www.jusletter-it.eu

Martina Arioli

# Risikomanagement nach der EU-Verordnung über Künstliche Intelligenz

Das Risikomanagementsystem ist eine der Hauptanforderungen für Hochrisiko-KI-Systeme nach der EU-Verordnung über Künstliche Intelligenz (KI-VO). Die KI-VO folgt dem EU-Konzept des New Legislative Framework (NLF) für die Produktsicherheit. Dabei werden grundlegende Anforderungen – wie insbesondere Anforderungen an ein KI-Risikomanagementsystem – in der KI-VO festgelegt, während die technische Umsetzung dieser Anforderungen durch «harmonisierte Normen» gewährleistet werden soll. Der vorliegender Beitrag soll einen Überblick über die Anforderungen an das Risikomanagementsystem nach der KI-VO sowie über bestehende Normen zum Risikomanagementsystem für KI verschaffen.

Beitragsart: Beiträge

Zitiervorschlag: Martina Arioli, Risikomanagement nach der EU-Verordnung über Künstliche Intelligenz, in: Jusletter IT 4. Juli 2024

## Inhaltsübersicht



## 1. Einführung

[1] Künstliche-Intelligenz-Systeme (nachfolgend «KI-Systeme») und die Kontexte ihres Einsatzes sind häufig komplex, was das Erkennen von und die Reaktion auf Fehler erschwert. Vorteile, aber auch Risiken von KI-Systemen resultieren aus dem Zusammenspiel technischer Aspekte und gesellschaftlicher Faktoren. Dies umfasst die Art und Weise, wie ein KI-System genutzt wird, auf welchen Daten das System trainiert und weiterentwickelt wird, seine Interaktionen mit anderen KI-Systemen, die Verantwortlichkeiten der Akteure und den gesellschaftlichen Kontext, in dem das KI-System eingesetzt wird. KI-Systeme sind inhärent sozio-technisch, sie beschlagen gleichzeitig soziale Komponenten (Menschen und deren kulturelles Umfeld und gesellschaftliche Dynamik) wie auch technische Aspekte.

[2] Risiken im Zusammenhang mit KI können sich von traditionellen Software-Risiken unterscheiden oder diese verstärken. KI-Systeme arbeiten auf einer weitaus komplexeren Ebene als andere Technologien, was zu einer grösseren Anzahl von Risikoquellen führt. So bilden die Daten, die für die Entwicklung eines KI-Systems verwendet werden, den Kontext oder den Verwendungszweck des KI-Systems möglicherweise nicht wahrheitsgetreu oder nicht angemessen ab. Schädliche Verzerrungen und andere Probleme mit der Datenqualität können die Vertrauenswürdigkeit und Verlässlichkeit von KI-Systemen beeinträchtigen. Datensätze, die zum Training von KI-Systemen verwendet werden, können aus ihrem ursprünglichen und beabsichtigten Kontext herausgelöst werden oder im Vergleich zum Einsatzkontext veraltet sein. Beabsichtigte oder unbeabsichtigte Änderungen während des Trainings des KI-Systems können die Leistung von KI-Systemen grundlegend verändern. Zudem kann die statistische Unsicherheit bei der Verwendung vortrainierter Modelle erhöht sein und Probleme mit dem Bias-Management, der wissenschaftlichen Validität und der Reproduzierbarkeit verursachen. Ferner sind KI-Systeme gegenüber herkömmlicher Software häufig allein aufgrund der Vielzahl von Entscheidungspunkten weitaus komplexer. Auch dürfte die regelmässige Durchführung von Tests sowie die Entscheidung, was überhaupt getestet werden soll, herausfordernd sein.[1]

[3] Ohne angemessene Kontrollen können KI-Systeme ungerechte oder unerwünschte Ergebnisse für Individuen und Personengruppen generieren, verstärken, perpetuieren oder verschlimmern.[2]

---

1 Vgl. NIST AI RMF Version 1.0, Januar 2023, https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10 und https://doi.org/10.6028/NIST.AI.100-1, S. 38.

2 «Without Research and Development breakthroughs, even well-meaning developers may inadvertently create AI systems that pursue unintended goals: The reward signal used to train AI systems usually fails to fully capture the intended objectives, leading to AI systems that pursue the literal specification rather than the intended outcome. Additionally, the training data never captures all relevant situations, leading to AI systems that pursue undesirable

Hingegen können adäquate Kontrollmechanismen dazu beitragen, solche ungerechten Ergebnisse zumindest zu mildern und zu steuern. Ein umfassendes Verständnis und ein effektiver Umgang mit den Risiken von KI-Systemen sind entscheidend, um deren Vertrauenswürdigkeit («*trustworthiness*») zu erhöhen. Ein geeignetes Risikomanagementsystem, d.h. die koordinierten Aktivitäten zur Identifikation, Analyse Bewertung, Steuerung und Kontrolle einer Organisation in Bezug auf Risiken[3], kann hierzu einen Beitrag leisten.

[4] Im Folgenden werden die Anforderungen an das Risikomanagementsystem von Hochrisiko-KI-Systemen gemäss Art. 9 KI-VO untersucht. Das Risikomanagement bildet nur eine von vielen Anforderungen, die die KI-VO an die Entwicklung und Nutzung von Hochrisiko-KI-Systemen stellt. Sie gehört jedoch zu den zentralen Anforderungen und muss deshalb mit der gebotenen Umsicht angegangen werden.

[5] «KI-Modelle mit allgemeinem Verwendungszweck und systemischem Risiko» («*General Purpose AI with systemic risk*») gemäss Art. 51 KI-VO können obgenannte KI-Risiken verschärfen und spezifische Risiken erzeugen, wie beispielsweise Konfabulation (gemeinhin als «Halluzination» bezeichnet), Toxizität, Verzerrung und Homogenisierung. Eine detaillierte Untersuchung dieser spezifischen Risiken im Zusammenhang mit «KI-Modellen mit allgemeinem Verwendungszweck» und deren Risikomanagement würde jedoch den Rahmen dieser Untersuchung sprengen.

## 2. KI-VO als Verordnung über die Produktsicherheit

[6] Zweck der KI-VO ist es, das Funktionieren des Binnenmarkts zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz zu fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf **Gesundheit**, **Sicherheit** und die in der Charta verankerten **Grundrechte**, einschliesslich Demokratie, Rechtsstaatlichkeit und Umweltschutz, vor schädlichen Auswirkungen von KI-Systemen in der EU zu gewährleisten und die Innovation zu unterstützen (Art. 1 Abs. 1 KI-VO).

[7] Als «KI-System» definiert Art. 3 (1) KI-VO ein **maschinengestütztes** System, das für einen in unterschiedlichem Grade **autonomen** Betrieb ausgelegt ist, das nach seiner Betriebsaufnahme **anpassungsfähig** sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele **ableitet** [«inference»], wie Ausgaben, wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen, erstellt werden, die physische oder virtuelle Umgebungen **beeinflussen** können.[4]

[8] Es sollen nur **sichere und anderweitig konforme Produkte auf den EU-Markt** gelangen. Deshalb sollen die Sicherheitsrisiken, die ein Produkt als Ganzes aufgrund seiner digitalen Kompo-

goals in new situations encountered after training.», Yoshua Bengio et al., «Managing extreme risk of AI in rapid progress», Science 10.1126/science.adn0117 (2024).

3 ISO 31000:2018.

4 Viele der derzeit eingesetzten Systeme erfüllen vermutlich nicht kumulativ alle Elemente der Definition gemäss Artikel 3 (1) KI-VO. Dies liegt daran, dass ihnen entweder die Anpassungsfähigkeit fehlt, sie nicht autonom arbeiten oder keine eigentliche Ableitung («inference») stattfindet. Selbst wenn diese Definitionselemente als graduelle Konzepte betrachtet werden, dürfte die «inference» das entscheidende Merkmal sein, das ein KI-System von einem konventionellen Algorithmus unterscheidet, denn eine reine Muster- oder Objekterkennung erfüllt die Bedingungen aus Art. 3 KI-VO vermutlich nicht. Die nächste Welle der KI-Entwicklung wird wahrscheinlich durch neuromorphe Systeme in Verbindung mit Quantencomputing geprägt sein, die neuronale Strukturen nicht nur simulieren, sondern tatsächlich aufbauen können.

nenten, einschliesslich KI-Systeme, mit sich bringen kann, angemessen vermieden und gemindert werden.[5] Die EU KI-VO ist in eine Vielzahl von EU-Regularien zur **Produktesicherheit und Martküberwachung**[6] eingebettet und folgt in Konzept, Logik, Prozessen und Terminologie dem New Legislative Framework («NLF»)[7] für Produktsicherheit. Das NLF wird im «Blue Guide on the implementation of EU product rules 2022»[8] der Europäischen Kommission erläutert und in der KI-VO in einer Vielzahl von Erwägungsgründen referenziert.[9] Gemäss NLF kann auf ein Produkt mehr als ein Rechtsakt der Harmonisierungsrechtsvorschriften der EU anwendbar sein. Dies bedeutet, dass ein Produkt nur dann bereitgestellt oder in Betrieb genommen werden kann, wenn es mit *allen* anwendbaren Harmonisierungsrechtsvorschriften der EU übereinstimmt.[10]

[9] Dem konzeptionellen Ansatz des NLF folgend sind in der KI-VO die Ziele und grundlegenden Anforderungen an die Produktsicherheit festgelegt, während Operationalisierung und Umsetzung dieser Anforderungen durch harmonisierte, freiwillige Normen gewährleistet werden soll. Ferner muss die Produktsicherheit nach den Grundprinzipien des NLF von allen Akteuren im Lebenszyklus eines Produkts sichergestellt werden. Dementsprechend hat die Europäische Kommission die europäischen Normungsorganisationen im Mai 2023 aufgefordert, europäische harmonisierte Normen zur Unterstützung der KI-VO zu entwickeln.[11] Diese Normen sollen als wesentliche Instrumente für die Umsetzung der KI-VO sowie für die Konformitätsvermutung dienen. Der Standardisation Request dürfte angesichts dessen, dass die KI-VO nunmehr in der definitiven Form vorliegt, nochmals überarbeitet werden.

[10] Damit setzt die Regulierung stark auf technische Lösungen, Standardisierungsprozesse und Selbsteinschätzungen der Akteure. Ob dieser Regulierungsansatz bei KI-Systemen, die wesentliche grundrechtliche Problemdimensionen mit sich bringen können, effektiven Schutz zu leisten vermag, wird sich weisen.[12] Die standardisierten Normen werden primär technischer und organisatorischer Natur sein. Sie werden weitgehend unter Ausschluss der Zivilgesellschaft erlassen und wohl ohne Einbezug von Expertise im Bereich des Grundrechtsschutzes. Wie oben ausgeführt, ist es jedoch nicht in erster Linie eine spezifische Technologie, die mit Risiken einhergeht, sondern die Verwendung dieser Technologie in einem bestimmten Kontext und zu einem bestimmten Zweck.[13]

---

5 Erw. 47 KI-VO.

6 Siehe insbesondere den Vorschlag für eine Verordnung des europäischen Parlaments und des Rates über die allgemeine Produktsicherheit, welche die bisherige Richtlinie 2001/95/EG des europäischen Parlaments und des Rates vom 3. Dezember 2001 über die allgemeine Produktsicherheit ersetzen soll, sowie Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates zur Anpassung der Vorschriften über ausservertragliche zivilrechtliche Haftung an künstliche Intelligenz (Richtlinie über KI-Haftung) COM/2022/496 final.

7 Regulation (EU) 2019/1020 of the European Parliament and of the Council Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products (New Legislative Framework).

8 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:C:2022:247:TOC.

9 So beispielsweise in Erw. 9, 46 und 64 KI-VO.

10 Vgl. Art. 6 Abs. 1 KI-VO, wonach die in Annex I KI-VO aufgezählten Regularien und die KI-VO sich gegenseitig nicht beinträchtigen sollen.

11 Commission Implementing Decision on a Standardisation Request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence («Standardisation Request»), Annex 1 und 2, 22. Mai 2023, https://ec.europa.eu/growth/tools-databases/enorm/mandate/593_en.

12 Angela Müller, «Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz», Zeitschrift für Europarecht (EuZ) 1/2022, S. 13.

13 Angela Müller (Fn. 12), S. 17.

## 3. Der risikobasierte Ansatz der EU KI-VO

[11] Die EU verfolgt mit der KI-VO einen risikobasierten Ansatz, indem KI-Systeme entsprechend dem Grad ihrer Risiken für Gesundheit, Sicherheit und Grundrechte, mit denen sie einhergehen, klassifiziert werden. Entlang von vier (4) Risikokategorien werden unterschiedliche Vorschriften und Anforderungen formuliert, die an die Intensität und den Umfang der von diesen Systemen ausgehenden Risiken angepasst sind. Die KI-VO definiert «Risiko» in Art. 3 (2) als die «Kombination von Wahrscheinlichkeit eines Schadenseintritts und Schwere des Schadens».[14]

[12] Bestimmte Praktiken im Bereich der KI werden als inakzeptabel eingestuft und sind daher gemäss Artikel 5 der KI-VO verboten.

[13] Im Gegensatz dazu sind Hochrisiko-KI-Systeme gemäss Artikel 6 KI-VO zwar zulässig, jedoch besonderen Anforderungen unterworfen, und den verschiedenen Akteuren werden in Bezug auf Hochrisiko-KI-Systeme sanktionierbare[15] Pflichten auferlegt. Zu den Hochrisiko-KI-Systemen gehören einerseits all jene Systeme, die ein von bestimmten anderen EU-Verordnungen erfasstes Produkt sind oder als Sicherheitskomponente in einem solchen Produkt eingesetzt werden, wenn dieses einer Konformitätsbewertung durch Dritte unterworfen ist (Art. 6 Abs. 1 KI-VO). Andererseits fallen bestimmte «*stand-alone*» KI-Systeme gemäss Art. 6 Abs. 2 KI-VO i.V.m. Anhang III KI-VO aus den Bereichen biometrische Identifizierung / Kategorisierung, kritische Infrastrukturen, Bildung, Arbeitsplatz, Zugang zu Dienstleistungen, Strafverfolgung, Migration und Asyl sowie Rechtspflege und demokratische Prozesse in die Kategorie der Hochrisiko KI-Systeme.[16] Als hochriskant sollten nur solche KI-Systeme eingestuft werden, die erhebliche schädliche Auswirkungen auf die Gesundheit, die Sicherheit und die Grundrechte von Personen in der EU haben (Art. 6 Abs. 3 KI-VO im Umkehrschluss, sowie Erw. 46 KI-VO). Schliesslich unterliegen KI-Systeme mit begrenzten Risiken, die mit natürlichen Personen interagieren, spezifischen Transparenzvorschriften.

[14] Keiner spezifischen Regulierung unterliegen sodann KI-Systeme mit keinem oder minimalen Risiko. Es handelt sich hier um die Kategorie der allgemeinen Anwendungen von KI wie z.B. Spamfilter oder KI-gestützte Rechtschreibprüfungen.

## 4. Das Risikomanagementsystem als Pflicht des Anbieters

[15] Anbieter von Hochrisiko-KI-Systemen gemäss Art. 6 KI-VO und Annex III KI-VO sind nach Art. 9 KI-VO verpflichtet, während des gesamten Lebenszyklus eines Hochrisiko-KI-Systems ein Risikomanagementsystem einzurichten, anzuwenden, zu dokumentieren und aufrechtzuerhalten (Art. 9 Abs. 1 KI-VO).[17]

---

14 Andere EU-Regularien verwenden entsprechend dem NLF dieselbe Definition, vgl. beispielsweise Art. 2 (23) EU-Verordnung 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte.

15 Geldbussen von bis zu 15 Mio. EUR oder – im Falle von Unternehmen – von bis zu 3% des gesamten weltweiten Jahresumsatzes des vorangegangenen Geschäftsjahres, vgl. Art. 99 Abs. 4 KI-VO.

16 Gemäss Art. 7 KI-VO kann die EU-Kommission die Liste in Anhang III gestützt auf Art. 7 KI-VO jederzeit mit einem delegierten Rechtsakt anpassen.

17 Man beachte die ähnliche Formulierung in Art. 10 Abs. 2 der Verordnung (EU) 2017/745 des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, welche das Risikomanagementsystem in Anhang I als eine der grundlegenden Sicherheits- und Leistungsanforderungen bezeichnet.

[16] Die Pflicht zur Implementierung eines Risikomanagementsystems ist eingebettet in weitere Pflichten des Anbieters, die bei Hochrisiko-KI-Systemen zwingend erfüllt werden müssen, so insbesondere die Implementierung eines kontinuierlichen Qualitätsmanagementsystems, die Anforderungen an die Qualität und die Relevanz der verwendeten Datensätze (Datengovernance, Art. 10 KI-VO), die technische Dokumentation (Art. 11 KI-VO), die Aufbewahrung von Aufzeichnungen (Aufzeichnungspflicht Art. 12 KI-VO), die Transparenz und die Bereitstellung von Informationen für die Betreiber (Art. 13 KI-VO), die menschliche Aufsicht (Art. 14 KI-VO) sowie die Genauigkeit, Robustheit und Cybersicherheit (Art. 15 KI-VO). Die Massnahmen, die die Anbieter zur Erfüllung der vorgenannten verbindlichen Anforderungen der KI-VO ergreifen, sollten dem **allgemein anerkannten Stand der Technik im Bereich der künstlichen Intelligenz** Rechnung tragen sowie **verhältnismässig** und **wirksam** sein, um die Ziele der KI-VO Verordnung zu erreichen (Art. 8 Abs. 1 KI-VO i.V.m. Erw. 64).[18]

[17] Ferner muss der Anbieter Pflichten im Zusammenhang mit dem Konformitätsbewertungsverfahren und der Konformitätserklärung, der Registrierungspflicht, den Korrekturmassnahmen und der Information der Betreiber sowie der nationalen zuständigen Behörde und der Anbringung der CE-Kennzeichnung einhalten.

[18] Als «Anbieter» gilt nach Art. 3 (3) KI-VO eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System oder ein KI-Modell für allgemeine Verwendungszwecke entwickelt oder entwickeln lässt und es unter ihrem eigenen Namen oder ihrer eigenen Marke in Verkehr bringt oder in Betrieb nimmt, unabhängig davon, ob dies entgeltlich oder unentgeltlich geschieht. Es ist zu beachten, dass Akteure wie insbesondere Betreiber, aber auch Importeure und Händler gemäss Art. 25 KI-VO in nachfolgenden Konstellationen als Anbieter betrachtet werden und dementsprechend den vorgenannten Pflichten für Anbieter nachkommen müssen. Eine solche Rollenänderung tritt auf, wenn:

- sie ein Hochrisiko-KI-System unter ihrem eigenen Namen oder Markenzeichen auf den Markt bringen oder in Betrieb nehmen,
- sie den vorgesehenen Zweck eines bereits auf dem Markt befindlichen oder in Betrieb genommenen Hochrisiko-KI-Systems ändern,
- sie eine wesentliche Änderung an dem Hochrisiko-KI-System vornehmen.

## 5. Das Risikomanagementsystem nach Art. 9 KI-VO

[19] Das Risikomanagementsystem versteht sich als ein kontinuierlicher iterativer Prozess, der während des gesamten Lebenszyklus eines Hochrisiko-KI-Systems geplant und durchgeführt wird und eine regelmässige systematische Überprüfung und Aktualisierung erfordert (Art. 9 Abs. 2 KI-VO und Erw. 65 KI-VO). Damit soll eine dynamische und kontinuierliche Überprüfung und Anpassung von Massnahmen zur Ermittlung und Minimierung von Restrisiken für **Gesundheit, Sicherheit und Grundrechte** und ein hohes Mass an **Vertrauenswürdigkeit** gewährleistet werden (Erw. 64 KI-VO). Folglich müssen nicht nur die behandelten Risiken kontinuierlich ak-

18 Erw. 64 KI-VO.

tualisiert werden, sondern auch das Risikomanagementsystem selbst. Das Risikomanagement ist daher eine dynamische Aufgabe.[19]

[20] Das Risikomanagementsystem bildet Bestandteil des **Qualitätsmanagementsystems** nach Art. 17 KI-VO sowie der **technischen Dokumentation** nach Annex IV. Das Risikomanagementsystem muss nachvollziehbar dokumentiert werden, sodass es von den Aufsichtsbehörden des betreffenden Mitgliedstaates überprüft werden kann. Die Dokumentation ist während 10 Jahren ab Inverkehrbringen aufzubewahren (Art. 18 KI-VO).

[21] Was ein Risikomanagementsystem ist, definiert die KI-VO nicht[20], es kann aber analog zur Definition des Qualitätsmanagementsystems nach Art. 17 KI-VO als «System», das systematisch und ordnungsgemäss in Form schriftlicher Regeln, Verfahren und Anweisungen dokumentiert wird, verstanden werden. Art. 9 KI-VO beschreibt das Risikomanagementsystem vielmehr als einen Prozess in vier Schritten:

a. die **Ermittlung und Analyse** der bekannten und vernünftigerweise vorhersehbaren Risiken, die vom Hochrisiko-KI-System für die Gesundheit, Sicherheit oder Grundrechte ausgehen können, wenn es entsprechend seiner Zweckbestimmung verwendet wird (Art. 9 Abs. 2 lit. a KI-VO);
b. die **Abschätzung und Bewertung** der Risiken, die entstehen können, wenn das Hochrisiko-KI-System entsprechend seiner Zweckbestimmung oder im Rahmen einer vernünftigerweise vorhersehbaren Fehlanwendung verwendet wird (Art. 9 Abs. 2 lit. b KI-VO);
c. die Bewertung anderer möglicherweise auftretender Risiken auf der Grundlage der Auswertung der Daten aus dem in Artikel 72 genannten System zur Beobachtung **nach dem Inverkehrbringen** («*post-market monitoring system*», Art. 9 Abs. 2 lit. c KI-VO i.V.m Art. 72 KI-VO);
d. die Ergreifung geeigneter und gezielter **Risikomanagementmassnahmen** zur Bewältigung der gemäss Art. 9 Abs. 2 lit. a KI-VO ermittelten Risiken (Art. 9 Abs. 2 lit. d KI-VO).

[22] Dass die KI-VO das Risikomanagementsystem als «iterativen Prozess» sieht, ist durchaus sinnvoll, denn es muss Ziel des Risikomanagements sein, dass bei jedem Schritt auf die Ergebnisse des vorherigen Schritts aufgebaut wird und eine schrittweise Annäherung an das Ziel der Risikoeliminierung oder -minimierung durch die wiederholte Anpassung und Verfeinerung erreicht wird.

[23] Die zu identifizierenden Risiken sind auf bekannte und vernünftigerweise vorhersehbare Risiken beschränkt, wobei aber die KI-VO keine Konkretisierung dafür bietet, was als «bekannt» und «vernünftigerweise vorhersehbar» gelten soll. Dies kann zu einem grossen Aufwand führen, bleibt doch unklar, ob und ab wann aufgehört werden darf, nach neuen Risiken zu suchen. Auf der anderen Seite sollen Anbieter den Schutzzweck von Art. 9 nicht mit dem Argument unterlaufen können, das Risiko sei nicht vorhersehbar gewesen.[21]

---

19 Nadja Braun Binder/Catherine Egli, «Kommentar zu Art. 9 KI-VO», in: KI-VO, Künstliche Intelligenz-Verordnung, Kommentar, Hrsg. Mario Martini/Christiane Wendehorst, Verlag C.H.Beck 2024, N 19 zu Art. 9, im Zeitpunkt der Publikation des vorliegenden Beitrags noch nicht veröffentlicht.

20 Das Risikomanagementsystem kann analog der Definition des Qualitätsmanagementsystems nach Art. 17 KI-VO als «System», das systematisch und ordnungsgemäss in Form schriftlicher Regeln, Verfahren und Anweisungen dokumentiert wird, definiert werden.

21 Vgl. Braun Binder/Egli (Fn. 19), N 23 zu Art. 9.

[24] Bei der Identifikation der «vernünftigerweise vorhersehbaren Fehlanwendung» von Hochrisiko-KI-Systemen gemäss Art. 9 Abs. 2 lit. b KI-VO sollte der Anbieter die Verwendungen von KI-Systemen erfassen, die zwar nicht unmittelbar der Zweckbestimmung entsprechen und in seiner Betriebsanleitung vorgesehen sind[22], bei denen jedoch nach vernünftigem Ermessen davon auszugehen ist, dass sie sich aus einem leicht absehbaren menschlichen Verhalten im Zusammenhang mit den spezifischen Merkmalen und der Verwendung eines bestimmten KI-Systems oder einer vernünftigerweise vorhersehbaren Interaktion mit anderen Systemen, auch anderen KI-Systemen[23], ergeben. Wenn aber das System nicht wie vorgesehen verwendet oder gar in unvorhersehbarer Weise missbraucht wird, müssen diese Risiken bei der Risikoermittlung nach Art. 9 Abs. 2 lit. b KI-VO nicht berücksichtigt werden. Damit soll sichergestellt werden, dass der Anbieter nur für Risiken verantwortlich ist, die er auch kontrollieren kann, was die Rechtssicherheit für den Anbieter erhöht. Unklar bleibt aber in diesem Zusammenhang Art. 9 Abs. 3 KI-VO, wonach im Rahmen des Risikomanagements nur solche Risiken gemeint sein sollen, die durch die Entwicklung oder Konzeption des Hochrisiko-KI-Systems oder durch die Bereitstellung ausreichender technischer Informationen angemessen gemindert oder behoben werden können.

[25] Es ist einem Anbieter zu empfehlen, bereits früh im Entwicklungsprozess das Risikomanagementsystem zu konzipieren und dabei potenzielle Nutzerinnen und Nutzer, beabsichtigte Verwendungen und vernünftigerweise vorhersehbare Missbräuche zu identifizieren.[24] Alle bekannten oder vorhersehbaren Umstände bezüglich der Verwendung des Hochrisiko-KI-Systems im Einklang mit seiner Zweckbestimmung oder einer vernünftigerweise vorhersehbaren Fehlanwendung, die zu Risiken für die Gesundheit und Sicherheit oder die Grundrechte führen können, sollten vom Anbieter in der Betriebsanleitung aufgeführt werden.[25]

[26] Die **Risikobewertung** kann zwischen KI-Systemen, die für den direkten menschlichen Kontakt entwickelt oder eingesetzt werden, und solchen, die dies nicht tun, unterschiedlich ausfallen. Eine höhere anfängliche Priorisierung kann in Umgebungen erforderlich sein, in denen das KI-System auf grossen Datensätzen trainiert wird, die sensible oder geschützte Daten wie personenbezogene Informationen enthalten, oder in denen die Ergebnisse der KI-Systeme direkte oder indirekte Auswirkungen auf Menschen haben. KI-Systeme, die ausschliesslich für die Interaktion mit Computersystemen konzipiert sind und auf nicht-sensiblen Datensätzen (z.B. Daten, die aus der physischen Umgebung gesammelt wurden) trainiert werden, erfordern möglicherweise eine geringere anfängliche Bewertung. Dennoch bleiben die regelmässige Bewertung und Priorisierung von Risiken basierend auf dem Kontext wichtig, da auch nicht-menschlich-interagierende KI-Systeme nachgelagerte Sicherheits- oder soziale Auswirkungen haben können.[26]

[27] Im Rahmen des Risikomanagementsystems muss der Anbieter sodann die geeigneten und gezielten **Risikomanagementmassnahmen** ergreifen, wobei die Auswirkungen und möglichen Wechselwirkungen, die sich aus der kombinierten Anwendung der Anforderungen der Artikel 8 bis 15 (Abschnitt 2) KI-VO ergeben, gebührend berücksichtigt werden müssen, um die Risiken

22 Vgl. Definition von «Zweckbestimmung» in Art. 3 (12) KI-VO.

23 Vgl. Definition der «vernünftigerweise vorhersehbaren Fehlanwendung» in Art. 3 (13) KI-VO.

24 Jonas Schuett, «Risk Management in the Artificial Intelligence Act», in: European Journal of Risk Regulation (2023), S. 11, doi:10.1017/err.2023.1.

25 Vgl. Erw. 65 KI-VO.

26 Vgl. NIST AI RMF 1.0 (Fn. 1), S. 8.

wirksamer zu minimieren und gleichzeitig ein angemessenes Gleichgewicht bei der Durchführung der Massnahmen zur Erfüllung dieser Anforderungen sicherzustellen (Art. 9 Abs. 4 KI-VO).

[28] Der Anbieter hat bei der Festlegung der am besten geeigneten Risikomanagementmassnahmen sicherzustellen, dass die ermittelten und bewerteten Risiken, durch eine geeignete Konzeption und Entwicklung des Hochrisiko-KI-Systems beseitigt oder verringert werden, soweit dies technisch möglich ist (Art. 9 Abs. 5 lit. a KI-VO). Dabei sind angemessene Minderungs- und Kontrollmassnahmen zur Bewältigung nicht auszuschliessender Risiken anzuwenden (Art. 9 Abs. 5 lit. b KI-VO).

[29] Darüber hinaus müssen die **Risikomanagementmassnahmen** so gestaltet sein, dass jedes mit einer bestimmten Gefahr verbundene relevante **Restrisiko** sowie das Gesamtrestrisiko der Hochrisiko-KI-Systeme als **vertretbar** beurteilt wird (Art. 9 Abs. 5 KI-VO). Der Zweck der Umsetzung von Risikomanagementmassnahmen besteht also darin, Risiken so zu reduzieren, dass jedes Noch-Restrisiko **vertretbar** ist. Folglich wird die Risikominimierungspflicht dadurch relativiert, dass von den Anbietern kein völlig fehlerfreies Hochrisiko-KI-System erwartet wird. Vielmehr geht der europäische Gesetzgeber davon aus, dass gewisse verbleibende Restrisiken vertretbar sein können.[27] «Vertretbares Risiko» kann als das Risikoniveau definiert werden, das in einem bestimmten Kontext, basierend auf den aktuellen gesellschaftlichen Werten, akzeptiert wird. Um zu bestimmen, ob ein Risiko vertretbar ist, müssen Anbieter die Risiken und Vorteile abwägen. Im Allgemeinen ist ein Risiko akzeptabel, wenn die Vorteile die Risiken (deutlich) überwiegen. Da die KI-VO Act jedoch den Schutz von Gesundheit, Sicherheit und Grundrechten zum Ziel hat, ist das Mass an Risiko, das Anbieter akzeptieren können, begrenzt – es hängt nicht nur von ihrer eigenen Risikobereitschaft («*risk appetite*») ab.[28] Im Hinblick auf Art. 9 Abs. 9 KI-VO muss der Massstab für die Vertretbarkeit von Restrisiken, die sich auf Personen unter 18 Jahren oder gegebenenfalls andere schutzbedürftige Gruppen nachteilig auswirken können, besonders hoch sein. Das Abwägen von Risiken und Vorteilen beinhaltet nicht nur rein technische Entscheidungen, sondern auch empirische Unsicherheiten und schwierige normative, kontextabhängige Entscheidungen, insbesondere auch angesichts dessen, dass die Festlegung normativer Schwellenwerte weiterhin ein ungelöstes Problem in der KI-Ethik bleibt. Selbst wenn beispielsweise ein anerkannter Prozess zur Bewertung der Auswirkungen auf die Grundrechte dem Anbieter ermöglichen würde, klarer zu erkennen, wessen Sicherheit und Rechte durch ein Hochrisiko-KI-System betroffen sein könnten, muss der Anbieter dennoch eine Wertentscheidung darüber treffen, ob das Risiko dieser Auswirkungen akzeptabel ist.[29] Zudem dürfte sich die Risikotoleranz wohl im Laufe der Zeit ändern, da sich KI-Systeme, Richtlinien und Normen weiterentwickeln. Unterschiedliche Organisationen können aufgrund ihrer spezifischen organisatorischen Prioritäten und Ressourcenüberlegungen unterschiedliche Risikotoleranzen aufweisen. Solange die Herausforderungen bei der Festlegung der Risikotoleranzen für KI bestehen bleiben, kann es Kontexte geben, in denen ein Risikomanagementsystem noch nicht ohne weiteres zur Minderung negativer KI-Risiken führt.[30]

---

27 Nadja Braun Binder/Catherine Egli (Fn. 19), N 33 zu Art. 9 KI-VO.

28 Jonas Schuett, (Fn. 24), S. 14.

29 Henry Fraser/José-Miguel Bello y Villarino, «Where residual risks reside – A comparative approach to art 9(4) of the EU's Proposed AI Regulation», Ziffer 1.4, September 2021, https://ssrn.com/abstract=3960461.

30 Vgl. NIST AI RMF 1.0 (Fn. 1), S. 7.

[30] Darüber hinaus muss der Anbieter die gemäss Art.13 KI-VO erforderlichen Informationen für die Betreiber bereitstellen und gegebenenfalls die Betreiber entsprechend schulen (Art. 9 Abs. 5 lit. c KI-VO). Dies bedeutet u.a. dass die vom Hochrisiko-KI-System erzielten Ergebnisse für die Betreiber transparent und nachvollziehbar zu sein haben und der Anbieter dem Betreiber eine **Betriebsanleitung** nach den Vorgaben von Art. 13 Abs. 3 KI-VO abgeben muss. Der Anbieter hat gemäss Art. 9 Abs. 5 in fine KI-VO zudem die technischen Kenntnisse, die Erfahrungen und den Bildungsstand, die vom Betreiber erwartet werden können, sowie den voraussichtlichen Kontext, in dem das System eingesetzt werden soll, gebührend zu berücksichtigen. Der Anbieter ist laut Erw. 65 KI-VO jedoch nicht verpflichtet, als Risikominderungsmassnahmen in Bezug auf vorhersehbare Fehlanwendungen spezifische zusätzliche Schulungen für das Hochrisiko-KI- System beim Betreiber durchzuführen. Dennoch stellt sich die Frage, wie ein Anbieter die KI-Kompetenz («*AI literacy*», vgl. Art. 3 (56) KI-VO) des Betreibers korrekt einschätzen können soll.

[31] Ein weiteres essentielles Element des Risikomanagementsystems ist das Durchlaufen von **Tests**. Gemäss Art. 9 Abs. 6 KI-VO müssen Hochrisiko-KI-Systeme getestet werden, um die am besten geeigneten gezielten Risikomanagementmassnahmen zu ermitteln. Durch das Testen wird sichergestellt, dass Hochrisiko-KI-Systeme stets im Einklang mit ihrer Zweckbestimmung funktionieren und die Anforderungen von Art. 8–15 KI-VO (Abschnitt 2) erfüllen. Die Testverfahren können einen Test unter Realbedingungen gemäss Artikel 60 KI-VO umfassen (Art. 9 Abs. 7 KI-VO).

[32] Das Testing kann gemäss Art. 9 Abs. 8 KI-VO während des gesamten Entwicklungsprozesses des Hochrisiko-KI-Systems durchgeführt werden, muss aber zwingend vor der Inbetriebnahme erfolgen. Somit erstreckt sich die Testpflicht nicht auf den gesamten Lifecycle des Hochrisiko-KI-Systems.

[33] Das Testen erfolgt anhand vorab festgelegter Metriken und Wahrscheinlichkeitsschwellenwerte, die für die Zweckbestimmung des Hochrisiko-KI- Systems geeignet sind (vgl. Art. 9 Abs. 8 KI-VO). «Metrik» umfasst Bewertungskriterien, Benchmarks und Schlüsselindikatoren. «Wahrscheinlichkeitsschwellenwerte» stellen eine besondere Art von Metrik dar, die eine Eigenschaft auf einer probabilistischen Skala mit einem oder mehreren vordefinierten Schwellenwerten bewertet. KI-Systeme verhalten sich jedoch oft anders, wenn sich die Anwendungsumgebung von der Trainingslandschaft unterscheidet («*distributional shift*»). Anhand der Testverfahren können die Anbieter daher erkennen, in welchen Anwendungsumgebungen das Hochrisiko-KI-System wahrscheinlich Mühe haben wird, im Einklang mit seiner Zweckbestimmung zu agieren («*out-of-distribution detection*»), und entsprechende Massnahmen ergreifen.[31]

[34] Eine Herausforderung dürfte für die Anbieter sein, dass es keine allgemeingültigen Prinzipien gibt, um zu entscheiden, welche Metrik oder welcher Wahrscheinlichkeitsschwellenwert zu verwenden ist, da deren Angemessenheit sehr kontextspezifisch ist und es bisher keine bewährten Verfahren gibt.[32] Dies kann zu Unsicherheiten führen, ob die in der Vergangenheit verwendeten Metriken in der Zukunft möglicherweise noch geeignet sind. Zu den potenziellen Fallstricken bei der Messung negativer Risiken oder Schäden gehört, dass die Entwicklung von Metriken oft ein institutionelles Unterfangen ist und unbeabsichtigt Faktoren widerspiegeln kann, die mit den zugrunde liegenden Auswirkungen nicht im Zusammenhang stehen. Darüber hinaus können Metri-

31 Nadja Braun Binder/Catherine Egli (Fn. 19) N 43 zu Art. 9 KI-VO.

32 Jonas Schuett (Fn. 24), S. 16.

ken übermässig vereinfacht, manipuliert oder unkritisch angewendet werden, wodurch wichtige Nuancen verloren gehen. Dies führt zu einer unzureichenden Berücksichtigung der Unterschiede zwischen den betroffenen Gruppen und Kontexten. Ferner sind Ansätze zur Messung der Auswirkungen auf Personengruppen dann am effektivsten, wenn sie berücksichtigen, dass Kontexte eine bedeutende Rolle spielen, dass Schäden verschiedene Personengruppen unterschiedlich betreffen können und dass Personengruppen, die möglicherweise geschädigt werden könnten, nicht immer direkte Nutzer eines Systems sind.[33] Der gegenwärtige Mangel an Konsens über robuste und überprüfbare Messmethoden für Risiko und Vertrauenswürdigkeit sowie deren Anwendbarkeit auf verschiedene KI-Anwendungsfälle stellt eine wesentliche Herausforderung für das KI-Risikomanagementsystem dar. Mit den derzeitig verfügbaren Testmethoden können Probleme leicht übersehen werden.

[35] Das Risikomanagement muss, wie erwähnt, während des gesamten Lifecycles des Hochrisiko-KI-Systems, von der Konzeption bis zur Stilllegung kontinuierlich durchgeführt werden. Dies bedeutet, dass Hochrisiko-KI-Systeme auch nach ihrem Inverkehrbringen weiterhin überwacht und relevante Daten im Rahmen des «*post-market monitoring*» nach Art. 72 KI-VO gesammelt und analysiert werden müssen. Dadurch dass der Anbieter das Risikomanagementsystem während des gesamten Lebensyzklus des KI-Systems einsetzen und laufend aktualisieren muss, soll die Dauerhaftigkeit der Risikoerkennung, -bewertung und -analyse gewährleistet werden.[34]

[36] Bei Anbietern von Hochrisiko-KI-Systemen, die den Anforderungen an interne Risikomanagementprozesse gemäss anderer einschlägiger Bestimmungen des EU-Rechts unterliegen[35], können die in den Absätzen 1 bis 9 Art. 9 KI-VO enthaltenen Aspekte Bestandteil der nach diesem Recht festgelegten Risikomanagementverfahren sein oder mit diesen Verfahren kombiniert werden (Art. 9 Abs. 10 KI-VO). Zur Gewährleistung der Kohärenz und zur Vermeidung eines unnötigen Verwaltungsaufwands und unnötiger Kosten sollten die Anbieter eines Produkts, das ein oder mehrere Hochrisiko-KI-Systeme enthält, über einen gewissen Spielraum für operative Entscheidungen verfügen. Diese Flexibilität gilt insbesondere für Produkte, die sowohl den Anforderungen der KI-VO als auch den auf dem NLF basierenden EU-Harmonisierungsrechtsvorschriften gemäss Anhang I KI-VO entsprechen müssen. Dies ermöglicht es den Anbietern, ihre Produkte in optimaler Weise mit allen geltenden harmonisierten Rechtsvorschriften der EU in Einklang zu bringen.[36]

[37] Diese Flexibilität soll Anbietern die Möglichkeit geben, einen Teil der erforderlichen Prüf- und Meldeverfahren, Informationen und Unterlagen, die nach KI-VO erforderlich sind, in bereits bestehende Unterlagen und Verfahren zu integrieren, die nach anderen EU-Regularien auf der Grundlage des NLF erforderlich und in Anhang I KI-VO aufgeführt sind. Die EU stellt sich denn auch vor, dass ein «extra layer» in bestehende Prozesse wie insbesondere dem Risikomanagementsystem nach Art. 9 KI-VO eingebaut werden könne.

---

33 Artificial Intelligence Risk Management Framework (AI RMF 1.0) National Institute of Standards and Technology, U.S. Department of Commerce, 26. Januar 2023, S. 6.

34 Vgl. Nadja Braun Binder/Catherine Egli (Fn. 19), N 17 zu Art. 9 KI-VO.

35 Bspw. im Finanzdienstleistungssektor oder im Zusammenhang mit Medizinprodukten.

36 Vgl. Erw. 64 KI VO.

## 6. Normen und Standards

[38] Die EU erachtet Normen und Standards als wichtige Instrumente zur Unterstützung der Umsetzung der KI-VO sowie zur Gewährleistung eines hohen Schutzniveaus für Sicherheit und Grundrechte bei der Förderung sicherer und vertrauenswürdiger KI-Systeme. Normen und Standards sollen dazu beitragen, gleiche Wettbewerbsbedingungen und ein «*level playing field*» für das Design und Entwicklung von KI-Systemen zu schaffen, insbesondere für kleine und mittlere Unternehmen, die KI-Lösungen entwickeln. Mit Normen und Standards soll eine harmonisierte technische Grundlage für die Umsetzung der KI-VO geschaffen werden. Zudem soll durch die Einbindung internationaler Standards und die Berücksichtigung europäischer Werte eine globale Harmonisierung angestrebt werden, die den Schutz der Grundrechte und die Sicherheit der Nutzerinnen und Nutzer gewährleistet.[37]

[39] Wenn Standards und Normen den Vorgaben der KI-VO entsprechen, kann die EU-Kommission diese als «harmonisierte Normen» im Sinne von Verordnung (EU) Nr. 1025/2012 anerkennen. Wenn Hochrisiko-KI-Systeme mit den harmonisierten Normen gemäss Art. 40 KI-VO übereinstimmen, wird Konformität mit den Pflichten gemäss Abschnitt 2 (Art. 8–15) KI-VO vermutet, soweit diese Anforderungen oder Verpflichtungen von den Normen abgedeckt sind.

[40] Am 22. Mai 2023 erteilte die Europäische Kommission dem Comité Européen de Normalisation (CEN) und dem Comité Européen de Normalisation Électrotechnique (CENELEC) den Auftrag[38], zur Unterstützung der Politik der Union im Bereich Künstliche Intelligenz Normen in den folgenden Bereichen zu erstellen, welche sich zu einem grossen Teil auf Art. 8–15 (Abschnitt 2) KI-VO beziehen:

1. Risikomanagementsysteme für KI-Systeme;
2. Verwaltung und Qualität von Datensätzen, die für den Aufbau von KI-Systemen verwendet werden («*data and data governance*»);
3. Aufzeichnung durch Protokollierungsfunktionen («*logging*») von KI-Systemen;
4. Transparenz und Information der Betreiber von KI-Systemen;
5. menschliche Aufsicht über KI-Systeme («*human oversight*»);
6. Anforderungen an Spezifikationen zur Genauigkeit von KI-Systemen;
7. Robustheitsspezifikationen für KI-Systeme;
8. Cybersicherheitsspezifikationen für KI-Systeme;
9. Qualitätsmanagementsysteme für Anbieter von KI-Systemen, einschliesslich Verfahren zur Überwachung nach dem Inverkehrbringen;
10. zur Konformitätsbewertung von KI-Systemen.[39]

37 Standardisation Request (Fn. 11), Annex 1 und 2.

38 Standardisation Request (Fn. 11), Annex 1 und 2.

39 Standardisation Request (Fn. 11), Annex 1, S. 1.

[41] Die Normen sollen «*state of the art*»[40] sein, mit dem EU-Recht betr. Grundrechten und Datenschutz in Einklang stehen und technologie-, prozess- oder methodengestützte technische Spezifikationen für den Entwurf und die Entwicklung von KI-Systemen, einschliesslich Verifikations-, Validierungs- und Testverfahren sowie objektiv nachprüfbarer Kriterien und praktikabler Methoden zur Bewertung der Einhaltung dieser Spezifikationen enthalten.[41] Ferner haben CEN und CENELEC sicherzustellen, dass die erstellten europäischen Normen und europäischen Normungsprodukte gegebenenfalls mit bestehenden und künftigen europäischen und allgemeinen Normen übereinstimmen, die in den verschiedenen einschlägigen Sektoren entwickelt wurden, insbesondere mit solchen, die sich auf Produkte beziehen, die unter bestehende Sicherheitsvorschriften der EU, einschliesslich der in Richtlinie 2001/95/EG über die allgemeine Produktsicherheit, fallen.[42]

[42] CEN und CENELEC sollen der Kommission bis zum 30. April 2025 den gemeinsamen Abschlussbericht vorlegen.[43] Angesichts dessen, dass der Auftrag der EU-Kommission am 22. Mai 2023 zeitlich vor dem Erlass der KI-VO erfolgte, ist davon auszugehen, dass der Auftrag vom 22. Mai 2023 nochmals angepasst wird/wurde.

[43] Aktuell bestehen bereits internationale Leitlinien für KI-Risikomanagement: Am 26. Januar 2023 veröffentlichte das National Institute of Standards and Technology (NIST) das Artificial Intelligence Risk Management Framework (AI RMF)[44], ein Leitdokument zur Nutzung durch Organisationen, die KI-Systeme entwerfen, entwickeln oder verwenden. Das AI RMF zielt darauf ab, einen praktischen Rahmen für die Messung und den Schutz vor potenziellen Schäden durch KI-Systeme bereitzustellen, indem Risiken gemindert, Chancen genutzt und die Vertrauenswürdigkeit von KI-Systemen erhöht werden. Das Rahmenwerk soll freiwillig, rechtserhaltend, nicht sektorspezifisch und anwendungsunabhängig sein, so dass Organisationen jeder Grösse, in allen Sektoren und in der gesamten Gesellschaft die Flexibilität haben, die Ansätze des Rahmenwerks umzusetzen.

[44] Am 6. Februar 2023 veröffentlichte die ISO mit ISO/IEC 23894:2023, eine Leitlinie zum Risikomanagement für KI, welche auf ISO 31000:2018 aufbaut und bereits am 22. Februar 2024 wieder revidiert wurde.[45] Ähnlich dem NIST AI RMF zielt die ISO-Leitlinie 23894:2024 darauf ab, Organisationen, die KI-Systeme entwickeln, einsetzen oder nutzen, bei der Einführung von Best Practices im Risikomanagement zu unterstützen. Die Leitlinie skizziert eine Reihe von Leitprinzipien, wonach das Risikomanagement für KI-Systeme integriert, strukturiert und umfassend, massgeschneidert, inklusiv, dynamisch, auf den besten verfügbaren Informationen basierend, menschliche und kulturelle Faktoren berücksichtigend und kontinuierlich verbessert werden sollte.

---

40 Die EU-Kommission versteht unter dem «Stand der Technik» ein zu einem bestimmten Zeitpunkt entwickelter Stand der technischen Leistungsfähigkeit von Produkten, Verfahren und Dienstleistungen, der auf den einschlägigen gesicherten Erkenntnissen von Wissenschaft, Technik und Erfahrung beruht und als gute fachliche Praxis anerkannt ist. Der Stand der Technik bedeutet nicht notwendigerweise die neueste wissenschaftliche Forschung, die sich noch im Versuchsstadium befindet oder noch nicht ausgereift ist, vgl. Standardisation Request (Fn. 11), Annexe 1 und 2, S. 2.

41 Standardisation Request (Fn. 11), Annexe 1 und 2, S. 3.

42 Standardisation Request (Fn. 11), Annexe 1 und 2, S. 4.

43 Der erste Report ist einsehbar unter https://publications.jrc.ec.europa.eu/repository/handle/JRC132833.

44 NIST AI RMF (Fn. 1).

45 ISO/IEC JTC 1/SC 42, das Subcommittee der International Organization for Standardization (ISO), hat in den letzten Jahren eine Vielzahl an Standards und technische Reporte im Zusammenhang mit KI erstellt.

[45] ISO/IEC 23894:2024 definiert ähnlich dem NIST AI RMF Prozesse und Richtlinien für die Anwendung des Risikomanagements bei KI und baut auf ISO 31000:2018, indem diese um KI-spezifische Kontexte ergänzt werden. Der Zweck des Risikomanagements soll in der Schaffung und dem Schutz von «Werten» bestehen. Die Leitlinien umfassen die Kommunikation und Konsultation, die Festlegung des Kontexts, die Bewertung, Behandlung, Überwachung, Überprüfung, Dokumentation und Berichterstattung von Risiken, die mit der Entwicklung und Nutzung von KI-Systemen verbunden sind. Dabei wird Risikomanagement als integraler Bestandteil aller organisatorischen Aktivitäten verstanden. Ein strukturierter Ansatz soll zu konsistenten und vergleichbaren Ergebnissen führen. Einen besonderen Fokus legt IO/IEC 23894:2024 auf dynamische Anpassungen an diverse Veränderungen sowie den angemessenen und rechtzeitigen Einbezug aller relevanten «*stakeholder*», um menschliche und kulturelle Faktoren einzubeziehen.

[46] Der eigentliche Risikomanagement-Prozess umfasst die folgenden Elemente:

1. **Kommunikation und Konsultation**: Einbeziehen aller relevanten «*stakeholder*»;
2. **Festlegen des Umfangs, Kontexts und der Kriterien:** Definition der relevanten Parameter und Risikokriterien;
3. **Risikobewertung:** Identifikation, Analyse und Bewertung von Risiken;
4. **Risikobehandlung:** Auswahl und Implementierung geeigneter Massnahmen zur Risikominimierung;
5. **Überwachung und Überprüfung:** Kontinuierliche Überwachung und Anpassung der Risikomanagementprozesse;
6. **Aufzeichnung und Berichterstattung:** Systematische Dokumentation und Berichterstattung der Risikomanagementaktivitäten.

[47] Zu beachten ist, dass «Risiko» gemäss ISO 31000:2018 eine negative oder positive Abweichung von den Zielen bedeutet.[46] Da ISO/ IEC 23894:2024 somit eine andere Definition von Risiko als die KI-VO enthält, den Fokus nicht auf das KI-System als Produkt mit entsprechenden Risiken für die Gesundheit, Sicherheit und die Grundrechte, sondern vielmehr auf organisatorische Prozesse zur Vermeidung von Risiken für die Organisation selbst legt, kann ISO/ IEC 23894:2024 den Anforderungen der KI-VO nicht genügen.

## 7. Abgrenzung zur Grundrechte-Folgenabschätzung

[48] Das Risikomanagementsystem gemäss Art. 9 KI-VO ist nicht mit der Grundrechte-Folgenabschätzung («*fundamental rights impact assessment*», abgekürzt «FRIA») für Hochrisiko-KI-Systeme gemäss Art. 27 KI-VO zu verwechseln. Mit der Grundrechte-Folgenabschätzung muss vor der Inbetriebnahme eines Hochrisiko-KI-Systems eine Abschätzung der Auswirkungen, die die Verwendung eines solchen Systems auf die Grundrechte haben kann, durchgeführt werden. Diese Abschätzung muss eine Beschreibung der Verfahren über den Einsatz des Hochrisiko-KI-Systems, eine Beschreibung des Zeitraums und der Häufigkeit, innerhalb dessen bzw. mit der ein Hochrisiko-KI-System verwendet werden soll, die Kategorien der natürlichen Personen und Personengruppen, die von seiner Verwendung im spezifischen Kontext betroffen sein könnten, die

46 «Risk» involves a negative or positive deviation from the objectives, ISO/IEC 23894:2023(E), S. 2.

Identifikation spezifischer Schadensrisiken für die betroffenen Personen, sowie die Massnahmen der menschlichen Aufsicht sowie Risikominimierungsmassnahmen umfassen.

[49] Diese Grundrechte-Folgenabschätzung für Hochrisiko-KI-Systeme ist nicht vom Anbieter, sondern vom Betreiber (bzw. von bestimmten Betreibern von bestimmten Hochriskio-KI-Systemen) zu erstellen. Allerdings ist der Anbieter gemäss Art. 13 KI-VO verpflichtet, dem Betreiber insbesondere für die Identifikation spezifischer Schadensrisiken für die betroffenen Personen entsprechende Informationen zur Verfügung zu stellen. Welche Informationen dies im Hinblick auf das vom Anbieter zu implementierende Risikomanagementsystem konkret sind, bleibt unter Art. 27 i.V.m. Art. 13 KI-VO allerdings noch unklar.

## 8. Bedeutung der Pflicht eines Risikomanagementsystems für Schweizer Unternehmen

[50] Bekanntlich hat die KI-VO extraterritoriale Wirkung: Gemäss Art. 2 (1) lit. a KI-VO gilt die KI-VO für Anbieter, die in der EU KI-Systeme in Verkehr bringen oder in Betrieb nehmen, unabhängig davon, ob diese Anbieter in der EU oder in einem Drittland niedergelassen sind. Ferner gilt die KI-VO gemäss Art. 2 (1) lit. c KI-VO für Anbieter (und Betreiber) von KI-Systemen, die ihren Sitz in einem Drittland haben oder sich in einem Drittland befinden, wenn die vom KI-System hervorgebrachte Ausgabe («*output*») in der EU verwendet wird.

[51] Wird Art. 2 (1) lit. c KI-VO wörtlich genommen, so kann die extraterritoriale Wirkung enorm sein: allein der Umstand, dass der «*output*» in der EU genutzt wird, soll für die Anwendung der KI-VO ausreichend sein. Dies bedeutet, dass Schweizer Anbieter, die nicht ausschliessen können, dass der «*output*» ihrer Hochrisiko KI-Systeme in der EU verwendet werden, grundsätzlich sämtliche Anbieterpflichten gemäss KI-VO zu erfüllen haben, also auch die Anforderungen an das Risikomanagementsystem nach Art. 9 KI-VO. Dies gilt eben selbst dann, wenn sie ihre Hochrisiko-KI-Systeme weder in der EU in Verkehr gebracht haben noch durch einen Kunden in der EU in Betrieb haben nehmen lassen. Erw. 22 ist im Hinblick auf Art. 2 (1) lit. c KI-VO kaum dienlich, wonach die KI-VO auch für in einem Drittland niedergelassene Anbieter (und Betreiber) gelten, soweit *beabsichtigt* wird, dass die von ihren KI-Systemen erzeugte Ausgabe («*output*») in der EU verwendet wird: Art. 2 (1) lit. c KI-VO referenziert keine wie auch immer geartete Absicht. Es gibt sogar Stimmen die behaupten, dass selbst vertragliche Ausschlüsse der Verwendung des «*outputs*» in der EU irrelevant wären.[47] Es bleibt zu abzuwarten, ob EU-Guidelines die extraterritoriale Wirkung von Art. 2 KI-VO auf ein vernünftiges Mass einschränken werden.

## 9. Würdigung

[52] Für die Entwicklung und das Inverkehrbringen eines Hochrisiko-KI-Systems ist ein wirksames und dokumentiertes Risikomanagement durch den Anbieter essentiell. Das KI-Risikomanagement kann dabei auf allenfalls bereits etablierte Risikomanagementsysteme in einem Unternehmen aufbauen. Sollten solche nicht bestehen, müssen diese nun neu bis zur Geltung der

47 Tim Hickman/Thomas Harper, «The EU AI Act's extraterritorial scope – Part 2», Data Protection Ireland, dpjournals.ie, Volume 17, Issue 13, https://www.whitecase.com/insight-our-thinking/eu-ai-acts-extraterritorial-scope-part-2.

Bestimmungen über Hochrisiko-KI-Systeme nach Inkrafttreten der KI-VO, spätestens also 24 Monate für Hochrisiko-KI-Systeme nach Art. 6 Abs. 2 KI-VO und 36 Monate für Hochrisiko-KI-Systeme nach Art. 6 Abs. 1 KI-VO nach Inkrafttreten der KI-VO eingerichtet werden. Sollten dannzumal harmonisierte Normen nach Art. 40 KI-VO oder andere Leitlinien seitens der EU bestehen, können diese sehr hilfreich sein, nicht zuletzt deshalb, weil Art. 9 KI-VO einige klärungsbedürftige Termini verwendet. Dass die zu erstellenden Normen dereinst, sofern von der EU-Kommission akzeptiert, «bloss» harmonisierte Normen sein werden, deren Einhaltung eine blosse Vermutung für die Einhaltung der Vorgaben der KI-VO erzeugt, statt deren Einhaltung zwingend vorzuschreiben, ist zu begrüssen. Sobald die harmonisierten Normen verfügbar sind, könnte die Berufung auf die Einhaltung von bestehenden ISO-Normen 310008 und 23894:2024 betr. KI und Risikomanagement den Anforderungen eines Risikomanagementsystems nach Art. 9 KI-VO nicht genügen.

[53] Ganz allgemein dürfte es zumindest für private Unternehmen eine Herausforderung sein, einerseits die Risiken der von ihnen entwickelten und angebotenen Hochrisiko-KI-Systeme für Gesundheit, Sicherheit und Grundrechte ohne externe Expertise zu identifizieren, gehören doch Grundrechtsanalysen über eine Datenschutz-Folgenabschätzung gemäss Art. 35 DSG-VO hinaus nicht unbedingt zu deren Kernkompetenzen. Eine weitere Herausforderung stellt zudem die Einschätzung von Risiken der Hochrisiko-KI-Systeme im Rahmen des Risikomanagementsystems dar, vor allem hinsichtlich der vernünftigerweise vorhersehbaren Fehlanwendungen. Hier könnte eine Vielzahl von Verwendungen in Frage kommen. Um diese Risikoeinschätzung vorzunehmen, ist der Einbezug von divers zusammengesetzten Testgruppen oder anderen Stakeholdern ausserhalb des Unternehmens bei der Entwicklung und beim Testing zu empfehlen.[48] Schliesslich dürfte es für Anbieter als auch deren Kunden als Betreiber ungewohnt sein, dass sich ein Anbieter ein Bild darüber verschaffen soll, ob und inwiefern die Mitarbeitenden des Kunden über «*AI literacy*» verfügen.

[54] Auch kann ganz grundsätzlich fraglich sein, ob im Zusammenhang mit dem Risikomanagementsystem der gewählte Ansatz des Produktsicherheitsrechts für die Einhaltung von Grundrechten der richtige Ansatz ist, zumal die in ihren Grundrechten gegebenenfalls gefährdeten Personen als konkret betroffene Personen lediglich über ein Recht auf Beschwerde bei einer Marktüberwachungsbehörde gemäss Art. 85 KI-VO sowie über ein Recht auf Erläuterung der Entscheidungsfindung im Einzelfall gegenüber dem Betreiber nach Art. 86 KI-VO, nicht aber gegenüber dem Anbieter verfügen.[49]

---

Martina Arioli ist Partnerin in der auf digitale Transformation, IT-Recht und Datenrecht spezialisierten Boutique-Anwaltskanzlei Arioli Law in Zürich. Sie ist Dozentin für IT-Recht und Datenschutzrecht an der Universität Zürich, Europa Institut.

Alle Websites wurden zuletzt am 15. Juni 2024 geprüft.

---

48 Ein interessanter Vorschlag in diesem Zusammenhang ist das sog. «*Stakeholder Impact Assessment*», macht David Leslie, «Understanding artificial intelligence ethics and safety – A guide for the responsible design and implementation of AI systems in the public sector», Public Policy Programme, The Alan Turing Institute, 2019, S. 26.

49 Immerhin sieht der «Proposal for a Regulation of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)» COM (2022) 496 final gemäss Art. 4 Abs. 2 vor, dass Ansprüche gestützt auf ausservertragliche Haftung erleichtert werden sollen, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496.