www.jusletter-it.eu

Christine Kaufmann

Neue OECD-Instrumente zu künstlicher Intelligenz

Wege zu vertrauenswürdiger künstlicher Intelligenz

Der Beitrag gibt einen ersten summarischen Überblick über neue, noch wenig bekannte Entwicklungen in der OECD zur Unterstützung von vertrauenswürdiger KI durch Regierungen und Unternehmen. Die Autorin fasst die aktuellen Diskussionen zusammen und bettet sie in den Zusammenhang mit den wegleitenden Instrumenten der OECD zur Sorgfaltspflicht (due diligence) von Unternehmen ein. Sie kommt zum Schluss, dass die laufenden Arbeiten an einem Leitfaden zur Sorgfaltsprüfung für KI das Potential haben, der zunehmenden Fragmentierung von KI-bezogenen Standards in einem spezifischen Bereich entgegenzuwirken.

Beitragsart: Beiträge
Rechtsgebiete: Artificial Intelligence & Recht

Zitiervorschlag: Christine Kaufmann, Neue OECD-Instrumente zu künstlicher Intelligenz, in: Jusletter IT 4. Juli 2024

ISSN 1664-848X, jusletter-it.weblaw.ch, Weblaw AG, info@weblaw.ch, T +41 31 380 57 77

Inhaltsübersicht

1. Auftakt: Empfehlung der OECD zu künstlicher Intelligenz 2019
2. Rezeption: OECD-Leitsätze für multinationale Unternehmen zu verantwortungsvollem unternehmerischem Handeln
   2.1. Grundzüge und Rechtsnatur
   2.2. Aktualisierung der Leitsätze 2023: Relevanz für künstliche Intelligenz
3. Weiterführung: Aktualisierung der OECD-Empfehlung zu KI 2023/2024
   3.1. Aktualisierung der Definition von KI-Systemen
   3.2. Weiterentwicklung der KI-Prinzipien in der Empfehlung 2024
4. Ausblick: Konsolidierung der Standards für KI-Risikomanagement?
   4.1. Zunehmende Proliferation und Fragmentierung
   4.2. Ausblick: Konsolidierung in Sicht?

1. Auftakt: Empfehlung der OECD zu künstlicher Intelligenz 2019

[1] 2019 wurden vom Rat der OECD mit der Empfehlung zu künstlicher Intelligenz (KI)[1] die ersten zwischenstaatlichen politischen Leitlinien für künstliche Intelligenz (KI) verabschiedet. Ziel der Empfehlung war es, einen sowohl innovativen als auch verantwortungsvollen Umgang mit KI zu fördern. Um dieses Ziel zu erreichen, bekannten sich die unterzeichnenden Staaten – die damals 36 OECD-Mitgliedstaaten sowie Argentinien, Brasilien, Kolumbien, Costa Rica, Peru und Rumänien – dazu, ihre Verantwortung in der Steuerung von vertrauenswürdiger KI unter Beachtung der Menschenrechte und der Demokratie wahrzunehmen. Um die Staaten dabei zu unterstützen, formulierte die Empfehlung eine Reihe von Grundsätzen, die im Austausch mit einer Multistakeholder-Expertengruppe erarbeitet wurden. Diese Gruppe setzte sich aus Vertretern von Regierungen, Wissenschaft, Unternehmen, Zivilgesellschaft, internationalen Gremien, der Tech-Community und Gewerkschaften zusammen.

[2] Der erste Abschnitt der Empfehlung enthielt wertebasierte Grundsätze für eine verantwortungsvolle Steuerung vertrauenswürdiger KI. Als fünf wegleitende Grundwerte genannt wurden

- Inklusives Wachstum, nachhaltige Entwicklung und Lebensqualität
- Menschenzentrierte Werte und Fairness
- Transparenz und Erklärbarkeit
- Robustheit und Sicherheit
- Rechenschaftspflicht

[3] Der zweite Abschnitt der Empfehlung widmete sich einzelstaatlichen Massnahmen und der internationalen Zusammenarbeit für vertrauenswürdige KI und formulierte fünf Handlungsempfehlungen:

- In KI-Forschung und -Entwicklung investieren
- Ein digitales Ökosystem für KI fördern
- Ein für KI günstiges Politikumfeld schaffen

[1] Empfehlung des Rates zu künstlicher Intelligenz (Recommendation of the Council on Artificial Intelligence) vom 22. Mai 2019 (inoffizelle deutsche Übersetzung: https://www.oecd.org/berlin/presse/Empfehlung-des-Rats-zu-kuenstlicher-Intelligenz.pdf); Karen Yeung, Recommendation of the Council on Artificial Intelligence (OECD), International Legal Materials 59/2020, S. 27 ff.

- Die Kompetenzen der Menschen stärken und sich auf den Wandel des Arbeitsmarktes vorbereiten
- Internationale Zusammenarbeit für vertrauenswürdige KI

[4] Ergänzt wurden diese, im Hinblick auf die sich rasch entwickelnde Technologie flexibel formulierten Grundsätze mit den notwendigen Definitionen, um ein gemeinsames Verständnis von Schlüsselbegriffen wie «KI-System», «Lebenszyklus eines KI-Systems» oder «KI-Akteure» im Kontext der Empfehlung sicherzustellen.

[5] Zusammenfassend sollten die zehn Grundsätze Regierungen, Organisationen und Einzelpersonen dabei unterstützen, KI-Systeme so zu gestalten und zu betreiben, dass die Interessen der Menschen im Zentrum stehen. Zudem sollten sie sicherstellen, dass die Entwickler und Betreiber von KI-basierten Systemen und Produkten für das ordnungsgemässe Funktionieren zur Rechenschaft gezogen werden.

[6] Der Generalsekretär der OECD prophezeite den neuen KI-Prinzipien der OECD anlässlich deren Verabschiedung 2019, zum globalen Bezugspunkt für eine glaubwürdige KI zu werden, da sie einem klaren Ziel dienten: Regierungen sollten die Entwicklung von KI-Systemen sicherstellen, die Werte und Gesetze respektieren, damit die Menschen darauf vertrauen können, dass ihre Sicherheit und ihre Privatsphäre an erster Stelle stehen.[2]

[7] Tatsächlich wurden die zehn KI-Prinzipien der OECD von 2019 nicht nur von der Europäischen Kommission unterstützt, sondern kurze Zeit später auch von den Staaten der G-20, die am Gipfeltreffen in Osaka die *G-20 Principles on Artificial Intelligence* verabschiedeten, welche sich auf die KI-Empfehlung der OECD beriefen.[3]

[8] Empfehlungen des OECD-Rates[4] sind rechtlich nicht verbindlich, sondern bringen ein politisches Bekenntnis der unterzeichnenden Staaten zu den in ihnen enthaltenen Grundsätzen zum Ausdruck. Mit einer Empfehlung ist die Erwartung verbunden, dass die unterzeichnenden Mitgliedstaaten sie umsetzen. Deshalb löst eine Empfehlung eine Berichterstattungspflicht des für das Thema sachlich zuständigen Ausschusses an den Rat aus. Für die KI-Empfehlung wurde der erste Bericht nach fünf Jahren, im Mai 2024, vorgelegt.[5]

---

2 Lancierung der OECD Empfehlung zu künstlicher Intelligenz, Pressemitteilung der OECD vom 22. Mai 2019, OECD, Paris (https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm).

3 G-20 AI Principles, Annex 8 to the G-20 Osaka Leaders» Declaration, Osaka 2019 (https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf).

4 Art. 5 lit. b Übereinkommen über die Organisation für Wirtschaftliche Zusammenarbeit und Entwicklung (OECD-Konvention) vom 14. Dezember 1960, SR 0.970.4.

5 Ziff. VIII.d OECD KI-Empfehlung 2019 (Fn. 1). Report on the Implementation of the OECD Recommendation on Artificial Intelligence, 24 April 2024, C/Min(2024)17 (https://one.oecd.org/document/C/MIN(2024)17/en/pdf).

2. Rezeption: OECD-Leitsätze für multinationale Unternehmen zu verantwortungsvollem unternehmerischem Handeln

2.1. Grundzüge und Rechtsnatur

[9] Die OECD-Leitsätze für multinationale Unternehmen zu verantwortungsvollem unternehmerischem Handeln (nachfolgend OECD-Leitsätze)[6] sind als Anhang I Teil der Investitionserklärung der OECD von 1976.[7] Diese hat kurz gefasst zum Ziel, Investitionen mit fairen Rahmenbedingungen zu fördern. Sie anerkennt, dass Investitionen einen positiven Beitrag zur Entwicklung in den Gastländern leisten, aber auch mit negativen Auswirkungen verbunden sein können. Die Staaten, die der Investitionserklärung beitreten, verpflichten sich deshalb, ihre Unternehmen zu verantwortungsvollem Handeln, wie es in den Leitsätzen umschrieben ist, anzuhalten.

[10] Diese Leitsätze sind nicht rechtsverbindlich, sondern Empfehlungen der Regierungen an Unternehmen. Anlässlich der Aktualisierung 2023 wurde dies ausdrücklich festgehalten:

> «4. In den Leitsätzen wird zum Ausdruck gebracht, welche gemeinsamen Erwartungen die Teilnehmerstaaten im Hinblick auf verantwortungsvolles unternehmerisches Handeln an die Unternehmen hegen, die auf ihrem Staatsgebiet tätig sind oder von dort aus operieren. Ausserdem dienen sie den Unternehmen und anderen Akteuren als eine massgebliche Orientierungshilfe. Den Unternehmen wird darin empfohlen, risikoabhängige Due-Diligence-Prüfungen durchzuführen, um tatsächliche und potenzielle negative Auswirkungen auf die in den Leitsätzen behandelten Themen zu ermitteln, zu verhüten und zu mindern sowie Rechenschaft darüber abzulegen, wie sie diesen Effekten begegnen. In dieser Hinsicht ergänzen und stärken die Leitsätze private und öffentliche Initiativen zur Definition und Umsetzung von Massstäben für ein verantwortungsvolles unternehmerisches Handeln.
>
> 5. Die Leitsätze legen auf dem Prinzip der Freiwilligkeit beruhende Grundsätze und Massstäbe für ein verantwortungsvolles und dem geltenden Recht und international anerkannten Normen entsprechendes unternehmerisches Handeln fest. [...]»[8]

[11] Das von den Leitsätzen erwartete verantwortungsvolle Handeln bezieht sich auf *alle* zentralen Themenbereiche, in denen Unternehmen mit der Gesellschaft in Berührung kommen: Die elf Kapitel der Leitsätze befassen sich neben Grundlagen und allgemeinen Grundsätzen[9] mit den

6 OECD, OECD-Leitsätze für multinationale Unternehmen zu verantwortungsvollem unternehmerischem Handeln, OECD Publishing, Paris 2023 (https://doi.org/10.1787/abd4d37b-de).

7 OECD, Erklärung über internationale Investitionen und multinationale Unternehmen vom 21. Juni 1976, zuletzt aktualisiert am 27. Oktober 2023 (https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0144).

8 OECD-Leitsätze (Fn. 6), Einführung, Ziff. 4 und 5.

9 OECD-Leitsätze (Fn. 6), Kapitel I Begriffe und Grundlagen, Kapitel II Allgemeine Grundsätze, Kapitel III Offenlegung von Informationen.

inhaltlichen Themenbereichen Menschenrechte[10], Arbeit[11], Umwelt[12], Korruption[13], Konsumentenschutz[14], Technologie und Digitalisierung[15], Wettbewerb[16] und Steuern[17].

[12] Kernstück der Leitsätze ist die Sorgfaltsprüfung. Sie verlangt seit 2011 von Unternehmen, das für die Abschätzung wirtschaftlicher, finanzieller oder rechtlicher Risiken bekannte Modell der «due diligence» auch auf Risiken für Mensch, Umwelt und Gesellschaft anzuwenden. Dabei geht es im Kern darum, die Auswirkungen der unternehmerischen Aktivitäten, d.h. die möglichen Risiken für Mensch und Umwelt, zu identifizieren, sie einzuschätzen – welche Menschenrechte sind betroffen, wie gravierend sind die Auswirkungen? – und dann gezielte Massnahmen zu treffen, um sie wenn möglich zu vermeiden oder, wenn das nicht möglich ist, zu mildern. Zur Sorgfaltsprüfung gehört auch ein Bericht über die getroffenen Vorkehrungen sowie unter bestimmten Voraussetzungen das Bereitstellen oder die Teilnahme an einem Wiedergutmachungsmechanismus. Die Kapitel IX bis XI zu Technologie, Wettbewerb und Besteuerung wurden allerdings 2011 nicht von der Sorgfaltsprüfung erfasst. Um Unternehmen die Umsetzung der erwarteten Sorgfaltsprüfung zu erleichtern, wurden ein allgemeiner sowie verschiedene themenspezifische Leitfäden, welche die einzelnen Schritte im Detail und erläutern und mit Beispielen verdeutlichen, erarbeitet.[18]

[13] Schliesslich verpflichtet ein Beschluss des Rates der OECD alle Beitrittsstaaten *verbindlich*, eine sog. Nationale Kontaktstelle einzurichten.[19] Diese Stellen haben die Aufgabe, die Umsetzung der Leitsätze zu fördern und Betroffenen ein wirksames Abhilfeverfahren («remedy») bei möglichen Verstössen durch Unternehmen zur Verfügung zu stellen. Es handelt sich dabei um aussergerichtliche, lösungsorientierte und vorwärts gerichtete Foren für einen Dialog.

[14] Da sich das Verständnis, was unter unternehmerischer Verantwortung zu verstehen ist, im Verlauf der Zeit verändert hat, wurden die Leitsätze in regelmässigen Intervallen angepasst, zuletzt 2023. Neben den 38 Mitgliedstaaten der OECD sind ihnen 13 weitere Staaten beigetreten.[20]

2.2. Aktualisierung der Leitsätze 2023: Relevanz für künstliche Intelligenz

[15] Hintergrund der Aktualisierung von 2023 waren die seit der letzten Revision 2011 stark veränderten Rahmenbedingungen. Dazu zählten u.a. Entwicklungen im Technologiebereich inkl. künstliche Intelligenz und die Herausforderungen der digitalen und technologischen Transfor-

10 OECD-Leitsätze (Fn. 6), Kapitel IV Menschenrechte.

11 OECD-Leitsätze (Fn. 6), Kapitel V Beschäftigung und Beziehungen zwischen den Sozialpartnern.

12 OECD-Leitsätze (Fn. 6), Kapitel VI Umwelt.

13 OECD-Leitsätze (Fn. 6), Kapitel VII Bekämpfung von Bestechung und sonstigen Korruptionsformen.

14 OECD-Leitsätze (Fn. 6), Kapitel VIII Verbraucherinteressen.

15 OECD-Leitsätze (Fn. 6), Kapitel IX Wissenschaft, Technologie und Innovation.

16 OECD-Leitsätze (Fn. 6), Kapitel X Wettbewerb.

17 OECD-Leitsätze (Fn. 6), Kapitel XI Besteuerung.

18 OECD, OECD-Leitfaden für die Erfüllung der Sorgfaltspflicht für verantwortungsvolles unternehmerisches Handeln, Paris 2018 sowie weitere themenspezifische Leitfäden: https://mneguidelines.oecd.org/due-diligence-guidance-for-responsible-business-conduct.htm.

19 OECD, Beschluss des Rates zu den Leitsätzen für multinationale Unternehmen zu verantwortungsvollem unternehmerischem Handeln, Ziff. I.1. Für die Schweiz ist dies der beim SECO angesiedelte Nationale Kontaktpunkt (NKP).

20 Die aktuelle Liste findet sich auf https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0144#adherents.

mation. Die Aktualisierung betraf alle Kapitel der Leitsätze, besonders signifikante Änderungen finden sich in Kapitel IX, das in vielen Teilen überholt war.

[16] Mit der Aktualisierung ist die Sorgfaltsprüfung nun auch auf das Kapitel IX. Wissenschaft, Technologie und Innovation anwendbar und damit auch auf digitale Technologien, Dienstleistungen und Produkte, inkl. KI. Von Unternehmen wird erwartet, dass sie auch im Bereich Technologie und Digitalisierung eine Sorgfaltsprüfung bei der Entwicklung, Finanzierung, dem Verkauf, der Lizenzierung, dem Handel und der Benutzung von Technologien, inkl. dem Sammeln und Nutzen von Daten durchführen. Darin eingeschlossen ist auch die Nutzung von Produkten und Dienstleistungen, d.h. die nachgelagerte Wertschöpfungskette. Es ist zu erwarten, dass in Zukunft vermehrt Verfahren vor Nationalen Kontaktstellen eingeleitet werden, welche Probleme im Zusammenhang mit der Digitalisierung etwa beim Einsatz von KI-basierten Systemen thematisieren.

3. Weiterführung: Aktualisierung der OECD-Empfehlung zu KI 2023/ 2024

3.1. Aktualisierung der Definition von KI-Systemen

[17] Es bedarf keiner langen Ausführungen, um zu erläutern, dass die OECD-Empfehlung zu KI angesichts der rasanten technologischen Entwicklung, insbes. im Bereich generativer KI-Systeme, nach vier Jahren einer Überarbeitung bedurfte. Ein erster Schritt auf diesem Weg wurde 2023 mit einer vom Rat der OECD verabschiedeten Aktualisierung der Definition «KI-System» unternommen.

Vergleich der Definitionen von KI-Systemen in der OECD-Empfehlung 2019 und 2024

Empfehlung 2019	Empfehlung 2024
Ein KI-System ist ein maschinenbasiertes System, das für bestimmte von Menschen definierte Ziele Voraussagen machen, Empfehlungen abgeben oder Entscheidungen treffen kann, die das reale oder virtuelle Umfeld beeinflussen. KI-Systeme können mit einem unterschiedlichen Grad an Autonomie ausgestattet sein.	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.

[18] Mit der neuen Definition sollten die Ziele eines KI-Systems präzisiert und klargestellt werden, dass diese impliziter oder expliziter Natur sein können. Insbesondere ging es darum,[21]

- die Rolle des von Menschen oder Maschinen bereitgestellten Inputs zu unterstreichen;

[21] OECD, Meeting of the Council at Ministerial Level, 2–3 May 2024, Report on the implementation of the OECD Recommendation on artificial intelligence, C/MIN(2024)17, 24 April 2024, Ziff. 14. https://one.oecd.org/document/C/MIN(2024)17/en/pdf.

- 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 eine mit anderen internationalen Initiativen kohärente Terminologie zu verwenden, und
- zu berücksichtigen, dass es KI-Systeme gibt, die sich auch nach Design und Bereitstellung noch weiterentwickeln können.

3.2. Weiterentwicklung der KI-Prinzipien in der Empfehlung 2024

[19] Der bei Empfehlungen des Rates regelmässig zu erstattende Bericht über die Umsetzung, Verbreitung und Relevanz der Empfehlung[22] kam 2024 zum Schluss, dass die Empfehlung von 2019 sich zu einem wichtigen internationalen Referenzwerk entwickelt hatte, das sich nicht zuletzt bei der Gestaltung von nationalen KI-Politiken als nützlich erwies. Der Bericht zeigte aber auch, dass in verschiedenen Bereichen Anpassungsbedarf bestand, um zum einen die technologische Entwicklung (generative KI) zu integrieren, zum andern die Umsetzung zu erleichtern und neuen regulatorischen Entwicklungen Rechnung zu tragen.

[20] 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, so dass 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 die 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.[23]

22 Siehe vorne Fn. 5.

23 OECD, Recommendation on Artificial Intelligence, vom 22. Mai 2019, aktualisiert am 3. Mai 2024, (https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449), Empfehlung des Rates zu künstlicher Intelligenz, inoffzielle Übersetzung (https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449#translations), Hintergrundinformationen, S. 4.

[21] Die Gliederung der Empfehlung blieb unverändert, die zehn Prinzipien wurden aber inhaltlich aktualisiert. Der erste Abschnitt *Grundsätze einer verantwortungsvollen Steuerung vertrauenswürdiger KI* folgt weiterhin dem Konzept der bereits 2019 entwickelten Prinzipien, legt aber im Sinne des Berichts von 2024 die Akzente stärker auf die Rolle des Menschen, die Missbrauchsbekämpfung und eine erweiterte Definition des Lebenszyklus von KI:

> (1) *Inklusives Wachstum, nachhaltige Entwicklung und Lebensqualität* als Ziele einer vertrauenswürdigen KI: Bei diesem Prinzip wurde das Ziel der Förderung der ökologischen Nachhaltigkeit hinzugefügt.
>
> (2) *Rechtsstaatlichkeit, Menschenrechte und demokratische Werte, insbesondere Fairness und Schutz der Privatsphäre* als von KI-Akteuren einzuhaltende Prinzipien während des ganzen Lebenszyklus: Menschenzentrierte Werte werden stärker betont. Unter Wahrung des Rechts auf freie Meinungsäusserung und anderer durch das geltende Völkerrecht geschützter Rechte und Freiheiten sind Anstrengungen gegen die Verbreitung von Falsch- und Desinformation zu intensivieren. Beschränkte sich die Empfehlung von 2019 noch auf dem Kontext angemessene und dem neuesten Stand der Technik entsprechende Mechanismen und Schutzmassnahmen, ohne die möglichen Risiken zu benennen, werden nun beispielhaft die Risiken, die sich aus Zweckentfremdung, Missbrauch oder unbeabsichtigtem Fehlgebrauch ergeben, erwähnt.
>
> (3) *Transparenz und Erklärbarkeit*: Angesichts des rapiden technologischen Fortschritts und der damit verbundenen Komplexität wurden die Anforderungen an Transparenz und Erklärbarkeit ergänzt. Neu müssen Informationen aussagekräftig sein und insbesondere auch das Verständnis der Möglichkeiten und Grenzen von KI-Systemen fördern. Da KI immer weitere Lebensbereiche durchdringt, wird näher ausgeführt, wie Menschen, die mit einem KI-System zu tun haben, zu informieren sind.
>
> (4) *Robustheit und Sicherheit* von KI-Systemen: Diese Bestimmungen wurden wesentlich überarbeitet. Neu werden Sicherheitsmechanismen verlangt, die gegebenenfalls gewährleisten, dass KI-Systeme gefahrlos ausser Kraft gesetzt, repariert und/oder stillgelegt werden können. Neu aufgenommen ist zudem die Informationsintegrität, welche durch Mechanismen, soweit technisch machbar und unter Wahrung der Meinungsfreiheit gestärkt werden soll.
>
> (5) *Rechenschaftspflicht:* Dieses Prinzip wurde gegenüber der Formulierung 2019, die eine Rechenschaftspflicht von KI-Akteuren ohne weitere Details verlangte, stark erweitert, indem die 2019 noch unter dem Titel Robustheit und Sicherheit aufgeführten Prinzipien der Rückverfolgbarkeit und des Risikomanagements nun als Elemente der Rechenschaftspflicht erfasst werden. Neu wird explizit auf verantwortungsvolles unternehmerisches Handeln referenziert und damit die Verbindung zu den OECD-Leitsätzen für multinationale Unternehmen hergestellt.

[22] Der zweite Abschnitt widmet sich wie 2019 *einzelstaatlichen Massnahmen und der internationalen Zusammenarbeit für vertrauenswürdige KI*.

> (1) Wenige präzisierende Änderungen finden sich in der Empfehlung an die Staaten, in *KI-Forschung und -Entwicklung* zu investieren.
>
> (2) Während die Empfehlung 2019 noch zur Förderung eines digitalen Ökosystems für KI aufrief, sind die Staaten neu gehalten, *ein inklusives und KI-freundliches Öko-*

system zu fördern. Die Elemente eines solchen Ökosystems werden in Einklang mit internationalen Entwicklungen und Arbeiten der OECD näher ausgeführt.

(3) Aus dem Appell von 2019, ein für KI günstiges Politikumfeld zu schaffen, wird neu die Empfehlung, *Governance und Politikumfeld KI-freundlich und interoperabel zu gestalten*. Neu wird zudem angeregt, die Interoperabilität der Governance und des Politikumfelds durch Zusammenarbeit auf inner- und zwischenstaatlicher Ebene zu fördern und damit einer zunehmenden Fragmentierung entgegenzuwirken. Diese Empfehlung deckt sich mit der *Empfehlung des Rates zur Rolle des Staates bei der Förderung verantwortungsvollen unternehmerischen Handelns*, welche die Kohärenz von Politiken und Regulierungen auf diesem Gebiet und insbesondere im Bereich der Sorgfaltsprüfung fördern will.[24]

(4) Mit der Betonung des menschenzentrierten Ansatzes kommt dem vierten Prinzip *Die Kompetenzen der Menschen stärken und Vorkehrungen für Umbrüche am Arbeitsplatz treffen* besondere Bedeutung zu. Mit verschiedenen Präzisierungen wird u.a. sichergestellt, dass berufliche Massnahmen für Mitarbeitende aller Altersstufen zu treffen sind.

(5) Weitgehend redaktionelle Änderungen wurden bei der Empfehlung, *die internationale Zusammenarbeit für vertrauenswürdige KI zu pflegen*, angebracht.

[23] Anlässlich der Verabschiedung der überarbeiteten Empfehlung zu KI war den beitretenden Staaten – den 38 OECD Mitgliedstaaten sowie acht weiteren Staaten[25] und der Europäischen Union klar, dass dies nicht den Abschluss der Arbeiten für eine vertrauenswürdige KI bilden würde, sondern dass weitere Massnahmen, insbes. im Hinblick auf die praktische Umsetzung, folgen müssten. Entsprechend wurde dem zuständigen Ausschuss in der OECD von den Ministern der Mitgliedstaaten der Auftrag erteilt, die Arbeiten fortzusetzen, dies aber nicht isoliert zu tun, sondern die Initiativen anderer internationaler Gremien einzubeziehen.[26] Zudem sollten praktische Leitlinien zur Umsetzung der Empfehlung erarbeitet resp. bestehende aktualisiert werden.[27] Angesichts der zunehmenden Fragmentierung von KI-Standards sind diese beiden Punkte für eine wirkungsvolle Steuerung von KI zentral.

4. Ausblick: Konsolidierung der Standards für KI-Risikomanagement?

4.1. Zunehmende Proliferation und Fragmentierung

[24] Wie die Beiträge in diesem Jusletter IT zeigen, haben sich in jüngster Zeit verschiedene Foren mit dem Management der mit KI verbundenen Risiken befasst. Die Vielzahl der daraus resultierenden, oft nicht koordinierten Ansätze zum Risikomanagement in Form von nationalen und internationalen Regulierungen, Standards und Initiativen erschwert die Umsetzung durch die

24 OECD Recommendation on the role of government in promoting responsible business conduct vom 12. Dezember 2022 (https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0486), inoffizielle deutsche Übersetzung (https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0486#translations).

25 Ägypten, Argentinien, Brasilien, Malta, Peru, Rumänien, Singapur, Ukraine.

26 OECD Empfehlung zu KI (Fn. 23), Ziff. VIII.a.

27 OECD Empfehlung zu KI (Fn. 23), Ziff. VIII.b.

relevanten Akteure. Als Beispiele genannt seien in Ergänzung zu der bereits in diesem Jusletter besprochenen Konvention des Europarats[28] und der KI-Verordnung (AI Act) der EU[29] der *G7 Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems*, der als freiwilliges Instrument die G7-Prinzipien zu KI ergänzt,[30] der *AI Risk Management Framework* des US National Institute of Standards and Technology,[31] eine Reihe nationaler Vorstösse, darunter der neue kanadische *Artificial Intelligence and Data Act*[32] sowie als Industriestandard die *ISO Artificial Intelligence Guidance on Risk Management*.[33]

[25] Vor diesem Hintergrund und in Umsetzung des Auftrags zur Umsetzung der OECD-Empfehlung zu KI, lancierte die OECD ein Projekt, um die von Unternehmen erwartete Sorgfaltsprüfung bzgl. KI so zu konkretisieren, dass sie in der Praxis angewandt werden können. Der noch laufende Prozess besteht aus vier Phasen:[34]

[26] In *Phase 1* geht es um die Erhebung der führenden relevanten Standards und Rahmenwerke zum Risikomanagement von KI sowie der relevanten Akteure und zentralen Risiken *(Mapping)*. Diese Phase wurde bereits abgeschlossen und zeigt, dass sich die verwendeten Ansätze zwar in begrifflichen Details und teilweise der Reichweite unterscheiden, insgesamt aber weitgehend gleiche Zielsetzungen verfolgen und in den verwendeten Grundkonzepten zur Risikoanalyse kompatibel sind. Die grössten Unterschiede wurden bei der Lenkungsfunktion von Rahmenwerken (*governance*) geortet.[35]

[27] Daran schliesst mit *Phase 2* die Identifizierung allfälliger Lücken bei der verwendeten Terminologie und den anwendbaren Konzepten in diesen Instrumenten, welche die Umsetzung vertrauenswürdiger KI in der Praxis erschweren könnten, an (*Gap analysis*). Zentral ist dabei ein gemeinsames Verständnis der KI-Wertschöpfungskette.[36]

[28] Eine Orientierung am Konzept der Sorgfaltsprüfung, wie sie in den OECD-Leitsätzen enthalten ist, ermöglicht es dabei, die relevanten Akteure entlang der KI-Wertschöpfungskette in drei Gruppen zu kategorisieren und damit deren systematische Erfassung im Risikomanagement zu erleichtern: Eine erste Gruppe könnte Akteure, die KI-Wissen und KI-Ressourcen generieren, umfassen, eine zweite Akteure, die aktiv in die Entwicklung und Verbreitung von KI-Systemen involviert sind, und die dritte die Benutzer von KI-Systemen.

[29] *Phase 3*, die gegenwärtig in Gang ist, besteht in der «Übersetzung» der Resultate aus *Phase 1* und *2* und deren Nutzbarmachung für die Praxis. Zu diesem Zweck soll ein Leitfaden zu verantwortungsvollem unternehmerischem Handeln im Bereich KI erarbeitet werden. Ausgangspunkt

28 David Marti, Die KI-Konvention des Europarats: Ursprung, Inhalt, Ausblick, in diesem Jusletter IT.

29 Martina Arioli, Risikomanagement nach der EU-Verordnung über Künstliche Intelligenz, in diesem Jusletter IT.

30 G7 Action Plan for promoting global interoperability between tools for trustworthy AI, 2023, https://www.meti.go.jp/press/2023/04/20230430001/20230430001-ANNEX5.pdf.

31 National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (ARI RMF 1.0), Januar 2023, https://doi.org/10.6028/NIST.AI.100-1.

32 Verabschiedet am 24. April 2024, https://www.parl.ca/legisinfo/en/bill/44-1/c-27.

33 International Organization for Standardization (ISO), Information technology – Artificial intelligence – Guidance on risk management (ISO/IEC 23894), Februar 2023.

34 OECD, Common guideposts to promote interoperability in AI risk management, OECD Artificial Intelligence Papers, No. 5, November 2023, S. 14. (https://doi.org/10.1787/ba602d18-en).

35 OECD, Common guideposts (Fn. 34), S. 15.

36 OECD, Common guideposts (Fn. 34), S. 35.

ist das aus den Leitsätzen für Multinationale Unternehmen bekannte und vielen Unternehmen weltweit bereits vertraute Konzept der Sorgfaltsprüfung.[37]

4.2. Ausblick: Konsolidierung in Sicht?

[30] Indem der zu entwickelnde Leitfaden auf einem breit abgestützten und etablierten Konzept der Sorgfaltsprüfung zu Risiken für Mensch, Umwelt und Planet aufbaut, erleichtert er die Rezeption in der Praxis. Der Ansatz, nicht neue Standards, sondern mit in der Praxis erhobenen Beispielen für *best practices* eine Anleitung zur kohärenten Umsetzung bestehender Instrumente zu geben, hat sich auf anderen Gebieten wie etwa dem Rohstoffsektor[38] oder im Finanzbereich[39] bewährt. Die Arbeiten zeigen zum einen, dass das Bedürfnis der Praxis, der zunehmenden Fragmentierung entgegenzuwirken von den Mitgliedstaaten ernst genommen wird. Zum andern bietet die Form eines praxisorientierten Leitfadens, der nicht den Anspruch erhebt, einen neuen Standard zu setzen, die Möglichkeit, für unterschiedliche regulatorische Ansätze und Ambitionen akzeptabel zu sein. Es lohnt sich deshalb, die Arbeiten in der OECD weiterzuverfolgen, nicht zuletzt im Hinblick auf aktuelle und bevorstehende Diskussionen in der Schweiz zur Regulierung von KI und der Sorgfaltspflicht von Unternehmen.

Prof. Dr. Christine Kaufmann, Professorin für Öffentliches Recht, Völker- und Europarecht an der Universität Zürich und Vorsitzende der OECD Working Party on Responsible Business Conduct. Die Autorin dankt Rashad Abelson, OECD Centre for Responsible Business Conduct, für wertvolle Informationen bei der Erarbeitung dieses Beitrags. Der Beitrag reflektiert ausschliesslich die persönliche Ansicht der Autorin.

Alle zitierten Internetseiten sind am 20. Juni 2024 besucht worden.

37 Vorne Rz. 12.

38 OECD-Leitfaden für die Erfüllung der Sorgfaltspflicht zur Förderung verantwortungsvoller Lieferketten für Minerale aus Konflikt- und Hochrisikogebieten, 3. Ausgabe 2019 (in mehreren Sprachen verfügbar https://mneguidelines.oecd.org/mining.htm).

39 OECD-Leitfäden zur Erfüllung der Sorgfaltspflicht für institutionelle Investoren (2017) und für das Kredit- und Emissionsgeschäft (2019) (jeweils in mehreren Sprachen); Responsible Business Conduct Due Diligence for Project and Asset Finance Transactions (2022); Managing Climate Risks and Impacts Through Due Diligence for Responsible Business Conduct – A Tool for Institutional Investors (https://mneguidelines.oecd.org/rbc-financial-sector.htm).