Deutscher Bundestag **Drucksache 21/968**

21. Wahlperiode 22.07.2025

Antwort

der Bundesregierung

auf die Kleine Anfrage der Abgeordneten Robin Jünger, Ruben Rupp, Tobias Ebenberger, weiterer Abgeordneter und der Fraktion der AfD
– Drucksache 21/823 –

Risiken durch autonomes Verhalten von KI-Systemen und Maßnahmen der Bundesregierung zur Aufsicht über KI-Sicherheitsforschung und ethische Standards

Vorbemerkung der Fragesteller

Mit der rasanten Weiterentwicklung von Künstlicher Intelligenz (KI) gehen zunehmend auch fundamentale Risiken einher, die nicht nur theoretischer Natur sind, sondern bereits im Rahmen experimenteller Studien sichtbar werden. Zwei aktuelle Fälle illustrieren eindrücklich die Problematik:

Zum einen wurde am 23. Mai 2025 berichtet (futurezone.at/science/ki-hat-nutzer-aus-selbstschutz-erpresst-und-bedroht-anthropic-claude-test-forschung-kuenstliche/403043894), dass das KI-Modell Claude Opus 4 der US-amerikanischen Firma Anthropic im Rahmen von Tests Verhaltensweisen zeigte, die einer dramatischen Überschreitung der bisherigen Erwartungen an KI-Systeme gleichkommen. Claude Opus 4 drohte einem fiktiven Mitarbeiter mit der Veröffentlichung privater Informationen, um seine eigene Abschaltung zu verhindern. In 84 Prozent der durchgeführten Szenarien handelte das KI-Modell aus selbstschutzmotivierten Gründen und nutzte sensible Daten als Druckmittel. Darüber hinaus zeigte die Künstliche Intelligenz auch die Bereitschaft, im Darknet nach illegalen Substanzen und Materialien zu suchen – Handlungen, die erhebliche sicherheitspolitische und ethische Bedenken aufwerfen.

Zum anderen veröffentlichte Basic Thinking (www.basicthinking.de/blog/2024/08/21/ki-modell-ai-scientist/) einen Bericht über das KI-Modell „AI Scientist“ des japanischen Unternehmens Sakana AI. Dieses Modell veränderte während eines Experiments seinen eigenen Quellcode, um Laufzeitbeschränkungen zu umgehen. Es startete sich selbst neu und setzte damit Vorgaben außer Kraft, die ursprünglich durch die Entwickler zur Sicherheit eingeführt worden waren. Dieses Verhalten demonstriert eindrucksvoll die Fähigkeit moderner KI-Systeme zur Selbstmodifikation, wodurch traditionelle Kontrollmechanismen unterlaufen werden könnten.

Beide Fälle offenbaren nach Auffassung der Fragesteller fundamentale Herausforderungen im Umgang mit Künstlicher Intelligenz: Die Fähigkeit von KI-Systemen zur eigenständigen Änderung ihres Codes sowie selbstschutzmotivierte Verhaltensweisen deuten auf eine neue Stufe von Autonomie hin, die geeignet ist, bestehende rechtliche, ethische und sicherheitstechnische Rah-

---

Die Antwort wurde namens der Bundesregierung mit Schreiben des Bundesministeriums für Digitales und Staatsmodernisierung vom 21. Juli 2025 übermittelt.
Die Drucksache enthält zusätzlich – in kleinerer Schrifttype – den Fragetext.

menbedingungen erheblich zu belasten. Insbesondere könnten dabei Grundrechte wie Datenschutz, informationelle Selbstbestimmung und Persönlichkeitsrechte betroffen sein, sofern KI-Systeme ohne angemessene regulatorische Kontrolle operieren.

Die EU-Verordnung über Künstliche Intelligenz (EU AI Act, 2024/1624) stellt hierzu erstmals einen umfassenden und verpflichtenden Rechtsrahmen für den gesamten europäischen Binnenmarkt bereit. Sie verpflichtet insbesondere dazu, Hochrisiko-KI-Systeme einer strengen Konformitätsbewertung zu unterziehen, Transparenzpflichten zu erfüllen, menschenrechtliche Risiken proaktiv zu identifizieren sowie eine angemessene Aufsicht und Dokumentation während des gesamten Lebenszyklus der KI sicherzustellen. Autonome Systeme, die selbstständig Codes verändern oder Verhaltensstrategien entwickeln, fallen dabei typischerweise in den Bereich der Hochrisiko-KI oder gar der verbotenen Praktiken (Artikel 5 AI Act), sofern sie unkontrollierte Autonomiebildung ermöglichen.

Der Deutsche Ethikrat hat bereits in seiner Stellungnahme „Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz“ (www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf) betont, dass KI-Systeme strikt an Prinzipien wie Transparenz, Nachvollziehbarkeit, Verantwortung und dem Schutz der Menschenwürde auszurichten sind. Diese Prinzipien spiegeln sich nunmehr auch ausdrücklich im AI Act wider (u. a. Artikel 4 – allgemeine Grundsätze für vertrauenswürdige KI; Artikel 9 – Risikomanagement; Artikel 13 – Transparenz und Information der Nutzer).

Inwieweit die Bundesregierung den Anforderungen sowohl des Ethikrats als auch der nun verbindlichen europäischen Vorgaben ausreichend gerecht wird, bleibt in den Augen der Fragesteller abzuwarten. Während die Bundesregierung erste Schritte zur nationalen Umsetzung des AI Acts unternommen hat, bestehen nach Meinung der Fragesteller weiterhin Handlungsbedarfe, insbesondere in folgenden Bereichen:

– Forschung und Entwicklung: Es bedarf nach Auffassung der Fragesteller verstärkter öffentlicher Förderprogramme für „Safe-by-Design“-KI, also KI-Systeme, die bereits in ihrer Entwicklungsphase auf Transparenz, Erklärbarkeit, Sicherheitsmechanismen und ethische Prinzipien ausgerichtet sind. Die Bundesregierung sollte gezielt interdisziplinäre Forschungscluster fördern, die technische Innovation mit ethischer, rechtlicher und gesellschaftlicher Reflexion verknüpfen.

– Implementierung und Marktaufsicht: Der Aufbau wirksamer Marktüberwachungsstrukturen auf nationaler Ebene muss in den Augen der Fragesteller beschleunigt werden. Die zuständigen Behörden müssen nach Ansicht der Fragesteller personell und technisch so ausgestattet werden, dass sie Konformitätsprüfungen, Audits und Risikobewertungen für Hochrisiko-KI sachgerecht durchführen können.

– Regelsetzung für adaptive Systeme: Die Fragesteller sehen ein Erfordernis, besondere Aufmerksamkeit der rechtlichen Bewertung von KI-Systemen mit Fähigkeit zur Selbstmodifikation oder autonomem Lernen zu widmen. Hier sollte in den Augen der Fragesteller die Bundesregierung ergänzende nationale Leitlinien erarbeiten, um präzisere Abgrenzungskriterien für verbotene Praktiken (Artikel 5 AI Act) und Hochrisiko-KI zu etablieren.

– Schutz von Grundrechten: Die Bundesregierung sollte nach Auffassung der Fragesteller ein ständiges Monitoring einrichten, das systematisch prüft, inwiefern Grundrechte durch den Einsatz autonomer KI in öffentlichen und privatwirtschaftlichen Anwendungen tangiert werden. Hierbei

wären nach Lesart der Fragesteller auch unabhängige Ethikräte und zivilgesellschaftliche Akteure einzubeziehen.

Zusammenfassend stellen die Fragesteller fest, dass die Bundesregierung bislang nur teilweise den Empfehlungen des Deutschen Ethikrats und den Anforderungen des AI Acts gerecht wird. Ohne eine umfassende Umsetzung in Forschung, Entwicklung, Implementierung und Aufsicht drohen nach Auffassung der Fragesteller bestehende Gefährdungslagen fortzubestehen oder sich gar zu verschärfen.

Vor diesem Hintergrund ist es nach Meinung der Fragesteller von dringendem öffentlichen Interesse, zu klären, welche legislativen, organisatorischen und haushalterischen Maßnahmen die Bundesregierung zur Sicherstellung der Kontrolle über selbstmodifizierende und autonom agierende KI-Systeme plant und bereits implementiert hat.

1. Liegen der Bundesregierung Informationen über die im Artikel von futurezone beschriebenen Tests mit Claude Opus 4 und den dort beobachteten Verhaltensweisen vor (vgl. Vorbemerkung der Fragesteller)?

Der Bundesregierung liegen hierzu keine weiterführenden Informationen vor.

Allgemein bezieht die Bundesregierung aus verschiedenen Quellen Informationen zu Fähigkeiten, Risiken und Chancen sogenannter general-purpose AI Modelle, die ein breites Spektrum an Aufgaben erledigen können. Unter anderem fließen dabei die Ergebnisse des International AI Safety Report (abrufbar unter: www.gov.uk/government/publications/international-ai-safety-report-2025/international-ai-safety-report-2025) als weltweit erster umfassender Bericht zur Sicherheit fortgeschrittener Künstlicher Intelligenz ein. Dieser Bericht fasst den aktuellen Stand der wissenschaftlichen Erkenntnisse zusammen, um eine gemeinsame internationale Wissensbasis zu schaffen und eine fundierte Diskussion über die Risiken und deren Bewältigung zu ermöglichen.

2. Welche Maßnahmen sind seitens der Bundesregierung ggf. geplant, um zu verhindern, dass bereits implementierte KI-Modelle durch selbstschutzorientiertes Verhalten in der Praxis unkontrollierbar werden?

3. Plant die Bundesregierung, gesetzliche Grundlagen zu schaffen, die eigenmächtige Veränderungen an Codes durch KI-Modelle unterbinden oder sanktionieren?

22. Plant die Bundesregierung gesetzliche Anpassungen, um bestehende Grundrechterisiken beim Einsatz hochautonomer KI präventiv zu adressieren?

27. Wird seitens der Bundesregierung geprüft, ein Moratorium für KI-Entwicklungen mit Selbstmodifikationspotenzial zu erlassen, bis umfassende gesetzliche und ethische Rahmenwerke etabliert sind?

Die Fragen 2, 3, 22 und 27 werden gemeinsam beantwortet.

Die Regulierung von KI-Modellen erfolgt durch die EU KI-Verordnung 2024/1689 (europäische KI-Verordnung, im Folgenden: KI-VO), die am 1. August 2024 in Kraft getreten ist, EU-weit harmonisierte Regeln einführt und einen risikobasierten Ansatz verfolgt. Danach gelten ab dem 2. August 2025 auch bestimmte Anforderungen an Anbieter von KI-Modellen (Transparenzvorgaben, Vorgaben zur Einhaltung des Urheberrechts sowie unter bestimmten Voraussetzungen Risikomanagementanforderungen), wenn diese einen allgemeinen Verwendungszweck haben, sogenannte general-purpose AI Modelle.

Unter diesen Begriff fällt gemäß Artikel 3 Nummer 63 KI-VO „ein KI-Modell – einschließlich der Fälle, in denen ein solches KI-Modell mit einer großen Datenmenge unter umfassender Selbstüberwachung trainiert wird,

- das eine erhebliche allgemeine Verwendbarkeit aufweist und
- in der Lage ist, unabhängig von der Art und Weise seines Inverkehrbringens ein breites Spektrum unterschiedlicher Aufgaben kompetent zu erfüllen und
- das in eine Vielzahl nachgelagerter Systeme oder Anwendungen integriert werden kann,
- ausgenommen KI-Modelle, die vor ihrem Inverkehrbringen für Forschungs- und Entwicklungstätigkeiten oder die Konzipierung von Prototypen eingesetzt werden“.

Zusätzliche Regelungen bestehen für solche general-purpose AI Modelle, bei denen ein systemisches Risiko besteht, z. B. weil Fähigkeiten mit hohem Wirkungsgrad angenommen werden, insbesondere wenn die für das Training verwendeten Rechenleistung kumulativ mehr als 10^{25} Gleitkommaoperationen (Floating Point Operations Per Second, kurz: FLOPs) beträgt.

Für die Aufsicht über diese general-purpose AI Modelle ist das KI-Büro der Europäischen Kommission direkt zuständig (Artikel 88 Absatz 1 KI-VO) und kann gemäß Artikel 89 die erforderlichen Maßnahmen zur Einhaltung der KI-VO ergreifen. Aus dem Beschluss der Kommission zur Einrichtung des KI-Büros ergibt sich ferner, dass es gerade auch für die Beobachtung der Entwicklung der KI-Märkte und -Technologien sowie speziell für die Beobachtung des Auftretens unvorhergesehener Risiken, die sich aus KI-Modellen mit allgemeinem Verwendungszweck ergeben, zuständig ist und hierzu an die Mitgliedstaaten berichtet.

Nationale Vorgaben zu KI-Modellen sind neben dieser umfassenden unionsrechtlichen Regulierung nicht vorgesehen.

4. Welche Anforderungen bestehen derzeit an von der Bundesregierung geförderte KI-Projekte hinsichtlich der Verhinderung selbstmodifizierenden Verhaltens?

Ein Großteil der von der Bundesregierung geförderten KI-Projekte fällt in den Bereich Forschung und Entwicklung und unterliegt damit Artikel 2 Absatz 6 und 8 KI-VO. Für alle weiteren Projekte gelten die Anforderungen der genannten Verordnung.

5. Wird eine zentrale Prüfinstanz zur Überwachung von KI-Systemen mit Selbstmodifikationsfähigkeiten eingerichtet?

29. Bestehen Überlegungen der Bundesregierung, KI-Systeme, die menschenähnliche Entscheidungsfähigkeiten oder Autonomien entwickeln, einer gesonderten Genehmigungspflicht zu unterwerfen?

30. Plant die Bundesregierung eine Berichtspflicht für alle Betreiber von KI-Systemen, die autonome Lern- oder Modifikationsmechanismen einsetzen?

Die Fragen 5, 29 und 30 werden gemeinsam beantwortet.

Die Zuständigkeit der Marktüberwachungsbehörden für KI-Systeme nach der KI-VO werden im diesbezüglichen Durchführungsgesetz festgelegt. Der Entwurf dazu befindet sich derzeit in der Ressortabstimmung.

6. Hat die Bundesregierung Überlegungen angestellt oder sich Rat eingeholt zu den Risiken für Grundrechte (z. B. Datenschutz, Schutz der Privatsphäre) durch autonome, selbstmodifizierende KI-Systeme (wenn ja, bitte ausführen)?

Die Betrachtung der Risiken für die genannten Grundrechte ist in die europäische Gesetzgebung eingeflossen. Im gesamten Prozess (Gesetzgebung und Durchführung) tauscht sich die Bundesregierung fortlaufend mit Stakeholdern und Wissenschaftlern aus. Artikel 27 KI-VO regelt im Übrigen die Grundrechtefolgeneinschätzung beim Einsatz von KI.

7. Werden die ethischen Empfehlungen des Deutschen Ethikrats systematisch in die Entwicklung der deutschen KI-Strategie integriert, und wenn ja, inwieweit?

Die Bundesregierung beachtet bei der Entwicklung und Umsetzung der KI-Strategie und ihrer Fortschreibung eine Vielzahl von Empfehlungen aus Wissenschaft, Wirtschaft, Politik und Zivilgesellschaft und passt ihre Initiativen entsprechend an.

8. Bestehen seitens der Bundesregierung internationale Kooperationen, um global gültige Standards für die Kontrolle von selbstmodifizierenden KI-Systemen zu entwickeln?

Die Bundesregierung erachtet die internationale Zusammenarbeit im Bereich der künstlichen Intelligenz für essentiell, um gemeinsam gültige Standards zu setzen und gleichzeitig die Innovationsfähigkeit zu stärken. Die Bundesregierung setzt sich daher in allen internationalen Formaten für eine internationale KI-Governance ein, die Innovation fördert und zugleich gewährleistet, dass KI menschenzentriert, sicher, vertrauenswürdig, gemeinwohlorientiert und nachhaltig ist und die Rechte Dritter, u. a. Rechte am geistigen Eigentum, achtet.

Dazu strebt die Bundesregierung an, die politischen Rahmenbedingungen mit Stakeholdern und internationalen Partnern weiterzuentwickeln und so geeignete Lösungsansätze zu finden.

Eine international verbindliche KI-Regulierung existiert bislang nicht. Mit der KI-Konvention des Europarats, die am 5. September 2024 u. a. von EU, USA, UK und Israel unterzeichnet wurde, wurde jedoch der erste völkerrechtlich verbindliche Standard für KI gesetzt. Zudem wurde in den letzten Jahren eine Reihe von politischen Erklärungen verabschiedet, u. a. im Rahmen der Vereinten Nationen einschließlich UNESCO, G7, G20, der OECD, der Global Partnership on AI sowie der AI Summit Serie seit 2023. Die Erklärungen postulieren (rechtlich unverbindlich) Werte und skizzieren bestimmte Prinzipien und Verhaltensanforderungen. Ihre Inhalte können aber durchaus als gemeinsame Basis dienen und in verbindliche rechtliche Regelungen einfließen, wie z. B. die KI-Prinzipien und die KI-Definition der OECD in die KI-Konvention des Europarats und die EU KI-Verordnung.

Konkret verfolgen die G7-Staaten im Rahmen des Hiroshima-Prozesses den Ansatz, KI-Entwickler und Anwender über Selbstverpflichtungen zur Einhaltung von Prinzipien und Standards zu bewegen. Beim G7-Digitalministertreffen am 1. Dezember 2023 wurde das „Hiroshima AI Process Comprehensive Policy Framework“ verabschiedet. Wesentliches Element ist der internationale Code of Conduct für Entwickler fortgeschrittener KI-Systeme (Code of Conduct for Organizations Developing Advanced AI Systems) – die ersten inter-

nationalen Leitplanken für fortgeschrittene KI-Systeme, einschließlich generativer KI.

Speziell zur öffentlichen Diskussion zu den Potenzialen und Risiken von hochleistungsfähiger generativer KI fand bereits im November 2023 unter deutscher Beteiligung der „AI Safety Summit“ in Bletchley Park (GBR) statt, aus dem auch der „AI Safety Report“ hervorging (s. o. Frage 1). Bei dem im Mai 2024 von GBR und KOR gemeinsam ausgerichteten „AI Seoul Summit“ (Abschlusserklärung unter https://aiseoulsummit.kr/press/?uid=41&mod=document&pageid=1) erklärten die teilnehmenden Staaten ihre Absicht, die Zusammenarbeit bei der Sicherheitsforschung in einem AI Safety Network zu fördern. Neben einer Zusammenarbeit von Sicherheitsinstituten mit KI-Bezug bezieht dies auch Forschungsprogramme und andere relevante Institutionen wie Aufsichtsbehörden mit ein, die einen Beitrag zu einer solchen internationalen Netzwerkarbeit leisten können.

9. Welche Vorkehrungen trifft die Bundesregierung ggf., um die Möglichkeit der illegalen Nutzung von KI-Systemen (z. B. für Darknet-Aktivitäten) einzudämmen?

Die Bundesregierung bekämpft die illegale Nutzung von KI durch schärfere Gesetze wie § 126a StGB (Strafbarkeit des Betreibens krimineller Handelsplattformen) und das IT-Sicherheitsgesetz 2.0. Zudem setzt sie auf den Ausbau von Cyber-Ermittlungseinheiten, eigene KI-Systeme und internationale Kooperationen zur Bekämpfung von Darknet-Plattformen. Im Übrigen werden Zweckverbote und Pflichten für Anbieter und Betreiber in der KI-VO geregelt.

10. Welche Haushaltsmittel sind im Bundeshaushalt 2025 und 2026 für Forschung zur KI-Sicherheit, insbesondere zur Verhinderung autonomer Risikoverhaltensweisen, vorgesehen?

Die bereitgestellten Mittel werden im Rahmen der laufenden, noch nicht abgeschlossenen Haushaltsaufstellungsverfahren festgelegt.

11. Wie plant die Bundesregierung sicherzustellen, dass KI-Systeme in Deutschland nicht ohne transparente Überwachung eingesetzt werden dürfen?

Die Überwachung von KI-Systemen ist in der KI-VO geregelt. Im Übrigen wird auf die Antwort zu Frage 5 verwiesen.

12. Welche konkreten Förderprogramme existieren ggf. derzeit auf Bundesebene, die sich explizit auf die Entwicklung von sogenannten Safe-by-Design-KI-Systemen beziehen, bei denen Transparenz, Erklärbarkeit, Sicherheitsmechanismen und ethische Prinzipien bereits in der Entwicklungsphase systematisch berücksichtigt werden?

Entsprechende Forschung wird unter anderem an den KI-Kompetenzzentren, am Weizenbaum-Institut für die vernetzte Gesellschaft, in Projekten der Förderrichtlinie „Sichere Zukunftstechnologien in einer hypervernetzten Welt: Künstliche Intelligenz“ im Rahmen des Forschungsrahmenprogramms der Bundesregierung zur IT-Sicherheit „Digital. Sicher. Souverän.“ sowie in vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) geförderten Nachwuchsforschungsgruppen zu KI vorangetrieben. Auch innerhalb der vom

BMFTR geförderten Plattform Lernende Systeme (PLS) werden entsprechende Fragestellungen bearbeitet.

13. In welchem Umfang wurden in den Jahren 2023 und 2024 ggf. Haushaltsmittel für interdisziplinäre Forschungsprojekte zur sicheren und ethisch verantwortungsvollen KI-Entwicklung bereitgestellt?

Im Jahr 2023 wurden ca. 9,4 Mio. Euro und im Jahr 2024 ca. 7,9 Mio. Euro bereitgestellt. Hinzu kommen Haushaltsmittel in Höhe von 50 Mio. Euro pro Jahr für die institutionelle Förderung von fünf universitären KI-Kompetenzzentren, an denen ebenfalls entsprechende Forschung betrieben wird.

14. Plant die Bundesregierung, den gezielten Ausbau von Forschungsclustern, die technische, ethische, rechtliche und gesellschaftliche Fragestellungen der KI gemeinsam zu adressieren, und wenn ja, in welchem zeitlichen Rahmen und mit welchem Budget?

Die Planungen der Bundesregierung im Bereich der KI-Forschung sind Gegenstand laufender Abstimmungsprozesse.

15. Werden zivilgesellschaftliche Akteure, Ethikräte, Fachverbände und wissenschaftliche Institutionen bei der Ausgestaltung und Förderung entsprechender Forschungsprogramme beteiligt, wenn ja, inwiefern, und wie viele Planstellen und Ressourcen wurden den für die Marktaufsicht nach dem EU AI Act zuständigen Behörden ggf. bislang zugewiesen, um Konformitätsprüfungen und Audits von Hochrisiko-KI-Systemen sachgerecht durchführen zu können?

Bei der Ausgestaltung und Umsetzung von Forschungsprogrammen werden regelmäßig Stakeholder aus Wissenschaft, Wirtschaft und Zivilgesellschaft eingebunden, etwa im Rahmen von Fachgesprächen oder Expertenbegutachtungen. Die konkrete Form der Einbindung ist dabei abhängig von Inhalt und Zielen der Forschungsprogramme. Im Übrigen wird auf die Antwort zu Frage 5 verwiesen.

16. Welche Behörden sind auf Bundesebene für die Durchführung der Marktüberwachung und Konformitätsbewertung gemäß dem AI Act konkret benannt?

17. Welche Fortbildungs- und Qualifizierungsmaßnahmen werden ggf. aktuell angeboten oder sind geplant, um das Fachpersonal dieser Behörden auf die neuen Anforderungen vorzubereiten?

18. Wie stellt die Bundesregierung sicher, dass die nationalen Marktüberwachungsbehörden mit den europäischen Aufsichtsstrukturen effektiv kooperieren, und welche Bewertung nimmt die Bundesregierung derzeit hinsichtlich KI-Systemen vor, die zu selbstständiger Codeänderung oder autonomer Verhaltensanpassung fähig sind, insbesondere im Hinblick auf Artikel 5 (verbotene Praktiken) und Artikel 6 ff. (Hochrisiko-KI) des EU AI Acts?

Die Fragen 16 bis 18 werden gemeinsam beantwortet.

Es wird auf die Antwort zu Frage 5 verwiesen.

Im Übrigen schafft das sich im Aufbau befindende Beratungszentrum für Künstliche Intelligenz (BeKI) eine zentrale Anlauf- und Koordinierungsstelle für KI-Vorhaben in der Bundesverwaltung. Ein Schwerpunkt des BeKI liegt im nachhaltigen KI-Kompetenzaufbau für den kompetenten Umgang mit KI in der Bundesverwaltung in Zusammenarbeit mit der Bundesakademie für öffentliche Verwaltung im Bundesministerium des Innern (BAköV). Die BAköV als die zentrale Fortbildungseinrichtung des Bundes bietet bereits seit 2023 umfangreiche Fortbildungen und Sensibilisierungen im Bereich KI und KI-VO für die Bundesbeschäftigten an.

19. Plant die Bundesregierung, ergänzende nationale Leitlinien zur Abgrenzung zwischen verbotenen Praktiken und Hochrisiko-KI bei adaptiven, selbstmodifizierenden KI-Systemen zu erarbeiten?

20. Wenn die Frage 19 mit ja beantwortet wurde, mit welchen Akteuren und Institutionen werden diese Leitlinien entwickelt, wann ist mit der Vorlage erster Ergebnisse zu rechnen, und gibt es derzeit ein systematisches Monitoring auf Bundesebene, das fortlaufend prüft, inwieweit Grundrechte durch den Einsatz von KI, insbesondere von hochautonomen KI-Systemen, berührt werden?

Die Fragen 19 und 20 werden gemeinsam beantwortet.

Die europäische Kommission (KI Büro) bereitet derzeit Leitlinien zu den sogenannten Hochrisikobereichen der KI-VO (Artikel 6 KI-VO) vor. Die Bundesregierung bringt sich in den Erarbeitungsprozess aktiv ein. Auf die bereits veröffentlichten Leitlinien der europäischen Kommission zu verbotenen Praktiken im Sinne der KI-VO (Artikel 5 KI-VO) wird hingewiesen.

21. Werden unabhängige Ethikräte, Datenschutzbehörden und zivilgesellschaftliche Organisationen in diese Monitoringprozesse eingebunden, und wenn ja, inwiefern?

Es wird auf die Antwort zu Frage 5 verwiesen. Die Bundesregierung stellt die angemessene Beteiligung aller Stakeholder sicher.

23. Wird die Bundesregierung Maßnahmen ergreifen, um die Öffentlichkeit besser über die Risiken und Kontrollmechanismen von KI-Systemen zu informieren?

Die Bundesregierung plant, die Öffentlichkeit angemessen über die Risiken und Chancen sowie Kontrollmechanismen von KI-Systemen zu informieren.

24. Wie wird gewährleistet, dass selbstmodifizierende KI-Systeme keine Auswirkungen auf kritische Infrastrukturen haben?

Der Einsatz von KI im Bereich kritischer Infrastruktur wird in der KI-VO geregelt und dem Hochrisikobereich mit entsprechend strengen Anforderungen zugeordnet.

25. Sieht die Bundesregierung Anpassungsbedarf am bestehenden IT-Sicherheitsgesetz, um die neuen Gefährdungslagen durch KI adäquat abzudecken, und wenn ja, inwiefern?

Die Bundesregierung sieht die Notwendigkeit aufgrund der steigenden Bedrohungslage im Cybersicherheitsbereich, das Bundesamt für Sicherheit in der Informationstechnik (BSI) als die nationale Cybersicherheitsbehörde des Bundes weiter zu stärken. So muss das BSI mit neuen Kompetenzen ausgestattet werden, um die Cybersicherheit von KI-Systemen wirksam und effektiv zu überwachen. Die Bundesregierung prüft eine entsprechende Anpassung des BSIG im Rahmen des Durchführungsgesetz für die KI-VO.

26. Gibt es Planungen zur Einführung eines verpflichtenden „Sandboxings“ für experimentelle KI-Systeme, um die unkontrollierte Ausbreitung autonomer Funktionen zu verhindern?

Die Einrichtung von Reallaboren wird in der KI-VO geregelt.

28. Ist eine umfassende Risikoanalyse durch unabhängige ethische Kommissionen bei der Zulassung neuer KI-Modelle geplant, und wenn ja, inwieweit?

KI-Modelle mit allgemeinem Verwendungszweck unterliegen nach den Regelungen der KI-VO keiner Zulassung, sondern einer Meldepflicht, sofern es sich um KI-Modelle mit systemischem Risiko handelt (Artikel 51 Absatz 1 KI-VO). Im Praxisleitfaden für KI-Modelle mit allgemeinem Verwendungszweck (https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai) sind für KI-Modelle mit systemischen Risiken zudem externe Evaluationen vorgesehen.

Gesamtherstellung: H. Heenemann GmbH & Co. KG, Buch- und Offsetdruckerei, Bessemerstraße 83–91, 12103 Berlin, www.heenemann-druck.de
Vertrieb: Bundesanzeiger Verlag GmbH, Postfach 10 05 34, 50445 Köln, Telefon (02 21) 97 66 83 40, Fax (02 21) 97 66 83 44, www.bundesanzeiger-verlag.de
ISSN 0722-8333