

Sichere Anwendung künstlicher Intelligenz in Fahrzeugen

Vorstandsbeschluss vom 06.10.2026 auf der Basis einer Empfehlung des Vorstandsausschusses Fahrzeugtechnik

Einführung

Künstliche Intelligenz (KI) ist im Straßenverkehr längst keine Zukunftsvision mehr, sondern bereits heute Bestandteil vieler Fahrzeuge. Zum Einsatz kommt KI bspw. bei der Verkehrszeichen- und Fußgängererkennung oder der Erkennung von Ablenkung und Müdigkeit der Fahrenden. KI verarbeitet Daten mithilfe mathematischer Modelle und gibt darauf basierend Empfehlungen oder Steuerungsimpulse.

Für höhere Automatisierungsstufen (SAE Level 2 und darüber) werden leistungsfähige KI-Methoden benötigt, die komplexe Verkehrssituationen selbstständig erfassen und bewerten. Anders als klassische regelbasierte Systeme treffen solche lernenden Systeme ihre Entscheidungen auf Basis von Mustern, die sie aus großen Datenmengen ableiten. Ihre Ergebnisse sind zwar eindeutig bestimmt, für den Menschen jedoch häufig schwerer nachvollziehbar.

Darüber hinaus ändert sich für höhere Automatisierungsstufen der Regelkreis Mensch-Fahrzeug-Umwelt und somit die Risiken für die Verkehrssicherheit. Mit zunehmender Fahrzeugautomatisierung verändern sich die Anforderungen an die Fahrkompetenz. Die Aufgaben der Fahrzeuglenkenden verschieben sich vom aktiven Fahren hin zur Überwachung der Assistenzsysteme. Der Mensch bleibt jedoch ein notwendiger Sicherheitsbaustein für Level-2-Systeme. Die Fahrausbildung sollte daher zukünftig nicht nur Fahrzeugbedienung vermitteln, sondern auch den kompetenten Umgang mit Fahrerassistenz- und Automatisierungssystemen schulen. Entscheidend ist dabei, dass die Lernenden die Bedienung und die Grenzen der Funktionen sowie ihre eigene Verantwortung kennen und verstehen. Der Einsatz von Fahrsimulatoren in der Fahrausbildung ist weiter voranzutreiben.

Mit diesem wachsenden Funktionsumfang gehen steigende Anforderungen an die Vertrauenswürdigkeit und Sicherheit von KI-Systemen einher. Die europäische KI-Verordnung schafft hierfür erstmals einen verbindlichen Rechtsrahmen und stellt insbesondere an sogenannte Hochrisiko-KI-Systeme, etwa vollautomatisierte Fahrfunktionen oder sicherheitsrelevante Assistenzsysteme, strenge Anforderungen. Gleichzeitig zeigen sich beim Einsatz von KI im Fahrzeug neue Risiken, wie fehlerhafte oder unausgewogene Trainingsdaten, eingeschränkte Erklärbarkeit (für den Kunden) und Nachprüfbarkeit (für die Technischen Dienste) von Entscheidungen sowie neue Angriffsflächen für Cyberangriffe, etwa durch gezielt manipulierte Sensordaten oder verfälschte Trainingsdaten.

Auch das bestehende Verkehrsrecht stößt an seine Grenzen. Viele Vorschriften der Straßenverkehrsordnung sind auf menschliches Verhalten und menschliches Situationsverständnis zugeschnitten und lassen sich nicht ohne Weiteres eindeutig auf KI-Systeme übertragen.

Die sichere Integration von KI im Fahrzeugsektor setzt eine Kombination aus klaren Regulierungen, unabhängiger Prüfung und harmonisierten Testverfahren voraus. Für eine sichere und transparente Einführung von KI im Straßenverkehr sind insbesondere folgende Maßnahmen zu ergreifen:

- Einheitliche Definitionen und ein klarer Rechtsrahmen für sicherheitskritische KI im Einklang mit der UNECE-Perspektive und der europäischen KI-Verordnung
- Verpflichtende Zertifizierungsverfahren sowie periodische Genehmigungsverfahren für Änderungen an allen sicherheitskritischen KI-Modellen analog der UN-Regelung Nr. 156, notwendige sicherheitsrelevante Softwareupdates müssen aber ohne vermeidbare Verzögerung in die Fahrzeugflotte gelangen.
- Aufbau einer europäischen, unabhängigen Infrastruktur für KI-Daten und Simulationen zur Validierung und Erprobung von KI-Systemen
- Integration von KI-Prüfungen in die bestehenden Verfahren der Fahrzeugtypgenehmigung sowie der Softwarestände in die Hauptuntersuchung.

Empfehlungen

Vor diesem Hintergrund empfiehlt der Deutsche Verkehrssicherheitsrat (DVR) e.V. den deutschen und europäischen Gesetz- und Verordnungsgebern sowie den Fahrzeugherstellern und Technischen Diensten die Umsetzung folgender Maßnahmen:

1. Harmonisierung von KI-Definitionen und Klassifizierungen

Hochrisiko-KI-Systeme, die als sicherheitskritisch gelten, müssen in der Fahrzeugtypgenehmigung eindeutig definiert und reguliert werden. Dabei sollten die Begriffe aus der UNECE-Resolution, wie Modelldrift, Verifizierbarkeit, Robustheit und Vorhersagbarkeit, berücksichtigt werden. Zudem ist eine klare Unterscheidung zwischen symbolischer KI (regelbasierte Systeme) und subsymbolischer KI (z.B.

neuronale Netze)¹ erforderlich. Da Letztere keine prüfbaren Zwischenschritte offenlegen, bedarf es hierfür besonderer Prüf- und Sicherheitsmechanismen.

2. Einführung von Prüf- und Genehmigungsprozessen

Der DVR empfiehlt, für sicherheitskritische KI-Systeme im Straßenverkehr harmonisierte Prüf- und Genehmigungsverfahren einzuführen. Vor dem Inverkehrbringen sind die Sicherheit und die Wirksamkeit des Gesamtsystems durch die Typgenehmigungsbehörden mit geeigneten Methoden zu beurteilen.

Genehmigungsrelevante Änderungen, insbesondere sicherheitskritische Software- und Modellupdates, müssen vor ihrer Aktivierung einem geregelten Änderungs-Genehmigungsverfahren unterliegen.²

Die Verfahren sind so auszugestalten, dass unabhängige Prüfungen der Softwarestände auch im Rahmen der regelmäßigen Hauptuntersuchung (HU) bzw. während der Nutzungs- und Betriebsdauer der Systeme gewährleistet sind. Die Vorschriftenentwicklung sollte so erfolgen, dass sowohl die Typgenehmigung als auch die periodische Fahrzeugüberwachung konsequent mitbetrachtet werden.

3. Sicherstellung der Datenqualität und Datenneutralität

Die Qualität der Trainingsdaten hat direkte Auswirkungen auf die Sicherheit von KI-Systemen. Daher ist es notwendig, eine unabhängige Datenprüfung der verwendeten Daten zu ermöglichen. Trainings- und Testdaten³ müssen aus repräsentativen, realistischen Fahrsituationen gewonnen werden, um zu verhindern, dass Fehler aus KI-generierten Daten durch die KI weitergegeben werden. Datensätze müssen repräsentativ, realistisch und divers sein, um unfaire oder diskriminierende KI-Entscheidungen (Modellbias bzw. Modellverzerrungen) zu verhindern. Validierungs- und Trainingsdaten müssen unabhängig voneinander sein. Dies ist nachzuweisen.

Anforderungen an Trainingsdaten müssen bereits in der Entwicklung systematisch festgelegt und umgesetzt werden. Hierzu sind geeignete Qualitätscharakteristiken

¹ Symbolische KI bildet Wissen in expliziten, lesbaren Regeln ab; der Entscheidungsweg ist unmittelbar nachvollziehbar. Subsymbolische KI nutzt eine implizite Kodierung von Informationen, die aus Trainingsdaten abgeleitet werden, sodass die Prüfung nur am beobachtbaren Verhalten ansetzen kann.

² Siehe z.B. Genehmigungsprozess gemäß UN-Regelung Nr. 155 und Nr. R156

³ Testdaten werden beim Test der ML-Komponente verwendet, um deren Qualität anhand bislang ungesehener und von Trainings- und Validierungsdaten unabhängiger Daten zu messen. Trainingsdaten werden während des Modelltrainings und der Modellabstimmung verwendet, um die optimalen Gewichte für eine gegebene neuronale Netzwerkstruktur zu bestimmen.

heranzuziehen, wie z.B. Genauigkeit, Vollständigkeit, Repräsentativität, Nachvollziehbarkeit und Überprüfbarkeit. Diese Kriterien müssen bei der Datenerhebung, -aufbereitung, -kennzeichnung und -verwendung berücksichtigt werden, um sicherzustellen, dass die Trainingsdaten für den vorgesehenen Einsatzbereich geeignet sind und die Sicherheit des KI-Systems unterstützen.

Neben der Qualität ist die Regelkonformität der Trainingsdaten zu berücksichtigen. Regelwidrige Inhalte, beispielsweise verkehrsrechtswidrige Fahrmanöver in aufgezeichneten Fahrdaten, können andernfalls unbemerkt in das Systemverhalten übernommen werden und sind im nachgelagerten Testen, insbesondere in seltenen Randfällen (Edge Cases), nur eingeschränkt erkennbar. Dies gilt gleichermaßen für synthetisch erzeugte oder simulierte Daten, da auch diese Daten regelwidriges Verhalten abbilden können.

Das System darf regelwidriges Verhalten nicht als erwünschtes Zielverhalten erlernen. Wenn ein KI-System im Straßenverkehr sicher und rechtskonform handeln soll, müssen auch die Trainingsdaten daraufhin geprüft werden, ob sie regelkonformes oder regelwidriges Verhalten enthalten.

Erforderlich ist demnach nicht die pauschale Entfernung solcher Daten, sondern ein transparenter Umgang mit ihnen: Trainingsdaten müssen hinsichtlich ihrer Regelkonformität bewertet und gekennzeichnet werden, beispielsweise durch Labels oder Bewertungssignale, sodass steuerbar ist, welches Verhalten das System erlernt.

Der Verordnungsgeber wird gebeten, hierfür einen entsprechenden Rechtsrahmen zu schaffen.

4. Gewährleistung der Sicherheit und Prüfbarkeit von KI-Systemen über die Fahrzeuglebensdauer

Die sichere Nutzung von KI-Systemen im Fahrzeug muss über den gesamten Lebenszyklus - von der Typgenehmigung bis zum Ende der Nutzungsdauer – im Zusammenwirken von Fahrzeugherstellern, Technischen Diensten und Genehmigungsbehörden gewährleistet werden.

Der DVR empfiehlt, die Sicherheit und Prüfbarkeit sicherheitskritischer KI-Systeme über die gesamte Nutzungs- und Betriebsdauer durch harmonisierte Verfahren sicherzustellen. Typgenehmigung, genehmigungsrelevante Software- und

Modelländerungen sowie die technische Überwachung (PTI/HU, perspektivisch CTI⁴) sind dabei als aufeinander abgestimmte Prüf- und Überwachungsstufen auszugestalten. Hierfür müssen sicherheitsrelevante Systemzustände, Auffälligkeiten sowie die eingesetzten Software- und Modellversionen über standardisierte Schnittstellen auslesbar und anhand geeigneter technischer Informationen durch unabhängige Stellen prüfbar sein.

Subsymbolische KI-Systeme für sicherheitskritische Fahrfunktionen müssen modular aufgebaut werden, damit sie prüfbare Zwischenergebnisse liefern. Monolithische bzw. nicht modulare Systeme dürfen nur eingesetzt werden, wenn ein unabhängiges, regelbasiertes Überwachungssystem ihre Ausgaben fortlaufend auf Plausibilität und Regelkonformität prüft. Bei Abweichungen muss dieses System sicher eingreifen können.

5. Weiterentwicklung von KI-Systemen

In sicherheitskritischen Funktionen sind datengetriebene Modellanpassungen während des bestimmungsgemäßen Fahrzeugbetriebs auszuschließen, da sie dazu führen können, dass sich Fahrzeuge mit gleicher Ausgangssoftware unterschiedlich entwickeln. Eine Änderung der Modellstruktur oder der lernbaren Parameter darf auf diesem Weg nicht stattfinden. Neue Modellversionen dürfen erst nach vollständiger Prüfung und formaler Freigabe aktiviert werden.

6. Verkehrsrecht

Für Fahrzeugnutzende muss verständlich sein, welche Fahrfunktionen KI nutzen.⁵

Das Bundesverkehrsministerium bzw. die Bundesregierung sollten sich bei der Europäischen Kommission dafür einsetzen, dass zeitnah die Anforderungen der europäischen KI-Verordnung in das Typgenehmigungsrecht überführt werden bzw. auf entsprechende UN-Regelungen verwiesen wird.

Landesdatenschutzbehörden sollten bei der Verarbeitung von personenbezogenen Daten zum Zwecke des KI-Trainings der Verkehrssicherheit im Rahmen der Grundrechtsabwägung die angemessene Bedeutung zukommen lassen.

⁴ Bei der kontinuierlichen technischen Überprüfung eines Fahrzeugs (CTI) werden während des Betriebs fortlaufend sicherheitsrelevante Komponenten und Funktionen überwacht. Ziel ist es, Fehler oder Verschleiß frühzeitig zu erkennen und die Verkehrssicherheit über die gesamte Fahrzeuglebensdauer sicherzustellen.

⁵ Z.B. beim Kauf, im Betriebshandbuch, etc.

 Erläuterungen**Erläuterung zu Punkt 1 - Harmonisierung von KI-Definitionen und Klassifizierungen**

Sicherheitskritische KI-Systeme im Fahrzeug müssen klar geregelt sein, weil Fehler dieser Systeme direkte Auswirkungen auf die Sicherheit von Fahrenden und anderen Verkehrsteilnehmenden haben können. Deshalb sollte bereits in der Fahrzeugtypgenehmigung eindeutig festgelegt werden, welche KI-Systeme besonders kritisch sind und welche Sicherheitsanforderungen für sie gelten. Dabei sollten unter anderem Veränderungen des Systems über die Zeit („Modelldrift“), seine Überprüfbarkeit, Widerstandsfähigkeit gegenüber Störungen („Robustheit“) und ein möglichst vorhersehbares Verhalten berücksichtigt werden.

Außerdem ist es wichtig, zwischen klassischen regelbasierten KI-Systemen und lernenden KI-Systemen, wie neuronalen Netzen zu unterscheiden. Lernende Systeme sind oft deutlich komplexer und ihre Entscheidungen schwerer nachzuvollziehen. Deshalb benötigen sie zusätzliche Prüfungen und Sicherheitsmaßnahmen, um auch unter unterschiedlichen Bedingungen zuverlässig und sicher zu funktionieren.

Erläuterung zu Punkt 2 - Einführung von Prüf- und Genehmigungsprozessen

Die Prüfung sollte sich nicht ausschließlich auf einzelne Komponenten oder Modelle beschränken, sondern das KI-System einschließlich seiner Einbindung in das Fahrzeug beziehungsweise in die Verkehrsinfrastruktur als Gesamtsystem berücksichtigen. Dabei sind insbesondere die Robustheit und Stabilität des Systems, die Nachvollziehbarkeit und Verifizierbarkeit sicherheitsrelevanter Entscheidungen sowie mögliche Veränderungen der Systemleistung zu bewerten.

Die Prüfverfahren müssen geeignet sein, Leistungsverschlechterungen, beispielsweise infolge von Modell- oder Konzeptdrift, zu erkennen. Zudem muss sichergestellt werden, dass sicherheitsrelevante Software- und Modelländerungen erst nach erfolgreicher Prüfung und Freigabe aktiviert werden.

Eine Harmonisierung der hierfür erforderlichen Test- und Bewertungsmethoden ist dringend erforderlich. Harmonisierte Testmethoden sollten an bestehenden Standards ausgerichtet werden, darunter ISO 26262 (funktionale Sicherheit), ISO 21434 (Cybersecurity), ISO/TR 4804 (szenarienbasiertes Testen), ISO PAS 8800, UNECE NATM (New Assessment/Test Method for Automated Driving) sowie SOTIF (ISO 21448) und UN-Regelung Nr. 185.

Erläuterung zu Punkt 3 - Sicherstellung der Datenqualität und Datenneutralität

Eine pauschale Bereinigung der Trainingsdaten um alle nicht-regelkonformen Inhalte ist weder praktikabel noch wünschenswert. Reale Daten enthalten naturgemäß auch regelwidriges Verhalten, etwa anderer Verkehrsteilnehmender, mit dem ein sicheres System umgehen können muss. Entscheidend ist daher die transparente Kennzeichnung der Regelkonformität, beispielsweise durch Labels, Merkmale oder Bewertungssignale, sodass im Trainingsprozess steuerbar ist, welches Verhalten übernommen werden soll. Die Konformitätsbewertung muss auf die zugrunde liegenden Regelwerke rückführbar sein.

Für die Gewinnung von Trainingsdaten müssen grundsätzlich die datenspezifischen Sicherheitseigenschaften nach ISO/PAS 8800, Abschnitt 11.4.3.2 (wie z.B. Genauigkeit, Vollständigkeit, Repräsentativität) angewendet werden. Begründetes „Tailoring“ ist möglich; die systematische Berücksichtigung dieser Properties erhöht jedoch die Qualität der Trainingsdaten und stärkt damit die Qualität und Sicherheit des KI-Systems.

Erläuterung zu Punkt 4 - Gewährleistung der Sicherheit und Prüfbarkeit von KI-Systemen über die Fahrzeuglebensdauer

Generell gilt die Prämisse: Unabhängige Stellen müssen die Einhaltung sicherheitsrelevanter Anforderungen anhand nachvollziehbarer, auf die zugrunde liegenden Rechtsquellen rückführbarer Kriterien prüfen können. Die Fahrzeughersteller, die Technischen Diensten und die Genehmigungsbehörden haben hierbei unterschiedliche Verantwortungen.

- Fahrzeughersteller müssen den Nachweis der fortdauernden Sicherheit ihrer KI-Systeme im Rahmen eines aktiven Field Monitoring führen, bei Bedarf durch Software-Update-Maßnahmen nachsteuern und die hierfür notwendigen technischen Informationen – Ausfälle und Fehlfunktionen, sicherheitsrelevante Auffälligkeiten sowie Software- und Modellversionen inklusive Integritätsmerkmalen – in standardisierter, technisch interpretierbarer Form bereitstellen. Für KI-Systeme gehören dazu auch Informationen über die Trainingsdatensätze. Werden Systeme in mehreren Ländern eingesetzt, ist die Konformität für jedes Einsatzland gesondert sicherzustellen.
- Technische Dienste benötigen eine Regulierung, die – aufbauend auf der laufenden UNECE-Perspektive – KI-Systeme in der Fahrzeugtypprüfung so gestaltet, dass sie auf Basis validierter und geprüfter Trainingsdaten angelern und überprüfbar bleiben. Die Überprüfbarkeit muss sowohl bei modularen als auch bei End-to-End-Systemen transparent, nachvollziehbar und vor allem reproduzierbar sein: gleiche Prüfbedingungen müssen zu gleichen, dokumentierten Prüfergebnissen führen. Ist

die Prüfung des Gesamtsystems aufgrund seiner Komplexität nicht möglich, sind Prüfansätze und messbare, transparente und nachvollziehbare Prüfkriterien für Einzelkomponenten zu entwickeln, die einen belastbaren Rückschluss auf die Sicherheit des Gesamtsystems zulassen. Die Prüfung durch Technische Dienste muss organisatorisch unabhängig von der Entwicklung erfolgen, um Bestätigungsfehler zu vermeiden und die Aussagekraft der Prüfergebnisse zu sichern. Systematische Dokumentation und durchgängige Rückverfolgbarkeit – von den Gefährdungs- und Sicherheitsanalysen über die Sicherheitszielen, die Trainingsdaten und Modellversionen bis zu den Prüfergebnissen – bilden die Grundlage, auf der Technische Dienste und Genehmigungsbehörden ihre Bewertung nachvollziehbar und auditierbar stützen können; dies gilt in besonderem Maße auch für extern bezogene KI-Komponenten und Trainingsdaten, um intransparente „Blackbox“-Abhängigkeiten in der Lieferkette zu vermeiden.

- Genehmigungsbehörden sollten gemeinsam mit Herstellern und Technischen Diensten harmonisierte Verfahren für die Typgenehmigung, Überwachung von Software- und Modelländerungen sowie die periodische technische Überwachung entwickeln. Ziel muss sein, dass die Sicherheit von KI-Systemen nicht nur zum Zeitpunkt der Genehmigung, sondern während der gesamten Betriebsdauer nachgewiesen werden kann. Dazu sind standardisierte Schnittstellen, geeignete Protokollierungsmechanismen und prüffähige Nachweise über Software- und Modellversionen erforderlich.

Die Zulässigkeit des Systemverhaltens bemisst sich am jeweils geltenden Recht des Einsatzortes und der konkreten Anwendung. Für die automatisierte Fahrzeugführung sind insbesondere die UN-Regelungen Nr. 157, Nr. 171 und Nr. 185, die EU-Durchführungsverordnung 2022/1426 sowie das nationale Verkehrsrecht maßgeblich; für das Fahrerzustandsmonitoring ergeben sich Anforderungen u. a. aus der Verordnung (EU) 2019/2144 sowie dem Datenschutzrecht. Cybersicherheitsanforderungen ergeben sich anwendungsübergreifend aus den UN-Regelungen Nr. 155 und Nr. 156 sowie ISO 21434. Übergreifend gelten die Anforderungen der KI-Verordnung an Hochrisiko-KI-Systeme.

Modulare Systeme zerlegen die Fahrfunktion in einzelne, prüfbare Verarbeitungsschritte. Dadurch lassen sich Fehler besser erkennen, Ursachen nachvollziehen und Sicherheitsanforderungen gezielt prüfen. Monolithische subsymbolische Systeme bilden Sensordaten unmittelbar auf ein Ergebnis (bspw. Fahrentscheidung) ab. Ihnen fehlen prüfbare Zwischenergebnisse. Ihre Bewertung beruht daher überwiegend auf statistischen Tests. Das begrenzt die Nachvollziehbarkeit und reicht für sicherheitskritische Funktionen allein nicht aus.

Erläuterung zu Punkt 5 - Weiterentwicklung von KI-Systemen

Sicherheitskritische KI-Systeme müssen im Fahrzeugbetrieb nachvollziehbar, reproduzierbar und sicher arbeiten. Deshalb sollte sich das Modell während des bestimmungsgemäßen Betriebs nicht aufgrund neu eingehender Daten verändern. In der Forderung wird der Ausdruck „datengetriebene Modellanpassung während des Fahrzeugbetriebs“ verwendet um Missverständnisse durch unpräzise oder oftmals falsch verwendete Begrifflichkeiten („autonomer“ Betrieb) zu vermeiden.

Mit diesem Ausdruck sind Änderungen an Modellparametern oder der Modellstruktur gemeint, die durch neue Betriebsdaten ausgelöst werden. Dazu gehören beispielsweise Online-Lernen, Continual Learning und föderiertes Lernen mit anschließender automatischer Übernahme eines neuen Modells. ISO/IEC 22989 beschreibt Continual Learning als eine schrittweise Aktualisierung eines KI-Modells mit neuen Daten.

Eine solche Anpassung kann dazu führen, dass sich Fahrzeuge mit ursprünglich gleicher Software unterschiedlich entwickeln. Dadurch werden Prüfungen, Fehleranalysen und die Bewertung der Systemleistung erschwert. Außerdem kann nicht mehr eindeutig nachvollzogen werden, welche Modellversion zum Zeitpunkt eines sicherheitsrelevanten Ereignisses aktiv war.

Die Nutzung von Verkehrsdaten soll deshalb nicht ausgeschlossen werden. Die Daten können zur Leistungsüberwachung, zur Erkennung von Modell- und Konzeptdrift sowie für die Entwicklung neuer Modellversionen verwendet werden. Vor dem Training müssen sie jedoch einen dokumentierten Prüf- und Aufbereitungsprozess durchlaufen. Dabei sind unter anderem Datenqualität, Herkunft, Repräsentativität, Annotation, Vollständigkeit und mögliche Verzerrungen zu prüfen. Diese Anforderungen ergeben sich insbesondere aus Artikel 10 der KI-Verordnung.

Eine neue Modellversion darf erst nach einer vollständigen Prüfung aktiviert werden. Dazu gehören die technische Verifikation, die Validierung unter repräsentativen Bedingungen, eine Robustheitsprüfung, die Dokumentation, die formale Freigabe sowie eine versionierte und rücksetzbare Ausbringung. Diese Anforderungen stehen insbesondere im Zusammenhang mit Artikel 9, 11, 12, 15 und 17 der KI-Verordnung.

Das gilt auch für föderiertes Lernen. Dass die Trainingsdaten auf mehrere Fahrzeuge verteilt sind, ändert nichts daran, dass das daraus entstehende Modell vor der Aktivierung geprüft und freigegeben werden muss.

Besonders kritisch sind Rückkopplungsschleifen. Werden die Ausgaben eines KI-Systems ohne Prüfung als neue Trainingsdaten oder Trainingslabels verwendet, können sich Fehler

und Verzerrungen verstärken. Artikel 15 Absatz 4 der KI-Verordnung fordert deshalb Maßnahmen gegen solche verzerrten Rückkopplungen.

Von der datengetriebenen Modellanpassung sind begrenzte, deterministische Parametrierungen zu unterscheiden. Dazu gehören beispielsweise Kalibrierwerte, Filterparameter oder Entscheidungsschwellen. Solche Anpassungen können zulässig sein, wenn ihre Auslösebedingungen, Wertebereiche und Wirkungen vorab festgelegt, technisch begrenzt, testbar, protokollierbar und rücksetzbar sind. Eine Änderung der Modellstruktur oder lernbarer Modellparameter darf dadurch nicht erfolgen.

Ändert ein Update die sicherheitsrelevante Funktion, die Robustheit, die Leistungsgrenzen oder die Zweckbestimmung des Systems, muss es einem dokumentierten Änderungs- und Freigabeverfahren unterzogen werden. Bei einer wesentlichen Veränderung kann nach Artikel 23 und Artikel 43 Absatz 4 der KI-Verordnung eine erneute Konformitätsbewertung erforderlich sein.

gez.

Manfred Wirsch
Präsident