



Empfehlung des Rates zu künstlicher Intelligenz

Inoffizielle Übersetzung

OECD-Rechtsinstrumente



Das vorliegende Dokument wird unter der Verantwortung des Generalsekretärs der OECD veröffentlicht. Es gibt ein OECD-Rechtsinstrument wieder und kann zusätzliches Material enthalten. Die im zusätzlichen Material zum Ausdruck gebrachten Meinungen und Argumente spiegeln nicht zwangsläufig die offizielle Einstellung der Mitgliedsländer der OECD wider.

Dieses Dokument sowie die darin enthaltenen Daten und Karten berühren weder den völkerrechtlichen Status von Territorien noch die Souveränität über Territorien, den Verlauf internationaler Grenzen und Grenzlinien sowie den Namen von Territorien, Städten oder Gebieten.

Die offiziellen und aktuellen Texte der OECD-Rechtsinstrumente sowie andere damit verbundene Informationen sind im Kompendium der OECD-Rechtsinstrumente unter <http://legalinstruments.oecd.org> verfügbar.

Bitte zitieren Sie dieses Dokument wie folgt:

OECD, *Empfehlung des Rates zu künstlicher Intelligenz*, OECD/LEGAL/0449

Reihe: OECD-Rechtsinstrumente

Übersetzung durch den Deutschen Übersetzungsdienst der OECD.

Bildnachweis: © OECD

© OECD 2024

Dieses Dokument wird kostenlos zur Verfügung gestellt. Es darf kostenlos reproduziert und verteilt werden, ohne dass weitere Genehmigungen erforderlich sind, solange es nicht in irgendeiner Weise verändert wird. Es darf nicht verkauft werden.

Dies ist keine amtliche Übersetzung. Obwohl die größtmöglichen Anstrengungen unternommen wurden, um die Übereinstimmung mit den Originaltexten zu gewährleisten, sind der englische und der französische Text die einzigen amtlichen Fassungen. Sie sind auf der OECD-Website <https://legalinstruments.oecd.org> verfügbar.

Hintergrundinformationen

Die Empfehlung zu künstlicher Intelligenz (KI) (im Folgenden als die „Empfehlung“ bezeichnet) – der erste zwischenstaatliche Standard auf dem Gebiet der KI – wurde vom Rat der OECD auf Minister-Ebene auf seiner Tagung am 22. Mai 2019 auf Vorschlag des Ausschusses für Digitalpolitik (Digital Policy Committee – DPC, vormals Ausschuss für digitale Wirtschaft [Committee on Digital Economy Policy – CDEP]) angenommen. Die Empfehlung soll den innovativen und vertrauensvollen Umgang mit KI fördern, indem sie die verantwortungsvolle Steuerung vertrauenswürdiger KI unter Beachtung der Menschenrechte und Werte der Demokratie unterstützt. Aus der Empfehlung leiten sich die Grundsätze der G20 zu KI ab, die von den Staats- und Regierungschef*innen der G20 auf ihrem Gipfeltreffen im Juni 2019 in Osaka begrüßt wurden.

Am 8. November 2023 nahm der Rat der OECD eine überarbeitete Version der Empfehlung an. Sie enthält eine aktualisierte Definition des Begriffs „KI-System“, die gewährleisten soll, dass die Empfehlung auch weiterhin dem Stand der Technik entspricht und neue technologische Entwicklungen widerspiegelt, insbesondere auf dem Gebiet der generativen KI. Auf der Grundlage des 2024 vorgelegten Berichts an den Rat über die Umsetzung, Verbreitung und weitere Relevanz der Empfehlung hat der Rat der OECD auf Minister-Ebene auf seiner Tagung vom 3. Mai 2024 eine weitere überarbeitete Version der Empfehlung angenommen. Ihre Anpassungen dienen dazu, weiteren technologischen Neuerungen u. a. auf dem Gebiet der generativen KI Rechnung zu tragen, politische Entwicklungen zu berücksichtigen und die Umsetzung der Empfehlung zu erleichtern.

Arbeiten der OECD zu KI

KI ist eine Querschnittstechnologie, die das Potenzial besitzt, die Wohlfahrt und die Lebensqualität der Menschen zu verbessern, zu einer positiven und nachhaltigen globalen Wirtschaftstätigkeit beizutragen, die Innovationstätigkeit und die Produktivität zu steigern und einen Beitrag zur Bewältigung dringender globaler Herausforderungen zu leisten. Sie kommt in vielen Bereichen zum Einsatz, im Fertigungs-, Bildungs- und Finanzsektor ebenso wie im Transport- und Gesundheitswesen oder auch im Bereich der Sicherheit.

Allerdings bringt KI nicht nur Vorteile, sondern auch gesellschaftliche und wirtschaftliche Herausforderungen mit sich, insbesondere was wirtschaftliche Veränderungen und Ungleichheiten, den Wettbewerb, Umbrüche auf dem Arbeitsmarkt sowie Konsequenzen für Demokratie und Menschenrechte betrifft.

Die Unterstützung der politischen Diskussion zum Thema KI durch empirische und politische Beiträge der OECD begann bereits 2016 mit einem Forum für Technikvorausschau (Technology Foresight Forum), auf das 2017 die internationale OECD-Konferenz „AI: Intelligent Machines, Smart Policies“ folgte. Zudem führte die Organisation Analysen und quantitative Arbeiten durch, um einen Überblick über den Stand der KI-Technik zu geben, wirtschaftliche und gesellschaftliche Effekte der KI und ihrer Anwendung darzulegen, grundsatzpolitische Fragen zu formulieren und KI-Initiativen von Regierungen und anderen Stakeholdern auf nationaler und internationaler Ebene zu beschreiben.

Dabei wurde deutlich, dass stabile politische Rahmenbedingungen auf internationaler Ebene erforderlich sind, um dafür zu sorgen, dass KI in der Gesellschaft auf Vertrauen stößt und Verbreitung findet. Vor diesem Hintergrund hat der Rat der OECD auf Vorschlag des DPC eine Empfehlung angenommen, die einen menschenzentrierten Ansatz für vertrauenswürdige KI fördert, die Forschung unterstützt, wirtschaftliche Innovationsanreize wahrt und sich an alle Stakeholder wendet.

Inklusiver und partizipativer Entstehungsprozess der Empfehlung

Die Ausarbeitung der Empfehlung folgte einem partizipativen Ansatz, während des gesamten Prozesses wurden Beiträge aus unterschiedlichsten Quellen berücksichtigt. Im Mai 2018 beschloss der DPC, eine Expertengruppe mit der Formulierung von Grundsätzen zu beauftragen, die dazu beitragen, dass KI auf Vertrauen stößt und Verbreitung findet. Ein entsprechender Empfehlungsentwurf sollte im Laufe des Jahres 2019 erarbeitet werden. Dies führte zur Gründung der informellen KI-Expertengruppe bei der OECD, an der mehr als 50 Expert*innen aus verschiedenen Fachbereichen und Sektoren (Regierung, Industrie, Zivilgesellschaft, Gewerkschaften, Technik und

Wissenschaft) beteiligt sind. Die vollständige Mitgliederliste ist unter <http://www.oecd.org/going-digital/ai/oecd-aigo-membership-list.pdf> einsehbar. Zwischen September 2018 und Februar 2019 kam die Gruppe zu vier Sitzungen zusammen. Als vorteilhaft für die Zusammenarbeit erwiesen sich dabei die Sorgfalt, das Engagement und die umfassenden Beiträge der Beteiligten, die große Bandbreite der vertretenen Stakeholder und die interdisziplinäre Zusammensetzung der Gruppe.

Auf der Grundlage des Schlussdokuments der informellen Gruppe arbeitete der DPC in Abstimmung mit weiteren einschlägigen OECD-Gremien einen Empfehlungsentwurf aus, den er in einer Sondertagung am 14. und 15. März 2019 annahm. Der Rat der OECD verabschiedete die Empfehlung im Rahmen seiner Tagung auf Ministerebene am 22. und 23. Mai 2019.

Anwendungsbereich der Empfehlung

Die Empfehlung stellt speziell KI betreffende Politikfragen in den Mittelpunkt. Damit ergänzt sie bereits bestehende OECD-Standards, die ebenfalls für KI relevant sind, weil sie z. B. den Schutz der Privatsphäre und den Datenschutz, das Risikomanagement im Bereich der digitalen Sicherheit oder verantwortungsvolles unternehmerisches Handeln betreffen. Die Empfehlung soll einen umsetzbaren Standard bieten, der flexibel genug ist, um trotz des raschen Wandels auf dem Gebiet der KI auf lange Sicht relevant zu bleiben. Sie enthält fünf Teile mit allgemeinen und wertebasierten Grundsätzen sowie fünf Empfehlungen für einzelstaatliche Maßnahmen und internationale Zusammenarbeit. Hinzu kommen Definitionen, die ein gemeinsames Verständnis von Schlüsselbegriffen wie „KI-System“, „Lebenszyklus eines KI-Systems“ oder „KI-Akteure“ im Kontext der Empfehlung gewährleisten sollen.

Die Empfehlung enthält im Wesentlichen zwei große Abschnitte:

1. **Grundsätze einer verantwortungsvollen Steuerung vertrauenswürdiger KI:** Der erste Abschnitt besteht aus fünf Teilen mit Grundsätzen, die sich ergänzen und für alle Stakeholder relevant sind: *i)* Inklusives Wachstum, nachhaltige Entwicklung und Lebensqualität; *ii)* Rechtsstaatlichkeit, Menschenrechte und demokratische Werte, insbesondere Fairness und Schutz der Privatsphäre; *iii)* Transparenz und Erklärbarkeit; *iv)* Robustheit und Sicherheit; *v)* Rechenschaftspflicht. Der Abschnitt enthält außerdem die Aufforderung an alle KI-Akteure, diese Grundsätze im Rahmen ihrer Rollen zur Geltung zu bringen und umzusetzen.
2. **Einzelstaatliche Maßnahmen und internationale Zusammenarbeit für vertrauenswürdige KI:** Der zweite Abschnitt enthält fünf mit den genannten Grundsätzen im Einklang stehende Empfehlungen für einzelstaatliche Maßnahmen und die internationale Zusammenarbeit der Mitglieds- und Nichtmitgliedsländer, die der Empfehlung zugestimmt haben (nachfolgend als die „zustimmenden Länder“ bezeichnet): *i)* In KI-Forschung und -Entwicklung investieren; *ii)* Ein inklusives und KI-freundliches Ökosystem fördern; *iii)* Governance und Politikumfeld KI-freundlich und interoperabel gestalten; *iv)* Die Kompetenzen der Menschen stärken und Vorkehrungen für Umbrüche am Arbeitsmarkt treffen; *v)* Internationale Zusammenarbeit für vertrauenswürdige KI pflegen.

Überarbeitungen der Empfehlung in den Jahren 2023 und 2024

2023 ergab sich die Gelegenheit, die Relevanz der Empfehlung mit einer Aktualisierung ihrer Definition eines „KI-Systems“ zu stützen. In einer gemeinsamen Sitzung am 16. Oktober 2023 einigte sich der DPC mit seiner Arbeitsgruppe AI-Governance (Working Party on AI Governance – AIGO) auf den Entwurf einer überarbeiteten Definition. Der Rat der OECD verabschiedete die überarbeitete Definition eines „KI-Systems“ in seiner Sitzung vom 8. November 2023. Die Änderungen an der Definition zielten u. a. darauf ab,

- die Ziele eines KI-Systems zu präzisieren (und klarzustellen, dass es sich um implizite oder explizite Ziele handeln kann),
- die Rolle des von Menschen oder Maschinen bereitgestellten Inputs zu unterstreichen,
- klarzustellen, dass sich die Empfehlung auch auf generative KI-Systeme bezieht, die „Inhalte“ erzeugen,
- den Begriff „reale“ durch den Begriff „physische“ zu ersetzen, um Unklarheiten zu beseitigen und die Wortwahl an anderen internationalen Initiativen auszurichten, und

- zu berücksichtigen, dass es KI-Systeme gibt, die sich auch nach Design und Bereitstellung noch weiterentwickeln können.

Im Einklang mit den Schlussfolgerungen des Berichts an den Rat 2024 ([2024 Report to Council](#)) wurde die Empfehlung während der OECD-Ministerratstagung 2024 einer weiteren Überarbeitung unterzogen, um fünf Jahre nach ihrer Verabschiedung ihre weitere Relevanz zu gewährleisten und ihre Umsetzung zu erleichtern. Die Anpassungen zielten insbesondere darauf ab,

- zu berücksichtigen, dass die Bekämpfung von Falsch- und Desinformation und die Wahrung der Informationsintegrität angesichts der generativen KI an Bedeutung gewinnen,
- Zweckentfremdungen, Missbrauch und unbeabsichtigten Fehlgebrauch zu adressieren,
- zu klären, welche Informationen KI-Akteure hinsichtlich ihrer KI-Systeme bereitstellen sollten, damit Transparenz und verantwortungsvolle Offenlegung gewährleistet sind,
- auf Sicherheitsbedenken einzugehen, sodass Menschen KI-Systeme sicher außer Kraft setzen, reparieren und/oder stilllegen können, wenn KI übermäßigen Schaden verursacht oder unerwünschtes Verhalten an den Tag legt,
- zu betonen, dass der gesamte Lebenszyklus von KI-Systemen mit verantwortungsvollem unternehmerischem Handeln einhergehen muss, was auch Konsequenzen für die Zusammenarbeit mit Lieferanten von KI-Wissen und -Ressourcen, den KI-Nutzern und anderen Stakeholdern hat,
- zu unterstreichen, dass die Staaten angesichts der weltweit steigenden Zahl der politischen Initiativen im Bereich KI zusammenarbeiten müssen, um hinsichtlich der Governance und politischen Rahmenbedingungen der KI Interoperabilität zu fördern, und
- explizit auf ökologische Nachhaltigkeit Bezug zu nehmen, da dieses Thema seit der Verabschiedung der Empfehlung im Jahr 2019 deutlich an Bedeutung gewonnen hat.

Um die Verständlichkeit zu verbessern, wurden außerdem einige der Überschriften der Grundsätze und Empfehlungen ausführlicher formuliert. Die Textabschnitte zu Rückverfolgbarkeit und Risikomanagement wurden erweitert und sind nun Teil der Ausführungen zum Prinzip der Rechenschaftspflicht, zu dem diese Begriffe am besten passen.

Weitere Informationen: [oecd.ai](#)

Kontakt: ai@oecd.org

Umsetzung

In der Empfehlung wird der DPC angewiesen, dem Rat fünf Jahre nach ihrer Verabschiedung und im Anschluss daran regelmäßig über die Umsetzung, Verbreitung und weitere Relevanz der Empfehlung Bericht zu erstatten.

Bericht an den Rat 2024

Fünf Jahre nach der Verabschiedung der Empfehlung hat die AIGO für den DPC den [Bericht an den Rat 2024](#) über die Umsetzung, Verbreitung und anhaltende Relevanz der Empfehlung erarbeitet. Dessen Schlussfolgerungen bildeten die Grundlage für Vorschläge des DPC für eine Überarbeitung der Empfehlung.

In dem Bericht kam der DPC zu dem Schluss, dass die Empfehlung ein wichtiges und nützliches internationales Referenzwerk für die Gestaltung der KI-Politik auf nationaler Ebene darstellt. Die Empfehlung wird von den zustimmenden Ländern umgesetzt, hat weite Verbreitung gefunden und ist nach wie vor umfassend relevant, u. a. als solider Rahmen für die Analyse technologischer Weiterentwicklungen wie z. B. der Neuerungen rund um generative KI.

Allerdings gilt es laut dem Bericht, Anpassungen vorzunehmen, um die Essenz mancher Bestimmungen klarer zu machen, die Umsetzung der Empfehlung zu erleichtern, ihre Relevanz zu erhöhen, und um zu gewährleisten, dass sie die Entwicklungen in Bezug auf generative KI und andere wichtige technologische Neuerungen widerspiegelt.

Weitere Arbeiten zur Förderung der Umsetzung der Empfehlung

Der DPC soll nicht nur dem Rat über die Umsetzung der Empfehlung Bericht erstatten, sondern auch seine Arbeit im Bereich KI fortsetzen. Dabei soll er auf der Empfehlung aufbauen und den Arbeiten Rechnung tragen, die andere internationale Foren leisten, z. B. die UNESCO, die Europäische Union, der Europarat und die Initiative zur Einrichtung eines internationalen Panels für KI (vgl. <https://www.pm.gc.ca/en/news/backgrounders/2018/12/06/mandate-international-panel-artificial-intelligence> und <https://www.gouvernement.fr/en/france-and-canada-create-new-expert-international-panel-on-artificial-intelligence>).

Der Rat hat den DPC außerdem beauftragt, die Umsetzung der Empfehlung zu unterstützen, indem er Leitlinien für die Umsetzungspraxis erarbeitet, ein Forum für den Informationsaustausch über KI-Maßnahmen und -Aktivitäten anbietet und den interdisziplinären Dialog zwischen verschiedenen Stakeholdern fördert.

Um ein inklusives Forum für den Austausch von Informationen zu KI-Politik und -Aktivitäten anzubieten sowie den interdisziplinären Dialog zwischen verschiedenen Stakeholdern zu fördern, hat die OECD im Februar 2020 das [AI Policy Observatory](#) (OECD.AI) und das informelle OECD Network of Experts on AI (ONE AI) eingerichtet.

OECD.AI soll den Staaten als inklusiver Wissenshub zur KI-Politik helfen, die verantwortungsvolle Entwicklung vertrauenswürdiger KI-Systeme im Dienst der Gesellschaft zu fördern, zu unterstützen und zu überwachen. Das Observatory gibt OECD-Ländern und Partnern aus dem Kreis der Stakeholder die Möglichkeit, ihre Ressourcen zu bündeln, um KI-Politik multidisziplinär und evidenzbasiert zu analysieren. Es bietet auch eine Echtzeitdatenbank, die es den Ländern und anderen Stakeholdern erlaubt, aktuelle Informationen zu ihren KI-Strategien, -Maßnahmen und -Initiativen bereitzustellen. Auf diese Weise können die zentralen Elemente interaktiv verglichen werden. Die Datenbank wird kontinuierlich mit KI-Metriken und anderen Messdaten sowie Angaben zu Maßnahmen und bewährten Verfahren auf den neuesten Stand gebracht, woraus sich weitere Anpassungen der Leitlinien für die Umsetzungspraxis ergeben.

Das OECD.AI Network of Experts (ONE AI) ist eine informelle Gruppe von KI-Expert*innen aus Regierung und Verwaltung, Wirtschaft, Wissenschaft und Zivilgesellschaft, die die OECD mit politischer Fachkompetenz zum Thema KI unterstützt und berät. ONE AI bietet der internationalen KI-Community ein Forum, um gemeinsame Chancen und Herausforderungen der KI-Politik eingehend zu erörtern.

Angesichts der steigenden Bedeutung und Reichweite seiner Arbeiten zum Thema KI gründete der DPC im Jahr 2022 eine Arbeitsgruppe, die sich speziell diesem Thema widmet, die Working Party on AI Governance (AIGO). Sie unterstützt die Umsetzung der Empfehlung, erarbeitet Leitlinien für die Umsetzungspraxis, nutzt das OECD.AI Policy Observatory als Forum und Online-Plattform für den Austausch von Informationen zu KI-Politik und -Aktivitäten, und sie fördert den interdisziplinären Dialog verschiedener Stakeholder über die Expertengruppen des OECD.AI Network of Experts.

DER RAT,

GESTÜTZT AUF Artikel 5 Buchstabe b des Übereinkommens über die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung vom 14. Dezember 1960;

GESTÜTZT AUF die von der OECD erarbeiteten Standards zum Schutz der Privatsphäre, für digitale Sicherheit, Verbraucherschutz und verantwortungsvolles unternehmerisches Handeln;

GESTÜTZT AUF die in der Agenda 2030 für nachhaltige Entwicklung aufgeführten und von der Generalversammlung der Vereinten Nationen verabschiedeten Ziele für nachhaltige Entwicklung ([A/RES/70/1](#)) sowie die 1948 verabschiedete Allgemeine Erklärung der Menschenrechte;

IM HINBLICK AUF die derzeit in anderen internationalen Regierungs- und Nichtregierungsforen durchgeführten wichtigen Arbeiten im Bereich der künstlichen Intelligenz (im Folgenden als „KI“ bezeichnet);

IN DER ERKENNTNIS, dass KI tiefgreifende, weitreichende und globale Auswirkungen hat, die Gesellschaften, Wirtschaftszweige und die Arbeitswelt grundlegend verändern, und dass sich diese Entwicklung in Zukunft voraussichtlich noch verstärken wird;

IN DER ERKENNTNIS, dass KI das Potenzial besitzt, die Wohlfahrt und die Lebensqualität der Menschen zu verbessern, zu einer positiven und nachhaltigen globalen Wirtschaftstätigkeit beizutragen, die Innovationstätigkeit und die Produktivität zu steigern und einen Beitrag zur Bewältigung dringender globaler Herausforderungen zu leisten;

IN DER ERKENNTNIS, dass diese Transformationen gleichzeitig sowohl innerhalb der Gesellschaften und Volkswirtschaften als auch zwischen ihnen verschiedenste Effekte haben können, insbesondere im Hinblick auf wirtschaftliche Veränderungen, den Wettbewerb, Umbrüche auf dem Arbeitsmarkt und Ungleichheiten, und dass sie sich auf die Demokratie, die Einhaltung der Menschenrechte, den Schutz der Privatsphäre und den Datenschutz sowie die digitale Sicherheit auswirken können;

IN DER ERKENNTNIS, dass Vertrauen eine wichtige Triebkraft der digitalen Transformation ist, dass die Vertrauenswürdigkeit von KI-Systemen ein entscheidender Faktor für die Verbreitung und den Einsatz von KI ist, auch wenn die künftigen Arten und Auswirkungen von KI-Anwendungen u. U. nur schwer vorherzusehen sind, und dass eine fundierte, gesamtgesellschaftliche öffentliche Debatte notwendig ist, um das positive Potenzial der Technologie auszuschöpfen und zugleich die damit verbundenen Risiken zu begrenzen;

UNTER HINWEIS DARAUF, dass es bereits bestimmte nationale und internationale rechtliche und regulatorische Regelwerke und Politikrahmen gibt, die für KI relevant sind, indem sie sich beispielsweise auf die Menschenrechte, den Verbraucherschutz und den Schutz personenbezogener Daten, geistige Eigentumsrechte, verantwortungsvolles unternehmerisches Handeln und den Wettbewerb beziehen, und in Anerkennung der Tatsache, dass möglicherweise einige Regelwerke im Hinblick auf ihre Eignung überprüft und neue Ansätze entwickelt werden müssen;

IN DER ERKENNTNIS, dass angesichts der raschen Entwicklung und Einführung von KI ein stabiles Politikumfeld erforderlich ist, das einen menschenzentrierten Ansatz für vertrauenswürdige KI fördert, die Forschung unterstützt, wirtschaftliche Innovationsanreize bewahrt und alle Stakeholder ihren jeweiligen Rollen und dem jeweiligen Kontext entsprechend berücksichtigt;

IN DER ERWÄGUNG, dass es von entscheidender Bedeutung ist, die Chancen, die KI-Anwendungen bieten, zu nutzen, die daraus resultierenden Herausforderungen anzugehen, und die Stakeholder zu befähigen sich einzubringen, um dazu beizutragen, dass vertrauenswürdige KI in der Gesellschaft Verbreitung findet, und um dafür zu sorgen, dass die Vertrauenswürdigkeit von KI ein Wettbewerbsfaktor auf dem Weltmarkt wird;

Auf Vorschlag des Ausschusses für Digitalpolitik:

I. **KOMMT ÜBEREIN**, dass die folgenden Begriffe für die Zwecke dieser Empfehlung folgendermaßen zu verstehen sind:

- *KI-System*: Ein KI-System ist ein maschinenbasiertes System, das expliziten oder impliziten Zielsetzungen dient und aus erhaltenen Inputs darauf schließt, wie Vorhersagen, Inhalte, Empfehlungen, Entscheidungen oder andere Outputs zu erzeugen sind, die die physische oder virtuelle Umgebung beeinflussen können. KI-Systeme unterscheiden sich hinsichtlich ihrer Autonomie und Anpassungsfähigkeit nach Einführung.
- *Lebenszyklus eines KI-Systems*: Der Lebenszyklus eines KI-Systems besteht typischerweise aus mehreren Phasen, u. a.: Planung und Design; Datensammlung und -verarbeitung; Entwicklung des Modells und/oder Anpassung eines oder mehrerer bestehender Modelle an spezifische Aufgaben; Test, Evaluierung, Verifizierung und Validierung; Bereitstellung/Einführung; Betrieb und Monitoring; Außerbetriebnahme/Stilllegung. Diese Phasen finden häufig iterativ statt und sind nicht zwangsläufig sequenziell. Die Entscheidung, ein KI-System außer Betrieb zu nehmen, kann zu jedem Zeitpunkt der Betriebs- und Monitoringphase erfolgen.
- *KI-Akteure*: KI-Akteure sind Akteure, die im Lebenszyklus des KI-Systems eine aktive Rolle spielen, einschließlich Organisationen und Personen, die KI-Systeme einführen oder betreiben.
- *KI-Wissen*: KI-Wissen bezieht sich auf die Kompetenzen und Ressourcen, die erforderlich sind, um den Lebenszyklus des KI-Systems zu verstehen, an ihm mitzuwirken sowie seine Risiken zu steuern, und umfasst u. a. Daten, Quelltext, Algorithmen, Modelle, Forschung, Know-how, Ausbildungsprogramme, Governance, Prozesse und empfehlenswerte Praktiken.
- *Stakeholder*: Stakeholder umfassen alle Organisationen und Personen, die direkt oder indirekt an KI-Systemen mitwirken oder von ihnen betroffen sind. KI-Akteure sind eine Untergruppe der Stakeholder.

Abschnitt 1: Grundsätze einer verantwortungsvollen Steuerung vertrauenswürdiger KI

II. **EMPFIEHLT**, dass die Mitglieder und Nichtmitglieder, die dieser Empfehlung zustimmen (im Folgenden als die „zustimmenden Länder“ bezeichnet) die folgenden für alle Stakeholder relevanten Grundsätze einer verantwortungsvollen Steuerung vertrauenswürdiger KI zur Geltung bringen und umsetzen.

III. **FORDERT** alle KI-Akteure **AUF**, die folgenden Grundsätze einer verantwortungsvollen Steuerung vertrauenswürdiger KI entsprechend ihren jeweiligen Rollen zur Geltung zu bringen und umzusetzen.

IV. **UNTERSTREICHT**, dass sich die folgenden Grundsätze ergänzen und als Ganzes betrachtet werden sollten.

1.1. Inklusives Wachstum, nachhaltige Entwicklung und Lebensqualität

Die Stakeholder sollten sich proaktiv für eine verantwortungsvolle Steuerung vertrauenswürdiger KI einsetzen, die darauf abzielt, einen Nutzen für die Menschen und den Planeten zu erzielen – wie eine Erweiterung der menschlichen Fähigkeiten und eine Steigerung der Kreativität, eine stärkere Einbindung unterrepräsentierter Bevölkerungsgruppen, eine Verringerung der wirtschaftlichen, sozialen, geschlechtsspezifischen und sonstigen Ungleichheiten sowie den Schutz der natürlichen Umwelt – und damit ein inklusives Wachstum, die Lebensqualität, eine nachhaltige Entwicklung und ökologische Nachhaltigkeit zu fördern.

1.2. Rechtsstaatlichkeit, Menschenrechte und demokratische Werte, insbesondere Fairness und Schutz der Privatsphäre

- a) Die KI-Akteure sollten während des gesamten Lebenszyklus eines KI-Systems das Prinzip der Rechtsstaatlichkeit, die Menschenrechte sowie demokratische und menschenzentrierte Werte achten. Hierzu zählen u. a. Nichtdiskriminierung und Gleichbehandlung, Freiheit, Würde und Selbstbestimmung, Schutz der Privatsphäre und Datenschutz, Vielfalt, Fairness, soziale Gerechtigkeit und international anerkannte Arbeitsrechte. Dazu gehört auch, sich unter Wahrung des Rechts auf freie Meinungsäußerung und anderer durch das geltende Völkerrecht geschützter Rechte und Freiheiten gegen die Verbreitung von Falsch- und Desinformation zu wenden, die durch KI zunimmt.

- b) Zu diesem Zweck sollten die KI-Akteure Mechanismen und Schutzmaßnahmen einführen, mit denen u. a. die Risiken, die sich aus Zweckentfremdung, Missbrauch oder unbeabsichtigtem Fehlgebrauch ergeben, dem Kontext angemessen und gemäß dem neuesten Stand der Technik adressiert werden können, z. B. durch menschliche Eingriffe und Überwachung.

1.3. Transparenz und Erklärbarkeit

Die KI-Akteure sollten sich in Bezug auf KI-Systeme zu Transparenz und verantwortungsvoller Offenlegung verpflichten. Zu diesem Zweck sollten sie aussagekräftige, dem Kontext angemessene und dem neuesten Stand der Technik entsprechende Informationen bereitstellen, um

- i. ein allgemeines Verständnis von KI-Systemen sowie ihrer Möglichkeiten und Grenzen zu fördern,
- ii. den Stakeholdern ihre Interaktionen mit KI-Systemen bewusst zu machen, z. B. wenn sie am Arbeitsplatz mit KI zu tun haben,
- iii. Menschen, die mit einem KI-System zu tun haben, klar und verständlich über die Datenquellen/den Input, die Einflussfaktoren, Prozesse und/oder Logik zu informieren, die der Vorhersage, den Inhalten, der Empfehlung oder der Entscheidung eines KI-Systems zugrunde liegen, soweit dies realisierbar und hilfreich für das Verständnis des Outputs ist, und
- iv. die von negativen Auswirkungen eines KI-Systems Betroffenen durch die Bereitstellung von Informationen zu befähigen, dessen Output infrage zu stellen.

1.4. Robustheit und Sicherheit

- a) KI-Systeme sollten während ihres gesamten Lebenszyklus robust und sicher sein, sodass sie bei normaler oder vorhersehbarer Verwendung ebenso wie bei Missbrauch bzw. unter sonstigen ungünstigen Bedingungen angemessen funktionieren und kein unzumutbares Sicherheitsrisiko darstellen.
- b) Gegebenenfalls sollten Sicherheitsmechanismen gewährleisten, dass KI-Systeme gefahrlos außer Kraft gesetzt, repariert und/oder stillgelegt werden können, wenn dies nötig ist, weil die Gefahr besteht, dass sie übermäßigen Schaden verursachen oder unerwünschtes Verhalten an den Tag legen.
- c) Darüber hinaus sollten weitere Mechanismen, soweit technisch machbar, die Informationsintegrität stärken, ohne dabei das Recht auf freie Meinungsäußerung zu gefährden.

1.5. Rechenschaftspflicht

- a) Die KI-Akteure sollten für den einwandfreien Betrieb der KI-Systeme und die Achtung der oben genannten Grundsätze rechenschaftspflichtig sein; die Rechenschaftslegung sollte auf ihren Rollen und dem jeweiligen Kontext basieren und dem neuesten Stand der Technik entsprechen.
- b) Zu diesem Zweck sollten KI-Akteure die Rückverfolgbarkeit gewährleisten – u. a. in Bezug auf Datensätze, Prozesse und die im Lebenszyklus eines KI-Systems getroffenen Entscheidungen –, um eine Analyse des Outputs und der Antworten des KI-Systems zu ermöglichen, die dem Kontext angemessen ist und dem neuesten Stand der Technik entspricht.
- c) Um den Risiken im Zusammenhang mit KI-Systemen zu begegnen, sollten die KI-Akteure in allen Phasen des Lebenszyklus eines KI-Systems kontinuierlich und ihren Rollen, dem Kontext und ihrer Handlungsfähigkeit entsprechend einen systematischen Risikomanagementansatz verfolgen und verantwortungsvoll unternehmerisch handeln; dafür sollten sie gegebenenfalls auch mit anderen KI-Akteuren, Lieferanten von KI-Wissen und -Ressourcen, Nutzern von KI-Systemen und weiteren Stakeholdern zusammenarbeiten. Die betreffenden Risiken können beispielsweise von schädlichen Bias ausgehen oder auch die Sicherheit, die Privatsphäre, andere Menschenrechte oder auch Arbeitsrechte und geistige Eigentumsrechte betreffen.

Abschnitt 2: Einzelstaatliche Maßnahmen und internationale Zusammenarbeit für vertrauenswürdige KI

V. EMPFIEHLT, dass die zustimmenden Länder im Rahmen ihrer nationalen Politik und ihrer internationalen Zusammenarbeit, im Einklang mit den in Abschnitt 1 genannten Grundsätzen sowie unter besonderer Berücksichtigung kleiner und mittlerer Unternehmen (KMU) die folgenden Empfehlungen umsetzen.

2.1. In KI-Forschung und -Entwicklung investieren

- a) Die Länder sollten im Hinblick auf Forschung, Entwicklung und offene Wissenschaft, einschließlich interdisziplinärer Anstrengungen, langfristige öffentliche Investitionen in Betracht ziehen und Anreize für private Investitionen schaffen, um Innovationen im Bereich vertrauenswürdige KI zu fördern, deren Fokus auf schwierigen technischen Fragen sowie auf sozialen, rechtlichen und ethischen Implikationen bzw. Politikproblemen im Zusammenhang mit KI liegt.
- b) Die Länder sollten außerdem im Hinblick auf Open-Source-Tools und offene Datensätze, die repräsentativ sind und dem Schutz der Privatsphäre sowie dem Datenschutz gerecht werden, öffentliche Investitionen in Betracht ziehen und Anreize für private Investitionen schaffen, um ein KI-Forschungs- und -Entwicklungsumfeld ohne schädlichen Bias zu fördern und die Interoperabilität sowie die Anwendung von Standards zu verbessern.

2.2. Ein inklusives und KI-freundliches Ökosystem fördern

Die Länder sollten die Entwicklung und Zugänglichkeit eines inklusiven, dynamischen, nachhaltigen und interoperablen digitalen Ökosystems für vertrauenswürdige KI fördern. Ein solches Ökosystem umfasst u. a. Daten, KI-Technologien, Rechenkapazitäten und Netzinfrastruktur sowie Mechanismen für einen angemessenen Austausch von KI-Wissen. Diesbezüglich sollten die Länder die Förderung von Mechanismen wie Datentreuhandmodelle in Betracht ziehen, um den sicheren, fairen, legalen und ethisch korrekten Austausch von Daten zu stärken.

2.3. Governance und Politikumfeld KI-freundlich und interoperabel gestalten

- a) Die Länder sollten ein agiles Politikumfeld fördern, das den Übergang vertrauenswürdiger KI-Systeme von der Forschungs- und Entwicklungsphase hin zur Bereitstellung und Betriebsphase unterstützt. Zu diesem Zweck sollten sie die Einrichtung von Experimentierräumen in Betracht ziehen, damit KI-Systeme in einem kontrollierten Umfeld getestet und gegebenenfalls skaliert werden können. Sie sollten außerdem ergebnisorientierte Ansätze verfolgen, die einen flexiblen Umgang mit Governancezielen erlauben, sowie durch Zusammenarbeit auf inner- und zwischenstaatlicher Ebene die Interoperabilität der Governance und des Politikumfelds fördern.
- b) Die Länder sollten ihre für KI-Systeme geltenden Politik-, Regulierungsrahmen und Evaluierungsmechanismen überprüfen und gegebenenfalls anpassen, um Innovationen und Wettbewerb in Bezug auf vertrauenswürdige KI zu fördern.

2.4. Die Kompetenzen der Menschen stärken und Vorkehrungen für Umbrüche am Arbeitsmarkt treffen

- a) Die Länder sollten eng mit den Stakeholdern zusammenarbeiten, um sich für den Wandel der Arbeitswelt und der Gesellschaft zu wappnen. Sie sollten die Menschen befähigen, KI-Systeme des gesamten Anwendungsspektrums effektiv zu nutzen und mit ihnen zu interagieren, u. a. indem sie sie mit den erforderlichen Kompetenzen ausstatten.
- b) Die Länder sollten u. a. im Rahmen eines sozialen Dialogs gewährleisten, dass sich die Arbeitskräfte bei Einführung von KI zu fairen Bedingungen umstellen können, beispielsweise durch berufliche Fort- und Weiterbildungsprogramme für alle Altersstufen, soziale Sicherung, Unterstützungsmaßnahmen für entlassene Arbeitskräfte sowie die Eröffnung neuer Arbeitsmarktchancen.
- c) Die Länder sollten darüber hinaus eng mit Stakeholdern zusammenarbeiten, um einen verantwortungsvollen Einsatz von KI in der Arbeitswelt zu fördern, die Sicherheit der Arbeitskräfte zu verbessern, die Beschäftigungs- und die öffentliche Dienstleistungsqualität zu erhöhen, unternehmerische Initiative und Produktivität zu stärken sowie sich zu bemühen, eine faire Teilhabe breiter Bevölkerungsschichten am Nutzen von KI zu gewährleisten.

2.5. Internationale Zusammenarbeit für vertrauenswürdige KI pflegen

- a) Die Länder, einschließlich der Entwicklungsländer, sollten aktiv zusammenarbeiten und auch die Stakeholder einbeziehen, um die hier genannten Grundsätze zur Geltung zu bringen und bei der verantwortungsvollen Steuerung vertrauenswürdiger KI Fortschritte zu erzielen.
- b) Die Länder sollten im Rahmen der OECD und anderer globaler und regionaler Foren zusammenarbeiten, um den Austausch von KI-Wissen angemessen zu fördern. Sie sollten internationale, sektorübergreifende und offene Multi-Stakeholder-Initiativen unterstützen, um ihre KI-Expertise nachhaltig zu erweitern.
- c) Die Länder sollten die konsensbasierte Ausarbeitung globaler technischer Multi-Stakeholder-Standards für interoperable und vertrauenswürdige KI fördern.
- d) Die Länder sollten darüber hinaus auf die Entwicklung international vergleichbarer Indikatoren zur Evaluierung der KI-Forschung, -Entwicklung und -Einführung hinwirken, diese Indikatoren selbst nutzen und eine Evidenzbasis aufbauen, um die Fortschritte zu evaluieren, die bei der Umsetzung der genannten Grundsätze erzielt wurden.

VI. ERSUCHT den Generalsekretär und die zustimmenden Länder, diese Empfehlung zu verbreiten.

VII. ERSUCHT die nicht zustimmenden Länder, diese Empfehlung gebührend zur Kenntnis zu nehmen und ihr zuzustimmen.

VIII. WEIST den Ausschuss für Digitalpolitik **AN**, über seine Working Party on AI Governance

- a) aufbauend auf dieser Empfehlung seine wichtige Arbeit im Bereich künstliche Intelligenz fortzusetzen und dabei den Arbeiten anderer internationaler Foren Rechnung zu tragen sowie den Rahmen zur Bewertung der evidenzbasierten KI-Politik weiterzuentwickeln;
- b) weitere praktische Leitlinien zur Umsetzung dieser Empfehlung zu erarbeiten bzw. zu erneuern, um aktuellen Entwicklungen und neuen politischen Prioritäten in angemessener Weise Rechnung zu tragen;
- c) ein Forum für den Austausch von Informationen über KI-Maßnahmen und -Aktivitäten, einschließlich der Erfahrungen mit der Umsetzung dieser Empfehlung, bereitzustellen und einen interdisziplinären Dialog zwischen verschiedenen Stakeholdern zu fördern, um das Vertrauen in KI zu stärken und ihre Einführung voranzutreiben;
- d) dem Rat in Abstimmung mit anderen einschlägigen Ausschüssen spätestens fünf Jahre nach der Überarbeitung der Empfehlung und im Anschluss mindestens alle zehn Jahre Bericht über die Umsetzung, Verbreitung und weitere Relevanz dieser Empfehlung zu erstatten.

Über die OECD

Die OECD ist ein in ihrer Art einzigartiges Forum, in dem Regierungen gemeinsam an der Bewältigung der wirtschaftlichen, sozialen und ökologischen Herausforderungen der Globalisierung arbeiten. Darüber hinaus unterstützt die OECD die Regierungen maßgeblich dabei, Antworten auf neue Entwicklungen und Fragestellungen – z. B. Corporate Governance, Informationsökonomie und die Herausforderungen der Bevölkerungsalterung – zu finden. Die OECD bietet den Regierungen einen Rahmen, um Erfahrungen aus verschiedenen Politikbereichen auszutauschen, Lösungen für gemeinsame Probleme zu erarbeiten, gute Praktiken aufzuzeigen sowie nationale und internationale Maßnahmen zu koordinieren.

Die OECD-Mitgliedsländer sind: Australien, Belgien, Chile, Costa Rica, Dänemark, Deutschland, Estland, Finnland, Frankreich, Griechenland, Irland, Island, Israel, Italien, Japan, Kanada, Kolumbien, Korea, Lettland, Litauen, Luxemburg, Mexiko, Neuseeland, die Niederlande, Norwegen, Österreich, Polen, Portugal, Schweden, die Schweiz, die Slowakische Republik, Slowenien, Spanien, die Tschechische Republik, die Republik Türkei, Ungarn, das Vereinigte Königreich und die Vereinigten Staaten. Die Europäische Union beteiligt sich an der Arbeit der OECD.

OECD-Rechtsinstrumente

Seit Gründung der OECD im Jahr 1961 wurden in ihrem Rahmen über 500 Rechtsinstrumente ausgearbeitet. Dazu gehören die Beschlüsse und Empfehlungen, die der Rat der OECD gemäß dem Übereinkommen über die OECD verabschiedet hat, aber auch andere Rechtsinstrumente, z. B. Erklärungen und internationale Übereinkommen.

Alle Rechtsinstrumente der OECD, ob in Kraft oder aufgehoben, sind im Online-Kompendium der OECD-Rechtsinstrumente aufgeführt. Sie gliedern sich in fünf Kategorien:

- **Beschlüsse** werden vom Rat verabschiedet und sind für alle Mitgliedsländer rechtlich bindend, außer für diejenigen Länder, die sich zum Zeitpunkt der Verabschiedung enthalten haben. Sie legen konkrete Rechte und Pflichten fest und können Monitoring-Mechanismen enthalten.
- **Empfehlungen** werden vom Rat verabschiedet und sind nicht rechtlich bindend. Sie stellen ein politisches Bekenntnis zu den darin enthaltenen Grundsätzen dar und sind mit der Erwartung verbunden, dass sich die Teilnehmer nach Kräften bemühen, die Empfehlungen umzusetzen.
- **Abschlussdokumente** werden zum Abschluss einer Tagung des Rates auf Ministerebene oder einer sonstigen hochrangigen Tagung im Rahmen der Organisation von den jeweils aufgeführten Teilnehmern und damit nicht von einem OECD-Organ verabschiedet. Sie legen in der Regel allgemeine Grundsätze oder langfristige Ziele fest und haben einen feierlichen Charakter.
- **Internationale Übereinkommen** werden im Rahmen der Organisation verhandelt und vereinbart. Sie sind für alle Vertragsparteien rechtlich bindend.
- **Vereinbarungen, Verständigungen und sonstige Rechtsinstrumente:** Im Laufe der Zeit wurden im Rahmen der OECD mehrere andere Rechtsinstrumente erarbeitet, wie die Vereinbarung über staatlich geförderte Exportkredite, die Internationale Vereinbarung über Grundsätze des Seeverkehrs (International Understanding on Maritime Transport Principles) und die Empfehlungen des Entwicklungsausschusses (DAC).