

02.25

Lizenziert für IFA der DGUV.
Die Inhalte sind urheberrechtlich geschützt.

In Kooperation mit:



76. Jahrgang
Jahrgang 2025
ISSN 2199-7330
1424

www.SISdigital.de

sicher ist sicher

Ihr Manual zur Anlagensicherheit



Anlagensicherheit im Betrieb

Herausgegeben von
Prof. Dr. Dirk Sohn und Dr. Michael Au

Online informieren und bestellen:
www.ESV.info/23707



Produktsicherheit und Betriebs-
sicherheit 54
Arbeitsschutz und Gesundheits-
förderung 65

KI-basierte Assistenz-
systeme 75
Tod durch fehlerhafte
Ingenieurplanung 81

ESV ERICH
SCHMIDT
VERLAG



© Titelbilder: IFA der DGUV

MARCEL BECKERS · THOMAS BÖMER · THOMAS SEIFEN

KI-basierte Assistenzsysteme im Maschinenschutz als Chance für die Reduzierung von Arbeitsunfällen

KI-basierte Systeme, etwa zur Objekterkennung, können im Maschinenschutz herkömmliche Sicherheitstechnik erweitern und so zur Sicherheit beitragen. Aktuell ist es aus technischer und regulatorischer Sicht allerdings schwierig, einen rigorosen Sicherheitsnachweis für solche Systeme zu führen und damit KI in formalen Sicherheitsfunktionen einzusetzen. Um trotzdem schon von den neuen Möglichkeiten der Technologie zu profitieren, können Assistenzsysteme eingesetzt werden. Dieser Beitrag beschreibt Einsatzmöglichkeiten und Anforderungen an solche Systeme.

1. Einleitung

Künstliche Intelligenz (KI) sorgt aktuell in fast allen Bereichen für Aufsehen. Besonders in der Öffentlichkeit stehen dabei seit einiger Zeit sogenannte *Large Language Models* (LLMs), wie beispielsweise die Chatbots ChatGPT und Gemini. Gleichzeitig eröffnen sich durch den Einsatz von KI aber auch im Maschinenschutz neue Möglichkeiten. So gibt es z.B. für Systeme, die zur Objekterkennung eingesetzt werden können, eine Reihe von vorstellbaren Anwendungsmöglichkeiten bei der Absicherung von Gefahrenstellen, etwa durch die Erkennung von Personen oder einzelnen Körperteilen. Durch den Einsatz von Machine Learning (ML) Modellen, insbesondere

mit tiefen neuronalen Netzen (DNNs), ist diese Art der Objekterkennung in Bildern im Gegensatz zu klassisch programmierter Software auch bei einer deutlich höheren Diversität von Anwendungsfällen möglich. Dennoch ist es zumindest nach heutigem Stand der Technik schwierig, mit KI-Systemen einen Sicherheitsnachweis etwa nach den in der funktionalen Sicherheit verwendeten Normen ISO 13849-1 oder IEC 61508-1/2/3 zu führen. Die Gründe hierfür sind zahlreich und sollen nur kurz gestreift werden. Ein wichtiger Punkt ist hierbei, dass ein trainiertes, tiefes neuronales Netz nicht von Menschen „verstanden“ werden kann, d.h. es ist nicht möglich, nachzuvollziehen, was das Netz aus den Trainingsdaten

DIE AUTOREN

**Dr. Marcel Beckers**

Wissenschaftlicher Mitarbeiter im Referat „Intelligente technische Systeme und Arbeitswelt“, Abteilung „Unfallprävention: Digitalisierung – Technologien“ am Institut für Arbeitsschutz (IFA) der Deutschen Gesetzlichen Unfallversicherung
Marcel.Beckers@dguv.de

**Thomas Bömer**

Leiter des Referats „Intelligente technische Systeme und Arbeitswelt“, Abteilung „Unfallprävention: Digitalisierung – Technologien“ am Institut für Arbeitsschutz (IFA) der Deutschen Gesetzlichen Unfallversicherung
Thomas.Boemer@dguv.de

**Thomas Seifen**

Prüflaborleiter für Schutzeinrichtungen im Referat „Intelligente Technische Systeme und Arbeitswelt“, Abteilung „Unfallprävention: Digitalisierung – Technologien“ am Institut für Arbeitsschutz (IFA) der Deutschen Gesetzlichen Unfallversicherung
Thomas.Seifen@dguv.de

„gelernt“ hat. Dem kann sich nur angenähert werden, indem ein solches Netz mit ihm unbekannten Daten getestet wird, d.h. im Fall einer Objekterkennung die Vorhersage des Netzes gegen eine vorher festgelegte Grundwahrheit (*Ground Truth*) geprüft wird. Hierbei ist jedoch eine weitere Herausforderung, dass die Gesamtmenge an Testdaten auch bei einer großen Anzahl die Realität immer nur annäherungsweise abbilden kann. Es ist aber unbestritten, dass die technische Forschung und Entwicklung z.B. zu Methoden der erklärbaren KI (*explainable AI/XAI*), Metriken für Testabdeckung und Robustheit sowie Diversität und Redundanz von neuronalen Netzen voranschreiten. Eine ausführliche Übersicht und einen Einstieg in die Literatur zu bereits vorhandenen Möglichkeiten, aber auch bestehenden Lücken, bietet der Review-Beitrag [1]. Gleichzeitig entwickeln sich auch die Normung und Gesetzgebung weiter, so dass davon auszugehen ist, dass es zukünftig KI-spezifische Anforderungen auch für sicherheitsrelevante Anwendungen im Kontext funktionaler Sicherheit geben wird.

Aktuell bieten insbesondere so genannte Assistenzsysteme die Möglichkeit, von den Chancen, die KI für den Arbeitsschutz bietet, zu profitieren, ohne einen rigorosen Sicherheitsnachweis führen zu müssen. Auch solche Systeme sollten aber nach gewissen Anforderungen und Qualitätsstandards entwickelt werden. Daher haben deutsche Prüfstellen aus dem Verbund von DGUV Test, unter anderem auch das Institut für Arbeitsschutz (IFA) mit dem GS-BAU-70 „Grundsätze für die Prüfung und Zertifizierung von sicherheitsrelevanten Assistenzsystemen an Maschinen und Nutzfahrzeugen“ [2] einen Prüfgrundsatz für Assistenzsysteme erarbeitet. Dieser Beitrag beschreibt, wie sich die dort enthaltenen allgemeinen Anforderungen auch für die Entwicklung und Prüfung von KI-basierten Assistenzsystemen nutzen lassen und welche spezifischen Herausforderungen dabei bestehen. Außerdem wird ein kurzer Ausblick in die zukünftige Regulatorik aus Maschinenverordnung und KI-Verordnung für sicherheitsrelevante KI-Systeme gegeben.

2. Definition von Assistenzsystemen nach dem GS-BAU-70

Im Allgemeinen wird im Prüfgrundsatz GS-BAU-70 [2] sinngemäß die nachfolgende Definition für ein Assistenzsystem beschrieben.

„Ein Assistenzsystem ist ein Unterstützungssystem für die Benutzenden stationärer und mobiler Maschinen, das zu Bedienungsaufgaben eine assistierende Funktion leistet bzw. in Bedienungsaufgaben Fehlern und Fehlverhalten entgegenwirken soll.“

Da nicht alle Assistenzsysteme das gleiche Wirkungsziel haben, sind die nachfolgenden Klassen von Assistenzsystemen definiert:

Assistenzsystem Klasse A

Ein Assistenzsystem der Klasse A ist ein informierendes System mit indirekter Wirkung.

Assistenzsystem Klasse B

Ein Assistenzsystem der Klasse B ist ein warnendes System mit indirekter Wirkung, das mit Nachdruck informiert.

Assistenzsystem Klasse C

Ein Assistenzsystem der Klasse C ist ein assistierendes System mit direkter Wirkung, das vom Benutzenden übersteuerbar ist, falls dies seitens der Anwendung erforderlich ist. Ein solches System greift aktiv minimierend/begrenzend ein, macht dem Benutzenden ggf. einen Handlungsvorschlag und ist nicht Teil der Risikominderungsstrategie.

Assistenzsystem Klasse D

Ein Assistenzsystem der Klasse D ist ein teilautomatisiertes System mit direkter Wirkung, das vom Benutzenden übersteuerbar ist, falls dies seitens der Anwendung erforderlich ist. Ein solches System greift aktiv minimierend/begrenzend ein, ist Teil der Risikominderungsstrategie und kann z. B. einen Teil der Bedienungsaufgabe übernehmen.

Assistenzsystem Klasse E

Ein Assistenzsystem der Klasse E ist ein vollautomatisiertes System mit direkter Wirkung, das vom Benutzenden nicht übersteuerbar ist.

Für die hier betrachteten KI-basierten Assistenzsysteme sind besonders die Klassen A bis C relevant.

Der Prüfgrundsatz legt eine Reihe von Anforderungen an solche Assistenzsysteme fest. Hier sind insbesondere folgende Anforderungen wichtig:

- ▶ Das Assistenzsystem muss eine Zuverlässigkeit „innerhalb der von den Benutzenden vernünftigerweise zu erwartenden Grenzen“ aufweisen.
- ▶ Die Funktion und die Zuverlässigkeit müssen den Bedienenden transparent gemacht werden.
- ▶ Das Assistenzsystem muss so ausgelegt werden, dass es die vorhersehbare Anwendung an der Maschine berücksichtigt.

Als Beispiel für ein KI-basiertes Assistenzsystem, welches heute schon einen Anteil an der Reduzierung schwerer Unfälle leisten kann, dient ein kamerabasiertes System an einer Formatkreissäge.

Das Kamerasystem kann bei einer Annäherung der Hand an die Gefahrstelle einen sicheren Zustand einleiten. Um eine Hand zu erkennen, hat der Hersteller eine Software entwickelt, die mithilfe von KI Kamerabilder auf die Anwesenheit menschlicher Hände untersucht (s.a. Abb. 1). Sobald die KI eine oder mehrere Hände im

Kamerabild erkennt, die sich auf das Sägeblatt zubewegen und so eine gefährliche Situation diagnostiziert, wird das Sägeblatt in Bruchteilen von Sekunden unter den Arbeitstisch versenkt, um so einen Kontakt zwischen Mensch und Sägeblatt zu verhindern. Aus diesem sicheren Zustand kann die Maschine innerhalb kurzer Zeit wieder in den Arbeitszustand versetzt werden, und die Arbeiten lassen sich ohne Beschädigungen an der Formatkreissäge oder am zu verarbeitenden Material fortsetzen. Normativ ist für eine solche Formatkreissäge zum Schutz der Hände des Bedieners eine höhenverstellbare Schutzhaube vorgesehen. Diese ist in dem beschriebenen System nach wie vor vollständig enthalten.

Das System hat eine direkte Wirkung, leitet bei Aktivierung einen sicheren Zustand ein, aber gleichzeitig erfolgt formal kein Beitrag zur Risikominderung im Sinne der DIN EN ISO 12100 [3]. Die normativ geforderte Schutzeinrichtung ist weiterhin vorhanden. Dieses System ersetzt also nicht die vorhandene Schutzeinrichtung, sondern stellt einen zusätzlichen Schutz für die Bedienenden dar. Dadurch erfüllt das KI-basierte System die Definition eines Assistenzsystems der Klasse C.

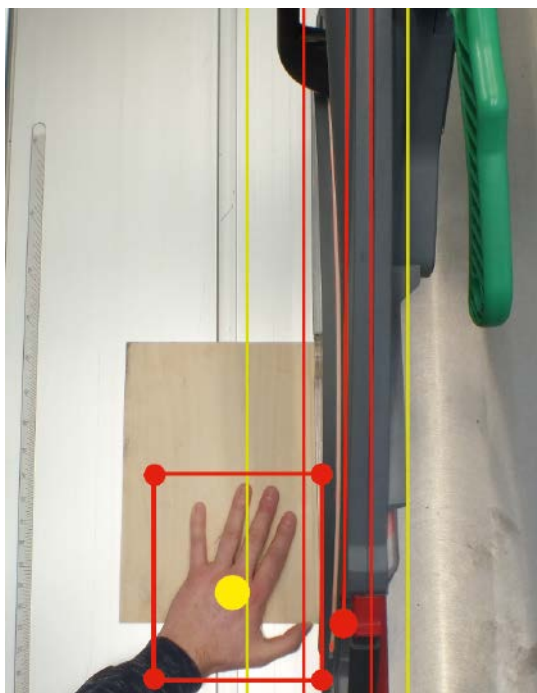


Abb. 1: Darstellung der Handerkennung durch ein KI-basiertes Assistenzsystem an einer Formatkreissäge © IFA der DGUV

In der betrieblichen Praxis stellt u.a. ein so genannter Anlauftest sicher, dass das System verlässlich funktioniert. Nach Start der Maschine muss es beide Hände der Bedienperson erkennen, andernfalls kann das Sägeblatt der Maschine nicht gestartet werden.

Das hier beschriebene Assistenzsystem wurde am Institut für Arbeitsschutz (IFA) geprüft.

3. Anforderungen an KI-basierte Assistenzsysteme

Da sich der GS-BAU-70 nicht an eine bestimmte Technologie, insbesondere auch nicht an KI-basierte Komponenten richtet, beinhaltet er auch keine konkreten technischen Anforderungen für diese. Es muss ausgeschlossen werden können, dass von Assistenzfunktionen neue Gefährdungen ausgehen, das gilt insbesondere auch beim Einsatz neuer Technologie wie KI. Außerdem müssen die durch den GS-BAU-70 festgelegten allgemeinen Anforderungen eingehalten werden, insbesondere was Zuverlässigkeit und Transparenz angeht. Um dies zu realisieren, sind KI-spezifische Ausfallmodelle relevant, welche die Zuverlässigkeit und Robustheit des KI-Systems beeinflussen können und somit einen direkten Einfluss auf dessen Sicherheit haben.

Robustheit ist eine neue Herausforderung im Zusammenhang mit KI-Systemen, da diese Systeme für sehr komplexe Aufgaben in komplexen Nutzungsumgebungen eingesetzt werden, die ein gewisses Maß an Unsicherheit beinhalten. Tiefe neuronale Netze stellen eine besonders schwierige Herausforderung dar, da sie sowohl schwer zu erklären sind als auch manchmal ein unerwartetes Verhalten zeigen. Mit der Robustheit eines KI-Systems ist in der Regel gemeint, dass dieses auch mit Anwendungsfällen zurechtkommt, die nicht exakt den im Training verwendeten Daten entsprechen. Es geht also um die Generalisierungsfähigkeit eines KI-Systems. Diese hat auch bei sehr weit fortgeschrittenen neuronalen Netzen und Trainingsmethoden immer Grenzen. Beispiele für bilderkennende Systeme sind etwa Robustheit gegenüber bestimmten Bildfehlern, Schwankungen in Helligkeit und Kontrast oder auch Umweltbedingungen wie Verschmutzungen. Eine Übersicht über verschiedene sogenannte „Corruptions“, gegen die ein bilderkennendes KI-System getestet werden kann, findet sich z. B. in Ref. [4]. Um so weit wie möglich sicherzustellen, dass ein KI-System wie erwartet funktioniert, ist es von besonderer Wichtigkeit, die Einsatzumgebung eines solchen Systems möglichst genau zu spezifizieren und bestenfalls einzuschränken. Im Kontext von KI wird hier in der Regel von einer genau definierten „Operational Design Domain“ (ODD) gesprochen. Diese kann bei bildverarbeitenden Systemen etwa bestimmte Voraussetzungen an die Beleuchtung, den Kamerawinkel oder die Benutzergruppe beinhalten. Ein hoher Detaillierungsgrad in der Beschreibung der ODD ist auch für die Erstellung und Überprüfung des Testdatensatzes relevant.

Um die im GS-BAU-70 geforderten Kriterien an die Zuverlässigkeit eines Assistenzsystems zu erreichen, muss die KI-Komponente „so gut wie realistisch möglich“ sein, ohne aber einem rigorosen

Sicherheitsnachweis entsprechen zu müssen. Dies gilt umso mehr, da die Definition im GS-BAU-70 über die „von den Benutzenden vernünftigerweise zu erwartenden Grenzen“ durch die öffentlichkeitswirksame Verbreitung von angeblich immer neuen Höchstleistungen durch KI geprägt sein kann. Hier gilt es deshalb auch besonders, ein transparentes System zu schaffen, um den Bediennenden über dessen Grenzen zu informieren.

Für KI-Systeme lassen sich aufgrund der aktuellen technischen, regulativen und normativen Besonderheiten vier aufeinander aufbauende Grundprinzipien für die Entwicklung ableiten, die jeweils mit entsprechenden Maßnahmen in Verbindung stehen (s. a. Abb. 2):

(1) Gute ingenieurwissenschaftliche und datenwissenschaftliche Praxis

Um ein Mindestmaß an Qualität für ein KI-System sicherzustellen, muss sich die Entwicklung an etablierten grundlegenden Praktiken orientieren. Dazu gehört die Einhaltung eines konsistenten Prozesses, der alle Phasen der Entwicklung eines KI-Systems umfasst. Neben der Sicherstellung eines guten Entwicklungsprozesses gehört zu den grundlegenden Praktiken auch die Anwendung von Normen zur Datenqualität und der Einsatz von Programmierrichtlinien.

(2) Anwendung/Interpretation bestehender Normen der funktionalen Sicherheit

Je nach Komplexität des gewählten Modells können einzelne Methoden zur Vermeidung und Reduzierung systematischer Fehler während der Softwareentwicklung aus den bestehenden Normen der funktionalen Sicherheit für KI-Systeme uminterpretiert werden (s. dazu etwa Anhang A.2 aus [5]).

(3) Minderung KI-spezifischer Risiken durch weiterführende Maßnahmen

Ergeben sich aus spezifischen Ausfallmodellen (z. B. fehlende Robustheit oder fehlende Transparenz) innerhalb der geplanten Anwendung des KI-Systems Gefährdungen, sollten präventiv verschiedene spezialisierte Methoden während der Entwicklung eingesetzt werden, um diese KI-spezifischen Ausfallrisiken zu minimieren. Beispiele wären hier etwa robustes Training oder Methoden der Erklärbarkeit. Oft ist die erreichte Ausfallsicherheit der KI-Komponente aber auch nach dem Einsatz solcher Methoden noch nicht ausreichend oder lässt sich nur eingeschränkt quantifizieren.

(4) Gesamtsystemische Absicherung und Laufzeitüberwachung

Auch nach Einsatz spezieller Techniken und Maßnahmen während des Entwicklungsprozesses bleibt nach aktuellem Stand der Technik beim maschinellen Lernen immer eine (Rest-)Unsicherheit bestehen. Wird die Unsicherheit während der Laufzeit erfasst und in der Implementierung berücksichtigt, lässt sich so auch für ein KI-System ein Maß für die Zuverlässigkeit bestimmen. Neben den modellbedingten Unsicherheiten können Umweltbedingungen dazu beitragen, dass die Vorhersagen eines Modells nicht zuverlässig sind. So wird beispielsweise ein kamerabasiertes System zur Personenerkennung bei fehlender Beleuchtung mit hoher Konfidenz vorhersagen, dass sich keine Person in seinem Sichtfeld befindet. Die Absicherung eines solchen KI-Systems sollte daher auf einer gesamtsystemischen Ebene betrachtet werden. Im genannten Beispiel könnte etwa ein klassischer Helligkeitssensor verwendet werden, um die Glaubhaftigkeit des Modells hinsichtlich dieses speziellen möglichen Fehlers zu überprüfen. Ist die Gesamtunsicherheit des KI-Systems hoch, sollte das Prinzip des sicheren Ausfalls angewendet werden. Dazu könnte das System etwa auf einen klassischen Entscheider zurückgreifen oder zunächst in den sicheren Zustand übergehen und eine weitere Überprüfung durch einen Menschen anfordern. Kann über eine solche gesamtsystemische Absicherung zur Laufzeit die Gesamtunsicherheit des Systems zu jeder Zeit glaubhaft quantifiziert werden, so dass ein Ausfall sicher erkannt wird, wäre ein wichtiger Schritt hin zu einem wirklichen Sicherheitsnachweis und damit dem Einsatz von KI in Sicherheitsfunktionen gemacht. Im aktuellen Stand der Technik ist dies aber noch nicht nachweisbar. Für Assistenzsysteme ist diese Form des Nachweises nach [2] nicht als Anforderung ableitbar. Dennoch lassen sich über Laufzeitüberwachung auch jetzt schon viele Unsicherheiten erkennen, wodurch potenziell gefährliche Situationen vermieden werden können, weshalb ihr Einsatz auch bei Assistenzfunktionen empfehlenswert ist.

3.1 Funktionsprüfungen des Gesamtsystems

Neben einer Entwicklung nach den oben beschriebenen Prinzipien liegt bei der Bewertung solcher Assistenzsysteme ein weiterer Schwerpunkt auf Funktionsprüfungen, so genannten Black-Box-Tests. Bei solchen Tests wird das System mit verschiedenen Ausgangssituationen, die in der Anwendung der betrieblichen Praxis rele-

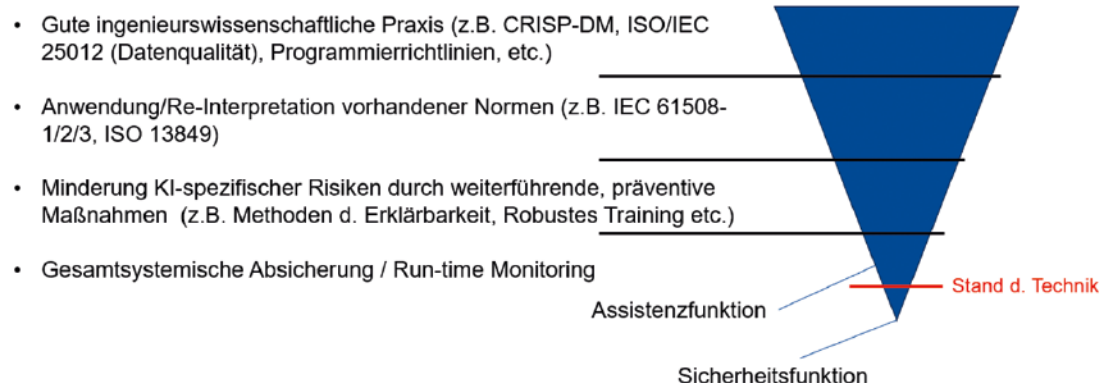


Abb. 2: Grundprinzipien bei der Entwicklung sicherheitsrelevanter KI-Systeme



Abb. 3: Einflussfaktoren auf die Erkennung bei einer KI-basierten Handerkennung (Verschmutzung, Tattoo, Handstellung)

© IFA der DGUV

vant sind, konfrontiert und die Einhaltung des erwarteten Systemverhaltens überprüft. Bei den Ausgangssituationen müssen bei KI-basierten Systemen besonders die unterschiedlichen Einflussfaktoren auf die Erkennung abgedeckt werden. In Bezug auf die Handerkennung an einer Formatkreissäge sind zum Beispiel die Hautfarbe, die Hautverschmutzung, Tätowierungen, die Beleuchtung und die verschiedenen möglichen Handstellungen Einflussfaktoren (Abb. 3). Die zuvor genannte Aufzählung stellt nur eine Auswahl von Einflussfaktoren dar und ist keine Wiedergabe der vollständigen Prüfscenarien.

4. KI-basierte Assistenzsysteme in der aktuellen KI-Normung

Im Bereich der Normung befasst sich der Technical Report ISO/IEC TR 5469 *Functional Safety and AI Systems* [5] mit dem Einsatz von KI-Komponenten z.B. im Maschinenschutz. Konkret werden dort die Themen Einsatz von KI als Sicherheitsbauteile, Einsatz von Nicht-KI Elementen zur Absicherung von KI-basierten Steuerungen sowie der Einsatz von KI-basierten Tools zum Design und zur Entwicklung von Sicherheitsfunktionen behandelt.

Der Technical Report beschreibt dabei sogenannte *Usage Level* für KI-Systeme, die eine Einteilung abhängig vom Einsatzkontext des Systems bieten. Hierbei werden *Usage Level* von A1 bis D definiert, wobei A1 einem System entspricht, in dem KI vollautomatisiert in der Sicherheitsfunktion eines Systems eingesetzt wird und D einem System, wo KI zwar eingesetzt wird, aber keinen Einfluss auf die Sicherheitsfunktion hat.

Innerhalb dieses Klassifizierungssystems fällt eine Assistenzfunktion durch KI in den *Usage Level* C. Dieser ist definiert als eine Anwendung, in der die KI-Technologie nicht Teil der Sicherheitsfunktion ist, aber einen indirekten Einfluss auf die Funktion haben kann. Eine Notiz präzisiert diese Definition, indem klargestellt wird, dass hierunter auch Systeme fallen, die eine zusätz-

liche Risikominderung bieten können, deren Funktion aber nicht ausschlaggebend für das Erreichen des akzeptablen Restrisikos ist.

ISO/IEC TR 5469 legt keine Anforderungen für die verschiedenen Systeme fest, sondern gibt lediglich allgemeine Hinweise zu deren Einsatz und speziellen Herausforderungen. Aktuell befindet sich der Nachfolger dieses Dokuments in Entwicklung, die Technical Specification ISO/IEC TS 22440. Dieses Dokument wird das Klassifizierungssystem des ISO/IEC TR 5469 weitestgehend übernehmen und abhängig vom Einsatzkontext entsprechende Anforderungen an das KI-System formulieren. Daher kann erwartet werden, dass dort auch konkretisierte, KI-spezifische Anforderungen an Systeme des Usage Level C, also auch Assistenzfunktionen, genannt werden. Die Veröffentlichung des neuen Dokuments ist für 2026 geplant.

5. Ausblick: Regulatorische Anforderungen an KI-basierte Sicherheitsbauteile in Maschinenverordnung und KI-Verordnung

Die neue EU-Maschinenverordnung (Verordnung (EU) 2023/1230, MV) enthält erstmals Aussagen über die Verwendung von KI in Sicherheitsbauteilen. So werden in Anhang II der Verordnung (Nicht erschöpfende Liste der Sicherheitsbauteile) unter Nr. 19. „Sicherheitsbauteile mit vollständig oder teilweise selbstentwickelndem Verhalten unter Verwendung von Ansätzen des maschinellen Lernens“ aufgeführt. In Anhang I, Teil A wird weiterhin festgelegt, dass Maschinen oder Produkte, die solche Sicherheitsbauteile enthalten, einer Konformitätsbewertung nach Artikel 25, Abs. 2 unter Einbeziehung einer notifizierten Stelle unterzogen werden müssen, also z.B. einer EU-Baumusterprüfung mit interner Fertigungskontrolle bedürfen. Eine Definition von vollständig oder teilweise selbstentwickelndem Verhalten wird innerhalb der Maschinenverordnung nicht gegeben, ebenso wenig konkrete, KI-spezifische Anforderungen. Diese ergeben sich aber über die Verbindung der MV mit der EU KI-Verordnung (Verordnung (EU) 2024/1689, KI-VO). Die KI-VO basiert auf einem risikobasierten Ansatz, der KI-Systeme entweder in die Kategorien kein Risiko, niedriges Risiko, hohes Risiko oder inakzeptables Risiko einteilt und dementsprechend Anforderungen festlegt bzw. im Falle der letzten Kategorie solche Systeme verbietet. Die KI-VO bindet sich über Artikel 6, Abs. 1 direkt an die MV (oder aktuell die EU-Maschinenrichtlinie (MRL) und andere Verordnungen), indem sie festschreibt, dass ein KI-System automatisch als Hochrisikosystem zu klassifizieren ist, wenn

- a) das KI-System als Sicherheitsbauteil eines Produkts verwendet wird, das unter die MV fällt und

b) dieses Produkt nach MV einer Konformitätsbewertung durch Dritte bedarf.

Dies trifft aktuell z.B. für alle KI-basierten Systeme zu, die dazu dienen, Personen oder Körperteile zu erkennen und die für die Realisierung von Sicherheitsfunktionen eingesetzt werden sollen sowie zudem die sonstigen Bestimmungen für Sicherheitsbauteile erfüllen. Für diesen Fall legt Artikel 43, Abs. 3 der KI-VO fest, dass die Konformitätsbewertung nach dem Verfahren der MV (oder der entsprechenden anderen harmonisierten Verordnung oder Richtlinie) durchgeführt wird, dabei jedoch die in der KI-VO festgelegten Anforderungen an Hochrisiko-KI-Systeme in diese Bewertung mit einbezogen werden müssen. Durch diese Verbindung gelten also die Anforderungen für Hochrisiko-KI-Systeme automatisch für alle Sicherheitsbauteile mit KI in Produkten, die unter die MV fallen und eine Konformitätsbewertung durch Dritte benötigen. Die Notwendigkeit einer separaten Konformitätsbewertung zur KI-VO entfällt somit für diese Produkte nach jetzigem Stand.

5.1 Anforderungen der KI-VO für Hochrisiko-KI-Systeme

Abb. 4 zeigt die Anforderungskategorien, welche die KI-VO für Systeme vorschreibt, denen ein hohes Risiko bescheinigt wird. Werden Anforderungen bereits durch die umschließende Konformitätsbewertung nach MV oder MRL gefordert wie beispielsweise ein Risikomanagement, kann diese Kategorie auch nach MV/MRL abgehandelt werden.

Zu allen Anforderungskategorien läuft aktuell ein *standardization request* der EU-Kommission an die europäischen Normungsorganisationen CEN/CENELEC. Dort sollen Normen entstehen,

welche die Anforderungen auf technischer Ebene konkretisieren und die im weiteren Verlauf zu harmonisierten Normen unter der KI-VO werden sollen.

6. Zusammenfassung

Um KI-Komponenten im Maschinenschutz als Sicherheitsfunktionen einzusetzen, wird ein rigoroser Sicherheitsnachweis benötigt. Dieser ist zum aktuellen Stand der Technik, Normung und Gesetzgebung noch schwer zu erreichen. Um KI-Komponenten trotzdem schon heute zur Verbesserung der Sicherheit bei der Arbeit einsetzen zu können und gleichzeitig Erfahrungen mit diesen Komponenten zu sammeln, bietet es sich an, diese als Assistenzsysteme z.B. zusätzlich zu einer bestehenden Sicherheitsfunktion zu verwenden. Zur Zertifizierung einer solchen Assistenzfunktion kann auf den Prüfgrundsatz GS-BAU 70 zurückgegriffen werden, der allerdings keine konkreten Hinweise und Anforderungen für KI-Komponenten enthält. Um die allgemeinen Anforderungen des Prüfgrundsatzes zu erfüllen, können die in diesem Beitrag beschriebenen Schritte für die Ableitung KI-spezifischer Kriterien verwendet werden. Besondere Relevanz haben dabei die Bestimmungen des Prüfgrundsatzes zu Zuverlässigkeit und Transparenz.

Für die zukünftige Nutzung von KI-Komponenten im Maschinenschutz wird nicht nur die EU-Maschinenverordnung, sondern über die skizzierte Verbindung beider Regelungen auch die EU KI-Verordnung eine wichtige Rolle spielen. Deren Anforderungen für Hochrisiko-KI-Systeme müssen für Sicherheitsfunktionen durch KI im Maschinenschutz unter den o.g. Voraussetzungen eingehalten werden. ■

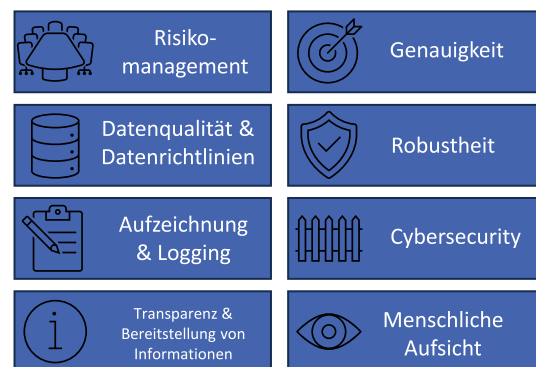


Abb. 4: Anforderungskategorien der KI-VO an Hochrisiko-systeme

© IFA der DGUV

LITERATUR

- [1] S. Mohseni et al., *Taxonomy of Machine Learning Safety: A Survey and Primer*. ACM Comput. Surv. 55, 8 (2023)
- [2] GS-BAU-70, Grundsätze für die Prüfung und Zertifizierung von sicherheitsrelevanten Assistenzsystemen an Maschinen und Nutzfahrzeugen, DGUV Test, Prüf- und Zertifizierungsstelle, Fachbereich Bauwesen, Stand: 2022-12.
- [3] DIN EN ISO 12100, Sicherheit von Maschinen – Allgemeine Gestaltungsleitsätze – Risikobeurteilung und Risikominderung (03.11). Beuth, Berlin 2011
- [4] D. Hendrycks, T. Dietterich, *Benchmarking Neural Network Robustness to Common Corruptions and Perturbations*, (2019), <https://arxiv.org/abs/1903.12261>
- [5] ISO/IEC TR 5469, Artificial intelligence – Functional safety and AI systems (01.24). ISO/IEC, Genf 2024