die Union beruht, sind die **Achtung der Menschenwürde, Freiheit, Demokratie, Gleichheit, Rechtsstaatlichkeit und die Wahrung der Menschenrechte** einschließlich der Rechte der Personen, die Minderheiten angehören⁴. Diese Werte sind den Gesellschaften aller Mitgliedstaaten gemeinsam, die sich durch Pluralismus, Nichtdiskriminierung, Toleranz, Gerechtigkeit, Solidarität und Gleichheit auszeichnen. Überdies sind in der **EU-Grundrechtecharta** alle persönlichen, bürgerlichen, politischen, wirtschaftlichen und sozialen Rechte der Menschen innerhalb der EU in einem einzigen Text zusammengefasst.

Die EU verfügt über einen **starken Rechtsrahmen**, der weltweite Standards für eine menschenzentrierte KI setzen wird. Die Datenschutz-Grundverordnung gewährleistet ein hohes Maß an Schutz personenbezogener Daten und schreibt verbindliche Vorkehrungen vor, die den „Datenschutz durch Technik“ und „datenschutzfreundliche Voreinstellungen“⁵ einschließen. Die Verordnung über den freien Verkehr nicht personenbezogener Daten beseitigt Hindernisse, die der freien Weitergabe nicht personenbezogener Daten entgegenstehen, und bewirkt, dass alle Datenkategorien überall in Europa verarbeitet werden können. Der erst kürzlich erlassene Rechtsakt zur Cybersicherheit wird dazu beitragen, das Vertrauen in die Online-Welt zu stärken. Die vorgeschlagene e-Datenschutz-Verordnung⁶ dient ebenfalls diesem Ziel.

⁴ Außerdem ist die EU Vertragspartei des Übereinkommens der Vereinten Nationen über die Rechte von Menschen mit Behinderungen.

⁵ Verordnung (EU) 2016/679. Die Datenschutz-Grundverordnung (DSGVO) gewährleistet den freien Verkehr personenbezogener Daten in der Union. Sie enthält Bestimmungen über Entscheidungsprozesse, die ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhen. So haben die betroffenen Personen das Recht, über das Bestehen automatischer Entscheidungsprozesse informiert zu werden und aussagekräftige Auskünfte über die Logik der automatisierten Entscheidungsfindung sowie über die Bedeutung und die vorgesehenen Folgen der Verarbeitung für sie zu erhalten. Außerdem haben sie in solchen Fällen Anspruch auf eine menschliche Überprüfung, bei der sie ihren Standpunkt darlegen und der Entscheidung widersprechen können.

⁶ COM(2017) 10.

Dennoch bringt die KI neue Herausforderungen mit sich, weil sie es Maschinen ermöglicht, eigenständig „zu lernen“ und ohne menschliches Zutun Entscheidungen zu treffen und umzusetzen. Es wird nicht mehr lange dauern, bis diese Funktionsweise bei vielen Arten von Waren und Dienstleistungen – von Smartphones bis hin zu automatisierten Fahrzeugen, Robotern und Online-Anwendungen – zum Standard wird. Entscheidungen, die von Algorithmen getroffen werden, können jedoch aus Daten resultieren, die unvollständig und daher nicht verlässlich sind; sie können durch Cyber-Angreifer manipuliert werden oder einfach fehlerhaft sein. Ein unbedachter Einsatz der sich entwickelnden Technik würde daher problematische Ergebnisse hervorbringen und dazu führen, dass die Bürger solche Technik nur zögerlich akzeptieren und nutzen.

Stattdessen sollte KI-Technik so entwickelt werden, dass sie die Menschen in ihren Mittelpunkt rückt und daher das Vertrauen der Öffentlichkeit verdient. Das bedeutet, dass KI-Anwendungen nicht nur rechtmäßig sein müssen, sondern auch ethische Grundsätze einhalten und sicherstellen sollten, dass bei ihrer Anwendung kein unbeabsichtigter Schaden entstehen kann. Die Vielfalt in Bezug auf Geschlecht, Rasse oder ethnische Herkunft, Religion oder Weltanschauung, Behinderung und Alter sollte in jeder Phase der KI-Entwicklung gewährleistet sein. KI-Anwendungen sollten die mündigen Bürger stärken und deren Grundrechte wahren. Ihr Einsatz sollte dem Ziel dienen, die Fähigkeiten der Menschen zu verbessern, und nicht, sie zu ersetzen. Und auch Menschen mit Behinderungen sollten sich die Möglichkeiten der KI voll zunutze machen können.

Deshalb brauchen wir **Ethik-Leitlinien**, die auf dem bestehenden Rechtsrahmen aufbauen und von KI-Entwicklern, -Anbietern und -Nutzern im Binnenmarkt gleichermaßen befolgt werden, damit in allen Mitgliedstaaten gleiche Wettbewerbsbedingungen herrschen. Aus diesem Grund hat die Kommission eine **hochrangige Expertengruppe für KI**⁷ eingesetzt, in der ein breites Spektrum von Interessenträgern vertreten ist. Sie hat die Gruppe mit der Aufstellung von Ethik-Leitlinien für die KI und mit der Ausarbeitung von Empfehlungen für eine umfassendere KI-Politik beauftragt. Gleichzeitig wurde die **Europäische KI-Allianz**⁸ als eine offene Multi-Stakeholder-Plattform mit mehr als 2700 Mitgliedern eingerichtet, die einen auf breiterer Grundlage fußenden Beitrag zur Arbeit der hochrangigen Expertengruppe für KI leisten soll.

Die hochrangige Expertengruppe für KI veröffentlichte im Dezember 2018 einen ersten Entwurf der Ethik-Leitlinien. Nach einer **Konsultation der Interessenträger**⁹ und nach **Zusammenkünften mit Vertretern der Mitgliedstaaten**¹⁰ legte die KI-Expertengruppe der Kommission im März 2019 eine überarbeitete Fassung der Leitlinien vor. In ihren bisherigen Beiträgen begrüßten die Interessenträger insgesamt den praktischen Charakter der Leitlinien und die konkreten Hinweise, die sie Entwicklern, Anbietern und Nutzern von KI in Bezug darauf geben, wie Vertrauenswürdigkeit gewährleistet werden kann.

⁷ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

⁸ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

⁹ Zu dieser Konsultation gingen Stellungnahmen von 511 Organisationen, Verbänden, Unternehmen, Forschungsinstituten, Einzelpersonen und anderen ein. Eine Zusammenfassung der Beiträge ist abrufbar unter: https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf.

¹⁰ Die Arbeit der Expertengruppe wurde von den Mitgliedstaaten positiv aufgenommen. In den Schlussfolgerungen des Rates vom 18. Februar 2019 wurden unter anderem die bevorstehende Veröffentlichung der Ethik-Leitlinien begrüßt und die Bemühungen der Kommission um die Etablierung eines Ethik-Ansatzes der EU auf weltweiter Ebene unterstützt: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf>.

2.1 Ethik-Leitlinien für eine vertrauenswürdige KI – ausgearbeitet von der hochrangigen Expertengruppe für KI

Die von der hochrangigen Expertengruppe für KI ausgearbeiteten Ethik-Leitlinien, auf die sich diese Mitteilung bezieht¹¹, beruhen insbesondere auf den Vorarbeiten der Europäischen Gruppe für Ethik der Naturwissenschaften und der neuen Technologien und der Agentur der Europäischen Union für Grundrechte.

In den Leitlinien wird davon ausgegangen, dass zur Verwirklichung einer „vertrauenswürdigen KI“ drei Komponenten notwendig sind: 1) sie sollte rechtmäßig sein, 2) sie sollte ethischen Grundsätzen entsprechen und 3) sie sollte robust sein.

Auf der Grundlage dieser drei Komponenten und der in Abschnitt 2 dargelegten europäischen Werte werden in den Leitlinien sieben Kernanforderungen genannt, die von KI-Anwendungen zu erfüllen sind, damit sie als vertrauenswürdige gelten können. Darüber hinaus enthalten die Leitlinien eine Bewertungsliste, mit deren Hilfe überprüft werden kann, ob diese Anforderungen erfüllt sind.

Die sieben Kernanforderungen sind:

- Vorrang menschlichen Handelns und menschlicher Aufsicht
- Technische Robustheit und Sicherheit
- Privatsphäre und Datenqualitätsmanagement
- Transparenz
- Vielfalt, Nichtdiskriminierung und Fairness
- Gesellschaftliches und ökologisches Wohlergehen
- Rechenschaftspflicht

Diese Anforderungen sollen zwar generell für alle KI-Systeme in verschiedenen Umfeldern und Branchen gelten, bei ihrer konkreten und verhältnismäßigen Umsetzung sind aber der jeweilige spezifische Anwendungskontext und die möglichen Folgen zu berücksichtigen. So wäre beispielsweise eine KI-Anwendung, die ein unpassendes Buch vorschlägt, weit weniger gefährlich als etwa eine Anwendung, die eine falsche Krebsdiagnose stellt, und könnte daher einer weniger strengen Aufsicht unterliegen.

Die von der hochrangigen Expertengruppe für KI ausgearbeiteten Leitlinien haben einen unverbindlichen Charakter und begründen daher als solche keine neuen rechtlichen Verpflichtungen. Mehrere dieser Kernanforderungen finden sich bereits in vielen bestehenden (nutzungs- oder bereichsspezifischen) Bestimmungen des Unionsrechts, z. B. in Sicherheitsvorschriften und in Vorschriften zum Schutz personenbezogener Daten und zum Schutz der Privatsphäre oder zum Umweltschutz.

Die Kommission begrüßt die Arbeit der hochrangigen Expertengruppe für KI und betrachtet ihre Arbeitsergebnisse als wertvollen Beitrag zu ihrer Politik.

2.2 Kernanforderungen an eine vertrauenswürdige KI

Die Kommission unterstützt die folgenden Kernanforderungen an eine vertrauenswürdige KI, die auf europäischen Werten beruhen. Sie ruft die Interessenträger auf, diese Anforderungen umzusetzen und die Bewertungsliste, die diese praktisch anwendbar

¹¹ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

macht, selbst zu testen, um so das richtige Umfeld für die erfolgreiche Entwicklung und Nutzung künstlicher Intelligenz zu schaffen. Die Kommission begrüßt etwaige Rückmeldungen der Interessenträger, um besser einschätzen zu können, ob die in den Leitlinien enthaltene Bewertungsliste weiter angepasst werden muss.

I. Vorrang menschlichen Handelns und menschlicher Aufsicht

KI-Systeme sollten die einzelnen Menschen dabei unterstützen, im Einklang mit ihren eigenen Zielen bessere, fundiertere Entscheidungen zu treffen. Sie sollten einer florierenden und gerechten Gesellschaft dienen, indem sie das menschliche Handeln und die Wahrung der Grundrechte unterstützen, keinesfalls aber sollten sie die Autonomie der Menschen verringern, beschränken oder fehlleiten. Die Richtschnur für das Funktionieren von KI-Systemen sollte das allgemeine **Wohl des Nutzers** sein.

Die menschliche Aufsicht hilft, dafür zu sorgen, dass ein KI-System die menschliche Autonomie nicht untergräbt oder sich sonst nachteilig auswirkt. In Abhängigkeit vom konkreten KI-gestützten System und dessen Anwendungsgebiet sollte es geeignete Stufen von **Kontrollmaßnahmen** geben, die auch Aspekte der Anpassbarkeit, Genauigkeit und Erklärbarkeit KI-gestützter Systeme beinhalten¹². Die **Aufsicht** kann durch Lenkungs- und Kontrollmechanismen wie die Gewährleistung der interaktiven Einbindung eines Menschen („*Human-in-the-Loop*“), der Überprüfung und Kontrolle durch einen Menschen („*Human-on-the-Loop*“) oder der Gesamtsteuerung durch einen Menschen („*Human-in-Command*“) erreicht werden¹³. Auf jeden Fall muss dabei aber sichergestellt werden, dass die Behörden stets in der Lage sind, die Aufsicht im Einklang mit ihrem jeweiligen Auftrag auszuüben. Für ein ansonsten gleiches System muss gelten: je weniger Aufsicht ein Mensch über ein KI-System ausüben kann, desto ausführlicher muss es zuvor getestet werden und desto strenger muss die Lenkung und Kontrolle sein.

II. Technische Robustheit und Sicherheit

Eine vertrauenswürdige KI setzt Algorithmen voraus, die sicher, verlässlich und robust genug sind, um Fehler oder Unstimmigkeiten in allen Phasen des Lebenszyklus des KI-Systems zu bewältigen und mit fehlerhaften Ergebnissen angemessen umzugehen. KI-Systeme müssen **verlässlich** und hinreichend sicher sein, sodass sie sowohl offenen Angriffen als auch subtileren Manipulationsversuchen an ihren Daten oder eigenen Algorithmen **standhalten**, und sie müssen für den Fall von Problemen eine **Rückfallstrategie** haben. Ihre Entscheidungen müssen **genau** sein oder zumindest den Grad der Genauigkeit richtig angeben, und ihre Ergebnisse sollten **reproduzierbar** sein.

Überdies sollten KI-Systeme konzeptuell integrierte Schutz- und Sicherheitsvorkehrungen aufweisen, damit sie in jeder Phase **nachprüfbar sicher** sind, wobei es auf die körperliche

¹² Die Datenschutz-Grundverordnung gibt Personen das Recht, 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 (Artikel 22 der DSGVO).

¹³ „*Human-in-the-Loop*“ (HITL) bedeutet, dass in jeden Entscheidungszyklus des Systems ein Mensch eingebunden sein muss, was in vielen Fällen weder möglich noch wünschenswert wäre. „*Human-on-the-Loop*“ (HOTL) bedeutet die Möglichkeit des Menschen, in den Entwurfszyklus des Systems einzugreifen und den Systembetrieb zu überwachen. „*Human-in-Command*“ (HIC) bedeutet die Möglichkeit, den Gesamtbetrieb des KI-Systems zu beaufsichtigen (einschließlich seiner weiteren wirtschaftlichen, gesellschaftlichen, rechtlichen und ethischen Auswirkungen), sowie die Fähigkeit zu entscheiden, wann und wie das System in einer bestimmten Situation eingesetzt werden soll. Dies beinhaltet auch die Entscheidungsmöglichkeit, ein KI-System in einer bestimmten Situation nicht einzusetzen, beim Einsatz des Systems ein bestimmtes Maß an menschlichem Ermessen zuzulassen oder eine vom System getroffene Entscheidung außer Kraft zu setzen.

und geistige Sicherheit aller Beteiligten ankommt. Dies schließt das Minimieren und möglichst das Rückgängigmachen unbeabsichtigter Folgen oder Fehler im Systembetrieb ein. Es sollten Verfahren zur Klärung und Bewertung potenzieller Risiken im Zusammenhang mit dem Einsatz von KI-Systemen in verschiedenen Anwendungsbereichen eingerichtet werden.

III. Privatsphäre und Datenqualitätsmanagement

Der Schutz der Privatsphäre und der **Datenschutz** müssen in **allen Phasen** des Lebenszyklus eines KI-Systems gewährleistet sein. Aus digitalen Aufzeichnungen über das menschliche Verhalten können KI-Systeme nicht nur auf persönliche Vorlieben sowie Alter und Geschlecht einzelner Menschen, sondern auch auf die sexuelle Ausrichtung und religiöse oder politische Ansichten schließen. Damit die Menschen Vertrauen in die Datenverarbeitung haben können, muss sichergestellt sein, dass sie die volle Kontrolle über ihre eigenen Daten behalten und dass die sie betreffenden Daten nicht dazu verwendet werden, sie zu schädigen oder zu diskriminieren.

Neben dem Schutz der Privatsphäre und der personenbezogenen Daten müssen hochwertige KI-Systeme weitere strenge Anforderungen erfüllen. Die Qualität der verwendeten Datensätze ist für die Leistungsfähigkeit von KI-Systemen von entscheidender Bedeutung. Bei der Erfassung der Daten kann es sozial bedingte Verzerrungen geben, oder die Daten können Ungenauigkeiten, Fehler und andere Mängel aufweisen. Solche Probleme müssen behoben werden, bevor ein KI-System mit einem bestimmten Datensatz ausgebildet wird. Darüber hinaus muss die **Integrität** der Daten gewährleistet sein. Die verwendeten Prozesse und Datensätze müssen in allen Schritten wie Planung, Ausbildung, Erprobung und Einsatz getestet und dokumentiert werden. Dies sollte auch für KI-Systeme gelten, die nicht intern entwickelt, sondern von außerhalb erworben werden. Schließlich muss auch der **Zugang** zu den Daten angemessen geregelt und kontrolliert werden.

IV. Transparenz

Die **Rückverfolgbarkeit** der KI-Systeme muss sichergestellt werden. Dazu müssen sowohl die vom System getroffenen Entscheidungen selbst als auch der gesamte Prozess, der zu der Entscheidung geführt hat, protokolliert und dokumentiert werden (einschließlich Beschreibung der Datenerfassung und Datenbenennung und des verwendeten Algorithmus). In diesem Zusammenhang sollte – soweit dies möglich ist – eine für die beteiligten Personen verständliche **Erklärung** des algorithmischen Entscheidungsprozesses bereitgestellt werden. Die laufenden Forschungsarbeiten zur Entwicklung von Erklärungsmechanismen sollten fortgesetzt werden. Darüber hinaus sollten Erläuterungen dazu vorliegen, inwieweit ein KI-System den Entscheidungsprozess der Organisation beeinflusst und formt, aber auch über Entscheidungen zum Systementwurf und die Gründe für dessen Einführung (sodass nicht nur die Daten- und Systemtransparenz, sondern auch die Transparenz des Geschäftsmodells gewährleistet wird).

Schließlich ist es wichtig, den verschiedenen Beteiligten die Fähigkeiten und Beschränkungen des KI-Systems in einer Weise **mitzuteilen**, die für die jeweilige Nutzung angemessen ist. Außerdem sollten KI-Systeme als solche erkennbar sein, damit die Nutzer stets wissen, dass sie es mit einem KI-System zu tun haben, und wer dafür verantwortlich ist.

V. Vielfalt, Nichtdiskriminierung und Fairness

Die von KI-Systemen (sowohl zur Ausbildung als auch im Betrieb) verwendeten

Datensätze können unbeabsichtigte historische Verzerrungen aufweisen, unvollständig sein und auf schlechte Lenkungs- und Kontrollmodelle zurückgehen. Die Fortschreibung solcher Verzerrungen könnte (in)direkte Diskriminierungen zur Folge haben. Schaden kann aber auch aus einer beabsichtigten Ausnutzung von Vorurteilen (der Verbraucher) oder durch unlauteren Wettbewerb entstehen. Auch die Art und Weise, wie KI-Systeme entwickelt werden (z. B. wie der Programmcode eines Algorithmus geschrieben wird), kann durch gewisse Einflüsse beeinträchtigt sein. Auf solche Bedenken sollte vom Beginn der Systementwicklung an eingegangen werden.

Die Bildung **vielfältig zusammengesetzter Entwurfsteams** und die Einrichtung von Mechanismen der **Beteiligung** (vor allem auch der Bürger) an der KI-Entwicklung können ebenfalls dazu beitragen, solche Bedenken auszuräumen. Ebenso ist es ratsam, all jene Beteiligten zu konsultieren, die von dem System während seines gesamten Lebenszyklus direkt oder indirekt betroffen sein können. KI-Systeme sollten dem gesamten Spektrum menschlicher Fähigkeiten, Fertigkeiten und Anforderungen Rechnung tragen und die Barrierefreiheit durch ein universelles Entwurfskonzept gewährleisten, um einen gleichberechtigten Zugang für Menschen mit Behinderungen zu erreichen.

VI. Gesellschaftliches und ökologisches Wohlergehen

Damit die KI vertrauenswürdig sein kann, sollten ihre Auswirkungen auf die **Umwelt und auf andere fühlende Wesen** berücksichtigt werden. Im Idealfall sollten alle Menschen, auch die künftigen Generationen, in biologischer Vielfalt und einer bewohnbaren Umwelt leben können. Die Nachhaltigkeit und die **ökologische Verantwortlichkeit** von KI-Systemen sollten daher gefördert werden. Dasselbe gilt für KI-Lösungen, die sich mit Belangen von weltweiter Bedeutung befassen, wie z. B. mit den Zielen der Vereinten Nationen für nachhaltige Entwicklung.

Darüber hinaus sollten die Auswirkungen von KI-Systemen nicht nur aus individueller, sondern auch aus **gesamtgesellschaftlicher** Sicht betrachtet werden. Der Einsatz von KI-Systemen sollte insbesondere im Zusammenhang mit dem demokratischen Prozess – also in der öffentlichen Meinungsbildung, im politischen Entscheidungsprozess oder in Bezug auf Wahlen – sorgfältig geprüft werden. Außerdem sollten die **sozialen Auswirkungen** der KI berücksichtigt werden. So wie KI-Systeme zur Verbesserung sozialer Kompetenzen eingesetzt werden können, können sie auch zu ihrer Verschlechterung beitragen.

VII. Rechenschaftspflicht

Es sollten Mechanismen geschaffen werden, die die Verantwortlichkeit und Rechenschaftspflicht für KI-Systeme und deren Ergebnisse vor und nach ihrer Umsetzung gewährleisten. Die **Nachprüfbarkeit** von KI-Systemen ist in dieser Hinsicht von entscheidender Bedeutung, denn die Bewertung von KI-Systemen durch interne und externe Prüfer und das Vorliegen solcher Bewertungsberichte trägt beträchtlich zur Vertrauenswürdigkeit der Technik bei. Die externe Nachprüfbarkeit sollte insbesondere bei Anwendungen sichergestellt sein, die sich auf Grundrechte auswirken, sowie bei sicherheitskritischen Anwendungen.

Mögliche negative Auswirkungen von KI-Systemen sollten ermittelt, bewertet, dokumentiert und minimiert werden. Folgenabschätzungen können diesen Prozess erleichtern. Solche Folgenabschätzungen sollten aber in einem angemessenen Verhältnis zu der Höhe der Risiken stehen, die von den KI-Systemen ausgehen. **Kompromisse** zwischen den verschiedenen Anforderungen, die oft unvermeidbar sind, sollten rational und methodisch angegangen und berücksichtigt werden. Sollte es schließlich doch zu

ungerechten und nachteiligen Auswirkungen kommen, sollten leicht zugängliche Mechanismen vorgesehen sein, die einen **angemessenen Rechtsschutz** gewährleisten.

2.3 Nächste Schritte: Pilotphase unter breitestmöglicher Beteiligung der Interessenträger

Der Konsens über diese Kernanforderungen an KI-Systeme ist ein erster wichtiger Meilenstein auf dem Weg zu Leitlinien für eine ethisch vertretbare KI. Als nächsten Schritt wird die Kommission dafür sorgen, dass diese Leitlinien in der Praxis getestet und umgesetzt werden können.

Zu diesem Zweck wird sie nun eine gezielte Pilotphase einleiten, um strukturierte Rückmeldungen von den Interessenträgern einzuholen. Dies wird vor allem anhand der Bewertungsliste geschehen, die die hochrangige Expertengruppe für jede der Kernanforderungen erstellt hat.

Diese Arbeiten werden in zwei Teile gegliedert sein: i) eine Pilotphase für die Leitlinien unter Einbeziehung von Interessenträgern, die KI entwickeln oder einsetzen, einschließlich öffentlicher Verwaltungen, und ii) die fortlaufende Konsultation der Interessenträger sowie die Sensibilisierung in den Mitgliedstaaten und in verschiedenen Gruppen von Interessenträgern, einschließlich Industrie und Dienstleistungssektor:

- i) Ab Juni 2019 werden alle Interessenträger und Einzelpersonen aufgerufen sein, die Bewertungsliste zu testen und Rückmeldungen darüber zu geben, wie sie verbessert werden kann. Darüber hinaus wird die hochrangige Expertengruppe für KI zusammen mit den Interessenträgern aus dem privaten und dem öffentlichen Sektor eine eingehende Überprüfung beginnen, um detailliertere Rückmeldungen darüber zu erhalten, wie die Leitlinien in einer Vielzahl von Anwendungsbereichen umgesetzt werden können. Alle Rückmeldungen zur Praxistauglichkeit und Umsetzbarkeit der Leitlinien werden bis Ende 2019 bewertet.
- ii) Parallel dazu wird die Kommission weitere Informationsmaßnahmen ergreifen und Vertretern der hochrangigen Expertengruppe für KI die Möglichkeit geben, die Leitlinien den einschlägigen Interessenträgern in den Mitgliedstaaten, auch in der Industrie und im Dienstleistungssektor, vorzustellen, damit diese eine zusätzliche Gelegenheit erhalten, sich zu den KI-Leitlinien zu äußern und einen Beitrag dazu zu leisten.

Die Kommission wird auch die Arbeit der Expertengruppe für Ethik beim vernetzten und automatisierten Fahren¹⁴ berücksichtigen und bei der Umsetzung der Kernanforderungen mit den von der EU finanzierten Forschungsprojekten zum Thema KI sowie mit einschlägigen öffentlich-privaten Partnerschaften zusammenarbeiten¹⁵. Beispielsweise wird die Kommission – in Abstimmung mit den Mitgliedstaaten – den Aufbau einer gemeinsamen medizinischen Bilddatenbank unterstützen, die ursprünglich für die häufigsten Arten von Krebserkrankungen bestimmt war, um Algorithmen zu entwickeln, die anhand von Symptomen mit hoher Präzision Diagnosen stellen können. Ebenso kann dank der Zusammenarbeit zwischen Kommission und Mitgliedstaaten eine wachsende Zahl grenzüberschreitender Korridore für die Erprobung vernetzter und automatisierter Fahrzeuge eingerichtet werden. Die Leitlinien sollten bei diesen Projekten angewendet und getestet werden. Die Ergebnisse werden dann in den Bewertungsprozess einfließen.

¹⁴ Siehe die Mitteilung der Kommission zur automatisierten Mobilität, COM(2018) 283.

¹⁵ Im Rahmen des Europäischen Verteidigungsfonds wird die Kommission ebenfalls besondere Ethik-Leitlinien für die Bewertung von Projektvorschlägen im Bereich der KI für Verteidigungszwecke ausarbeiten.

Die Pilotphase und die Konsultation der Interessenträger werden vom Beitrag der Europäischen KI-Allianz und vom Projekt AI4EU, der KI-Abruf-Plattform, profitieren. Das im Januar 2019 ins Leben gerufene Projekt AI4EU¹⁶ wird Algorithmen, Werkzeuge, Datensätze und Dienste zusammenführen, um Organisationen, insbesondere kleine und mittlere Unternehmen, bei der Umsetzung von KI-Lösungen zu unterstützen. Die Europäische KI-Allianz wird zusammen mit dem Projekt AI4EU weiterhin das KI-Ökosystem in ganz Europa mobilisieren, auch im Hinblick auf die Erprobung der Ethik-Leitlinien für die KI und die Förderung einer auf den Menschen ausgerichteten („menschenzentrierten“) KI.

Aufbauend auf der Bewertung der Rückmeldungen aus der Pilotphase **wird die hochrangige Expertengruppe für KI Anfang 2020 die Leitlinien überprüfen und aktualisieren**. Auf der Grundlage der Überprüfung und der gewonnenen Erfahrungen **wird die Kommission das Ergebnis bewerten und nächste Schritte vorschlagen**.

Bei einer ethisch vertretbaren KI gibt es nur Gewinner. Für die Achtung der Grundwerte und Grundrechte zu sorgen, ist nicht nur für sich genommen wichtig, sondern erleichtert auch die Akzeptanz in der Öffentlichkeit und erhöht den Wettbewerbsvorteil der europäischen KI-Unternehmen, indem eine auf den Menschen ausgerichtete vertrauenswürdige KI mit ethisch vertretbaren und sicheren Produkten als Marke etabliert wird. Dies baut allgemein auf dem gefestigten guten Ruf europäischer Unternehmen bei der Bereitstellung sicherer und hochwertiger Produkte auf. Die Pilotphase wird dazu beitragen, dass die KI-Produkte diese Erwartungen erfüllen.

2.4 Hin zu internationalen KI-Ethik-Leitlinien

Die internationale Debatte über die KI-Ethik hat sich intensiviert, nachdem Japan im Rahmen seiner G7-Präsidentschaft das Thema im Jahr 2016 ganz oben auf die Tagesordnung gesetzt hatte. Angesichts der gegenseitigen internationalen Abhängigkeiten bei der Entwicklung der KI bezüglich der Weitergabe von Daten, der Entwicklung von Algorithmen und der Forschungsinvestitionen **wird die Kommission ihre Bemühungen fortsetzen, den Ansatz der Union auch weltweit zur Geltung zu bringen und einen Konsens über eine auf den Menschen ausgerichtete KI zu erzielen**¹⁷.

Dank der Arbeit der hochrangigen Expertengruppe für KI sowie vor allem der Liste der Anforderungen und der Einbindung der Interessenträger, verfügt die Kommission über zusätzliche wertvolle Beiträge als Grundlage für die internationalen Gespräche. Die Europäische Union kann bei der Ausarbeitung internationaler KI-Leitlinien und, wenn möglich, eines entsprechenden Bewertungsmechanismus eine Führungsrolle übernehmen.

Die Kommission wird daher

– **die Zusammenarbeit mit gleich gesinnten Partnern verstärken:**

- Untersuchung, inwieweit Konvergenz mit den Ethik-Leitlinienentwürfen von Drittländern (z. B. Japan, Kanada, Singapur) erreicht werden kann, und – aufbauend auf dieser Gruppe gleich gesinnter Länder – Vorbereitung einer breiteren Diskussion, unterstützt durch

¹⁶ <https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019>

¹⁷ Die Hohe Vertreterin der Union für Außen- und Sicherheitspolitik wird mit Unterstützung der Kommission die Konsultationen im Rahmen der Vereinten Nationen, im *Global Tech Panel* (globaler Technikbeirat) und in anderen multilateralen Foren fortsetzen und insbesondere Vorschläge zur Bewältigung der betreffenden komplexen Sicherheitsfragen koordinieren.

Maßnahmen zur Durchführung des Partnerschaftsinstruments für die Zusammenarbeit mit Drittstaaten¹⁸,

- Untersuchung, wie Unternehmen aus Drittländern und internationale Organisationen durch Erprobung und Validierung zur „Pilotphase“ der Leitlinien beitragen können;

– **weiterhin eine aktive Rolle bei internationalen Diskussionen und Initiativen spielen:**

- Beitrag zu multilateralen Foren wie der G7 und der G20,
- Dialoge mit Drittländern und Ausrichtung bilateraler und multilateraler Treffen, um einen Konsens über eine auf den Menschen ausgerichtete KI zu erzielen,
- Beitrag zu einschlägigen Normungstätigkeiten in internationalen Normungsorganisationen zur Förderung dieser Vision und
- Stärkung der Erfassung und Verbreitung von Erkenntnissen über staatliche Maßnahmen, die gemeinsam mit einschlägigen internationalen Organisationen durchgeführt werden.

3. SCHLUSSFOLGERUNGEN

Die EU beruht auf einer Reihe von Grundwerten, auf deren Grundlage sie einen starken und ausgewogenen Rechtsrahmen geschaffen hat. Aufbauend auf diesem bestehenden Rechtsrahmen sind wegen der Neuheit der KI und der besonderen Herausforderungen, die mit dieser Technologie verbunden sind, Ethik-Leitlinien für ihre Entwicklung und Nutzung erforderlich. Nur wenn KI unter Achtung der weitgehend gemeinsamen ethischen Werte entwickelt und genutzt wird, kann sie auch als vertrauenswürdig angesehen werden.

Im Hinblick auf dieses Ziel begrüßt die Kommission die Beiträge der hochrangigen Expertengruppe für KI. Ausgehend von den Kernanforderungen, die eine als vertrauenswürdig anzusehende KI erfüllen muss, wird die Kommission nun eine gezielte Pilotphase einleiten, um dafür zu sorgen, dass die sich daraus ergebenden KI-Ethik-Leitlinien für die Entwicklung und Nutzung der KI in der Praxis angewandt werden können. Die Kommission wird ferner darauf hinarbeiten, unter anderem mit allen Beteiligten und unseren internationalen Partnern einen breiten gesellschaftlichen Konsens über eine auf den Menschen ausgerichtete KI zu erzielen.

Die ethische Dimension der KI ist kein Luxus oder ergänzender Zusatz, sondern muss fester Bestandteil der KI-Entwicklung sein. Durch das Bemühen um eine auf den Menschen ausgerichtete KI, die auf Vertrauen beruht, gewährleisten wir, dass unsere gesellschaftlichen Grundwerte gewahrt bleiben und Europa und seine Industrie als Vorreiter auf dem Gebiet der innovativen KI eine unverwechselbare Marke schaffen können, die überall auf der Welt Vertrauen genießt.

¹⁸ Verordnung (EU) Nr. 234/2014 des Europäischen Parlaments und des Rates vom 11. März 2014 zur Schaffung eines Partnerschaftsinstruments für die Zusammenarbeit mit Drittstaaten (ABl. L 77 vom 15.3.2014, S. 77). So wird beispielsweise das geplante Projekt für ein „internationales Bündnis für ein auf den Menschen ausgerichtetes Herangehen an die künstliche Intelligenz“ gemeinsame Initiativen mit gleich gesinnten Partnern zur Förderung von Ethik-Leitlinien und zur Festlegung gemeinsamer Grundsätze und operativer Schlussfolgerungen erleichtern. Im Rahmen des Projekts werden die EU und gleich gesinnte Länder operative Schlussfolgerungen aus den von der Expertengruppe vorgeschlagenen Ethik-Leitlinien für die KI erörtern können, um zu einem gemeinsamen Ansatz zu gelangen. Darüber hinaus wird im Rahmen des Projekts die weltweite Einführung der KI-Technologie beobachtet. Schließlich werden als Teil des Projekts Maßnahmen im Bereich der Öffentlichkeitsarbeit bei internationalen Veranstaltungen organisiert, z. B. im Rahmen der G7, der G20 und der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung.

Um die ethisch vertretbare Entwicklung der KI in Europa in einem breiteren Kontext zu gewährleisten, verfolgt die Kommission einen umfassenden Ansatz, der insbesondere folgende Maßnahmen umfasst, die bis zum dritten Quartal 2019 umgesetzt werden sollen:

- Sie wird mithilfe von „Horizont 2020“ mehrere **Netze von KI-Spitzenforschungszentren** einrichten. Sie wird bis zu vier Netze auswählen, deren Schwerpunkt auf großen wissenschaftlichen oder technischen Herausforderungen liegt, wie z. B. der Erklärbarkeit von KI-Systemen und der fortgeschrittenen Mensch-Maschine-Interaktion, die für eine vertrauenswürdige KI von entscheidender Bedeutung sind.
- Sie wird **Netze digitaler Innovationszentren**¹⁹ einrichten, die sich auf KI in der Fertigung und auf Big Data konzentrieren.
- Sie wird vorbereitende Gespräche mit den Mitgliedstaaten und Interessenträgern führen, um **ein Modell für eine gemeinsame Datennutzung und die bestmögliche Verwendung gemeinsamer Datenräume** zu entwickeln und umzusetzen, wobei der Schwerpunkt insbesondere auf den Bereichen Verkehr, Gesundheitswesen und industrielle Fertigung liegen wird²⁰.

Darüber hinaus arbeitet die Kommission derzeit an einem Bericht über die mit KI verbundenen Herausforderungen im Zusammenhang mit dem Sicherheits- und Haftungsrahmen und einem Leitfaden zur Umsetzung der Produkthaftungsrichtlinie²¹. Gleichzeitig wird im Rahmen des gemeinsamen Unternehmens für europäisches Hochleistungsrechnen (EuroHPC)²² die nächste Generation von Supercomputern entwickelt, da Rechenkapazitäten für die Verarbeitung von Daten und für die Ausbildung von KI-Anwendungen eine wesentliche Rolle spielen und Europa die gesamte digitale Wertschöpfungskette selbst meistern muss. Die laufende Partnerschaft mit den Mitgliedstaaten und der Industrie im Bereich der Mikroelektronikkomponenten und -systeme (ECSEL)²³ sowie die Initiative für europäische Prozessoren²⁴ werden zur Entwicklung von Niedrigenergieprozessortechnik für vertrauenswürdige und sichere Hochleistungsrechnen und modernste Spitzencomputer beitragen.

Ebenso wie die Arbeit an den Ethik-Leitlinien für die KI beruhen all diese Initiativen auf einer **engen Zusammenarbeit aller beteiligten Akteure**, einschließlich der Mitgliedstaaten, der Industrie, der gesellschaftlichen Kräfte und der Bürger. Insgesamt zeigt das europäische Konzept für die künstliche Intelligenz, wie die wirtschaftliche Wettbewerbsfähigkeit und das Vertrauen der Gesellschaft auf denselben Grundwerten beruhen und sich gegenseitig verstärken müssen.

¹⁹ <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>

²⁰ Die erforderlichen Mittel werden aus dem Programm „Horizont 2020“ (vorgesehen sind knapp 1,5 Mrd. EUR für KI im Zeitraum 2018–2020), dessen Nachfolgeprogramm „Horizont Europa“, dem digitalen Teil der Fazilität „Connecting Europe“ und insbesondere dem künftigen Programm „Digitales Europa“ bereitgestellt werden. Die Projekte werden sich auch auf Mittel aus dem Privatsektor und aus den Programmen der Mitgliedstaaten stützen.

²¹ Siehe die Mitteilung der Kommission zur künstlichen Intelligenz für Europa, COM(2018) 237.

²² <https://eurohpc-ju.europa.eu>

²³ www.ecsel.eu

²⁴ www.european-processor-initiative.eu