Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme

Thomas Jürgensohn[1], Christina Platho[1], David Stegmaier[2], Matthias Hartwig[2], Mathilde Krampitz[2], Lorenz Funk[2], Timon Plass[2], Heiko Ehrlich[3]

baua: Fokus

Der Einsatz von Maschinen mit KI-Algorithmen in der Industrie scheint möglich. Inwieweit sich KI-Systeme auf die Sicherheit der Beschäftigten auswirken und ob Änderungen an der bestehenden Rechtsordnung notwendig sind, ist jedoch noch nicht systematisch untersucht worden.
In dem Projektes F2432 wurde eine Taxonomie entwickelt, die sicherheitsrelevante Faktoren in KI-Systemen übersichtlich dargestellt. Auf dieser Basis werden Vorschläge zur Weiterentwicklung des Produktsicherheitsrechts gemacht. Durch die Einführung unterschiedlicher rechtlicher Konstrukte kann der Rechtsrahmen an die spezifischen Eigenschaften von KI-Systemen angepasst werden.

Inhalt

1 Einleitung

Zentrale Fragestellung des Forschungsvorhabens war, ob und inwieweit der Einsatz von KI-Algorithmen und anderer Software mit Autonomiemerkmalen in physischen Systemen der Industrie, die einer Sicherheitsbewertung bedürfen, Änderungen an der bestehenden Rechtsordnung wie beispielsweise Produktsicherheits- und Betriebssicherheitsrecht erforderlich machen. Diese potentiell gefährdenden physischen Systeme, deren Verhalten durch Software bestimmt ist, die zu einem Teil aus KI-Algorithmen besteht, werden im Forschungsvorhaben

[1] HFC Human-Factors-Consult GmbH. [2] IKEM Institut für Klimaschutz, Energie und Mobilität e.V., [3] TÜV Nord

baua: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

als „software-physische (KI)Systeme“ bezeichnet. Da in industriellen Bereichen wie etwa der Automatisierungstechnik, der Robotik oder dem Maschinenbau KI-Algorithmen bisher nur sehr rudimentär eingesetzt werden, wurde auch die Thematik hochautomatisierter (autonomer) Kraftfahrzeuge bei der Projektdurchführung berücksichtigt.

Aus rechtlicher Sicht galt es im Forschungsvorhaben im ersten Schritt detailliert zu prüfen, ob sich mit der beschriebenen Klasse neuer Systeme zusätzliche Gefährdungen für Arbeitnehmer und Verbraucher ergeben, die im Recht bisher keine Berücksichtigung gefunden haben oder die mit den bestehenden Regelungen nicht mehr zweckmäßig erfasst und beherrscht werden können. Im zweiten Schritt sollte geprüft werden, welche rechtlichen Rahmenbedingungen bei erkannten potentiellen Gefährdungen eine Gewährleistung der Sicherheit garantieren können. Es sollte ferner untersucht werden, ob die Verantwortungsallokation unter den Beteiligten (Hersteller, Betreiber/ Verwender, Arbeitnehmer) nach dem präventiven Ordnungsrecht (insb. Produktsicherheits- und Betriebssicherheitsrecht) bzw. dem repressiven Haftungsrecht diesen neuen Systemen noch gerecht wird.

2 Entwicklung einer Taxonomie zur Kategorisierung von KI-Systemen

Um diese zentralen Fragen des Forschungsvorhabens beantworten zu können, wurde ein als Taxonomie ausgearbeitetes Kategoriensystem entwickelt, das Faktoren der Sicherheit in software-physischen (KI-)Systemen unter besonderer Berücksichtigung neuerer technischer Entwicklungen kategorial zusammenfasst. Wesentliche Grundlage dieser Taxonomie waren umfangreiche Expertenbefragungen und ergänzende Inhalte aus der Literatur. Die Expertenbefragungen erstreckten sich über einen Zeitraum von über einem Jahr und wurden als Einzelinterviews von 1 – 2 Stunden Dauer mit 33 Experten durchgeführt. Einige der Experten wurden mehrfach befragt. Die Experten kamen aus den Bereichen Robotik, Smart-Home, KI, Automatisierungstechnik, Automotive, Funktionale Sicherheit und Security. Ein Großteil der Experten waren Safety-fachleute, die in den unterschiedlichsten Normungsgremien mit der Thematik Safety in Verbindung mit KI tätig waren. Eine vollständige Doppelexpertise sowohl im Bereich KI als auch im Bereich Sicherheitsbewertung hatten aber nur wenige der Experten. Der Gegenstandsbereich KI und Sicherheit im Sinne Safety ist noch sehr jung und außerhalb des Bereichs hochautomatisierter Fahrzeuge im Gegensatz zum Themenfeld KI in der Informationstechnik bislang kaum bearbeitet. Von einem etablierten Wissensgebiet KI-Safety kann bisher noch nicht gesprochen werden.

3 Die Dimensionen der Taxonomie

Die Taxonomie beschreibt insgesamt 40 Ausprägungen (Faktoren oder Eigenschaften) von physischen Systemen und deren Umgebungen, geordnet in 7 Dimensionen und 14 Unterkategorien, die Einfluss auf die Sicherheit (Safety) bzw. damit zu tun haben. Die oberste Ebene des Kategoriensystems wird durch die Merkmale „Veränderbarkeit“, „Transparenz“, „Vernetzung“, „Kontrollierbarkeit“, „Widerstandsfähigkeit“, „Involviertheit des Menschen“, und „Schadensfolgen“ gebildet. Diese fassen jeweils eine Reihe von Unterkategorien und Merkmale zusammen. Beispielhaft sind dabei zu nennen: „Spezifizierbarkeit“, „Nachvollziehbarkeit“, „Vorhersehbarkeit“ oder „Autonomie“, „Emergenz“, „Robustheit“, „Resilienz“. Die folgende Tabelle gibt einen Überblick über die Ausprägungen der Taxonomie.

Tab 1: Grafische Übersicht über die Taxonomie

Veränderbarkeit		Vernetzung		Involviertheit des Menschen	
System	Keine Veränderung nach Auslieferung	**Intern**	Zentrale Vernetzung	**Handelnder**	Mensch als sicherheitsgewährender sowie fehlerbehafteter Teil der Prozesskette
	Vorgesehene, kontrollierte Veränderung nach Auslieferung		Dezentrale Vernetzung		Fehlbenutzung entgegen bestimmungsgemäßer Verwendung
	Adaptivität (geringfügige Veränderung weniger Parameter im Betrieb	**Nach außen**	Abgesprochene Daten, nur informativ oder handlungsbeeinflussend		Bewusste Störung der intendierten Funktionsweise bzw. Schädigung
	Bedeutende Veränderungen im Betrieb		Unabgesprochene Daten		
Umfeld	Geringe Änderungen in einem kontrollierbaren Umfeld	**Kontrollierbarkeit**		**Gefährdeter**	Eingewiesene Angestellte
	Nicht vorhersehbare Änderungen in einem komplexen Umfeld	**Emergenz**	Autonomie (relative Unabhängigkeit durch Gütekriterien, Zielhierarchien)		Nutzer
Transparenz			Selbstorganisation (Ordnung durch interne Interaktion)		Verbraucher
Experten	Spezifizierbarkeit	**Beschränkungen**	Systembeschränkungen durch virtuelle Schutzzäune/konventionelle Systeme		Dritte ohne Beziehung zum System
	Beschreibbarkeit der Funktionsgrenzen		Beschränkungen im Umfeld oder Einsatzbereich	**Schadensfolgen**	
	Nachvollziehbarkeit		Beschränkungen und Kontrolle im Datenfluss	**Personenschäden**	Keine Personenschäden
	Vorhersehbarkeit	**Widerstandsfähigkeit**			Geringfügige Personenschäden
Beteiligte	Kenntnis des Einsatzbereichs und der Grenzen des Systems	**Robustheit**	Stabilität bei kleinen Änderungen des Inputs		Erhebliche Personenschäden
	Vorhersehbarkeit der Dynamik bzw. des Prozesszustands des Systems		Bewältigung unbekannter Situationen bzw. unvorhergesehener Ereignisse	**Sonstige Schäden**	Sachschäden
	Verständnis der Systemfunktionalität		Security		Umweltschäden
		Resilienz	Passive Wirkungsbegrenzung nach Fehlern		Immaterielle Schäden
			Aktive Wirkungsminderung nach Fehlern		

3.1 Dimension Veränderbarkeit

Die Dimension „Veränderbarkeit“ beschreibt technisch-physische Systeme in Bezug auf Änderungen von Eigenschaften bzw. von Verhalten des Systems im Betrieb oder in Bezug auf Änderungen des Umfelds. Dabei werden vier Stufen der Veränderbarkeit des Systems im Betrieb unterschieden. Art und Ausmaß von Veränderbarkeit können sehr unterschiedlich gestaltet sein. Dazu gehören zum einen „vorgesehene und kontrollierte Veränderungen nach Auslieferung“ des Systems. Diese sind entweder von außen durch Nutzer bzw. Hersteller/Betreiber initiiert oder aber sind von letzteren im System selbst angelegt. Eine neue Qualität der Veränderbarkeit ergibt sich bei „bedeutenden Veränderungen“ während des Betriebs, wenn die Konsequenzen der Veränderungen nicht mehr vollständig vorhersehbar sind. Das Systemverhalten kann sich mit der Zeit fundamental verändern, weil das System während des Betriebs anfallende Daten für die Modifikation des eigenen Verhaltens (in der Regel eine Optimierung) nutzt. Eine Veränderung während des Betriebs auf Basis zusätzlicher Daten wurde in dem Projekt als „Weiterlernen“ bezeichnet. Für die rechtliche Bewertung sind weiterlernende Systeme von besonderer Bedeutung.

3.2 Dimension Transparenz

Die Dimension „Transparenz“ der Taxonomie beschreibt die „Verständlichkeit“ des Systems bzw. seines Verhaltens und zwar aus zwei Perspektiven: Nach innen wird die Transparenz

aus Expertensicht, d. h. dem Entwickler oder Sicherheitsingenieur gegenüber beschrieben, während sie nach außen als die am Bedarf gemessene Vollständigkeit und Verständlichkeit von Informationen zur generellen und situationsspezifischen Funktionsweise eines Systems für einen Nutzer, Bediener, Instandhalter oder anderweitig Beteiligten definiert wird. Bei der Transparenz aus Expertensicht handelt es sich dabei um die Spezifizierbarkeit, die Beschreibbarkeit der Funktionsgrenzen, die Nachvollziehbarkeit und die Vorhersehbarkeit des Systemverhaltens. Für einen Beteiligten ist hingegen die Transparenz durch eine grundlegende Kenntnis des Einsatzbereichs und der Grenzen des Systems und seiner Dynamik von Bedeutung.

3.3 Dimension Vernetzung

Die Dimension „Vernetzung“ beschreibt Eigenschaften von Systemen in Bezug auf interne Prozesse des Austauschs von digitalen Daten und bezüglich des Einflusses äußerer digitaler Daten auf systeminternes Verhalten. „Vernetzungen“ über niedrigdimensionale, digital vermittelte Sensordaten zur Erfassung von Temperatur, Druck, etc. sind dabei nicht einbezogen. Inhaltlich abzugrenzen ist der Aspekt der Vernetzung auch von der Kommunikation bzw. Interaktion mit Beteiligten oder Nutzern zum Austausch von Informationen über Status, Aufgaben oder Intentionen des Systems. Eine „interne Vernetzung“ charakterisiert den Austausch digitaler Daten zwischen Subsystemen. Liegt eine rein interne Vernetzung vor, so besteht diese entweder mit einer zentralen Instanz, die Aktionen des Systems (und ggf. anderer Systeme) im Sinne einer übergeordneten Zielsetzung orchestriert (z. B. zur bedarfsorientierten Optimierung des Einsatzgebiets der Systeme), oder sie ist dezentral. Vernetzung ist insofern für Sicherheitsüberlegungen von Bedeutung, weil – insbesondere bei dezentraler Vernetzung – die Gefahr einer fehlenden Kontrollierbarkeit besteht.

3.4 Dimension „Kontrollierbarkeit“

Die Dimension „Kontrollierbarkeit“ beschreibt die Eigenschaften von Systemen und ihres Umfeldes in Bezug auf das Vermögen des Herstellers oder Betreibers, das Systemverhalten determinieren zu können. Die Kontrollierbarkeit eines Systems hängt von dem Grad der Emergenz eines Systems ab, d. h. der Befähigung zu Verhalten aus sich selbst heraus, sowie von Art und Ausmaß der Beschränkungen, die man dem System oder dem Datenfluss von außen auferlegt. Systeme gelten im Sinne der vorliegenden Taxonomie dann als emergent, wenn sie Befähigung zu Autonomie oder Selbstorganisation aufweisen. Die Facette „Beschränkungen“ der Dimension „Kontrollierbarkeit“ umfasst die Maßnahmen, die dabei helfen, auch bei hoher Emergenz und dadurch verminderter Kontrollierbarkeit sicheres Verhalten zu gewährleisten, sowie bei nach außen vernetzten Systemen jene Maßnahmen, die die Kontrolle des Systemverhaltens bei Datenfluss von außen sicherstellen. Zu den Systembeschränkungen gehören beispielsweise Systembeschränkungen durch technologische Maßnahmen, deren Ziel es ist, schlecht kontrollierbare Teilsysteme zu überwachen.

3.5 Dimension Widerstandsfähigkeit

Die Dimension „Widerstandsfähigkeit“ beschreibt das Vermögen von Systemen, trotz Störungen frei von sicherheitsrelevanten Fehlern zu agieren und etwaige sicherheitswirksame Fehlfunktionen abzuwenden oder zumindest deren Folgen abzuschwächen. Das Vermeiden von Fehlern ist unter dem Begriff „Robustheit“, das Vermeiden oder die Minderung von Fehlerfolgen hingegen unter dem Begriff „Resilienz“ gefasst. „Robustheit“ ist in diesem Sinne synonym zu „abfangbar“ zu verstehen, wobei zwischen dem Abfangen äußerer Einflüsse (Security) und dem Abfangen innerer Abweichungen unterschieden wird. Im Gegensatz zur Robustheit, die auf das Vermeiden eines systemseitigen Fehlverhaltens oder von Angriffen von außen mit Auswirkungen auf die Safety zielt, beschreibt die „Resilienz“ das Potential von Systemen, die Folgenschwere zu reduzieren – bis hin zur vollständigen Vermeidung von Folgen.

3.6 Dimension Involviertheit des Menschen

Eine Beurteilung der Sicherheit von Systemen kann nicht ohne Bezug zum Menschen erfolgen. Dieser kann einerseits als Handelnder die Sicherheit des Gesamtsystems aktiv beeinflussen und ist andererseits derjenige, den es vor etwaigen Gefährdungen zu schützen gilt. Diese beiden Aspekte werden im Rahmen der Taxonomie in der Dimension „Involviertheit des Menschen“ berücksichtigt. Mit Blick auf den Menschen als Handelndem werden in der Taxonomie unterschiedliche Aspekte betrachtet. Es wird zum einen die sicherheitsbeeinflussende Wirkung des Menschen beim Umgang mit dem System betrachtet. Ein bedeutender Teilaspekt sind auch etwaige Abweichungen von bestimmungsgemäßen Verwendungen von Systemkomponenten durch den Menschen. Schließlich kann der Mensch als Saboteur auch als bewusster Schädiger in Erscheinung treten. Blickt man auf den Menschen als Gefährdeten, der etwaigen unerwünschten Folgen nicht intendierten Systemverhaltens ausgesetzt ist, so kann er in verschiedene Klassen eingeordnet werden, die durch seine Rolle im Systemkontext und durch Schutzmöglichkeiten beschrieben sind.

3.7 Dimension Schadensfolgen

Im Fokus der Dimension „Schadensfolgen“ steht die Betrachtung der Personenschäden. Die Betrachtung von Sachschäden, Umweltschäden oder immateriellen Schäden steht eher am Rande. Ist kein Szenario möglich, in dem der Einsatz des Systems zu einer Verletzung des Menschen führt, sind die anderen Dimensionen im Zusammenhang mit Sicherheitsüberlegungen in der Regel ohne Bedeutung. Die Taxonomie wurde mit Blick auf Anwendungen im industriellen Bereich entwickelt; eine Übertragbarkeit auf andere Anwendungsbereiche – bspw. Robotik im Smart Home-Bereich oder der medizinischen Versorgung – ist möglich, erfordert aber Modifikationen.

4 Herausforderungen bei der Beschreibung von KI-Systemen

In der Folge der Befragungen stellte sich heraus, dass die Safety-Experten oft nur eine vage Vorstellung über Grenzen und Fähigkeiten von KI und die dort verwendeten Begrifflichkeiten aufwiesen, und die KI-Experten analog kaum Vorstellung über Vorgehen, Denkweisen und regulatorische Rahmenbedingungen im Themenfeld Produktsicherheit hatten. Aus diesem Grunde wurde im Projekt als Fundament der eigenen, sich in der Taxonomie und in den rechtlichen Bewertungen widerspiegelnden Begriffsverwendung eine umfangreiche Analyse der Bedeutung der wichtigsten relevanten Begriffe geleistet.

5 Begriffe und Definitionen

5.1 autonom und automatisiert

Bei der Analyse der Begrifflichkeiten stellte sich heraus, dass die Verwendung des Begriffes „autonom“ und daraus abgeleitet des Begriffes „autonomes System“ domänenspezifisch ist. Darüber hinaus ist eine Abgrenzung zum Begriff „automatisiert“ schwierig. Im Projekt wurde aus einer Begriffsanalyse eine sehr allgemeine Definition von „autonom“ erarbeitet, die in sehr verschiedenen Kontexten eingesetzt werden kann. Demnach beschreibt „autonom“ das unabhängig/selbstständig sein von etwas und bezüglich etwas, und ein autonomes System bezeichnet folglich ein System, das gleichsam selbstständig von etwas und bezüglich etwas ist - es agiert in einem definierten Sinne unabhängig. Es gibt eine starke semantische Nähe von Autonomie und Automatisierung. Bei beiden Begriffen ist es die Eigenständigkeit, die begrifflich adressiert wird. Allerdings liegt bei der Automatisierung der Schwerpunkt der Betrachtung auf Eigenständigkeit im Ersatz menschlicher Handlungen, bei der Autonomie kann die Unabhängigkeit auch von anderen Maschinen beziehungsweise Softwaresystemen möglich sein.

5.2 Intelligenz und Künstliche Intelligenz

Im Zuge des Projektes wurde auch erarbeitet, was im Zusammenhang mit KI unter „Intelligenz“ zu verstehen ist. Eine gängige Definition aus der Psychologie beschreibt „Intelligenz“ als „sehr allgemeine geistige Kapazität, die – unter anderem – die Fähigkeit zum schlussfolgernden Denken, zum Planen, zur Problemlösung, zum abstrakten Denken, zum Verständnis komplexer Ideen, zum schnellen Lernen und zum Lernen aus Erfahrung umfasst. Es ist nicht reines Bücherwissen, keine enge akademische Spezialbegabung, keine Testerfahrung. Vielmehr reflektiert Intelligenz ein breiteres und tieferes Vermögen, unsere Umwelt zu verstehen, ‚zu kapieren‘, ‚Sinn in Dingen zu erkennen‘ oder ‚herauszubekommen‘, was zu tun ist“[3].

Der Begriff „Intelligenz“ ist zur Beschreibung von Unterschieden zwischen Menschen entwickelt worden und daher auch auf diesen Zweck hin ausgelegt. Eine Analyse der Verwendung des Begriffes im Kontext der KI, aber auch in anderen, nicht auf den Menschen bezogenen Zusammenhängen ergab, dass sich inzwischen ein zweiter Intelligenzbegriff etabliert hat, der sich von dem psychologischen unterscheidet. Einer der wesentlichen Unterschiede liegt darin, dass eine Intelligenzzuordnung in der KI durch Beschreibung von Systemkonstituenten, also diskreten Merkmalen und weniger durch messbare Eigenschaften gebildet wird. Für „Intelligenz“ gibt es dort – im Gegensatz zur Psychologie – keine Metrik, mit deren Hilfe beispielsweise der Intelligenzgrad verschiedener Maschinen verglichen werden könnte. Auch ist die Operationalisierung nicht sinnvoll auf Maschinen übertragbar.

Wie der Begriff „Intelligenz“ in neuer Semantik ist auch der Begriff „Künstliche Intelligenz, KI“ nur schwer konkret zu fassen. Neben der Beschreibung eines Gegenstandsbereichs oder eines Systems wird der Begriff „Künstliche Intelligenz“ auch für ein Teilgebiet der Informatik verwendet, welches allerdings durchaus als stark heterogen zu bezeichnen ist. So gehören dazu beispielsweise die Symbolische KI, Maschinelle Lernverfahren, Konnektionismus und viele weitere Gebiete. Eine inhaltlich begründete Umgrenzung ist schwierig, auch deshalb, weil die Anzahl der Teilgebiete der „Künstlichen Intelligenz“ wächst. Aufgrund der Schwierigkeit, die der (neue) Intelligenzbegriff mit sich bringt, wurde KI im Projekt definiert als: „Ein heterogenes Teilgebiet der Informatik, das aus dem ursprünglichen Ziel, menschliches Denken nachzubilden, entstanden ist“. Ein „KI-System“ ist dementsprechend ein System, das Komponenten enthält, die der KI zuzuordnen sind.

6 Einsatz von sicherheitskritischer KI in KMUs

Die Neuheit des Wissensgebiets KI-Safety wird dadurch unterstrichen, dass nach Meinung der Experten KI in sicherheitskritischen Anwendungen bisher noch nicht Einzug gehalten hat, und gerade erst die ersten Versuche diesbezüglich gestartet werden. Eine Abschätzung über Zeithorizonte konnte von keinem der Experten vorgenommen werden. Als Hinderungsgründe für eine schnelle Einführung wurden übereinstimmend fehlende Normen und fehlende Best-Practice-Beispiele genannt. Insbesondere für die – im Vergleich zur Automobilindustrie – eher mittelständisch orientierten Unternehmen bedarf es dieser an Vorgehensmodellen orientierten Herangehensweise zur Absicherung des eigenen Verhaltens. Umgekehrt wird aus den Normungsgremien heraus beklagt, dass der erste Schritt von den Unternehmen getätigt werden müsse – das übliche Vorgehen der Normung sei das Zusammenfassen und Ordnen von realisierten Best-Practice-Beispielen.

[3] Die vollständige Definition findet man im Gesamtbericht.

7 Besonderheiten von Methoden der KI am Beispiel der künstlichen neuronalen Netze

Die Automobilindustrie ist hier mit der Entwicklung von hochautomatisierten Fahrzeugen mit dem Endziel des fahrerlosen Fahrzeugs wesentlich weiter. KI-Algorithmen werden in hochautomatisierten Fahrzeugen, aber auch beispielsweise bei mobilen Robotern an mehreren unterschiedlichen Stellen eingesetzt. Im Zentrum der Diskussion stehen allerdings hier Algorithmen, die zur Auswertung von Kamerabildern und weiteren Sensoren eingesetzt werden, um die Umgebung des Fahrzeugs mit allen statischen und dynamischen Objekten zu erkennen und daraus dann das situationsangepasste Verhalten der fahrzeug- oder roboterlenkenden Algorithmen abzuleiten. Für die Auswertung der Kameradaten werden häufig tiefe neuronale Netze genutzt, eine Untergruppe der künstlichen neuronalen Netze (KNN), die wiederum zu den maschinellen Lernverfahren zu rechnen sind – ein Teilgebiet der KI. Das Besondere der hier eingesetzten KNN ist eine sehr große Zahl von verhaltensdeterminierenden Variablen, deren Werte durch Daten bestimmt werden, die in einem Lernprozess für das Eingangs-Ausgangsverhalten bestimmt werden. Wegen der hohen Anzahl von Eingangsvariablen, die im realen Einsatz ganz unterschiedlich belegt sein können, ist in Verbindung mit der hohen Anzahl von möglichen Zuständen des KNN eine genaue Vorhersage des Ausgangs bei bekanntem Eingang nicht möglich. Der Designprozess der Funktion einer Software verschiebt sich beim Einsatz von KNN deshalb von der Programmierung von Funktionalität hin zur Auswahl von Eingangs- bzw. Trainingsdatensätzen.

7.1 Notwendige Sicherheitsnachweise werden komplex

Auswahl und Qualität der funktionsbestimmenden Trainingsdaten determinieren also die Sicherheit von Systemen, die auch von dem KNN-Verhalten bestimmt sind. Dies ist ein völlig neues Faktum und im traditionellen Vorgehen des Sicherheitsnachweises bisher unbekannt. Datengetriebene Algorithmen hoher Komplexität werden eingesetzt, weil mit ihnen Systemverhalten erzeugt werden kann, das mit anderer, traditioneller Software nicht möglich ist. Ein Nachteil der Algorithmen ist, dass die Nachvollziehbarkeit des Systemverhaltens reduziert ist: Das Verhalten „ergibt sich“, und die Frage, „warum“ ein spezielles Verhalten auftritt, ist nur sehr eingeschränkt beantwortbar. Von vielen Experten wurde diese Besonderheit dieser speziellen KI-Algorithmen betont und als fehlende „Transparenz“ bezeichnet. Eine weitere Besonderheit der KNN in komplexen Anwendungsszenarien ist die im Vergleich zu nicht-datengetriebener Software verhältnismäßig geringe Robustheit, womit gemeint ist, dass kleine Änderungen am Eingang u. U. zu großen Änderungen am Ausgang führen. Dies ist ein Verhalten, das man bei klassischen Algorithmen nicht oder nur in wesentlich geringerem Ausmaß kennt, und das dazu führt, dass das Systemverhalten in einem gewissen Maße un-vorhersehbar wird.

Für einen Sicherheitsnachweis jedoch spielt die Vorhersehbarkeit und Nachvollziehbarkeit von Systemverhalten eine große Rolle. Zwar wird in der Forschung derzeit intensiv nach Möglichkeiten der Erhöhung der Nachvollziehbarkeit gearbeitet, es zeigt sich aber, dass bei datenerzeugtem Systemverhalten u.U. ein neuer Prozess der Gewährleistung der Sicherheit entwickelt werden muss. Eng mit Vorhersehbarkeit und Nachvollziehbarkeit verbunden ist der Prozess der Spezifizierung von Systemeigenschaften. Von einigen Experten wurde betont, dass datengetriebene Algorithmen hoher Mächtigkeit insbesondere bei Aufgabenstellungen mit hoher Komplexität – beispielsweise bei hochautomatisierten Fahrzeugen – eingesetzt werden, die mit klassischer Software nicht gelöst werden können. Die Komplexität der Aufgabenstellung führt aber dazu, dass die Funktionalität der Software nicht mehr in dem Maße spezifizierbar ist, wie es in der herkömmlichen Produktentwicklung üblich ist. Dies wiederum führt dazu, dass natürlich auch nicht mehr gegenüber einer Spezifikation über Tests in gleichem Maße verifiziert werden kann, wie in dem klassischen Vorgehen. Die Unmöglichkeit der genauen Spezifizierung in bestimmten Bereichen verändert also den klassischen Weg des

Sicherheitsnachweises mit den Stufen Spezifikation – Verifikation (Test) – Validierung. Die Minderung der Spezifizierbarkeit hat seine Ursache allerdings in der Komplexität der Anwendung. Sind irgendwann in der Zukunft auch andere als KI-Algorithmen in der Lage, dieselben Aufgaben zu lösen, wird folglich auch bei diesen die Spezifizierbarkeit reduziert sein.

8 Zwischenfazit

8.1 Taxonomie und Einfluss auf die rechtliche Betrachtung

Die eben angesprochenen Besonderheiten von bestimmten KI-Algorithmen bei Veränderbarkeit, der unter Transparenz gefassten Nachvollziehbarkeit, Spezifizierbarkeit und Vorhersehbarkeit sowie der Kontrollierbarkeit und der darunter gefassten Robustheit sind tragende Elemente der im Forschungsprojekt entwickelten Taxonomie. Weitere Elemente sind über das Anwendungsfeld der Robotik (autonome Roboter, fahrerlose Transportsysteme, kollaborierende Roboter) sowie die sich dadurch ergebenden Interaktionen mit Menschen (Involviertheit des Menschen) eingeflossen. Ferner sind Elemente eingeflossen, die sich aus den heutigen Möglichkeiten der Vernetzung von Teilsystemen im Innern (cyberphysische Systeme), aber auch durch eine Vernetzung nach außen ergeben. Für die rechtlichen Betrachtungen besonders relevant ist die Tatsache, dass Systemverhalten, das sich aus der Verwertung vieler Daten in einem Trainingsprozess ergibt, durch Weiterlernen während des betrieblichen Einsatzes optimal an die spezifischen Einsatzverhältnisse angepasst werden kann. Aus der Anwendung von KI-Algorithmen bei Spracherkennungsalgorithmen ist dieses Verfahren des Weiterlernens bekannt. Ein Nachweis der Sicherheit ist dann aber wegen der Veränderbarkeit im Betrieb und den eben genannten Defiziten in der Vorhersehbarkeit und Robustheit entscheidend erschwert.

8.2 Ergebnisse der rechtlichen Betrachtung von sicherheitskritischen (KI) Systemen

Ein bedeutender Teil der rechtlichen Betrachtungen in vorliegendem Forschungsvorhaben behandelt eine mögliche Wandelbarkeit von Systemen während des Betriebs und die Vernetzung von Systemen. Nach dem Produktsicherheitsrecht ist der Hersteller für das Inverkehrbringen und die Inbetriebnahme von Maschinen verantwortlich und hat dafür zu sorgen, dass die Maschinen den Sicherheits- und Gesundheitsschutzanforderungen des Produktsicherheitsrechts entsprechen. Zum Zeitpunkt des Inverkehrbringens bzw. der Inbetriebnahme müssen alle formellen und materiellen Voraussetzungen vorliegen. Insbesondere weiterlernende Systeme – also solche, die sich im Betrieb auf Basis neuer Daten verändern – sind jedoch rechtlich nicht mehr nach den Kriterien des Produktsicherheitsrechts beherrschbar, da der maßgebliche Zeitpunkt der Risikobeurteilung (das Inverkehrbringen bzw. die Inbetriebnahme) die Veränderungen des Systems nach diesem Zeitpunkt ausblendet.

9 Risikobeurteilungen von KI-Systemen

Die Risikobeurteilung nimmt den ganzen Lebenszyklus des Produkts in den Blick und berücksichtigt über die Zeit auftretende Veränderungen, z. B. durch Verschleiß, bestimmt eine bestimmungsgemäße Verwendungsart und -dauer und berücksichtigt auch davon abweichende Verwendungen. Diese Beurteilung bildet die Grundlage für die „Integration der Sicherheit“. Die produktsicherheitsrechtliche Pflicht des Herstellers konzentriert sich auf den benannten Zeitpunkt des Inverkehrbringens bzw. der Inbetriebnahme. Die Risikobeurteilung nimmt zwar auch die Zeit danach in den Blick, ist jedoch im maßgeblichen Zeitpunkt abgeschlossen. Treten danach Umstände zutage, die einem Inverkehrbringen des Produkts entgegengestanden hätten, wären sie im Zeitpunkt des Inverkehrbringens bekannt gewesen, werden die Marktüberwachungsbehörden die erforderlichen Maßnahmen ergreifen, um der von dem Produkt ausgehenden Gefahren zu begegnen. In manchen Produktbereichen wird diese Marküberwa-

chung flankiert durch eine Produktbeobachtungspflicht des Herstellers, so z. B. bei Verbraucherprodukten. Dieses Regel-Ausnahme-Verhältnis von abgeschlossener und umfassender Risikobeurteilung einerseits und unvorhergesehenen, nach Inverkehrbringen auftretenden Mängel und Eingreifen der Marktüberwachung andererseits wird bestimmten KI-Systemen jedoch nicht mehr gerecht. Denn bei einem hohem Grad an Veränderbarkeit im Betrieb bei gleichzeitig geringer Transparenz und Kontrollierbarkeit für den Verwender bzw. Betroffenen ist das Auftreten unvorhergesehenen Systemverhaltens nach Inverkehrbringen nicht mehr die Ausnahme, sondern die Regel.

Es ist deshalb unvorhergesehen, weil es im herkömmlichen maßgeblichen Zeitpunkt des Inverkehrbringens bzw. der Inbetriebnahme nur bedingt oder gar nicht im Rahmen der Risikobeurteilung bestimmt und als Risiko beschrieben werden kann. Bei diesem unvorhersehbaren Systemverhalten handelt es sich auch nicht um ein Restrisiko. Vielmehr sind diese Ausprägungen des Systems gerade erwünscht, um z. B. im Betrieb stets den optimalen Wirkungsgrad zu erreichen oder flexibel auf sich häufig ändernde äußere Umstände reagieren zu können. Mit solchem Systemverhalten im Betrieb umzugehen sollte jedoch nicht der Marktüberwachungsbehörde obliegen, sondern dem Hersteller. Seine Reaktion hierauf sollten jedoch nicht, wie bei der herkömmlichen Produktbeobachtung, die Warnung, der Rückruf oder die Rücknahme sein. Vielmehr könnte die Ausrichtung der Pflicht des Herstellers zur Risikobeurteilung auf den gesamten Lebenszyklus bestimmter KI-Systeme sinnvoll sein.

10 Anpassung der Begriffe Produkt und Gesamtheit der Maschine

Zudem scheinen die Begriffe des Produkts und der Gesamtheit von Maschinen nicht mehr auf hochgradig vernetzte Systeme zu passen. Sofern sich das System nach Inbetriebnahme mit anderen Systemen vernetzen kann, so ist dies zwar durch den Hersteller bei der Konstruktion seines Systems zu beachten – durch die nachträgliche Vernetzung muss jedoch nicht zwangsläufig ein neues Gesamtprodukt entstehen. Wenn die sicherheitsrelevante Vernetzung mit anderen Systemen/Maschinen nach Inverkehrbringen bzw. Inbetriebnahme spontan erfolgt, ggf. durch das System selbst initiiert wird und ggf. nicht vorhersehbar ist, mit welchen externen Systemen eine Vernetzung eingegangen wird und wie lange sie aufrechterhalten wird, dann ist es für den Hersteller schwer möglich, seine Risikobeurteilung hinsichtlich dieser Art von Vernetzung abzuschließen. Es stellt sich auch die Frage, ob durch eine solche sicherheitsrelevante Vernetzung ein neues Gesamtsystem entsteht und ob dafür neue Herstellerpflichten entstehen (und dann natürlich, wer hier zum Hersteller wird). Gleichzeitig muss durch den Hersteller jedes Produkts sichergestellt werden, dass eine sicherheitsrelevante Vernetzung nur erfolgt, wenn ein gewisses Maß an IT-Sicherheit bei allen vernetzten Systemen gewährleistet ist.

11 Neue Rechtsbegriffe: „wandelbares Produkt“ und „Produktbegleitungskonzept“

Im Forschungsprojekt wurden als Konsequenz der eben kurz angerissenen Problematik bei veränderbaren Produkten zwei aufeinander aufbauende Lösungsansätze formuliert:

- Schaffung einer Definition für „wandelbare“ Produkte
- Schaffung einer Pflicht des Herstellers zur Einführung eines „Produktbegleitungskonzepts“

In dem Vorschlag „Anpassung des Produktbegriffs“ wird der Produktbegriff dahingehend ergänzt, dass eine Veränderung des Produkts im Rahmen einer „bestimmungsgemäßen“

Veränderbarkeit nicht zur Herstellung eines neuen Produkts führt. Damit sollen besonders veränderbare, wenig transparente und wenig kontrollierbare Produkte (sog. „wandelbare" Produkte) regulatorisch erfasst werden, ohne das mit ihnen einhergehende unvorhersehbare Systemverhalten allein zum Gegenstand der Marktüberwachung bzw. Produktbeobachtung zu machen. Im Zuge der Ausarbeitung alternativer Rechtsvorschriften wird der Begriff „wandelbar" als eine Kombination der Taxonomiedimensionen Veränderbarkeit, Transparenz und Kontrollierbarkeit definiert. Wichtig ist, dass „bestimmungsgemäß" für das jeweilige Produkt genau definiert ist.

In einem zweiten Ansatz wurde ein Konzept für „zeitraumbezogene Pflichten" des Herstellers zur Erhaltung der Sicherheit mit engem Begriff des „wandelbaren" Produkts erarbeitet. Der Hersteller bekommt nach diesem Ansatz die Möglichkeit, auch nach Bereitstellung eines hochgradig veränderbaren, wenig kontrollierbaren und intransparenten Produkts seine Risikobeurteilung zu „aktualisieren" und auf dieser Grundlage die bereits bestehenden Maßnahmen zur Gewährleistung der Sicherheit anpassen zu können.

Als Rechtsfolge eines „wandelbaren Produkts" erarbeitet der Hersteller ein angepasstes „Produktbegleitungskonzept". Den Hersteller treffen nach diesem Ansatz Pflichten zur Gewährleistung der Sicherheit über den gesamten bestimmungsgemäßen Produktlebenszyklus. Das Produktbegleitungskonzept umfasst sowohl das Sammeln von Informationen im Betrieb des Produkts und deren Auswertung als auch das Ergreifen von Maßnahmen. Es geht hier also um Beobachtung und Reaktion auf unvorhergesehenes, aber vorausgesetztes Systemverhalten – in Abgrenzung zur herkömmlichen Produktbeobachtung, die auf unerkannte, aber unerwünschte Risiken reagiert. Dieser regulatorische Ansatz trifft damit Regelungen, die das Verhältnis von Hersteller und Verwender betreffen und ergänzt damit das Vertragsrecht um eine ordnungsrechtliche Dimension.

12 Herausforderungen von vernetzten Produkten und Datenqualität

Um möglichen Problemen bei der rechtlichen Bewertung bei bestimmten externen Vernetzungen zu begegnen, wurden im Projekt Vorschläge für einen

- Produktbegriff für vernetzte Produkte sowie für ein
- „Produktsicherheitsrecht für Daten"

skizziert. Um zu verhindern, dass eine bestimmungsgemäße sicherheitsrelevante Vernetzung eines Produkts mit externen Systemen zu einem neuen Produkt führt, könnte der Produktbegriff im Produktsicherheitsrecht dahingehend konkretisiert werden, dass eine solche bestimmungsgemäße (also durch den Hersteller ausdrücklich vorausgesetzte) Vernetzung nicht zur Fertigung eines neuen Produkts führt. Damit würde klargestellt, dass es allein Aufgabe des Herstellers des Ausgangsprodukts ist, die sicherheitstechnischen Implikationen einer solchen Vernetzung in die Risikobeurteilung einfließen zu lassen. Um ihm dies zu erleichtern, könnte ein „Produktsicherheitsrecht für Daten" hilfreich sein.

Für Datendienstleister kann eine Pflicht zur Gewährleistung und Dokumentation eines bestimmten Sicherheitsniveaus von Daten eingeführt werden, die in sicherheitsrelevanter Weise vernetzten Systemen zur Verfügung gestellt werden. Durch entsprechende Zertifizierungen wird den Herstellern und Verwendern hochgradig vernetzter KI-Produkte die ausführliche Prüfung der Datenqualität abgenommen. Dies erleichtert dem Hersteller eines derart vernetzten Produkts die Risikobeurteilung. Er setzt die Verwendung zertifizierter Daten im Rahmen der

Vernetzung voraus, sodass er bei der Risikobeurteilung von einer bestimmten Datenqualität ausgehen kann. So kann ein regulierter Markt für solche sicherheitsrelevanten Daten geschaffen werden, in dem ein eigenes Produktsicherheitsrecht gilt, das für Rechtssicherheit sorgt.

13 Die ePerson als neues Rechtssubjekt?

Die rechtlichen Erörterungen werden ergänzt durch eine Untersuchung des Begriffs der ePerson. In der öffentlichen Diskussion über „autonom agierende Systeme“, die über „Intelligenz“ verfügen, wird seit einigen Jahren im Zusammenhang mit Entscheidungen, die durch die Systeme „selbst“ vorgenommen werden und von direktem menschlichem Einfluss abgekoppelt sind, von der Notwendigkeit gesprochen, wegen der inhaltlichen Abkopplung auch eine juristische Abkopplung – zumindest in Haftungsfragen – vorzunehmen. Ergebnis dieser Überlegungen ist eine „elektronische Person“. Dieses Konzept wird insbesondere im Zusammenhang mit Robotern diskutiert und würde auf eine umfassende Reform hinauslaufen, bei der neben natürliche und juristische Personen eine neue Art von Rechtssubjekten gestellt würde.

Die rechtliche Analyse kommt zu dem Fazit, dass das Konzept der ePerson die komplexen haftungsrechtlichen Probleme beim Einsatz von KI-Systemen nicht befriedigend lösen kann. Ihre Einführung würde keinen Gewinn an Rechtssicherheit bedeuten, sondern voraussichtlich eine Reihe neuer Rechtsunsicherheiten aufwerfen, insbesondere zur Identität der ePerson, ihrer praktischen Umsetzung im Recht und den drohenden Unwuchten bei der Verantwortungsverteilung. Sinnvoller erscheint es, sektorspezifisch für KI-Systeme das Recht dort weiterzuentwickeln, wo die Technik tatsächlich auf eine Lösung drängt. Mit einem sektorspezifischen Ansatz lassen sich gezieltere und sachgerechtere Lösungen entwickeln, die auf bestehende Haftungskonzepte der Beweislastumkehr und Gefährdungshaftung in anderen Rechtsgebieten aufsetzen können. Dem Gesetzgeber bietet sich damit ein breites Instrumentarium zur Regulierung der KI-Technologie, wobei viele vermeintliche Probleme bereits mit dem bestehenden Recht und vorsichtigen Anpassungen ausgezeichnet gelöst werden können.

Zitiervorschlag
Jürgensohn, Thomas et al.: 2021. Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. baua: Fokus.

Im Text wird eine geschlechtergerechte Sprache verwendet. Dort, wo das nicht möglich ist oder die Lesbarkeit eingeschränkt würde, gelten die personenbezogenen Bezeichnungen für alle Geschlechter.

Impressum | Herausgeber: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Friedrich-Henkel-Weg 1-25, 44149 Dortmund | Telefon: 0231 9071-2071, E-Mail: *info-zentrum@baua.bund.de*, Internet: *www.baua.de*, Gestaltung: R. Grahl | DOI: 10.21934/baua:fokus20210602 | Juni 2021

baua: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin