Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme

baua: Bericht

baua
Bundesanstalt für Arbeitsschutz
und Arbeitsmedizin

Forschung
Projekt F 2432

T. Jürgensohn
C. Platho
D. Stegmaier
M. Hartwig
M. Krampitz
L. Funk
T. Plass
H. Ehrlich

Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme

1. Auflage 2021
Dortmund

Diese Veröffentlichung ist der Abschlussbericht zum Projekt F 2432 "Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme" im Auftrag der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

Zitiervorschlag:
Jürgensohn, Thomas; Platho, Christina; Stegmaier, David; Hartwig, Matthias; Krampitz, Mathilde; Funk, Lorenz; Plass, Timon und Ehrlich, Heiko, 2021. Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. baua: Bericht, F 2432.

Autorinnen/Autoren: Thomas Jürgensohn, Christina Platho
HFC Human-Factors-Consult GmbH

David Stegmaier, Matthias Hartwig, Mathilde Krampitz
Lorenz Funk, Timon Plass
Institut für Klimaschutz, Energie und Mobilität e.V.

Heiko Ehrlich
TÜV Nord

Titelfoto: nay/iStock.com

Umschlaggestaltung: Susanne Graul
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

Herausgeber: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)
Friedrich-Henkel-Weg 1 – 25, 44149 Dortmund
Postanschrift: Postfach 17 02 02, 44061 Dortmund
Telefon 0231 9071-2071
Telefax 0231 9071-2070
E-Mail *info-zentrum@baua.bund.de*
Internet *www.baua.de*

doi:10.21934/baua:bericht20210423 (online)

www .baua.de/dok/8859246

Inhalt

Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme

Kurzreferat

Zentrale Fragestellung des Forschungsvorhabens war, ob und inwieweit der Einsatz von KI-Algorithmen (KI – Künstliche Intelligenz) in physischen Systemen der Industrie, Änderungen am bestehenden Rechtsrahmen wie beispielsweise Produktsicherheits- und Betriebssicherheitsrecht erforderlich machen. Aus der Verwendung datengetriebener Algorithmen, wie sie in maschinellen Lernverfahren, einem Teilgebiet der KI, zu finden sind, ergeben sich Besonderheiten gegenüber herkömmlicher Software. Diese betreffen Aspekte, wie z.B. die Nachvollziehbarkeit, die Vorhersehbarkeit, die Spezifizierbarkeit, die Robustheit oder die Transparenz. Für die rechtlichen Betrachtungen besonders relevant ist die neue Verfahrensweise, bei der durch die Verwertung vieler Daten in einem Trainingsprozess die Möglichkeit besteht, diese durch Erweiterung des Trainingsvorgangs in die Zeit des betrieblichen Einsatzes an die spezifischen Einsatzverhältnisse anzupassen. So sind beispielsweise weiterlernende Systeme – also solche, die sich im Betrieb auf Basis neuer Daten verändern – rechtlich nicht mehr nach den Kriterien des Produktsicherheitsrechts beherrschbar, da der maßgebliche Zeitpunkt der Risikobeurteilung (die Inbetriebnahme) die Veränderungen des Systems nach der Inbetriebnahme ausblendet.

Als Bewertungsgrundlage wurde ein als Taxonomie ausgearbeitetes Kategoriensystem entwickelt, das Faktoren der Sicherheit in software-physischen (KI-)Systemen unter besonderer Berücksichtigung neuerer technischer Entwicklungen übersichtlich darstellt. Wesentliche Grundlage dieser Taxonomie waren umfangreiche Expertenbefragungen und ergänzende Inhalte aus der Literatur.

Im Rahmen des Projekts wurden verschiedene Vorschläge zur Weiterentwicklung des Produktsicherheitsrechts ausgearbeitet. In dem Vorschlag „Anpassung des Produktbegriffs“ wird der bestehende Rechtsbegriff ‘Produkt‘ dahingehend ergänzt, dass eine Veränderung des Produkts im Rahmen einer „bestimmungsgemäßen“ Veränderbarkeit nicht zur Herstellung eines neuen Produkts führt. Wichtig ist, dass „bestimmungsgemäß“ für das jeweilige Produkt genau definiert ist. Im Zuge der Ausarbeitung alternativer Rechtsvorschriften wird der Begriff „wandelbar“ als eine Kombination mehrerer Dimensionen der entwickelten Taxonomie definiert.
Als Rechtsfolge eines „wandelbaren Produkts“ wird ein angepasstes „Produktbegleitungskonzept“ erarbeitet. Den Hersteller sollten nach diesem Ansatz Pflichten zur Gewährleistung der Sicherheit über den gesamten bestimmungsgemäßen Produktlebenszyklus treffen. Das Produkt*begleitung*skonzept 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 – in Abgrenzung zur herkömmlichen Produkt*beobachtung*.

Schlagwörter:
KI, autonom, Sicherheit, Safety, funktionale Sicherheit, Intelligenz, Taxonomie, wandelbare Maschine, Produktsicherheitsrecht

Legal framework for making available autonomous and AI systems

Abstract

The goal of the project presented here was to identify necessary changes to existing legal regulations such as the product safety law and the industrial safety law due to the use of AI-based algorithms and other software with autonomy features in physical systems. The overview is limited to industrial systems for which a safety assessment is required. In order to be able to answer these central questions of the research project, a system of categories (i.e. taxonomy) was developed, which summarizes safety-related factors in recently developed software-physical (AI-based) systems. The taxonomy was based on extensive expert interviews and supplemented by literature analysis.
Complex data-driven algorithms as found in machine learning, a subfield of AI, differ from conventional software. They are characterized by aspects such as understandability, predictability, ability for specification, robustness or transparency. Particularly relevant for the legal considerations is the fact that data-driven system behavior can be tailored it to its specific application by extending the data-based training process into its operational use. Systems that continue to learn - i.e. systems that use data collected in operational use to modify their behavior - are not considered in product safety law, since changes to the system after commissioning are not taken account of in the risk assessment, which is carried out in the process of commissioning (and not afterwards).
Within this project, various proposals for the enhancement of product safety law were developed. In the proposal "modification of product definition", the product definition is extended by stating that changes to the product do not create a new product as long as the changeability is intended. It is important that the functionality intended is precisely defined for the product in question. During the course of drafting this alternative legislation, the term "mutability" is defined as a combination of several dimensions of the developed taxonomy.
As a legal consequence of a "mutable product", an adapted "product support concept" is developed. According to this approach, the manufacturer is obliged to ensure safety over the entire intended product life cycle. The product support concept includes both the collection and evaluation of information during the operation of the product as well as taking appropriate measures. In contrast to conventional product monitoring, this approach is based on both observing and reacting.

Keywords:
AI, autonomous, safety, safety, functional safety, intelligence, taxonomy, mutable machine, product safety law.

1 Einleitung

1.1 Ausgangslage

Die zunehmend günstiger werdenden Prozess- und Mikrorechner, die Möglichkeiten der Datenvernetzung – auch zu Sensoren – sowie die damit einhergehenden Fortschritte in der Softwareentwicklung haben dazu geführt, dass einerseits die Komplexität vieler technischer Systeme immer mehr steigt und dass sich andererseits der Charakter der Systeme von dem eines deterministischen Automaten, dessen Zustände berechenbar und vorhersagbar sind, hin zu Systemen zu wandeln beginnt, die zwar deterministisch sein können, wegen der Komplexität der internen Zusammenhänge aber nach außen den Eindruck stochastischer, chaotischer oder in einem anderen Sinne regelloser Systeme vermitteln. Dies trifft insbesondere auf Systeme mit Algorithmen der KI und hier im Besonderen auf solche mit künstlichen neuronalen Netzen zu. Insbesondere bei selbstoptimierenden Systemen, die sich im laufenden Betrieb an die konkreten Rahmenbedingungen des Systems anpassen (sog. weiterlernende Systeme), ist die Vorhersage des konkreten Verhaltens im Einzelnen erheblich erschwert.
Adaptive und (weiter)lernende Systeme haben den Vorteil, dass sie in unterschiedlichen Anwendungsfeldern einsetzbar sind und sich an die prozessoralen und situativen Gegebenheiten anpassen können. Sie weisen aber auch den eben erwähnten Nachteil auf, dass ihre Reaktionen nicht mehr deterministisch berechenbar sind. Zulassungs-, Genehmigungs-, Zertifizierungs- oder Konformitätsprüfungsverfahren sind so vor eine große Herausforderung gestellt, da ihre Ziel- und Qualitätsdefinitionen, Prüfungsmaßstäbe und -verfahren sich an dieserart deterministischen Aktions-Reaktions-Erwartungen ausrichtet.
Diese Änderungen in der Art von Maschinen haben zunächst Auswirkungen auf die Prozeduren von technischen Prüfungen, wie sie beispielsweise von technischen Sachverständigen und Prüfungsstellen durchgeführt werden. Welche Eigenschaften des Systems werden durch das Prüfsiegel belegt? Sie haben aber möglicherweise auch Auswirkungen auf die rechtliche Bewertung. Wer ist schuld, wenn ein lernfähiger Roboter einen Schaden verursacht – der Hersteller, der Besitzer/Betreiber, der Inbetriebsetzende, die Gesellschaft? Intensiv wird dies aktuell im Zusammenhang mit autonom fahrenden Fahrzeugen diskutiert. Wegen der aktuellen Regelungen in diesem Bereich sind diese Systeme bisher so ausgelegt, dass bei allen bisherigen Konzepten immer der Fahrer die letzte Handlungsinstanz innehält und so im Falle eines Versagens Hauptbezugspunkt der rechtlichen Verantwortungsallokation bleibt.
Das Fehlen einer deterministischen Nachvollziehbarkeit eines Softwarecodes ist eines der charakteristischen Merkmale von KI-Systemen in Form von künstlichen neuronalen Netzen. Dies gilt auch dann, wenn sie sich nicht im Betrieb durch Weiterlernen verändern, sondern nach Auslieferung strenggenommen deterministische Automaten sind. Das Verhalten dieser KI-Systeme ergibt sich aus übergeordneten Gütemaßen oder Zielkriterien auf Grundlage von vielen Datensätzen, die in einem Trainings- bzw. Lernprozess das Eingangs-Ausgangs-Verhalten bestimmen. Das Verhalten ist also nicht explizit vorgegeben, sondern entsteht implizit im Iterationsprozess. Je abstrakter diese Zielvorgaben sind, desto mehr entfernt sich das Softwareverhalten von der expliziten Vorgabe hin zu Verhalten „aus sich heraus“. Nach Außen entsteht der Eindruck, dass die Software selbstständig agiert. Dieses in

vorliegendem Bericht als „Emergenz“ bezeichnete Phänomen zeigen alle auf Gütekriterien fußenden, iterativ entstehenden Algorithmen. Aus sich heraus agierende Softwaresysteme zeigen das Merkmal von Unabhängigkeit, da sich das Verhalten mehr oder weniger losgelöst von expliziten Vorgaben durch Code „verselbstständigt“ ergibt. Generell kann Systemen Autonomie zugesprochen werden, wenn sie eine gewisse Unabhängigkeit zeigen – so beispielsweise Unabhängigkeit von einer menschlichen Kontrolle wie bei vollautomatisierten Fahrzeugen oder wie bei mobilen Robotern ohne Fernsteuerung.

Wenn technische Systeme, die zunehmend Emergenz bzw. Selbstständigkeit zeigen, den Anschein machen, sie würden eigene Bewertungen durchführen und eigene Entscheidungen fällen [1], liegt die Vermutung nahe, der Mensch verschwände als einzig möglicher Adressat rechtlicher Verhaltensanforderungen (präventiv) und Verantwortungsallokation (repressiv) aus dem Betrieb des jeweiligen Systems. Dadurch könnten insbesondere rechtliche Anforderungen aus dem Bereich der Produkt- und Betriebssicherheit, aber auch verhaltenssteuernde Regelungen an Bedienungs- und Pflegepersonal etc. keinen Adressaten mehr finden. Die Anforderungen an die Produktsicherheit wären dadurch erheblich komplexer. Rechtsstandards könnten bei vollautonomen Systemen letztlich nur noch vor bzw. bei Inbetriebnahme (Genehmigung, Zulassung) sowie anlass- oder zeitpunktbezogen (Produktüberprüfung und -wartung) geprüft und garantiert werden. Alle betroffenen Rechtsbereiche müssten u.U. auf dieses Grundphänomen hin gänzlich neu durchdacht werden, damit Anforderungen der Produktsicherheit, aber auch der grundlegenden Produktfunktionalität (insbesondere reibungslose Einpassung in die für das Produkt vorgesehenen Abläufe) bzw. Produktleistungsfähigkeit eingehalten werden können.

Das geltende Produktsicherheitsrecht geht in seinen Sicherheitsstandards und den dafür vorgesehenen Prüfprozessen von deterministischen Produktreaktionen (aus A folgt immer B) aus. Bei emergenten und KI-Systemen mit ihrer Quasi-Indeterminiertheit [2] kann jedoch weder der Normgeber noch der Produktprüfer die Reaktion des Produkts sicher voraussagen. Konnte in einer technischen Anforderung an ein Produkt bisher ein bestimmtes Ergebnis von einem Produkt erwartet und gefordert werden, wird für solche Systeme zunehmend nur noch ein Ergebnisraum mit einer stochastischen Wahrscheinlichkeit vorausgesagt werden. Da auch das Recht bisher grundsätzlich final strukturiert ist (ein fester Tatbestand zieht eine feste Rechtsfolge nach sich), stellen quasi-nicht-deterministische Systeme völlig neue Anforderung an das Produktsicherheitsrecht.

Die eben geschilderte rechtliche Problemstellung wird noch weiter dadurch verschärft, dass der Übergang zwischen herkömmlichen Produkten und emergenten KI-Systemen fließend ist. Wenn es also beispielsweise Änderungen im Produktsicherheitsrecht für emergente KI-Systeme geben sollte, müsste geklärt werden, ab wann ein System in diese Klasse fallen würde. Dies ist vor dem Hintergrund der enormen Unterschiede in der Komplexität der Systeme mit großen Herausforderungen verbunden. Weiterhin muss geprüft werden, ob allein die Datenakquise als Merkmal eines solchen Systems verstanden werden darf. Beispielsweise zeichnen sich autonom und selbstständig

[1] An dieser Stelle soll der Gebrauch von Begriffen, wie „selbst“ und „eigen“ in Verbindung mit Maschinen nicht problematisiert werden. In der Tat bedarf es aber wegen der im Vergleich zur Verwendung im Zusammenhang mit Menschen verbogenen Semantik einer Klärung.

[2] Meint: Die Systeme sind eigentlich deterministisch, haben aber wegen der Komplexität in der praktischen Wirkung nichtdeterministische Züge.

agierende Roboter dadurch aus, dass sie die Topologie ihres Einsatzbereichs – z. B. eine Maschinenhalle – durch selbstständiges oder geführtes Abfahren lernen und sich auf Basis der so ermittelten topologischen Karte zielgerichtet in der gelernten Umgebung bewegen können. Dieses Lernen ist sicherlich von anderer sicherheitsbezogenen Qualität als ein Roboter, der aus gelernten Erfahrungen seine maximale Fahrgeschwindigkeit selbst heraufsetzt und damit seine Gefährlichkeit selbstständig erhöht.
Die in der Öffentlichkeit geführte Diskussion über die mit KI-Systemen verbunden Besonderheiten und Gefahren werden dominiert von Problemen, wie sie in reinen Informationssystemen wie Spracherkennern oder Entscheidungshilfesystemen zu finden sind. Hier spielen häufig Probleme des Datenschutzes oder von Diskriminierung eine Rolle [3]. Untersucht werden dort Aspekte wie Erklärbarkeit und Nachvollziehbarkeit und Kontrolle.[4] Andere Aspekte kommen aber hinzu, wenn diese Softwaresysteme Bestandteil physischer Systeme sind. Von diesen software-physischen Systemen kann potentiell eine primäre Gefährdung für Leib oder Gut ausgehen. Obwohl es auch bei reinen Informationssystemen indirekt durch psychische Wirkung zu körperlichen Schäden kommen kann, fehlt das Element der unmittelbaren Schädigung, das oft mit Unfällen in Verbindung steht. Beispiele potentiell gefährdender software-physischer Systeme sind hochautomatisierte Kraftfahrzeuge, selbstfahrende Industrieplattformen oder auch komplexe Industriesteuerungen. Die Gefahren können bis hin zu der Explosion von Kernkraftwerken reichen.
Wegen der mitunter hohen negativen Folgen sind die Anforderungen an das Inverkehrbringen software-physischer Systeme bezüglich des Nachweises ihres sicheren Verhaltens traditionell sehr hoch, wobei die Anforderungen mit zunehmender Folgenhöhe stark steigen. Es ist klar, dass der Nachweis sicheren Verhaltens wesentlich beeinträchtigt ist, wenn das Verhalten eines Systems nur im statistischen Sinne bewertbar ist. Einzelne Ausreißer in einem statistisch vermeintlich sicheren System könnten zu Fehlern mit u.U. verheerende Folgen führen.
Die Untersuchung der Problematik des Einsatzes von KI-Algorithmen in software-physischen Systemen mit den möglichen Folgen für Leib und Gut ist im Vergleich zur Betrachtung von KI in Informationssystemen bislang stark vernachlässigt. Anliegen dieses Projektes ist es, diese Lücke zu füllen.

1.2 Ziele des Projektes

Ziel des Forschungsvorhabens ist es einerseits, eine Abschätzung zu geben, wie sich der aktuelle technologische Entwicklungsstand von software-physischen Systemen mit KI-Komponenten darstellt, und zudem zu untersuchen, welche Besonderheiten sich durch Einbeziehung von KI-Komponenten auftun. Der Schwerpunkt der Untersuchungen liegt im Bereich industrieller Anwendungen, beispielsweise der Automatisierungstechnik, Industrie 4.0, etc., die stark durch das Produktsicherheitsgesetz geprägt sind. Da dort KI-Systeme bisher nur sehr rudimentär

[3] Whitepaper „Vertrauenswürdiger Einsatz von Künstlicher Intelligenz“, Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIS (Hrsg.) St. Augustin, 2019
[4] Zweig, K. A. (2019). *Algorithmische Entscheidungen: Transparenz und Kontrolle*. Konrad-Adenauer-Stiftung.

eingesetzt werden, wurde auch die Thematik hochautomatisierter (autonomer) Kraftfahrzeuge bei der Projektdurchführung berücksichtigt. Im vorliegenden Projekt stehen also Folgen im Vordergrund, die aus der Safety-Sicht [5] in Verbindung mit KI stehen.
Ziel des Forschungsvorhabens ist andererseits zu untersuchen, inwieweit der Einsatz von KI in software-physischen Systemen mit den oben angedeuteten Besonderheiten, Änderungen im Rechtssystem notwendig machen. Es soll detailliert geprüft werden, ob sich mit der beschriebenen Klasse neuer Systeme zusätzlichen 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. Es soll konkret 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. Zudem soll betrachtet werden, wie das Recht mit derartigen quasi-nichtdeterministischen Systemen umgeht. Die stete Veränderbarkeit sog. weiterlernender Systeme in Verbindung mit einer entsprechenden Intransparenz und verringerter Kontrollierbarkeit wirft insbesondere die Frage auf, welchen Wert zeitpunktbezogene Prüfungen noch haben und wer dauerhaft für die Sicherheit solcher Systeme verantwortlich sein soll. Es sollen Analysen des Anpassungsbedarfs des bestehenden rechtlichen und administrativen Rahmens durchgeführt und Vorschläge für geänderte Vorschriften und Regelwerke erarbeitet werden. Dabei werden verschiedene Rechtsbereiche in den Blick genommen und insgesamt als Regelungsgefüge verstanden.
Ein weiteres Ziel im Projekt ist die Entwicklung einer Taxonomie software-physischer (KI-)Systeme. Diese Taxonomie sollte als Ordnungsschema für die rechtliche Bewertung dienen.

1.3 Vorgehen im Projekt

Bevor das Kernthema der rechtlichen Bewertung angegangen werden konnte, wurden im Rahmen des Projektes gegenwärtige Entwicklungen der KI und ihre Anwendung im industriellen Kontext und bei fahrerlosen Fahrzeugen im Verkehr untersucht. Dazu wurden Experten und Expertinnen aus den Bereichen KI, Smart-Home, Robotik und funktionale Sicherheit bezüglich ihrer Einschätzung gegenwärtiger und zukünftiger Entwicklungen sowie bezüglich möglicher Gefahren in Interviews befragt. Eine weitere Kernfrage in den Interviews war, inwieweit sich durch den Einsatz von KI in der Industrie die Notwendigkeit ergibt, die etablierten Vorgehensweises des Nachweises von Sicherheit zu modifizieren und ab dies mit der Notwendigkeit zu Änderungen rechtlicher Art einhergehen könnte.
Um für die Rechtsuntersuchungen einen Handlungsrahmen zu schaffen, wurde ein als Taxonomie ausgearbeitetes Kategoriensystem entwickelt, das Faktoren der Sicherheit in software-physischen (KI-)Systemen unter besonderer Berücksichtigung neuerer

[5] Schwerpunkt der Untersuchungen in diesem Bericht ist Sicherheit im Sinne von Safety. Sicherheit und Safety werden im Folgenden synonym verwendet und inkludieren Felder wie Produktsicherheit, funktionale Sicherheit, Betriebssicherheit, Anlagensicherheit, Sicherheit der beabsichtigten Funktionalität (SOTIF), etc

technischer Entwicklungen kategorial zusammenfasst. Die Struktur wurde vornehmlich auf die Belange der rechtlichen Bewertungen in diesem Bericht ausgerichtet, kann aber auch darüber hinaus für andere Anwendungskontexte eingesetzt werden. Im Gegensatz zu anderen, aus der Untersuchung von Einflüssen der KI entstammenden Taxonomien, deren Fokus auf reiner Datenverarbeitung liegt [6], ist die entwickelte Taxonomie konsequent auf die Belange der Sicherheit in dem eben festgelegten Sinne von Safety ausgelegt.
Neben Einzelinterviews von Experten und Expertinnen wurde ein Expertenworkshop durchgeführt, in dem Ergebnisse der rechtlichen Bewertungen und die entwickelte Taxonomie diskutiert und zur Korrektur oder Erweiterung ebenjener Taxonomie genutzt wurden.

1.4 Festlegung des Gegenstandsbereichs des Projektes

Zentrale Frage des Projektes ist, ob und wenn ja, wie die rechtlichen Bedingungen, die den Nachweis der Sicherheit von Maschinen, Anlagen, etc. regeln, modifiziert werden müssen, wenn deren Verhalten nicht vollständig explizit vorgegeben und demnach „autonom“ (siehe Kapitel 2.1) ist oder der KI zuzuordnen ist (siehe Kapitel 2.2). Die Autonomie als ein Merkmal der „relativen Unabhängigkeit im Systemverhalten“ erschwert den Nachweis der Sicherheit und erfordert u. U. andere als die bekannten, traditionellen Herangehensweisen des Sicherheitsnachweises.
Eine derartige Problematik findet sich beispielsweise bei Systemen, die künstliche neuronale Netze und hier insbesondere tiefe neuronale Netze enthalten. Die möglichen Zustände sind so zahlreich, dass sie nicht mehr vollständig kontrolliert werden können. Die Einflussmöglichkeiten von Entwicklern auf das Systemverhalten verlagert sich von der Festlegung des Codes in Richtung der Festlegung von verhaltensbestimmenden Daten.
Nicht bei allen KI-Algorithmen treten diese potentiellen Probleme auf. Beispielsweise können Algorithmen der symbolischen KI bezüglich der Sicherheitsüberprüfung wie herkömmliche Software behandelt werden. Ebenso ist das Merkmal der Autonomie nicht an KI-Systeme gebunden.
Der Schwerpunkt der Anwendungsdomänen dieses Berichts liegt im Bereich Industrie, Automatisierungstechnik, Anlagen, etc. Allerdings werden dort bisher kaum Systeme eingesetzt, die sich durch außergewöhnliche Autonomie (im obigen Sinne) auszeichnen oder KI-Systeme im sicherheitsrelevanten Einsatz integriert haben. Deshalb wurden auch die Erfahrungen aus dem Automobilbereich zu selbstfahrenden Fahrzeugen in die Überlegungen eingebunden. Als Analogon wurden selbststeuernde mobile Roboter oder mobile Plattformen mit in den Betrachtungshorizont einbezogen.
Dem Schwerpunkt der Anwendungsdomänen entspricht die Schwerpunktsetzung bei den untersuchten Rechtsgebieten. Die Untersuchung konzentriert sich auf das Produktsicherheitsrecht, das Recht des technischen Arbeitsschutzes, das Immissionsschutzrecht, das IT-Sicherheitsrecht und das Haftungsrecht. Neben einer Betrachtung des aktuellen Standes dieser Rechtsgebiete werden Ansätze entwickelt und erörtert, wie diese Rechtsgebiete weiterentwickelt werden können, um den

[6] Z. B. Whitepaper „Vertrauenswürdiger Einsatz von Künstlicher Intelligenz“, Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIS (hrsg.) St. Augustin, 2019

identifizierten Herausforderungen mit den untersuchten KI-Systemen gerecht werden zu können.

2 Begriffsklärungen

Gemäß dem Titel des Projektes „Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme“ stehen im Mittelpunkt des Projektes Systeme, die mit den Attributen „autonom“ und „KI“ in Verbindung belegt sind und zwar speziell im Betrachtungskontext von Sicherheit (Safety) und noch konkreter im Lichte juristischer Überlegungen und zusätzlich in einem sehr breiten Anwendungsgebiet, in dem sehr unterschiedliche software-physische Systeme möglich sind. In dieser Kombination treffen drei sehr unterschiedliche Fachrichtungen aufeinander, die häufig ein unterschiedliches Verständnis der Nomenklatur aufweisen. „Autonom“ und „KI“ sind Begriffe, deren Semantik in dem durch den Projekttitel aufgespannten Rahmen originär aus der Informatik und dort speziell aus einem Teilbereich, der „Künstlichen Intelligenz“ (KI) stammt. Safety, insbesondere in dem Bereich der Maschinensicherheit, ist durch die Ingenieurwissenschaften geprägt, in denen sich indes eine andere Nomenklatur etabliert hat. So wird dort beispielsweise die Übernahme von Tätigkeiten, die vormals der Mensch durchgeführt hat, durch Maschinen, als Automatisierung bezeichnet. In der KI sind Systeme ohne menschlichen Eingriff durch das Attribut „autonom“ gekennzeichnet. Entsprechend finden wir in der einen Domäne „Stufen der Automatisierung“, die in der anderen Domäne als „Stufen der Autonomie“ bezeichnet werden. In beiden Fällen werden die Stufen durch die zunehmende Ersetzung des Menschen charakterisiert. Die Problematik wird zusätzlich dadurch verschärft, dass außerdem noch die Fachrichtung Rechtswissenschaft mit ihren eigenen Fachbegriffen und festgelegten Semantiken hinzukommt. So kann aus rechtlicher Hinsicht nur ein Mensch „handeln“[7], in der durch die KI getriebenen öffentlichen Diskussion wird aber teilweise auch schon von „autonom handelnden Maschinen“ gesprochen[8].

Zu den bezüglich ihrer Bedeutung zu klärenden Begriffen KI und autonom kommen weitere Begriffe wie „selbst“ und „Entscheidung“ hinzu, die fest im allgemeinen Sprachgebrach verankert sind, im Zusammenhang mit KI in dem dortigen Gebrauch aber missverstanden werden können.

Der vorliegende Bericht hat den Anspruch, fachrichtungsunabhängig verständlich zu sein. Dazu wurden teilweise fachspezifische Begriffe vermieden und verschiedene eigene Definitionen entwickelt. Es ließ sich nicht verhindern, dass dabei Auffassungen über Begriffe entstanden sind, die mitunter von den Fachdisziplinen u.U. so nicht mitgetragen werden.

Bevor also der durch den Titel des Projektes gefasste Gegenstandsbereich umrissen werden kann, sollen zunächst die wichtigsten Schlüsselbegriffe bezüglich der hier vertretenen Auffassung geklärt werden.

---

[7] Handlung ist jedes willensgetragene menschliche Verhalten, durch das die Außenwelt verändert wird (Zivilrecht). Handlung als jedes willensgetragene, bewusst vom Ziel her gelenktes (zweckgerichtetes) menschliches Verhalten, durch das die Außenwelt verändert wird (Strafrecht). Ein Roboter handelt dagegen nach einem Programm und hat Entscheidungsspielräume nur soweit sie programmiert wurden. Es fehlt die Willensfreiheit das erforderliche Mindestmaß an Reflektion unter Kenntnis und Möglichkeit der Abwägung der Folgen. Daher können Computer und Roboter in rechtlicher Sicht nicht handeln.

[8] Siehe z.B Bundesministerium für Wirtschaft und Energie (Hrsg.), Technologieszenario „Künstliche Intelligenz in der Industrie 4.0“

2.1 Festlegung der Begriffe Autonomie, autonom

In der Öffentlichkeit stark verbreitet und bekannt ist das Attribut „autonom“ im Zusammenhang mit Kraftfahrzeugen, die ohne Fahrereingriff im Straßenverkehr fahren. Ein autonomes Fahrzeug ist dabei die finale Endstufe zunehmender Automatisierung. Allerdings gibt es dafür auch andere Bezeichnungen wie „vollständig automatisiertes Fahrzeug“. So wird in dem Klassifikationsschema der SAE J3016 von 2016[9] die oberste der 6 Stufen, die den Grad der Automatisierung in Bezug auf die Rolle des Menschen in einem Kraftfahrzeug angeben, als *Full Driving Automation* bezeichnet. In derselben Veröffentlichung wird darauf hingewiesen, dass autonom keine passende Bezeichnung im Zusammenhang mit automatisiertem Fahren sei, weil es um den Prozess der Ersetzung des Menschen durch technische Systeme geht, also um Automatisierung im gebrauchten Wortsinn[10].
Auch in Automatisierung oder Automat steckt das auto (selbst) – αὐτόματος[11] bedeutet ein sich „selbstständig“ Bewegendes oder „von selbst Geschehendes“[12]. Ein Automat ist ein von Menschen konzipiertes technisches System, das etwas selbstständig tut. Im allgemeinen Sprachgebrauch passiert etwas „automatisch“, wenn es von selbst geschieht.
Im Gebrauch beider Begriffe automatisch und autonom wird das „unabhängig-von“ und „selbstständig“ adressiert. Bei beiden Begriffen ist es häufig die Unabhängigkeit vom menschlichen Bediener oder von menschlichem Einfluss, die ausgedrückt werden soll. In vielen Diskussionen zum Autonomiebegriff in der KI-Literatur wird versucht, eine Abgrenzung zu „automatisch“ dadurch zu begründen, dass in autonom auch automatisch enthalten ist, der Begriff autonom aber noch darüber hinaus gehende Semantik enthält. Häufig sind dies die Aspekte „sein eigenes Verhalten kontrollieren“ oder „seine Ziele verfolgen“[13]. Die Nähe von automatisch und autonom bzw. Automation und Autonomie wird deutlich, wenn man sich den Sprachgebrauch zweier technischer Systeme anschaut. Ein System, das eine Notbremsung unabhängig vom Fahrer einleitet, wird sowohl „Automatisches Notbremssystem“[14] als auch „Autonomes Notbremssystem“[15] genannt. Ein fahrerloses Transportsystem, das Waren in

---

[9] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE J 3016, 2016

[10] Automatisierung wird in der Regel als technische Maßnahme bezeichnet, bei der technische Systeme die Arbeit von Menschen übernehmen. Nach DINV 19233 bezeichnet Automatisierung „Das Ausrüsten einer Einrichtung, so dass sie ganz oder teilweise ohne Mitwirkung des Menschen bestimmungsgemäß arbeitet.“ (DIN V 19233: Leittechnik – Prozessautomatisierung – Automatisierung mit Prozessrechensystemen, Begriffe. Deutsches Institut für Normung e. V.). Sehr oft wird in diesem Zusammenhang davon gesprochen, dass Maschinen *selbstständig* etwas leisten.

[11] https://de.wiktionary.org/wiki/Automat

[12] „Im Jahr 1745 erfand der englische Schmied Edmund Lee eine frühe Vorrichtung zur Automatisierung, durch die sich Windmühlen selbstständig in den Wind drehen“ (aus https://www.wortbedeutung.info/Automatisierung/).

[13] Z.B. in MÜLLER, Vincent C. Autonomous cognitive systems in real-world environments: Less control, more flexibility and better interaction. Cognitive Computation, 2012, 4. Jg., Nr. 3, S. 212-215.

[14] System von ZF https://www.zf.com/products/de/lcv/products_51121.html, Zugriff 27.3.2020

[15] Wabco https://www.autoservicepraxis.de/nachrichten/autobranche/fahrerassistenz-wabco-praesentiert-autonomes-notbremssystem-2507176. Zugriff 27.3.2020

Fabrikhallen transportiert wird sowohl „Automated Guided Vehicle“ (AGV) [16] als auch „Autonomous Guided Vehicle“ (AGV)[17] genannt.
Ein anderes Beispiel, das die semantische Nähe von automatisch und autonom demonstriert sind Versuche, eine durch technischen Fortschritt induzierte Transition der menschlichen Aufgabenerfüllung von ausschließlich vom Menschen und völlig ohne maschinell-informatorische Unterstützung (z.B. durch einen Computer) ausgeführt, bis hin zu rein maschinell ohne menschliches Zutun, in Stufen zu fassen. Schon 1978 teilen Sheridan und Verblank[18] diesen Übergang in 10 Stufen der „Automation bei der Mensch-Computer-Entscheidungsfindung“ ein. Dieses Stufen-Modell wird bis heute in der ingenieurwissenschaftlichen Literatur als Grundlage der Charakterisierung des Grades der Aufgabenallokation in der Entscheidungsunterstützung technischer Systeme herangezogen. Parasuraman et al. erweitern dieses Modell 2000[19] auf ein mehrdimensionales Modell, das diese Automatisierungsstufen auf die Dimensionen „Informationsaufnahme“, „Informationsanalyse“, „Entscheidungsauswahl“, „Handlungsausführung“ auffächert. Onnasch et al. (2016)[20] übernehmen dies in einer Taxonomierung der Mensch-Roboter-Interaktion, wobei sie explizit Automatisierungsgrad und Autonomiegrad gleichsetzen.
Diese Gleichsetzung finden wir auch beispielsweise in BMWi (2019)[21], bei dem die oben erwähnten 6 Automatisierungsstufen (0 bis 5) hoch- bis vollautomatisierter Fahrzeuge (SAE-Level) eins zu eins auf die industrielle Produktion übertragen und dort als „Autonomie-Stufen“ bezeichnet werden. Es stellt sich folglich die Frage: Ist also autonom und automatisiert gleichzusetzen?
Der Unterschied der Begriffsverwendung (aktuell und historisch) ist in vielen Anwendungen weniger durch logische Unterschiede als durch die Provenienz der beteiligten Akteure und parallel und isoliert verlaufende Wissenschaftsentwicklungen zu begründen. Im Maschinenbau geht es vornehmlich um den Arbeitskontext, also das Ersetzen des Menschen bei bestimmten Aufgaben, so in der Fabrikarbeit, aber auch beim Autofahren. In der Informatik liegt der Fokus hingegen mehr auf der Beschreibung von Systemeigenschaften.
Da der Begriff „autonom“ in der heutigen Diskussion originär aus der Informatik kommt, soll kurz auf den historischen Werdegang eingegangen werden.
Der Ursprung heutiger Definitionen von „autonomes System“ liegt nach Auffassung der Autoren in der Informatik und ist zweigeteilt. Eine Quelle der inhaltlichen Belegung des Begriffs „autonom“ sind „autonome Agenten“, die teilweise auch als autonome Systeme[22] bezeichnet werden. Bei Softwareprogrammen werden damit Subsysteme

[16] Sick https://www.sick.com/ag/en/industries/port/container-terminal/ground-transportation/automated-guided-vehicle-agv/c/g366055, Zugriff 27.3.2020
[17] https://acim.nidec.com/motors/motion-control/industries/agv-and-amr. Zugriff 27.3.2020
[18] Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. (Technical Report). Cambridge, MA: MIT Man-Machine Systems Laboratory
[19] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 30(3), 286–297.
[20] Onnasch, L; Maier, X & Jürgensohn, T.. Mensch-Roboter-Interaktion–Eine Taxonomie für alle Anwendungsfälle (BAuA: fokus). Berlin: BAuA, 2016
[21] Bundesministerium für Wirtschaft und Energie (Hrsg.), Technologieszenario „Künstliche Intelligenz in der Industrie 4.0“
[22] Hage, Jaap. Theoretical foundations for the responsibility of autonomous agents. Artificial Intelligence and Law, 2017, 25. Jg., Nr. 3, S. 255-271.

bezeichnet, die eine relative Ausführungsunabhängigkeit gegenüber einer beauftragenden Software oder in einem Verbund mit anderen Agenten haben[23]. Wegen dieser Unabhängigkeit werden sie mit dem Attribut autonom belegt. Weil Ausführungsunabhängigkeit eine Eigenschaft ist, die auf sehr viele unterschiedliche Systeme anwendbar ist, gibt es eine Vielzahl von Definitionen für Agenten[24]. Häufig werden diese Agenten dann durch ein eingrenzendes Adjektiv spezifiziert. Allen diesen unterschiedlichen Definitionen ist aber das Attribut Autonomie in dem eben gemeinten Sinne von selbstständig agierender Agent gemein.

Ab etwa Anfang der 70er Jahren entwickelte sich in der KI die an die Kybernetik angelehnte, die „verhaltensbasierte KI“ als Gegenentwurf zur „symbolischen KI“. Verhaltensbasierte KI ist eine KI, in der Zeit eine Rolle spielt und es einen Output (Verhalten) gibt, der durch die Verarbeitung von Sensordaten berechnet wird. Im Zuge dessen entwickelt sich auch der Begriff des „autonomen Agenten“ als Bezeichnung für ein abstraktes System oder ein Programm, das Information aus der Umwelt aufnimmt und dort bis zu einem gewissen Grad unabhängig von äußeren Steuerungen mit einem Auftrag agiert.

Das Attribut autonom ist wiederum wie bei allgemeinen Agenten der Unabhängigkeit geschuldet[25]. Allerdings werden zusätzliche Anforderungen postuliert, die eine Zuschreibung von KI-Software als „autonomer Agent“ rechtfertigt. Eine der Anforderungen ist beispielsweise, dass autonome Agenten Ziele haben und Pläne verfolgen bzw. sogar neue erzeugen. Die zusätzlichen Charakteristika autonomer Agenten, die über der Eigenschaft des „unabhängig von“ gehen, sind Eigenschaften, die im Zuge der Weiterentwicklungen in der KI entstanden sind, aber nicht aus dem Begriff „autonom“ abgeleitet sind. Im Gegenteil: die semantische Auffassung von „autonom“ hat sich durch einen allmählichen Semantikdrift an den speziellen Gebrauch angepasst.

Die Eigenschaften „autonomer Agenten“ finden sich häufig auch zur Charakterisierung „intelligenter Agenten“ (siehe unten). Eine Abgrenzung dieser beiden Agententypen ist häufig schwierig. Der Gebrauch in der Literatur scheint auf weitgehende Synonymität hinzuweisen.

Die zweite Quelle heutiger Definitionen von „autonomes System“ sind autonome mobile Roboter, die am Anfang eine Domäne der Ingenieurwissenschaften war. Dies waren entweder zoomorphe Roboter[26] oder selbstfahrende Roboter, die sich mit Sensoren und Motoren ausgestattet auf Basis einer Steuereinheit in einer Umwelt bewegen. Dazu gehören auch Roboter mit Rädern, die auch als Fahrzeuge aufzufassen sind. Das Attribut autonom wurde ihnen zugesprochen, weil sie unabhängig von einer externen steuernden Entität – sei es ein Mensch oder eine

---

[23] Agent meint Beauftragter. Luc Steels: When are robots autonomous agents? Robotics and Autonomous Systems Volume 15, Issues 1–2, July 1995, Pages 3-9.

[24] Einige der Definitionen finden sich in Franklin, S., & Graesser, A. (1996, August). Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In International Workshop on Agent Theories, Architectures, and Languages (pp. 21-35). Springer, Berlin, Heidelberg.

[25] Luc Steels: When are robots autonomous agents? Robotics and Autonomous Systems Volume 15, Issues 1–2, July 1995, Pages 3-9

[26] Als einer der ersten Systeme findet sich in der Literatur oft eine elektromechanische schildkrötenartige „Kreatur“ mit dem Namen Machina speculatrix, die sich selbstständig (autonom) bewegte. Der Autor bezeichnet die Schildkröte selbst allerdings nicht als Roboter, sondern als „Synthetik Life“ und benutzt den Begriff autonom nicht. William Grey Walter An Electromechanical Animal, Dialectica (1950) 4(3):206-213

Maschine – also ohne Fernsteuerung agieren[27]. Autonom ist hier folglich wie auch bei den hochautomatisierten Fahrzeugen als Synonym zu „unabhängig vom Menschen“ zu verstehen. Dabei können verschiedene Grade von Autonomie auftreten. So sind die Mars-Rover eigentlich ferngesteuert, haben aber auch in bestimmten Funktionen eine Autonomie. Wenn die mobilen Roboter über Solarzellen verfügen und damit ihre eigene Energieversorgung gewährleisten, sind sie zusätzlich noch autark – eine Autonomie bezüglich eines anderen Merkmals[28]. Wegen der Ähnlichkeit des Übergangs eines nicht mehr ferngesteuerten Roboters zu einem nicht mehr von Menschenhand gelenkten Auto liegt es nahe, auch vollautomatisierte Kraftfahrzeuge mit dem Attribut „autonom“ zu belegen.
Die Autonomieauffassung als „unabhängig vom Menschen“ und die im Begriff „autonome Agenten“ entstandenen zusätzlichen Attribute mischten sich im Zuge der Weiterentwicklungen in der Robotik, die zunehmend Methoden der KI einsetzten. So finden wir heute die Kernelemente der historisch so erwachsenen Vorstellung, was ein „autonomer Agent“ sei, als Sense-Plan/Think-Act bei sehr vielen Beschreibung von autonomen Fahrzeugen, autonomen Systemen oder autonomen Robotern[29].
Allerdings sind diese Attribute wie geschildert nicht aus der Ableitung des Begriffs Autonomie entstanden, sondern als Addendum dazu gekommen. Vor dem Hintergrund der angestrebten disziplinunabhängigen Klärung der Begrifflichkeiten, soll im Folgenden eine eigene Definition von „autonom“ entwickelt werden.
Das obige Beispiel des Mars-Rovers zeigt, dass Autonomie nicht eine absolute Eigenschaft ist, sondern des Zusatzes „bezüglich was“ bedarf. Auch ein selbstfahrendes Fahrzeug ist nur bezüglich bestimmter Aspekte autonom (i.S.v. vom Menschen unabhängig). Die Reisezielvorgabe geschieht in der Regel durch den Menschen, wobei auch für derartige strategische Entscheidungen mit eine erweiterte Autonomie zumindest vorstellbar ist [30]. Auch bei den autonomen Agenten finden wir den Aspekt der Unabhängigkeit, nämlich von exakten Vorgaben des Verhaltens von außen.
In dem ursprünglichen Autonomiebegriff griechischen Ursprungs bezieht sich Autonomie auf die politische Selbstgesetzgebung von Menschen oder Menschengruppen [31]. Wie bei vielen Begriffen hat sich der Gebrauch inzwischen aber gewandelt. Während sich der Autonomiebegriff bei Kant zunächst zum „ethischen Prinzip menschlichen Handelns“[32] wandelt, ist er heute vielfach nur synonym zu

---

[27] Todd, D. J. (1986). Economic and Social Aspects of Robotics. In Fundamentals of Robot Technology (pp. 227-235). Springer, Dordrecht. Siehe auch „An autonomous robot would not be completely controllable and observable by an outside agent (this is what we usually mean by autonomy) aus . McFarland, D. (1992). Animals as cost-based robots. International Studies in the Philosophy of Science, 6(2), 133-153.
[28] Selbstfahrende Automobile, die als autonome Systeme bezeichnet werden, sind beispielsweise in mittelfristiger Zukunft nicht autark.
[29] Z. B. Ein Referenzmodell für vertrauenswürdige KI - Vorstellung VDE-AR-E 2842-61 © DKE
[30] Völlige Autonomie scheint hingegen ausgeschlossen, weil von der Menschheit nicht gewollt.
[31] Autonomie = eigenes Recht: Etwa Mitte des 5. Jahrhunderts v. Chr. ist mit der Forderung griechischer Stadtstaaten nach Autonomie jene nach Selbstständigkeit gemeint, sowie das Recht auf Unabhängigkeit von anderen Mächten. Pohlmann, Rosemarie: Autonomie, in: HWPh Bd. 1 (A-C) 1971, S.. 701-719. Historisches Wörterbuch der Philosophie, Joachim Ritter (Hrsg.), Bd. 1-13, Basel 1971-2007.
[32] Erduana Shala (2014), Die Autonomie des Menschen und der Maschine. Hochschulschrift. am KIT als Magisterarbeit.

„unabhängig von etwas“ gedacht[33]. Bei Kant bezeichnet Autonomie die Eigengesetzgebung des freien Willens, die unabhängig aller empirischen Bedingungen ist und damit erst ethisches Handeln ermöglicht. In heutige Terminologie übersetzt ist es die individuelle Autonomie gegenüber einem rein reizbezogenen Handeln. Kants Autonomiebegriff beschreibt also eine moralische und nicht eine rechtlich-politische Autonomie. Den politischen Aspekt findet man auch im heutigen Gebrauch, wenn von einem Landesteil, der Autonomie erhält, oder einer autonomen Region in einem Staat gesprochen wird. In vielen heutigen Verwendungen bleibt aber nur noch das „unabhängig von“ bzw. das „selbstständig“ als sinnstiftendes Element erhalten – insbesondere, wenn es um Übertragungen auf Nichtmenschliches geht. So wird beispielsweise bei einer „autonomen Differentialgleichung“ autonom nur noch als Synonym zu „unabhängig von“ verwendet. Ähnlich wird eine „autonome Kaufentscheidung“ als eine Entscheidung unabhängig von Bezugsgruppen verstanden[34]. Auch im Politischen gibt es autonome Entscheidungen, wie beispielsweise unabhängig vom Verteidigungsbündnis getroffene politische Entscheidungen eines Bündnispartners. Autonom meint hier ebenfalls „unabhängig von“. Andere Beispiele mit dieser Auffassung von autonom sind: autonome Transposition[35], autonome Musik[36] oder Begriffe aus der Medizin: autonomes Nervensystem[37], Autonomes Adenom[38], Autonome Dysreflexie[39].
Fasst man die Analyse des Autonomie-Begriffs in anderen Kontexten und die Verwendung in „autonome Roboter“, „autonome Fahrzeuge“ und „autonome Agenten“ zusammen, dann ist es das Selbstständig/Unabhängig-Sein von irgendetwas, das den Kern des Begriffs ausmacht. Es lässt sich also zunächst definieren:

Definition **autonom:**

Autonom bezeichnet das unabhängig/selbstständig sein von etwas und bezüglich etwas.

Ein autonomes System ist analog dazu zu verstehen als:

Definition: **autonomes System**

[33] Das Sprachverständnis ist an dieser Stelle zweigeteilt: Texte bis zum Aufkommen „autonomer Agenten“ (siehe weiter unten) in der KI benutzen autonom (beispielsweise auch in der KI) als Synonym für unabhängig. In späteren Texten wird teilweise ein anderes Sprachgefühl sichtbar.
[34] „Die autonome Entscheidung wird im Kaufverhalten verwendet und sagt aus, dass ein Konsument unabhängig von den anderen Familienmitgliedern eine Kaufentscheidung trifft.“ http://www.mein-wirtschaftslexikon.de/a/autonome-entscheidung.php
[35] LTR-lose Retrotransposons mit der Fähigkeit zur autonomen Transposition; d. h. sie sind selbst in der Lage sich im Genom zu verbreiten.
[36] nicht an äußere Zwecke gebunden.
[37] Unglücklicherweise gibt im Englischen noch den Begriff autonomic, der im Zusammenhang mit medizinischen Begriffen wie die eben vorgestellten genutzt wird. Ein autonomes Nervensystem ist ein autonomic nervous systems, autonome Musik ist aber autonomous music.
[38] Erkrankung der Schilddrüse. Autonome, nicht mehr von der Hirnanhangdrüse gesteuerte Funktionen.
[39] Left and right brain hemisphere, independently, i.e. they are autonomous. J. O. WISDOM, A NEW MODEL FOR THE MIND-BODY RELATIONSHIP, The British Journal for the Philosophy of Science, Volume II, Issue 8, February 1952, Pages 295–301, https://doi.org/10.1093/bjps/II.8.295

> Ein autonomes System bezeichnet ein System, das selbstständig von etwas und bezüglich etwas ist. Es agiert in einem definierten Sinne unabhängig

Die Erläuterungen zeigen, dass es keine generelle, „natürliche“ Hierarchie von Autonomiegraden und nicht einmal von Autonomieniveaus geben kann, weil immer der zusätzliche Kontext „von etwas“ und „bezüglich etwas“ festgelegt sein muss. In dem Beispiel der hochautomatisierten/autonomen Fahrzeuge ist es die Unabhängigkeit von einem menschlichen Fahrer bezüglich dessen Tätigkeit des Führens auf Straßen die zu den oben erwähnten SAE-Stufen der „Driving Automation” geführt haben. Die Zahl der Stufen dieser Taxonomie ist den speziellen technischen Randbedingungen und den Entwicklungsständen der Industrie geschuldet.
Eine Übertragung auf andere Bereiche, z.B. die zunehmende Automatisierung im Industriekontext, erfordert eine sorgfältige Prüfung und Anpassung.

2.1.1 Unterscheidung autonom – automatisiert

Wegen der starken Ausrichtung des vorliegenden Projektes auf den Maschinenbau und die Industrie soll zum Abschluss des Kapitels noch versucht werden, Unterschiede zwischen autonom und automatisiert herauszuarbeiten.
Bei beiden Begriffen (autonom und automatisch) ist es die Eigenständigkeit, die begrifflich adressiert wird. Dennoch könnte man einen kleinen semantischen Unterschied konstruieren. Während bei der Automatisierung der Schwerpunkt der Betrachtung auf Eigenständigkeit durch Ersatz menschlicher Handlungen liegt, liegt der Fokus bei der Autonomie auf der Unabhängigkeit vom Menschen oder von anderen Maschinen beziehungsweise Softwaresystemen. Jede Maschine ist autonom bezüglich einer bestimmten Fähigkeit, die im Zuge der Automatisierung vom Menschen übernommen wurde. Deshalb zeigen alle Assistenzsysteme im Automobil Autonomie, auch wenn sie keine KI-Algorithmen verwenden. So hält ein Tempomat im Auto autonom (selbstständig, automatisch) die vorher festgelegte Geschwindigkeit, ist also eine Automatisierung der Aufgabe „Geschwindigkeitsregelung“. Bezüglich der Aufgabe der Geschwindigkeitsregelung (ansonsten Aufgabe des Fahrers) agiert der Tempomat autonom, bezüglich der Geschwindigkeitswahl hingegen nicht. Das ACC-System wählt zusätzlich noch die zu fahrende Geschwindigkeit in Abhängigkeit zum vorausfahrenden Verkehr und agiert bezüglich der Aufgabe der Geschwindigkeitswahl autonom. Nicht autonom ist dasselbe System jedoch bezüglich der Einstellung des Komfortabstandes und bezüglich seiner Aktivierung.
Bei Automatisierung betrachtet man mehr den Ersatz, bei Autonomie mehr die Unabhängigkeit des Systems. Autonomie im industriellen Umfeld wird einem System zugesprochen, das als Automat nach dem Prozess der Automatisierung Aufgaben durchführt, die vorher der Mensch gemacht hat. Wenn diese Aufgabe beispielsweise das Bewegen von Objekten ist, nennt man die Maschine aus historischen, aber nicht aus logischen Gründen Roboter.

2.2 Intelligenz, Künstliche Intelligenz (Gegenstandsbereich)

KI ist die Abkürzung von „Künstlicher Intelligenz“ und ist ein Teilgebiet der Informatik. Für das vorliegende Projekt unmittelbar relevant ist, dass Künstliche Intelligenz mit Software verbunden ist. KI-Systeme sind demnach Systeme, die Software aus dem Bereich KI enthalten. Im Safety-Kontext dieses Projektes seien als KI-Systeme physische Systeme verstanden, die Softwarekomponenten auf Basis von Algorithmen der KI enthalten. Das sind beispielsweise Steuerungen von Maschinen, fahrerlose Transportsysteme, Sicherheitssysteme, Systeme zur Qualitätsüberprüfung, etc.
Software auf Basis von KI-Algorithmen unterscheidet sich wie in der Einleitung geschildert von herkömmlicher Software bezüglich ihrer Fähigkeiten, aber auch bezüglich der Transparenz, Kontrollierbarkeit und Überprüfbarkeit teilweise drastisch. Um den dadurch erwachsenen besonderen Herausforderungen gerecht zu werden, gibt es zurzeit zahlreiche Bemühungen, in Normungsgremien die Besonderheiten von KI-Algorithmen zu fassen und möglicherweise Vorgehensrichtlinien zu erstellen. Ein großer Teil der in diesem Projekt interviewten Experten war oder ist in solchen Normungsgremien mit KI als Thema tätig. Von allen wurden die großen Schwierigkeiten und Kontroversen bei der Definition des Begriffs „Künstliche Intelligenz“ berichtet. Die EU-Kommission definiert beispielsweise Künstliche Intelligenz als „*Systeme mit einem „intelligenten“ Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen.“* [40]. Problematisch an der obigen Definition ist das Attribut „intelligent“. In den eben erwähnten Diskussionen in Normungsgruppen ist es gerade die Frage, was „intelligent“ ist, die zu Kontroversen führt. Um den Gegenstandsbereich für dieses Projekt abstecken zu können, soll versucht werden, diese Frage zu beantworten.

2.2.1 Intelligenz

Der Begriff „Intelligenz“ ist im 19 Jahrhundert entstanden und charakterisiert bestimmte psychische Fähigkeiten des Menschen, die eine differenzierende Bewertung zwischen Gruppen von Menschen oder zwischen Individuen zulassen. So wurde einer der ersten Intelligenztests, der Binet-Simon-Test[41] 1905 dazu benutzt, diejenigen Kinder in Klassenverbünden zu identifizieren, die einer Förderung bedürfen. Dieses Merkmal als Metrik ist auch heute noch ein bedeutendes Merkmal von Intelligenz: Es gibt Menschen mit viel Intelligenz und Menschen mit wenig. ‚Keine Intelligenz‘ kann man allerdings ebenso wenig haben wie ‚kein Gewicht‘. Damit die Aussage *„intelligentes Verhalten“,* auf den Menschen angewandt, irgendeinen Sinn ergibt, muss damit eine implizite Quantifizierung verbunden sein. Sinnvoll erscheint, dass analog zu Aussagen wie „großer Mensch“ oder „müder Mensch“ damit ein „Verhalten eines Menschen mit hohem Intelligenzwert“ gemeint sein muss. Analog dazu gibt es unintelligenes Verhalten von Menschen. Das quantitative Maß des Intelligenzquotienten gilt jeweils nur für Gruppen von Menschen und ist mit dem Wert 100 auf die jeweilige

[40] MITTEILUNG DER KOMMISSION AN DAS EUROPÄISCHE PARLAMENT (2018), Künstliche Intelligenz für Europa
[41] Rozencwajg, P. (2006). Quelques réflexions sur l'évaluation de l'intelligence générale: un retour à Binet? *Pratiques psychologiques*, *12*(3), 395-410.

Grundgesamtheit normiert. Die Intelligenz von Fünfjährigen ist also anders festgelegt, als die von Zwanzigjährigen.

Nach Rost[42] ist Intelligenz das „am besten erforschte Merkmal der Psychologie“ mit mehr als 100 Jahren Forschung. Intelligenz beschreibt sowohl in der Psychologie als auch im Alltagsverständnis ein Maß für „denken können“ also die Leistungsfähigkeit bei bestimmten geistigen bzw. mentalen Fähigkeiten, die heute mit Kognition, in der philosophischen Tradition mit Verstand bezeichnet wird. Auch Tieren wird teilweise Intelligenz im Sinne kognitiver Leistungsfähigkeit zugesprochen[43]. Nicht zur Intelligenz werden z. B. Persönlichkeitsmerkmale oder etwa auch die Aggressivität gezählt. Gleichsam nicht Kern der klassischen Vorstellung von Intelligenz sind ferner hochautomatisierte sensumotorische Fähigkeiten, die Intelligenz nur zu einem eher kleinen Teil erfordern. Ein Beispiel dafür ist die für das Fahrradfahren notwendige Intelligenz[44]. Menschliche Intelligenz kann man schwerlich am Fahrradfahren erkennen.

In einem Zeitraum von 120 Jahren Forschung ist eine Vielzahl von Intelligenzdefinitionen entstanden. Allerdings gibt es inzwischen eine „wissenschaftliche Mehrheitsmeinung“ von 52 Experten aus dem Jahr 1994:

> *„Intelligenz ist eine 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“*[45]

In dieser Definition sind alle „Zutaten“ in ihrer Bedeutung für das Konstrukt Intelligenz der Reihe nach aufgeführt. Zu Intelligenz gehört in erster Linie „abstrakt-logisches Denken“, „Problemlösefähigkeit“ und die „Kapazität, sich Wissen anzueignen“ in zweiter Linie „Anpassung an die Umwelt“[46] und „Gedächtnis“.

Nach Rost (2013) gehen *„praktisch alle psychometrisch arbeitenden Intelligenzforscher und alle neueren Intelligenzmodelle von einer hierarchisch organisierten Struktur der Intelligenz aus“*. In der psychologischen Forschung am meisten akzeptiert[47] ist die *hierarchische Theorie der drei Intelligenzschichten* von

---

[42] Rost, D. H. (2013). Handbuch Intelligenz. Weinheim: Beltz.

[43] Reznikova, Z. (2007). Animal intelligence: From individual to social cognition. Cambridge University Press.

[44] Auch Bären können Fahrradfahren

[45] Zitiert nach Rost (2013) aus: Gottfredson, L. S. (Ed.). (1997). Intelligence and social policy. Intelligence, 24 (1). Deutsch: Eysenck, H. J. (2004). Die IQ-Bibel. Intelligenz verstehen und messen. Stuttgart, DE: Klett-Cotta, S. 368–377.

[46] Damit ist nicht eine biologische Anpassung oder das Anpassen beim Bewegen in einer dynamischen Umwelt gemeint, sondern wiederum das Lösen von Problemen, die aus Anforderungen der Umwelt entspringen.

[47] Es gibt inzwischen Erweiterungen geringeren Ausmaßes wie das Cattell-Horn-Carroll Modell,

Carroll[48]. Die oberste Schicht wird dort durch einen einzigen Generalfaktor G[49] gebildet, so etwas wie eine „generelle Intelligenz“. Die zweite Schicht wird von acht General-Sekundärfaktoren gebildet, die unterschiedlich stark zu der allgemeinen Intelligenz beitragen. Hier finden sich Faktoren (in absteigender Wichtigkeit zur Erklärung von G) wie Fluide Intelligenz (schlussfolgerndes Denken, Induktion und Deduktion, Klassifikation und Begriffsbildung, etc.) kristalline Intelligenz (deklaratives Wissen, Alltagerfahrungen, etc.), aber auch Gedächtnisfähigkeit, auditive Wahrnehmung, kognitive Verarbeitungsgeschwindigkeit oder Reaktionsgeschwindigkeit, die alle in einer gewichteten Summe das Konstrukt Intelligenz determinieren. Auf der untersten Schicht befinden sich fast 70 Primärfaktoren[50], die wiederum gewichtet die Faktoren der zweiten Ebene determinieren.

Für die Überlegungen in Verbindung mit Künstlicher Intelligenz soll hervorgehoben werden, dass ein wesentliches Merkmal der obigen Intelligenzdefinition darin besteht, dass sich die Höhe der Intelligenz durch die Höhe der Werte in allen zugrundeliegenden Kriterien zusammen ergibt. Inselbegabte, sogenannte Savants, die z.B. über eine überragende Gedächtnisleistung verfügen, können teilweise nicht ohne Hilfe eigenverantwortlich leben und sind eher unterdurchschnittlich intelligent. Ein weiteres wesentliches Merkmal der obigen Intelligenzdefinition ist das Ausschließen von Fähigkeiten der Fertigkeiten, die durch langes Üben, Lernen oder Training entstanden sind. Überdurchschnittliche Fertigkeiten sind noch kein Indiz für hohe Intelligenz[51].

Die teilweise als „unfair“[52] aufgefasste Einengung der Intelligenzdefinition (eben den Fokus auf kognitionsbezogene Leistungsfähigkeit) hat neben der zitierten wissenschaftlichen Mehrheitsmeinung zu einer Reihe „alternativer“ Intelligenztheorien geführt. Bekannte Theorien sind „soziale“ Intelligenz[53], „multiple“ Intelligenzen[54],

---

[48] Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge, NY, USA: Cambridge University Press.

[49] Der entspricht dem von Spearman schon 1923 postulierten Generalfaktor g (Spearman, C.,1923. The nature of ‘intelligence’ and the principles of cognition. London, GB: Macmillan.). Der Generalfaktor sagt aus, dass der Wert eines einzigen Konstrukts die Leistungen in einem sehr breiten Bereich menschlichen Tuns erklären kann.

[50] Z. B. „Geschwindigkeit schlussfolgernden Denkens“ „Gedächtnisspanne“, „Lernfähigkeit“

[51] Eine gute Leistung im Schachspielen wird in der Regel nicht als gutes Indiz für Intelligenz gewertet (Rost, D. H. (2009). Mehr multiple Perspektiven–mehr multiple Irritationen? Replik auf die Kritik von Kim & Hoppe-Graff. Zeitschrift für Pädagogische Psychologie, 23(1), 75-83.). (Anmerkung: Interessanterweise lagen die ersten Ziele in der KI darin, Maschinen zu konstruieren, die eine „typische intelligente Leistung des Menschen, das Schachspielen“ meistern können. Inzwischen sind Maschinen in diesem Bereich ja auch besser als der Mensch.)

[52] Kim, H.-O. & Hoppe-Graff, S. (2009). Multiple Intelligenzen, multiple Perspektiven. Kommentar zu Rost, Multiple Intelligenzen, multiple Irritationen. Zeitschrift für Pädagogische Psychologe, 23, 65–74.

[53] Thorndike, E. L. (1920). Intelligence and its uses. Harper’s Magazine, 140, 227–235.

[54] Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. New York, NY, USA: Basic Books.

„emotionale“ Intelligenz[55] oder „praktische“ Intelligenz[56], die ihren Eingang auch in populärwissenschaftliche Veröffentlichungen gefunden haben. Rost (2013) spricht in diesem Zusammenhang von einer Inflation der Intelligenzen und beklagt eine fast uferlose Ausweitung des Intelligenzbegriffs, weit weg vom Grundkonsens der Intelligenzforschung. Das Definieren vieler besonderer Fähigkeiten als eine Art „Intelligenz“ führe zur Verwässerung des Intelligenzbegriffs [57]. Größter Kritikpunkt an der Einführung sehr unterschiedlicher Intelligenzarten ist aber ihre fehlende belastbare Operationalisierbarkeit – es gibt schlichtweg kein Instrument diese Intelligenzen (Fähigkeiten) trennscharf zu messen. Süß (2006, S. 8–11)[58] formulierte in diesem Zusammenhang einen Anforderungskatalog, den belastbare Intelligenzkonstrukte erfüllen sollten[59], und den die eben aufgezählten alternativen Intelligenzen eben nicht erfüllen. Ohne eine solche Operationalisierung haben diese Intelligenzkonstrukte allenfalls plakativen Wert.
Das Ziel der Intelligenztheorien ist es, ein Instrumentarium zur Unterscheidung von menschlicher mentaler Leistung – z. B. bei der Identifikation von Hochbegabten oder zur Erklärung von Unterschieden im beruflichen Erfolg – aufbauen zu können. Dies beinhaltet auch die Möglichkeit der operativen Testung und die damit verbundenen Forderungen nach empirischer Fundierung. Die Theorien können inzwischen als relativ gefestigt gelten. Man weiß also, wie (menschliche) Intelligenz am sinnvollsten zu charakterisieren ist. Die gemeine Auffassung von Intelligenz bei Nichtpsychologen ist zwar nicht so genau wie die professionelle, stimmt aber bis auf Verwässerung durch die oben genannten alternativen Intelligenztheorien insbesondere bei den wichtigsten Faktoren damit überein. Wir können uns beispielsweise gut vorstellen, was ein besonders intelligentes Kind ausmacht. Auch welches Verhalten bei Tieren als intelligent bezeichnet werden kann, ist gut vermittelbar.

2.2.2 Künstliche Intelligenz

Das Attribut „künstlich“ im Begriff „Künstliche Intelligenz“ verdeutlicht, dass hier „Intelligenz“ im Zusammenhang mit künstlich entwickelten Maschinen betrachtet wird. Viel zitiert ist die erste Verwendung des Begriffs von Intelligenz zusammen mit Fähigkeiten eines Computers in einem Essay des Mathematikers Turing von 1950[60]. Allerdings kommt hier der Begriff Intelligenz in Verbindung mit einem Computer im Wesentlichen nur im Titel seines Essays vor. Im Text stellt er nicht die Frage, ob

---

[55] Salovey, P. & Mayer, J. D. (1990). Emotional intelligence. Imagination, Cognition and Personality, 9, 185–211.
[56] Sternberg, R. J. & Wagner, R. K. (Eds.). (1986). Practical intelligence: Nature and origins of competence in the everyday
world. Cambridge, GB: Cambridge University Press.
[57] Rost, D. H. (2009). Mehr multiple Perspektiven–mehr multiple Irritationen? Replik auf die Kritik von Kim & Hoppe-Graff. Zeitschrift für Pädagogische Psychologie, 23(1), 75-83.)
[58] Süß, H.-M. (2006). Eine Intelligenz – viele Intelligenzen. Neue Intelligenztheorien im Widerstreit. In H. Wagner & Thomas-Morus-Akademie Bensheim (Hrsg.), Intellektuelle Hochbegabung. Aspekte der Diagnostik und Beratung. Tagungsbericht (S. 7–39). Bad Honnef, DE: Bock.
[59] Zitiert nach Rost (2013)
[60] A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.

Maschinen *intelligent* sein können, sondern ob Maschinen *denken* können, wobei dabei natürlich menschliches Denken gemeint ist. Auch seine Zeitgenossen, die seinen Aufsatz referierten, bezogen sich auf die von ihm gestellte Frage, ob Maschinen denken können[61]. Ziel Turings war also der Entwurf „Künstlichen Denkens". Eine detaillierte Beschreibung der Einzelheiten des sog. Turing-Tests wäre an dieser Stelle nicht zielführend; angemerkt werden soll lediglich, dass es sich dabei um ein Frage-Antwort-Spiel mit nichtphysischer Sprachinteraktion handelt, das menschliche Bewerter als Evaluatoren der „Menschähnlichkeit" einer Maschine bezüglich bestimmter, durch Fragen erfassbare psychische Fähigkeiten nutzt. Obwohl dies in der KI-Literatur häufig behauptet[62] wird, ist der Test von Turing kein Intelligenztest im obigen psychologischen Sinne: es wird nicht getestet, *wie* intelligent die Maschine ist, sondern *ob* sie intelligent ist, wobei hier unter intelligent „denkt wie ein Mensch" verstanden wird – unabhängig davon, welches Intelligenzniveau diesem „Vergleichsmensch" zugesprochen wird[63]. Die Gleichsetzung von „intelligent" mit „ähnlich zu menschlichem Denken und menschlicher Informationsverarbeitung" zieht sich durch einen Großteil der KI-Literatur.

Der Begriff „Künstliche Intelligenz" entstand im Zuge der Beantragung von Fördermitteln für ein zweimonatiges Sommerprojekt in Dartmouth College in Hanover, New Hampshire[64]. Die Beantragenden diskutierten, wie das Ziel erreicht werden könnte, spezielle, bis dato nur vom Menschen leistbare Fähigkeiten durch Maschinen zu simulieren und diese schließlich sogar besser werden zu lassen als der Mensch. Als Beispiele für Ziele, die erreicht werden sollten, wurden u.a. „Beweisen von Theoremen", „Komponieren" oder „Schach spielen" genannt. Es ging darum „intelligentes" Verhalten nachzubilden, wobei der Begriff „Intelligenz" analog zu Turing offensichtlich wiederum als Synonym menschlicher kognitiver Leistungen gebraucht wurde.

Die Begriffsbildung „Künstliche Intelligenz" war ein Stück weit auch zufälliger Natur. McCarthy, einer der Initiatoren des Projektes, wählte den Begriff Künstliche Intelligenz als Alternative zu „Automatentheorie", weil er zur selben Zeit mit Shannon einen Sammelband mit diesem Namen herausgab[65]. Einige Teilnehmer der Konferenz waren mit dem Begriff nicht einverstanden und benutzen ihn wie Simon und Newell zunächst nicht. Stattdessen wählten sie lieber den für die Arbeiten der ersten Jahre passenderen Begriff „complex information processing". Bei der Konferenz war auch der Begriff „berechenbare Rationalität" im Gespräch, setze sich aber auch nicht

---

[61] Shannon, C., and McCarthy, J., eds. 1956. Automata Studies. Annals of Mathematical Studies, 34. Vorwort, Princeton, N.J.: Princeton University Press.

[62] Der Ansatz von Turing wird noch heute als „ausreichend Operationalisierung von Intelligenz bei der Adaption auf Maschinen angesehen. Norwig, P., & Russel, S. (2010). Artificial Intelligence. A modern approach. Willliams.

[63] In dem Turing-Test muss die Maschine im Spiel durchaus enorme Fähigkeiten haben, die auch für einen Menschen herausfordernd sind. Als Intelligenztest taugt der Test wegen der fehlenden Quantifizierbarkeit dennoch nicht.

[64] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine, 27(4), 12-12.

[65] Aus: McCorduck, P. 2004. Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence. Natick, MA: A. K. Peters, Ltd., 2nd edition.

durch[66]. Einmal in die Welt gesetzt, blieb die Begrifflichkeit „Künstliche Intelligenz“ an dem Wissenschaftsgebiet haften.

Die Verbindung zum Intelligenzbegriff in der Psychologie ist folglich als lose zu bezeichnen. Der Intelligenzbegriff in der KI unterscheidet sich fundmental von dem psychologischen. Marvin Minsky formuliert im Jahr 1966:

> „*Artificial Intelligence is the science of making machines do things that would require intelligence if done by men*.[67]“ wobei er als Intelligenz “*the ability to solve hard problems*”[68] annahm.

Ähnlich ist die Definition McCarthys und Hayes (1969)[69], die dabei auch beschreiben, was „intelligent“ im Zusammenhang mit Maschinen bedeutet.

> „*a machine is intelligent if it solves certain classes of problems requiring intelligence in humans, or survives in an intellectually demanding environment*“.

Wichtig ist der zweite Teil der Definition, der sehr Vielem – auch Maschine oder Tier – das Potenzial zum Attribut „intelligent“ zusprechen. Dieser Bezug zum Agieren in der Umwelt ist auch wesentlicher Teil der Definition von Nilsson 2009[70]:

> „*For me artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment“.*

Der Bezug der Maschinenintelligenz auf das Für-den-Menschen-Schwierige geht ins Leere, wie das Beispiel des wissenschaftlichen Taschenrechners zeigt, der sicherlich nicht als intelligent bezeichnet wird, aber Aufgaben lösen kann, für die es bei menschlichen Akteuren zumindest gehobener Intelligenz bedarf. Das zeigt sich auch beim Schachcomputer, der heute nicht mehr als intelligent bezeichnet wird.[71] Diese Ernüchterung gegenüber KI-Technologien bezüglich der Zuschreibung von Intelligenz nachdem sie sich etabliert haben, wird als KI-Effekt bezeichnet: „KI ist das, was noch nicht gemacht wurde“ [72].

Problematisch an der Zuschreibung einer Intelligenz an dem Zurechtkommen in der Umwelt ist, dass – auch bei Berücksichtigung, dass dies Voraussicht enthält – damit praktisch alle Tiere als intelligent bezeichnet werden müssen.

In 80er Jahren wird der Begriff „intelligent“ zunehmend in Zusammenhang mit dem Begriff „Agent“ verwendet[73]. In dem Lehrbuch „Artificial Intelligence. A modern

---

[66] Aus Norwig, P., & Russel, S. (2010). Artificial Intelligence. A modern approach. Willliams

[67] M Minsky (ed) (1968) Semantic Information Processing, The MIT Press, Cambridge, Mass.

[68] Minsky, M. 1985. The Society of Mind. New York: Simon and Schuster.

69 McCarthy, J.; Hayes, P.J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer und D. Michie, Hrsg., Machine Intelligence 4, Seiten 463–502. Edinburgh University Press, 1969.

[70] Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press.

[71] Wie erwähnt wird auch menschliches Schachspielen nicht als Ausdruck menschlicher Intelligenz angesehen.

[72] Schank, R. C. 1991. Where is the AI? AI Magazine 12(4):38–49..

[73] Der Begriff intelligenter Agent wurde schon viel früher verwendet, zunächst als „General Intelligent Agent“ im „general problem solver“ GPS. Newell, A. & Simon, H.A. (1972). Human Problem Solving und später für spezialisierte Expertensysteme zur Unterstützung von Wissenschaftlern oder Ärzten, z. B. DENDRAL Expertensystem für Massenspektrometer. Feigenbaum, E.A. "Artificial Intelligence Research: What is it? What has it achieved? Where is it going? " invited paper, Symposium on Artificial Intelligence, Canberra, Australia. 1974.

approach“ von Norwig und Russel (2010) werden intelligente Agenten ganz allgemein als etwas eingeführt, das über Sensoren Daten aus einer Umwelt empfangen und auf dieser Basis mit Hilfe von Aktuatoren in dieser Welt handeln kann. Auf oberster von vier Intelligenzstufe stehen Agenten, die Ziele verfolgen, auf Basis von Modellen und Gütefunktionalen (Nützlichkeit) operieren und lernen. Allerdings werden auch einfache reflexartig Strukturen als „intelligente Agenten“ bezeichnet. Daraus folgert dann, dass auch Temperaturregler oder Geißeltierchen als „intelligent“ – wenn auch auf der untersten Stufe stehend – bezeichnet werden müssen.

2.2.3 „Intelligent“ als Attribut von Software und Maschinen

Das vorherige Kapitel hat gezeigt, dass in der Fachrichtung „Künstliche Intelligenz“ eine Vorstellung von Intelligenz entstanden ist, die sich fundamental von der der Psychologie unterscheidet. Beschreibt der psychologische Intelligenzbegriff Unterschiede zwischen Menschen bezüglich eines (mehr oder minder gut) messbaren Merkmals, ist es in der KI einerseits der Vergleich von Maschinen mit dem Menschen und andererseits die Annahme bestimmter Informationsverarbeitungsprozesse, die die Vergabe des weitgehend „binären“[74] Labels „intelligent“ rechtfertigt. Menschliches Handeln und Denken ist in der KI per se intelligent und eine Maschine, die intelligent genannt werden will, muss psychische Fähigkeiten oder Fertigkeiten haben, die der Mensch hat, wobei die Fähigkeiten nicht eingegrenzt sind. In letzter Konsequenz ist es die Fähigkeit des Menschen, in einer Umwelt handeln zu können. Dieses „in einer Umwelt auf Basis von Sensordaten adäquat handeln“ ist das Wesen von den oben beschriebenen „intelligenten Agenten“, was aber in letzter Konsequenz dazu führt, dass auch Temperaturreglern, Lichtschranken oder Fröschen das Attribut „intelligent“ verliehen werden muss. Da jedes Softwareprogramm ein Ergebnis produziert und jede Maschine irgendetwas macht, ist es – neben der Menschähnlichkeit – in letzter Konsequenz die Tatsache eines Inputs über Sensoren, die Intelligenz festlegt. Der Input kann sogar reiner Datennatur sein, wie etwa bei intelligenten Softwareagenten[75]. Der Begriff „intelligent“ ist in der KI-Literatur allgegenwärtig und wird nicht sichtbar in Frage gestellt. Zunehmend problematisch ist die Frage, was *nicht* intelligent ist. Wird das Verhalten eines Temperaturreglers als intelligent akzeptiert, müssten eigentlich alle Systeme der Automatisierungstechnik als „intelligent“ bezeichnet werden. Weiterhin ist es problematisch, unterschiedliche intelligente Systeme bezüglich ihrer (synthetischen) Intelligenz zu vergleichen. Was ist intelligenter, ein Schachcomputer oder ein autonom fahrender Roboter, und wie verhält es sich bei beiden im Vergleich etwa zu einem intelligenten Softwareagenten? Oder sind die Fragen falsch? Bedeutet vielleicht in der KI das Label „intelligent“ nichts anderes als „enthält KI-Algorithmen“?

---

[74] Eben den eben beschriebenen Stufen von Intelligenz, die sich an dem Ausmaß von Planungsoperationen verankert, wird teilweise die Stufigkeit von „Intelligenz“ auch an Fortschritten in der technischen Entwicklung festmacht. Je moderner die Entwicklung, desto intelligenter ist das System (siehe z.B. Wahlster, W., Winterhalter, Ch. (eds.) (2020): Deutsche Normungsroadmap Künstliche Intelligenz. S.12)

[75] Buettner, R. (2011). Automatisiertes Headhunting im Web 2.0: Zum Einsatz intelligenter Softwareagenten als Headhunting-Robots. INFORMATIK 2011 - Informatik schafft Communities, 41. Jahrestagung der Gesellschaft für Informatik, 4.-7.10.2011, Berlin.

Problematisch an der divergierenden Auffassung von Intelligenz bei Menschen und bei Maschinen ist auch die Wertung bei gleichen Aufgaben. Ein selbstfahrender Mähroboter mit KI-Algorithmen hat eindeutig die Merkmale eines intelligenten Agenten, u. U. sogar höherer Stufe, und wird auch als „intelligent“ vermarktet. Ein Tier (z.B. ein Schaf), das dieselbe Tätigkeit leistet, wird aber in der Regel ob dieser Leistung keineswegs als intelligent bezeichnet. Und auch bezogen auf Menschen ist das Mähen kein Indikator von Intelligenz, bzw. nicht einmal in das Maßkonstrukt einzuordnen.
Dieser neue (synthetische) Intelligenzbegriff und die Diskussion in der Öffentlichkeit haben dazu geführt, dass „intelligent“ zunehmend auch als Marketingbegriff benutzt wird. Veröffentlich sind:

- [76] *intelligente Stromnetzkonzepte* (Smart Grid),
- *intelligentes Energiemanagement*,
- *intelligente Verpackungen*,
- *intelligente Arbeitsplätze*,
- *intelligentes Verkehrsmanagement*,
- *intelligente Sensorik* oder
- *intelligentes, multinutzerfähiges Zuhause*.

Man findet

- *intelligente Fabriken*[77],
- *intelligente Produkte*[78] die über das Wissen ihres Herstellungsprozesses und künftigen Einsatzes verfügen,
- *intelligente Lieferketten*[79]*,*
- *intelligente Lichtschranken*[80],
- *intelligente Haushaltsgeräte*[81],
- *intelligente Bohrmaschinen*[82] oder
- *intelligente Rasenmäher*[83].

In vielen Beispielen ist es allein die Tatsache, dass die Systeme oder Produkte Komponenten mit KI-Algorithmen enthalten, die das Attribut „intelligent“ rechtfertigen.

---

[76] Alle folgenden Beispiele aus: Fachforum Autonome Systeme, acatech (Hrsg) (2016) Das Fachforum Autonome Systeme im Hightech-Forum der Bundesregierung–Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft. Abschlussbericht. München

[77] Kagermann, H., Wahlster, W., & Helbig, J. (2013). Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. *Abschlussbericht des Arbeitskreises Industrie*, *4*(5), 1-9.

[78] Kasper, B. Materialsammlung zu den Arbeitspaketen 1 und 2 im BAuA-Projekt F 2474 „Sicherheitstechnische Methoden Industrie 4.0 (SMI4. 0)“

[79] Bitkom, D. (2017). Entscheidungsunterstützung mit Künstlicher Intelligenz: Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung. [online]

[80] „Intelligente Lichtschranken“ werden durchaus schon beworben. https://www.md-automation.de/themen/automation/mit-intelligenten-lichtschranken-sicher-detektieren-und-effizient-automatis

[81] https://www.lupus-electronics.de/de/blog/smartes-leben-intelligente-haushaltsgeraete/

[82] https://www.ke-next.de/news/intelligente-bohrmaschine-mit-igus-gleitlagern-richtet-knochen-122.html

[83] Multiple Referenzen

Zunehmend wird „intelligent“ aber nur als Synonym für „hochmodern“ oder „fortschrittlich“ mit entsprechender Marketingwirkung verwendet. Dies wird besonders deutlich in Beispielen wie:

- *intelligente Kieze*[84]*,*
- *intelligente Brücken*[85] *oder*
- *intelligente Steine*[86]*.*

Besonders bei den intelligenten Steinen, die nicht einmal elektronische Komponenten oder Software enthalten, wird die Verselbstständigung des Begriffes intelligent außerhalb des Kontextes Mensch deutlich. Intelligent wird zunehmend zu einem Synonym für „fortschrittlich, an vorderster Front technologischer Entwicklung stehend“ – auch dann, wenn kein KI-Algorithmus darin enthalten ist.

2.2.4 Konsequenzen für das Projekt

Die Verwendung des Begriffes „intelligent“ in Verbindung mit Algorithmen oder Maschinen ist in der KI ausreichend definiert und abgegrenzt, um im Wissenschaftsgebiet widerspruchsfrei verwendet werden zu können und als Charakterisierung von Systemeigenschaften dienen zu können. In diesem Sinne ist auch die oben zitierte Definition der EU-Kommission „*Systeme mit einem „intelligenten“ Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen.“* zu verstehen. Die Definition greift im Grunde auf, was zurzeit in der KI erforscht oder entwickelt wird. Die Begrifflichkeit „Intelligentes Verhalten“ in der Definition gibt keine zusätzlichen Informationen, sondern nimmt die Beschreibung im Nachsatz auf. Im Grunde hätte man auch definieren können „*Künstliche Intelligenz (KI) bezeichnet Systeme mit einem „intelligenten“ Verhalten. Intelligent handelnde System analysieren ihre Umgebung und handeln mit einem gewissen Grad an Autonomie*“
Mag der Begriff „intelligent“ innerhalb der KI auch ausreichend bekannt sein, so muss aber wegen des multidisziplinären Ansatzes des vorliegenden Projektes gefragt werden, ob sich angesichts des fundamentalen Unterschieds zu dem psychologischen Intelligenzbegriffs nicht möglicherweise Missverständnisse auftun, die eine Fehlinterpretation der Fähigkeiten von als „intelligent“ bezeichneten KI-Algorithmen zur Folge haben. Obwohl auch der psychologische Intelligenzbegriff nicht jedem Protagonisten aus dem Bereich Safety und jedem Juristen bekannt sein dürfte, kann man vermuten, dass eher die Anschauung aus der eigenen oder berichteten Erfahrung mit Intelligenztests gegenüber der Kenntnis der KI-Literatur überwiegt. Bei der Propagierung des Intelligenzbegriffs der KI besteht die Gefahr des Überschätzens der Fähigkeiten von KI-Algorithmen. Die KI-Spezialistin Katharina Morik (eine im Rahmen

---

[84] https://www.inforadio.de/programm/schema/sendungen/wissenswerte/202009/13/gespraech-wissenschaft-staedteplanung-quartier-vernetzung-technik-berlin.html, abgerufen 20.1.2021

[85] Fischer, J., Schneider, R., Thöns, S., Rücker, W., & Straub, D. (2014). Intelligente Brücke-Zuverlässigkeitsbasierte Bewertung von Brückenbauwerken unter Berücksichtigung von Inspektions- und Überwachungsergebnissen.

[86] https://www.topagrar.com/energie/news/ziegel-sind-energiespeicher-12135078.html, abgerufen 20.1.2021

des Projektes interviewte Expertin), warnt in diesem Zusammenhang von einer Überschätzung: *„Menschen gehen davon aus, wenn der [Computer] z. B. Schach spielen kann, dann beherrscht er auch alles, was Schachspieler sonst noch können. Die Vorstellung von dem, was ein Rechner kann, ist zu sehr geprägt von der Analogie zum Menschen“.*

Die Ausführungen über die in der Psychologie untersuchte Intelligenz haben gezeigt, dass eine Übertragung auf Maschinen wegen der fehlenden Möglichkeiten der Operationalisierung nicht möglich ist. Intelligenz als ein ganzheitliches Maß ist sehr auf den Menschen zugeschnitten. Menschen sind zwar auch unterschiedlich; diese Unterschiede sind aber sehr viel geringer als Unterschiede zu Maschinen. Selbst eine Bewertung von Teilen von Menschen, wie beispielsweise das neuronale Netze des primären visuellen Cortex, ist bezüglch des Konstruktes Intelligenz nicht möglich. Und auch die Übertragung auf Tiere gelingt nur in Ansätzen, weil adaptierte Tests durchführbar sind. Auch hier ist allerdings ein quantitativer Vergeich zu der Intelligenz von Menschen nicht möglich.

Das Konstrukt Intelligenz ist so angelegt, dass man die Intelligenz bei Maschinen nur dadurch testen könnte, dass man sie mit den beim Menschen angewandten Intelligenztests abfragt. In letzter Konsequenz würde das nur gehen, wenn man einen Menschen einszueins nachbaut, eine Vision, die als starke künstliche Intelligenz bekannt ist. Aber auch hier wäre es wahrscheinlich besser, eine spezielle Operationalisierung von Intelligenz zu entwickeln, mit der verschiedene Ausprägungen dieser – sicherlich nicht vollständig gleichen – „künstlichen Menschen“ verglichen werden könnten. Auch dann hätten wir es aber wieder mit einer *anderen* Intelligenz zu tun.

Summarisch bleibt festzuhalten, dass der in der KI verwendete Begriff von „Intelligenz“ anders belegt ist, als der in der Psychologie. Differierende Bedeutungsbelegung von Begriffen in unterschiedlichen Disziplinen und Sprachkontexten ist allgegenwärtig. Für das vorliegende Projekt ist bedeutsam, dass bei einem Disziplinwechsel nicht fehlerhafte Schlussfolgerungen wegen Unkenntnis über die wahre disziplinfremde Bedeutung entstehen.

2.3 Festlegung der Begriffe Künstliche Intelligenz (Fachrichtung), KI, KI-System

Künstliche Intelligenz, KI: Neben der Beschreibung eines Gegenstandsbereichs oder eines Systems wie in der obigen EU-Definition wird der Begriff „Künstliche Intelligenz“ auch für ein Teilgebiet der Informatik verwendet, welches allerdings als durchaus 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. So werden beispielsweise vormals der Mathematik zugeordnete Themenfelder inzwischen teilweise auch unter „Künstlicher Intelligenz“ subsummiert. Aufgrund der dargelegten Schwierigkeit, die der

Intelligenzbegriff mit sich bringt (in Form einer möglichen Überinterpretation des Intelligenzbegriffs), werden in vorliegendem Bericht sowohl bei der Taxonomie als auch im Rechtsgutachten die Termini „intelligent“ sowie „Künstliche Intelligenz“ vermieden. Für letzteres soll stattdessen synonym der Begriff „KI“ Anwendung finden, der definiert sei als:

Definition: **KI**

Ein heterogenes Teilgebiet der Informatik, das aus dem ursprünglichen Ziel, menschliches Denken nachzubilden, entstanden ist.

Ein KI-System wird im Bericht „*A definition of Artificial Intelligence: main capabilities and scientific disciplines*“ der EU[87] als:

Ein KI-System ist jede KI-basierte Komponente.

definiert.
Wir schließen uns im Wesentlichen dieser Auffassung an und definieren auch in Hinblick auf den dauernden Wandel in der KI:

Definition: **KI-System**

Ein KI-System ist ein System ist, das Komponenten enthält, die der KI zuzuordnen sind.

2.4 „selbst“, „selbstständig“

Wenn in der Öffentlichkeit von Risiken, aber auch Gefahren der künstlichen Intelligenz gesprochen wird, wird häufig thematisiert, dass KI-Systeme „selbst“ nach eigenem Ermessen entscheiden[88] oder „selbstständig“ effizient Probleme lösen[89]. Es gibt *selbst*lernende künstliche neuronale Netze. Man findet Sätze wie:

- „*Dass Computer und elektronische Systeme selbstständig Entscheidungen treffen, ist mittlerweile eine ihrer größten Funktionen*“ (Computerweekly, Sep. 2019)[90]
- „*Autonome Systeme können selbstständig Entscheidungen treffen, auch auf der Grundlage eigener Lernprozesse*“.

[87] A definition of Artificial Intelligence: main capabilities and scientific disciplines, 8.4.2019 https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

[88] Z.B. Christen, M., Mader, C., Čas, J., Abou-Chadi, T., Bernstein, A., Binder, N. B., ... & Thouvenin, F. (2020). *Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz* (Vol. 72). vdf Hochschulverlag AG.

[89] Mainzer, K. (2016). Künstliche Intelligenz – wann übernehmen die Maschinen? Technik im Fokus. Wien: Springer, S. 3.

[90] https://www.computerweekly.com/de/feature/Die-Gratwanderung-zwischen-kuenstlicher-Intelligenz-und-Ethik Zugriff am 20.3.2020

- „*Dank intelligenter und selbstständiger Module sowie standardisierter Schnittstellen wird die Produktion in Echtzeit angepasst“* [91].
- Beim Entwurf *„Selbststeuerung in Produktion und Logistik*“ wird davon gesprochen, dass sich „*Autorückleuchten selbstständig durch die Montage steuern*“ [92].

Die Wörter „selbst“ und „selbstständig“ suggerieren, dass diese Systeme ähnlich wie ein Mensch mit „freiem Willen“ operieren. Häufig wird diese Suggestion auch eingesetzt, um die besondere Mächtigkeit der Systeme zu propagieren oder im Gegenzug Ängste von unkontrollierbaren Handlungen oder einer „ungewollten“ Übernahme zu schüren.
Der Begriff „selbst“ ist sehr stark mit psychologischen und philosophischen Begriffen verbunden. Beim Menschen gibt es das „Selbst“ als Zentrum der Persönlichkeit. Es gibt in der Psychologie zahlreiche Wortverbindungen mit „selbst“ wie Selbstvertrauen, Selbstachtung, Selbstsicherheit. Allerdings wird das Attribut „selbst“ auch bei Bezügen verwendet, bei denen sich „selbst“ nicht auf den Menschen bezieht. Es gibt selbstreinigende Fassadenfarbe, einen Selbstzünder, Selbstbestäubung, selbstschließende Tore. Es gibt selbstbezügliche Aussagen, selbstheilende Kondensatoren oder selbstheilende Schichten [93], die Selbstheilungskräfte der Natur und natürlich selbstfahrende Fahrzeuge.
Definition: **selbst**

> „selbst“ bezieht sich auf etwas, das normalerweise von etwas (jemand) anderem durchgeführt oder gesteuert wird

Wie beim Begriff „Autonomie“ wird auch bei „selbst“ immer die Benennung des „normalen“ Alternativvorgangs benötigt. Gibt es dieses „Normale“ nicht bzw. ist nicht klar, was es sein kann, ist der der Begriff „selbst“ sinnlos. Wenn also von Maschinen gesprochen wird, die „selbst“ Entscheidungen treffen, muss man fragen, was denn der „Normalvorgang“ ist. Wenn die Entscheidungen vorher der Mensch getroffen hat, dann ist die Übernahme durch eine Maschine ein klassischer Vorgang der Automatisierung. Eine „selbstentscheidende“ Maschine ist also eine Maschine, bei der Entscheidungen automatisiert sind. Aber was sind „Entscheidungen“? Ist es schon eine selbstständige Entscheidung, wenn eine Lichtschranke „eine Aktion auslöst, sobald der Lichtstrahl unterbrochen wird“ (Computerweekly, Sep. 2019) [94]. Wenn dem so wäre, gibt es fast kein Automatisierungssystem, das keine „selbstständigen“ Entscheidungen trifft.
In der Regel ist es aber nicht der ersetzte Mensch, der das oben genannte „Normal“ zur Rechtfertigung der Verwendung von „selbst“ oder „selbstständig“ darstellt, sondern der Übergang von einer Technologie zu einer anderen. In dem obigen Beispiel der Autorückleuchten, die sich „selbstständig“ durch die Montage steuern, ist das „Normale“ ein zentrales Steuerprogramm, das jetzt von dezentralen Einheiten ersetzt wird. Selbstverständlich sind die dezentralen Einheiten so programmiert, dass sie fehlerfrei und sicher funktionieren.

---

[91] Autonome Systeme, F. A. acatech (Hrsg) (2016) Das Fachforum Autonome Systeme im Hightech-Forum der Bundesregierung–Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft. Zwischenbericht. München.
[92] https://youtu.be/evazDOmVKq0
[93] https://www.weltderphysik.de/gebiet/materie/news/2018/selbstheilend-wie-haut-hart-wie-zahnschmelz/
[94] https://www.computerweekly.com/de/feature/Die-Gratwanderung-zwischen-kuenstlicher-Intelligenz-und-Ethik Zugriff am 20.3.2020

Wie bei der Diskussion des Intelligenzbegriffes schon dargestellt, besteht die große Gefahr, dass eine freie Benutzung auch im Alltag besetzter Begriffe zu Missverständnissen – in beide Richtungen – führen kann. In Kapitel 6 wird über die Gefahren der „Vermenschlichung“ von Maschinen im Zusammenhang mit dem Rechtsbegriff ePerson referiert. Die Verwendung der Begriffe „selbst“ und „selbstständig“ in diesem Bericht sind so ausgelegt, dass damit keine Vermenschlichung bis hin zu einem „freien Willen“ mitgedacht ist.

3 Darstellung der Ergebnisse der Experten-befragungen

Wie in der Einleitung erläutert, war Ziel des Projekts die Erstellung eines Rechtsgutachtens zur Überprüfung der Notwendigkeit einer Erweiterung des Rechtsrahmens zur sicheren Einführung gegenwärtiger und zukünftiger KI-Entwicklungen in der Industrie. Die Voraussetzung dafür war eine Beschreibung der Merkmale, durch die aktuelle und in Entwicklung befindliche KI-Systeme gekennzeichnet sind unter besonderer Berücksichtigung der Aspekte, die sich diese auf den sicheren Einsatz dieser KI-Systeme auswirken. Zudem sollten die sicherheitsbezogenen Eigenschaften systematisiert werden, so dass sie eine sinnhafte Struktur in Form einer speziell entwickelten Taxonomie als strukturierendes Gerüst für das geplante Rechtsgutachten bieten.

3.1 Vorgehen

Aufgrund der Zukunftsorientierung der Fragestellung bezüglich derzeit noch nicht eingesetzter KI-Systeme wurde eine Befragung von Experten durchgeführt, die in aktuelle Entwicklungen im Bereich KI, Robotik und insbesondere an der Schnittstelle von KI und und funktionaler Sicherheit im industriellen Bereich involviert sind. Zusätzlich wurden Experten eingebunden, die sich mit der vergleichbaren Fragestellung der sicheren Einführung von KI-Systemen im Automobilbereich, der aufgrund seiner Entwicklungen von autonomen Fahrfunktionen als besonders progressiv und damit zukunftsweisend gilt, beschäftigen. Aufgrund der Breite des Themenfelds und der unterschiedlichen Expertiseschwerpunkte der Befragten wurde von den anfänglich geplanten strukturierten Befragungsansätzen Abstand genommen. Stattdessen wurden halbstrukturierte Interviews geführt mit dem Ziel, eine Übersicht über die derzeitig implementierten sowie in Entwicklung befindlichen Anwendungen von KI-Systemen im industriellen Bereich, von denen mögliche Gefährdungen ausgehen, aufzuzeigen, sowie die Faktoren zu identifizieren, die Einfluss auf die Sicherheit eines KI-Systems haben. Die Leitfragen, die den Kern der halbstrukturierten Interviews bildeten, lauteten:

1) Was versteht man unter KI?
2) Worin unterscheiden sich KI-Systeme von konventionellen Systemen?
3) Welche Herausforderungen birgt KI für die Gewährleistung der Sicherheit?
4) Welche Entwicklungen liegen vor uns?

Dabei entschied man sich bewusst für ein generisches, d. h. nicht Use-Case-basiertes Vorgehen, um das auf der Systematisierung der Befragungsergebnisse aufbauende Rechtsgutachten ebenfalls generisch für das breite Feld der industriellen Anwendungen entwerfen zu können.

Für die Befragungen wurden 33 Teilnehmer mit Expertise in vorwiegend industriellen Anwendungsfeldern mit KI-basierten Anwendungen ausgewählt. Davon wiesen 13 Teilnehmer Expertise im Bereich funktionaler Sicherheit auf und 20 Teilnehmer Expertise bei Forschung und Entwicklung von KI-Anwendungen. Kenntnisse im jeweils anderen Expertisefeld (KI bzw. funktionale Sicherheit) waren bei den Teilnehmern unterschiedlich stark ausgeprägt. Der Anteil der Experten aus der funktionalen Sicherheit an der Gesamtheit der Experten je Anwendungsfeld verteilte sich wie folgt:

- Maschinenbau/Automatisierungstechnik (8/10)
- Automotive (2/7)
- Forschung zu KI (0/4)
- Eingebettete/vernetzte Systeme (0/4)
- Robotik (0/3)
- Normung (2/2)
- Smart Home (0/2)
- Anlagensicherheit (1/1)

Die Expertenbefragungen erstreckten sich über einen Zeitraum von 1,5 Jahren. Zwischen Februar 2019 und September 2020 wurden die Experten in die anfangs Face-to-Face, später virtuell durchgeführten Befragungen eingebunden. Die insgesamt ca. 44 Interviewstunden wurden aufgezeichnet, transkribiert und anschließend systematisiert. Zentrales Ergebnis der Interviews stellt die Erarbeitung einer Taxonomie sicherheitsbezogener Faktoren dar, die in Kapitel 3 (S. 35 ff.) erörtert wird. Vorab erfolgte eine Zusammenfassung der wichtigsten Ergebnisse der Experteninterviews. Die Schwerpunkte der nachfolgend dargestellten Ergebnisübersicht wurden mit Blick auf die Sicherheitsproblematik bei Einsatz von KI-Systemen gesetzt. Einzelne Zitate aus der Gesamtheit der durchgeführten Interviews wurden dann herausgestellt, wenn sie einen vorherrschenden Konsens innerhalb der KI-Experten bzw. der Experten aus der funktionalen Sicherheit widerspiegeln. Jedes Zitat steht damit prototypisch für eine Vielzahl inhaltlich vergleichbarer Aussagen. Ergänzungen seitens der Autoren des vorliegenden Berichts fließen nicht in die Darstellung ein.

3.2 Ergebnisse[95]

3.2.1 Was versteht man unter KI?

KI wird als sehr breit gefasster Begriff beschrieben, bei dem sich trefflich darüber streiten lässt, was er umfasst. Die vorliegenden Definitionen gelten entweder als umstritten oder als zu allgemein, um ein weithin geteiltes Begriffsverständnis herzustellen. Das ist jedoch nicht der einzige Grund, aus dem der Begriff KI bei vielen Befragten auf teils deutliche Ablehnung stößt. Diese rührt auch daher, dass der Begriff KI eine Projektion menschlicher Eigenschaften auf technische Systeme auslöst. Diese

[95] Kursive Aussagen sind Zitate

wird mutmaßlich durch die dem KI-Begriff inhärente Metapher der Intelligenz getriggert und verführt dazu, falsche Parallelen zwischen menschlichem Entscheiden und Verhalten und systemseitigen Ausgaben zu ziehen[96]. Die Vorstellung von dem, was KI kann, so heißt es, ist geprägt von der Analogie zum Menschen. Diese Vorstellung findet sich vor allem bei denjenigen, die von außen auf das System schauen, und legt sich mit zunehmendem Verständnis über die Funktionsweise der Systeme.

> ***„Die KI, wie wir jetzt kennen, die hat ja kein Bewusstsein, die trifft ja keine Entscheidungen, das ist ja vorprogrammiert."***
>
> **[Zitat aus dem KI-Expertenkreis]**

Ein weiterer Grund für die Ablehnung des KI-Begriffs geht darauf zurück, dass der Begriff einem zeitlichen Wandel unterliegt. Das, was einstig als KI bezeichnet wurde (beispielsweise symbolische KI, Expertensysteme), wird derzeit nur von wenigen noch als KI angesehen. Formuliert man diesen Sachverhalt etwas pointierter, lässt sich sagen, dass das, was noch nicht geht, KI ist, und sobald es die gewünschte Leistung zeigt, den Status der KI verliert und nur noch als Algorithmus gilt[97]. Demzufolge hätte man noch nie KI implementiert und würde es auch nie. Auch mit diesem Begriffsverständnis verliert die Begrifflichkeit KI die Beschreibungskraft für das aktuell Gegebene, wenn er nur auf das zukünftig Erreichbare anwendet wird.

> ***„Ich finde den Begriff KI ungeschickt, weil er beschreibt, was Computer noch nicht können aber demnächst können werden. In 20 Jahren wird man etwas anderes darunter verstehen als heute."***
>
> **[Zitat aus dem KI-Expertenkreis]**

Anstelle von KI wird daher bevorzugt die technische Begrifflichkeit des Maschinellen Lernens (ML) verwendet, der heute vielfach mit KI gleichgesetzt wird. ML beschreibt den „*Prozess, aus der Umwelt heraus Daten zu extrahieren, die man final nutzt, um Algorithmen zu implementieren, die technische Systeme handlungsfähig machen*", eine „*hochdynamische nichtlineare Abbildfunktion, die nicht intelligent ist*", einen „*Optimierungsprozess mit hoher Rechenleistung*". Als der KI zugehörige Verfahrensgruppen werden exemplarisch Algorithmen aus der Robotik, Entscheidungsbäume, agentenbasierte Systeme, Bayessche Netze, Reinforcement Learning sowie (künstliche) neuronale Netze angeführt. Eine klare Einordnung von ML-Verfahren als KI oder Nicht-KI fällt jedoch auch denjenigen schwer, die sich mit dem Gegenstand KI intensiv beschäftigen. Unstrittig erscheint jedoch die Einordnung neuronaler Netze als KI.

> ***„Ich fühle mich überfordert, genau immer zuzuordnen, welches ML-Verfahren zur KI gehört."***
>
> **[Zitat aus dem KI-Expertenkreis]**

---

[96] Siehe dazu die Diskussion in Kap. 2.2, S. 18 ff.
[97] Siehe dazu die Anmerkung zu dem „KI-Effekt" in Kap. 2.2.2

3.2.2 Warum unterscheiden sich KI-Systeme von konventionellen Systemen?

Ausgangspunkt für den Einsatz von KI-Systemen ist die Komplexität der Aufgabe, die es zu lösen gilt. Bei bestimmten Anwendungsfällen ist die Aufgabe zu schwierig oder die zu bewältigenden Situationen so vielfältig, um sie noch regelbasiert beschreiben und von Hand implementieren zu können. Der vielleicht verbreitetste Anwendungsfall ist hier die Erkennung und Klassifikation von Objekten versus Personen (per se schützenswert), eine für Menschen einfach durchzuführende Aufgabe, die jedoch nur mit immensen Aufwand und nicht zufriedenstellendem Erfolg in explizite Regeln zum Anlernen eines Systems gegossen werden kann. Die Komplexität der Anwendung verlangt folglich ein andersgeartetes Vorgehen zur Lösung der Aufgabe. Anstelle einer expliziten Implementierung von Hand wird die Aufgabe an einen Algorithmus übertragen, der die Verbindung zwischen den Eingaben (Beispielbilder) und den Ausgaben (Klassifikation) selbst herstellt. Dieser Prozess der impliziten Regelbildung wird als Selbstlernen bezeichnet. Dieses Selbstlernen von KI-Systemen erfolgt auf Basis umfangreicher und heterogener Beispieldaten, die den Anwendungsbereich möglichst umfassend abbilden. Die Fähigkeit zur Generalisierung auf Situationen, die nicht in dem Lerndatensatz vorkommen, ist dabei prinzipiell gegeben, nimmt aber insbesondere mit steigender Abweichung zu den gelernten Daten ab. Ein Nachlernen auf Basis neuer Daten kann demnach erforderlich sein, um neuartige Situationen bewältigen zu können.

Die Grundlage des Lernens wird jedoch entwicklungsseitig geschaffen. Dazu gehören die Bereitstellung des Lerndatensatzes, die Wahl des Lernverfahrens und die Vorgabe der Lernziele bzw. der Optimierungs- oder Gütekriterien. Der datengetriebene Entwicklungsansatz per se sowie die Entscheidungen, die in Bezug auf Datensatz, Verfahren und Gütekriterien entwicklungsseitig getroffen werden, beeinflussen maßgeblich die Eigenschaften eines KI-Systems. Zu den mächtigsten Lernverfahren gehört die Gruppe der neuronalen Netze.

> ***„Das präsenteste Beispiel ist die Bilderkennung, da es keine alternativen Verfahren gibt, die so gut funktionieren. Dafür werden neuronale Netze eingesetzt. Man kann die auch für sämtliche anderen Prozesse nutzen, beispielsweise zur Überwachung bei der Herstellung, um zu schauen, ob das Produkt gut ist oder Merkmale hat, die nicht so gut funktionieren.“***
>
> **[Zitat aus dem KI-Expertenkreis]**

Gleichzeitig gehören sie zu den Verfahren, die sich in ihrem Eigenschaftsprofil besonders stark von konventionellen nicht-KI-basierten Verfahren unterscheiden.

3.2.3 Worin unterscheiden sich KI-Systeme von konventionellen Systemen?

Wie vorangehend dargelegt ermöglicht das datengetriebene Vorgehen bei KI-Systemen eine Möglichkeit zur Beherrschung komplexer Anwendungsfälle, die mittels konventioneller Vorgehensweisen, die auf einer expliziten Beschreibung des Situationsraums basieren, nicht bzw. nicht in vergleichbarer Güte lösbar wären.

> ***„Man kann den Situationsraum nicht mehr beschreiben und sagen: mach dies oder das, weil die Eingaben zu groß sind. Wenn man versuchen würde, es regelbasiert zu beschreiben, dann würde man bei dem Herunterbrechen in eine explizite Spezifikation Fehler machen und dann ist die korrekte Implementierung unsicher.“***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Der Datensatz ist die Grundlage, auf der die implizite Regelbildung eines KI-Systems aufbaut, und damit zentrales Bestimmungsstück seiner Performanz. Im Gegensatz zu konventionellen Systemen kommt damit dem Aspekt der Datengüte bei der Entwicklung und Weiterentwicklung von KI-Systemen eine zentrale Bedeutung zu. Dazu gehört zum einen eine repräsentative Abbildung der vom System zu beherrschenden Situationen in den Lerndaten einschließlich seltener, insbesondere kritischer Ereignisse. Zusätzlich erforderlich ist ein verlässliches Labelling der Lerndaten, um die korrekte Zuordnung zur Kategorie (bspw. „Objekt“ vs. „Mensch“) gut anlernen zu können. Dieses Labelling bedarf auch einer menschlichen Einschätzung. Da die Güte eines KI-Systems entscheidend davon beeinflusst wird, welche Situationen in der Lernphase vorgelegt wurden, fällt es schwer, die Generalisierungsfähigkeit vom Gelernten und damit das Qualitätslevel des KI-Systems zu bestimmen.

> ***„Die Funktionalität eines KI-Systems hängt weniger vom Algorithmus ab als von den gelernten Eingangsdaten. Aus diesem Grund kann ein Hersteller nie sagen: Ok, das System ist auf einem hohen Qualitätslevel“***
>
> **[Zitat aus dem KI-Expertenkreis]**

Wenn bei Inbetriebnahme eines solchen Systems damit zu rechnen ist, dass bislang unbekannte Situationen auftreten – je komplexer das Umfeld, desto wahrscheinlicher – ist die Möglichkeit oder vielmehr Notwendigkeit gegeben, die Systemfunktionalität durch ein Nachtrainieren zu verändern bzw. zu erweitern.

Mit Einsatz eines KI-Systems verlagert sich das Problem der nicht vollständigen Beschreibbarkeit des komplexen Anwendungsfalls durch explizite Regeln auf die Eigenschaften des Systems. Das KI-System, das die Bewältigung der komplexen Aufgabe implizit gelernt hat, verfügt folglich nicht über einen expliziten und inhärent nachvollziehbaren Code. Da die KI den Code generiert hat und nicht der Mensch, ist schwer nachzuvollziehen, was gelernt wurde und ebenso, warum es gelernt wurde.

Zwischen der Fähigkeit zur Komplexitätsbewältigung und der Nachvollziehbarkeit eines Systems besteht ein nicht auflösbarer Widerspruch.

> ***„Es gibt keine Lernverfahren, die diese Komplexität beherrschen können und gleichzeitig transparent sind. Deswegen setzt man die Systeme ja ein, weil sie die Komplexität beherrschen können, ohne dass man das beschreiben muss.“***
>
> **[Zitat aus dem KI-Expertenkreis]**

Manche weniger komplexe und auch weniger mächtige Ansätze verfügen noch über eine ihnen inhärente Transparenz, die dem Anwender die Nachvollziehbarkeit ihres Verhaltens ermöglicht. In diesem Falle fällt oftmals der Begriff der „White-Box-Modelle“ in Abgrenzung zu „Black-Box-Modellen“ wie bspw. neuronalen Netzen, die keinen derartigen Einblick erlauben. Die Zuordnung von White-Box-Modellen (bspw. Bayessche Netze als nicht dateninferierte Modelle) als der KI zugehöriges Verfahren ist auch aus diesem Grunde unter Experten nicht unstrittig.

> ***„Es ist etwas anderes, ob ich Bayessche Netze [nicht dateninferiert] habe oder neuronale Netze oder sonstige Techniken, die man potenziell mit KI in Zusammenhang bringen könnte.“***
>
> **[Zitat aus dem KI-Expertenkreis]**

Die KI-basierten Regeln in neuronalen Netzen folgen weder Semantik noch Verständnis, sondern der schlichten Optimierung eines vorgegebenen Gütekriteriums auf Basis von Mustern in den Datensätzen. Lernen und Fähigkeiten basieren auf diesen Mustern, und bilden keine Kausalzusammenhänge oder Wissen ab.

> ***„Eine KI kann sehr gut Korrelationen finden, aber keine Kausalitäten.“***
>
> **[Zitat aus dem KI-Expertenkreis]**

Daraus resultieren für den menschlichen Betrachter schwer nachvollziehbare Fehler. Auf scheinbar vergleichbare Situationen kann das System unterschiedlich reagieren. Dieses Phänomen hat dazu geführt, dass man KI-Systemen fälschlicherweise ein nicht deterministisches Verhalten zugeschrieben hat.

> ***„Das ist ja ein großes Fehlurteil, dass ein selbstlernendes System keinem Determinismus unterworfen ist. Es ist nur so komplex, dass es keiner durchschaut. Zeigen wir einem angelernten System hundert Mal dasselbe Bild, kommt hundert Mal dasselbe raus. Das ist ein klarer Determinismus, das Problem ist die Komplexität, in der wir diese Anwendungen einsetzen."***
>
> **[Zitat aus dem KI-Expertenkreis]**

Dieser scheinbare Nichtdeterminismus ist Folge dessen, was als mangelnde Robustheit oder Fehlertoleranz beschrieben wird, d. h. die Fähigkeit eines Systems, seine Funktionsweise auch dann aufrechtzuerhalten, wenn unvorhergesehene Eingaben oder Fehler auftreten. Bei manchen KI-Verfahren kann sich durch die Veränderung eines Parameters oder gar eines einzigen Pixels das ganze Ergebnis

verändern. Bei anderen haben kleine Eingangsänderungen weniger drastische Auswirkungen, so dass bei ähnlichen Eingaben auch große Ähnlichkeit in den Ausgaben besteht. Eine zentrale Frage ist daher, wieviel Ungenauigkeit ein Algorithmus vertragen kann bzw. wie sensibel er darauf reagiert. Bei konventionellen Systemen hingegen ist es durch den Entwicklungsprozess gemäß Best-Practice nahezu ausgeschlossen, dass ein System ein Verhalten zeigen kann, das nicht gewünscht ist.

> ***„Bei konventionellen Systemen hatte man Fehler in der Spezifikation und dadurch einen Fehler im Produkt. Jetzt ist ein [KI-]System so komplex, dass ein Hauch reicht, um die Entscheidung zu kippen."***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Die Verhaltensvariabilität und damit auch die eingeschränkte Kontrollierbarkeit eines KI-Systems hängt entscheidend davon ab, wieviel Raum zur Veränderung eingeräumt wird. Bei geschlossenen Systemen wird diese Möglichkeit mit Abschluss der Lernphase unterbunden. Zwei nach Abschluss der Lernphase identische Systeme werden sich also auch nach Einsatz in unterschiedlichen Umwelten auf dieselben Eingaben vergleichbar reagieren. Es ist jedoch auch möglich, nach Abschluss der Lernphase ein Reagieren auf variable Umgebungsbedingungen zuzulassen. Je abstrakter das Optimierungsziel entwicklungsseitig formuliert wird, desto größer ist der Handlungsspielraum des Systems. Der simplen Selbstoptimierung eines singulären Parameters steht die Befähigung zum komplexen Weiterlernen gegenüber. In dem einen Fall handelt es sich um eine Adaption, deren Art und Umfang man vorher absehen kann; in der Industrie sind das eingekesselte und stark limitierte Prozesse. Im anderen Fall handelt es sich um eine nicht absehbare, und damit nicht ohne zusätzliche Maßnahmen kontrollierbare Weiterentwicklung des Systems, bei dem erwünschtes und nicht erwünschtes Verhalten erst während der Laufzeit unterschieden werden kann.

Eine weitere Quelle nicht gezielt steuerbarer Verhaltensvariabilität ist die Möglichkeit zu emergentem Verhalten im System selbst bzw. im Verbund mit anderen Systemen. Emergenz beschreibt das Phänomen, dass die Analyse der Teilkomponenten eines Systems oder Systemverbunds nicht das zu erklären vermag, was aus der Interaktion dieser Komponenten resultiert.

> ***„Es gibt natürlich immer die Eigenschaft der Emergenz: das Gesamt ist größer als die Summer ihrer Teile. Einzeln betrachtet ist das Verhalten klar, das Gesamtverhalten ist es aber nicht."***
>
> **[Zitat aus dem KI-Expertenkreis]**

Diese sicherheitsbezogenen Eigenschaften von KI-Systemen bergen, wie in den folgenden Kapiteln ausgeführt, erhebliche Herausforderungen für ihren Einsatz im industriellen Bereich.

3.2.4 Welche Herausforderungen birgt das für die Gewährleistung der Sicherheit?

Systeme, die in der Industrie zum Einsatz kommen sollen, müssen nachweislich sicher gestaltet sein (siehe Kapitel 5). Da eine gänzliche Vermeidung jedweder Gefährdung nicht realisierbar ist, gilt es, das Restrisiko unter die noch akzeptierte Schwelle zu drücken. Je höher das mit dem System verbundene Risiko, desto höher sind auch die Sicherheitsanforderungen an das System. Im Industriebereich gilt beim Einsatz risikomindernder Maßnahmen zur sicheren Systemgestaltung nach wie vor das TOP-Prinzip. Zuerst muss die gefahrbringende Maschine inhärent sicher gestaltet sein. Erst danach dürfen organisatorische Maßnahmen greifen, gefolgt von Maßnahmen zum Selbstschutz durch den Menschen. Damit steht der Mensch nicht als Rückfallebene zur Verfügung, um ein etwaiges sicherheitskritisches Systemversagen aufzufangen.

> ***„In der Sicherheitstechnik der Industrieautomatisierung wird der Mensch nicht als risikomindernder Faktor berücksichtigt."***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

In diesem Faktor unterscheidet sich die Sicherheitsbetrachtung in der Industrie von der Sicherheitsbetrachtung im Automobilbereich, bei der im Rahmen der Risikoanalyse zusätzlich zu den Eintrittswahrscheinlichkeiten und Auswirkungen eines Systemversagens noch die mutmaßliche Kontrollierbarkeit durch den Fahrer berücksichtigt wird. Durch die Annahme seines zeitnahen und angemessenen Reagierens können die Sicherheitsanforderungen an ein System und damit die Aufwände des Sicherheitsnachweises erheblich reduziert werden. In der Industrie hingegen ist eine derartige Teildelegation der Verantwortung an den Menschen für seine eigene Sicherheit selten.

Primat ist folglich die inhärent sichere Gestaltung eines Systems. Voraussetzung für den Einsatz eines Systems ist, dass der Hersteller die Erfüllung der Sicherheitsanforderungen nachweisen kann. Formal erfolgt dieser Nachweis zumeist über die Konformitätserklärung zur Norm und in selteneren Fällen über Freigaben durch offizielle Prüfstellen[98]. In beiden Fällen ist eine Sicherheitsprüfung erforderlich, die einem vorgegebenen Prozess folgt, aber gleichzeitig hinreichend Spielraum zur Ausgestaltung der Prüfung für den jeweiligen Anwendungsfall lässt. Die Kenntnis des Kernprinzips des erforderlichen Prüfprozesses ist elementar, um die Herausforderung nachvollziehen zu können, die mit der Einführung von KI-Systemen für die Sicherheit und insbesondere den Nachweis dieser Sicherheit verbunden sind. Die nachfolgenden Ausführungen beschränken sich auf eine vereinfachte Darstellung dieses Kernprinzips, ohne die Vielschichtigkeit dieses Prozesses im Detail nachzuzeichnen.

[98] Anmerkung der Autoren: Der elaborierte Prozess und das Instrumentarium zur Gewährleistung der Sicherheitsanforderungen kann im Rahmen des Berichts nicht näher ausgeführt werden. Die vorliegende Darstellung reduziert die Ausführungen der Experten auf die Kernprinzipien, die beim Einsatz KI-basierter Systeme nicht aufrecht zu erhalten wären.

Das Herzstück der Sicherheitsprüfung ist die Spezifikation, in der die technischen und funktionalen Aspekte des Systems festgehalten sind. Hier werden die Anwendungsfälle des Systems beschrieben und festgehalten, bei welchen Eingaben welche Ausgaben des Systems erfolgen sollen. Die Gesamtheit dieser Funktionen wird anschließend in einzelne Funktionen untergliedert, von denen jede einer spezifischen Überprüfung unterzogen wird. Anschließend erfolgt die schrittweise Zusammenführung zu Funktionsgruppen, die dann auf ihr anforderungsgemäßes Zusammenwirken überprüft werden. Dieses schrittweise Prüfen der Anforderungserfüllung bezeichnet man als Verifikation. Die Verifikation besagt, dass das System der gegebenen Spezifikation folgt und die Prüfergebnisse mit den theoretischen Anforderungen übereinstimmen. Auf die Verifikation folgt im letzten Schritt die Validierung, bei der geprüft wird, ob das finale System in seiner Anwendung die gewünschten Nutzungsziele erreicht.

Elementare Voraussetzung dieses etablierten Sicherheitsnachweises ist demnach die Spezifikation des Systems einschließlich der daraus ableitbaren Zergliederung des Systems in einzeln prüfbare Funktionen. Bei Systemen, bei denen dieser Prüfprozess durchführbar ist, kann der Sicherheitsnachweis erbracht werden. Dazu können unter bestimmten Voraussetzungen (genannt wurden ein begrenztes, nicht variables Einsatzumfeld sowie ein hoher Entwicklungsaufwand) auch Systeme gehören, die im Rahmen der vorliegenden Befragung zumindest von einigen Befragten als KI-Systeme eingeordnet wurden (konkret genannt wurden hier Bayessche Netze). Bei komplexen KI-Systemen wie neuronalen Netzen mit dem im vorherigen Kapitel (3.2.3) beschriebenen Eigenschaftsprofil funktioniert das konventionelle Vorgehen des Sicherheitsnachweises jedoch nicht mehr. Grund dafür ist die datengetriebene Systementwicklung, die zur Beherrschung komplexer Anwendungsfälle genau deswegen eingesetzt wird, weil die explizite Spezifikation zu aufwändig und voraussichtlich fehlerbehaftet wäre, während das datengetriebene Vorgehen eine bessere Erfüllung der Nutzungsziele erreicht als konventionelle Systeme.

> ***„Man setzt neuronale Netze ein, da sie in der Lage sind, zu abstrahieren, von konkreten Problemen zu verallgemeinern, und so muss ich es gar nicht mehr definieren. Wenn ich bspw. von der Personenerkennung im Industriekontext ausgehe, beschreibe ich nicht mehr, was eine Person ist und wie sie aussieht. Das macht es auch schwierig zu überprüfen, wenn ich es nicht genau spezifizieren kann."***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Diese Spezifikation fehlt jedoch im folgenden Schritt als Grundlage des traditionellen Sicherheitsnachweises. Der Use-Case bzw. die Nutzungsziele sind bekannt (bspw.: „Personenerkennung“), aber der Einsatzbereich ist nicht konkret definierbar. Die Szenarien können nicht erschöpfend beschrieben werden, wenn man zugrunde legt, dass jeder unberücksichtigte Parameter bzw. jede unberücksichtigte Ausprägung eines Parameters ein neues Szenario bilden können (siehe Kap. 3.2.3). Somit können nicht alle Eingänge in das System benannt und durch darauf zugeschnittene Tests abgeprüft werden. Die bisherigen Testmethoden, die sich im Bereich der Sicherheit

etabliert haben, reichen folglich nicht mehr aus, da sie darauf basieren, eine spezifizierte Eingabe mit einer spezifizierten Ausgabe zu vergleichen.

> ***„Also das ist der Kern der ganzen Diskussion, dass sich die Methodik, an das Testen heranzugehen, sich durch den KI-Einsatz grundlegend ändert. Da, wo wir konkrete, automatisierte Tests definieren konnten, im Sinne von: „Bitte stimuliere mir gewisse Inputs meines Systems mit folgenden verschiedenen Vektoren“. Und diese Vektoren sind auch ausreichend, da ich sie auf eine Spezifikation, eine Anforderung rückführen kann. Genau das funktioniert jetzt nicht mehr. Das ist der Kern der ganzen Problematik. Das liegt aber nicht nur an der KI, sondern auch am komplexen Use-Case des Systems.“***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Damit ist die Verifikation, die eine Erfüllung der an das System gestellten Anforderungen nachzuweist, nicht mehr durchführbar. Damit fehlt aber das grundlegende Verständnis über die Funktionsweise des Systems, das die Basis des Sicherheitsnachweises bildet. Eine Validierung der KI-Systeme wäre zwar weiterhin möglich, jedoch gilt gemäß dem traditionellen Vorgehen zur Sicherheitsprüfung ein alleiniger Nachweis auf Basis empirischer Feldtests in der Anwendung unter Verzicht auf vorhergehende Verifikation nach den Maßstäben der funktionalen Sicherheit als „zu instabil."

> ***„Es gibt kein Entwickeln ohne Verifizieren in der Sicherheitstechnik. Das ist etwas, was die funktionale Sicherheit so nicht mehr vorsieht.“***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Das Kernproblem bei KI-Systemen besteht demnach im Unvermögen, das bei konventionellen Systemen bewährte Vorgehen des Sicherheitsnachweises auf KI-Systeme übertragen zu können und damit das zur Sicherheitsargumentation erforderliche Systemverständnis herzustellen. Nähme man an, es gäbe ein nicht erklärbares KI-System, das stets dasselbe Verhalten zeige wie ein konventionelles System, so ließe sich trotzdem ausschließlich das konventionelle System als sicher verargumentieren.

> ***„Man muss ja Sicherheit irgendwie argumentieren. Und dieses Argumentieren läuft normalerweise direkt oder indirekt über das Systemverständnis.“***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Unabhängig von der Bedeutung der Sicherheitsargumentation über das Systemverständnis, das elementar für die etablierte Sicherheitsprüfung ist, gibt es auch inhaltliche Gründe, die Inbetriebnahme eines Systems nicht allein von empirischen Tests bzw. Feldtests abhängig zu machen. Schließlich ist nicht vorhersehbar, wie sich das System in all den Situationen verhält, die aufgrund der Komplexität des Anwendungsfalls nicht vorhersehbar waren oder nicht geprüft werden konnten. Daher wird an Ansätzen gearbeitet, die das Wissen um die sichere Bewältigung der Situationen des Anwendungsfalls, das dem datengetriebene

Vorgehen fehlt, entweder im Lernprozess zu gewährleisten oder bei der Überprüfung des KI-Systems nachträglich wiederherzustellen. Zu diesen Ansätzen gehören eine indirekte Spezifikation des Systems über die Beschaffenheit des Lerndatensatzes, Analysen zur Beschreibung und Überprüfung der Systemfunktionalität sowie die mehr oder minder weitgreifende Aufhebung der einst strikten Trennung zwischen Entwicklung und der Inbetriebnahme eines Systems.

Bei dateninferierten KI-Systemen dienen die Daten einerseits dem Anlernen des Systems und andererseits seiner Validierung. Die aus der Validierung ableitbaren Metriken (z. B. Erkennungswahrscheinlichkeit in Prozent) beziehen sich jedoch zwangsläufig auf einen verfügbaren und gelabelten Datensatz und sind nur zu dem Maße auf den Anwendungsfall übertragbar, in dem der Validierungsdatensatz selbigen widerspiegelt. Daher werden hohe Anforderungen an eine vollumfängliche Abbildung des Anwendungsfalls, der erwartbaren Situationen und deren Verteilung im Datensatz gestellt (vgl. UL 4600 „Standard for Safety for the Evaluation of Autonomous Products“ im Automobilbereich). In diesem Kontext hört man oft, dass der Datensatz einschließlich des Labellings die Spezifikation ersetze. Diese Aussage impliziert jedoch zumindest aus Safety-Perspektive keine Gleichwertigkeit zwischen der eigentlich erwünschten expliziten Systemspezifikation, die ein grundlegendes Verständnis der Systemfunktionalität ermöglicht, und einer datenbasierten Spezifikation, bei der die Beschaffenheit der Daten über Performanz und Sicherheit entscheiden. Vielleicht ließe sich auch sagen, die Systematisierung der Daten sei nur ein erneuter Anlauf, etwas zu spezifizieren, das eigentlich zu aufwändig zur Spezifikation ist.

> ***„Die Sicherheit des Systems hängt dann von den Daten ab, die eingesetzt werden. Das ist ein spezieller Fall, den man normalerweise in einem Safety-Kontext vermeidet.“***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Das Systemverständnis gilt nicht nur aus Voraussetzung für die Sicherheit, sondern auch bei nicht sicherheitskritischen Anwendungen als Voraussetzung dafür, die Akzeptanz der Nutzer oder die Fairness KI-basierter Entscheidungen sicherstellen zu können. Dieses Verständnis ist, wie in Kapitel 3.2.3 beschrieben, bei komplexen dateninferierten KI-Methoden wie neuronalen Netzen nicht inhärent gegeben. Daher wird an Methoden gearbeitet, diese Black-Box im Anschluss an den Lernvorgang zu durchleuchten, um verstehen zu können, wie sich die Ausgaben des Systems erklären zu lassen. Dazu gehören beispielsweise Ansätze, die nachträglich kenntlich machen, welche Teile eines neuronalen Netzes welche Aspekte eines Bildes als Grundlage seiner Erkennung und Klassifikation nutzt. Andere Ansätze bemühen sich um eine Analyse der Struktur neuronaler Netze, um von der Klassifikation rückwärts aufzuzeigen, wie diese zustande kam. Um diese Bestrebungen entstand unter dem Schlagwort „Explainable AI“ ein breites Forschungsfeld mit bedeutsamen Beiträgen zur verbesserten Nachvollziehbarkeit komplexer KI. Ob diese als ausreichend erachtet wird, hängt vom Anwendungskontext ab. Grundsätzlich gilt, dass bei kritischen

Anwendungsfällen ein höheres Maß an Systemverständnis erforderlich ist als bei unkritischen. Für einen Einsatz in sicherheitskritischen Anwendungen gilt der Grad des durch diese Methoden nachträglich erzielbaren Verständnisses von der Funktionsweise des KI-Systems jedoch als nicht ausreichend.

> ***„Ich kenne keinen Ansatz, der hinreichend genau ist."***
>
> **[Zitat aus dem KI-Expertenkreis]**

Vorab wurde ausgeführt, dass die Herausforderung beim Einsatz von KI-Systemen nicht darin besteht, dass sie mehr Fehler machen als konventionelle Systeme, sondern darin, dass das zur Sicherheitsargumentation erforderliche Systemverständnis nicht hergestellt werden kann. Die Aussage bezieht sich auf die Problematik des Sicherheitsnachweises und ist in diesem Kontext gültig. Es bedeutet jedoch nicht, dass KI-Systeme per se fehlerfreier funktionieren als konventionelle Systeme. Würde man einen pauschalen Vergleich ziehen wollen, würde man sagen, dass KI-Systeme komplexe Aufgaben wie beispielsweise die Erkennung von Menschen besser lösen als konventionelle Systeme. Konventionelle Systeme werden in ihrer Anwendung jedoch auf das zugeschnitten, was verlässlich umsetzbar und prüfbar ist. Die herausfordernde Aufgabe der Erkennung von Menschen wird daher reduziert auf eine vergleichsweise schlichte, aber mit konventionellen Methoden umsetzbare und prüfbare Unterscheidung zwischen „Hindernis" und „kein Hindernis" – mit der Folge, dass auch jedes unbelebte Hindernis einen Stillstand des Systems hervorruft.

> ***„Die meisten sagen, man kriegt die Fehlertoleranz klassischer Safety-Systeme nicht hin. Das liegt auch an der Problemstellung. Der Algorithmus ist natürlich auch fehlerbehaftet, aber Dinge mit Bildverarbeitung zu lösen sind ja komplexere Probleme als das, was man in Sicherheitssystemen sonst implementiert."***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Gegenwärtig werden systematische Methoden entwickelt, um der aufgrund der höheren Aufgabenkomplexität geringeren Fehlertoleranz (vgl. Kapitel 3.2.3) zu begegnen. Dazu gehört die gezielte Veränderung der Lerndaten, bspw. durch das Löschen von Pixeln oder durch Erzeugung von Rauschen bei der Bilderkennung, um festzustellen, inwieweit diese Abweichungen bei den Eingaben zu Fehlklassifikationen führen. Beispielsweise können in simulationsbasierten Ansätzen systematisch einzelne Parameter variiert werden, um zu sehen, ob das gewünschte Ergebnis dadurch verändert wird. Bei kleinen neuronalen Netzen können damit auch Robustheitsmetriken erzeugt werden. Je weniger das System durch diese Störungen bzw. Variationen beeinträchtigt wird, desto robuster ist es und desto geringer das Risiko einer Fehlklassifikation bei unvorhergesehenen Situationen oder auch gezielten, böswilligen Manipulationen in der späteren Anwendung. Ein derartiger empirischer Nachweis der Robustheit ist für den Einsatz eines KI-Systems in sicherheitskritischen industriellen Anwendungen jedoch nicht ausreichend für den Sicherheitsnachweis.

„Methoden zur Erhöhung der Robustheit und Analyseansätze [zur Schaffung von Transparenz] gehen in die richtige Richtung, aber sie reichen noch nicht aus, um funktionale Sicherheit nachzuweisen.“

[Zitat aus dem Expertenkreis für funktionale Sicherheit]

Schließlich ist nicht vorhersehbar, wie sich das System in unerwarteten und damit nicht abgeprüften Situationen in der späteren Anwendung verhält. Aus diesem Grund wird eine Fortführung des Lernens über die Inbetriebnahme angestrebt, um das KI-System auf Basis der im Einsatz gesammelten Erfahrungen verbessern zu können. Ein derartiger Ansatz eines im Betrieb veränderbaren Systems löst die strikte Trennung zwischen Implementierung und Inbetriebnahme auf und stellt damit nicht nur eine mögliche Antwort auf die Problematik einer nicht vollständigen Beschreibbarkeit aller denkbaren Situationen des Use-Cases dar, sondern auch eine zusätzliche Herausforderung für den Sicherheitsnachweis in Abhängigkeit von der Kontrollierbarkeit dieser Veränderung (siehe Kapitel 5).

„Je mehr Adaptionen ich zur Laufzeit habe, desto mehr muss ich das Safety-Engineering in die Laufzeit verlagern.“

[Zitat aus dem Expertenkreis für funktionale Sicherheit]

Grundsätzlich lässt sich dabei zwischen einer offline vorgenommenen Aktualisierung der im Betrieb erfassten Daten und systemseitigen Adaptionen, die während der Laufzeit ohne vorhergehende Kontrolle erfolgen, unterscheiden. Die offline vorgenommene Aktualisierung bietet die Möglichkeit, den Lernprozess im Rahmen der mit einem KI-System gegebenen Möglichkeiten zu kontrollieren. Der Ansatz wird bereits im Automobilbereich bei Entwicklungen autonomer Fahrfunktionen verfolgt, bei der während der Fahrt Daten erfasst werden, auf deren Basis gezielt Verbesserungen vorgenommen werden können. Darüber hinaus lässt sich das angemessene Funktionieren des Systems bezogen auf Einsatzstunden oder Fahrleistung konkret beziffern. In industriellen Anwendungen erscheint ein vergleichbares Vorgehen weniger erfolgversprechend, da die Anwendungsfälle heterogener sind und es deutlich kleinere Stückzahlen gibt, so dass eine Vergleichbarkeit der Daten und damit der Lernerfahrungen nicht gewährleistet ist.

Grundsätzlich erscheint jedoch der Ansatz, eine Offline-Aktualisierung auf Basis der im Betrieb gesammelten Daten vorzunehmen, der vom Menschen initiiert und kontrolliert wird, gleichermaßen notwendig wie auch – zumindest im Vergleich zu nicht kontrolliert weiterlernenden Systemen – zukünftig umsetzbar, sofern sich ein (derzeit noch nicht ausdetaillierter) Prozess zur Sicherheitsprüfung etabliert (vgl. Absatz 1.2.5).

„Dies ist auch der einzige Weg, der irgendwann gangbar wäre"

[Zitat aus dem Expertenkreis für funktionale Sicherheit]

Während geringfügige Adaptionen in einem oder zumindest wenigen Parametern in engen Grenzen möglich sind, gelten nennenswerte Erweiterungen während der

Laufzeit ohne menschliche Kontrolle in sicherheitskritischen Anwendungen jetzt und zukünftig aus Sicht der Experten als undenkbar. Als Grund dafür wird angeführt, dass weder die Selektion noch die Qualität der Lerndaten steuerbar ist, wenn das System sich diese selbst wählt, und damit nicht die Möglichkeit gegeben ist – zumindest nicht ohne zusätzliche rigide Kontrollmaßnahmen – sicheres Verhalten in der Anwendung zu garantieren.

3.2.5 Welche Entwicklungen liegen vor uns?

Die Frage, welche zukünftigen Entwicklungen sicherheitskritischer KI-Systeme im industriellen Bereich absehbar sind, ließ sich im Rahmen der Befragung nicht vollumfänglich beantworten. Stattdessen lautete die vordringliche Frage vielmehr, unter welchen Voraussetzungen derartige Systeme derzeit und zukünftig in der Industrie überhaupt Einsatz finden können. Bislang werden sicherheitskritische KI-Systeme kaum bis gar nicht im industriellen Bereich eingesetzt. Grund dafür ist die fehlende Rechtssicherheit und damit auch die fehlende Betriebshaftpflicht, die bei konventionellen Systemen bei normkonformen Vorgehen über die Eigenerklärung als Hersteller gegeben ist. Dieses normkonforme Vorgehen ist jedoch aus den im vorherigen Kapitel erläuterten Herausforderungen nicht auf KI-Systeme anwendbar. Stattdessen werden KI-Systeme in der Sicherheitsnorm nur insoweit berücksichtigt, als ihr Einsatz als bisher noch als „not recommended“ (IEC 61508) gilt. Dieser Ausdruck hält zwar die Möglichkeit eines KI-Einsatzes offen, verbindet sie jedoch mit der Notwendigkeit, diesen Einsatz besonders gut zu verargumentieren. Folglich müsste für KI-Systeme ein überzeugender Sicherheitsnachweis geführt werden, ohne dass auf die in der funktionalen Sicherheit etablierten Prozesse der Sicherheitsargumentation zurückgegriffen werden kann. Der alternative Weg, KI-Systeme zur Zulassung zu bringen, wäre die Freigabe durch eine externe Prüfstelle. Dieser Weg wäre für kleine Hersteller per se nicht praktikabel, da dieser ein aufwändigeres Sicherheitsengineering erfordert als ein Sicherheitsnachweis über die Normkonformität. Beiden Ansätzen gemein ist jedoch die Problematik, dass bis heute kein systematisches Vorgehen vorliegt, um eine Sicherheitsbewertung vorzunehmen und konkrete Vorgaben zur Prüfung sicherheitskritischer KI-Systemen zu machen, die ein definiertes Qualitätslevel nachweisen können. Dafür fehlt es bislang noch an den Erfahrungen, die bei dem inzwischen etablierten Vorgehen des Sicherheitsnachweises konventioneller Systeme inzwischen gegeben ist.

> ***„Die Norm zur funktionalen Sicherheit (IEC 61508) hat sich auf Erfahrungen gestützt, die nicht neu waren. Man hat sichere Systeme so konstruiert, und dann hat man eine Norm geschrieben. Und bei KI sind wir in der parallelen Entwicklung. Man versucht gleichzeitig Dokumente zu verfassen, lernt aber erst noch, wie man das zur Zulassung bringen kann.“***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Das bedeutet, dass die Hersteller von Systemen sich nach den aktuell gegebenen Normen richten und solange keine KI-Systeme einsetzen werden, bis sie über die Konformität zu einem neuen Normenwerk – oder etwaige andere Entwicklungen – die

erforderliche rechtliche Absicherung erhalten. In dieser Orientierung an Regularien sieht man einen deutlichen Gegensatz zu den derzeitigen Entwicklungen von KI-Anwendungen im Automobilbereich, in dem man progressiver vorgeht und sich schrittweise einen veränderten Stand der Technik verschafft.

> ***„Wenn sich an den Normen nichts ändert, wird sich auch bei uns nichts ändern."***
>
> **[Zitat aus dem KI-Expertenkreis]**

Ein weiteres Hindernis für den Einsatz von KI-Systemen in der Industrie können die Performanzverluste darstellen, die aus den zur Gewährleistung der Sicherheit gewählten Maßnahmen resultieren. Beispielhaft genannt wurde hier der Einsatz der Mensch-Roboter-Interaktion, bei der die Sicherheit nicht durch den Nachweis sicheren Verhaltens, sondern durch eine Beschränkung von Kraft und Geschwindigkeit (vgl. ISO TS 15066) erreicht wird. Dies kann je nach anvisiertem Einsatzzweck nicht nur eine Verringerung des theoretisch möglichen Nutzens, sondern die Nivellierung jedweden Zusatznutzes zur Folge haben.

> ***„Wenn ich die ISO-Norm einhalte, bringt das keine Vorteile, weil ich 10 Minuten statt 2:30 Minuten [für die kollaborative Fertigung eines Produkts] brauche."***
>
> **[Zitat aus dem KI-Expertenkreis]**

Eine andere Möglichkeit, KI in sicherheitskritischen Anwendungen einzusetzen, besteht bei der zusätzlichen Absicherung durch ein konventionelles System. In diesen Fällen stellt die KI einen Teil des eigentlichen Produkts, dessen Sicherheit durch ein konventionelles Sicherheitssystem gewährleistet wird. Die KI-Komponente übernimmt die nicht sicherheitskritischen Aufgaben, während die sicherheitskritische Abschirmung vom Menschen durch prüfbare, nicht KI-basierte Sicherheitsstrategien (Umfeld- und Kontaktsensoren, Laserscanner o. ä.) gewährleistet ist.

> ***„Die Sicherheitstechnik muss immer der KI vorausgehen und sie stoppen können. D.h. KI ist für mich immer ein Teil des Gesamtsystems, das bevormundet wird durch die Sicherheit. Anders geht's nicht, sonst geht die Personensicherheit den Bach runter."***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Damit muss nur nachgewiesen werden, dass die konventionelle Sicherheitstechnik das Gesamtsystem in einem sicheren Zustand hält, während ein Nachweis für die KI-Komponente nicht erforderlich ist. Dies wäre erst dann erforderlich, wenn die KI die Sicherheitsfunktion übernehmen und beurteilen soll, ob eine Situation kritisch wird oder nicht – wenn also beispielsweise eine Kamera anstelle eines LIDAR-Scanners eingesetzt würde und ein neuronales Netz die Erkennung leisten sollte. Dieses Vorgehen wäre mitunter produktiver als die Eingrenzung der KI durch ein konventionelles System, aber führt zu eben dargestellter Problematik der noch fehlenden Möglichkeit des Sicherheitsnachweises.

Bislang ist jedoch ein Einsatz von KI-Systemen im sicherheitsrelevanten Kontext nur unter Einschränkung ihrer Performanz durch nachweislich sicherheitsgewährleistende Maßnahmen (inhärente Leistungsbeschränkungen oder Begrenzung durch

konventionelle Systeme) möglich. Darüber hinaus besteht noch die theoretische Möglichkeit, den Einsatz von KI-Systemen auf eng begrenzte und damit beherrschbare Anwendungsfälle zu beschränken. Diskutiert wird diese Möglichkeit insbesondere im Automobilbereich unter dem Begriff Operative Design Domain (ODD). Als Beispiel dient dort die Zulassung autonomer Fahrfunktionen ausschließlich auf Autobahnen. Mit dem Konzept der ODD ist jedoch die Anforderung verbunden, stets erkennen zu können, ob sich das System innerhalb oder außerhalb selbiger befindet. Inwiefern es eine nutzbringende Übertragung dieses Prinzips auf Anwendungsfälle in der Industrie gibt, lässt sich auf Basis der vorliegenden Ergebnisse nicht beurteilen. In nicht sicherheitskritischen Anwendungen hingegen bestehen vergleichsweise niedrige Hürden für den Einsatz von KI-Systemen. In der Qualitätssicherung oder der Instandhaltung von Anlagen werden KI-basierte Verfahren mit Erfolg eingesetzt. In diesen Anwendungsfällen erscheint auch vieles möglich, was bei etwaigen zukünftigen Entwicklungen von Normenwerken zur Prüfung sicherheitskritischer Systeme weithin als ausgeschlossen gilt (bspw. der Einsatz weiterlernender Systeme). Beim Blick auf derartige Entwicklungen ist jedoch zu beachten, dass Demonstrationen neuer Use-Cases von KI nicht zwangsläufig unmittelbar oder mittelfristig bevorstehende Entwicklung anzeigen, sondern perspektivisch weit in der Zukunft liegen (bspw. selbstverändernde Fabriken).

> ***„In publikumswirksamen Darstellungen habe ich schon vieles gehört, wo man sagt, wow, das müsste schon alles da sein. Aber gesehen habe ich das nie.“***
>
> **[Zitat aus dem KI-Expertenkreis]**

Eine Norm mit Vorgaben zur Prüfung von KI-Systeme gilt als wichtigste Voraussetzung für deren Einführung in sicherheitskritischen Anwendungsbereichen. Die derzeitige Situation wird als eine Wartehaltung beschrieben, insofern als der Einsatz neuer Technologien weniger von der technischen Reife als von den erforderlichen Rahmenbedingungen zur Inverkehrbringung abhängt. Einen Entwicklungsvorsprung sehen einige in den Bereichen, in denen das Fehlen eines derartigen Regelwerks die Entwicklung weniger zu bremsen vermag. Dazu gehören zum einen die Entwicklungen in China oder den USA, die regulatorisch andere Maßstäbe und eine andere Sicherheitskultur haben, und zum anderen die Entwicklungen im Automobilbereich, in denen durch das aktive Vorantreiben autonomer Fahrfunktionen schrittweise ein neuer Stand der Technik geschaffen wird.

Die Schwierigkeit bei den laufenden Bemühungen um eine Sicherheitsnorm für KI-Anwendungen liegt zum einen darin, dass die bewährten Verfahren nicht anwendbar sind und die Erfahrungswerten bezüglich eines praktikablen und zielführenden Vorgehens fehlen, die einstig bei der Erstellung des heute gültigen Normenwerks die Grundlage bildeten. Zum anderen wird die Übereinkunft über neue Vorgehensweisen dadurch erschwert, dass bei dieser Aufgabe die Fachdisziplinen KI und Safety mit unterschiedlichen und in diesem Fall entgegengesetzten Leitprinzipien aufeinandertreffen. Aus Safetysicht herrscht das Primat der Sicherheit vor, das man durch den Einsatz von KI-Systemen verletzt sieht.

> ***„In der Zukunft scheitern wir daran, dass der Mensch nicht verletzt werden soll. Das ist das Unabänderliche. Da geht keiner dran und sagt: Wir haben eine neue Technik, da kann zwar einer draufgehen, aber das machen wir jetzt mal.“***
>
> **[Zitat aus dem Expertenkreis für funktionale Sicherheit]**

Gleichzeitig wird der mit dem Einsatz von KI verbundene Nutzen nicht so hoch bewertet wie es in der Fachdisziplin KI der Fall ist. Das liegt auch daran, dass man bisher noch keine Use-Cases für KI im industriellen Bereich sieht, die auch nur ansatzweise den Nutzwert erreichen, der im Bereich Kommunikation oder Internet durch KI ermöglicht wird. Die Fertigung einschließlich der individualisierten Fertigung erscheint bereits so optimiert, dass sich der Mehrwert durch den Einsatz von KI nicht offenbart. Damit gesellt sich zu den hohen Kosten (Gefährdung) auch ein als gering erachteter Nutzen und damit wenig Grund, von den etablierten und vor allem in der Praxis bewährten Prozessen abzuweichen. Das hohe Niveau der in der Industrieautomatisierung seit langem etablierten Sicherheitstechnik ist mit ein Grund für die mangelnde Bereitschaft, den bewährten Weg zugunsten des Einsatzes neuartiger Technologien zu verlassen.

Die Fachdisziplin KI hingegen sieht sich in erster Linie mit den Hindernissen für den Einsatz von KI-Systemen konfrontiert. Der Nutzen von KI-Systemen wird höher angesetzt, dahingehend, dass ihr Einsatz entweder deutliche Performanzsteigerungen verspricht oder gar einen Einsatz neuer Technologien überhaupt erst wirtschaftlich und damit möglich macht (bspw. Mensch-Roboter-Kollaboration).

> ***„Wenn wir die Prozesse zur Einhaltung der Sicherheit unverändert lassen, dann werden wir das Werteversprechen der Industrie 4.0 nur bedingt einhalten.“***
>
> **[Zitat aus dem KI-Expertenkreis]**

Gleichzeitig fällt der Blick auf die Gefährdung zwangsläufig anders aus, da der Spezifikation von Systemen und der darauf aufbauenden Verifikation nicht das gleiche Gewicht beigemessen wird. Zum einen wird Kritik an dem Prinzip der Spezifikation als Kernpunkt des Sicherheitsnachweises geäußert, da die Verifikation nicht die Sicherheit per se nachweist, sondern die Erfüllung der festgelegten Anforderungen, die sich in der Praxis als unvollständig erweisen können. Zum anderen werden empirisch gewonnene Erfahrungen höher gewertet. Aus diesem Grund erscheint auch die Zulassung von KI-Systemen auf Basis empirischer Prüfungen denkbar, gerade wenn die Zulassung schrittweise erfolgt, das Sammeln von Erfahrung ermöglicht und die Bewährung in der Praxis die Voraussetzung dafür ist, die Stückzahl der Systeme anzuheben oder die Performanz weiter zu steigern. Damit würde sich die Sicherheitsprüfung zumindest teilweise in den Einsatz verlagern und eine Produktbeobachtung erfordern – ein reaktiver Ansatz, der dem proaktiven Sicherheitsverständnis der Safety widerspricht.

3.2.6 Zusammenfassung

Zusammenfassend lässt sich sagen, dass mit dem schwer fassbaren, über die Zeit veränderlichen Begriff KI keine klare Abgrenzung möglich ist, welche Verfahrensgruppen bzw. Systeme darunterfallen und welche nicht. Folglich ist KI in

erster Linie als Arbeitsbegriff zu verwenden, der aus heutiger Sicht mit Blick auf die operative Komponente des maschinellen Lernens eine neuartige Engineeringmethode mit anderen Eigenschaften als konventionelle Systeme beschreibt. Diese Eigenschaften gehen überwiegend auf den dateninferierten Entwicklungsprozess zurück. Vorliegen sowie Ausprägung dieser Eigenschaften unterscheiden sich zwischen verschiedenen KI-Verfahren und sind bei datengetriebenem Vorgehensweisen wie neuronalen Netzen besonders ausgeprägt. Die vorgenommene Beschreibung der Eigenschaften von KI-Systemen erfolgte ausschließlich mit Blick auf ihre mögliche Bedeutsamkeit für deren sicheren Einsatz und bildet die Grundlage für die im folgenden Kapitel dargestellte Taxonomierung. Nicht sicherheitsbezogene Eigenschaften, die als wünschenswert oder gar zwingend erforderlich erachtet werden können (bspw. Fairness, Diskriminierungsfreiheit oder Nachhaltigkeit), blieben dabei unberücksichtigt.
Aufgrund der besonderen Eigenschaften von KI-Systemen (bspw. mangelnde Robustheit und Transparenz aufgrund des datenbasierten Entwicklungsprozesses) kann das zur Sicherheitsargumentation erforderliche Systemverständnis nicht hinreichend herzustellt werden und damit das bei konventionellen Systemen bewährte Vorgehen des Sicherheitsnachweises nicht auf KI-Systeme übertragen werden. Ein Einsatz von Systemen, bei denen die Sicherheit nicht auf Basis des Systemverständisses verargumentiert werden kann, ist bislang nicht möglich. Die grundlegende Frage ist daher, inwieweit man vom etablierten Prozess zur Gewährleistung der Sicherheit, der für komplexere KI-Systeme nicht einsetzbar ist, abrücken kann bzw. sollte, um die mit dem Einsatz von KI erwarteten Potenziale auszuschöpfen – und wie ein neuer Prozess zu gestalten wäre. Während man entwicklungsseitig auf neue Wege zur Sicherheitsbewertung von KI-Systemen dringt, steht man in der funktionalen Sicherheit der Abkehr vom langjährig Bewährten ablehnend gegenüber. Als erforderliche Voraussetzung für eine Verbindung beider Perspektiven erscheint ein geteiltes Verständnis bei der Definition und Bewertung von Sicherheit – ob im Vergleich zu bisherigen Systemen (nicht möglich bei neuartigen Anwendungsfällen), dem Menschen (dessen Fehler anders bewertet werden als ein identischer Fehler eines Systems) oder einem anderen normativen Maßstab (bspw. absoluten Schwellwerten). Diese zentrale Frage des Sicherheitsverständnisses, das Leitplanken für die weitere Entwicklung und Überprüfung zukünftiger KI-Systeme setzt, sollte dabei nicht nur aus technischer, sondern auch aus ethischer Perspektive betrachtet werden.

3.3 Weiterverwendung der Befragungsergebnisse als Grundlage der Taxonomie

Aus den vorliegenden Ergebnissen der Befragung wurden die zentralen Eigenschaften KI-basierter Systeme einschließlich der sicherheitsbezogenen Merkmale ihrer Anwendung abgeleitet und in Kriterien eines sicheren Funktionierens übersetzt. Entfernt wurden Kriterien ohne jedweden Bezug zur Sicherheit (bspw. Fairness). Die verbleibenden Kriterien wurden systematisiert und zur besseren Beschreibbarkeit heuristisch als Unterkategorie verschiedenen übergeordneten Dimensionen zugeordnet. In einem iterativen Prozess erfolgte ein Abgleich der sicherheitsbezogenen Kriterien mit der rechtlichen Beurteilung eines Systems mit eben

dieser Eigenschaft und deren anschließende Präzisierung der Auswahl, Definition und heuristischen Zuordnung.

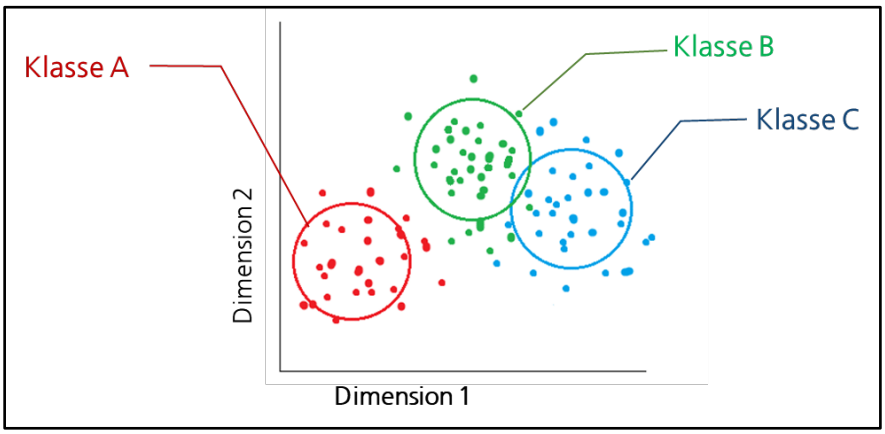
Die finale Version in der Taxonomie enthält sieben Dimensionen mit jeweils zwei Unterkategorien, die sicherheitsbezogene Merkmale des Systems auch unter Berücksichtigung seines Einsatzgebiets beschreiben. Dabei handelt es sich vorwiegend, aber nicht ausschließlich um Merkmale, die den Unterschied zwischen konventionellen Systemen und KI-Systemen beschreiben. Aus diesem Grund wird auch von der „Taxonomie sicherheitsbezogener software-physischer (KI) Systeme“ anstelle KI-basierter Systeme gesprochen. Da die Taxonomie mit Blick auf Anwendungen im industriellen Bereich entwickelt wurde, ist eine Übertragbarkeit auf andere Anwendungsbereiche – bspw. Robotik im Smart Home-Bereich oder der medizinischen Versorgung – zum jetzigen Zeitpunkt nicht geprüft. Vermutlich wären Modifikationen zu deren Anpassung an ein neues Einsatzfeld mit anderem Anforderungsprofil erforderlich.

4 Taxonomie software -physischer KI-Systeme

4.1 Vorbemerkungen

Im Folgenden wird ein als Taxonomie bezeichnetes Schema vorgestellt, das den Ordnungsrahmen für das systematische Vorgehen bei der Erstellung des Rechtsgutachtens (Kapitel 5, S. 85 ff.) liefert. Mit Hilfe dieses Rahmens können Szenarien mit potentiell sicherheitskritischen physischen Systemen (Maschinen, Anlagen, Fahrzeuge, etc.), die Softwarekomponenten mit KI-Anteilen enthalten, einer rechtsrelevanten strukturierten Analyse unterzogen werden. Die Entwicklung der Taxonomie folgt dieser Zielsetzung. Ein anderer als rechtlicher Fokus hätte eine andere, u.U. dafür passendere Taxonomie zur Folge gehabt.
Das Ziel vieler Taxonomien ist eine Reduzierung der Komplexität eines umgrenzten Gegenstandsbereiches, um Aussagen (in dieser Studie rechtliche Konsequenzen) über einzelne Systeme ableiten zu können (die Zielbewertung), ohne diese jeweils detailliert untersuchen zu müssen. Die Frage ist, ob es Merkmale oder Merkmalskombinationen von Systemen aus dem Gegenstandsbereich gibt, die erste – oder im Idealfall vollständige – Rückschlüsse bezüglich der Zielbewertung ermöglichen. Wenn also wie in vorliegender Studie die Frage nach rechtlichen Konsequenzen bei der Einführung von KI-Systemen gestellt wird, werden diese Systeme stellvertretend für sehr viele unterschiedliche Eigenschaften bezüglich der rechtlichen Konsequenzen durch die Merkmale und Merkmalskombinationen der Taxonomie charakterisiert. Umgekehrt hilft die Taxonomie auch, den komplexen Gegenstandsbereich durch das aufgebaute Ordnungsprinzip leichter gedanklich durchdringen zu können.
Wie beispielsweise alle denkbaren Arten von Hämmern möglicherweise durch die Dimensionen „Material“, „Formtyp“, „Gewicht“, „Füllung“ und „Härte“ charakterisiert sind, so bilden in der im Folgenden dargestellten Taxonomie sieben Dimensionen das Gerüst zur Charakterisierung potentiell sicherheitskritischer software-physischer (KI-)Systeme, deren Softwarekomponenten möglicherweise auch KI-Anteile enthalten können. Der Geltungsbereich der Taxonomie erstreckt sich aber auch auf Systeme mit Softwarekomponenten, die nicht der KI zuzuordnen sind – insbesondere, wenn die damit realisierten Eigenschaften den KI-Algorithmen ähneln. Ein konkretes Anwendungsszenario mit einer speziellen Konstellation von KI-Systemen kann bezüglich der Eigenschaften bzw. Merkmale der Dimensionen charakterisiert werden. Resultat ist dann ein dem Anwendungsszenario zugeordnetes Tupel von Merkmalen. Dieses Tupel repräsentiert eine Vielzahl von Systemen und Szenarien mit ähnlichen Eigenschaftsmengen und bietet damit die Möglichkeit allgemeingültiger rechtlicher oder weiterer (z. B. Sicherheits-) Betrachtungen. Die Systeme oder Szenarien können zwar in vielen Beschreibungsebenen sehr unterschiedlich sein; bezüglich der hier interessierenden Ebenen der Sicherheit und des Rechtsaspektes zeigen sie aber durch die Dimensionen der Taxonomie definierte Ähnlichkeit.
Abbildung 1 demonstriert dieses Einordnen in das Taxonomieschema. Ein Punkt repräsentiert ein System im Situationskomplex, beschrieben durch die Merkmale der

Dimensionen der Taxonomie. Daraus können Cluster gebildet werden, die mit jeweils unterschiedlichen rechtlichen oder anderen Konsequenzen verbunden sind. Für einen neuen Fall kann dann allein aus der Verortung im Raum der Merkmalskombinationen



auf mögliche rechtliche oder andere Konsequenzen geschlossen werden.

Abb. 4.1 Verdeutlichung der Klassenbildung von Systemen im Merkmalsraum der Dimensionen der Taxonomie

Wie bemerkt ist die Taxonomie auf die Bewertung von Systemen ausgelegt, die KI-Komponenten enthalten. Einige Dimensionen, Unterkategorien und Ausprägungen sind nur vorhanden, da KI-Systeme Besonderheiten gegenüber anderer, in software-physischen Systemen verwendeter Software aufweisen. Generell sind auch KI-Systeme Softwaresysteme. Für letztere gibt es allerdings etablierte Methoden, um in Verbindung mit einem möglicherweise gefährdenden physischen System einen Sicherheitsnachweis zu führen. Für die Klasse von KI-Systemen, deren Verhalten in starkem Maße von Daten bestimmt sind und sich u. U. nach Inbetriebnahme auf Basis neuer Daten noch verändert, bedarf es aber geänderter Herangehensweisen. Obwohl es heute ausschließlich KI-Software ist, die diese Merkmale aufweist, gelten die Überlegungen in gleichem Maße auch für andere Software, also Nicht-KI-Software mit den gleichen Merkmalen. Triebfeder der Taxonomieerstellung war es, die Besonderheiten dieserart Software in einem speziell zugeschnittenen Merkmalsraum abzubilden. Es waren also die Rechtsrelevanz und die softwaretechnischen Besonderheiten bestimmter KI-Systeme, die leitend für die Auswahl der Dimensionen der Taxonomie und deren Attribute waren, und zwar stets unter dem speziellen Sichtwinkel der Sicherheit (Safety [99]). Nur am Rande ist die Sicherheit im Sinne von Security berücksichtigt.

Diese spezielle Sicht auf Safety unterscheidet sich von anderen Taxonomien in Verbindung mit KI. Beispielsweise wird in dem Whitepaper „Vertrauenswürdiger Einsatz von Künstlicher Intelligenz“[100] ein ebenfalls mehrdimensionales Klassifikationsschema vorgestellt, das in Hinblick auf eine Zertifizierung von KI mit dem Schwerpunkt einer KI in Informationssystemen entwickelt wurde. Das Thema Safety

[99] Wenn im Folgenden von Sicherheit gesprochen wird, ist damit immer Safety gemeint.
[100] Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIS (Hrsg.) St. Augustin, 2019

spielt dort allerdings nur eine Nebenrolle. Auf der anderen Seite kommen Hauptdimensionen des dortigen Klassifikationsschemas wie „Ethik“, „Datenschutz“ oder „Fairness“ in der hier vorgestellten Taxonomie nicht vor. Andere dort vorhandene Dimensionen wie „Transparenz“ sind in der vorliegenden Taxonomie wiederum etwas anders belegt.

Solcherart mehrdimensionale Klassifikationsschemata unterscheiden sich fundamental von eindimensionalen Entwicklungsstufen wie beispielsweise die 10 Stufen der Automatisierung von Entscheidungsunterstützung von Sheridan und Verplank [101]. Ebenso wie die Stufen der Entwicklung hochautomatisierter Fahrzeuge [102] charakterisieren diese Klassifikationen den Übergang technischer Entwicklungen bezüglich einer einzigen Dimension; in den genannten Beispielen wäre dies die Dimension des Anteils menschlicher Tätigkeit bei der Steuerung, Kontrolle oder Manipulation eines technischen Systems [103].

Hinsichtlich der in vorliegendem Bericht vorgestellten Taxonomie gilt zu beachten, dass der Bereich von durch Software beeinflussten physischen Systeme, für die funktionale Sicherheit bzw. Produkt- oder Betriebssicherheit eine Rolle spielen, gemäß der Aufgabenstellung des Projektes sehr breit ist. Für die Elemente der Taxonomie hat das zur Folge, dass in der Regel die Festlegung von Ausprägungen der ordnenden Kategorien nicht allgemein möglich ist. In der Anwendung der Taxonomie muss der Anwender die Ankerbildung oder möglicherweise eine quantitative Festlegung zugeschnitten auf das spezielle Anwendungsfeld selbst gestalten. Das bedeutet auch, dass das oben in Abbildung 1 schematisch erläuterte Vorgehen der Clusterbildung nicht so mathematisch exakt durchexerziert werden kann, wie die Abbildung suggeriert. Zudem ließen sich die Dimensionen wegen der Komplexität des Gegenstandsbereiches in Verbindung mit der vorgegeben Baumstruktur des Klassifikationsschemas auch nicht strikt disjunkt realisieren; es gibt inhaltliche Überschneidungen.

4.2 Übersicht über die Taxonomie

Die Taxonomie software-physischer KI-Systeme 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. mit Sicherheit zu tun haben. Die entweder mit Begriffen belegten oder beschriebenen Ausprägungen wurden aus den Auswertungen der Experteninterviews oder aus der Sichtung von Literatur abgeleitet.

---

[101] Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. (Technical Report). Cambridge, MA: MIT Man-Machine Systems Laboratory.

[102] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE J 3016, 2016

[103] Auch auf der untersten Stufe der SAE-Automatisierungsstufen, als „Manuelles Fahren“ bezeichnet, werden von technischen Systemen Ausgaben des Menschen übernommen. Rein manuell wäre ein Fahrzeug, wenn es auch vom Menschen mechanisch bewegt würde, wie das beispielsweise beim Fahrrad der Fall ist.

4.2.1 Dimension Veränderbarkeit

Veränderungen technischer Systeme im Betrieb sind allgegenwärtig, seien es ungewollte (z.B. Alterungsprozesse) oder gewollte, wie z.B. das „Einfahren“ von Kolbenringen. Mit den adaptiven Regelkreisen gibt es seit den 80er-Jahren [104] technische Systeme auf dem Markt, die bewusst so konzipiert sind, dass sie ihr Verhalten an das zu regelnde System anpassen, indem sie mit Hilfe von Systembeobachtung lernen (identifizieren), welche Eigenschaften dieses hat und sich selbst dann so modifizieren (adaptieren), dass das Gesamtverhalten optimiert ist [105]. Diese Veränderungen sind sehr kontrolliert und die Auswirkungen gut abschätzbar. Ursache dafür ist einerseits, dass die Anzahl der Größen, die ein System beschreiben und die dann Grundlage einer Veränderung des Reglers ist, in der Regel im einstelligen Bereich liegt. Andererseits sind die Mechanismen der adaptiven Änderungen detailliert festgelegt und die Grenzen berechenbar. Hinzu kommt, dass die Zeitdauer fundamentaler Änderungen relativ kurz ist und in der Regel nur bei der Inbetriebnahme und ersten Anpassung erfolgt. Im weiteren Verlauf gibt es dann eher kleine Änderungen – z.B. bei der Anpassung an Abnutzungs- oder Alterungsprozessen. Trotz dieses hohen Maßes an Kontrollierbarkeit zeigen auch adaptive Regler ein gewisses Maß an Emergenz[106], da die Änderungen gemäß einem Gütekriterium erfolgen.

Die Möglichkeit einer genauen Abschätzung der Folgen von Änderungen im Betrieb schwinden in dem Maß, in dem Menge und Vielfalt an Daten, die Ursache der Anpassung bzw. des Lernens sind, steigen, die Zeitdauer des Lernprozesses steigt und die Emergenz im Adaptionsprozess zunimmt. Je komplexer die Datengrundlage ist, je länger der Veränderungsprozess dauert und je weniger explizit die Adaptions- und Lernprozesse vorgegeben sind, desto weniger besteht die Möglichkeit der Vorhersage der Konsequenzen des Änderungsprozesses. Ein hohes Maß an Emergenz ist nun ein besonderes Merkmal künstlicher neuronaler Netze in der heutigen Anwendung. Allein daraus resultiert eine besonders schlechte Vorhersehbarkeit des Verhaltens im realen Einsatz. Da das Verhalten neuronaler Netze im Wesentlichen durch einen Lernprozess auf Grundlage von Daten charakterisiert ist, besteht in der Auswahl und Kontrolle von Lerndaten in Verbindung mit ausgiebigen Tests die Möglichkeit, die Vorhersehbarkeit des Verhaltens zu steigern. Da ein Mehr an Daten (bei entsprechender Kontrolle der Qualität) die Leistungsfähigkeit (z. B. die Klassifikationsgüte) der Netze steigert und zur Anpassung an konkrete Situationen notwendig ist, ist der Lernprozess – z. B. bei Spracherkennern – auch in den laufenden Betrieb eingebunden. Diese im Folgenden „Weiterlernende

---

[104] Anfänge gehen bis in die 50er-Jahre zurück. Z.B. Gabor D, Wilby WPL, Woodcock R (1959) A universal non-linear filter, predictor and simulator which optimizes itself by a learning process. Proc Inst Electron Eng 108(Part B):1061–, 1959, Gregory PC (ed) (1959) Proceedings of the self adaptive flight control symposium. Wright Air Development Center, Wright-Patterson Air Force Base, Ohio

[105] Z.B. K. J. Astrom, B. Wittenmark: *Adaptive Control*. 2. Auflage. Addison-Wesley, 1989.

[106] Emergenz wird hier als übergeordneter Begriff einer Nichtexplizitheit bzw. eines „Aus-sich-Heraus“ verstanden. Er beinhaltet Adaptivität, Autonomie, Selbstorganisation, Autopoiesis. Siehe Dimension „Kontrollierbarkeit“ und Glossar.

Systeme [107]“ genannten Vertreter stellen besondere Anforderungen für den Nachweis des sicheren Verhaltens, da u.U. die Vorhersagbarkeit ihres Verhaltens nicht ausreichend ist. Hinzu kommt, dass es sein kann, dass die Änderungen im laufenden Betrieb so stark sind, dass ein fundamental anderes Verhalten resultiert, was dann beispielsweise die Kontinuitätsannahme einer Maschine in Frage stellt. Es stellt sich folglich die Frage, ob bei derartig starken Änderungen noch die Annahme gerechtfertigt ist, dass die geänderte Maschine noch als die vom Hersteller ausgelieferte Maschine zu betrachten ist?

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**[108] unterschieden, während sich die Ausprägung der **Veränderbarkeit des Umfelds**[109] zwischen den beiden Extremen geringfügiger und hoher Komplexität aufspannt. Der Aspekt der Veränderbarkeit bezieht sich in diesem Kontext nur auf die Veränderbarkeit während des bestimmungsgemäßen Einsatzes des Systems. Veränderungen, die Teil des Entwicklungsprozesses sind, werden hierbei nicht berücksichtigt.

Keine Veränderbarkeit eines Systems liegt dann vor, wenn das Systemverhalten nach Auslieferung durch den Hersteller während der gesamten Betriebsphase unverändert bleibt. Folglich werden zum Auslieferungszeitpunkt identische Systeme unabhängig von der seit ihrer Inbetriebnahme verstrichenen Zeit und der Gegebenheiten ihres Einsatzortes in vergleichbaren Situationen stets das gleiche Verhalten zeigen. Ist dies nicht gegeben, gilt das System als veränderbar. Art und Ausmaß dieser Veränderbarkeit kann 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 ist von letzteren im System selbst angelegt. Die nutzerinitiierte Veränderung erfolgt beispielsweise im Rahmen eines gezielten Teachings mit dem Ziel, das System im Betrieb auf seine aktuelle Aufgabe hin anzulernen. Dies kann beispielsweise über eine manuelle Führung der Aktoren eines Roboterarms erfolgen oder über Demonstration durch den Nutzer und anschließende Imitation des Beobachteten durch das System. Gezielte Änderungen können auch auf Basis einer kontrollierten (Weiter-)Entwicklung durch Software-Updates in den Bestand ausgerollt werden, über Firmware-Änderungen bis hin zu applikativen Modifikationen. Solche Updates können auch auf Basis von im Betrieb gesammelten Daten stattfinden. Dieser Mechanismus wird derzeit beispielsweise bei den Fahrzeugen des Herstellers TESLA praktiziert. Durch dieses Vorgehen kann der

---

[107] In der Literatur werden sie oft nur „lernende Systeme“ oder „selbstlernende Systeme“ genannt. Im Zusammenhang mit dem Sicherheitsnachweis von software-physischen Systemen muss aber eindeutig der Systemstatus vor und nach Auslieferung unterschieden werden. Da lernende Systeme auch während des Entwicklungsprozesses lernen, muss eine Unterscheidung getroffen werden, ob die Systeme nur vor der Auslieferung lernen oder noch im Betrieb „weiterlernen“.

[108] In der englischsprachigen Literatur unter *model drift* gefasst

[109] In der englischsprachigen Literatur unter *concept drift* gefasst

Hersteller die Kontrolle über das Systemverhalten behalten, aber gleichzeitig die Vorteile der Systemverbesserung über die Auswertung zusätzlicher Daten nutzen.
Von den Entwicklern bewusst angelegt sind auch solche Systeme, die während des Betriebs anfallende Daten sammeln und zur Modifikation ihres Verhaltens heranziehen. Im gewissen Sinne ist dies schon bei sogenannten adaptiven Regelsystemen der Fall. Die Anpassung im Betrieb stellt also eine höhere Stufe der Veränderbarkeit im Betrieb dar. Das Ausmaß dieser Veränderbarkeit bei ihrem Einsatz kann theoretisch beträchtlich variieren; Einsatz im industriellen Bereich finden derzeit jedoch höchstens adaptive Systeme, deren Veränderbarkeit auf geringfügige Verhaltensänderungen zur Optimierung eines oder weniger Parameter in Abhängigkeit von der Umwelt bzw. Beteiligten begrenzt ist. Ein Beispiel hierfür ist die Anpassung der Dicke der aufzutragenden Lackschicht in Abhängigkeit von der gegebenen Witterung in der Werkshalle.
Eine neue Qualität der Veränderbarkeit ergibt sich bei **bedeutenden Veränderungen** während des Betriebs, wenn die Konsequenzen der Veränderungen wegen der hohen Komplexität nicht mehr vollständig vorhersehbar sind. Das Systemverhalten verändert sich mit der Zeit fundamental, indem das System während des Betriebs Betriebsdaten für die Modifikation des eigenen Verhaltens (in der Regel eine Optimierung) nutzt. Eine Veränderung während des Betriebs auf Basis zusätzlicher Daten ist deshalb ein „Weiterlernen“. Die Unterscheidung zur Adaptivität liegt in der Komplexität der Daten, die Basis der Veränderung sind, und in der Komplexität der durch die Daten erzeugten Verhaltensänderungen. Diese Systeme zeichnen sich in der Regel dadurch aus, dass die Wirkung von einzelnen Daten auf das Verhalten nur wenig abschätzbar ist. Generell ist ein Lernen während des Betriebs auch durch Nicht-KI-Software realisierbar, beispielsweise basierend auf modellbasierten Berechnungen. Die Vorhersehbarkeit des Systemverhaltens kann bei entsprechendem Wissen über die möglichen Lerndaten dann wesentlich höher sein.
Neben Veränderungen im System kann es während des Betriebes auch Änderungen im Umfeld geben, die infolge von Adaptions- und Lernprozessen Einfluss auf das Verhalten der Systeme im Betrieb haben. Hierbei ist zu unterscheiden zwischen dem Ausmaß und der Kontrollierbarkeit dieser Veränderungen.
Zu den **geringen Änderungen**, die in der Regel bei der Auslegung oder bei der Auswahl von Steuerungssystemen schon heute berücksichtigt werden, gehören beispielsweise Änderungen der Temperatur während des Betriebs. Das kann Auswirkungen auf das Systemverhalten haben, wird aber in der Regel abgefangen.
Im Zusammenhang mit Anwendungen von ML-Verfahren kommen aber mögliche **nicht vorhersehbare Änderungen** hinzu, die in einem nicht oder nur eingeschränkt kontrollierbaren, **komplexen Umfeld** auftreten. Wenn beispielsweise ein hochautomatisiertes Fahrzeug das Verhalten auf Basis trainierter Kamerabilder erlangt und zu einem großen Teil „alles mal gesehen“ haben muss, dann können systematische Änderungen in der Welt unter Umständen dazu führen, dass der Algorithmus mehr Fehler macht. Wenn also beispielsweise durch Training von Bildern oder Bildsequenzen von Fußgängern ein System entstanden ist, das Fußgänger beispielsweise von Autos oder Schildern unterscheiden kann, dann könnte die

Erkennungsrate wesentlich reduziert sein, wenn alle Fußgänger plötzlich Atemschutzmasken trügen.

VERÄNDERBARKEIT

DIMENSION	Veränderbarkeit beschreibt Änderungen der Eigenschaften bzw. des Verhaltens eines Systems oder seines Umfelds im Betrieb.

UNTERKATEGORIEN

Änderungen des Systems: Änderungen der Eigenschaften bzw. des Verhaltens eines Systems

Änderungen des Umfelds: Änderungen des Umfelds, die wegen Adaptions- und Lernprozessen Einfluss auf das Verhalten der Systeme im Betrieb haben

Änderungen des Systems

DEFINITION	Änderungen von Eigenschaften bzw. von Verhalten software--physischer Systeme im Betrieb.

AUSPRÄGUNGEN

Keine Veränderung nach Auslieferung vom Hersteller

Das System ist darauf angelegt, während der gesamten Lebensdauer unverändert zu funktionieren

Vorgesehene, kontrollierte diskrete Veränderungen nach Auslieferung vom Hersteller

Das System hat variable Elemente, die bei Installation oder im Betrieb durch Hardwareerweiterungen, Spezifizierung von Software oder durch Änderung von Software das Systemverhalten ändern (bspw. Anbau von Greifern bei Robotern, Teaching der Roboter oder Softwareupdates)

Begrenzte Änderung weniger Parameter (Adaptivität)

Das Systemverhalten adaptiert sich entweder während der Inbetriebnahme an vorhersehbare Randbedingungen oder während des Betriebs an Änderungen weniger Bedingungen (bspw. infolge von Temperaturänderungen, Abnutzung, Fabrikationsschwankungen)

Bedeutende Veränderungen im Betrieb, z. B. weiterlernende System

Das Systemverhalten verändert sich fundamental in der Zeit, indem es während des Betriebs Betriebsdaten für die Modifikation des eigenen Verhaltens (in der Regel eine Optimierung) nutzt. Die Unterscheidung zur Adaptivität liegt in der Komplexität der Daten, die Basis der Veränderung sind und in dem Umfang der durch die Daten erzeugten Verhaltensänderungen.

Änderungen des Umfelds

DEFINITION	Änderungen des Umfelds, die wegen Adaptions- und Lernprozessen Einfluss auf das Verhalten technisch-physischer Systemen im Betrieb haben.

AUSPRÄGUNGEN

Vorhersehbare Änderung weniger Größen

In Art und Ausmaß vorhersehbare Änderungen wie Temperaturschwankungen, Abnutzung, Bauteiletoleranzen

Nicht bekannte Änderungen hoher Komplexität

Nicht oder nicht vollständig vorhersehbare Änderungen des Umfelds, die bedeutenden Einfluss auf die Systemfunktionalität haben (bspw. systematische Änderungen des äußeren Erscheinungsbilds von Objekten in der Verkehrsumgebung hochautomatisierter Fahrzeuge, systematische Änderungen in Fabrikhallen wie z.B. Veränderung der Farbe der Weste aller Personen, die Zugang zur Halle haben)

4.2.2 Dimension Vernetzung

Bei unvernetzten, autarken Systemen beruht die Aufgabenausführung, also die Aktionen, allein auf dem, was sie sensorisch erfassen und verarbeiten. Dahingegen können vernetzte Systeme nur in ihrer Vernetztheit die Aufgabe erfüllen. Die Vernetzung eines Systems mit einer oder mehreren Instanzen verändert oder erweitert folglich die Funktionen eines Systems. Die unterschiedlichen Aspekte der Vernetzung sind in der ebenso benannten Dimension beschrieben.

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. Dabei sind „Vernetzungen“ niedrigdimensionaler Sensordaten zur Erfassung von Temperatur, Druck, etc. nicht in Betrachtungen der Dimension „Vernetzung“ einbezogen, selbst wenn sie in digitalisierter Form vorliegen. Inhaltlich abzugrenzen ist der Aspekt der Vernetzung zudem von der Kommunikation bzw. Interaktion mit Beteiligten oder Nutzern zum Austausch von Informationen über Status, Aufgaben oder Intentionen des Systems.

Die **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**. Diese Art der internen Vernetzung hat Einfluss auf den Grad der Kontrollierbarkeit. Eine sternförmige Vernetzungsstruktur mit einer zentralen Steuerungsinstanz erzeugt in der Regel einen höheren Grad der Kontrollierbarkeit als eine dezentrale Vernetzungsstruktur mit Untersystemen relativer Unabhängigkeit. Dezentrale Vernetzungsstrukturen in Verbindung mit KI-Algorithmen zeichnen sich häufig durch einen hohen Grad an Autonomie (siehe Dimension Kontrollierbarkeit) aus.

Aus rechtlicher Sicht sind solche Vernetzungen relevant bei der Feststellung der „Gesamtheit der Maschinen“ (siehe Abschnitt 5.3.2.1.2, S. 136). Von rechtlich besonderer Relevanz sind Datenvernetzungen nach außen. Wenn das Verhalten eines Systems in starkem Maße von Daten abhängen, die von außen geliefert werden, kann die Feststellung der „Gesamtheit der Maschinen“ besonders erschwert sein.

Die **Vernetzung nach außen** beschreibt die Beeinflussung des Systemverhaltens durch digitale Daten außerhalb des direkten Einflussbereichs des Betreibers. Dabei wird zwischen abgesprochenen und unabgesprochenen digitalen Daten unterschieden. **Abgesprochene Daten** sind Daten vordefinierten Art und Umfangs. Diese können sowohl **informativen Charakter** haben als auch **direkt handlungsdeterminierend** sein. Informative Daten können nach erfolgter Informationsverarbeitung auch auf das Systemverhalten wirken. Direkt

handlungsdeterminierende Daten hingegen wirken unmittelbar auf das Systemverhalten, vergleichbar mit der direkten Steuerung einer internen zentralen Vernetzung. Für informative wie für direkt handlungsdeterminierende Daten muss geklärt werden, durch welche Mechanismen die Kontrollierbarkeit der Datenqualität von dritten Datenlieferanten sichergestellt werden. Dies gilt insbesondere für Vernetzungen nach außen, wenn die Daten nicht in vordefinierter Art und Umfang geliefert werden – als Beispiel seien hier Daten über die aktuelle Ampelphase, die an automatisierte Fahrzeuge geliefert werden, genannt –, sondern wenn es sich um **unabgesprochene** Daten handelt, die nicht vordefiniert beispielsweise aus dem Internet abgeleitet sind. Aus rechtlicher Sicht würden dann *„ad hoc-Gesamtheiten von Maschinen“* entstehen, die dann u.U. zu im Betrieb veränderlichen Maschinen führen und für die der Nachweis sicheren Verhaltens sehr spezieller Methoden bedürfte. Ferner sind auch Zwischenformen zwischen abgesprochenen und unabgesprochenen Daten denkbar. Kriterium der Einordnung wäre in diesem Falle das Ausmaß der prinzipiellen Möglichkeiten eines Betreibers, die externen Daten informatorisch filtern zu können.

VERNETZUNG

DEFINITION	Die Dimension „Vernetzung“ beschreibt Eigenschaften von technischen physischen Systemen in Bezug auf interne Prozesse des Austauschs von systembeeinflussenden digitalen Daten und bezüglich des Einflusses systeminternen Verhaltens von äußeren digitalen Daten.

UNTERKATEGORIEN ***Interne Vernetzung***
Vernetzung nach außen

Interne Vernetzung

DEFINITION	Austausch systembeeinflussender digitaler Daten zwischen Subsystemen.

AUSPRÄGUNGEN

Zentrale Vernetzung

Eine zentrale Steuereinheit kontrolliert Subsysteme. Dies stellt (bei gleichen anderen Bedingungen) die unproblematischste Art der Vernetzung dar.

Dezentrale Vernetzung

Mehrere Subsysteme sind ohne eine zentral steuernde Einheit miteinander vernetzt und operieren in gewissem Umfang lokal unabhängig, dies aber im Sinne einer übergeordneten Zielsetzung.

Vernetzung nach außen

DEFINITION	Beeinflussung des Systemverhaltens durch digitale Daten von außerhalb des direkten Einflussbereichs des Betreibers.

AUSPRÄGUNGEN

Abgesprochene Daten

Daten vordefinierter Art Umfangs. Sie können entweder rein informativ oder direkt handlungsdeterminierend sein

Unabgesprochene Daten

Art oder Umfang der Daten sind nicht vordefiniert und u.U. nicht kontrollierbar.

4.2.3 Dimension Kontrollierbarkeit

Bei der Erörterung der Dimensionen „Veränderbarkeit“ und „Vernetzung“ war es der mögliche Verlust der Kontrollierbarkeit des Verhaltens der Systeme, der bei bestimmten Ausprägungen in diesen Dimensionen besondere Maßnahmen des Nachweises der Sicherheit erfordert oder u. U. sogar Änderungen an rechtlichen Rahmenbedingungen nötig macht. Generell bezeichnet Kontrollierbarkeit die Möglichkeit der Beeinflussung bzw. Einflussnahme. Man kann Kontrolle nach Einfluss auf das durch Systemeigenschaften bestimmte Verhalten und nach Einfluss auf von außen kommende Daten differenzieren. Bei unvernetzten Systemen erstreckt sich die Kontrollierbarkeit nur auf innere Systemeigenschaften, bei Vernetzung nach außen auch auf die Kontrolle von Daten, die von Dritten geliefert werden.
Die Dimension **KONTROLLIERBARKEIT** beschreibt die Eigenschaften von Systemen in Bezug auf das Vermögen des Herstellers oder Betreibers, das Systemverhalten determinieren [110] zu können, sowie hinsichtlich sicherheitsgewährleistender Maßnahmen wie eine mögliche Beschränkungen des Systemeinsatzes oder die Kontrolle eines Datenflusses, der von außen kommt. (siehe Kap. 5.7.6).
Die Kontrollierbarkeit eines Systems hängt von dem Grad der **Emergenz** eines Systems ab, d. h. der Befähigung zu Verhalten aus sich 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. Selbstorganisiertes Verhalten resultiert aus der Interaktion sehr vieler Teileinheiten, bei denen nur die Interaktion zu anderen Teileinheiten vorhanden oder vorgegeben ist. Ein Beispiel hierfür sind Verkehrssimulationen mit vielen Tausend Fahrer-Fahrzeug-Einheiten, die auf einem Straßennetz mit wenigen vorgegeben Regeln interagieren. Das Gesamtverkehrsverhalten „ergibt“ sich durch die Interaktionen und bedarf besonderer Maßnahmen, um ein spezielles Verhaltensmuster (z.B. Stau) zu erzielen.
Der Begriff der Autonomie wird selten einheitlich verwendet. Generell sind System bezüglich anderer Systeme autonom, wenn sie bezüglich eines bestimmten Merkmals unabhängig (selbstständig) von ihnen sind (siehe Erläuterungen in Kapitel 2.1. In der vorliegenden Taxonomie beschreibt **Autonomie** eine Eigenschaft im Systemverhalten, die sich aus der Vorgabe von Gütekriterien, Zielen oder Zielhierarchien ergibt. Die Unabhängigkeit bezieht sich in diesem Gebrauch des Begriffes auf die Unabhängigkeit des Systemverhaltens von konkreten Verhaltensvorgaben der Entwickler. Zur Veranschaulichung kann die Pferdmetapher (Flemisch et al. 2005) [111] dienen. Das Pferd reitet dorthin, wohin der Reiter es lenkt,

[110] determinieren meint hier „exakt bestimmen in allen Einzelschritten bzw. kausalen Abfolgen“
[111] FLEMISCH, F.O., SCHOMERUS, J., KELSCH, J., & SCHMUNTZSCH, U. (2005). Vom Fahrer zum Reiter? Zwischenbericht 2005 auf dem Weg von der H-Metapher zum H-Mode für Bodenfahrzeuge. In *VDI-Berichte. Fahrer im 21. Jahrhundert* (*Vol. 1919, S*. 63-74). Braunschweig: VDI-Verlag GmbH.

bewegt aber nach eigenem Plan die Beine und vermeidet selbstständig ein Stolpern über Äste. Die Selbstständigkeit steigt mit der Abstraktheit der Zielvorgaben bzw. der Tiefe der Zielhierarchie. Der Besitzer eines Pferdes hat das Ziel, damit Geld zu verdienen und erreicht dies dadurch, dass er jemanden einstellt, der das Pferd dann über den Parkour leitet. Generell sind viele Stufen zunehmender Abstraktheit in den Zielvorgaben denkbar. Die Autonomie steigt in der Regel mit der Komplexität des lernenden Systems, auch wenn keine Zielhierarchie vorliegt.
Die Kontrollierbarkeit gelingt in hohem Maße, wenn die Software zur Steuerung der physischen Systeme durch explizite Sequenzen mit genau bekannten Handlungsfolgen realisiert ist. Je geringer die Emergenz der Software ist, je weniger das Verhalten also nur implizit über die Vorgabe von Gütekriterien oder Zielen oder noch basaleren Mechanismen wie Selbstorganisation oder spontaner Musterbildung bestimmt ist, desto besser ist sie von Entwickler, Hersteller oder Betreiber kontrollierbar. Wie oben bei der Beschreibung der Dimension „Veränderbarkeit“ geschildert, zeichnen sich gerade Systeme auf Basis maschineller Lernverfahren und im Besonderen die im Betrieb weiterlernenden System durch ein Defizit an Kontrollierbarkeit aus und bedürfen deshalb besonderer Vorgehensweisen zur Gewährleistung der Sicherheit.
Die Facette **Beschränkungen** 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 zum einen **Systembeschränkungen durch technologische Maßnahmen**, deren Ziel es ist, schlecht kontrollierbare Teilsysteme zu überwachen. Dies kann beispielsweise mittels Kapselungen durch andere etablierte, explizit agierende Software geschehen, oder aber durch konventionelle Sicherheitssysteme wie Schranken, Zäune, Laserscanner, etc. Die Systeme werden zwar in ihrem Einsatzbereich eingeschränkt, dafür ist der Sicherheitsnachweis aber wesentlich leichter, da immer eine sichere Notfallebene vorhanden ist. Bei fahrerlosen Transportsystemen wird die intrinsische Sicherheit durch zertifizierte Sicherungssysteme gewährleistet, die ein Abschalten des Systems bei einer bestimmten Nähe zu einem Objekt in Fahrtrichtung auslösen. Andere Maßnahmen, um die Kontrollierbarkeit zu erhöhen, bestehen in **Beschränkungen des Umfelds oder des Einsatzbereichs** des Systems. Ein Beispiel ist die Beschränkung hochautomatisierter Fahrzeuge auf Autobahnen oder auf Maximal- bzw. Minimalgeschwindigkeiten. Ein anderes Beispiel ist eine Beschränkung auf Anwendungen oder Situationen, die nicht sicherheitsrelevant sind, in denen also keine Gefahren für Leib und Gut entstehen können. Als Beispiel sei die Adaption einer graphischen Nutzungsschnittstelle an Nutzergruppen über lernende KI-Algorithmen oder der Einsatz von mustererkennenden KI in Maintenance-Anwendungen genannt. **Beschränkungen und Kontrolle im Datenfluss** sind Maßnahmen, die bei einem Datenfluss von außen die Sicherheit des Systemverhalten gewährleisten.

KONTROLLIERBARKEIT

DEFINITION	Eigenschaften von Systemen in Bezug auf das Vermögen des Herstellers oder Betreibers, das Systemverhalten determinieren zu können, sowie Maßnahmen, um die Beherrschbarkeit sicherstellen zu können.

UNTERKATEGORIEN ***Emergenz***
Beschränkungen

Emergenz

DEFINITION	Verhalten aus sich heraus, nicht durch eine externe Instanz vorgegeben.

AUSPRÄGUNGEN

Autonomie

Durch die Vorgabe von Gütekriterien, Zielen oder Zielhierarchien induzierte Eigenschaft der relativen Unabhängigkeit im Systemverhalten

Selbstorganisation

Erzeugung von Ordnung und Verhalten von Systemen aus dem nicht orchestrierten (von außen oder zentral) verändernden Wechselspiel der Komponenten des Systems.

Beschränkungen

DEFINITION	Überwachen des externen Datenzugangs sowie schwer kontrollierbarer Teilsysteme oder Beschneidung deren Einsatzbereichs oder deren Funktionsumfangs.

AUSPRÄGUNGEN

Systembeschränkungen

Begrenzung oder Überprüfung eines Systems (typisch eines KI-Systems) durch begleitenden Einsatz konventioneller und überprüfbarer Software oder durch Hardwarekompnenten

Beschränkungen im Umfeld und im Einsatzbereich

Beschränkung des Gesamtsystems auf spezielle Umgebungen, wie z.B. Fabrikhallen mit bestimmter Ausstattung. Beschränkung auf bestimmte Einsatzbereich, wie z.B. Autobahnfahrt, Fabrikhallen mit definiertem Erscheinungsbild, Geometrien oder Funktionselemente (Spiegel, Trassen, Induktionsschleifen, etc.).

Beschränkungen und Kontrolle im Datenfluss

Kontroller externer Daten bei Vernetzung nach außen. Kontrolle der Datenqualität, Beschränkungen im Datenstrom

4.2.4 Dimension Transparenz

Transparenz als Begriff wird in sehr unterschiedlichen Kontexten als Metapher der physikalischen Durchsichtigkeit, also im Sinne des Durchschauens einer Hülle, der Möglichkeit des „Reinschauens“ oder im übertragenen Sinne des Durchschauens oder Verstehens von Verborgenem eingesetzt. Als Metapher auf Software übertragen

gedacht, wäre dann beispielsweise die Erhöhung der Transparenz eines KI-Systems mit Black-Box-Charakter durch Erhöhung der Einsicht in das System realisiert. Erhöhung der Transparenz hätte demnach den Charakter einer Offenlegung bzw. Verdeutlichung interner Mechanismen. Umgekehrt ist mit der Erhöhung der Transparenz ein Verbergen „unnötiger“ Details verbunden. Hier greift die Metapher des Durchscheinens [112]. Offenlegung hängt sowohl von den Systemeigenschaften als auch von demjenigen ab, dem offengelegt oder verdeutlicht wird. Eine Offenlegung hat mit der Komplexität oder Strukturiertheit des „Verborgenen“ zu tun, hängt aber auch von dem Wissen und Informationsbedarf des Rezipienten ab. In der öffentlichen Diskussion, in der eine „transparente“ KI gefordert wird, ist es der „normale“ Bürger, dem die „für ihn wichtigen“ Mechanismen des KI-Algorithmus – beispielsweise welche seiner persönlichen Daten für die Entscheidungsfindung der Kreditvergabe herangezogen werden – deutlich gemacht werden sollen. An diesem Beispiel wird zum einen deutlich, dass Transparenz immer in Bezug auf einen Empfänger gedacht werden muss, und zum anderen, dass Transparenz nicht absolut, sondern anforderungsbezogen zu betrachten ist.

Im Rahmen der Taxonomie wird daher die Dimension **TRANSPARENZ** als Verfügbarkeit und Verständlichkeit des Systems bzw. seines Verhaltens aus zwei Perspektiven definiert. 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. Die Transparenz wird folglich nicht isoliert als mehr oder minder gegebene Systemeigenschaft aufgefasst, wie es die in diesem Kontext verbreitete Bezeichnung der Black-Box-Systeme suggeriert, sondern in Bezug auf denjenigen betrachtet, der diese relevanten Informationen benötigt. Anders als bei der Expertensicht spielt bei der Transparenz aus Beteiligtensicht folglich der durch Ausbildung und Schulungsmaßnahmen beeinflussbare Kenntnisstand ebenfalls eine gesondert bedeutsame Rolle. Die bedarfsgemäße Festlegung des zu erreichenden Transparenzgrades, d.h. bezüglich welcher Aspekte ein bestimmtes Maß an Transparenz bestehen muss, bezieht sich im vorliegenden Fall nur auf die Aspekte, die unmittelbar oder mittelbar im Bezug zur Sicherheit stehen. Nicht adressiert wird beispielsweise Transparenz bezüglich der Verwendung personenbezogener Daten.

Im Rahmen der Taxonomie werden für beide Perspektiven (Experten und Beteiligte) verschiedene Facetten der Transparenz beschrieben, die ein zunehmend tieferes Verständnis des Systems widerspiegeln. Bei der Transparenz aus **Expertensicht** handelt es sich dabei um die Spezifizierbarkeit, die Beschreibbarkeit der Funktionsgrenzen,die Nachvollziehbarkeit und die Vorhersehbarkeit des Systemverhaltens. Die Kategorie **Spezifizierbarkeit** bezieht sich auf das Ausmaß, in dem Funktionalität, Einsatzbedingungen, Umgebungund auch Anwendungszenarien

[112] Die Forderung nach Verbergen von Einzelheiten, also höherer Transparenz wird z.B. im Zusammenhang mit Computer- und Netzwerktechnik benutzt, wenn der Nutzer die äußeren Schichten beim Zugriff auf Ressourcen nicht explizit angeben muss.

festgelegt werden können. Für komplizierte Systeme gilt auch heute schon, dass diese nicht vollständig spezifizierbar sind. Dennoch kann über bewährte Prozesse der Überprüfung, ob Spezifikationen eingehalten wurden, das Sicherheitsniveau der Systeme nachgewiesen werden. Die Unsicherheit besteht in einer Abweichung von der Spezifikation durch das Auftreten möglicher Fehler.
Bei sehr komplexen Systemen wie beispielsweise hochautomatisierten Fahrzeuge in einer beliebigen Umwelt, bei denen wegen der Komplexität in der Regel KI-Algorithmen eingesetzt werden, ist die Spezifizierbarkeit im konventionellen Sinne der industriellen Praxis entscheidend reduziert. Damit kann aber auch das Mittel des Sicherheitsnachweises über den Prozess der Überprüfung der Einhaltung der Spezifikationen nicht mehr aufrecht erhalten bleiben.
Aus diesem Grund wurde für den Bereich der Automobiltechnik das Konzept der SOTIF (Safety of the Intended Functionality) und ein dazugehöriger Standard in Form der ISO 21448 entwickelt. Bei weniger genau spezifizierbaren Systemen besteht die Notwendigkeit, die zusätzlichen, nicht durch die Spezifikation abgedeckte Risiken zu berücksichtigen. Die Festlegung der gewünschten Funktionalität entspricht einer Spezifikation auf einem höheren Abstraktionslevel. Dementsprechend werden auch die Fehler auf einem höheren Abstraktionslevel beschrieben. Für den Sicherheitsnachweis bedeutet dies eine Notwendigkeit der Festlegung und Hinzunahme weiterer, über den bisherigen Standard hinausgehenden Methoden bei der Festlegung von Vorgehensstandards der Entwicklung und der Prüfung, die als sicherheitsgewährend angesehen werden.
Bei hochkomplexen Aufgaben und dementsprechend komplexen Systemen zur Bewältigung dieser Aufgaben ist eine vollständige Spezifizierung und damit eine vollständige Testung des Systems nicht mehr möglich. In Verbindung mit der Beschreibbarkeit des Systemverhaltens steht die **Beschreibbarkeit der Funktionsgrenzen** im Sinne der Bedingungen, unter denen das System nicht mehr wie vorgesehen agieren kann. Diese stellt unter anderem die Voraussetzung für die Ergreifung ergänzender Sicherungsmaßnahmen dar, beispielsweise durch Beschränkung des Einsatzbereichs (vgl. Beschränkungen der Dimension Kontrollierbarkeit, Abschnitt 4.2.3). Über diese beiden Aspekte hinaus geht die **Nachvollziehbarkeit** des Systemverhaltens durch diejenigen, die aufgrund ihrer Expertise befähigt sind, den Prozess von der Informationsaufnahme des Systems bis zur Handlungsausführung nachzuverfolgen. Diese Nachvollziehbarkeit wird entscheidend von der Komplexität des Systems, seiner Aufgaben und des Umfelds mit beeinflusst. Ein spezifizierbares System gilt als vollständig nachvollziehbar, während nicht spezifizierbare Systeme in unterschiedlichem Maße nachvollziehbar sind. Dabei wird die Nachvollziehbarkeit qualitativ verändert, sobald ein System auf emergenten Algorithmen fußt.
Während die Nachvollziehbarkeit das nachträgliche Erklären von Verhalten, also den retrospektiven Aspekt umreißt, wird mit dem Aspekt der **Vorhersehbarkeit**, das Vermögen durch Anayse, Simulation, Berechnungen oder anderen Methoden, das Verhalten von Systemen prospektiv vorhersagen zu können. So verlangen beispielsweise Standards für die funktionale Sicherheit in der Regel eine statische

Codeanalyse und empfehlen, den Code einer Laufzeitanalyse (oder dynamischen Analyse) zu unterziehen[113]. Die Verfahren sind nicht sinnvoll auf künstliche neuronale Netze übertragbar. Hier fehlen es z. Z. noch akzeptierte Methoden. Eine alternative ist die Analyse über Simulationen, um ein Verständnis über das zukünftige Systemverhalten im realen Einsatz zu bekommen.
Die Transparenz aus Expertensicht kommt im Zuge einer kritischen Beurteilung der Sicherheit eines Systems zu tragen und nicht im (un)mittelbaren Umgang mit dem System während des Betriebs. Ein weiterer Aspekt der Transparenz ergibt sich jedoch, wenn der Mensch mit software-physischen Systemen interagiert bzw. diese bedient. Wenn Mensch und Maschine gemeinsam die Sicherheit determinieren, wird diese auch von dem Wissen des Interagierenden um die Funktionalität im System geprägt. Das Ausmaß des als erforderlich anzusehenden Wissens bestimmt sich gemäß den Eigenschaften des Systems und der Rolle des Menschen, die dieser im Prozess der Aufgabenbewältigung und dem System gegenüber einnimmt. Für die **Beteiligten** muss eine grundlegende **Kenntnis des Einsatzbereichs und der Grenzen des Systems** immer dann gegeben sein, wenn das Fehlen ebensolcher Wissensbestände sicherheitskritisch werden kann. Als Beispiel wäre ein fahrerloses Transportsystem zu nennen, das in einer Fabrikumgebung nur gefahrlos operiert, wenn alle Menschen in der Arbeitsumgebung eine gelbe Weste tragen. Unkenntnis bezüglich des Unvermögens des Systems, Andersgekleidete akkurat zu erkennen, kann zu falschen Annahmen bezüglich des Systemverhaltens und in der Konsequenz zu Unfällen führen.
Handelt es sich bei dem software-physischen System um ein System, das – wie ein Roboter – beliebige Bewegungen im Raum ausführen kann, kommt zu der durch ein logisches Gerüst von Handlungsbausteinen beschreibbaren Verhaltenslogik, noch die Bewegungstrajektorie in Raum und Zeit dazu. Deren Vorhersehbarkeit durch einen in der Nähe befindlichen Menschen beeinflusst ebenfalls die Sicherheit. Nicht erwartbare Bewegungsbahnen, bahnbezogene Geschwindigkeiten oder auch Achsendrehungen können zu Fehlern im Situationsbewusstsein und dadurch zu Sicherheitsproblemen führen. Dies macht den Aspekt der **Vorhersehbarkeit der Dynamik von Systemhandlungen** bedeutsam.
Fehler (Unfälle) entstehen hier nicht durch eine Fehlbenutzung, sondern durch einseitige oder zweiseitige Interaktionsfehler aufgrund falscher Prädiktion von System- bzw. Menschverhalten. Dies ist insbesondere in Szenarien der Mensch-Roboter-Kollaboration oder -Interaktion relevant. Analog zur in der Sicherheitstechnik fest verankerten[114] *Fehlbenutzung wegen Manipulationsanreizen außerhalb der bestimmungsmäßigen Verwendung* (siehe dazu Dimension „Involviertheit des Menschen“, Kap. 4.2.6) kann eine *Fehlinteraktion außerhalb der geplanten Interaktion* definiert werden.

---

[113] Z.B. Standard IEC 61508 *Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme*

[114] TRBS – 1151 Technische Regel für Betriebssicherheit – Gefährdungen an der Schnittstelle Mensch – Arbeitsmittel – Ergonomische und menschliche Faktoren, Arbeitssystem (TRBS 1151), GMBl. Nr. 17/18 Ausgabe: März 2015.

Es gilt anzumerken, dass der Aspekt der Vorhersehbarkeit gleichermaßen auch für nicht mobile Systeme gilt, bei denen nicht die bevorstehende Bewegung im Raum, sondern vielmehr der **nächste Prozesszustand** des Systems bedeutsam ist.
Über die grundlegende Kenntnis sowie die Vorhersehbarkeit des bevorstehenden Systemschritts hinaus geht der Aspekt eines tiefen **Verständnisses der Systemfunktionalität**. Ein derart elaborierteres mentales Modell des Menschen vom System ermöglicht es, die Verhaltenslogik des Systems auch in selten auftretenden Situationen nachzuvollziehen und angemessen damit umgehen zu können. Diese Anforderung an nutzerseitige Transparenz sollte insbesondere in hochgradig komplexen und sicherheitskritischen Aufgabenwelten gegeben sein.

TRANSPARENZ

DEFINITION	Zugänglichkeit, Erklärbarkeit, Nachvollziehbarkeit und Vorhersehbarkeit aller relevanten Informationen über Eigenschaften und Verhalten eines Systems für die relevanten Empfänger.

UNTERKATEGORIEN ***Expertensicht***
Sicht Beteiligte

Expertensicht

DEFINITION	Zugänglichkeit aller relevanten Informationen über Eigenschaften und Verhalten von Systemen für diejenigen, die mit Entwicklung und Prüfung eines Systems betraut sind.

AUSPRÄGUNGEN

Spezifizierbarkeit

Spezifizierbarkeit bezieht sich auf den Grad und das Abstraktionsniveau der Möglichkeit, Funktionen, Einsatzbereich, Umgebung, etc.für Systeme festlegen zu können.

Beschreibbarkeit der Funktionsgrenzen

Beschreibbarkeit der Bedingungen, unter denen das System nicht mehr wie vorgesehen agieren kann

Nachvollziehbarkeit

Maß für die Möglichkeit, Gründe für Systemverhalten geben zu können. Ein Algorithmus ist nachvollziehbar, wenn ein Mensch die Frage beantworten kann „Warum verhält sich das System so?“. Wegen des expliziten Bezugs auf einen menschlichen Nachvollziehenden hängt das Maß der Nachvollziehbarkeit von dem Maß der Verstehbarkeit ab, ist also an Wissen und Fähigkeiten des Analysierenden gebunden.

Vorhersehbarkeit

Maß für die Möglichkeit, Systemverhalten durch Analyse, Modellierung oder Simulation vorhersagen zu können

Sicht Beteiligter

DEFINITION	Zugänglichkeit aller relevanten Informationen über Eigenschaften und Verhalten des Systems für diejenigen, die unmittelbar mit dem System interagieren oder mittelbar von ihm betroffen sind.

AUSPRÄGUNGEN

Kenntnis des Einsatzbereichs und der Grenzen des Systems

Grundlegende Kenntnis des Verhaltens eines Systems und der Situationen, in denen selbiges an seine Grenzen stößt.

Vorhersehbarkeit der Dynamik von Systemhandlungen und des nächsten Prozesszustands des Systems

Vorhersehbarkeit des Systemverhaltens oder Prozesszustands in der unmittelbaren Zukunft für den bzw. die mit dem System interagierenden Menschen.

Verständnis der Systemfunktionalität

Wissen von Bedienern über die Verhaltenslogik von Systemen auch in selten auftretenden Sondersituationen

4.2.5 Dimension Widerstandsfähigkeit

In der Taxonomie beschreibt die Dimension **WIDERSTANDSFÄHIGKEIT** 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. Der Begriff der Robustheit spielt in der Diskussion von Systemen, die Komponenten basierend auf maschinellen Lernverfahren enthalten, eine starke Rolle. Häufig geht es dabei um die Aufgabe der Klassifikation auf der Grundlage von Bildsequenzen. Bei heutigen Verfahren können selbst geringfügige und für den Menschen unmerkliche oder vermeintlich irrelevante Änderungen im Datensatz zu Fehlklassifikationen führen. In der Forschung werden gegenwärtig Methoden entwickelt, um die Systeme gegenüber unterschiedlichen Bedingungen robuster, d.h. fehlerärmer zu gestalten. Als Beispiel seien Methoden genannt, eine sichere Objekterkennung auch bei verrauschten Bildern oder schwachen Kontrasten zu leisten. Unter die Thematik der Robustheit fällt aber auch die Fähigkeit, mit Unbekanntem fertig zu werden. Auf die Anwendung fahrerloser Transportsysteme bezogen, zählen z.B. auch unbekannte Objekte dazu, die neu in der Fahrumgebung auftreten.

Der Begriff Robustheit wird auch in anderen Disziplinen mit der ähnlichen Konnotation der Fehlertoleranz angewandt. So werden unter „robusten Regelsystemen“ beispielsweise Regelsysteme verstanden [115], die auch dann in genügendem Maße operieren, wenn keine ausreichende Kenntnis zur Anpassung des Reglers an das zu regelnde System vorhanden ist. Auch in der Statistik gibt es sogenannte robuste Testverfahren, die robust (im Sinne von wenig anfällig) auf Ausreißer reagieren [116].

Entsprechend dieses Robustheitsbegriffs beschreibt die Unterkategorie Robustheit die Widerstandsfähigkeit eines Systems gegenüber sowohl geringfügigen als auch größeren Änderungen im Einsatzumfeld. Es wird folglich die Fähigkeit des Systems adressiert, bei zumindest nicht im Detail vorhergesehenen inneren oder von außen kommenden Einflüssen dennoch ohne etwaige sicherheitswirksame Fehler zu agieren.

---

[115] J. Ackermann: *Robuste Regelung*. Springer-Verlag, 1993., Abschnitt 11.4

[116] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (2011). *Robust statistics: the approach based on influence functions* (Vol. 196). John Wiley & Sons..

Dazu gehören nicht nur Schwankungen in den Sensordaten, sondern z. B. auch in menschlicher Bedienung.

Robustheit ist in diesem Sinne synonym zu „abfangbar“ zu verstehen. Binnen dieser Unterkategorie wird nach einem Abfangen äußerer Einflüsse (**Security**) und dem Abfangen innerer Abweichungen unterschieden. Bei dem Abfangen innerer Abweichungen wird wiederum zwischen der Art der abfangbaren Abweichungen von bekannten und schon bewältigten Situationen unterschieden. Werden kleine Änderungen im Input oder den Rahmenbedingungen bewältigt – wie beispielsweise eine korrekte Klassifikation von Bildern trotz Rauschen, geänderten Lichtverhältnissen, o.ä. – so wird dies in der Taxonomie als **Stabilität** bezeichnet. Ergänzend dazu bedarf ein robustes System der Fähigkeit, mit Unbekanntem fertig zu werden. Bei hochautomatisiertem Fahren sind dies beispielsweise ungelernte Objekte, die neu in der Welt erscheinen. Diese Fähigkeit zur **Bewältigung unbekannter Situationen bzw. unvorhergesehener Ereignisse** setzt voraus, die Situation als unbekannt zu erkennen und damit ein angemessenes und sicheres Systemverhalten in der gegebenen Situation zu ermöglichen; beispielsweise über ein eigeninitiiertes Abschalten, eine informierte Rückmeldung bzw. Rückübergabe an einen menschlichen Bediener oder ein konventionelles System (vgl. Beschränkungen im Rahmen der Dimension Kontrollierbarkeit, s. Abschnitt 4.2.3). Die Unterscheidung zwischen der Bewältigbarkeit von kleinen Änderungen vs. unbekannten Situationen spiegelt keinen quantitativen Unterschied wider, da kleine Änderungen nicht schlicht leichter bewältigbar sind als gänzlich unbekannte Situationen – wie es für einen Menschen wohl der Fall wäre –, sondern eine qualitativ anders beschaffene Anforderung an ein System darstellen. Folglich bedeutet eine Bewältigbarkeit unbekannter Situationen nicht zwangsläufig, dass das System hohe Stabilität im obigen Sinne zeigt.

Als dritter Aspekt ist unter der Robustheit noch die **Security** hinzugenommen, soweit dies Einfluss auf die Safety hat. Die Security umfasst in diesem Zusammenhang die Fähigkeit des Systems, Angriffen von außen zu widerstehen bzw. trotz Angriffen fehlerfrei (oder fehlerminimiert) zu agieren.

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. Resilienz bezieht sich damit auf die Widerstandsfähigkeit gegenüber den Fehlerfolgen nach einem mehr oder weniger punktuellen Ereignis einer potentiell schädigenden Fehlfunktion (die oben beschriebene „Robustheit“ bezieht sich hingegen auf Systemeigenschaften vor dem Ereignis). Resilienz ist in diesem Sinne synonym zu „linderbar“ oder „heilbar“ aufzufassen.

Der Begriff Resilienz wird häufig auch in Zusammenhang mit der Sicherheitsforschung bei der Diskussion des Katastrophenschutzes verwendet [117]. Allerdings ist die

[117] Z. B. Fekete, A., Grinda, C., & Norf, C. (2016). Resilienz in der Risiko-und Katastrophenforschung: Perspektiven für disziplinübergreifende Arbeitsfelder. In *Multidisziplinäre Perspektiven der Resilienzforschung* (S. 215-231). Springer, Wiesbaden.

Verwendung des Begriffs dort sehr uneinheitlich. Fekete et al. (ebd.) beklagen die insbesondere in Deutschland sehr uneinheitliche Verwendung des Begriffs in Zusammenhang mit Naturkatastrophen, Terrorismus, Cyber-Angriffen oder Unfällen. Der hier vertretenen Auffassung, dass Resilienz sich auf die Folgen nach einem bedeutenden, widrigen Ereignis bezieht, kommt die Bemerkung von Scharte et al. 2016 [118] nahe, wenn die Autoren schreiben: „*Resiliente Gesellschaften zeichnen sich dahingehend aus, dass sie die Schäden widriger Ereignisse soweit möglich minimieren".*

Maßnahmen zur Erhöhung der Resilienz wirken einerseits unmittelbar folgenmindernd beim Auftreten des Ereignisses oder bezogen auf einen Zeitraum zur Widerherstellung und Erholung oder Minimierung von negativen Langzeitfolgen. Häufig sind Maßnahmen zur Erhöhung der Resilienz passiver Natur. So sind beispielsweise in der Kraftfahrzeugtechnik bautechnische Maßnahmen wie „Knautschzone", Air-Bag oder Sicherheitsgurt folgenmindernd und sind als „passive Sicherheit" der Resilienz zuzuordnen. In der kollaborativen Robotik sind es Maßnahmen zur Erhöhung der Flexibilität in den Gelenken, weiche Hüllen oder dauerhaft wirksame Kraft- oder Leistungsbegrenzungen, die die Resilienz erhöhen. Auch der Einbau von Sicherheitsbauteilen [119] wie etwa Schutzzäunen oder NOT-HALT-Befehlsgeräten stellt eine Maßnahme zur Erhöhung der Resilienz dar.

In der Taxonomie werden bei der im Begriff „Resilienz" abgebildeten Befähigung zur Abwendung der Folgen fehlerhaften Systemverhaltens zwei Aspekte unterschieden. Dabei handelt es sich zum einen um die **passive Wirkungsbegrenzung nach Fehlern** zum Abfangen unerwünschter Folgen von Systemfehlern**.** Hierbei kann die Schadensbegrenzung bzw. -vermeidung in wirkungsbegrenzenden Maßnahmen wie in den obigen Beispielen am System oder in Schutzmöglichkeiten des Umfelds bestehen.

Demgegenüber steht die **aktive Wirkungsminderung nach Fehlern,** bei der sich die Schadensminderung oder -vermeidung aus dem Verhalten des betrachteten Systems ergibt. Als Beispiel sei der Ausfall eines von drei redundanten Steuerrechnern genannt. Die internen Mechanismen wirken „heilend" und verhindern sicherheitswirkende Fehler. Zu dieser Ausprägung der Unterkategorie Resilienz sind z.B. auch das aktive Überführen in einen sicheren Zustand bei einem selbstfahrenden Fahrzeug oder das aktive Ausweichen eines Roboterarms vor einem sich nähernden Menschen zu nennen, der z. B. stolpert (Fehler des Menschen).

Eine exakte Trennung beider Arten Wirkungsminderung verschwimmt in Fällen, in denen aktive Maßnahmen die passive Wirkungsbegrenzung erhöhen. Als Beispiel seien Sicherheitsassistenzfunktionen genannt, die die passive Sicherheitswirkung nach Fehler aktiv modifizieren. Ein System, dass die Steifheit einer pneumatischen elastischen Schutzhülle eines Roboterarms an den berechneten Impuls einer (nicht vermeidbaren) Kollision adaptiert, würde dazu zu zählen sein.

---

[118] Scharte, B., & Thoma, K. (2016). Resilienz–Ingenieurwissenschaftliche Perspektive. In *Multidisziplinäre Perspektiven der Resilienzforschung* (pp. 123-150). Springer, Wiesbaden.

[119] Maschinenrichtlinie 2006/42/EG, Artikel 2 c

WIDERSTANDSFÄHIGKEIT

DEFINITION	Vermögen von Systemen, auch bei inneren oder äußeren Störungen fehlerfrei zu agieren und bei eintretenden, potentiell sicherheitswirksamen Fehlern oder anderen unerwünschten Ereignissen die Folgen zu minimieren.

UNTERKATEGORIEN ***Robustheit***
Resilienz

Robustheit

DEFINITION	Widerstandsfähigkeit gegenüber unbekannten, untypischen oder bekannten, aber ungewollten Gegebenheiten und Einflüssen.

Robustheit ist synonym zu „abfangbar“ zu verstehen

AUSPRÄGUNGEN

Stabilität bei kleinen Änderungen des Inputs

Ausmaß der Änderungen im Systemverhalten bei kleinen Änderungen von Input oder Rahmenbedingungen

Bewältigung unbekannter Situationen bzw. unvorhergesehener Ereignisse

Fähigkeit mit Unbekanntem fertig zu werden, z.B. ungelernte Objekte, die neu in der Welt erscheinen

Security

Fähigkeit eines Systems, Angriffen von außen zu widerstehen bzw. trotz Angriffen fehlerfrei (oder fehlerminimiert) zu agieren.

Resilienz

DEFINITION	Widerstandsfähigkeit gegenüber den Fehlerfolgen nach einem mehr oder weniger punktuellen Ereignis einer potentiell schädigenden Fehlfunktion.

Resilienz ist synonym zu „linderbar“ oder „heilbar“ zu verstehen

AUSPRÄGUNGEN

passive Wirkungsbegrenzung nach Fehlern

Schadensbegrenzung bzw. -vermeidung durch wirkungsbegrenzende Maßnahmen am System oder Schutzmöglichkeiten des Umfelds wie beispielsweise Sicherheitsbauteile.

aktive Wirkungsminderung nach Fehlern

Aktives systemimmanentes Verhalten, das nach Auftreten eines Fehlers sicherheitswirksame Folgen reduziert und/oder verhindert.

4.2.6 Dimension Involviertheit des Menschen

Der Mensch, der als Bediener eines Kraftfahrzeugs oder als mit einem Roboter kollaborierender Mitarbeiter in dynamische Prozesse eingebunden ist, beeinflusst damit die Sicherheit des Gesamtsystems, dessen unverzichtbarer funktionaler Teil er ist. Folglich kann eine Beurteilung der Sicherheit von Systemen 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 wird unterschieden zwischen seinen sicherheitsbeeinflussenden Möglichkeiten im Umgang mit dem System, etwaigen Abweichungen von der bestimmungsgemäßen Verwendung desselbigen sowie der Intention, die solchen Abweichungen zugrunde liegt. Blickt man auf den Menschen als Gefährdetem, der etwaigen unerwünschten Folgen nicht intendierten Systemverhaltens ausgesetzt ist, so werden in Abhängigkeit von der Rolle des Menschen – verbunden mit einem systemspezifischen Expertisegrad und etwaigen Schutzmöglichkeiten seines Umfelds im Umgang mit dem System – unterschiedliche Klassen von Gefährdeten unterschieden.

Der Mensch als **Handelnder** kann sowohl sicherheitsmindernd als auch sicherheitserhöhend wirken. In der Probabilistischen Sicherheitsanalyse (PSA) werden zur prospektiven Sicherheitsbeurteilung komplexer Systeme mit hohem Gefährdungspotential folglich auch Fehlerwahrscheinlichkeiten menschlicher Bediener berücksichtigt [120]. In diesen Analysen (Human Reliability Analysis, HRA) wird von diskreten Aufgaben ausgegangen, die vom Menschen entweder fehlerfrei oder fehlerbehaftet ausgeführt werden. Der Mensch ist in diesen Betrachtungen Funktionselement eines Mensch-Maschine-Systems, wenn er zum Betätigen und Überwachen von technischen Erzeugnissen herangezogen wird (ebd.). Ähnlich wie Ausfallraten von Maschinen können Fehlerwahrscheinlichkeiten, sogenannte HEPs [121] für bestimmte Teilaufgaben bestimmt werden [122]. Dieses Vorgehen ist aber nur möglich, wenn die Aufgaben des Menschen genau vorgegeben werden. Hier entstehen bei der Sicherheitsbewertung menschlicher Handlungen mit dessen natürlicher Intelligenz, eingebettet in ein maschinell oder technisch geprägtes Umfeld, dieselben Schwierigkeiten, die auch bei der Bewertung von Systemen der künstlichen Intelligenz entstehen:

> *Einen Fehler als Abweichung von einer vorgeschriebenen Handlungsausführung zu definieren, ist in Mensch-Maschine-Systemen nur für Handlungen in klar strukturierten technischen Systemen nützlich, also solchen, in denen eine korrekte Handlungssequenz als normative Arbeitsanweisung definiert werden kann. Je komplexer und dynamischer ein System ist, desto schwieriger wird es, Handlungssequenzen eindeutig festzulegen [123].*

---

[120] Giesa, H. G., & Timpe, K. P. (2002). Technisches Versagen und menschliche Zuverlässigkeit: Bewertung der Zuverlässigkeit in Mensch-Maschine-Systemen. *Mensch-Maschine-Systemtechnik*, 63-106.

[121] Human Error Probability

[122] Z.B. Noroozi, A., Khan, F., MacKinnon, S., Amyotte, P., & Deacon, T. (2014). Determination of human error probabilities in maintenance procedures of a pump. *Process Safety and Environmental Protection*, *92*(2), 131-141.

[123] Giesa, H. G., & Timpe, K. P. (2002). Technisches Versagen und menschliche Zuverlässigkeit: Bewertung der Zuverlässigkeit in Mensch-Maschine-Systemen. *Mensch-Maschine-Systemtechnik*, 63-106.

Zu den Stärken[124] des Menschen gehört es, auf wechselnde und unbekannte Bedingungen reagieren zu können. Menschliche Fähigkeit zur kreativen Adaption kann also sicherheitserhöhend wirken – ein bisher im Zusammenhang mit funktionaler Sicherheit nur wenig betrachteter Fakt, schlicht weil ihr Einfluss quantitativ schwer zu fassen ist. In der vorliegenden Taxonomie ist dieser Einfluss durch die Berücksichtigung des Menschen als „**Teil der Prozesskette**“ [125] einbezogen. Sicherheitsgewährleistend ist der Mensch, wenn er in Situationen die Sicherheit dadurch sicherstellt, dass er Aufgaben der ausgefallenen oder überforderten Technik übernimmt oder korrigierend eingreift. Sicherheitserhöhend operiert ein Mensch, wenn er in Sondersituationen seine Fähigkeiten als universeller Problemlöser mit Einfallsreichtum korrekt nutzt. Dies ist insbesondere dann wichtig, wenn es zu sicherheitskritischen Situationen kommt, die in der Entwicklung nicht vorhergesehen wurden und für die es deshalb keine technischen Lösungen gibt. Mit „Weltwissen“ und Erfahrung ausgestattet kann der Mensch u. U. schwerwiegendes Systemversagen verhindern oder zumindest die negativen Folgen von nicht erwünschten Systemzuständen verhindern oder mindern. Eine falsche Auslegung der Anforderungen an den Menschen hingegen kann dazu führen, dass die Sicherheit des Gesamtsystems nicht gewährleistet ist. Das wäre beispielsweise der Fall, wenn der Mensch als letzte Instanz bei einer Überforderung der Automatisierung gefordert würde, die Übergabe aber so gestaltet ist, dass auch der Mensch überfordert wird (wie beispielsweise beim hochautomatisierten Fahren, in denen die vom Fahrzeug übernommene Aufgabe an den Menschen zurückdelegiert werden kann).

Eine besondere Kategorie menschlicher Fehlhandlungen sind diejenigen, die entgegen der bestimmungsgemäßen Verwendung des Systems erfolgen. Dazu gehören zum einen bewusste Fehlhandlungen ohne Ziel einer Schädigung. Dies sind Handlungen entgegen von Vorgaben, die aus dem Ziel der Erhöhung der Effizienz oder aus Gründen der Arbeitserleichterung (z. B. Umgehen von Sicherheitssystemen [126]) oder der Verbesserung der Zielerreichung (z. B. zu schnelles Fahren bei Zeitdruck oder zur Erhöhung des Thrill [127]) initiiert werden. In der Taxonomie ist dies durch die Kategorie „**Fehlbenutzung wegen Manipulationsanreizen außerhalb der bestimmungsmäßigen Verwendung**“ abgedeckt [128]. Die Manipulation ist in diesem Fall keine bösartige Sabotage, sondern eher die „Ergreifung von Möglichkeiten“. Die Auslegung der technischen Systeme befördert oder hemmt solche Manipulationsmöglichkeiten. Zum anderen gehören zu den Fehlhandlungen entgegen der bestimmungsgemäßen Verwendung auch

[124] Die gleichzeitig seine Schwächen sein können.
[125] „In the loop“ und als Überwacher „on the loop“
[126] Siehe dazu TRBS – 1151 Technische Regel für Betriebssicherheit – Gefährdungen an der Schnittstelle Mensch – Arbeitsmittel – Ergonomische und menschliche Faktoren, Arbeitssystem (TRBS 1151), GMBl. Nr. 17/18 Ausgabe: März 2015, Anlage 6
[127] Holte, H. (2012). Einflussfaktoren auf das Fahrverhalten und das Unfallrisiko junger Fahrerinnen und Fahrer (Schriftenreihe der Bundesanstalt für Straßenwesen, M229). *Bremerhaven: Wissenschaftsverlag NW*.
[128] Siehe auch TRBS – 1151 Technische Regel für Betriebssicherheit – Gefährdungen an der Schnittstelle Mensch – Arbeitsmittel – Ergonomische und menschliche Faktoren, Arbeitssystem (TRBS 1151), GMBl. Nr. 17/18 Ausgabe: März 2015.

diejenigen, die gezielt durchgeführt werden im Wissen um die damit verbundenen schadhaften Folgen. Unter dem Aspekt dieser **Schädigung oder bewussten Störung der intendierten Funktionsweise** lassen sich physische Sabotageakte sowie Eingriffe in die Funktionsweise des Systems zusammenfassen, die mit etwaigen Folgen für die Sicherheit des Menschen verbunden sind. Wechselt man den Blickwinkel von diesem Aspekt menschlicher Involviertheit auf die Eigenschaften des davon getroffenen Systems (Störbarkeit durch den Menschen durch gezielte Eingriffe), so wäre das die Security des Systems. Wie in Abschnitt 4.2.5 dargelegt umfasst die Security die Fähigkeit des Systems, Angriffen zu widerstehen bzw. trotz Angriffen fehlerfrei (oder fehlerminimiert) zu agieren.

Der Mensch ist als **Gefährdeter** Angriffspunkt von Schädigungen und muss dementsprechend geschützt werden. Bedeutsam ist dabei die Unterscheidung, zu welcher Klasse die Gefährdeten gehören bzw. welche individuellen Fähigkeiten ihnen zugebilligt werden können. So kann z.B. involvierten Erwachsenen mehr Umsicht im Umgang mit Produkten zugemutet werden als Kindern. Ferner relevant ist auch die Frage, welche Vulnerabilität die Gefährdeten aufweisen, die berücksichtigt werden muss (z. B. von Kindern beim Spielen mit Spielzeug). Die rechtlichen Anforderungen orientieren sich zweckmäßigerweise an diesen Eigenschaften der potentiell Gefährdeten. In der Taxonomie sind die Subklassen **eingewiesene Arbeitnehmer, Nutzer, Verbraucher sowie Dritte ohne Beziehung zum System** definiert.

Diese Subklassen unterscheiden sich zunächst in ihrer Schutzbedürftigkeit. Die Schutzbedürftigkeit orientiert sich unter anderem an dem vorausgesetzten Wissen über Funktion und Gefahrenpotenziale des Systems, in das diese Menschen involviert sind. Weiter unterscheiden sie sich in ihren Sicherheitserwartungen an das System und in ihrer vorausgesetzten Fähigkeit und (ökonomisch und/oder sozial bedingten) Bereitschaft, mit Gefahrenpotenzialen adäquat (also ggf. auch aktiv sicherheitsfördernd) umzugehen. Zudem orientiert sich die Unterteilung an den relevanten Gesetzen.

INVOLVIERTHEIT DES MENSCHEN

DEFINITION	Mensch als Angriffspunkt von Schädigungen oder als Subjekt, das durch Interaktionen mit software-technischen Systemen Teil des Gesamtsystems wird und sicherheitsbeeinflussend wirkt.

UNTERKATEGORIEN ***Mensch als Handelnder***
Mensch als Gefährdeter

Mensch als Handelnder

DEFINITION	Einfluss des mit dem software-physischen System interagierenden Menschen auf die Sicherheit des Gesamtsystems.

AUSPRÄGUNGEN

Mensch als Teil der Prozesskette

Mensch als potentiell fehlerhaftes Element, aber auch als Sicherungssystem beim Einsatz software-physischer Systeme.

Fehlbenutzung entgegen bestimmungsgemäßer Verwendung

Fehlbenutzung wegen Manipulationsanreizen entgegen der bestimmungsmäßigen Verwendung, aber ohne Ziel der Schädigung

Schädigung bzw. bewusste Störung der intendierten Funktionsweise

Gezielte und nicht vorgesehene Einwirkung auf das software-physische System mit möglichen schadenbringenden Folgen für dessen sicheres Funktionieren (Sabotage)

.

Mensch als Gefährdeter

DEFINITION	Rolle des Menschen als durch software-physische Systeme gesundheitlich Gefährdeten.

Klassen Gefährdeter

DEFINITION	Unterscheidung von Gefährdeten ihre Schutzbedürftigkeit und entsprechend ihrer gesetzlichen Unterscheidung.

AUSPRÄGUNGEN

Eingewiesene Arbeitnehmer

Weisungsgebundene Beschäftigte, die im Rahmen ihrer entgeltlichen Tätigkeit in die untersuchten Systeme involviert sind und eine entsprechende Einweisung oder ein entsprechendes Training erhalten haben

Nutzer

Jeder, der das untersuchte System in Gang setzt, steuert oder daraus unmittelbaren Nutzen zieht und nicht in die Klasse der eingewiesenen Arbeitnehmer fällt

Verbraucher

Jeder Nutzer, der das System nicht im Rahmen überwiegend gewerblicher oder zu selbstständiger beruflicher Tätigkeiten nutzt

Dritte ohne Beziehung zum System

Menschen ohne unmittelbare Beziehung zum System, beispielsweise die Bewohner in der Umgebung eines Kraftwerks.

4.2.7 Dimension Schadensfolgen

Die vorab beschriebenen Dimensionen beschreiben Eigenschaften von Systemen in Bezug auf ihr Einsatzfeld unter Beteiligung des Menschen, die die Vermeidung fehlerhaften Systemverhaltens und die Möglichkeiten zur Abwendung dieser Folgen beeinflussen. Wie bedeutsam diese Befähigung letztendlich ist, hängt entscheidend davon ab, wie gravierend die Folgen dieses Fehlverhaltens sind. Folglich gilt es, die unerwünschten Folgen der Nutzung von bzw. Interaktion mit software-physischen Systemen in Wechselwirkung mit dem Umfeld ihres Einsatzes zu beschreiben. Neben den Fehlerfolgen eines bestimmungsgemäßen Einsatzes gilt es auch Fehlnutzungen ungeachtet der ihr zugrundeliegenden Intention (gezielte vs. unbeabsichtigter Schaden, vgl. Abschnitt 4.2.6 Dimension Involviertheit des Menschen) zu berücksichtigen.

Die Folgen können sich auf unterschiedliche geschützte Rechtsgüter beziehen. Im Fokus der Dimension **SCHADENSFOLGEN** steht die Betrachtung der **Personenschäden**. Die Betrachtung von Sachschäden 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 irrelevant – es sei denn man betrachtet auch andere Schäden (siehe unten). Unter besonderen Umständen mag diese Feststellung schwierig sein, wenn sich eine derartige Gefährdung nur in ganz speziellen, nicht vorbedachten Situationen ergibt. Sollten Personenschäden möglich sein, wird weiterhin zwischen **geringfügigen und erheblichen Personenschäden** unterschieden. Als geringfügige Personenschäden werden reversible Beeinträchtigungen der körperlichen Unversehrtheit bezeichnet, während irreversible Beeinträchtigungen der körperlichen Unversehrtheit – Todesfälle eingeschlossen – als erhebliche Personenschäden eingeordnet werden.
Ein gesamtheitliches Folgenbild erfordert es jedoch, auch **sonstige Schäden** zu betrachten, die in Verbindung mit bzw. zusätzlich zu Beeinträchtigungen der körperlichen Unversehrtheit denkbar sind. Dazu gehören **Sachschäden** (Schäden an Sachen und Tieren), **Umweltschäden** (Schäden an der Umwelt durch Emissionen) und **ideelle Schäden** (Verletzung immaterieller Güter, wie bspw. Daten).
Die Betrachtung der Schadensfolgen ist weniger durch die Eigenschaften der Steuerung eines Systems determiniert und damit auch unabhängig vom Einsatz neuer Methoden. Ausschlaggebend ist hier vielmehr die physische Ausgestaltung des Systems und seines Umfelds. Im Gegensatz zu den anderen Dimensionen der Taxonomie spiegelt diese Dimension damit keine qualitative oder quantitative Veränderung infolge neuartiger Entwicklungsmethoden von KI-Systemen wieder, da lediglich Art und Ausmaß der Schadensfolgen, nicht aber die durch das Systemverhalten beeinflussbare Wahrscheinlichkeit ihres Auftretens betrachtet wird. Art und Ausmaß der Schadensfolgen bestimmen jedoch die Bedeutsamkeit der Erfüllung der anderen sicherheitsbezogenen Kriterien. Zudem können mit gravierenden Schadensfolgen gesetzliche Maßnahmen strenger ausgestaltet werden. Die Erörterung der rechtlichen Implikationen der an der Taxonomie aufgespannten sicherheitsbezogenen Eigenschaften software-physischer Systeme erfolgt im anschließenden Kapitel.

SCHADENSFOLGEN

DEFINITION	Mögliche unerwünschte Folgen des Einsatzes software-physischer Systeme.

UNTERKATEGORIEN	***Personenschäden***
	Sonstige Schäden

Personenschäden

DEFINITION	Beeinträchtigung der körperlichen Unversehrtheit durch software-physische Systeme.

AUSPRÄGUNGEN

Keine Personenschäden

Keine Beeinträchtigung der körperlichen Unversehrtheit möglich

Geringfügige Personenschäden

Reversible Beeinträchtigungen der körperlichen Unversehrtheit (Klasse S1 [129]) möglich

Erhebliche Personenschäden

Irreversible Verletzungen oder Tod (Klasse S2 ebd.) möglich

Sonstige Schäden

DEFINITION	Weitere Schäden, die in Verbindung mit bzw. zusätzlich zu Beeinträchtigungen der körperlichen Unversehrtheit denkbar sind.

AUSPRÄGUNGEN

Sachschäden

Schäden an Sachen und Tieren

Umweltschäden

Schäden an der Umwelt durch Emissionen (Schall, Strahlung, Erschütterungen, Ausbringung von Umweltgiften etc.)

Ideelle Schäden

Verletzung immaterieller Güter (Nichtvermögensschäden, Belästigungen, Ehrverletzungen und Herabwürdigungen, Schmerzen)

[129] Siehe Mössner, T. (2012). Risikobeurteilung im Maschinenbau, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Forschung Projekt F 2216.

4.3 Grafischer Überblick über die Taxonomie

Veränderbarkeit	
System	Keine Veränderung nach Auslieferung
	Vorgesehene, kontrollierte Veränderung nach Auslieferung
	Adaptivität (geringfügige Veränderung weniger Parameter im Betrieb)
	Bedeutende Veränderungen im Betrieb
Umfeld	Geringe Änderungen in einem kontrollierbaren Umfeld
	Nicht vorhersehbare Änderungen in einem komplexen Umfeld

Transparenz	
Experten	Spezifizierbarkeit
	Beschreibbarkeit der Funktionsgrenzen
	Nachvollziehbarkeit
	Vorhersehbarkeit
Beteiligte	Kenntnis des Einsatzbereichs und der Grenzen des Systems
	Vorhersehbarkeit der Dynamik bzw. des Prozesszustands des Systems
	Verständnis der Systemfunktionalität

Vernetzung	
Intern	Zentrale Vernetzung
	Dezentrale Vernetzung
Nach außen	Abgesprochene Daten, nur informativ oder handlungsbeeinflussend
	Unabgesprochene Daten

Kontrollierbarkeit	
Emergenz	Autonomie (relative Unabhängigkeit durch Gütekriterien, Zielhierarchien)
	Selbstorganisation (Ordnung durch interne Interaktion)
Beschrän-kungen	Systembeschränkungen durch virtuelle Schutzzäune/konventionelle Systeme
	Beschränkungen im Umfeld oder Einsatzbereich
	Beschränkungen und Kontrolle im Datenfluss

Widerstandsfähigkeit	
Robustheit	Stabilität bei kleinen Änderungen des Inputs
	Bewältigung unbekannter Situationen bzw. unvorhergesehener Ereignisse
	Security
Resilienz	Passive Wirkungsbegrenzung nach Fehlern
	Aktive Wirkungsminderung nach Fehlern

Involviertheit des Menschen	
Handelnder	Mensch als sicherheitsgewährender sowie fehlerbehafteter Teil der Prozesskette
	Fehlbenutzung entgegen bestimmungsgemäßer Verwendung
	Bewusste Störung der intendierten Funktionsweise bzw. Schädigung
Gefährdeter	Eingewiesene Angestellte
	Nutzer
	Verbraucher
	Dritte ohne Beziehung zum System

Schadensfolgen	
Personen-schäden	Keine Personenschäden
	Geringfügige Personenschäden
	Erhebliche Personenschäden
Sonstige Schäden	Sachschäden
	Umweltschäden
	Immaterielle Schäden

Abb. 4.2 Grafischer Überblick über die Taxonomie

4.4 Beispielhafte Anwendung der Taxonomie

In der Darstellung der rechtlichen Erörterungen in Abschnitt 5 ist die Anwendbarkeit der entwickelten Taxonomie ausreichend demonstriert und an Szenarien exerziert. Im Folgenden werden weitere mögliche Anwendung vorgestellt. Diese sollen die Taxonomie demonstrieren und erheben nicht den Anspruch, finale Entwürfe zu sein.

Anwendung: Klassen von KI-Systemen

In einigen Normungsaktivitäten (wie z. B. in ISO/IEC JTC 1/SC 42/WG 3) wird gegenwärtig unabhängig von diesem Forsschungsprojekt versucht, KI-Systeme bezüglich Safety in Klassen einzuteilen und diese bezüglich ihres mögichen Einsatzes in mehr oder minder sicherheitskritischen Szenarien zu bewerten. Im Folgenden ist ein mögliches Vorgehen vorgestellt, dass die entwickelte Taxonomie als Grundlage heranzieht.
Dabei werden die beiden Dimensionen „Involviertheit des Menschen“ und der „Schadensfolgen“ sowie die Transparenz aus Beteiligtensicht herangezogen, um unterschiedliche Gefährdungsszenarien aufzuspannen. Zudem werden auf Basis der fünf verbleibenden Dimensionen unterschiedliche Merkmalsprofile gebildet und daraus Systemklassen gebildet. Die (in diesem Beispiel fünf) Systemklassen werden anschließend in Bezug auf ihren Einsatz in den (hier ebenfalls fünf) Szenarien bewertet. Im vorliegenden Beispiel erfolgt eine Bewertung in Bezug auf die Einsatzfähigkeit des Systems in ebenfalls fünf Stufen (A bis E) unter Nennung der dazu erforderlichen Voraussetzung.

A	Einsatz von KI problemlos möglich und heute schon realisiert
B	Einsatz von KI erfordert besondere Methoden der Sicherstellung der Robustheit und Transparenz, die aber heute schon in der Entwicklung sind und möglicherweise in einigen Jahren verfügbar sind.
C	Einsatz von KI erfordert besondere Methoden der Sicherstellung der Robustheit und Transparenz, darüber hinaus aber Konzepte zur Sicherstellung der Datenqualität und Datenintegrität. Entwicklungen in dieser Richtung laufen und sind mittelfristig verfügbar.
D	Einsatz von KI erfordert besondere Methoden der Sicherstellung der Robustheit und Sicherstellung von Transparenz, darüber hinaus aber U.U. Konzepte zur periodischen Überwachung durch Beobachtungsbehörden und/oder Hersteller. Entwicklungen in dieser Richtung sind noch nicht gestartet. Realisierung in langfristig möglich. Einsatz nicht empfohlen für Risikoszenarien mit hohem Schadenpotential
E	Einsatz von KI-Methoden wegen zu geringer Kontrollierbarkeit bis auf nicht abschätzbare Zeit nicht möglich

Klasse 1-KI-Systeme: KI-Systeme mit niedriger Anzahl von Variablen, z.B. Symbolische KI mit einem überschaubaren Satz an Regeln. Autonomie möglich, wird aber durch die Analysierbarkeit, Erklärbarkeit und daraus resultierenden Kontrollierbarkeit abgefangen

Veränderbarkeit	Im Betrieb nicht veränderlich
Transparenz	Das System ist voll erklärbar, nachvollziehbar und analysierbar. Funktionsgrenzen sind eindeutig beschreibbar
Vernetzung	Verhalten wird nicht von außen bestimmt
Kontrollierbarkeit	Voll kontrolliert, gewisse Autonomie, aber durch Analysierbarkeit, Erklärbarkeit und daraus resultierenden Kontrollierbarkeit abgefangen
Widerstandsfähigkeit	Die Software im System ist so robust wie konventionelle Software, das System ist so resilient, wie konventionelle Systeme

Klasse 2-KI-Systeme: KI-Systeme mit hoher Anzahl von Variablen, z.B. künstliche neuronale Netze aber im Betrieb nicht weiterlernend. Zeigt in gewissem Ausmaß Emergenz bzw. Autonomie. Die Robustheit ist gegenüber konventioneller Software reduziert. Die Transparenz ist reduziert. Die Kontrolle ist aber durch Beschränkungen gewährleistet. Verhalten wird nicht von außen bestimmt.

Veränderbarkeit	Im Betrieb nicht veränderlich
Transparenz	Das System ist nicht voll erklärbar, nachvollziehbar oder analysierbar. Funktionsgrenzen sind aber eindeutig beschreibbar
Vernetzung	Verhalten wird nicht von außen bestimmt
Kontrollierbarkeit	emergentes Verhalten bzw. Autonomie, aber Beschränkungen sorgen für sicherheitsrelevante Kontrolle, daher im sicherheitskritischen Einsatz so sicher wie konventionelles System
Widerstandsfähigkeit	Die Software im System ist weniger robust wie konventionelle Software, das System ist aber so resilient, wie konventionelle Systeme

Klasse 3a-KI-Systeme: KI-Systeme mit hoher Anzahl von Variablen, z.B. künstliche neuronale Netze aber im Betrieb nicht weiterlernend. Zeigt in gewissem Ausmaß Emergenz bzw. Autonomie. Die Transparenz ist durch neu entwickelte Analysemethoden ausreichend hoch. Die Robustheit ist gegenüber konventioneller Software zwar reduziert, die Kontrolle ist aber durch die hohe Transparenz gewährleistet. Aus diesem Grunde sind Beschränkungen nicht notwendig. Verhalten wird nicht von außen bestimmt. Komplexität aus Expertensicht gering oder kompliziert beherrschbar.

Veränderbarkeit	Im Betrieb nicht veränderlich
Transparenz	Das System ist durch neu entwickelte Methoden erklärbar, nachvollziehbar oder analysierbar. Funktionsgrenzen sind eindeutig beschreibbar.
Vernetzung	Verhalten wird nicht von außen bestimmt
Kontrollierbarkeit	emergentes Verhalten bzw. Autonomie, aber hohe Transparenz sorgt für sicherheitsrelevante Kontrolle, Beschränkungen in Form einer White-Box sind nicht notwendig. U.U. sind Beschränkungen im Einsatzbereich angezeigt.
Widerstandsfähigkeit	Das System ist durch neu entwickelte Methoden ausreichend widerstandsfähig

Klasse 3b-KI-Systeme: Wie 3a, aber, Verhalten wird auch von außen bestimmt.

Veränderbarkeit	Im Betrieb nicht veränderlich
Transparenz	Das System ist durch neu entwickelte Methoden erklärbar, nachvollziehbar oder analysierbar. Funktionsgrenzen sind eindeutig beschreibbar.
Vernetzung	Verhalten wird auch von außen bestimmt
Kontrollierbarkeit	emergentes Verhalten bzw. Autonomie, aber hohe Transparenz sorgt zwar zunächst für sicherheitsrelevante Kontrolle, die Tatsache der Vernetzung nach außen erfordert aber besondere Datensicherungsmaßnahmen.
Widerstandsfähigkeit	Das System ist durch neu entwickelte Methoden ausreichend widerstandsfähig

Klasse 4a/b-KI-Systeme: KI-Systeme mit hoher Anzahl von Variablen, z.B. künstliche neuronale Netze im Betrieb kontrolliert weiterlernend. Wird aber in Intervallen einer Beobachtung durch Behörden und/oder Hersteller unterzogen und dann neu analysiert. Zeigt in gewissem Ausmaß Emergenz bzw. Autonomie. Die Transparenz ist durch neu entwickelte Analysemethoden ausreichend hoch, bezieht sich aber immer nur auf den Auslieferungszustand bzw. den Zustand nach Neuanalyse in regelmäßigen Zeiträumen. Wegen des Weiterlernens schwindet die Transparenz und Kontrollierbarkeit. Die Robustheit und Transparenz entsprechen zwar prinzipiell der Klasse3-Systeme, die Kontrolle ist aber durch die hohe Autonomie nur beschränkt gewährleistet. Mit Beschränkungen würden sie zu Klasse-2-Systemen mutieren. Wenn das Verhalten von außen bestimmt ist, sind Maßnahmen wie in Klasse 3b notwendig.

Veränderbarkeit	Im Betrieb veränderlich (z.B. weiterlernend), wird aber in Intervallen neu analysiert
Transparenz	Das System ist durch neu entwickelte Methoden erklärbar, nachvollziehbar oder analysierbar, aber immer nur zu den Zeitpunkten, an denen es unter der Kontrolle des Herstellers ist. Funktionsgrenzen sind eindeutig beschreibbar.
Vernetzung	Verhalten wird nicht (Variante **a**) auch (Variante **b**) von außen bestimmt
Kontrollierbarkeit	Das System ist wegen neuer Methoden erklärbar, nachvollziehbar oder analysierbar, aber immer nur an den Zeitpunkten, an denen es unter der Kontrolle des Herstellers steht. Es gibt keine Systembeschränkungen. Wenn eine Vernetzung nach außen vorliegt, sind aber besondere Datensicherungsmaßnahmen notwendig.
Widerstandsfähigkeit	Das System ist durch neu entwickelte Methoden ausreichend widerstandsfähig

Klasse 5a/b-KI-Systeme: KI-Systeme mit hoher Anzahl von Variablen, z.B. künstliche neuronale Netze im Betrieb unkontrolliert weiterlernend. Zeigt Anzeichen von Selbstorganisation. Zeigt hochgradige Emergenz. Die Transparenz ist wegen der hohen Emergenz gering. Die Robustheit ist nicht kontrollierbar. Mit Beschränkungen

würden sie zu Klasse-2-Systemen mutieren. Wenn das Verhalten von außen bestimmt ist, kann die Kontrolle u.U. durch permanenten Zugriff erhöht sein.

Veränderbarkeit	Im Betrieb stark veränderlich (z.B. selbstständig weiterlernend)
Transparenz	Das System ist durch neu entwickelte Methoden erklärbar, nachvollziehbar oder analysierbar, aber immer nur zu den Zeitpunkten, an denen es unter der Kontrolle des Herstellers ist. Funktionsgrenzen sind eindeutig beschreibbar.
Vernetzung	Verhalten wird nicht (Variante **a**) auch (Variante **b**) von außen bestimmt
Kontrollierbarkeit	Hohe Emergenz und deswegen geringe Kontrollierbarkeit. Wenn eine Vernetzung nach außen vorliegt, sind zwar besondere Datensicherungsmaßnahmen erforderlich, die Kontrolle aber u.U. erhöht. Es gibt keine Systembeschränkungen.
Widerstandsfähigkeit	Das System ist durch neu entwickelte Methoden ausreichend widerstandsfähig

Zusammengefasst lässt sich dies dann wie folgt visualisieren:

Tab. 4.1 Matrix von Szenarien und KI-Klassen im Safety-Bezug

Szenarien	**Klasse 1**	**Klasse 2**	**Klasse 3**	**Klasse 4**	**Klasse 5**
Variation der Involviertheit des Menschen,der Transparenz aus Beteiligtensicht und der Schadensfolgen					
Erhebliche Schäden möglich, Mensch als Sicherheitsrisiko denkbar	**A**	**C**	**D**	**E**	**E**
Erhebliche Personenschäden möglich, auch Dritte potentiell gefährdet, erhebliche Schäden möglich, Transparenz aus Beteiligtensicht gering, Systemverhalten auch nach Schulung nur schwer vorhersehbar	**A**	**A**	**D**	**E**	**E**
Erhebliche Personenschäden möglich, Mensch potentiell gefährdet, aber insbesondere eingewiesene Angestellte. Einfluss des Menschen auf Sicherheit möglich.	**A**	**A**	**C**	**D**	**E**
Personenschäden möglich, Mensch potentiell gefährdet, Einfluss des Menschen auf Sicherheit möglich, Schäden möglich	**A**	**A**	**B**	**D**	**E**
Mensch nicht gefährdet, Kein Einfluss des Menschen auf Sicherheit möglich, keine Schäden möglich.	**A**	**A**	**A**	**A**	**A**

5 Rechtsgutachten

5.1 Zusammenfassung

In allen Lebensbereichen kommen zunehmend Produkte zum Einsatz, die mit mehr oder weniger komplexer Software ausgestattet sind. Diejenigen Rechtsgebiete, die der Gewährleistungen von Sicherheitsstandards von Produkten dienen, wurden für Produkte entwickelt, die (noch) nicht über eingebettete Software verfügten. Durch die Entwicklung immer komplexer werdender Software, bei der Algorithmen der Künstlichen Intelligenz (KI) verwendet werden, stellt sich die Frage, ob das Recht noch auf diese Produkte so angewendet werden kann, dass es einerseits weiterhin ein hohes Maß an Sicherheit garantiert, andererseits die technische Entwicklung solcher Systeme nicht verhindert.

Das Produktsicherheitsrecht und das Arbeitsschutzrecht setzen den ordnungsrechtlichen Rahmen für das Inverkehrbringen von Produkten und die Verwendung von Arbeitsmitteln im Betrieb. Sie bilden zusammen den Kern des Rechts des technischen Arbeitsschutzes. Beide sind durch europäische Richtlinien geprägt. Für unterschiedliche Produkttypen gibt es jeweils gesonderte Regelungen, die jedoch nach demselben Konzept entwickelt sind und sich daher im Aufbau und der dort verwendeten Regelungsinstrumente ähneln. Die für das Inverkehrbringen und die Inbetriebnahme von Maschinen maßgebliche Maschinenrichtlinie wird derzeit überarbeitet.[130] Sie soll unter anderem der Entwicklung neuer Technologien im Bereich der vernetzten Systeme („Internet of Things“), der zunehmend Digitalisierung im Maschinenbereich und auch dem Einsatz von KI-Systemen in Maschinen angepasst werden. Sie ist daher zentraler Gegenstand der vorliegenden Untersuchung. Da das europäische Produktsicherheitsrecht nach einem einheitlichen Konzept entwickelt ist, kann sie beispielhaft für die anderen produktspezifischen Richtlinien stehen. Gleichwohl soll nicht unterschlagen werden, dass sich die einzelnen Regelungen unterscheiden. Die Untersuchung konzentriert sich insbesondere wegen der angesprochenen Novellierung der Maschinenrichtlinie auf den Einsatz von KI-Systemen in Maschinen. Als weitere Rechtsgebiete werden zudem das Bundesimmissionsschutzgesetz sowie haftungsrechtliche Regelungen in den Blick genommen. Ergänzend werden die technischen Anforderungen dargestellt, die sich aus der Datenschutzgrundverordnung, dem BSI-Gesetz sowie weiterer europarechtlicher Regelungen ergeben. Zusammen lässt sich so ein Überblick über die Regelungen verschaffen, die den hier interessierenden Teil des Technikrechts bilden.

Die hier untersuchten KI-Systeme lassen sich bezogen auf eine rechtliche Relevanz anhand von sieben Kriterien in eine Taxonomie einordnen. Die unterschiedlichen Ausprägungen der Taxonomiekriterien und deren Kombination bergen unterschiedliche rechtliche Probleme bzw. Herausforderungen.

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. Im Zeitpunkt des Inverkehrbringens bzw. der Inbetriebnahme müssen alle formellen und materiellen Voraussetzungen vorliegen.

---

[130] Zum Stand der Überarbeitungsinitiative siehe https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/2019-Revision-of-the-Machinery-Directive (zuletzt abgerufen am 02.04.2020).

Insbesondere sich im Betrieb auf Basis neuer Daten verändernde Systeme, sog. weiterlernende Systeme, sind nicht mehr nach diesen Kriterien rechtlich beherrschbar. Denn der maßgebliche Zeitpunkt der Risikobeurteilung, die Inbetriebnahme, blendet die Veränderungen des Systems nach der Inbetriebnahme aus.
Vernetzte Systeme werfen zudem Probleme bei der Frage auf, ob es sich um einzelne Maschinen handelt oder um eine Gesamtheit von Maschinen. Daraus folgt die Frage, wer für die Gesamtheit als Hersteller verantwortlich ist.
Werden die Maschinen als Arbeitsmittel eingesetzt, muss der Arbeitgeber zudem dafür sorgen, dass von den Maschinen keine Gefährdungen für seine Beschäftigten ausgehen. In Abhängigkeit der Merkmale der Maschine sind die jeweils zu stellenden Anforderungen an die Sicherheit und die dafür erforderlichen Maßnahmen bei Konstruktion und Bau der Maschine durch den Hersteller unterschiedlich zu definieren. Entsprechend muss der Arbeitgeber die Umstände in seinem Betrieb und die Eigenschaften der verwendeten Maschine bei der Ermittlung der erforderlichen Schutzmaßnahmen beachten.
Bei den weiterlernenden Systemen, aber auch bei solchen, die durch Updates auch nach Inbetriebnahme durch den Hersteller in sicherheitsrelevanter Weise verändert werden, ist der Arbeitgeber trotzdem nach der jetzigen Rechtslage in der Verantwortung für die Sicherheit des Arbeitsmittels mit KI-Komponente.
Die Verantwortungsverteilung zwischen Hersteller und dem Arbeitgeber erscheint dann nicht mehr geeignet, die möglichen Risiken beherrschbar zu halten, wenn sich die Eigenschaften der Maschine nach dem für den Hersteller maßgeblichen Zeitpunkt der Inbetriebnahme ändern. Einerseits steht der Hersteller vor dem Problem, die Erfüllung der materiellen Anforderungen an die Beschaffenheit der Maschine zu diesem (frühen) Zeitpunkt nachzuweisen. Andererseits werden die Entwicklungsrisiken, die sich erst nach Inbetriebnahme zeigen, ganz in die Sphäre des Arbeitgebers verlegt.
Für die Pflicht des Anlagenbetreibers nach Immissionsschutzrecht, die genehmigungspflichtige Anlage entsprechend der erteilten Genehmigung zu betreiben, ergeben sich bei weiterlernenden Systemen entsprechende Fragen. Die durch das System initiierten Änderungen müssten von der erteilten Genehmigung erfasst sein. Im Immissionsschutzrecht lässt sich jedoch durch technische Gewährleistung der Einhaltung der Grenzwerte erreichen, dass auch eine veränderbare Anlage sich im Rahmen der Genehmigung bewegt. Etwas anderes gilt im Bereich der Störfall-Verordnung, wonach der Anlagenbetreiber die erforderlichen Maßnahmen zur Verhinderung von Störfällen ergreifen muss. Wie eng demnach z. B. eine Überwachung sein muss, wenn ein weiterlernendes System eingesetzt wird, ist durch das Recht nicht vorgegeben.
Im Haftungsrecht ist in Deutschland bei der richterrechtlich geprägten Produzentenhaftung bereits eine Pflicht des Herstellers zur Beobachtung des Produkts nach Inverkehrbringen etabliert, womit bei weiterlernenden Systemen das Risiko unkalkulierbarer Schäden durch das System vom Hersteller zu tragen ist. Dieses Konzept findet sich jedoch nicht in der verschuldensunabhängigen Produkthaftung, sodass hier ein Ungleichgewicht zulasten des Geschädigten gesehen werden kann. Zudem ist bei hochgradig vernetzten Systemen für den Geschädigten oft unklar, wem gegenüber er einen Schadensersatzanspruch hat.
Es erscheint angebracht, für bestimmte Produkttypen die Pflichten des Herstellers zu erweitern und das Produktsicherheitsrecht zu konkretisieren, um den Eigenschaften bestimmter KI-Systeme gerecht zu werden. Das betrifft insbesondere solche Systeme, die weiterlernen oder durch Updates durch den Hersteller (oder seinen

Bevollmächtigten) nach Inbetriebnahme verändert werden. Ein solcher Ansatz findet sich bereits im neuen europäischen Kaufrecht zu Waren mit „Embedded Software“ und wird derzeit für die Überarbeitung des Produkthaftungsrechts diskutiert.
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. Das kann für Maschinen entsprechend klargestellt werden. 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.
Daraus können zwei Lösungsansätze formuliert werden:

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

Was die Widerstandsfähigkeit von Systemen im Sinne der IT-Sicherheit angeht, trifft die DSGVO für solche Systeme bereits Regelungen, die personenbezogene Daten verarbeiten. KI-Systeme, die personenbezogene Daten verarbeiten, unterfallen schon jetzt den Regelungen der DSGVO. Bei der Weiterentwicklung des Produktsicherheitsrechts sind die dort bereits getroffenen Vorgaben zu berücksichtigen. Zugleich sieht das europäische Recht bereits jetzt ein Zertifizierungsverfahren für IT-Sicherheit vor. Auf dieses System kann zurückgegriffen werden, wenn für die hier untersuchten Systeme verpflichtende Anforderungen an die IT-Sicherheit formuliert werden sollen.
Damit ergibt sich folgender Ansatz:

- Schaffung einer verpflichtenden Zertifizierung für sicherheitsrelevante Daten bzw. Datenprodukte, also eines „Produktsicherheitsrechts für Daten“

Im Haftungsrecht kann die Pflicht des Herstellers zur Produktbeobachtung auch im Produkthaftungsrecht etabliert werden. Zudem kann für bestimmte Systeme eine Gefährdungshaftung des Nutznießers dieser Systeme in Betracht gezogen werden.

5.2 Einleitung

Mit der fortschreitenden Digitalisierung werden immer mehr Lebensbereiche mit digitalen Produkten durchdrungen und geprägt. Bereits jetzt entstehen durch von Verbrauchern genutzte Anwendungen im privaten Bereich große Datenmengen, z. B. über Smartphones. Gleiches gilt in zunehmendem Maße für den industriellen Bereich. Der Einsatz von sog. KI-Systemen wird damit auch in der Industrie interessanter, da sie große Datenmengen verarbeiten können.[131] Gleichzeitig sollen auch Small Data-Anwendungen in Zukunft eine größere Rolle spielen.[132] Unter Begriffen wie Arbeit 4.0,

[131] *Europäische Kommission* COM(2020) 65 final, Weißbuch – Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, S. 2, abrufbar unter: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (Stand: 19.02.2020, zuletzt abgerufen am 09.03.2020).
[132] *Bundesregierung*, Strategie Künstliche Intelligenz der Bundesregierung, S. 35, abrufbar unter https://www.bundesregierung.de/resource/blob/997532/1550276/3f7d3c41c6e05695741273e78b8039f

Arbeitsschutz 4.0 oder Industrie 4.0 wird neben verschiedenen Aspekten wie Datenschutz, Arbeitszeitregelungen und die menschengerechte Ausgestaltung von Arbeitsplätzen beim Einsatz von KI-Systemen insbesondere die Frage nach der Gewährleistung der Sicherheit am Arbeitsplatz diskutiert.[133] Im Bereich der Verbraucherprodukte macht es aus Sicht der Verbraucher für die erwartete Sicherheit keinen Unterschied, ob ein Produkt mit oder ohne KI auskommt[134]. Gleiches müssen auch Beschäftigte im Hinblick auf die Arbeitssicherheit beim Einsatz von KI-Systeme am Arbeitsplatz erwarten können.
Im Zentrum der Untersuchung steht daher der vorgreifende Gefahrenschutz durch Produktsicherheitsrecht und Arbeitsschutzrecht. Es stellt sich die Frage, ob die bestehenden Regeln des aus diesen beiden Rechtsbereichen gebildeten technischen Arbeitsschutzes ihrem Zweck noch gerecht werden können, wenn sie mit immer autonomeren Systemen mit mehr KI-Komponenten konfrontiert werden. Daraus folgt wiederum die Frage, wie das Recht auf die neuen Anforderungen derartiger Technologien reagieren kann. Die vorliegende Darstellung versucht, Antworten zu diesen Fragen zu aufzuzeigen.

5.2.1 Schadensfolgen als Eingangskriterium

Es sollen nur solche KI-Systeme betrachtet werden, die für die Sicherheit von Menschen oder Sachen von Bedeutung sind, wenn also Schadensfolgen (siehe Taxonomie) möglich sind. Es werden solche Systeme von der Untersuchung ausgenommen, die keinerlei Wirkung in der physischen Außenwelt haben können, der Menschen oder Sachen also nicht ausgesetzt sein können und deshalb keine Schadensfolgen drohen [135].
Neben den bereits benannten Personen- und Sachschäden sind grundsätzlich auch Schäden durch Verlust oder ungewollte Veränderung von personenbezogenen Daten möglich. Diese versucht das Datenschutzrecht zu vermeiden. Das Datenschutzrecht soll hier jedoch nicht wegen seiner Schutzrichtung betrachtet werden, da die vorliegende Untersuchung nur die Sicherheit von Menschen und Sachen zum Gegenstand hat. Die Einbeziehung erfolgt vielmehr vor dem Hintergrund, dass im Datenschutzrecht bereits jetzt verschiedene Regelungen zur Gestaltung von (automatisierten) Datenverarbeitungssysteme bestehen, die wiederum in den hier untersuchten KI-Systemen zum Einsatz kommen können.

5.2.2 Das Recht als Gegenstand der Untersuchung

Wegen der mehr oder weniger stark ausgeprägten Unvorhersehbarkeit des Verhaltens mancher Systeme sind auch mögliche Verzweigungen in Verhaltenssträngen oder sogar neues, nur implizit programmierbares, aber nicht in allen Einzelheiten vorhersehbares Verhalten möglich. Nach Außen machen diese Systeme deshalb den Eindruck, als wäre es „ihr“ Verhalten. Der Gedanke liegt daher nicht fern, solche

---

2/2018-11-15-ki-strategie-data.pdf?download=1 (Stand: November 2018, zuletzt abgerufen am 19.05.2020).

[133] *Bundesministerium für Arbeit und Soziales*, Weißbuch Arbeiten 4.0, S. 138, abrufbar unter: https://www.bmas.de/DE/Service/Medien/Publikationen/a883-weissbuch.html (Stand: März 2017, zuletzt abgerufen am 24.02.2020).

[134] *Europäische Kommission* COM (2020) 65 final, Weißbuch – Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, S. 11.

[135] Indirekte Folgen durch beispielsweise psychische Beeinträchtigung werden nicht betrachtet.

Systeme auch als Adressaten von Recht zu begreifen, sofern Recht als Regelungssystem zur Verhaltenssteuerung verstanden wird.[136] Schließlich ist es das KI-System, das vermeintlich eine „Entscheidung“ trifft. Recht hat jedoch stets den Menschen als Handlungsträger im Blick; der Mensch ist Rechtssubjekt.[137] Der Mensch ist es auch, der die KI-Systeme gestaltet, den Verhaltensspielraum der Systeme definiert, sich ihrer bedient, sie in ihr Arbeitsumfeld integriert, sie mit Daten versorgt etc.[138] Sie sind daher Werkzeug zur Entscheidungsfindung oder Delegierte, denen die Entscheidungsfindung in einer definierten Sache durch den Menschen übertragen wurde.[139] Der Mensch ist damit weiterhin rechtlich adressierbar, wenn es um die Steuerung der „Entscheidung“ des KI-Systems bzw. um die Zuordnung der Verantwortung für die „Entscheidung“ geht. Es stellt sich jedoch die Frage, wie das gegebene Recht diese Adressierung bewerkstelligt, wenn der Mensch durch die Delegierung von Aufgaben mehr und mehr in den Hintergrund tritt. Wann erreicht die Technologie die Grenze, hinter der eine Gewährleistung sicherer maschineller Aufgabenerfüllung durch Verhaltenssteuerung mit dem gegebenen Recht nicht mehr zu bewerkstelligen ist? In welche Richtung kann sich das Recht weiterentwickeln, um der Technologie zu folgen und dabei ihre Entwicklung auf einem Pfad zu halten, der mit den unserem Rechtssystem zugrundeliegenden Entscheidungen vereinbar ist?
Zur Annäherung an eine Beantwortung dieser Fragen ist zunächst zu ermitteln, welche Rechtsnormen überhaupt für die hier untersuchte Technologie relevant sind. So vielfältig wie die denkbaren Anwendungsfelder von KI-Systemen sind, so groß ist die Anzahl der relevanten Gesetze und Verordnungen.
Der in 5.2.1 skizzierte Untersuchungsgegenstand verengt dieses weite Feld möglicher Normen. Es sollen nur solche Regelungen berücksichtigt werden, die für mögliche Anwendungsbeispiele von KI-Systemen im Arbeitsbereich und zum Schutz insbesondere der Nutzer, z. B. Arbeitnehmer im industriellen Bereich, relevant sind. Der gesamte Lebensmittelbereich ist damit ebenso ausgenommen wie die Regelungsbereiche zu pflanzlichen und tierischen Produkten, Arzneimitteln, chemischen Produkten sowie Bauprodukten. Auch wird der gesamte Bereich der Medizinprodukte ausgeklammert.[140]
Das **Produktsicherheitsrecht** bietet für „herkömmliche“ Produkte einen Rahmen, um die Verantwortung für die Gewährleistung der Produktsicherheit zwischen den Produktverantwortlichen[141] zu verteilen, also zwischen den an der Produktion, dem Vertrieb und der Nutzung der Produkte Beteiligten. Im Arbeitsumfeld tritt neben das

---

[136] *Boehme-Neßler*, Die Macht der Algorithmen und die Ohnmacht des Rechts, NJW 2017, 3031, 3035.
[137] Die juristische Person als „nichtmenschliches“ Rechtssubjekt, insbesondere die Kapitalgesellschaft, ist auch menschlich geprägt, da in ihr als Personalverband zumindest mittelbar Menschen repräsentiert sind und diese das Verhalten der juristischen Person als Organwalter oder Shareholder beeinflussen. Selbst die Stiftung, in der keine Personen, sondern Vermögenswerte verselbstständigt sind, wird letztlich auch durch Menschen im Stiftungsgeschäft gegründet und ihr Verhalten im Rechtsverkehr damit durch Menschen vorbestimmt.
[138] Es wird jedoch diskutiert, ob es für zunehmend „intelligentere“ Maschinen (siehe dazu die Ausführungen über Intelligenz) nötig werden wird, diese als Rechtssubjekt zu klassifizieren, insbesondere um eine zweckmäßige Zuordnung von Haftung zu ermöglichen. Vergleiche hierzu (im Ergebnis ablehnend) *Lohmann* Ein europäisches Roboterrecht – überfällig oder überflüssig?,ZRP 2017, 168, 171.
[139] *Dettling/Krüger*, Erste Schritte im Recht der Künstlichen Intelligenz, MMR 2019, 211, 213.
[140] Hier sind gleichwohl Anwendungsfälle von KI-Systemen denkbar.
[141] Diesen zusammenfassenden Begriff verwendet *Schmidt am Busch* in: Eifert: Produktbeobachtung durch Private, 149, 150. Im harmonisierten Produktsicherheitsrecht werden Hersteller, Bevollmächtigte, Einführer und Händler als Wirtschaftsakteure bezeichnet, vgl. § 2 Nr. 29 ProdSG.

Produktsicherheitsrecht zudem das **Recht des technischen Arbeitsschutzes**, welches die Sicherheit der Arbeitnehmer gewährleisten soll. Inwieweit diese Materien schon jetzt den Besonderheiten von KI-Systemen gerecht werden, soll im Folgenden untersucht werden.
Daneben wird das **Bundesimmissionsschutzgesetz** untersucht, das wesentliche Regelungen zur Anlagensicherheit trifft und damit ebenfalls der Risikoverwaltung dient.
Wie bereits erwähnt, ist auch das Datenschutzrecht Gegenstand des Gutachtens. Dabei liegt der Fokus auf der **Datenschutzgrundverordnung**.
Ergänzend dazu wird auf das **Haftungsrecht** eingegangen.

5.2.3 Gang der Untersuchung

Die hier untersuchten software-physischen Systeme mit KI-Komponenten können genauer anhand der im Projekt entwickelten Taxonomie charakterisiert werden. Die einzelnen Merkmale der Taxonomie werden anhand des ermittelten Rechtsrahmens abstrakt auf ihre rechtliche Bedeutung untersucht. Dabei werden nur die wesentlichen Aspekte des jeweiligen Rechtsgebiets in den Blick genommen (5.3).
Der Rechtsrahmen wird in einem zweiten Schritt auf konkrete Beispiele für autonome und KI-Systeme angewendet. Die Beispiele zeichnen sich durch unterschiedlich ausgeprägte Merkmale der in diesem Projekt entwickelten Taxonomie für autonome und KI-Systeme aus. Diese beispielhafte Rechtsanwendung soll helfen zu ermitteln, welche Ausprägungen der Technologie nicht mehr mit dem bestehenden Recht erfasst werden können (5.4).
Darauf aufbauend können in einem dritten Schritt Vorschläge erarbeitet werden, wie die relevanten Regelungen weiterentwickelt werden könnten. Der aufgezeigte Rechtsrahmen gibt dabei vor, wo diese Ansätze zu suchen sind (5.6 und 5.7).

5.3 Kriterien der Taxonomie für autonome und KI-Systeme

Die hier zu untersuchenden KI-Systeme lassen sich anhand der verschiedenen Dimensionen der in diesem Projekt entwickelten Taxonomie charakterisieren. Die einzelnen Dimensionen können rechtlich von unterschiedlicher Relevanz sein. Auch die Kombination von Taxonomiedimensionen in unterschiedlicher Ausprägung hat unterschiedliche rechtliche Konsequenzen.
Die von der Taxonomie identifizierten sieben Dimensionen werden zunächst kurz dargestellt. Für jede Dimension erfolgt dann eine rechtliche Bewertung der nach den unterschiedlichen Rechtsgebieten jeweils kritischen Ausprägungen. Danach werden rechtlich problematische Kombinationen von Taxonomiekriterien erörtert und analysiert.
Unterschiedliche Kombinationen von Ausprägungen der Dimensionen sollen anhand von Beispielen veranschaulicht werden und damit einer rechtlichen Prüfung zugänglich gemacht werden. Auf diese Weise können rechtlich problematische Taxa veranschaulicht werden.

5.3.1 Veränderbarkeit im Betrieb

Die Systeme unterscheiden sich in der **Veränderbarkeit**, d. h. in Art und Ausmaß, in dem sich die **Eigenschaften des Systems** oder **des Umfelds** während ihres

Betriebes verändern. Unter Veränderbarkeit ist in diesem Kontext nur die Veränderbarkeit während des bestimmungsgemäßen Einsatzes oder der vorhersehbaren Fehlanwendung des Systems gemeint. Eine Veränderbarkeit der Systeme nach Inbetriebnahme wird entweder von außen durch Verwender bzw. Hersteller initiiert oder ist im System selbst angelegt.
Die von außen initiierte Veränderung kann z. B. durch Updates oder gezieltes Teaching erfolgen. So können Software-Updates in den Bestand ausgerollt werden; möglich sind Firmware-Änderungen bis hin zu applikativen Modifikationen. Solche Updates können auch auf Basis von im Betrieb gesammelten Daten stattfinden.
Die systeminhärente Veränderung im laufenden Betrieb kann wiederum unterschieden werden nach dem Grad der Veränderbarkeit. Es kann sich nur um eine begrenzte Veränderbarkeit handeln, bei der einzelne Parameter angepasst werden können (Adaptivität) oder um weitreichende Veränderbarkeit mit bedeutender Auswirkung auf Funktionalität oder Sicherheit des Systems. Hierunter fällt insbesondere das weiterlernende System.
Neben der Veränderbarkeit des Systems ist auch die Veränderbarkeit des Umfelds erfasst, sofern die Veränderung des Umfelds Einfluss auf Funktion oder Sicherheit des Systems haben kann (z. B. Außentemperatur, nichttrainierte Objekte).

5.3.1.1 Veränderbarkeit und Produktsicherheitsrecht

Für die rechtliche Bewertung der einzelnen Taxonomiedimensionen sollen zunächst das **Produktsicherheitsgesetz (ProdSG)**[142] und die auf dessen Grundlage erlassene **Maschinenverordnung (9. ProdSV)**[143] herangezogen werden.
Für die Begutachtung sollen – wie auch bei der Erörterung der anderen Rechtsgebiete – die maßgeblichen Regelungen des Produktsicherheitsrechts zunächst erläutert werden, um dann auf die Probleme mit bestimmten Ausprägungen der jeweiligen Dimension hinzuführen.

5.3.1.1.1 Anwendungsbereich des ProdSG

Der Hersteller eines Produkts und die anderen Produktverantwortlichen nach ProdSG und den ProdSV sind über den Lebenszyklus des Produkts chronologisch die ersten Regelungsadressaten, weshalb die rechtliche Untersuchung hier beginnt.
Das Produktsicherheitsrecht im Allgemeinen und das ProdSG mit der 9. ProdSV im Besonderen bilden für die Gewährleistung der Sicherheit bis zum Zeitpunkt des Inverkehrbringens die wesentliche Rechtsmaterie. Daher seien hier kurz der Anwendungsbereich und die Bedeutung des ProdSG für die vorliegende Untersuchung dargestellt.

Als zentrales Gesetz des Produktsicherheitsrechts kommt dem ProdSG die Rolle eines **„allgemeinen Teils“** zu.[144] In § 1 Abs. 4 ProdSG wird der Vorrang speziellerer Regelungen klargestellt. Das ProdSG trifft also Regelungen für solche Produkte, die

[142] Gesetz über die Bereitstellung von Produkten auf dem Markt (Produktsicherheitsgesetz – ProdSG), Artikel 1 des Gesetzes vom 08.11.2011, BGBl. I S. 2178, 2179, 2012 I 131, zuletzt geändert durch Artikel 301 der Verordnung vom 19.06.2020, BGBl. I S. 1328.
[143] Neunte Verordnung zum Produktsicherheitsgesetz (Maschinenverordnung – 9. ProdSV), Artikel 1 der Verordnung vom 12.05.1993, BGBl. I S. 704; zuletzt geändert durch Artikel 19 des Gesetzes vom 08.11.2011, BGBl. I S. 2178.
[144] So die Gesetzesbegründung BT-Drucks. 15/1620, S. 23 zum GSPG.

nicht in einem Sondergesetz bzw. -verordnung geregelt sind. Zudem schließt es Regelungslücken des jeweils einschlägigen besonderen Produktsicherheitsrechts[145].
Das **funktionale Verhältnis von ProdSG und den ProdSV bzw. den Durchführungsgesetzen zu EU-Richtlinien für bestimmte Produktgruppen**[146] kann zudem als Zwei-Säulen-Modell bezeichnet werden, bei dem das ProdSG neben allgemeinen Regelungen, wie den Begriffsbestimmungen, insbesondere die verwaltungsverfahrensbezogenen Vorschriften zur Marktüberwachung enthält und die ProdSV die inhaltlichen Anforderungen an die jeweiligen Produkte sowie Regelungen zum Konformitätsbewertungsverfahren festlegen.[147]
Im Anwendungsbereich des ProdSG unterscheidet § 3 ProdSG den harmonisierten Bereich nach § 3 Abs. 1 ProdSG und den nichtharmonisierten Bereich nach § 3 Abs. 2 ProdSG. Der harmonisierte Bereich umfasst alle Produkte, die in die Anwendungsbereiche der Produktsicherheitsverordnungen (ProdSV) nach § 8 Abs. 1 ProdSG sowie der Durchführungsgesetze fallen. Diese dienen der Umsetzung der europäischen Harmonisierungsvorschriften für Produktgruppen oder spezifische Gefahren.
Europarechtliche Grundlage der 9. ProdSV ist die Richtlinie 2006/42/EG (**Maschinen-RL)**. Sie regelt sowohl materielle als auch formelle Anforderungen an das Inverkehrbringen und die Inbetriebnahme von Maschinen. Der § 8 Abs. 1 ProdSG ist die Verordnungsermächtigung zum Erlass der ProdSV zur Umsetzung solcher Richtlinien für bestimmte Produktgruppen.[148]
Der **sachliche Anwendungsbereich** des ProdSG erfasst nach § 1 Abs. 1 ProdSG die Bereitstellung, Ausstellung oder erstmalige Verwendung von Produkten im Rahmen einer Geschäftstätigkeit, also mit Bezug zu einer wirtschaftlichen Unternehmung, die jedoch nicht auf Gewinnerzielung gerichtet sein oder entgeltlich erfolgen muss.[149] Damit ist der Anwendungsbereich einerseits tätigkeitsbezogen, andererseits auf den Begriff des Produkts begrenzt. Dieser wird wiederum definiert in § 2 Nr. 22 ProdSG als Waren, Stoffe oder Zubereitungen, die in einem Fertigungsprozess hergestellt worden sind. Reine Softwareanwendungen sind nicht vom Produktbegriff erfasst.[150] Erfasst sind hingegen auch gebrauchte Produkte.[151]
Das ProdSG ist auf die hier untersuchten software-physischen Systeme anwendbar, da es sich jedenfalls um ein verkörpertes System handelt, wie oben bei 5.2.1 erläutert. Es handelt sich um ein Produkt im Sinne des § 2 Nr. 22 ProdSG.

5.3.1.1.2 Anwendungsbereich der 9. ProdSV und Bedeutung für die Untersuchung

Die vorliegende Untersuchung konzentriert sich aus einer praktischen und einer formellen Überlegung heraus auf die 9. ProdSV. Die von der 9. ProdSV erfassten Maschinen werden oft auch als Betriebsmittel oder in immissionsschutzrechtlich

---

[145] *Klindt*, in: ders., Produktsicherheitsgesetz, § 1 Rn. 81.
[146] Z. B. das PSA-Durchführungsgesetz oder das Gasgerätedurchführungsgesetz.
[147] *Klindt* in: ders., Produktsicherheitsgesetz, § 3 Rn. 14.
[148] *Kapoor/Klindt*, Das neue deutsche Produktsicherheitsgesetz (ProdSG), NVwZ 2012, 719, 720.
[149] *Gauger*, Produktsicherheit und staatliche Verantwortung, S. 88 f.
[150] Etwas anderes gilt für Medizinprodukte, siehe § 3 Nr. 1 Medizinproduktegesetz, wonach auch Software erfasst ist.
[151] Es sei denn, sie müssen vor ihrer Verwendung in Stand gesetzt oder wiederaufgearbeitet werden, sofern der Wirtschaftsakteur denjenigen, an den sie abgegeben werden, darüber unterrichtet, so § 1 Abs. 3 Nr. 2 ProdSG.

geregelten Anlagen eingesetzt, sodass Maschinen als Untersuchungsgegenstand die Betrachtung des Produktsicherheitsrecht, des Rechts des technischen Arbeitsschutzes und des Immissionsschutzrechts ermöglichen. Die 9. ProdSV kann zudem formell als exemplarisch für den gesamten harmonisierten Bereich herangezogen gelten, da sich die europäischen Harmonisierungsvorschriften für die verschiedenen Produktgruppen im Aufbau ähneln. Für eine möglichst einheitliche Gestaltung der Harmonisierungsvorschriften wurde als Teil des „New Legislative Framework" der **Beschluss Nr. 768/2008**[152] gefasst. Er legt ein Muster für die einzelnen Harmonisierungsvorschriften fest, nach dem diese gestaltet werden. Als bloße politische Handlungsanweisung[153] an den europäischen Gesetzgeber entfaltet er gegenüber den Mitgliedstaaten oder Privaten keine Bindungswirkung.[154] Allerdings orientiert sich der europäische Gesetzgeber bei der Gestaltung der Harmonisierungsvorschriften an diesem Beschluss, sodass eine entsprechende Vereinheitlichung erfolgt.

Der **Anwendungsbereich** der 9. ProdSV konkretisiert den Anwendungsbereich des ProdSG hinsichtlich der Anforderungen der Maschinen-RL. Hier gilt wieder das Rangverhältnis ProdSG und ProdSV: Regelt die spezielle 9. ProdSV etwas nicht abschließend, greift das ProdSG. Der Anwendungsbereich der 9. ProdSV kann zudem nicht über den des ProdSG hinaus gehen.

Der **sachliche Anwendungsbereich** umfasst die Tätigkeiten Inverkehrbringen und Inbetriebnahme.

Inverkehrbringen ist die erstmalige Bereitstellung eines Produkts auf dem Markt, wozu auch die Einfuhr eines neuen Produkts in den Europäischen Wirtschaftsraum zählt, so § 2 Nr. 15 ProdSG.

Inbetriebnahme ist die erstmalige bestimmungsgemäße Verwendung einer Maschine, so § 2 Nr. 9 der 9. ProdSV.

Die 9. ProdSV gilt demnach **nur für neue Produkte**, gebrauchte Produkte sind nicht erfasst.

Da das ProdSG die Grenze des Anwendungsbereichs der 9. ProdSV mitbestimmt, ist entsprechend § 1 Abs. 1 S. 1 ProdSG ein *geschäftsmäßiges* Inverkehrbringen oder eine *geschäftsmäßige* Inbetriebnahme erforderlich.

Erfasst sind in sachlicher Hinsicht die in § 1 Abs. 1 der 9. ProdSV aufgelisteten Produkte. Dies sind einerseits die Maschinen *im weiteren Sinne*, die dort in den Nr. 1 bis 6 aufgezählt werden. Dazu gehören insbesondere Maschinen *im engeren Sinne* gemäß § 1 Abs. 1 Nr. 1 der 9. ProdSV. Solche Maschinen *im engeren Sinne* sind definiert in § 2 Nr. 2 der 9. ProdSV.

Vom **persönlichen Anwendungsbereich** sind der **Hersteller und sein Bevollmächtigter**[155] erfasst.

Hersteller im Sinne der 9. ProdSV ist nach § 2 Nr. 10 S. 1 der 9. ProdSV wer die Maschine konstruiert oder baut und für die Übereinstimmung der Maschine mit der 9.

---

[152] Beschluss Nr. 768/2008/EG des Europäischen Parlaments und des Rates vom 9. Juli 2008 über einen gemeinsamen Rechtsrahmen für die Vermarktung von Produkten und zur Aufhebung des Beschlusses 93/465/EWG des Rates, ABl. L 218 vom 13.8.2008, S. 82–128.

[153] Sog. Sui-generis-Beschluss, vgl. Bekanntmachung der *Kommission* – Leitfaden für die Umsetzung der Produktvorschriften der EU (sog. „Blue Guide"), Abl. C 272 vom 26.07.2016, Nr. 1.1.1.2.

[154] *Gauger,* Produktsicherheitsrecht, S. 81.

[155] Sofern im Gutachten nur von Hersteller gesprochen wird, sind damit alle nach der jeweiligen ProdSV bzw. dem ProdSG verantwortlichen Marktakteure gemeint, sofern sie die gleichen Pflichten treffen wie den Hersteller.

ProdSV im Hinblick auf das Inverkehrbringen unter ihrem eigenen Namen oder für den Eigengebrauch verantwortlich ist.

5.3.1.1.3 Problem: Neues Produkt durch Veränderung im Betrieb?

Damit kann auch die Konstruktion von Maschinen für den Eigengebrauch oder eine wesentliche Veränderung der Maschine zu Herstellerpflichten der 9. ProdSV führen.

> *Bei KI-Systemen, die sich während der Nutzung durch bestimmungsgemäße Aneignung neuer Datensätze, Updates, Teaching durch den Verwender etc. verändern, kann dies zu Herstellerpflichten führen. Voraussetzung ist, dass es sich um ein – im Gegensatz zur bisherigen Maschine – neues Produkt handelt.*

Tritt nun eine solche sicherheitsrelevante Änderung der Maschine ein, ist also zu prüfen, ob ein neues Produkt vorliegt.

Wann eine neue Maschine vorliegt, hängt vom Einzelfall ab, es gibt keine gesetzliche vorgegebenen Kriterien zur Beurteilung dieser Frage.[156] Für den harmonisierten Bereich stellt die Kommission im sog. „Blue Guide“ für die wesentliche Veränderung von Produkten fest, dass ein neues Produkt vorliegen *kann*, wenn erhebliche Veränderungen oder Überarbeitungen mit dem Ziel der Modifizierung seiner ursprünglichen Leistung, Verwendung oder Bauart vorgenommen worden sind, die sich wesentlich auf die Einhaltung der Harmonisierungsrechtsvorschriften der Union auswirken.[157] Das BMAS hat zu diesem Thema ein Interpretationspapier veröffentlicht. Demnach ist ausgehend von der bestehenden Risikobeurteilung für die konkrete Maschine zunächst zu prüfen, ob neue Gefährdungen oder Risiken entstanden sind, für die die vorhandenen Schutzmaßnahmen nicht mehr ausreichend oder geeignet sind. Dann ist eine neue Risikobeurteilung durchzuführen.[158] Ergibt diese neue Risikobeurteilung, dass ohne zusätzliche Schutzmaßnahmen die Maschine nicht mehr sicher ist und eine Risikominderung nicht durch einfache Schutzeinrichtungen erreicht werden kann, liegt eine wesentliche Veränderung vor und es gelten die Regelungen für neue Maschinen.[159]

> *Wer also gewerbsmäßig eine Maschine verwendet, bei der systeminhärente Veränderungen im Betrieb möglich sind, die über die Anpassung weniger Parameter ohne Sicherheitsrelevanz hinausgehen, muss bei Eintritt solcher Veränderungen eine neue Risikobeurteilung anstellen und ggf. alle Anforderungen der 9. ProdSV für die Inbetriebnahme neuer Maschinen beachten. Der Verwender kann so unversehens zum Hersteller werden.*
>
> *Anders wirkt sich diese Veränderung im ProdSG aus. Der § 1 Abs. 1 ProdSG erfasst zwar auch die erstmalige Verwendung, aber für den nichtharmonisierten Bereich stellt § 3 Abs. 2 S. 1 ProdSG keine Anforderungen an die erstmalige Verwendung, sondern die Bereitstellung am Markt. Die Veränderung des Produkts bei Verwendung löst im nichtharmonisierten Bereich also keine Herstellerpflichten, solange es danach nicht auf dem Markt bereitgestellt wird.*

Solange eine Veränderbarkeit des Systems von der Risikobeurteilung durch den Hersteller *abschließend* erfasst ist, es sich also nur um graduelle Veränderungen z. B.

---

[156] *Kommission*, Leitfaden für die Anwendung der Maschinenrichtlinie 2006/42/EG, § 72.
[157] *Kommission* – Leitfaden für die Umsetzung der Produktvorschriften der EU („Blue Guide“), Abl. C 272 vom 26.07.2016, Nr. 2.1.
[158] *BMAS*, Interpretationspapier „Wesentliche Veränderung von Maschinen“, S. 3.
[159] *BMAS*, Interpretationspapier „Wesentliche Veränderung von Maschinen“, S. 6.

durch gezieltes Teaching handelt, für die die Schutzmaßnahmen ausreichen, entsteht kein neues Produkt bzw. eine neue Maschine.

5.3.1.1.4 Materielle und formelle Voraussetzung nach der 9. ProdSV

Um eine Maschine in Verkehr bringen oder in Betrieb nehmen zu können, muss der Hersteller die Anforderungen des § 3 Abs. 1 ProdSG in Verbindung mit der für Maschinen gültigen 9. ProdSV erfüllen. Das Konformitätsbewertungsverfahren, die Ausstellung der Konformitätserklärung, die Anbringung der CE-Kennzeichnung und die Zusammenstellung der technischen Unterlagen sowie der Bedienungsanleitung muss vor dem maßgeblichen Zeitpunkt des Inverkehrbringens bzw. der Inbetriebnahme erfolgen.
Für KI-Systeme, die erst am Einsatzort, z. B. im Betrieb, installiert, in die Arbeitsumgebung integriert und abschließend konfiguriert werden müssen, können die Konformitätsbewertung, die Anbringung der CE-Kennzeichnung etc. erst nach diesen Arbeitsschritten abschließend erfolgen.
Für die rechtliche Beurteilung des Merkmals der Veränderbarkeit sind insbesondere die besonderen Sicherheitsanforderungen nach Anhang I der Maschinen-RL mit der **Risikobeurteilung** und das **Konformitätsbewertungsverfahren** § 4 der 9. ProdSV von Interesse. Zudem soll ein Blick auf die Bedeutung der technischen Normen und dort insbesondere der sog. harmonisierten Normen geworfen werden.

5.3.1.1.4.1 Spezielle Sicherheitsanforderungen und Risikobeurteilung

Die materiellen Anforderungen an Maschinen stellt § 3 Abs. 2 Nr. 1 der 9. ProdSV, der auf Anhang I der Maschinen-RL verweist, wo wiederum detaillierte Anforderungen gelistet sind, die Sicherheit und Gesundheitsschutz gewährleisten sollen.
Die Sicherheits- und Gesundheitsschutzanforderungen des Anhang I der Maschinen-RL sind auf die Vermeidung von Verletzungen und Gesundheitsschäden gerichtet, so Nr. 1.1.1 lit. a) Anhang I der Maschinen-RL.
Hier zeigt sich der regulatorische Ansatz des „Neuen Konzepts“: Der Gesetzgeber legt die zu erreichenden Ziele fest, wie der Hersteller diese erreicht, bleibt ihm überlassen.[160]
Der Anhang I der Maschinen-RL gibt dem Hersteller jedoch Grundsätze an die Hand, die er bei Konstruktion und Bau seiner Maschinen berücksichtigen muss. Der Gesetzgeber gibt hier auch Anweisungen, *wie* durch Verfahren die Sicherheit gewährleistet werden soll.
Zentral ist der erste Grundsatz, die **Risikobeurteilung.** Der Anhang I der Maschinen-RL stellt verschiedene Kriterien und Anforderungen an die Beurteilung des Risikos bzw. die Konstruktion, Fabrikation, Inverkehrbringen etc. der Maschine auf. Die genaue Ausgestaltung der Risikobewertung kann auch den einschlägigen harmonisierten Normen entnommen werden.
Anhang I legt unter Nr. 1 zunächst die allgemeinen Anforderungen an Maschinen fest, während unter den Nr. 2 bis 6 dann für bestimmte Maschinentypen weitere besondere Anforderungen aufgeführt werden. Der Anhang I stellt damit das **Grundlagenwerk des sicherheitsgerechten Konstruierens** dar.[161]

---

[160] *Gauger,* Produktsicherheitsrecht, S. 60 f.
[161] *Klindt*, in: ders., Produktsicherheitsgesetz, § 8, Rn. 32.

Die **Risikobeurteilung** erfolgt grundsätzlich entsprechend den Vorgaben des Anhangs I, Allgemeine Grundsätze Nr. 1 der Maschinen-RL. Demnach sind **fünf Schritte** vorgesehen:

- Bestimmung der Grenzen der Maschine;
- Gefährdungsermittlung;
- Risikoabschätzung, Einschätzung der ermittelten Risiken (Schwere der möglichen Schäden und Eintrittswahrscheinlichkeit);
- Risikobewertung der ermittelten Risiken hinsichtlich der Erforderlichkeit einer Risikominderung;
- Beseitigung oder Minderung der Risiken.

Im Anschluss an dieses iterative Verfahren ist dieses erneut durchzuführen, wenn im fünften Schritt Maßnahmen zur Beseitigung oder Minderung des ermittelten Risikos ergriffen wurden.
Zu den Schritten der Risikobeurteilung im Einzelnen:

- Die **Bestimmung der Grenzen der Maschinen** erfolgt auf Grundlage der *bestimmungsgemäßen Verwendung*. Es muss exakt definiert werden, wofür die Maschine verwendet werden soll, z. B. ist die maximale Traglast anzugeben.[162] Es ist auch die *vernünftigerweise erwartbare Fehlanwendung* zu berücksichtigen. Ob es sich um unabsichtliche oder absichtliche Fehlandwendungen handelt, ist nicht maßgeblich. Zur Ermittlung der bestimmungsfremden Anwendungen muss der Hersteller auch auf Erkenntnisse aus dem Betrieb von typenähnlichen Maschinen zurückgreifen. Bekannte typische Fehlanwendungen sind dann in die Risikobeurteilung mit einzustellen.[163] Den Hersteller trifft also eine **indirekte Produktbeobachtungsobliegenheit** bei der Weiterentwicklung bestehender Maschinen. Wie weit diese geht, hängt jedoch vom Einzelfall ab. Insbesondere ist sie nicht direkt aus dem Gesetz abzuleiten.

  > *Maschinen, die KI-Systeme enthalten, sind zwar in der Lösungsfindung mitunter nicht völlig im Vorhinein festgelegt, werden sich jedoch nicht über die (physischen oder logischen) Grenzen der Maschine hinwegsetzen können. Trotzdem ist für die weiteren Schritte der Risikobeurteilung die Bestimmung dieser Grenzen maßgebend. Das gilt besonders dann, wenn es sich um (vermeintlich) frei oder selbstständig handelnde Systeme dreht.*

- Die **Ermittlung der Gefährdungen**, die von der Maschine ausgehen können, erfolgt innerhalb der zuvor bestimmten Grenzen der Maschine.

  „Gefährdung" ist eine potenzielle Quelle von Verletzungen oder Gesundheitsschäden (Nr. 1.1.1 lit. a) Anhang I der Maschinen-RL).

  Damit ist jede potenzielle Quelle gemeint. Erfasst sind auch solche Quellen, die nicht ohne Weiteres für Menschen zugänglich sind (weil sie z. B. im Inneren der Maschine liegen). Es kommt also nicht auf die Wahrscheinlichkeit des Eintritts von Verletzungen oder Gesundheitsschäden an. [164]

---

[162] *Kommission*, Leitfaden für Anwendung der Maschinenrichtlinie 2006/42/EG, § 171.
[163] *Kommission*, Leitfaden für Anwendung der Maschinenrichtlinie 2006/42/EG, § 172.
[164] *Kommission*, Leitfaden für Anwendung der Maschinenrichtlinie 2006/42/EG, § 164.

- Zur **Abschätzung der ermittelten Risiken** bedarf es der Ermittlung von Schwere und Wahrscheinlichkeit der möglichen Verletzung.

  „Risiko“ ist die Kombination aus der Wahrscheinlichkeit und der Schwere einer Verletzung oder eines Gesundheitsschadens, die in einer Gefährdungssituation auftreten können (Nr. 1.1.1 lit. e) Anhang I der Maschinen-RL).

  Es kommt dabei auf die möglichen *Gefährdungssituationen* an, z. B. Kontakt von Menschen mit beweglichen Teilen der Maschine. Entscheidend ist stets der Einzelfall. Es sind alle innerhalb der im ersten Schritt ermittelten Grenzen der Maschine denkbaren Gefährdungssituationen über die verschiedenen Lebensphasen der Maschine zu berücksichtigen.[165] Es ist also die **gesamte Lebensdauer der Maschine** in den Blick zu nehmen. Dabei können auch **Informationen über in Betrieb befindliche ähnliche Maschinen** von Relevanz sein. Entwickelt ein Hersteller seine bereits vertriebene Maschine also weiter, sind bekannte Mängel, Unfälle mit der Maschine etc. in die Gefährdungsbeurteilung neuer Maschinen einzubeziehen.

  > *Diese Anforderung an die Ermittlung möglicher Gefährdungssituationen wird herausfordernder, je weiter die Grenzen der Maschine sind. Bei Maschinen, die KI-Systeme enthalten, die sich im Betrieb anpassen können, gilt dies umso mehr.*

- Die **Bewertung der Risiken** bedeutet, das zuvor ermittelte Risiko in Relation zu dem erwarteten Nutzen der Maschine zu setzen. Hierbei geht es um die Ermittlung des **Grenzrisikos**, also des noch akzeptablen Risikos. Maßgeblich ist dabei, ob sich die Person, die sich dem Risiko ausgesetzt sieht, diesem freiwillig oder unfreiwillig ausgesetzt hat. Wer sich freiwillig einem Risiko aussetzt, wird auch ein höheres Grenzrisiko hinnehmen. Unfreiwillige Exposition erfahren z. B. Arbeitnehmer im Betrieb bei der Arbeit an Maschinen. Auch können Faktoren wie Ausbildung der Verwender und besondere Empfindlichkeit der Verwender zu berücksichtigen sein.[166]

  > *Auch hier werden Maschinen, die veränderbare KI-Systeme enthalten, den Herstellern eine umfangreiche Untersuchung abverlangen. Die Gewichtung des Risikos fällt dann umso schwerer, wenn nicht klar ist, ob im Arbeitsablauf ein Vorgehen ergibt, das riskant ist. Beispielsweise kann es im System zu nicht vorhersehbaren Verhaltensverzweigungen kommen, die unterschiedlich riskant sind.*

- Zur **Beseitigung oder Minderung** dieser Risiken wird der Hersteller entsprechend Nr. 1.1.2 lit. b) Anhang I der Maschinen-RL in drei aufeinanderfolgenden Schritten vorgehen[167]:
  - Inhärent sichere Konstruktion;
  - technische und ergänzende Schutzmaßnahmen;
  - Benutzerinformationen.

  Vorrangig ist also eine entsprechend insgesamt inhärent sichere Konstruktion. Soweit diese nicht möglich ist, kommen auf der zweiten Stufe Schutzmaßnahmen z. B. an der Maschine und dann erst Benutzerinformationen in Betracht. Welche Maßnahmen für das konkrete Risiko in Betracht kommen, hängt nicht zuletzt von der Benutzergruppe

[165] *Kommission*, Leitfaden für Anwendung der Maschinenrichtlinie 2006/42/EG, § 168.
[166] *BAuA,* Risikobeurteilung im Maschinenbau, S. 16.
[167] *BAuA,* Risikobeurteilung im Maschinenbau, S. 20.

ab, für die die Maschine bestimmt ist bzw. von der eine Nutzung der Maschine zu erwarten ist. Hier werden also wieder die eingangs festgelegten **Grenzen der Maschine** relevant.[168] Verbleiben **Restrisiken**, die sich nicht beseitigen lassen, sind in der **Betriebsanleitung anzugeben**.

Bei der Konstruktion der Maschine muss nach Nr. 1.1.6 Anhang I der Maschinen-RL die **Ergonomie der Maschine** beachtet werden. Sie muss so konstruiert werden, dass bei der bestimmungsgemäßen Verwendung **Belästigung**, **Ermüdung** sowie **körperliche und psychische Fehlbelastung** auf das **nötige Mindestmaß reduziert** werden. Dabei sind an der Mensch-Maschine-Schnittstelle die voraussehbaren Eigenschaften des Bedienpersonals zu beachten.

> *Bei kollaborierenden Robotern, die KI-Systeme enthalten, ist also besonders auf eine vorausschauende Programmierung und ggf. auf technische und ergänzende Schutzmaßnahmen zu setzen, um den Anforderungen an die Ergonomie auch bei hochgradig veränderbaren Systemen gerecht zu werden.*[169]

5.3.1.1.4.2 Die Vermutungswirkung bei Verwendung harmonisierter Normen

Dieses allgemeine Schema der Risikobeurteilung wird durch verschiedene **harmonisierte Normen** konkretisiert. Der Hersteller kann dann das einschlägige Normenwerk anwenden und die Risikobeurteilung entsprechend durchführen. Auf diese Weise kann die **Vermutung begründet** werden, dass die Anforderungen an die Risikobeurteilung erfüllt wurden. Diese Vermutungswirkung gilt jedoch nur für solche Sicherheits- und Gesundheitsschutzanforderungen, die durch die jeweilige harmonisierte Norm abgedeckt werden. Dies folgt aus § 3 Abs. 5 der 9. ProdSV und findet sich letztlich in der hier umgesetzten Maschinen-RL, dort Art. 7 Abs. 2., sowie in § 4 Abs. 2 ProdSG.

Zur **Harmonisierung der Normsetzung** auf europäischer und mitgliedstaatlicher Ebene wurde die Verordnung (EU) Nr. 1025/2012[170] erlassen. Da sie als Verordnung auch in den Mitgliedstaaten unmittelbare Wirkung entfaltet, werden hier die dort in Art. 2 aufgestellten Definitionen der unterschiedlichen Normtypen verwendet. Norm ist nach Art. 2 Nr. 1 VO 1025/2012 eine von einer anerkannten Normungsorganisation[171] angenommene technische Spezifikation zur wiederholten oder ständigen Anwendung, deren Einhaltung nicht zwingend ist. Auf Ebene der EU wird der Normbegriff weiter differenziert: Nach Art. 2 Nr. 1 litt. b und c VO 1025/2012 ist eine europäische Norm eine Norm, die von einer europäischen Normungsorganisation angenommen wurde; sie wird zur harmonisierten Norm, wenn sie auf der Grundlage eines Auftrags der Kommission zur Durchführung von Harmonisierungsrechtsvorschriften der Union angenommen wurde. Die (rechtlich unverbindlichen) europäischen und harmonisierten

---

[168] *BAuA,* Risikobeurteilung im Maschinenbau, S. 23 f.

[169] *Kohte*, Arbeitsschutz in der digitalen Arbeitswelt, NZA 2015, 1417, 1419.

[170] Verordnung (EU) Nr. 1025/2012 des Europäischen Parlaments und des Rates vom 25. Oktober 2012 zur europäischen Normung, zur Änderung der Richtlinien 89/686/EWG und 93/15/EWG des Rates sowie der Richtlinien 94/9/EG, 94/25/EG, 95/16/EG, 97/23/EG, 98/34/EG, 2004/22/EG, 2007/23/EG, 2009/23/EG und 2009/105/EG des Europäischen Parlaments und des Rates und zur Aufhebung des Beschlusses 87/95/EWG des Rates und des Beschlusses Nr. 1673/2006/EG des Europäischen Parlaments und des Rates (VO 1025/2012), ABl. L 316 vom 14.11.2012, S. 12.

[171] CEN, CENELEC, ETSI, vgl. Anhang I der VO 1025/2012.

Normen konkretisieren die (rechtlich verbindlichen) Sicherheitsanforderungen in den Harmonisierungsrichtlinien.[172]
Hier zeigt sich dieses **Zusammenspiel aus hoheitlichem Recht, und technischer Normung durch private Normungsgremien**. Die Maschinen-RL gibt zwar einen umfangreichen Katalog an Sicherheitsanforderungen vor, wie diese konkret technisch dargestellt werden sollen, lässt sie jedoch offen. Die Konkretisierung der Anforderungen erfolgt dann durch technische Normen.[173]
Bei der **Risikobeurteilung** kann grundsätzlich zwischen **deduktiven** und **induktiven Ansätzen** unterschieden werden. Ein deduktiver Ansatz ist z. B. die Fehlerbaumanalyse. Diese verfolgt dabei jeweils eine (ungewollte) Folge im Sinne einer Gefährdung auf alle möglichen Ursachen zurück. Ein induktiver Ansatz ist die Ergebnisablaufanalyse, wobei ausgehend von der Ursache die möglichen Folgen ermittelt werden.[174]
Diese **gesetzliche Vermutungswirkung** des § 3 Abs. 5 der 9. ProdSV besteht nur bei der Einhaltung von **harmonisierten Normen** und auch nur **soweit diese eine Regelung** zu dem jeweiligen Schritt der Risikobeurteilung treffen. Es ist daher genau zu prüfen, ob die herangezogene Norm tatsächlich den Sicherheitsaspekt regelt, der durch den Hersteller erfüllt werden muss. Dabei sind alle verfügbaren Informationen über die anzuwendende harmonisierte Norm heranzuziehen.[175]
Die Anwendung von **harmonisierten Normen für ähnliche Maschinen** oder Sicherheitsaspekte zum Nachweis der Erfüllung der Sicherheitsanforderungen ist daher nicht möglich.
Handelt es sich bei der herangezogenen technischen Norm **nicht um eine harmonisierte Norm**, so besteht **keine Vermutungswirkung**. Der Hersteller muss dann im Einzelnen nachweisen, dass die konkrete Sicherheitsanforderung erfüllt ist.

5.3.1.1.4.3 Konformitätsbewertungsverfahren

Zur Feststellung der Konformität der Maschine mit den materiellen Sicherheitsanforderungen ist ein **Konformitätsbewertungsverfahren** durchzuführen. Die Konformität hat der Hersteller grundsätzlich in eigener Verantwortung durchzuführen. Welches Verfahren dabei zum Einsatz kommen kann, regelt die Maschinen-RL in Abhängigkeit des von dem jeweiligen Maschinentyp ausgehenden Risikopotenzials, sodass für bestimmte Typen von Maschinen andere Konformitätsbewertungsverfahren vorgeschrieben sind. 176 Diese Typen und die jeweils vorgesehen Konformitätsbewertungsverfahren sind in Anhang IV der Maschinen-RL geregelt, weshalb diese Maschinen als **Anhang IV-Maschinen** bezeichnet werden.
Das Konformitätsbewertungsverfahren richtet sich nach § 4 Abs. 2 bis 4 der 9. ProdSV. Von dort wird auf die Anhänge der Maschinen-RL verwiesen, in denen das durchzuführende Verfahren beschrieben ist:

> *Da für KI-Systeme häufig noch keine harmonisierten Normen bestehen, wird nach § 4 Abs. 4 der 9. ProdSV regelmäßig auf das EG-Baumusterprüfverfahren*

---

[172] *Gauger,* Produktsicherheitsrecht, S. 139.
[173] *Gauger*, Produktsicherheitsgesetz, S. 139.
[174] *BAuA*, Risikobeurteilung im Maschinenbau, S. 11.
[175] *Landessauschuss für Arbeitsschutz und Sicherheitstechnik,* Leitlinie zum Produktsicherheitsgesetz, 4/1, S. 17.
[176] Erwägungsgrund 20 der Maschinen-RL.

und das Verfahren zur umfassenden Qualitätssicherung zurückgegriffen werden.

Das **Verfahren zur umfassenden Qualitätssicherung** sieht die Einrichtung eines Systems zur Gewährleistung der Erfüllung der Sicherheits- und Gesundheitsschutzanforderungen durch die Maschinen auf Seiten des Herstellers vor. Dazu gehören nach Nr. 2.2 Abs. 2 dritter und fünfter Gedankenstrich Anhang X der Maschinen-RL auch **Konstruktionsprüfungs- und Verifizierungsverfahren** sowie **vor, während und nach der Konstruktion durchzuführende Prüfungen**. Dieses Qualitätssicherungssystem wird von einer **benannten Stelle** zugelassen und dessen Einhaltung regelmäßig überprüft.

Bei der Durchführung der erforderlichen Konformitätsbewertungsverfahren können **harmonisierte Normen** berücksichtigt werden. Auch hier gibt es eine **Konformitätsvermutung**. So sieht Nr. 2.3 Abs. 2 Anhang X der Maschinen-RL die Vermutung vor, dass das Qualitätssicherungssystem den Anforderungen der Nr. 2.2 Anhang X der Maschinen-RL entspricht, wenn es den einschlägigen harmonisierten Normen entspricht.

5.3.1.1.4.4 Maßgeblicher Zeitpunkt

Die materiellen und formellen Anforderungen an die Maschine bilden die Voraussetzungen für das Inverkehrbringen bzw. die Inbetriebnahme der Maschine. Sie müssen also zunächst in dem Moment vorliegen, in dem die Maschine in den Verkehr gebracht oder in Betrieb genommen wird. Mit dem Inverkehrbringen bzw. der Inbetriebnahme der Maschine tritt eine Zäsur ein. Der Hersteller wird insoweit aus der Verantwortung entlassen, als dass er dann **keinen weiteren Produktbeobachtungspflichten aus der 9. ProdSV mehr** unterliegt.

Zu beachten ist jedoch, dass der maßgebliche Zeitpunkt bei der Inbetriebnahme ein anderer ist als beim Inverkehrbringen. Die Inbetriebnahme erfordert die erstmalige zweckmäßige Verwendung der Maschine. Dann liegt der maßgebliche Zeitpunkt für das Vorliegen der Konformität ggf. **nach der Montage, dem Einbau** oder einer **sonst für die Verwendung der Maschine erforderlichen Handlung**.[177]

Hiervon **scharf zu trennen** ist jedoch die Berücksichtigung der **gesamten Lebensdauer** der Maschine im Rahmen der **Risikobeurteilung**. Hier muss der Hersteller auch später auftretende Gefährdungssituationen mit in die Beurteilung einstellen, z. B. besondere Gefährdungssituationen bei der Demontage im Rahmen der Ausmusterung der Maschine, die durch abgenutzte Teile, Ablagerungen gefährlicher Stoffe oder ähnliches entstehen. Ebenso müssen auf allen Ebenen der Risikobeurteilung Erfahrungen mit ähnlichen in Betrieb befindlichen Maschinen berücksichtigt werden. Es geht letztlich darum, den **Stand der Technik** als Maßstab anzulegen, und dabei kann nicht ignoriert werden, was bei in Betrieb befindlichen Maschinen an Erfahrungen gesammelt wird. Dieser Stand der Technik findet auch Ausdruck in den technischen Normen, die als harmonisierte Normen auch die dargestellte Vermutungswirkung begründen.

---

[177] *Kommission*, Blue Guide, Nr. 2.5.

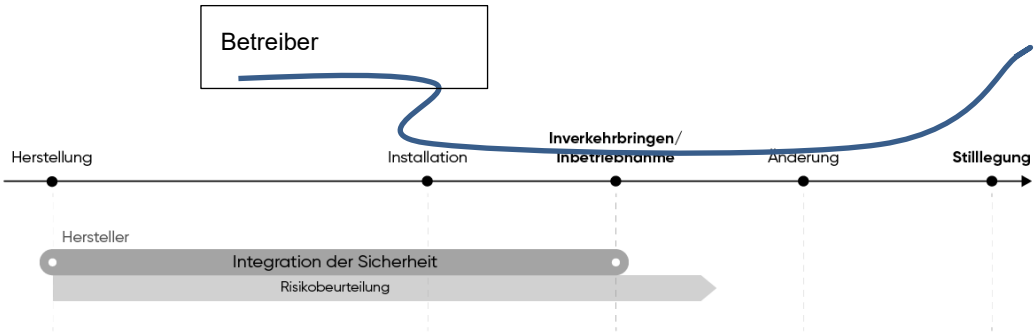


Abb. 5.1 Verantwortlichkeit des Herstellers nach der 9. ProdSV. Er ist nur bis zur Inbetriebnahme verpflichtet, die Maschine sicher zu gestalten. Die dabei maßgebliche Risikobeurteilung nimmt jedoch den gesamten Lebenszyklus der Maschine in den Blick.

Dabei ist zu bedenken, dass sich die Risikobeurteilung grundsätzlich auf die individuelle Maschine bezieht. Die Erfahrungen im Betrieb mit baugleichen Maschinen können jedoch in die Risikobeurteilung später zu fertigender Maschinen einfließen.

> *Bei KI-Systemen, die im Betrieb veränderbar bleiben, sind so für Risikobeurteilungen die Erfahrungen mit bereits in Betrieb befindlichen Systemen relevant. Hier stellt sich die Frage, wie der Stand der Technik zuverlässig ermittelt werden kann.*

Der Hersteller hat jedoch **gesetzliche Produktbeobachtungspflichten**, wenn es sich bei der Maschine um ein **Verbraucherprodukt** handelt. Dann muss er gemäß § 6 Abs. 3 ProdSG durch Stichproben, Beschwerdemanagement und Händlerinformationen auch nach Markteintritt der Maschine am Risikomanagement teilnehmen.[178]
Andere gesetzlich geregelte Produktbeobachtungspflichten finden sich beispielsweise in der Spielzeugverordnung (2. ProdSV)[179], wo gemäß § 3 Abs. 4 S. 3 der 2. ProdSV der Hersteller und gemäß § 6 Abs. 5 S. 1 der 2. ProdSV auch der Einführer Stichproben und Prüfungen der im Verkehr befindlichen Spielzeuge durchführen, ein Beschwerdemanagement einrichten und Händler regelmäßig über die Ergebnisse ihrer Produktüberwachung unterrichten müssen, sollten von dem Spielzeug Risiken ausgehen. Damit wird die hoheitliche Marktüberwachungsaufgabe teilweise auf die privaten Produktverantwortlichen übertragen, um sich dem Wissen insbesondere der Hersteller zu ihren Produkten und der Händler und Einführer zu den Vertriebsketten zunutze zu machen. So können Risiken besser erkannt und Maßnahmen zu ihrer Beseitigung effektiver gestaltet werden.[180]

Im **Business-to-Business-Bereich** bestehen diese **gesetzlichen Produktbeobachtungspflichten nicht**.

---

[178] *Eifert* in: Eifert, Produktbeobachtung durch Private, 9, 18.
[179] Verordnung über die Sicherheit von Spielzeug (2. ProdSV), Verordnung vom 07.07.2011, BGBl. I S. 1350, 1470 (Nr. 35); zuletzt geändert durch Artikel 1 der Verordnung vom 09.07.2018, BGBl. I S. 1093.
[180] *Appel* in: Eifert, Produktbeobachtung durch Private, 27, 33.

Vor dem Hintergrund der hoheitlichen Marktüberwachung kann es jedoch ratsam sein, auch ohne rechtliche Verpflichtung die Maschinen weiter zu beobachten, um Maßnahmen der Marktüberwachungsbehörden zuvor zu kommen.

5.3.1.1.4.5 Problem: Risikobeurteilung und Konformitätsbewertung bei weiterlernenden Systemen

Wenn die Konformitätsbewertung im maßgeblichen Zeitpunkt (Inverkehrbringen oder Inbetriebnahme) abgeschlossen wird, ist damit bestätigt, dass die Maschine in diesem Zeitpunkt und in diesem Zustand den Sicherheitsanforderungen entspricht. Bei veränderbaren Systemen besteht dann das Problem, dass das System als nicht mehr konform anzusehen ist, wenn wesentliche Änderungen vorgenommen werden oder auftreten. Es muss hier unterschieden werden, ob es sich um eine Veränderung im Sinne der bestimmungsgemäßen Verwendung bzw. der erwartbaren Fehlanwendung oder um eine außerhalb dieser Grenzen vorgenommene Veränderung handelt. Im erstgenannten Fall wird die Veränderung durch die Pflicht des Herstellers zur Integration der Sicherheit erfasst. Geht die Veränderung darüber hinaus, stellt sich die Frage, ob eine neue Maschine entstanden ist, für die erneut ein Konformitätsbewertungsverfahren etc. durchgeführt werden muss.[181]
Wenn das System selbst weiterlernt, also sicherheitsrelevante Parameter selbstständig ändert, dann stellt das den Hersteller bei der Risikobeurteilung und der Integration der Sicherheit vor große Herausforderungen: Die Bewertung der Risiken, für die auch die Wahrscheinlichkeit der Realisierung der ermittelten Gefährdungen maßgeblich ist, kann bei weiterlernenden Systemen nicht abschließend erfolgen. Jedenfalls können keine verifizierbaren Werte bestimmt werden. Eine Risikobeurteilung, die für die konkrete Maschine nicht über den Zeitpunkt der Inbetriebnahme hinaus verlässliche Angaben treffen kann, scheitert an der gesetzlichen Forderung, die Maschine den Anforderungen in Anhang I der Maschinen-RL entsprechend zu konstruieren und zu bauen. Solange mit den hergebrachten Methoden, also insbesondere mit technischen Normen, die die Erfahrungen der Praxis mit den einzelnen Maschinentypen und Gefährdungen dokumentieren, ermittelt werden kann, ob die Maschine die ermittelten Sicherheits- und Gesundheitsschutzanforderungen erfüllen wird, kann sich der Hersteller auf diese Methoden berufen. Wo aber die konkrete Maschine weiterlernt, steht der Hersteller wegen der Nachweispflicht vor dem Problem, dass seine Maschine mit Inbetriebnahme im Extremfall schon nicht mehr mit den Anforderungen der § 3 Abs. 1 ProdSG und § 3 der 9. ProdSV konform ist.
Der Hersteller muss im Rahmen der Risikobeurteilung nach dem Allgemeinen Grundsatz Nr. 1 in Anhang I der Maschinen-RL nach der Bestimmung der Grenzen der Maschine alle von der Maschine ausgehenden Gefährdungen ermitteln und darauf aufbauend das Risiko abschätzen. Risiko meint dabei die Kombination aus der Wahrscheinlichkeit und der Schwere einer Verletzung oder eines Gesundheitsschadens, die auftreten können (vgl. Nr. 1.1.1 lit. e) Anhang I der Maschinen-RL). Je größer die Ungewissheit hinsichtlich des künftigen „Verhaltens“ der KI-gesteuerten Maschine, desto weniger genau kann das Risiko abgeschätzt werden. Selbst wenn für die Risikobeurteilung im Rahmen des Konformitätsbewertungsverfahrens normierte Testverfahren für solche weiterlernenden Systeme bereitstünden, würden solche Systeme im Zeitpunkt der

[181] Vgl. dazu auch die Ausführungen oben bei 5.3.1.1.3

Inbetriebnahme nur begrenzt Aussagen über ihr künftiges „Verhalten“ im Betrieb erlauben. Eine hohe Ausprägung des Taxonomiemerkmals der Veränderbarkeit (bedeutende Veränderbarkeit) führt zu ungenaueren Aussagen im Zeitpunkt der Inbetriebnahme. Die dadurch bestehende Ungewissheit widerspricht jedoch dem zeitpunktbezogenen Konzept der Konformitätsbewertung. Der Hersteller kann bei bedeutend veränderbaren Systemen keine zweckmäßige Risikobeurteilung anstellen und auf dieser Grundlage sein Produkt nicht rechtskonform entwickeln. Auch ein Rückgriff auf technische Normen, die Test- und Simulationsverfahren vorsehen, hilft ab einem bestimmten Grad[182] der Veränderbarkeit nicht mehr. Denn der Verweis auf eine technische Norm kann nur insoweit die Vermutungswirkung begründen, dass die Anforderungen an die Sicherheit des Produkts erfüllt sind, wie die Norm diese Anforderungen auch abdeckt. Da das Gesetz einen Nachweis des Vorliegens der Sicherheitsanforderungen im maßgeblichen Zeitpunkt verlangt, können Testverfahren, die über das künftige „Verhalten“ des Produkts nur bedingt Aussagen treffen, auch nur eine bedingte Vermutungswirkung begründen.

5.3.1.1.5 Zwischenergebnis zu Veränderbarkeit und Produktsicherheitsrecht

- *Durch wesentliche Veränderung nach Inverkehrbringen bzw. Inbetriebnahme können ein neues Produkt und damit für den Verwender Herstellerpflichten entstehen (z. B. nach der 9. ProdSV, wenn die veränderte Maschine in Betrieb genommen wird).*
- *Eine abschließende Risikobeurteilung und eine entsprechende Konformitätsbewertung sind nicht möglich, wenn die künftigen Zustände des Produkts und damit die ihm innewohnenden Risiken im Zeitpunkt des Inverkehrbringens nicht mit hinreichend bestimmt werden können, weil spätere sicherheitsrelevante Änderungen möglich bleiben.*

Veränderbare Systeme, bei denen lediglich einzelne Parameter geändert werden sind hingegen nicht problematisch, da hier nicht von einer wesentlichen Veränderung ausgegangen werden muss. Zudem sind hier die Parameter klar, sodass es für die ordnungsgemäße abschließende Risikobeurteilung vor allem darauf ankommen wird, dass hinreichende Simulations- und Testverfahren zur Verfügung stehen.
Die hier identifizierten Probleme treten also nur bei bedeutend veränderbaren Systemen auf. Eine kritische Veränderbarkeit kann dann erreicht sein, wenn im Rahmen der Risikobeurteilung nach dem neuesten Stand der Technik wegen der Veränderbarkeit im maßgeblichen Zeitpunkt des Inverkehrbringens oder der Inbetriebnahme nicht abschließend bestimmt werden kann, ob und wenn ja welche Risiken (noch) bestehen. Wenn dann wegen des möglicherweise bestehenden Schadenspotenzials des Produkts diese Ungewissheit nicht tolerierbar ist, kann ein solches Produkt nicht produktsicherheitsrechtlich konform auf den Markt kommen bzw. in Betrieb genommen werden.

[182] Wann dieser Grad erreicht ist, lässt sich dem Gesetz jedoch nicht entnehmen.

5.3.1.2 Veränderbarkeit und Recht des technischen Arbeitsschutzes

Das Produktsicherheitsrecht bildet einen Teil des technischen Arbeitsschutzes. Im Arbeitsrecht treffen das **Arbeitsschutzgesetz (ArbSchG)** und die **Betriebssicherheitsverordnung (BetrSichV)** wesentliche Regelungen zur Gewährleistung der Sicherheit am Arbeitsplatz. Verpflichteter ist regelmäßig der Arbeitgeber. Die BetrSichV regelt den Einsatz von Arbeitsmitteln.

> *Die hier untersuchten KI-Systeme sind insbesondere auch solche, die am Arbeitsplatz zum Einsatz kommen, sei es als Teil einer industriellen Fertigungsanlage oder als Teil eines Roboters, der kollaborativ mit Menschen zusammenarbeitet. Es stellt sich daher die Frage, wie die Sicherheit der Arbeitnehmer am Arbeitsplatz gewährleistet wird.*

Im deutschen Recht existiert ein **duales System zur Gewährleistung des Arbeitsschutzes**, das einerseits aus staatlichen Regelungen besteht (Gesetze und Verordnungen), andererseits aus dem Satzungsrecht der selbstverwalteten Unfallversicherungsträger (Unfallverhütungsvorschriften).[183] Die staatlichen Regelungen finden sich insbesondere im Arbeitsschutzgesetz[184] und in den auf Grund dieses Gesetzes erlassenen Verordnungen, wie der Betriebssicherheitsverordnung[185]. Im weiteren Sinne gehört auch das oben dargestellte Produktsicherheitsrecht dazu, sofern im Arbeitsumfeld Produkte – wie die hier untersuchten KI-Systeme – zum Einsatz kommen. Es nimmt die Produktverantwortlichen in die Pflicht, Sicherheit im Produkt „mitzuliefern“ und leistet damit einen Beitrag zum Arbeitsschutz.[186]
Die Unfallverhütungsvorschriften der Unfallversicherungsträger, also der Berufsgenossenschaften und der Unfallkassen, sowie der DGUV als deren Spitzenverband regeln u.a. das Verhaltensrecht innerhalb von Betrieben, in denen Maschinen zum Einsatz kommen. Zudem stehen den Unfallversicherungsträgern hoheitliche Befugnisse gegenüber dem Arbeitgeber gemäß § 19 SGB VII[187] zu, um die Pflichten des Arbeitsgebers u.a. aus den Unfallverhütungsvorschriften durchzusetzen und besondere Unfall- und Gesundheitsgefahren abzuwehren.
Die vorliegende Untersuchung konzentriert sich auf das staatliche Recht des ArbSchG.

5.3.1.2.1 Arbeitsschutzgesetz und Betriebssicherheitsverordnung

Nach dem ArbSchG treffen den Arbeitgeber verschiedene Pflichten zur Gewährleistung der Sicherheit und des Gesundheitsschutzes am Arbeitsplatz. Hervorzuheben ist die Pflicht zur Gefährdungsbeurteilung aus § 5 Abs. 1 ArbSchG, aufgrund der der Arbeitgeber etwaige erforderliche Schutzmaßnahmen ermitteln muss.
Die Anforderungen des ArbSchG werden durch verschiedene Verordnungen auf Grundlage der §§ 18 und 19 ArbSchG konkretisiert. Für die Verwendung von Produkten im Betrieb ist die BetrSichV einschlägig.

---

[183] BT-Drucksache 13/3540, S. 12.
[184] Arbeitsschutzgesetz (ArbSchG), Artikel 1 des Gesetzes vom 07.08.1996, BGBl. I S. 1246; zuletzt geändert durch Artikel 113 des Gesetzes vom 20.11.2019, BGBl. I S. 1626.
[185] Betriebssicherheitsverordnung (BetrSichV), Artikel 1 der Verordnung vom 03.02.2015, BGBl. I S. 49 (Nr. 4); zuletzt geändert durch Artikel der 1 Verordnung vom 30.04.2019, BGBl. I S. 554.
[186] *Pauli* in: Kohte/Faber/Feldhoff, Handkommentar Arbeitsschutzrecht, BetrSichV Rn. 9.
[187] Siebtes Buch Sozialgesetzbuch (SGB VII), Artikel 1 des Gesetzes vom 07.08.1996, BGBl. I S. 1254; zuletzt geändert durch Artikel 35 des Gesetzes vom 12.12.2019, BGBl. I S. 2652.

Die BetrSichV regelt zur Gewährleistung der Sicherheit am Arbeitsplatz die Pflichten des Arbeitgebers hinsichtlich einer bestimmten Gefahrenquelle, nämlich der Arbeitsmittel. Die Verordnung kann als das „Grundgesetz des technischen Arbeitsschutzes“ bezeichnet werden.[188] Sie konkretisiert die Pflichten des Arbeitgebers aus dem ArbSchG und steht neben den anderen, ebenfalls auf Grundlage des § 18 Abs. 1 ArbSchG erlassenen Verordnungen.[189] Sie dient zudem als Verordnung gemäß § 19 ArbSchG der Umsetzung der Richtlinie 2009/104/EG über Arbeitsmittel[190] und trifft entsprechend § 34 ProdSG Regelungen für überwachungsbedürftige Anlagen. Es finden sich zudem Regelungen zu **überwachungsbedürftigen Anlagen** in §§ 15 – 18 BetrSichV. Sie gehen auf die Verordnungsermächtigung des § 34 Abs. 1 ProdSG zurück, sind also schon in formeller Hinsicht dem Produktsicherheitsrecht zuzuordnen. Hier zeigt sich die inhaltliche Verknüpfung von Produktsicherheitsrecht und Arbeitsschutzrecht.[191]
Als Arbeitsmittel erfasst sind gemäß § 2 Abs. 1 BetrSichV auch Maschinen, Geräte und Anlagen, die für die Arbeit verwendet werden, also auch die hier gegenständlich software-physischen KI-Systeme, wenn sie für die Arbeit verwendet werden.
Aus Sicht des Arbeitgebers beginnt die Auseinandersetzung mit dem Arbeitsmittel als Gefahrenquelle mit der Planung der Beschaffung und setzt sich fort über die Gefährdungsbeurteilung, weiter über die nach den dort gewonnenen Erkenntnissen erforderlichen Maßnahmen zum Schutz der Beschäftigten, über eine etwaige Prüfung vor der erstmaligen Verwendung bis hin zur Instandhaltung und Änderung des Arbeitsmittels sowie etwaiger wiederkehrender Prüfungen.
Für dieses Gutachten sollen kurz die **Gefährdungsbeurteilung**, die **erforderlichen Schutzmaßnahmen**, die **Änderung des Arbeitsmittels** und die **Prüfung vor erstmaliger Verwendung** dargestellt werden. Zudem wird kurz auf die Besonderheiten bei **überwachungsbedürftigen Anlagen eingegangen**.

5.3.1.2.2 Gefährdungsbeurteilung

In **§ 3 Abs. 1 S. 1 BetrSichV** wird der Arbeitgeber verpflichtet, vor Verwendung eines Arbeitsmittels eine **Gefährdungsbeurteilung** durchzuführen. Diese hat er gemäß § 3 Abs. 3 S. 1 BetrSichV auch schon vor Auswahl des Arbeitsmittels durchzuführen. Damit wird die Pflicht des Arbeitgebers aus § 5 ArbSchG zur Durchführung einer Gefährdungsbeurteilung für die Verwendung von Arbeitsmitteln konkretisiert. Die Gefährdungsbeurteilung gemäß § 3 Abs. 1 S. 1 BetrSichV überschneidet sich daher auch mit den in anderen Verordnungen zum ArbSchG geforderten Gefährdungsbeurteilungen.
Gefährdungsbeurteilung meint die systematische Ermittlung und Bewertung relevanter Gefährdungen der Beschäftigten mit dem Ziel, die erforderlichen Maßnahmen für Sicherheit und Gesundheit bei der Arbeit festzulegen.[192]

---

[188] *Wilrich*, Verantwortlichkeit und Pflichtenverteilung gemäß Betriebssicherheitsverordnung, NZA 2015, 1433.
[189] Z. B. die Arbeitsstättenverordnung und die Biostoffverordnung.
[190] Richtlinie 2009/104/EG des Europäischen Parlaments und des Rates vom 16. September 2009 über Mindestvorschriften für Sicherheit und Gesundheitsschutz bei Benutzung von Arbeitsmitteln durch Arbeitnehmer bei der Arbeit, ABl. L 260 vom 3.10.2009, S. 5.
[191] Sie dient zudem als Verordnung nach § 19 Abs. 1 ChemG. Auf diesen Regelungsbereich der BetrSichV wird im vorliegenden Gutachten nicht weiter eingegangen.
[192] *Gemeinsame Deutsche Arbeitsstrategie*, Leitlinie Gefährdungsbeurteilung und Dokumentation, S. 10.

Im Ergebnis muss der Arbeitgeber zum Schutz der Beschäftigten vor den mit ihrer Arbeit verbundenen Gefahren die erforderlichen Maßnahmen ergreifen, die er im Rahmen einer Gefährdungsbeurteilung ermittelt hat.[193]
Der Arbeitgeber hat gemäß § 3 Abs. 2 BetrSichV bei der Gefährdungsbeurteilung neben den von dem Arbeitsmittel ausgehenden Gefährdungen insbesondere die Umstände am Einsatzort des Arbeitsmittels zu berücksichtigen. Dazu gehören auch die von der Arbeitsumgebung und von den anderen Arbeitsmitteln ausgehenden Gefährdungen, mit denen das zu beurteilende Arbeitsmittel in Kontakt kommt oder interagiert.
Im Zusammenhang mit den hier untersuchten KI-Systemen als Arbeitsmittel sind zudem die § 3 Abs. 2 S. 2 Nr. 1 und 2 BetrSichV hervorzuheben. Demnach sind in der Gefährdungsbeurteilung insbesondere die **ergonomische Gestaltung des Arbeitsmittels** sowie **sicherheitsrelevante Zusammenhänge zwischen Arbeitsmittel einerseits und Arbeitsverfahren, -organisation, -ablauf, -zeit und -aufgabe andererseits** zu beachten.
Zur **Beschaffung der notwendigen Informationen** zur Gefährdungsbeurteilung kann der Arbeitgeber gemäß § 3 Abs. 4 S. 2 BetrSichV auf verschiedene Quellen zurückgreifen. Dazu zählen zunächst die gemäß § 21 Abs. 6 Nr. 1 BetrSichV vom Bundesministerium für Arbeit und Soziales bekannt gegebenen Regeln und Erkenntnisse des **Ausschusses für Betriebssicherheit**, die **Technischen Regeln für Betriebssicherheit (TRBS)**. Sie stellen den Stand der Technik, Arbeitsmedizin und Arbeitshygiene sowie gesicherte arbeitswissenschaftliche Erkenntnisse für die Verwendung von Arbeitsmitteln dar.
Für die Gefährdungsbeurteilung kann auf die **TRBS 1111** abgestellt werden. Demnach richtet sich die konkrete Ausgestaltung der Gefährdungsbeurteilung nach der Art des zu beschaffenden Arbeitsmittels. Dabei wird beispielhaft die Komplexität des Arbeitsmittels als maßgeblich für die Gefährdungsbeurteilung genannt.[194]
Die TRBS 1111 sieht zudem ein **Schema zur Durchführung der Gefährdungsbeurteilung** vor, das wie folgt gegliedert ist:

- Informationsbeschaffung über Verwendung und Beschaffenheit des Arbeitsmittels (z. B. aus Dokumenten des Herstellers);
- Gefährdungen ermitteln (unter Beachtung z. B. der Herstellerangaben);
- Gefährdungen bewerten (unter Beachtung des Standes der Technik, zu ermitteln aus spezifischen TRBS, DGUV-Regelwerken, Veröffentlichungen der Unfallversicherungsträger, der Länder, der BAuA oder ggf. unter Hinzuziehung von Experten);
- Festlegung der Schutzmaßnahmen (unter Beachtung der TOP-Rangfolge[195]);
- Umsetzung der Schutzmaßnahmen;
- Überprüfung der Wirksamkeit der Schutzmaßnahmen (§ 4 Abs. 5 BetrSichV);
- Dokumentierung der Ergebnisse (§ 3 Abs. 8 BetrSichV).

---

[193] *Wink* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, BetrSichV § 3 Rn. 2a.
[194] TRBS 1111, 4.2 Abs. 3 (Stand: März 2019).
[195] Erst technische, dann organisatorische, dann personenbezogene Schutzmaßnahmen.

Die Gefährdungsbeurteilung geht im Prinzip nach demselben Schema vor, wie die **Risikobeurteilung nach der Maschinen-RL**.
Dementsprechend kann der Arbeitgeber zur **Informationsbeschaffung** auf Informationen zurückgreifen, die der Hersteller mitliefert. Er darf nach § 3 Abs. 4 S. 4 BetrSichV dann davon ausgehen, dass diese Informationen zutreffend sind, sofern er nicht andere Erkenntnisse hat. So ermittelt der Hersteller im Rahmen der Risikobeurteilung die Grenzen der Maschine, die sich in der bestimmungsgemäßen Verwendung und der vorhersehbaren Fehlanwendung ausdrücken. Jedenfalls erstere ist in den technischen Unterlagen und der Betriebsanleitung dokumentiert. Weiter werden die nach Integration der Sicherheit nach ProdSG und der 9. ProdSV noch bestehenden Restrisiken in die Bedienungsanleitung aufgenommen.
Bei der **Bewertung der Gefährdung** ist der **Stand der Technik** maßgeblich. Auch hier findet sich eine Parallele zur Risikobewertung. Letztlich hat der Arbeitgeber auf eine Vielzahl von Informationsquellen zurückzugreifen, um durch die Gefährdungsbeurteilung letztlich die erforderlichen Maßnahmen zur Beseitigung der ermittelten Gefährdungen zu erarbeiten. Der Stand der Technik ist im Übrigen auch Maßstab für alle anderen aus der Gefährdungsbeurteilung gezogenen Konsequenzen, auf die im Folgenden noch einzugehen ist.

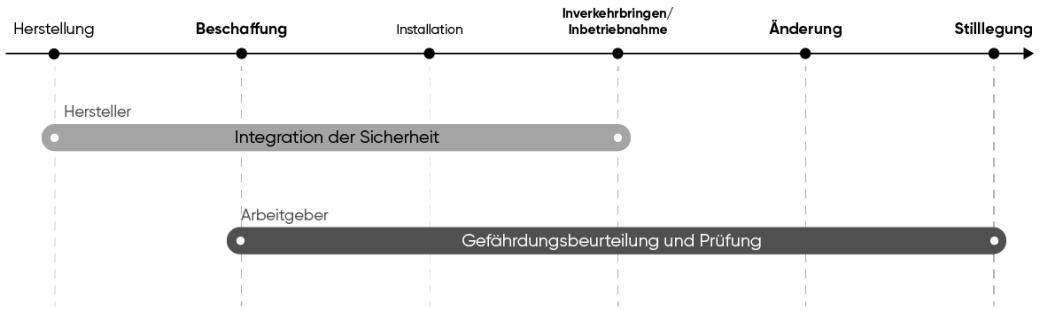


Abb. 5.2 Verantwortlichkeiten des Arbeitgebers (dunkel). Die Pflicht zur Gefährdungsbeurteilung deckt insbesondere die Zeit ab Inbetriebnahme ab, bei der der Hersteller einer Maschine nach der 9. ProdSV nicht mehr verantwortlich ist.

Gemäß § 3 Abs. 6 S. 1 BetrSichV hat der Arbeitgeber auf Grundlage der Gefährdungsbeurteilung zudem die **Fristen der wiederkehrenden Prüfungen** festzulegen, sofern sie erforderlich sind. Für diese wiederkehrenden Prüfungen hat er als Maßstab ebenfalls aufgrund der Gefährdungsbeurteilung den **Soll-Zustand** des Arbeitsmittels zu bestimmen. Der Soll-Zustand beschreibt den nach der Gefährdungsbeurteilung sicheren Zustand. Zu dessen Ermittlung sind auch etwaige Betriebs-, Bedienungs- und Gebrauchsanleitungen des Herstellers heranzuziehen.[196]
Entsprechend der kontinuierlichen Pflicht des Arbeitgebers zur Gewährleistung und Verbesserung der Sicherheit am Arbeitsplatz hat er die **Gefährdungsbeurteilung**

[196] TRBS 1111, 4.6 Abs. 1 (Stand: März 2019).

regelmäßig zu überprüfen und diese sowie die demnach erforderlichen Schutzmaßnahmen ggf. anzupassen, so § 3 Abs. 7 BetrSichV. Das ist insbesondere dann der Fall, wenn sicherheitsrelevante Ereignisse eintreten.[197] Nach der TRBS 1111 zählen zu diesen sicherheitsrelevanten Ereignissen z. B. Änderungen im Verfahren.[198] Die Überprüfung kann dabei auch ergeben, dass sich der **Soll-Zustand geändert hat**, was insbesondere bei den weiterlernenden Systemen der Fall sein kann.

5.3.1.2.3 Problem: Gefährdungsbeurteilung bei weiterlernenden Systemen

Bei veränderbaren KI-Systemen steht der Arbeitgeber bei der Gefährdungsbeurteilung vor besonderen Herausforderungen. Die ergonomische Gestaltung mit Menschen interagierender oder kollaborierender KI-Systeme zu beurteilen, kann sich als schwierig erweisen, wenn diese Systeme nach der Montage und Kalibrierung veränderbar bleiben und sich so an das Verhalten der mit ihnen interagierenden oder kollaborierenden Beschäftigten anpassen.

> *Bei adaptiven Systemen, die sich nur in wenigen Parametern ändern, mag dies durch eine entsprechende Gefährdungsbeurteilung praktisch beherrschbar sein. Bei einer höheren Stufe der Veränderbarkeit, also einem weiterlernenden System, kann in kurzen Zeitabständen eine Neubeurteilung erforderlich werden.*

Die Zusammenhänge zwischen den KI-Systemen als Arbeitsmittel und den Arbeitsverfahren etc. sind dann auch in der ersten Gefährdungsbeurteilung umfassend zu ermitteln, wovon auch die möglichen späteren Anpassungen durch das System selbst umfasst werden müssen.

> *Auch hier wird, wie bei der produktsicherheitsrechtlichen Risikobeurteilung durch den Hersteller, eine abschließende Gefährdungsbeurteilung bei zunehmender Veränderbarkeit des Systems immer schwieriger.*

Die **Gefährdungsbeurteilung** muss gemäß § 3 Abs. 7 BetrSichV insbesondere nach sicherheitsrelevanten Ereignissen **überprüft und ggf. aktualisiert** werden. Das ist bei wesentlichen Änderungen des Arbeitsmittels der Fall.

> *Bei veränderbaren Systemen wird also regelmäßig eine Überprüfung und ggf. Aktualisierung der Gefährdungsbeurteilung zu erfolgen haben. Bei weiterlernenden Systemen muss der Arbeitgeber das System engmaschig überwachen, um bei etwaigen Änderungen die Gefährdungsbeurteilung anpassen zu können.*

Auf Grundlage der Gefährdungsbeurteilung legt der Arbeitgeber außerdem fest, in welchem Intervall das Arbeitsmittel nach § 14 BetrSichV zu prüfen ist. Für diese **wiederkehrende Prüfung** bestimmt er den Soll-Zustand des Arbeitsmittels, der dann jeweils abgeprüft wird. Außerdem ist die Prüfung gemäß § 14 Abs. 3 S. 1 BetrSichV zu wiederholen, wenn prüfpflichtige Änderungen vorgenommen werden.

> *Bei KI-Systemen, die über ein hohes Maß an Veränderbarkeit verfügen, steht der Arbeitgeber also vor denselben Fragen, wie bei der Aktualisierung der Gefährdungsbeurteilung: Wann tritt eine prüfpflichtige Änderung bei einem weiterlernenden System ein? Muss dann nicht de facto dauerhaft geprüft werden?*

---

[197] *Wink* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, BetrSichV § 3 Rn. 7.
[198] TRBS 1111, 4.1 Abs. 3 (Stand: März 2019).

5.3.1.2.4 Schutzmaßnahmen

Der Arbeitgeber darf gemäß § 5 Abs. 1 bis 3 BetrSichV Arbeitsmittel nur dann zur Verfügung stellen und verwenden lassen, wenn sie sicher sind.
Dementsprechend darf der Arbeitgeber gemäß § 5 Abs. 3 S. 1 BetrSichV nur solche Arbeitsmittel zur Verfügung stellen und verwenden lassen, die den jeweils geltenden Rechtsvorschriften über Sicherheit und Gesundheitsschutz entsprechen. Dazu gehören nach § 5 Abs. 3 S. 2 BetrSichV auch die **Regelungen des jeweils einschlägigen Produktsicherheitsrechts**. Werden Maschinen als Arbeitsmittel verwendet, müssen sie also den Anforderungen von ProdSG, 9. ProdSV und Maschinen-RL entsprechen.
Hier zeigt sich, wie sich die **Pflichten der Produktverantwortlichen und des Arbeitgebers** für die Gewährleistung der technischen Sicherheit der Produkte bzw. Arbeitsmittel unterscheiden und ergänzen.
Die nach dem Produktsicherheitsrecht Verantwortlichen müssen bei der Integration der Sicherheit gemäß § 3 Abs. 1 Nr. 2 ProdSG **die bestimmungsgemäße oder vorhersehbare Verwendung des Produkts** und gemäß § 3 Abs. 1 der 9. ProdSV die **ordnungsgemäße Installation und Wartung** sowie die **bestimmungsgemäße Verwendung oder vorhersehbare Fehlanwendung** der Maschine berücksichtigen.
Der Arbeitgeber dagegen hat **alle konkreten Umstände der Verwendung des Arbeitsmittels** zu berücksichtigen. Seine Überwachungspflichten gehen also zunächst weiter als die der Produktverantwortlichen. Dies entspricht seiner Nähe zur Gefährdungsquelle, dem Arbeitsmittel, und dem daraus folgenden Informationsvorsprung gegenüber den Produktverantwortlichen.
Die entspricht auch dem historischen Verhältnis von Arbeitsschutzrecht und Produktsicherheitsrecht:
Das Produktsicherheitsrecht als Teil des Technikrechts ist aus dem technischen Arbeitsschutz entstanden. Allein bei den Pflichten der Arbeitgeber anzusetzen, durch Schutzmaßnahmen für die Sicherheit der Arbeitnehmer zu sorgen, reichte mit zunehmender Komplexität der eingesetzten Maschinen nicht mehr aus. Auch die Eingriffsbefugnisse der Gewerbeaufsicht und Berufsgenossenschaften konnten durch ihren repressiven Charakter nur im Einzelfall und bei akuten Gefahren wirken. Durch die Regulierung der technischen Geräte sollte direkt an der Gefahrenquelle angesetzt werden und der Wissensvorsprung der Hersteller genutzt werden.[199] Aus dieser Idee ist das heute bestehende Geflecht aus Produktsicherheit und Arbeitsschutz entstanden. Das Produktsicherheitsrecht ergänzt den Arbeitsschutz, der durch die Arbeitgeber zu gewährleisten ist, durch die produktbezogenen Pflichten der Produktverantwortlichen, insbesondere der Hersteller.[200]
Dass ein **CE-Kennzeichen** an das verwendete Arbeitsmittel angebracht ist, entbindet den Arbeitgeber gemäß § 3 Abs. 1 S. 2 BetrSichV nicht von der Pflicht, die Gefährdungsbeurteilung durchzuführen. Denn ein bei Inverkehrbringen im Sinne des Produktsicherheitsrechts sicheres Arbeitsmittel muss nicht unter den vom Arbeitgeber für dessen Einsatz im Betrieb vorgesehenen Umständen sicher sein. Die Art und Weise der konkreten Nutzung im Betrieb kann sich im Einzelfall von der durch den

[199] *Gauger,* Produktsicherheitsrecht, S. 42.
[200] *Pauli* in: Kohte/Faber/Feldhoff: Handkommentar Arbeitsschutzrecht, BetrSichV Rn. 9.

Hersteller bei der Risikobeurteilung nach Produktsicherheitsrecht vorausgesehenen unterscheiden.[201]
Erfüllt die Maschine die produktsicherheitsrechtlichen Anforderungen und sind nach der Gefährdungsbeurteilung weitere **Schutzmaßnahmen erforderlich**, hat der Arbeitgeber gemäß **§ 4 Abs. 3 S. 1 BetrSichV** bei der Ermittlung der erforderlichen Schutzmaßnahmen die **TRBS** zu beachten. Tut er dies, begründet dies gemäß § 4 Abs. 3 S. 2 BetrSichV die **Vermutung**, dass damit die Anforderungen der BetrSichV insoweit erfüllt sind.
Bei der Auswahl der erforderlichen Schutzmaßnahmen ist gemäß § 4 Abs. 2 S. 2 BetrSichV nach der **TOP-Rangfolge** vorzugehen. Zuerst sind also technische Maßnahmen am Arbeitsmittel selbst zu ergreifen. Reichen diese nicht aus, ist den Gefährdungen durch organisatorische Maßnahmen zu begegnen. Verbleiben dennoch Gefährdungen, kann auf personenbezogene Maßnahmen zurückgegriffen werden, wie beispielsweise persönliche Schutzausrüstung.
Der Arbeitgeber kann gemäß § 3 Abs. 5 BetrSichV zur Festlegung der erforderlichen Schutzmaßnahmen die Ergebnisse **vorangegangener Gefährdungsbeurteilungen**, insbesondere der **Risikobeurteilung** des Maschinenherstellers verwenden, soweit diese auf seinen Betrieb anwendbar sind.
Die **Ergebnisse** der Gefährdungsbeurteilung **vor der erstmaligen Verwendung** sind entsprechend § 3 Abs. 8 BetrSichV **zu dokumentieren**. Diese Dokumente sind einerseits eine wichtige Informationsquelle für spätere Prüfungen und weitere Gefährdungsbeurteilungen. Andererseits dienen sie zum Nachweis der Erfüllung arbeitsschutzrechtlicher Anforderungen an die Gefährdungsbeurteilung und die ergriffenen Schutzmaßnahmen gegenüber Aufsichtsbehörden.

5.3.1.2.4.1 Änderung von Arbeitsmitteln

In § 10 Abs. 5 BetrSichV werden die Pflichten des Arbeitgebers bei Änderung des Arbeitsmittels geregelt: Werden **Änderungen** an Arbeitsmitteln vorgenommen, ist § 10 Abs. 1 bis 3 BetrSichV zu beachten. Das bedeutet insbesondere, dass der Arbeitgeber die Änderungen unter Beachtung der Gefährdungsbeurteilung und der Betriebsanleitung durchzuführen hat.
Handelt es sich um **prüfpflichtige Änderungen**, ist eine Prüfung nach § 14 BetrSichV durchzuführen. Ob eine Änderung prüfpflichtig ist, hängt auch von der für das Arbeitsmittel durchgeführten Gefährdungsbeurteilung ab. Nach der **TRBS 1201** sind Änderungen prüfpflichtig, wenn sie z. B. Folgewirkung auf die Sicherheit des Arbeitsmittels haben oder neue Wechselwirkungen mit anderen Arbeitsmitteln, der Arbeitsumgebung oder den Arbeitsgegenständen, an denen Tätigkeiten mit Arbeitsmitteln durchgeführt werden, bewirken können.[202]
Der Arbeitgeber kann auch durch Änderung des Arbeitsmittels zum Hersteller werden. Voraussetzung dafür ist, dass **durch die Änderung ein neues Produkt entsteht**.[203]
Wenn er das Arbeitsmittel bzw. Produkt selbst weiter in seinem Betrieb nutzt oder verwenden lässt, so kann es sich nach § 1 Abs. 1, 3. Var. ProdSG um eine erstmalige Verwendung eines Produkts im Rahmen einer Geschäftstätigkeit handeln. Den Arbeitgeber treffen dann die **einschlägigen Herstellerpflichten**, er muss nach

[201] *Wink* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, BetrSichV § 3 Rn. 1b.
[202] TRBS 1201, 3.2.3 Abs. 1 (Stand: März 2019).
[203] *Wink* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, BetrSichV § 10 Rn. 5.

§ 5 Abs. 3 S. 2 BetrSichV jedenfalls die materiellen Sicherheitsanforderungen des einschlägigen harmonisierten Produktsicherheitsrechts erfüllen.

Nach dem Blue Guide kann ein Produkt, an dem erhebliche Veränderungen oder Überarbeitungen vorgenommen wurden, um die ursprüngliche Leistung, Verwendung oder Bauart zu verändern, kann als neues Produkt angesehen werden.[204] Wenn also durch die Veränderbarkeit eine solche Veränderung im Betrieb auftritt, könnte demnach ein neues Produkt vorliegen. Dem Zweck des Konzepts neuen Produkts entsprechend kommt es nicht darauf an, dass die Änderung nicht „vorgenommen" wird, wie im Blue Guide formuliert. Obwohl das Produkt sich selbst ändert, sollte eine Änderung in diesem Sinne angenommen werden. Verantwortlich wäre dann, in wessen Herrschaftsbereich das Produkt steht und wer es nutzt, wenn die Veränderung eintritt. Der Weiterbrieb nach Änderung stellt dann eine neue Inbetriebnahme dar. Ansonsten entstünde eine Schutzlücke. Denn das ProdSG wäre auf diese, vom Produkt selbst veranlasste, Änderungen nicht anwendbar. Ein Produkt, das gleichzeitig Arbeitsmittel ist, fällt dann zwar immer noch unter die Pflicht des Arbeitgebers, eine angepasste Gefährdungsbeurteilung vorzunehmen. Für ein derart verändertes Produkt, dass kein Arbeitsmittel ist, entstünden jedoch keine Herstellerpflichten, die Einhaltung der Sicherheitsanforderungen müsste nicht gewährleistet werden, der Schutzzweck des ProdSG wäre damit verfehlt. Daher sollte auch eine aus dem Produkt heraus veranlasste Änderung zu einem neuen Produkt führen können.
Handelt es sich bei dem neuen Arbeitsmittel um eine **Maschine**, sind also die materiellen Anforderungen der 9. ProdSV umzusetzen. Dann hat der Arbeitgeber bzw. Hersteller insbesondere die Anforderungen des § 3 der 9. ProdSV (Risikobeurteilung) zu erfüllen. 205

5.3.1.2.4.2 Problem: Wesentliche Änderung, die zu neuem Produkt führt

Für die Pflichten des Arbeitgebers ist hier bei der Veränderung durch kontrollierte Updates zu berücksichtigen, dass er nach § 13 ArbSchG für die Sicherheit im Betrieb stets zumindest mitverantwortlich ist. Auch wenn er die Veränderung des Arbeitsmittels an einen Dritten auslagert, der als Hersteller oder Dienstleister z. B. für die Wartung, Pflege und ggf. Anpassung der Embedded Software über Updates verantwortlich ist, bleibt der Arbeitgeber verantwortlich im Sinne des ArbSchG und der BetrSichV.

> *Lernt ein KI-System unüberwacht nach Inbetriebnahme weiter, können mitunter **neue Produkte entstehen**, wenn die vom Hersteller vorgegebenen Grenzen überschritten werden. Der Arbeitgeber kann dann zum Hersteller werden, mit den entsprechenden Pflichten aus ProdSG und der jeweils einschlägigen ProdSV. Zu den mit dieser Frage verbundenen Problemen sei auf die Ausführungen oben bei 5.3.1.1.3 verwiesen.*
> *Auch hier kann ein hoher Grad an Veränderbarkit von KI-Systemen den Arbeitgeber also vor besondere Herausforderungen stellen. Denn die KI-*

[204] *Kommission* – Leitfaden für die Umsetzung der Produktvorschriften der EU („Blue Guide"), Abl. C 272 vom 26.07.2016, Nr. 2.1.
[205] Dies gilt wegen § 2 Nr. 10 S. 2 der 9. ProdSV auch außerhalb des Anwendungsbereichs der BetrSichV immer dann, wenn eine Maschine so verändert wird, dass eine neue entsteht. Dann sind jedenfalls auch die formellen Voraussetzungen der 9. ProdSV zu erfüllen.

Systeme sind u. U. derart veränderbar, dass sie ihr „Verhalten“ in sicherheitsrelevanter Weise ändern, ohne dass der Arbeitgeber oder die Beschäftigten darauf Einfluss haben.

5.3.1.2.4.3 Prüfung vor der erstmaligen Verwendung oder bei Änderung; Kontrollen

Gemäß § 14 Abs. 1 S. 1 BetrSichV hat der Arbeitgeber Arbeitsmittel vor der erstmaligen Verwendung zu prüfen, wenn deren Sicherheit von den Montagebedingungen abhängt. Prüfinhalte, die bereits im Rahmen des (produktsicherheitsrechtlichen) Konformitätsbewertungsverfahrens geprüft wurden, sind nach § 14 Abs. 1 S. 3 BetrSichV nicht mehr Gegenstand der Prüfung. Bei den im Betrieb zu installierenden Arbeitsmitteln geht also die Pflicht des Herstellers zur Gewährleistung der Sicherheit vor.
Wie die Prüfung durchzuführen ist, kann der **TRBS 1201** entnommen werden. Demnach umfasst die Prüfung im Wesentlichen den Vergleich des **Istzustandes** mit dem **Sollzustand** des Arbeitsmittels.
Dabei sind in **technischer Hinsicht** die **sicherheitstechnisch relevanten Merkmale** des Arbeitsmittels auf Zustand, Vorhandensein und ggf. Funktionsfähigkeit zu untersuchen.[206] Zu den sicherheitsrelevanten Merkmalen gehören z. B. auch die sicherheitsrelevanten Mess-, Steuer- und Regeleinrichtungen (MSR-Einrichtungen).[207]

Ein in ein Arbeitsmittel integriertes KI-System kann eine solche sicherheitsrelevante MSR-Einrichtung darstellen und daher Gegenstand einer solchen Prüfung sein.

Gemäß § 3 Abs. 6 S. 1 BetrSichV hat der Arbeitgeber **Fristen für wiederkehrende Prüfungen** festzulegen. Dies ist für bestimmte Arbeitsmittel in § 14 Abs. 2 und 4 BetrSichV sowie für überwachungsbedürftige Anlagen in § 16 BetrSichV vorgeschrieben. Für andere Arbeitsmittel obliegt es dem Arbeitgeber, die Notwendigkeit und gegebenenfalls entsprechende Fristen für wiederkehrende Prüfungen in eigener Verantwortung im Rahmen der Gefährdungsbeurteilung zu ermitteln.[208]
Auch **Art und Umfang der Prüfung** sind vom Arbeitgeber in eigener Verantwortung im Rahmen der Gefährdungsbeurteilung festzulegen[209] und eine **Dokumentation** entsprechend **§ 3 Abs. 8** BetrSichV sicherzustellen.
Im Rahmen der Prüfung festgestellte **Abweichungen des Ist- vom Soll-Zustand** (Mängel) sind **durch die erforderlichen Maßnahmen zu beseitigen**. Die Durchführung der Prüfung und die Vornahme der erforderlichen Maßnahmen ist zu dokumentieren, so § 4 Abs. 4 BetrSichV.
Die **Prüfung** ist gemäß **§ 14 Abs. 3 S. 1 BetrSichV zu wiederholen**, wenn **prüfpflichtige Änderungen** vorgenommen werden.
Neben den Prüfungen sieht die BetrSichV in § 4 Abs. 5 S. 3 sogenannte **Kontrollen** vor. Hierbei handelt es sich um Sichtprüfungen auf offensichtliche Mängel und die regelmäßige Kontrolle der Funktionsfähigkeit von Schutz- und Sicherheitseinrichtungen vor Verwendung des Arbeitsmittels.[210] Fristen sowie Art und

[206] TRBS 1201, 2.4 (Stand: März 2019).
[207] TRBS 1201, 2.9 (Stand: März 2019).
[208] *Wink* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, BetrSichV § 3 Rn. 6a.
[209] TRBS 1201, 3.1 Abs. 1 (Stand: März 2019).
[210] TRBS 1201, 2.6 (Stand: März 2019).

Umfang der Kontrollen sind vom Arbeitgeber im Rahmen der Gefährdungsbeurteilung festzulegen.

5.3.1.2.4.4 Problem: Prüfpflichtige Änderungen bei weiterlernenden Systemen

Je nach Veränderbarkeit des Arbeitsmittels kann also häufig eine Änderung eintreten, die zu einer außerordentlichen Prüfung verpflichtet.

> *Bei weiterlernenden Systemen kann dann eine ständige Prüfung erforderlich werden. Hier können also die gleichen praktischen Probleme entstehen wie bei der Bestimmung des regelmäßigen Prüfintervalls im Rahmen der Gefährdungsbeurteilung.*

Gleichwohl kann bei weiterlernenden Systemen die sonst prüfpflichtige Änderung bereits Gegenstand der Gefährdungsbeurteilung gewesen sein. Voraussetzung dafür ist, dass das System ausreichend transparent ist.

5.3.1.2.4.5 Überwachungsbedürftige Anlagen

Für überwachungsbedürftige Anlagen treffen die §§ 15 – 18 BetrSichV zusammen mit der Anlage 2 der BetrSichV besondere Regelungen für die Prüfintervalle, die Prüfinhalte und die Anforderungen an die die Prüfung durchführenden Personen.
Zu den überwachungsbedürftigen Anlagen gehören gemäß Abschnitt 2 Nr. 2 lit. b) Anhang 2 der BetrSichV auch **Maschinen** nach Nr. 17 Anhang IV der Maschinen-RL zum Heben von Personen und Personen und Gütern, bei denen die Gefährdung eines Absturzes aus einer Höhe von mehr als 3 m besteht.
Für bestimmte Anlagen sieht § 18 Abs. 1 S. 1 BetrSichV eine Erlaubnispflicht für den Bau und die Errichtung sowie die sicherheitsrelevante Änderung der Bauart oder Betriebsweise bestimmter dort aufgeführter Anlagen vor. Erst nach behördlicher Prüfung und Erlaubnis dürfen diese Handlungen vorgenommen werden.
Insgesamt werden also die Anforderungen an die Prüfung durch den Arbeitgeber konkretisiert. Dabei gilt auch hier, dass **Doppelprüfungen** von Aspekten, die Gegenstand des **Konformitätsbewertungsverfahrens** waren, nicht nötig sind. Was der Hersteller geprüft hat, muss der Arbeitgeber also nicht noch einmal prüfen.

> *Es stellen sich jedoch dieselben Fragen wie zuvor bei der Gefährdungsbeurteilung und der Prüfung nach § 14 BetrSichV, wenn es sich um weiterlernendes System handelt.*

5.3.1.2.4.6 Zwischenergebnis zu Veränderbarkeit und Recht des technischen Arbeitsschutzes

- *Eine abschließende Gefährdungsbeurteilung ist bei weiterlernenden Systemen praktisch nicht möglich.*
- *Die erforderlichen Prüfintervalle werden bei weiterlernenden Systemen, die sich schnell ändern, so eng, dass sich die Pflicht zur wiederkehrenden Prüfung zur Pflicht zur dauerhaften Prüfung verdichtet.*
- *Bei wesentlichen Änderungen können Herstellerpflichten für den Arbeitgeber entstehen.*

Wie auch im Produktsicherheitsrecht stellt sich also das weiterlernende System als problematisch für den Arbeitgeber dar.
Den Arbeitgeber treffen, anders als den Hersteller, keine **zeitpunkt- sondern zeitraumbezogene Pflichten**. Strukturell bieten damit die Gefährdungsbeurteilung sowie die Prüfungen des Arbeitsmittels dem Arbeitgeber Möglichkeiten, auch ein hochgradig veränderbares System zu beherrschen. Allerdings kann auch hier die Beurteilung der Gefahren bei weiterlernenden Systemen nicht abschließend erfolgen. Entsprechend häufig muss die Gefährdungsbeurteilung aktualisiert werden.

5.3.1.3 Veränderbarkeit und Immissionsschutzrecht

Das Immissionsschutzrecht hat zum Ziel, schädliche Einwirkungen von Industrie- und Gewerbeanlagen zu vermeiden, und zwar sowohl **Allmählichkeitsschäden durch anhaltende Immissionen** als auch **plötzlich auftretende Schäden durch Störfälle**. Sie ist nicht in einem einzelnen Gesetz geregelt, sondern soll durch diverse Anforderungen aus unterschiedlichen Gesetzen gewährleistet werden, insbesondere dem **Bundesimmissionsschutzgesetz** (BImSchG)[211], der **EU-Seveso-Richtlinie** und der darauf basierenden **Störfall-Verordnung** (Störfall-VO, 12. BImSchV)[212].

5.3.1.3.1 Regelungsgegenstand des BImSchG

Das für das Umweltrecht zentrale BImSchG dient gemäß seines § 1 Abs. 1 BImSchG dem Schutz von Menschen, Tieren und Pflanzen, Boden, Wasser, Atmosphäre sowie Kultur- und sonstigen Sachgütern vor **schädlichen Umwelteinwirkungen** und der Vorbeugung solcher Einwirkungen (= Immissionen). Der Begriff der schädlichen Umwelteinwirkungen wird in § 3 Abs. 1 BImSchG legaldefiniert als „*Immissionen, die nach Art, Ausmaß oder Dauer geeignet sind, Gefahren, erhebliche Nachteile oder erhebliche Belästigungen für die Allgemeinheit oder die Nachbarschaft herbeizuführen"*, konkretisiert damit das baurechtliche Gebot der Rücksichtnahme und entspricht der wesentlichen Beeinträchtigung durch die Zuführung unwägbarer Stoffe nach § 906 Abs. 1 BGB.[213] Das Gesetz knüpft an unterschiedliche Arten von Immissionen an, namentlich an Luftverunreinigungen, Geräusche, Erschütterungen, Licht, Wärme, Strahlen und ähnliche Umwelteinwirkungen, § 3 Abs. 2 BImSchG.
Während §§ 4 - 31 BImSchG des Gesetzes Regelungen für die Errichtung und den Betrieb von **Anlagen** treffen, ist in §§ 32 – 37g BImSchG der **„produktbezogene Immissionsschutz"** geregelt, der immissionsschutzrechtliche Anforderungen auf der vorangehenden Stufe des Inverkehrbringens, der Einfuhr und der Herstellung u. a. von Anlagen (-teilen) normiert.[214] Die nachfolgenden Teile des Gesetzes behandeln **verkehrsbezogenen** (§§ 38 - 43) und **gebietsbezogenen** (§§ 44 - 47) Immissionsschutz.
Im BImSchG selbst sind nur grundlegende Anforderungen an den Immissionsschutz geregelt. Zur Konkretisierung dieser Anforderungen enthält das Gesetz zahlreiche

---

[211] Gesetz zum Schutz vor schädlichen Umwelteinwirkungen durch Luftverunreinigungen, Geräusche, Erschütterungen und ähnliche Vorgänge (Bundes-Immissionsschutzgesetz - BImSchG) neugefasst durch Beschluss vom 17.05.2013 BGBl. I S. 1274; zuletzt geändert durch Artikel 103 der Verordnung vom 19.06.2020 BGBl. I S. 1328.
[212] Zwölfte Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes (Störfall-Verordnung - 12. BImSchV) neugefasst durch Beschluss vom 15.03.2017 BGBl. I S. 483, 3527; zuletzt geändert durch Artikel 107 der Verordnung vom 19.06.2020 BGBl. I S. 1328.
[213] *Jarass*, BImSchG, § 3 Rn. 24.
[214] *Jarass,* BImSchG, § 32 Rn. 1.

Ermächtigungen zum Erlass von Verordnungen, die technische Einzelheiten regeln und unbestimmte Rechtsbegriffe des BImSchG auslegen. Zur Durchführung des BImSchG und dieser Bundesimmissionsschutzverordnungen (BImSchV) werden gem. § 48 BImSchG verschiedene, für die Praxis sehr bedeutsame, normkonkretisierende Verwaltungsvorschriften wie die Technischen Anleitungen (TA) Luft und die TA Lärm erlassen, die vor allem Grenzwerte für einzelne Immissionen enthalten.[215]
Aufgrund der umfassenden Genehmigungsvoraussetzungen und der aus § 13 BImSchG folgenden, weitreichenden Konzentrationswirkung der immissionsschutzrechtlichen Genehmigung[216] soll im Folgenden das Hauptaugenmerk im Rahmen der Anlagensicherheit auf die anlagenbezogenen Regelungen des BImSchG gelegt werden. Wegen der Relevanz für die Herstellung von Produkten wird anschließend auf den produktbezogenen Immissionsschutz des BImSchG eingegangen.

5.3.1.3.2 Anlagenbezogener Immissionsschutz

Der anlagenbezogene Immissionsschutz als Kernstück des BImSchG geht bis auf die Preußische Allgemeine Gewerbeordnung von 1845 zurück und hat seinen unmittelbaren Ursprung in den früheren Vorschriften für gefährliche Anlagen der §§ 16 - 25 GewO a. F. Während Regelungsdichte und Anwendungsbereich sich immer weiter ausgeweitet haben, ist die gewerberechtliche Struktur von genehmigungsbedürftigen und -freien Anlagen beibehalten worden. Es handelt sich also um das Genehmigungsrecht für Industrie- und Gewerbeanlagen.[217]

5.3.1.3.2.1 Genehmigungspflichtige Anlagen

Genehmigungspflichtig sind bestimmte Anlagetypen aufgrund ihres erhöhten Immissions- bzw. Gefahrenpotentials. Diese sind in der *Verordnung über genehmigungsbedürftige Anlagen* (4. BImSchV) abschließend in einem umfangreichen Katalog aufgelistet.[218] Es bestehen verschiedene Grundpflichten (z. B.: Schutz- und Abwehrpflicht, Vorsorgepflicht, § 5 I Nr. 1 und 2 BImSchG), die einerseits die grundlegenden materiellen Voraussetzungen für die Genehmigungserteilung und für spätere Maßnahmen darstellen, andererseits aber für die Anlagenbetreibenden unmittelbar verbindlich sind. Eine Konkretisierung der anlagenbezogenen Pflichten für verschiedene Teilbereiche geschieht etwa durch die Störfall-VO und die bereits genannten Verwaltungsvorschriften TA Luft und TA Lärm.[219] Maßstab für Vorsorgepflichten ist der **Stand der Technik** gem. § 5 I Nr. 2 BImSchG bzw. der **Stand der Sicherheitstechnik** gem. § 3 IV Störfall-VO.
Die **Erstgenehmigung** gem. § 4 BImSchG (im Gegensatz zur Änderungsgenehmigung nach § 16 BImSchG) ist umfassend, denn neben der Erfüllung der immissionsschutzrechtlichen Pflichten ist gem. § 6 I Nr. 2 BImSchG Voraussetzung für ihre Erteilung, dass andere öffentlich-rechtliche, anlagenbezogene Vorschriften und Belange des Arbeitsschutzes nicht entgegenstehen. So sind beispielsweise die auf das ArbSchG gestützten und die gem. § 34 ProdSG erlassenen

---

[215] *Jarass*, Grundstrukturen des Immissionsschutzrechts, JuS 2009, 608, 611.
[216] *Jarass,* BImSchG, § 13 Rn. 1.
[217] *Jarass*, BImSchG, § 4 Rn. 1 f; *Jarass*: Grundstrukturen des Immissionsschutzrechts, JuS 2009, 608, 611.
[218] *Jarass*, Grundstrukturen des Immissionsschutzrechts, JuS 2009, 608, 610 f.
[219] *Jarass*, Grundstrukturen des Immissionsschutzrechts, JuS 2009, 608, 611.

Rechtsverordnungen des Produktsicherheitsrechts zu beachten[220], also auch die oben behandelte BetrSichV, wodurch die anlagenbezogenen Sicherheitsanforderungen im immissionsschutzrechtlichen Erstgenehmigungsverfahren vollständig geprüft werden.
Die Genehmigungspflicht besteht aber nicht nur bei der erstmaligen Errichtung oder Inbetriebnahme einer in der 4. BImSchV aufgeführten Anlage, sondern gem. § 16 BImSchG auch bei **wesentlichen Änderungen** einer solchen genehmigungsbedürftigen Anlage, was in der Praxis sogar häufiger als Erstgenehmigungen vorkommt. Wird diese Wesentlichkeitsschwelle nicht überschritten und eine Änderungsgenehmigung daher nicht beantragt, so besteht gem. § 15 Abs. 1 BImSchG eine Anzeigepflicht, wenn sich die Änderung auf die in § 1 BImSchG genannte Schutzgüter auswirken kann, sog. **bedeutsame Änderung**.[221] **Unbedeutende Änderungen** sind hingegen anzeigefrei.[222]
Die §§ 15 f. BImSchG unterscheiden Änderungen in Lage, Beschaffenheit und Betrieb der Anlage. Eine Änderung der Beschaffenheit liegt vor, wenn ihr Zustand oder ihre konstruktiven Merkmale verändert werden, insbesondere wenn einzelne Teile ersetzt oder beseitigt werden oder zusätzliche Einrichtungen hinzukommen, während Betriebsänderungen die Modifizierung der Produktionsprozesse oder der Betriebsweise der Anlage meinen.[223]
Änderung im Sinne der Vorschriften meint immer eine Abweichung vom vorangegangenen Genehmigungsbescheid, wobei kleinere oder nur positive Änderungen oft vom Genehmigungsbescheid gedeckt sind.[224] Dagegen bedeutet die Nutzung von im Genehmigungsbescheid bereits mitgenehmigten Alternativen keine solche Änderung[225], ebenso wenig die bloße Instandsetzung, Reparatur oder Unterhaltung, die den genehmigten Zustand unverändert wiederherstellen.[226]
Eine **wesentliche Änderung** liegt gem. § 16 Abs. 1 Hs. 1 BImSchG dann vor, wenn durch die Änderung nachteilige Auswirkungen (auf die Schutzgüter des § 1) hervorgerufen werden können und die Möglichkeit besteht, dass diese auch im Rahmen einer Erstgenehmigung nach § 6 Abs. 1 Nr. 1 BImSchG erheblich wären. Nach Hs. 2 ist eine Änderung unabhängig von den Auswirkungen auch dann wesentlich, wenn sie für sich genommen die in der 4. BImSchV enthaltenen Schwellenwerte erreicht.[227]
Es handelt sich bei der rechtlichen Bewertung von Anlagenveränderungen also um eine zweistufige Systematik, die sich an zwei unterschiedlichen Bezugspunkten orientiert: Zunächst stellt sich bei der Einordnung einer Erweiterung oder Modifizierung **als Änderung** die Frage, inwieweit diese die Form oder die Arbeitsweise der Anlage umgestaltet. Bei der anschließenden Klassifizierung der Änderung **als unbedeutend, anzeige- oder genehmigungspflichtig** ist die Auswirkung der Änderung auf das Gefahrenpotential der Anlage der Maßstab.
Zum Betrieb im Sinne des BImSchG gehören u. a. die Produktionsverfahren, die Kapazität, die Arbeitsabläufe und die Betriebszeiten.[228]

---

[220] *Jarass*, BImSchG § 6 Rn. 41.
[221] *Jarass*, Grundstrukturen des Immissionsschutzrechts, JuS 2009, 608, 611.
[222] *Jarass*, BImSchG § 15 Rn. 3.
[223] *Jarass*, BImSchG § 15 Rn. 7.
[224] *Jarass*, BImSchG § 15 Rn. 11.
[225] *Jarass*, BImSchG § 15 Rn. 18.
[226] *Jarass*, BImSchG § 15 Rn. 16.
[227] *Jarass*, BImSchG § 16 Rn.15.
[228] *Jarass*, BImSchG § 4 Rn. 57.

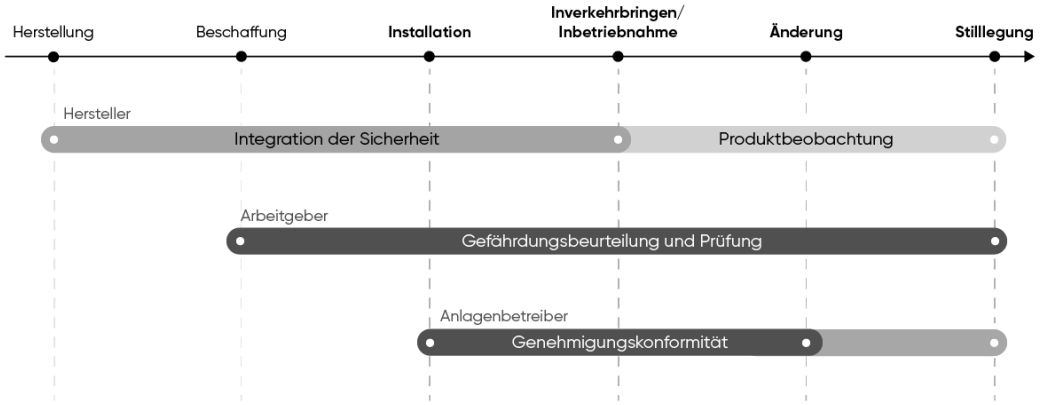


Abb. 5.3 Verantwortlichkeit des Anlagenbetreibers (unten). Nach Änderung der Anlage stellt sich die Frage, ob sie noch von der erteilten Genehmigung erfasst ist.

5.3.1.3.2.2 Problem: Genehmigung und wesentliche Änderung durch weiterlernendes System

Die Anlage muss auch nach Inbetriebnahme hinaus weiterhin die jeweiligen Anforderungen erfüllen, also im Rahmen der jeweiligen Genehmigung betrieben werden. Eine Änderung muss immer den Betrieb der Anlage betreffen, also Produktionsverfahren, die Kapazität, die Arbeitsabläufe oder die Betriebszeiten.

> *Änderungen durch weiterlernende KI-Systeme dürften ausschließlich als Änderungen im Betrieb der Anlage einzuordnen sein, wenn diese etwa eine Effizienzsteigerung durch eine Verfahrensänderung errechnen. Dies kann zudem die Kapazität des Betriebs steigern und sich auf Arbeitsabläufe und Betriebszeiten auswirken. Dann stellen sie Änderungen im Sinne der §§ 15 f. BImSchG dar, wenn sie nicht schon als Alternativbetriebe Gegenstand der Vorgenehmigung waren.*
>
> *Bezüglich weiterlernender KI-Systeme in Anlagen (-teilen) stellt sich also die Frage, ob die Veränderbarkeit bereits im für den erstmaligen Einsatz des Systems einzuleitenden Genehmigungsverfahren – sei es eine Erstgenehmigung nach § 4 oder eine Änderungsgenehmigung nach § 16 BImSchG – vollständig erfasst werden muss oder ob erst die tatsächliche (Eigen-) Veränderung des Systems, sobald absehbar, nach Maßgabe der §§ 15 f. BImSchG zu behandeln ist.*

Wenn diese Änderungen zur Steigerung von Immissionen führen (können), sind sie wesentlich und bedürfen der Genehmigung gem. § 16 BImSchG.

Die Veränderung darf nicht zu einer Überschreitung der Grenzwerte führen. Zum Genehmigungsverfahren gehört daher ggf. auch der Nachweis, dass trotz Veränderbarkeit z. B. initiiert durch einen Dritten keine schädlichen Umwelteinwirkungen eintreten, die sich nicht im erlaubten Rahmen bewegen. Solange dies der Fall ist, bestehen hier nicht dieselben Schwierigkeiten wie in der 9. ProdSV und dem ArbSchG bzw. der BetrSichV bei der Risikobeurteilung bzw. Gefährdungsbeurteilung.

5.3.1.3.2.3 Weitere Instrumente staatlicher Kontrolle im BImSchG

Auch nach Genehmigungserteilung können gem. § 17 BImSchG **nachträgliche Anordnungen** getroffen und dadurch faktisch[229] die Nebenbestimmungen der jeweiligen Genehmigung verschärft werden. Eine solche Anordnung kann jederzeit – auch wenn sämtliche (Neben-) Bestimmungen der Genehmigung befolgt werden – ergehen, wenn die Anlage bzw. deren Betrieb aktuellen immissionsschutzrechtlichen Anforderungen nicht (mehr) entspricht.

> *Solche Anordnungen können – ebenso wie herkömmliche Systeme – auch die in der Anlage eingesetzten KI-Systeme betreffen, auch wenn diese von der bestehenden Genehmigung erfasst sind oder eine durch sie hervorgerufene Änderung der Anlage angezeigt worden ist.*

Weitere Eingriffsmöglichkeiten sind die **Untersagung**, **Stilllegung** und **Beseitigung** sowie **Widerruf** und **Rücknahme** der Genehmigung, § 20 f. BImSchG bzw. § 48 VwVfG.[230]

5.3.1.3.2.4 Störfallrecht

Im Bereich der **Störfall-VO** gelten besondere Anforderungen an den Betrieb der davon erfassten Anlagen. Es handelt sich um besonders risikoträchtige Anlagen. Werden in ihnen veränderbare Systeme eingesetzt, steht der Betreiber vor vergleichbaren Problemen bei der Risikoermittlung wie der Hersteller eines Produkts bei der Integration der Sicherheit bei Konstruktion des Produkts. Denn der Betreiber muss nach § 3 Abs. 1 S. 1 Störfall-VO die angesichts der möglichen Gefahren erforderlichen Maßnahmen zur Verhinderung von Störfällen treffen. Dafür ist eine Ermittlung von Art und Maß der möglichen Gefahr und auf dieser Grundlage eine Ermittlung der erforderlichen Vorkehrungen durchzuführen. Jedenfalls bei weiterlernenden Systemen wird diese Ermittlung aus den bereits unter 3.1.2.3 dargestellten Gründen erschwert.
Neben der immissionsschutzrechtlichen Genehmigung besteht eine Genehmigungspflicht für **störfallrelevante Änderungen** gem. § 16a BImSchG bei genehmigungsbedürftigen Anlagen[231], wenn die Änderung nicht bereits durch die immissionsschutzrechtliche Änderungsgenehmigung des § 16 Abs. 1 S. 1 BImSchG erfasst ist, sowie bei nicht genehmigungsbedürftigen Anlagen eine **störfallrechtliche Anzeige- bzw. Genehmigungspflicht** gem. §§ 23a, 23b BImSchG.
Eine störfallrechtliche (Änderungs-) Genehmigung ist gem. § 16a S. 1 bzw. § 23b Abs. 1 S. 1 BImSchG dann einzuholen, wenn durch die störfallrelevante Errichtung, den störfallrelevanten Betrieb oder die störfallrelevante Änderung Sicherheitsabstände zu benachbarten Schutzobjekten (weiter) unterschritten werden oder eine erhebliche Gefahrenerhöhung ausgelöst wird. Der Änderungsbegriff ist derselbe wie der des § 16 BImSchG. Störfallrelevant ist die Änderung, wenn sie sich möglicherweise erheblich auf die Gefahren schwerer Unfälle auswirkt oder die Klasse des Betriebsbereichs durch sie geändert werden könnte. [232]

---

[229] Rechtlich bleibt die bestehende Genehmigung unberührt, *Jarass,* BImSchG, § 17 Rn. 2.
[230] *Jarass,* Grundstrukturen des Immissionsschutzrechts, JuS 2009, 608, 612.
[231] Als "genehmigungsbedürftige Anlagen" werden im Kontext des BImSchG nur die gem. § 4 BImSchG immissionsschutzrechtlich genehmigungsbedürftigen Anlagen bezeichnet, nicht aber diejenigen Anlagen, die einer störfallrechtlichen Genehmigung gem. § 23b BImSchG bedürfen, *Jarass*, BImSchG § 4 Rn. 14a.
[232] *Jarass*, BImSchG § 16a Rn. 4.

*Genehmigungsrelevante, durch KI-Systeme ausgelöste Änderungen dürften sich auch hier auf erhebliche Gefahrenerhöhungen durch Änderungen **im Betrieb** der Anlage beschränken. Dabei stellt sich wiederum die Frage, ob die Veränderbarkeit des Systems, soweit eine Vorgenehmigung vorliegt, bereits von dieser voll erfasst wird oder ob erst für die absehbare, tatsächliche Veränderung eine Änderungsgenehmigung eingeholt werden muss.*

5.3.1.3.3 Produktbezogener Immissionsschutz

In §§ 32 – 37 BImSchG wird ein produktbezogener Immissionsschutz vor schädlichen Umwelteinwirkungen geregelt. Insbesondere sollen hier die anlagenbezogenen §§ 32 und 33 BImSchG dargestellt werden. Beide stellen Verordnungsermächtigungen dar.
Der § 32 BImSchG sieht vor, dass durch Rechtsverordnung das Inverkehrbringen oder die Einfuhr von serienmäßig hergestellten Teilen von Betriebsstätten und sonstigen ortsfesten Einrichtungen sowie von Maschinen, Geräten und sonstigen ortsveränderlichen technischen Einrichtungen sowie Fahrzeugen und serienmäßig dafür hergestellte Teilen davon abhängig gemacht werden kann, dass bestimmten Anforderungen zum Schutz vor schädlichen Umwelteinwirkungen durch Luftverunreinigungen, Geräusche, Erschütterungen oder nichtionisierende Strahlen entsprechen. Diese Anforderungen adressieren also nicht den Anlagenbetreiber, sondern den Hersteller oder Einführer.[233] Damit können die Anforderungen nach der so erlassenen Verordnung mit dem Produktsicherheitsrecht in Konkurrenz treten. Häufig finden sich im einschlägigen Produktsicherheitsrecht jedoch Subsidiaritätsklauseln, sodass es hinter die entsprechende BImSchV insoweit zurücktritt, als dort andere oder weitergehende Anforderungen an das Produkt formuliert werden.[234] Eine solche Klausel findet sich in § 1 Abs. 3 der 9. ProdSV.
Als Verordnung auf Grundlage von (unter anderem) § 32 BImSchG ist die Geräte- und Maschinenlärmschutzverordnung (32. BImSchV)[235] zu nennen. Sie regelt neben Anforderungen an den Betrieb von Maschinen zur Vermeidung schädlicher Umwelteinwirkungen auch immissionsschutzrechtliche Anforderungen an das Inverkehrbringen und die Einfuhr von Maschinen und richtet sich damit auch an den Hersteller.

Insofern kann wegen der Herausforderungen des Herstellers von Produkten mit KI-Komponenten, die unter die 32. BImSchV fallen, auf die Ausführungen oben zur 9. ProdSV verwiesen werden. Es handelt sich um besonderes Produktsicherheitsrecht, das die Anforderungen des ProdSG und der jeweils einschlägigen ProdSV ergänzt bzw. verdrängt. Es stellt jedoch mit den Grenzwerten klare Anforderungen an das Produkt, sodass die Herausforderungen der Risikobeurteilung bei der Integration der Sicherheit insofern nicht vorliegen. Auch das weiterlernende System darf den Grenzwert nicht überschreiten.

Der § 33 Abs. 1 Nr. 1 und 2 BImSchG hingehen sieht eine Verordnungsermächtigung zur Einführung von Bauartzulassungen für Anlagen im Sinne des § 3 Abs. 5 Nr. 1 und 2 BImSchG vor, also für ortsfeste Anlagen und ortsveränderliche Anlagen wie

---

[233] *Jarass*, BImSchG § 32 Rn. 9.
[234] *Jarass*, BImSchG § 32 Rn. 2.
[235] 32. Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes (Geräte- und Maschinenlärmschutzverordnung - 32. BImSchV), Art. 1 der Verordnung vom 29.08.2002, BGBl. I S. 3478; zuletzt geändert durch Art. 83 der Verordnung vom 31.08.2015, BGBl. I S. 1474.

Maschinen. Diese können sich nach § 33 Abs. 1 Nr. 1 BImSchG auf die Errichtung und den Betrieb von Anlagen beziehen, oder nach § 33 Abs. 1 Nr. 2 BImSchG auf das Inverkehrbringen. Bei der Bauartzulassung handelt es sich um eine Produktgenehmigung, die die immissionsschutzrechtliche Zulässigkeit bestätigt.[236] Der Hersteller lässt sich also für die Serie bestätigen, dass das Muster den in der jeweiligen BImSchV geregelten Anforderungen entspricht. Wird die Verordnung mit einer Verordnung nach § 4 Abs. 1 S. 3, 2. Halbsatz BImSchG verbunden, also mit einer Verordnung betreffend die genehmigungsbedürftigen Anlagen, bedarf es für die Errichtung und den Betrieb der Anlage keiner Genehmigung mehr, wenn eine Bauartzulassung vorliegt.[237] Der Hersteller entlastet damit den Anlagenbetreiber. Von § 4 Abs. 1 S. 3, 2. Halbsatz BImSchG wurde bisher noch nicht Gebrauch gemacht, eine derartige Verordnung liegt also bisher nicht vor.[238]

5.3.1.3.4 Zwischenergebnis zu Veränderbarkeit und Immissionsschutzrecht

Auch hier ist erneut das weiterlernende System problematisch. Wie auch bei den Pflichten des Arbeitgebers kann der Anlagenbetreiber seine Anlage über die Zeit kontrollieren. Hier setzt das Konzept der Änderungsgenehmigung an. Diese zeitraumbezogene Pflicht unterscheidet seine Pflichten von denen des Herstellers.

- *Bei der Erstgenehmigung muss die Veränderbarkeit weiterlernender Systeme berücksichtigt werden. Das stellt Betreiber und Genehmigungsbehörde vor umfangreiche Nachweis- bzw. Prüfpflichten im Genehmigungsverfahren.*
- *Die Risikoermittlung nach Störfallrecht stellt den Betreiber vor vergleichbare Herausforderungen wie den Arbeitgeber bei der Gefährdungsbeurteilung, wenn er ein weiterlernendes System in der Störfallanlage einsetzt. Wie auch der Arbeitgeber hat er jedoch gegenüber dem Hersteller den Vorteil, dass er seine Risikoermittlung während des Betriebs anpassen kann. Ggf. ist dann jedoch (wie bei anderen genehmigungsbedürftigen Anlagen auch) eine Änderungsgenehmigung erforderlich.*

5.3.1.4 Veränderbarkeit und DSGVO

Die Datenschutzgrundverordnung (DSGVO)[239] schützt gemäß Art. 1 Abs. 1 und 2 DSGVO die Grundrechte und Grundfreiheiten natürlicher Personen bei der Verarbeitung personenbezogener Daten. Sofern personenbezogene Daten für die hier untersuchten KI-Systeme relevant sind, ist die Anwendbarkeit der DSGVO damit in Betracht zu ziehen. Das kann z. B. dann der Fall sein, wenn ein interagierendes KI-System zur Anpassung seiner Bewegungsabläufe die Bewegungsmuster der mit ihm kollaborierenden Menschen erfasst und verarbeitet oder ein automatisch selbstfahrender Roboter sich für Dialoge merken muss, mit wem er gerade gesprochen hat.

---

[236] *Jarass*, BImSchG § 33 Rn. 17.
[237] *Jarass*, BImSchG § 33 Rn. 20.
[238] *Jarass*, BImSchG § 4 Rn. 44, der die Norm daher als symbolisches Recht bezeichnet.
[239] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (DSGVO), ABl. L 119 vom 4.5.2016, S. 1–88.

5.3.1.4.1 Anwendungsbereich

Die DSGVO erfasst in **sachlicher Hinsicht** nach Art. 2 Abs. 1 DSGVO personenbezogene Daten im Sinne des Art. 4 Nr. 1 DSGVO. Damit sind alle Informationen gemeint, die sich einer identifizierten oder identifizierbaren natürlichen Person zuordnen lassen. Identifizierbar ist die Person nach der Legaldefinition in Art. 4 Nr. 1 2. Halbsatz DSGVO dann, wenn sie auch nur indirekt über eine Kennung, wie z. B. den Namen, über eine Kennnummer, über Standortdaten, Online-Kennung oder durch Zuordnung zu besonderen Merkmalen identifiziert werden kann.
Die DSGVO regelt gemäß Art. 2 Abs. 1 DSGVO die ganz oder teilweise automatisierte Verarbeitung dieser Daten. Zur Verarbeitung, die in Art. 4 Nr. 2 DSVGO legaldefiniert ist, gehört eine Vielzahl von Vorgängen, angefangen mit dem Erheben der Daten bis hin zu ihrer Löschung und Vernichtung. Der Katalog ist nicht abschließend, der Begriff der Verarbeitung ist also weit gefasst.
Beim Einsatz von KI-Systemen in Umgebungen, in denen sie mit Menschen derart in Kontakt treten, dass sie Informationen zu diesen Menschen sensorisch erfassen müssen oder aus Datenbanken benötigen, um bestimmungsgemäß und sicher zu funktionieren, kann daher der sachliche Anwendungsbereich der DSGVO eröffnet sein.

> *Die Datenaufnahme durch die Sensorik des KI-Systems oder der Abruf von personenbezogenen Daten z. B. über den interagierenden Menschen ist eine **automatisierte Verarbeitung** der Daten, da es sich um einen Vorgang unter Einbeziehung informationstechnischer Infrastruktur handelt, wenn diese Daten von der Hardware aufgenommen und an die verarbeitenden Systemteile übermittelt werden.*

Der **persönliche Anwendungsbereich** der DSGVO erfasst die Verantwortlichen der Datenverarbeitung und damit gemäß Art. 4 Nr. 7 DSGVO die Stellen, die allein oder mit anderen über die Zwecke der Verarbeitung entscheiden. Demzufolge kann jeder, der Daten für sich verarbeitet, Verantwortlicher in diesem Sinne sein.[240] Maßgeblich ist hier die Entscheidungsbefugnis über die Zwecke und Mittel der Verarbeitung. Diese kann sich aus einem Gesetz, aus tradierter Praxis oder aus tatsächlichen Umständen ergeben.[241]

5.3.1.4.2 Datenschutz-Folgenabschätzung

Nach Art. 35 Abs. 1 S. 1 DSGVO hat der Verantwortliche eine Datenschutz-Folgenabschätzung durchzuführen, wenn die Datenverarbeitung **hohe Risiken** für die Rechte und Freiheiten natürlicher Personen in sich birgt. Bei der Datenschutz-Folgenabschätzung handelt es sich um ein Instrument der Technikfolgenabschätzung. Sie dient dazu, die Folgen neuer Technologien auf die Gesellschaft zu untersuchen.[242] Zudem soll sie als Frühwarnsystem konkreten Risiken vorbeugen, indem Gefährdungen im Vorfeld erkannt und durch die erforderlichen Maßnahmen (vgl. Art. 25 und 32 DSGVO) verhindert werden können.[243]

---

[240] *Ernst* in: Paal/Pauly, DS-GVO, Art. 4 Rn. 55.
[241] *EDPS*, Guidelines on the concepts of controller, processor and joint controller under Regulation EU 2018/1725 vom 07.11.2019, S. 8 (abrufbar unter: https://edps.europa.eu/sites/edp/files/publication/19-11-07_edps_guidelines_on_controller_processor_and_jc_reg_2018_1725_en.pdf, zuletzt abgerufen am 18.03.2020).
[242] *Martini* in: Paal/Pauly, DS-GVO, Art. 35 Rn. 3.
[243] *Martini* in: Paal/Pauly, DS-GVO, Art. 35 Rn. 8.

Ob ein hohes Risiko vorliegt, bestimmt sich im Einzelfall anhand der Art, des Umfangs und der Zwecke der Verarbeitung. Der Verantwortliche hat dieses Risiko in eigener Verantwortung zu ermitteln.

> *Big Data-Anwendungen und darauf aufbauende KI-Systeme können aufgrund des hohen Datenvolumens bei der Verarbeitung ein solches hohes Risiko darstellen.*[244]

In Art. 35 Abs. 7 DSGVO findet sich ein Katalog mit **Mindestanforderungen an das Verfahren** der Datenschutz-Folgenabschätzung. Nach Art. 35 Abs. 7 lit. a) DSGVO gehört dazu auch eine **systematische Beschreibung der Datenverarbeitungsvorgänge**.

> *Das kann den Verantwortlichen bei dem Einsatz von KI-Systemen zu einer Darstellung der Arbeitsweise des KI-Systems zwingen.*

Welche Anforderungen an die Beschreibung zu stellen sind, ist vom konkreten Fall abhängig. Der Katalog des Art. 30 Abs. 1 DSGVO gibt einen Überblick, was von der Beschreibung umfasst sein muss.[245] Dazu gehört ggf. auch eine Beschreibung der eingesetzten Technik und ihrer Prozessabläufe.

Nach Art. 35 Abs. 7 lit. d) DSGVO besteht zudem eine **Dokumentationspflicht** über das Ergebnis der Datenschutz-Folgenabschätzung.

Gemäß Art. 35 Abs. 11 DSGVO ist die Datenverarbeitung erforderlichenfalls regelmäßig darauf zu überprüfen, ob sie noch entsprechend der Datenschutz-Folgenabschätzung erfolgt.

> *Das kann bei veränderbaren KI-Systemen relevant werden. Wann und wie sich das KI-System verändert, muss für den Verantwortlichen erkennbar sein, damit er seiner Pflicht zur Überprüfung der Datenschutz-Folgenabschätzung nachkommen kann.*[246]

5.3.1.4.3 Zwischenergebnis Veränderbarkeit und DSGVO

Im Anwendungsbereich der DSGVO kann eine hohe Veränderbarkeit also für den Verantwortlichen bei der Erstellung und Überprüfung der Datenschutz-Folgenabschätzung zu den gleichen Problemen führen, wie sie sich für den Arbeitgeber bei der Gefährdungsbeurteilung und für den Anlagenbetreiber bei der Risikoermittlung ergeben.

5.3.1.5 Veränderbarkeit und Haftungsrecht

Neben die bisher erörterten öffentlich-rechtlichen Regelungen treten zivilrechtliche Vorgaben. Für das vorliegende Gutachten werden jene zivilrechtlichen Regelungen untersucht, die die Sicherheit der hier gegenständlichen Technologie zum Gegenstand haben. Es werden zunächst die für den Hersteller, anschließend die für die Verwender der KI-Systeme relevanten Regelungen betrachtet. Das zivilrechtliche Haftungsregime wirkt nachsorgend, anders als das zuvor dargestellte vorsorgende Ordnungsrecht. Es hat außerdem nur im Einzelfall unmittelbare rechtliche Wirkung. Trotzdem ist das Haftungsrecht bereits in der Fertigung von Produkten maßgeblich, insbesondere wenn es sich um höchstrichterlich etablierte Haftungskonstellationen handelt. In der Fertigung der Produkte wird das Haftungsrecht durch den Hersteller antizipiert, um kostspielige Schadensersatzklagen zu vermeiden, wodurch das Haftungsrecht trotz

---

[244] *Martini* in: Paal/Pauly, DS-GVO, Art. 35 Rn. 18.
[245] *Martini* in: Paal/Pauly, DS-GVO, Art. 35 Rn. 47.
[246] *Conrad*, Künstliche Intelligenz, DuD 740, 744.

seines retrospektiven Ansatzes einen über den Einzelfall hinauswirkenden Steuerungseffekt hat.
Wo also öffentlich-rechtliche Regelungen zum Schutz bestimmter Rechtsgüter vor den von Produkten ausgehenden Gefahren fehlen oder im Einzelfall nicht hinreichend Schutz bieten, kann das auf Sicherung des Integritätsinteresses und Äquivalenzinteresses gerichtete Haftungsrecht einspringen.

5.3.1.5.1 Haftung des Herstellers

Im Folgenden werden kurz die den Hersteller potenziell betreffenden vertraglichen Haftungsregelungen nach dem Kaufrecht sowie die gesetzlichen Haftungsregelungen nach dem Produkthaftungsgesetz sowie der Produzentenhaftung dargestellt. Zudem wird erläutert, an welchen Stellen das Produktsicherheitsrecht haftungsrechtlich relevant wird.

5.3.1.5.1.1 Produkthaftungsgesetz

Unter Produkthaftung wird hier die Produkthaftung nach dem Produkthaftungsgesetz (ProdHaftG)[247] verstanden.[248] Es beruht auf der Umsetzung der Produkthaftungsrichtlinie[249].
Im Rahmen der Produkthaftung steht der Hersteller eines Produkts für Schäden an Sachen, Gesundheit und Leben des Geschädigten aufgrund eines in den Verkehr gebrachten fehlerhaften Produkts ein (sog. Mangelfolgeschäden). Haftungsrechtlicher Anknüpfungspunkt ist das Inverkehrbringen eines fehlerhaften Produktes. Dabei kann es sich um einen Konstruktions-, Fabrikations- oder Instruktionsfehler handeln.
Haftbar ist nach dem ProdHaftG der Hersteller. Für das ProdHaftG definiert § 4 Abs. 1 S. 1 Var. 1 ProdHaftG den Hersteller als denjenigen, der das Endprodukt herstellt. Darunter fallen auch Assembler, die die einzelnen Komponenten am Markt erwerben, zusammensetzen und dann in den Verkehr bringen.[250] Nach § 4 Abs. 1 S. 1 Var. 2 ProdHaftG ist auch der Hersteller eines Grundstoffs Hersteller im Sinne der Norm. Darunter fallen solche Produkte, die durch Modifizierung für den Einsatz im Endprodukt angepasst werden. Als Grundstoff kann ein weiterlernendes System[251] in Form eines Softwareagenten angesehen werden, das im Prozess der Herstellung des Endprodukt in dieses integriert und dann durch den Hersteller des Endprodukts konfiguriert wird. Denn der Agent ändert sich durch das Teaching nach Auslieferung an den Endhersteller und wird auf das Endprodukt angepasst. Der Hersteller des Agenten bleibt trotzdem Hersteller nach § 4 Abs. 1 S. Var. 2 ProdHaftG.[252] Zudem ist auch der Hersteller eines Teilprodukts nach § 4 Abs. 1 S. 1 Var. 3 ProdHaftGTeilprodukte sind solche Teile, die in das Endprodukt integriert werden, anders als der Grundstoff jedoch nicht ihre ursprünglichen Eigenschaften verlieren.[253]

---

[247] Gesetz über die Haftung für fehlerhafte Produkte (Produkthaftungsgesetz - ProdHaftG), Gesetz vom 15.12.1989 BGBl. I S. 2198; zuletzt geändert durch Artikel 5 des Gesetzes vom 17.07.2017 BGBl. I S. 2421.
[248] Es gibt daneben weitere spezialgesetzliche Produkthaftungsregelungen, wie z. B. im Arzneimittelrecht oder im Gentechnikrecht.
[249] Richtlinie 85/374/EWG des Rates vom 25. Juli 1985 zur Angleichung der Rechts- und Verwaltungsvorschriften der Mitgliedstaaten über die Haftung für fehlerhafte Produkte ABl. L 210 vom 07.8.1985, S. 29 – 33 (ProdHaft-RL).
[250] *Günther*, Roboter und rechtliche Verantwortung, S. 172.
[251] Hier im Sinne von nach Auslieferung lernend, siehe Glossar.
[252] *Günther*, a.a.O., S. 173.
[253] *Günther*, a.a.O.

Nach § 4 Abs. 1 S. 2, Abs. 2 und 3 ProdHaftG sind schließlich auch der Quasi-Hersteller, also wer durch Anbringen seines Namens o. ä. sich als Hersteller ausgibt, sowie der Importeur und der Lieferant des Endprodukts Hersteller im Sinne des ProdHaftG.

Die Unterscheidung in Herstellern von End- und Teilprodukten ist haftungsrechtlich deswegen relevant, weil die Haftung von Endherstellern und Teilherstellern nach dem Produkthaftungsgesetz nicht (gänzlich) deckungsgleich ist. Dies gilt insbesondere mit Blick auf Entlastungsmöglichkeiten und Haftungsumfang. Für die Fehlerfreiheit des Zulieferprodukts trägt der Teilhersteller die vollständige Verantwortung, neben ihm haftet allerdings auch der Hersteller des Endprodukts. Erfasst wird mithin die gesamt Zulieferkette, soweit es sich dabei jeweils um hergestellte (und fehlerhafte) Produkte handelt. Eine Konzentration der Haftung mit Blick auf den Endhersteller findet nicht statt.[254]

Ein Produkt im Sinne des ProdHaftG ist nach § 2 ProdHaftG eine bewegliche Sache. Auch Elektrizität ist als Produkt erfasst. Umstritten ist, ob Software dem Produktbegriff unterfällt. Dabei gibt es verschiedene Ansätze, die die Produkteigenschaft von Software von unterschiedlichen Kriterien abhängig machen. Zum einen wird auf die Verkörperung der Software auf einem gegenständlichen Datenträger, zum anderen auf die Standardisierung der Software abgestellt, um die Haftung zuzulassen.[255] Einen vergleichbaren Weg geht der Bundesgerichtshof bislang mit Blick auf die (analoge) Anwendung des Gewährleistungsrechts auf mangelhafte Computersoftware.[256] Stark umstritten ist die Frage nach der haftungsbegründenden Produkteigenschaft der Datenverarbeitungs- und Kommunikationssoftware daher vornehmlich, wenn die Software zum Zeitpunkt des Inverkehrbringens nicht verkörpert vorliegt, sondern beispielsweise eine Online-Anwendung darstellt und mithin nie auf dem Festspeicher fixiert wird oder erst später auf die Betriebssysteme aufgespielt wird.[257] Jedenfalls nach dem Wortlaut fallen solche Softwareprodukte mangels Sachqualität nicht unter den Produktbegriff des ProdHaftG, auch wenn mit Blick auf den Zweck der ProdHaft-RL in richtlinienkonformer Auslegung des ProdHaftG auch vertretbar erscheint, sie als Produkt anzusehen.

> *Bei KI-Systemen, die als Embedded Software Teil eines Produkts sind, kann damit im Rahmen der Haftung des Herstellers dieses Produkts relevant werden. Er muss dann bei der Produktion für eine sichere Integration des KI-Systems in das Produkt sorgen. Der Hersteller des KI-Systems wiederum, der dieses dem Hersteller des Produkts zur Verfügung stellt (und damit in den Verkehr bringt), ist nach dem engen Produktbegriff nicht haftbar nach dem ProdHaftG.*

Es handelt sich konzeptionell um eine Gefährdungshaftung, ein Verschulden wird nicht verlangt.[258] Das gilt jedenfalls für den Fabrikationsfehler. Damit sind auch sog. „Ausreißer“ haftungsbegründend, also solche Produkte einer Serie, die bei sonst einwandfreier Fabrikation ausnahmsweise Fehler aufweisen. Für den Konstruktions- und den Instruktionsfehler werden jedoch auch subjektive Tatbestandselemente verlangt. Insofern wird von einem hybriden Charakter des ProdHaftG gesprochen (strikte verschuldensunabhängige Haftung bei Fabrikationsfehlern einerseits,

---

[254] Ebenda, § 4 ProdHaftG Rn. 22.
[255] *Rebin* in: BeckOGK zum Produkthaftungsgesetz, § 2 ProdHaftG Rn. 49.
[256] BGH NJW 1988, 406.
[257] *Wagner* in: Münchener Kommentar BGB § 2 ProdHaftG Rn. 19 f.
[258] *Littbarski* in: Kilian/Heusser, Computerrechts-Handbuch, Produkthaftung Rn. 98 f.

verschuldensabhängige Haftung bei Konstruktions- und Instruktionsfehlern andererseits).[259]
Die Unterscheidung zwischen der verschuldensabhängigen Haftung und der verschuldensunabhängigen Haftung wegen des Inverkehrbringens eines fehlerhaften Produktes wirkt sich im Ergebnis nur selten aus, da ein Hersteller, der die im Verkehr erforderliche Sorgfalt einhält und mithin nicht fahrlässig handelt i. S. d. § 276 BGB[260], auch kein fehlerhaftes Produkt in den Verkehr bringt.[261] In Abhängigkeit von dem konkreten Produktfehler ist in diesem Zusammenhang auch ein Ausschluss der Produkthaftung nach § 1 Abs. 2 Nr. 5 ProdHaftG denkbar. Demgemäß ist die Haftung des Herstellers ausgeschlossen, wenn der Fehler[262] nach dem Stand der Wissenschaft und Technik im Zeitpunkt des durch den Hersteller erfolgten Inverkehrbringens des Produktes nicht erkannt werden konnte. Von der Haftung ausgenommen werden sollen sogenannte Entwicklungsrisiken.[263] Die verschuldensunabhängige Haftung des Herstellers ist begrenzt auf das objektiv Mögliche und mit Blick auf Gefahren auf die Umsetzung des Kenntnisstandes, der im Zeitpunkt des Inverkehrbringens vorlag, beschränkt. Gehaftet werden soll nicht für eine von einem Produkt ausgehende Gefahr, die in der Entwicklungs- und Konstruktionsphase bei Anwendung aller zumutbarer Sorgfalt nicht erkennbar war.[264]

5.3.1.5.1.2 Problem: Keine Produktbeobachtung nach ProdHaftG bei weiterlernenden Systemen

Mit dem Inverkehrbringen des Produkts endet also die produkthaftungsgesetzliche Verantwortung des Herstellers. Eine Produktbeobachtungspflicht kennt das ProdHaftG nicht. Ähnlich wie in der 9. ProdSV ist zwar bei Konstruktion des Produkts der gesamte Lebenszyklus in den Blick zu nehmen, eine unmittelbare Haftung für, sich erst nach dem Inverkehrbringen zeigende Fehler sieht das ProdHaftG nicht vor.

> *Damit kann für weiterlernende Systeme hinsichtlich der Haftung auf die Überlegungen zu den Herstellerpflichten nach der 9. ProdSV verwiesen werden: Je veränderbarer ein System ist, desto eher könnte bei Inverkehrbringen noch unvorhersehbares Verhalten des Systems zu Schäden führen, die nicht durch die Haftung nach ProdHaftG ausgeglichen werden können. Aber auch eine bei Konstruktion schon vorhersehbare risikobehaftete Veränderbarkeit kann zur Haftung führen, wenn sie falsch eingeschätzt wird und dementsprechend unzureichende Sicherungsmaßnahmen ergriffen werden. Es können auch strengere Anforderungen an die Instruktion zu stellen sein.[265]*

---

[259] *Seibl* in: BeckOGK zum Produkthaftungsgesetz, § 1 Rn. 18 ff.
[260] Bürgerliches Gesetzbuch (BGB), neugefasst durch Beschluss vom 02.01.2002 BGBl. I S. 42, 2909; 2003, 738; zuletzt geändert durch Artikel 24 des Gesetzes vom 20.11.2019 BGBl. I S. 1724.
[261] *Wagner*, Produkthaftung für autonome Systeme, Archiv für die civilistische Praxis, 2017, 709, 712.
[262] Erfasst sind nur Instruktionsfehler und Konstruktionsfehler, vgl. *Seibl* in: BeckOGK zum Produkthaftungsgesetz, § 1 Rn. 121.
[263] BGH NJW 1995, 2162.
[264] BT-Drucksache 11/2447, S. 7, 15.
[265] *Kreutz* in: Oppermann/Stender-Vorwachs, Autonomes Fahren, 3.1.2 Rn. 37.

5.3.1.5.1.3 Deliktische Haftung nach § 823 BGB

Neben dem Produkthaftungsrecht nach dem ProdHaftG steht die Haftung aus unerlaubter Handlung nach § 823 BGB, die sogenannte Produzentenhaftung[266]. Geschützt sind nach § 823 Abs. 1 BGB Leben, Körper, Gesundheit, die Freiheit, das Eigentum oder sonstige Rechte. Der Kreis der geschützten Rechtsgüter geht also weiter als bei der Haftung nach dem ProdHaftG. Es gibt zudem keine Haftungsobergrenze. Auch wird ein anderer, weiterer Produktbegriff verwendet. So sind auch Dienstleistungen und Software aus der Cloud Produkte im Sinne der Produzentenhaftung.[267]
Im Rahmen der Produzentenhaftung gemäß § 823 Abs. 1 BGB ist die schuldhafte Verletzung von Verkehrssicherungspflichten des Herstellers eines Produkts haftungsbegründend.[268] Wie diese Verkehrssicherungspflicht ausgestaltet ist, hängt vom Einzelfall ab. Voraussetzung für die deliktische Produzentenhaftung nach § 823 BGB ist jedenfalls ein Verschulden.[269]
Nach § 823 Abs. 2 BGB kann die Produzentenhaftung durch die Verletzung eines Verbotsgesetzes begründet werden. Als Verbotsgesetz kommen insbesondere die Regeln des ProdSG und der ProdSV in Betracht. Im Folgenden wird kurz erläutert, wie die beiden Haftungstatbestände des § 823 BGB mit dem öffentlich-rechtlichen Produktsicherheitsrecht zusammenwirken.
Die Verletzung von Verkehrssicherungspflichten kann die Haftung nach § 823 Abs. 1 BGB begründen. Die bei der Produzentenhaftung relevanten Pflichtverletzungen werden gegenstandsbezogen formuliert. Eine Pflichtverletzung liegt vor, wenn das Produkt fehlerhaft ist, also wenn ein Konstruktions-, Fabrikations-, Instruktions- oder Produktbeobachtungsfehler vorliegt.[270] Welcher Sorgfaltsmaßstab anzulegen ist, bestimmt sich nach dem Verhältnis zwischen den Kosten für die Maßnahme zur Gewährleistung der Sicherheit und dem durch sie vermiedenen Schaden. Es ist diejenige Sorgfalt zu verlangen, die objektiv erforderlich und zumutbar ist. Je größer der drohende Schaden ist, desto höher sind die Anforderungen an die zu treffenden Maßnahmen.[271] Was erforderlich ist, bestimmt sich laut BGH nach dem neuesten Stand der Wissenschaft und Technik. Eine Maßnahme ist demnach erforderlich, wenn sie konstruktiv möglich und geeignet und genügend ist, um Schäden zu vermeiden. Die Möglichkeit der Gefahrvermeidung ist gegeben, wenn nach gesichertem Fachwissen der einschlägigen Fachkreise praktisch einsatzfähige Lösungen zur Verfügung stehen. Diese Lösungen müssen Serienreife haben, lediglich theoretisch verfügbare oder in der Erprobung befindliche Technologien sind nicht erforderlich.[272]
Die so ermittelte Maßnahme muss auch objektiv zumutbar sein. Hier ist die bereits angesprochene Abwägung zwischen Kosten und Nutzen der Maßnahme anzustellen. Dabei sind die wirtschaftlichen Auswirkungen der Maßnahme maßgeblich, im Rahmen derer insbesondere die Verbrauchergewohnheiten, die Produktionskosten, die

---

[266] Zur unterschiedlichen Verwendung der Begriffe Produkthaftung und Produzentenhaftung vgl. *Günther*, Roboter und rechtliche Verantwortung, S. 114 f.
[267] *Wagner* in: Münchener Kommentar BGB, § 823 Rn. 784.
[268] *Redeker,* IT-Recht, Rn. 825, 6. Auflage 2017 (Beck-Online).
[269] Daher wird die Produkthaftung teilweise als verschuldensunabhängig, die Produzentenhaftung als verschuldensabhängig bezeichnet, *Günther*, Roboter und rechtliche Verantwortung, S. 171. Beachte aber, dass auch die Haftung nach ProdHaftG bei Konstruktions- und Instruktionsfehlern subjektive Tatbestandsmerkmale aufweist.
[270] *Wagner* in: Münchener Kommentar BGB, § 823 Rn. 806.
[271] BGH NJW 2009, 1669 (1670).
[272] BGH NJW 2009, 2952 (2953).

Absatzchancen für ein entsprechend verändertes Produkt sowie die Kosten-Nutzen-Relation zu berücksichtigen sind. Diese Abwägung kann auch ergeben, dass eine Vermarktung des Produkts angesichts der drohenden Gefahren zu unterbleiben hat.[273] Die Sicherheitsmaßnahmen sind dabei an den Sicherheitserwartungen der hypothetischen Verwender des Produkts auszurichten. Einem fachkundigen Verwender können andere produktseitige Sicherheitsmaßnahmen zugetraut werden als einem Laien.[274]
Unterliegt das Produkt dem öffentlich-rechtlichen Sicherheitsrecht, also z. B. dem ProdSG und den ProdSV, finden sich dort Sorgfaltsmaßstäbe, die im Schadensersatzprozess herangezogen werden können. Aus einer Verletzung der dort normierten Anforderungen muss jedoch nicht zwingend auf eine Verletzung der Verkehrssicherungspflicht im Sinne des § 823 Abs. 1 BGB geschlossen werden. Umgekehrt garantiert die Einhaltung der produktsicherheitsrechtlichen Anforderungen nicht, dass eine spätere Haftung für trotzdem eintretende Schäden ausgeschlossen ist. Der Produzent darf sich nicht blind auf die öffentlich-rechtlichen Normen verlassen.[275] Gleiches gilt für die technischen Normen, die die Sicherheitsanforderungen des ProdSG und der ProdSV konkretisieren. Hinkt die technische Norm der technischen Entwicklung hinterher oder haben sich im konkreten Fall Gefahren realisiert, die von den technischen Normen nicht oder nicht hinreichend adressiert werden, reicht eine Berufung auf die Einhaltung dieser Normen nicht aus, um dem Vorwurf der Pflichtverletzung zu begegnen.[276]
Wie auch bei der Haftung nach ProdHaftG sind nur solche Gefahren durch Sicherungsmaßnahmen abzuwenden, die ex ante vorhersehbar und in zumutbarer Weise vermeidbar waren. Der Produzent soll nicht für Entwicklungsrisiken haften. Waren Risiken also im Zeitpunkt der Entwicklung nicht objektiv erkennbar, so kann das dem Produzenten in einem späteren Schadensersatzprozess rückblickend nicht zum Vorwurf gemacht werden.[277] Gleiches gilt, wie bereits dargestellt, für sog. Entwicklungslücken, also solche Risiken, die zwar erkennbar waren, aber nach dem Stand der Wissenschaft und Technik im Zeitpunkt der Herstellung nicht beseitigt werden konnten. Die Grenze sind dabei solche unvermeidbaren Risiken, die einer Vermarktung des Produkts schlicht entgegenstehen. Hier kommt es wieder auf die bereits dargestellte Abwägung an. In zeitlicher Hinsicht endet also die Verantwortung des Herstellers für Entwicklungsrisiken mit der Vermarktung des Produkts. Trotzdem können sich bei einer Produktserie aus den Erfahrungen mit im Verkehr befindlichen Modellen des Produkts gesteigerte Anforderungen an die Sicherheit ergeben. Gegenstand der Haftung ist immer das konkrete Produkt, dass den Schaden verursacht, nicht die ganze Serie. Macht der Produzent aber Erfahrungen mit den Produkten am Markt, muss er ggf. darauf reagieren und seine Produktion anpassen. Handelt es sich um ein gänzlich neues Produkt, gelten umso strengere Anforderungen für die Risikoermittlung. Der Produzent muss dann aktiv nach möglichen Risiken forschen.[278]
Den Produzenten können zudem hinsichtlich der in Verkehr befindlichen Produkte **Produktbeobachtungspflichten** treffen. Die Rechtsprechung hat diese Pflicht des Produzenten entwickelt, da die auf den Zeitpunkt des Inverkehrbringens oder der

[273] BGH NJW 2009, 2952 (2954).
[274] *Wagner* in: Münchener Kommentar BGB, § 823 Rn. 811.
[275] *Wagner* in: Münchener Kommentar BGB, § 823 Rn. 444.
[276] BGH NJW 1994, 3349 (3350).
[277] BGH NJW 2009, 2952 (2955).
[278] *Wagner* in: Münchener Kommentar BGB, § 823 Rn. 816 f.

Inbetriebnahme gerichteten ordnungsrechtlichen Regelungen zur Risikosteuerung sowie das Produkthaftungsrecht nur die Konstruktionsrisiken abdecken, nicht jedoch die Entwicklungsrisiken. Erst im Nachhinein entdeckte Risiken, die den Produkten innewohnen, können mit der Produktbeobachtungspflicht beherrscht werden. Sie obliegt dem Produzenten, weil er regelmäßig am besten in der Lage ist, Informationen über die Nutzungen und Auswirkungen seiner Produkte in der Praxis zu sammeln und auszuwerten.[279] Die Produktbeobachtungspflicht kann sich bei entsprechender Informationslage auch einer **Reaktionspflicht** verdichten, die dem Hersteller Maßnahmen zur Risikominimierung aufgibt.[280] Kommt er dieser Pflicht nicht nach, haftet er für durch das Unterlassen entstandene Schäden. Hierin liegt der Unterschied zu Konstruktions-, Fabrikations- und Instruktionspflichten: Der haftungsbegründende Vorwurf an den Produzenten lautet bei Verletzung der Produktbeobachtungspflicht nicht, dass er Entwicklungsrisiken nicht erkannt und das so mit verdeckten Risiken behaftete Produkt in Verkehr gebracht hat. Vielmehr hat er es unter Verletzung seiner Produktbeobachtungspflicht unterlassen, auf diese verdeckten Risiken in erforderlicher Weise zu reagieren.

> *Für die Verkehrssicherungspflichten des Produzenten nach § 823 Abs. 1 BGB gilt bei weiterlernenden Systemen das zur Haftung nach ProdHaftG Ausgeführte. Daneben tritt die Produktbeobachtungspflicht. Weiterlernende Systeme können zu strengerer Produktbeobachtung verpflichten. Hier kann die Haftung des Herstellers dann neben die Haftung des Betreibers bzw. Arbeitgebers treten, der das KI-System nutzt.*

5.3.1.5.1.4 Problem: Umfangreiche Produktbeobachtungspflicht nach Produzentenhaftung bei weiterlernendem System

Ein **weiterlernendes System** kann zu umfangreichen Produktbeobachtungspflichten des Herstellers im Rahmen der Produzentenhaftung nach § 823 BGB führen. Da die Produzentenhaftung jedoch richterrechtlich geprägt ist, bergen solche Systeme damit jedenfalls solange eine hohe Rechtsunsicherheit für den Hersteller, wie es keine Normen oder gar ordnungsrechtliche Vorgaben für die Beschaffenheit des Systems und die Integration der Sicherheit auch über den Zeitpunkt des Inverkehrbringens hinaus gibt. Diese wären zwar für das die Verkehrssicherungs- bzw. Produktbeobachtungspflicht im Einzelfall feststellende Gericht nicht bindend bei der Frage, worin diese Pflichten konkret bestanden. Praktisch fungierten derartige Vorgaben aber als wichtige Anhaltspunkte dafür, was dem Stand der Wissenschaft und Technik enstpräche. Wenn bei der Herstellung des Produkts also das mögliche Haftungsrisiko bestimmt werden soll, hat das Produktsicherheitsrecht eine doppelte Funktion. Erstens dient es als Filter, indem es nur solche Produkte auf dem Markt erlaubt, die einen gewissen Sicherheitsstandard erfüllen. Der Hersteller muss sich bei der Herstellung nicht an Einzelfallentscheidungen der Gerichte zur Haftung orientieren, sondern kann grundsätzlich auf die allgemeinen Sicherheitsanforderungen des Produktsicherheitsrechts vertrauen. Im Einzelfall kann jedoch auch ein legal auf dem Markt zirkulierendes Produkt zur Haftung führen, wenn der Hersteller konkret eine Verkehrssicherungspflicht verletzt hat. Hier kommt die zweite Funktion des Produktsicherheitsrechts zum Tragen. Zur Bestimmung der Verkehrssicherungspflichten kann sich das Gericht an den Sicherheitsanforderungen des Produktsicherheitsrechts sowie dem in den technischen Normen dokumentierten

---

[279] *Wagner* in: Münchener Kommentar BGB, § 823 Rn. 837.
[280] *Wagner* in: Münchener Kommentar BGB, § 823 Rn. 836.

Stand der Technik orientieren. Das Produktsicherheitsrecht kann also auch hier eine Orientierung für den Hersteller bieten. Da das Produktsicherheitsrecht jedoch keine Sicherheitsanforderungen stellt, die die Veränderbarkeit von Produkten adressieren, besteht für den Hersteller auch bei der Frage nach drohenden Haftungsrisiken Ungewissheit.
Anders verhält es sich mit der Haftung nach ProdHaftG. Dort besteht keine weitergehende Produktbeobachtungspflicht.

5.3.1.5.1.5 Exkurs: Verhältnis Produktsicherheitsrecht und Produkthaftung bzw. Produzentenhaftung

Der eingangs erwähnte Steuerungseffekt des zivilen Haftungsrechts kann auch über den des Produktsicherheitsrechts hinausgehen und von den Produktverantwortlichen mehr verlangen, als ihnen das Produktsicherheitsrecht auferlegt. So kann eine Haftung auch dort entstehen, wo keine öffentlich-rechtlichen Pflichten des Herstellers bestehen. In diese „Lücken“ stößt dann das Haftungsrecht, um im Einzelfall dort für einen Ausgleich zu sorgen, wo das öffentliche Ordnungsrecht Sicherheitsanforderungen nicht geregelt hat oder wo es im Einzelfall aus zivilrechtlicher Sicht zu anderen Ergebnissen kommt. Das Zivilrecht folgt den öffentlich-rechtlichen Vorgaben an die Technik nicht einfach nach und setzt sie zwischen Privaten um, sondern steht selbstständig neben dem öffentlichen Recht. Nach ständiger Rechtsprechung des BGH kann daher ein produktsicherheitsrechtlich zulässiger Sachverhalt vor den Marktüberwachungsbehörden bestehen, in einem zivilrechtlichen Schadensersatzprozess jedoch anders bewertet werden und zur Haftung des Herstellers führen.[281]
Die dadurch auf den ersten Blick scheinbar widersprüchlichen Bewertungen ein und desselben Sachverhalts durch das Recht erklären sich mit den unterschiedlichen Regelungszwecken und -systematik der ihnen zugrundeliegenden Normen. Das zeigt der Blick auf das Produktsicherheitsrecht als Teil des öffentlichen Rechts und der Herstellerhaftung nach dem Produkthaftungsgesetz als Teil des Zivilrechts. Das europarechtlich geprägte ProdSG bezweckt, entsprechend der ProdS-RL, die EU-weit harmonisierte Regelung des Marktzugangs von Produkten unter Einhaltung eines hohen Schutzstandards. Es wird also der Marktzugang in den Blick genommen. Der Hersteller wird in die Verantwortung genommen, sein Produkt entsprechend den Sicherheitsanforderungen zu konstruieren und die formellen Anforderungen zu erfüllen, bevor er es in Verkehr bringt. Tut er dies nicht, ist er polizeipflichtig und kann Adressat von Marktüberwachungsmaßnahmen werden, die darauf abzielen, die von dem rechtswidrig in Verkehr gebrachten Produkt ausgehenden Risiken zu beseitigen oder zu minimieren. Das ProdHaftG, wiederum in Umsetzung der europäischen Produkthaftungsrichtlinie, bezweckt die Harmonisierung der Haftung des Herstellers, um einen unverfälschten Wettbewerb auf dem Binnenmarkt bei gleichzeitig hohem Schutzniveau der Verbraucher vor Schädigung ihrer Gesundheit oder ihres Eigentums durch industriell gefertigte Produkte zu gewährleisten.[282] Bringt der Hersteller ein fehlerhaftes Produkt in Verkehr und schädigt dieses aufgrund des Fehlers einen Verwender, so besteht ein Schadensersatzanspruch. Das ProdSG unterliegt (jedenfalls hinsichtlich der Marktüberwachung durch die Behörden) der Logik der Eingriffsverwaltung. Der Staat hat weitreichende Eingriffsrechte und teils auch -

[281] Vgl. statt vieler BGH NJW 1998, 2905 (2906).
[282] Erwägungsgrund 1 der Produkthaftungsrichtlinie.

pflichten, sofern die Eingriffstatbestände erfüllt sind. Die staatliche Marktüberwachung ist insbesondere an den Verhältnismäßigkeitsgrundsatz gebunden. Maßnahmen zur Gefahrenabwehr unterliegen einer strengen Prüfung. Das Untermaßverbot kann auf der anderen Seite wegen der grundgesetzlichen Schutzpflichten des Staates zu einem Eingreifen zwingen. Das Regelungsregime des Produktsicherheitsrecht hat dabei keinen umfassenden Anspruch, sämtliche Einzelfälle detailliert zu regeln. Einerseits führte das in der Vergangenheit bereits zu überbordenden Regelwerken, die durch die Schwerfälligkeit der Gesetzgebungsverfahren innovationshemmend wirkte. Andererseits kann ein allumfassender Regulierungsanspruch des öffentlichen Rechts dazu führen, dass die mit dem Vollzug beauftragten Behörden personell und sachlich an ihre Grenzen stoßen. Entsprechend wurde im Produktsicherheitsrecht die Informationsgewinnung, die im Verwaltungsverfahren der Vorbereitung einer Entscheidung dient, an die Hersteller ausgelagert, z. B. durch Dokumentationspflichten wie in § 3 Abs. 2 Nr. 2 der 9. ProdSV oder durch Produktbeobachtungspflichten wie in § 6 Abs.
Das Haftungsrecht dagegen überlässt die Regulierung von Schäden, in denen sich Technikrisiken realisiert haben, den Privaten. Der Anspruchsteller hat alle Anspruchsvoraussetzungen darzulegen und zu beweisen. Der Gesetzgeber hilft ihm mitunter durch Beweiserleichterungen, wie die Beweislastumkehr in § 280 Abs. 1 S. 2 BGB hinsichtlich des Verschuldens. In der Produzentenhaftung ist eine Beweislastumkehr ebenfalls richterrechtlich anerkannt: Dort besteht eine (begrenzte) Beweislastumkehr hinsichtlich des Verschuldens des Produzenten, geschuldet dem Umstand, dass sich der Geschädigte mangels Kenntnis der innerbetrieblichen Vorgänge regelmäßig in Beweisnot befindet.[283]. Zudem steht dem Geschädigten mit § 1 Abs. 1 ProdHaftG außerdem ein Anspruch aus Gefährdungshaftung zur Verfügung, der ihm den Beweis des Verschuldens ganz erspart.

5.3.1.5.2 Haftung der Verwender

Als Verwender kommen hier, entsprechend der Darstellung zu den öffentlich-rechtlichen Regelungen, der Arbeitgeber als Verwender von Arbeitsmitteln und der Anlagenbetreiber einer Anlage im Sinne des BImSchG in Betracht. Zudem regelt das Haftpflichtgesetz (HaftPflG)[284] die Haftung des Betreibers bestimmter Anlagen. Sie soll hier auch kurz dargestellt werden.

5.3.1.5.2.1 Vertragliche Gewährleistung der Sicherheit

Vertragliche Pflichten zur Gewährleistung der Sicherheit erwachsen insbesondere dem Arbeitgeber gegenüber dem Arbeitnehmer, dem gegenüber er durch den Arbeitsvertrag (§ 611a BGB) vertraglich gebunden ist. Nach § 618 Abs. 1 BGB ist er verpflichtet, die Räume, Vorrichtungen und Gerätschaften zur Verrichtung der Dienste so einzurichten und zu unterhalten, dass der Arbeitnehmer gegen Gefahren für Leben und Gesundheit soweit geschützt ist, wie es die Natur des Dienstverhältnisses gestattet.
Dazu muss er jedenfalls die Anforderungen des ArbSchG und insbesondere der BetrSichV erfüllen. Sie werden über § 618 Abs. 1 BGB in das zivilrechtliche

[283] BGH, Beschluss vom 24-04-1990 - VI ZR 358/89, NJW 1992, 41, 42.
[284] Haftpflichtgesetz, neugefasst durch Beschluss vom 04.01.1978 BGBl. I S. 145; zuletzt geändert durch Artikel 9 des Gesetzes vom 17.07.2017 BGBl. I S. 2421.

Schuldverhältnis aus dem Arbeitsvertrag transformiert.[285] Zudem sind die nach § 15 SGB VII erlassenen UVV, die Pflicht zur menschengerechten Gestaltung der Arbeit nach § 2 Abs. 1 ArbSchG sowie der medizinische und soziale Arbeitsschutz Teil der vertraglichen Pflichten des Arbeitgebers.[286] Die praktische Bedeutung des § 618 Abs. 1 BGB als Grundlage für die Haftung des Arbeitgebers ist indes gering, da die öffentlich-rechtlichen Pflichten des Arbeitgebers aus dem ArbSchG bereits einen weitreichenden Schutz bieten. Sie sind bußgeldbewehrt und ihre Befolgung wird durch Aufsichtsbehörden und den Betriebsrat überwacht. Außerdem schließt das System der Schadensregulierung durch die gesetzliche Unfallversicherung nach § 104 Abs. 1 und 2 SGB VII die Haftung nach § 618 Abs. 1 und 2 BGB aus.[287] Bei grober Fahrlässigkeit oder Vorsatz seitens des Arbeitgebers hinsichtlich des den Versicherungsfall herbeiführenden Handelns greift jedoch § 110 Abs. 1 SGB VII. Dann steht dem Unfallversicherungsträger der Regressanspruch des Geschädigten zu. Hier wird es dann – sofern nicht Vorsatz vorliegt – auf die Bestimmung der Verkehrssicherungspflichten ankommen, die der Arbeitgeber grob fahrlässig verletzt hat.

> *Zur Bestimmung der Verkehrssicherungspflichten kommt es (auch) auf die jeweils maßgeblichen ordnungsrechtlichen Pflichten des Arbeitgebers an, sodass hier auf die Probleme bei der Gefährdungsbeurteilung zu verweisen ist.*

Bei der Verwendung von Vorrichtungen und Gerätschaften, die als Produkte dem ProdSG unterliegen, erfüllt der Arbeitgeber in der Regel seine Pflicht aus § 618 Abs. 1 BGB, wenn sie den produktsicherheitsrechtlichen Anforderungen entsprechen.[288] Für das Einrichten und Unterhalten sind insbesondere die Anforderungen des ArbSchG und der BetrSichV maßgeblich.[289]

> *Im Ergebnis stellen sich damit in der vertraglichen Haftung zwischen Arbeitgeber und Arbeitnehmer bei der Verletzung der Pflichten hinsichtlich der Sicherheit der Arbeitsmittel, in denen die hier untersuchten KI-Systeme zum Einsatz kommen, dieselben Fragen wie bei den produktsicherheitsrechtlichen und arbeitsschutzrechtlichen Anforderungen.*

5.3.1.5.2.2 Gesetzliche Haftung des Arbeitgebers

Eine Haftung aus Gesetz kann insbesondere aus § 823 BGB abgeleitet werden, wenn der Arbeitgeber es in schuldhafter, also nach § 276 Abs. 1 BGB mindestens fahrlässiger Weise versäumt hat, seinen Verkehrssicherungspflichten (§ 823 Abs. 1 BGB) oder seinen gesetzlichen Pflichten insbesondere aus ArbSchG und BetrSichV nachzukommen (§ 823 Abs. 2 BGB) und dadurch eines der von § 823 BGB geschützten Rechtsgüter, insbesondere Leben und Gesundheit eines anderen, verletzt hat. Auch hier gilt wieder das Haftungsprivileg des § 104 Abs. 1 und 2 SGB VII, weshalb der Haftung gegenüber dem Arbeitnehmer insofern nur eine geringe Bedeutung zukommt, wie bereits dargestellt. Zu beachten ist auch hier, dass bei vorsätzlichem oder grob fahrlässigem Verhalten des Arbeitgebers wieder der Regressanspruch nach § 110 Abs. 1 SGB VII bestehen kann.

---

[285] *Henssler* in: Münchener Kommentar BGB, § 618 Rn. 9.
[286] *Henssler* in: Münchener Kommentar BGB, § 618 Rn. 11.
[287] *Henssler* in: Münchener Kommentar BGB, § 618 Rn. 6 f.
[288] *Henssler* in: Münchener Kommentar BGB, § 618 Rn. 40.
[289] *Henssler* in: Münchener Kommentar BGB, § 618 Rn. 45.

5.3.1.5.2.3 Gesetzliche Haftung des Anlagenbetreibers

Der Anlagenbetreiber wird mit dem Geschädigten regelmäßig nicht vertraglich verbunden sein, sodass hier in erster Linie die gesetzlichen Haftungstatbestände betrachtet werden sollen.

Im Nachbarverhältnis ist dabei grundsätzlich § 906 BGB zu beachten, wonach ein Grundstückseigentümer unwesentliche sowie ortsübliche Immissionen zu dulden hat, die von einem benachbarten Grundstück ausgehen. Immissionen, die nach § 906 BGB als unwesentlich oder ortsüblich zu dulden sind, sind auch gerechtfertigt und können damit keine Haftung nach § 823 Abs. 1 BGB für Schäden an Sachen des Nachbarn der emittierenden Anlage begründen.[290] Dem Nachbarn steht jedoch ein Aufopferungs- und Entschädigungsanspruch nach § 906 Abs. 2 S. 2 BGB zu, wenn die wesentlichen ortsüblichen Beeinträchtigungen nach dem Stand der Technik nicht vermeidbar oder Vorkehrungen dagegen wirtschaftlich dem Verwender nicht zumutbar sind.

> *Bei hochgradig veränderbaren KI-Systemen kann der sie nutzende Nachbar nach dem Stand der Technik aber auch wirtschaftlich nicht dazu in der Lage sein, wesentliche Emissionen seiner Anlage mit KI-Komponenten zu verhindern. Das kann z. B. der Fall sein, wenn das KI-System als Steuerungseinrichtung die emittierende Anlage trotz technischer Sicherheitsvorkehrungen und entsprechender Gestaltung des KI-Systems ausnahmsweise so hochfahren kann, dass die Beeinträchtigungen des Nachbarn wesentlich werden. Dann trifft den Verwender der Anlage der Anspruch aus § 906 Abs. 2 S. 2 BGB.*

Nach § 906 Abs. 1 S. 2 und 3 BGB besteht hier ein Gleichlauf mit den öffentlich-rechtlichen Anforderungen an die Anlage. Nach der Rechtsprechung hat Wesentlichkeit nach § 906 BGB dieselbe Bedeutung wie Erheblichkeit nach §§ 3 Abs. 1, 22 Abs. 1 BImSchG.[291] Was nach dem BImSchG unerheblich ist, ist auch nach § 906 BGB unwesentlich. Das ist jedoch kein zwingender Schuss. Wie auch bei der Produzentenhaftung nach § 823 BGB steht das Haftungsrecht selbstständig neben dem Ordnungsrecht. Maßstab zur Bewertung der Wesentlichkeit der Immissionen ist das Empfinden eines durchschnittlichen Menschen, wobei Natur und Zweckbestimmung des von der Beeinträchtigung betroffenen Grundstücks in seiner konkreten Beschaffenheit eine entscheidende Rolle spielen.[292] Das private Haftungsrecht kann so im Einzelfall zu einer feineren Interessenabwägung kommen.[293] De facto ist die Bedeutung der zivilrechtlichen Haftung hier jedoch gering, da in der Regel unerhebliche Immissionen auch unwesentlich sind.[294]

Erwähnt sei daneben die Haftung des Betreibers nach dem **Haftpflichtgesetz (HaftPflG)[295]**. Dort ist nach § 1 Abs. 1 HaftPflG eine Gefährdungshaftung des Betreibers von Schienenbahnen oder Schwebebahnen sowie nach § 2 Abs. 1 HaftPflG des Betreibers von bestimmten Stromleitungs- oder Rohrleitungsanlagen vorgesehen. Der § 1 Abs. 1 HaftPflG geht historisch auf § 25 des preußischen Gesetzes über Eisenbahn-Unternehmungen – Preußisches Eisenbahngesetz (prEG) – vom

---

[290] BGH NJW 1965, 2099.
[291] BGH NJW 1990, 2465 (2466), BVerwG NJW 1988, 2396 (2397).
[292] BGH NJW 1990, 2465.
[293] *Brückner* in: Münchener Kommentar BGB, § 906 Rn. 21.
[294] *Brückner* in: Münchener Kommentar BGB, § 906 Rn. 36.
[295] Haftpflichtgesetz, neugefasst durch Beschluss vom 04.01.1978 BGBl. I S. 145; zuletzt geändert durch Artikel 9 des Gesetzes vom 17.07.2017 BGBl. I S. 2421.

3.11.1838 zurück.[296] Den bahntypischen Gefahren soll mit einer Gefährdungshaftung des Betreibers begegnet werden. Sie ist verschuldensunabhängig. Der Betreiber haftet nicht für verbotenes Verhalten, wie im Deliktsrecht, sondern aufgrund der Schaffung einer besonderen Gefahr. Wer zur Förderung seiner Zwecke eine solche Gefahr erlaubtermaßen schafft oder unterhält, muss unabhängig vom Verschulden auch für etwaige Schäden einstehen, die bei dem gefahrträchtigen Betrieb auch bei Einhaltung der erforderlichen Sorgfalt entstehen, sofern nicht besondere Ausnahmetatbestände, wie etwa höhere Gewalt, vorliegen.[297]
Weiter sieht § 3 HaftPlfG eine Repräsentantenhaftung vor. Demnach haftet der Betreiber eines Bergwerks, seines Steinbruchs, einer Grube oder einer Fabrik für den Tod oder die Körperverletzung eines Menschen, der oder die durch Verschulden eines Bevollmächtigten oder Repräsentanten oder einer zur Leitung oder Beaufsichtigung des Betriebes oder der Arbeiter angenommenen Person hervorgerufen wurde. Es handelt sich also nicht um eine Gefährdungshaftung, sondern um eine Haftung für fremdes Verschulden, also um eine Gehilfenhaftung: Wer die Vorteile der Arbeitsteilung nutzt, soll auch die Nachteile tragen, also haften.[298] Wegen des Haftungsausschlusses in § 104 Abs. 1 S. 1 SGB VII spielt diese Haftung im Verhältnis zu Angestellten des Betreibers quasi keine Rolle. Relevant wird die Norm daher nur gegenüber nicht unfallversicherten Dritten. Praktisch kommt der Norm jedoch kaum Bedeutung zu.[299]

5.3.1.5.2.4 Exkurs: Pflichtversicherung und Deckungsvorsorge

Für bestimmte Haftungstatbestände hat der Gesetzgeber eine flankierende Pflicht zum Abschluss einer entsprechenden Haftpflichtversicherung eingeführt. Ein bekanntes Beispiel hierfür ist die Kraftfahrzeughaftpflichtversicherung, zu deren Abschluss der Halter eines Kraftfahrzeugs oder Anhängers nach § 1 Pflichtversicherungsgesetz[300] verpflichtet ist. Die Umstände, die zur Regelung einer solchen Pflicht bewegen, sind unterschiedlich. Unter anderem werden dazu das Vorliegen einer Gefährdungshaftung, die Bedrohung besonders wichtiger Rechtsgüter durch eine ausgeübte Tätigkeit oder die massenhafte Existenz insolvenzgefährdeter Haftpflichtiger gezählt. Letztgenannter Umstand ist der Grund für die Versicherungspflicht von KFZ-Haltern.[301] Eine solche Versicherungspflicht findet sich in § 13 Abs. 8 ProdSG für die Konformitätsbewertungsstellen.
Einen ähnlichen Ansatz verfolgt die Pflicht zur Deckungsvorsorge, die sich beispielsweise in § 94 Arzneimittelgesetz (AMG)[302] findet. Der pharmazeutische Unternehmer, den nach § 84 Abs.1 S. 1 AMG eine Gefährdungshaftung treffen kann, muss sich zur Deckung etwaiger Schadensersatzansprüche absichern, was er z. B. durch den Abschluss einer Haftpflichtversicherung tun kann.

5.3.1.5.3 Zwischenergebnis zu Veränderbarkeit und Haftung

---

[296] *M. Vogeler* in: beckOGK, HaftPflG § 1 Rn. 2.
[297] BGH NJW-RR 2009, 959 (961).
[298] *Ballhausen* in: beckOGK, HaftPflG § 3 Rn. 2.
[299] *Ballhausen* in: beckOGK, HaftPflG § 3 Rn. 3.
[300] Pflichtversicherungsgesetz, neugefasst durch Beschluss vom 05.04.1965, BGBl. I S. 213; zuletzt geändert durch Artikel 1 der Verordnung vom 06.02.2017, BGBl. I S. 147.
[301] *Brand* in: Langheid/Wand, Münchener Kommentar zum VVG, Vor § 113 Rn. 4.
[302] Arzneimittelgesetz (AMG), neugefasst durch Beschluss vom 12.12.2005, BGBl. I S. 3394; zuletzt geändert durch Artikel 2 Abs. 1 des Gesetzes vom 25.06.2020, BGBl. I S. 1474.

- *Weiterlernende Systeme begründen für den Hersteller Rechtsunsicherheiten hinsichtlich des Umfangs seiner Produktbeobachtungspflicht aus Produzentenhaftung.*
- *Fehlende ordnungsrechtliche Regulierung führt zu einer größeren Relevanz des auf Einzelfälle bezogenen Haftungsrechts.*

Der zweite Aspekt, die Verlagerung der Regulierung ins Haftungsrecht, ist nicht per se als Problem zu identifizieren. Im Gegenteil kann die Behandlung des Einzelfalls und die juristische „Erschließung" einer neuen Technologie durch die Justiz zu praktikablen und rechtssicheren Lösungen führen. Hier soll lediglich auf die Konsequenz der Abwesenheit ordnungsrechtlicher Regulierung hingewiesen werden. Andererseits muss ordnungsrechtliche Regulierung nicht per se zu weniger häufigen und weniger komplexen Haftungsprozessen führen. Zwischen den verschiedenen Rechtsmaterien des Ordnungsrechts und des Haftungsrechts sowie zwischen den an deren Umsetzung beteiligten staatlichen Organen bestehen komplexe Verhältnisse und Abhängigkeiten, die hier nicht dargestellt werden können. Ein Aspekt jedenfalls kann sein, dass ein abstrakter ordnungsrechtlicher Rahmen viele Fragen löst, die ohne einen solchen im Zivil- oder Strafprozess zu klären wären.[303]
Für die Verwender von weiterlernenden Systemen folgen die Haftungstatbestände vor allem den ordnungsrechtlichen Pflichten der Verwender, sodass insoweit auf die Schwierigkeiten des Arbeitgebers und des Anlagenbetreibers bei Verwendung von weiterlernenden Systemen verwiesen werden kann.

5.3.2 Vernetzung

Eine Vernetzung des Systems ist entweder intern oder nach außen möglich.
Interne Vernetzung bedeutet die Vernetzung mehrerer Teilsysteme zu einem neuen Gesamtsystem.
Die Vernetzung eines Systems nach außen mit einer oder mehreren externen Instanzen kann zunächst abgesprochen erfolgen. Dann kommt es weiter darauf an, ob diese abgesprochen Vernetzung die Funktionen eines Systems verändert oder erweitert oder ob es sich um eine rein informative Vernetzung handelt. Mit der Bedeutung dieser Vernetzung für Art und Güte der Aufgabenerfüllung müssen auch deren Folgen im Falle eines Systemausfalls oder eines fehlerhaften Verhaltens berücksichtigt werden. Vernetzte Systeme unterscheiden sich darin, ob und in welchem Maße ihre Aufgabenbewältigung durch eine Vernetzung mit anderen Systemen beeinflusst wird. Bei unvernetzten, autarken Systemen beruht die Aufgabenausführung allein auf dem, was sie sensorisch erfassen und verarbeiten. Das andere Extrem sind Systeme, die aus vielen miteinander vernetzten Systemen bestehen, welche nur in ihrer Vernetztheit die ihnen zugewiesene Aufgabe erfüllen können.[304] Eine Verbindung zur Dimension der **Veränderbarkeit** eröffnet sich bei Systemkomplexen (z. B. viele autonome Transportsysteme in der Industrie), die den

[303] Vgl. dazu oben 5.3.1.5.1.5
[304] Inhaltlich abzugrenzen ist dieser Aspekt der Vernetzung von der Kommunikation bzw. Interaktion mit Verwendern zum Austausch von Informationen über den Status oder Aufgaben/Intentionen des Systems. Dies wird unter dem Aspekt der **à TRANSPARENZ** adressiert.

Grad der Vernetzung während des Betriebs verändern können und die Erlerntes weitergeben können.
Dagegen kann eine unabgesprochene Vernetzung ad hoc erfolgen. Hier ist insbesondere die Vernetzung über das Internet zu nennen (obwohl auch eine abgesprochene Vernetzung darüber möglich ist). Auch hier kommt es wieder darauf an, ob die durch die Vernetzung von außen kommenden Daten nur informativ oder funktionsbeeinflussend sind.

5.3.2.1 Vernetzung und Produktsicherheitsrecht

Rechtlich bedeutet die **interne Vernetzung** im Anwendungsbereich der 9. ProdSV, dass es sich um eine **Gesamtheit von Maschinen** gemäß § 2 Nr. 2 der 9. ProdSV handelt. Bei den für die einzelnen Teilmaschinen jeweils vorzunehmenden **Risikobeurteilungen** ist zu beachten, wie die Teilmaschinen interagieren. Der Hersteller der Gesamtheit erstellt entsprechend die Gesamtrisikobeurteilung und führt das Konformitätsbewertungsverfahren für die Gesamtheit.

5.3.2.1.1 Gesamtheit von Maschinen

Hersteller ist auch, wer eine **Gesamtheit von Maschine** nach § 2 Nr. 2 der 9. ProdSV konstruiert oder baut. Wann eine Gesamtheit vorliegt, ist mitunter schwierig zu bestimmen. Daher hat das Bundesministerium für Arbeit und Soziales in einer Arbeitsgruppe mit der BAuA und dem Ministerium für Umwelt, Naturschutz und Verkehr des Landes Baden-Württemberg in Abstimmung mit den Marktüberwachungsbehörden der Länder, der Deutschen Gesetzlichen Unfallversicherung e.V. (DGUV), einzelnen Unfallversicherungsträgern sowie dem Verband Deutscher Maschinen- und Anlagenbau ein Interpretationspapier zur Auslegung des Begriffs der Gesamtheit von Maschinen erarbeitet. Demnach bilden mehrere (unvollständige) Maschinen eine Gesamtheit, wenn ein **produktionstechnischer** und ein **sicherheitstechnischer Zusammenhang** zwischen ihnen besteht.[305]
Ein **produktionstechnischer Zusammenhang** besteht demnach, wenn die (unvollständigen) Maschinen als geschlossene Einheit angeordnet sind, als Gesamtheit zusammenwirken, also auf einen gemeinsamen Zweck ausgerichtet sind, und als Gesamtheit betätigt werden, also über eine gemeinsame oder übergeordnete, funktionale Steuerung oder Befehlseinrichtung verfügen.
Ein **sicherheitstechnischer Zusammenhang** besteht dann, wenn ein Ereignis, das nur einen Bestandteil der Gesamtheit betrifft, auf einen anderen Bestandteil wirkt und für diese Gesamtheit wirkende Maßnahmen ergriffen werden müssen, um diese Gesamtheit wieder in einen sicheren Zustand zu bringen. Ist dagegen zwischen in einem produktionstechnischen Zusammenhang wirkenden Maschinen eine Übertragung von Gefährdungen ausgeschlossen, sind beide als Einzelmaschinen zu behandeln. Ist eine mögliche Übertragung der Gefährdungen durch einfache Sicherheitsmaßnahmen auszuschließen, spricht dies auch dafür, dass es sich um Einzelmaschinen handelt.[306]

---

[305] *BMAS*, Interpretationspapier „Gesamtheit von Maschinen“, S. 1.
[306] *BMAS*, Interpretationspapier Gesamtheit von Maschinen, S. 3.

5.3.2.1.2 Problem: Gesamtheit von Maschinen und Risikobeurteilung bei Vernetzung nach außen

Bei einer Vernetzung nach außen mit anderen Systemen ist fraglich, ob eine Gesamtheit entsteht. Ist dies der Fall, sind die einzelnen Risikobeurteilungen und Konformitätsbewertungen für die beteiligten Systeme entsprechend umfangreicher.

> *Sobald sicherheits- und funktionsrelevante Daten von außen in die Maschine eingebunden werden sollen, kann eine Gesamtheit von Maschinen entstehen. Durch eine Vernetzung kann (zumindest für eine Zeit) eine geschlossene Einheit entstehen, deren Teilmaschinen als Gesamtheit wirken. Sofern eine Teilmaschine als „Prozessor" zentral agiert, obliegt ihr die Steuerung der Gesamtheit. Es kann bei sicherheitsrelevanter Vernetzung zudem ein sicherheitstechnischer Zusammenhang bestehen, wenn z. B. die Kompromittierung einer Teilmaschine die Sicherheit der Gesamtheit beeinflusst.*
>
> *Bei extern vernetzten Systemen kann die Abgrenzung zudem schwierig sein, wenn sich die Einzelmaschinen unabgesprochen vernetzen können bzw. vernetzt werden können. Dann stellt sich die Frage, wie die Sicherheit dieser **„ad-hoc-Gesamtheit"** in der Integration der Sicherheit und der Konformitätsbewertung der Einzelmaschinen Niederschlag finden kann.*

Es stellt sich also in einem ersten Schritt die Frage, wie weit die **Gesamtheit der Maschinen** in diesem Fall zu fassen ist. Ob der Maschinenbegriff des § 2 Nr. 2 der 9. ProdSV für eine abschließende und umfassende Risikobeurteilung noch geeignet ist, hängt vom Einzelfall ab. Der Extremfall der unabgesprochenen Vernetzung über das Internet zeigt das mögliche Ausmaß der zu beurteilenden Maschinenkomponenten: Es erfolgt eine Vernetzung mit einer unüberschaubaren Vielzahl von Teilkomponenten dieser neuen (entgrenzten) Gesamtheit. Die Grenzen der im Einzelfall zu konstruierenden Maschine werden konturlos. Dementsprechend wird auch die Risikobeurteilung dieser konkreten Maschine ausgedehnt. Die umfassende und erforderliche Vernetzung wird daher durch die rechtlichen Anforderungen an die Risikobeurteilung und jedenfalls durch die Nachweispflichten des Herstellers im Konformitätsbewertungsverfahren erschwert.

5.3.2.2 Vernetzung und Recht des technischen Arbeitsschutzes

Ein ähnliches Problem wie im Produktsicherheitsrecht für den Hersteller stellt sich für den **Arbeitgeber**, wenn nicht klar ist, was Gegenstand der **Gefährdungsbeurteilung** sein soll. Die Bestimmung der Grenze des Arbeitsmittels ist jedoch solange kein Problem, wie alle sich vernetzenden Arbeitsmittel in der Verantwortung desselben Arbeitgebers liegen. Kommt jedoch eine Vernetzung mit externen Systemen in Betracht, muss der Arbeitgeber dies in die Gefährdungsbeurteilung einfließen lassen. Insbesondere Vernetzung mit externen, also betriebsfremden Systemen, wirft die Frage auf, wie die Sicherheit des konkret zu beurteilenden Arbeitsmittels abschließend bewertet werden soll, wenn die Vernetzung *nach* der Gefährdungsbeurteilung erfolgt. Eine solche Vernetzung kann den Arbeitgeber dann bei der praktischen Umsetzung der Gefährdungsbeurteilung vor Probleme stellen, da er nicht unmittelbar Einfluss auf die externen Systeme hat, gleichzeitig aber für die Sicherheit seiner Arbeitsmittel einstehen muss, die in sicherheitsrelevanter Weise von den externen Systemen abhängen.

5.3.2.3 Vernetzung und Immissionsschutzrecht

Für den Anlagenbetreiber stellt sich eine **externe Vernetzung** bei der **Genehmigung nach BImSchG** dann als problematisch dar, wenn nicht bestimmbar ist, wo die Grenzen der Anlage liegen. Steuerungsimpulse „von außen", also von einem mit der zu genehmigenden Anlage vernetzten System, die zu schädlichen Umwelteinwirkungen führen können, können der Genehmigung entgegenstehen.

5.3.2.4 Vernetzung und DSGVO

Bei der Vernetzung von Systemen, die personenbezogene Daten verarbeiten, ist für Verantwortlichen dieser Systeme die DSGVO zu beachten.
Sofern mehrere Verantwortliche einer Datenverarbeitung gemeinsam über die Zwecke und Mittel der Verarbeitung entscheiden, handelt es sich um **gemeinsame Verantwortliche** nach Art. 26 Abs. 1 S. 1 DSGVO. Die gemeinsam Verantwortlichen müssen in einer Vereinbarung festlegen, wer für welche Teile der Verarbeitung zuständig ist. Damit soll entsprechend dem **Transparenzgebot des Art. 5 Abs. 1 lit. a) DSGVO** den von der Verarbeitung betroffenen Personen ermöglicht werden, zu erkennen, wer die Daten wie verarbeitet und über die Verarbeitung entscheidet. Diese Regelung soll einer zunehmenden Vernetzung datenverarbeitender Dienstleister- auch in der Industrie - gerecht werden.[307] Der EuGH hat einen relativ weiten Begriff der gemeinsamen Entscheidung über die Datenverarbeitung entwickelt.[308] Dies führt dazu, dass bei der Vernetzung von informationstechnischen Systemen, welche personenbezogene Daten verarbeiten und dabei von verschiedenen Betreiber gehostet werden, stets an eine gemeinsame Verantwortlichkeit nach Art. 26 DSGVO und damit auch an die nach Art. 83 Abs. 4 lit. a) DSGVO bußgeldbewehrte Pflicht zur Vereinbarung von Zuständigkeiten zwischen den gemeinsam Verantwortlichen nach Art. 26 Abs. 1 S. 2 DSGVO gedacht werden muss.
Von den gemeinsamen Verantwortlichen zu unterscheiden ist der Auftragsverarbeiter. Dieser hat nach Art. 4 Nr. 8 DSGVO über die vorgenommene Verarbeitung keine Entscheidungsbefugnis, sondern ist ein Dritter, der für den jeweils entscheidenden Verantwortlichen z. B. im Rahmen einer vertraglichen Geschäftstätigkeit die Daten verarbeitet.

> *Bei Systemen, die nach außen vernetzt werden und personenbezogene Daten verarbeiten, stellt sich damit die Frage der Verantwortlichkeit für die Verarbeitung von Daten. So kann der Hersteller des KI-Systems als Dienstleister auftreten und die Lerndatensätze des KI-Systems von außen pflegen und damit die dort enthaltenen personenbezogenen Daten verarbeiten oder zur Fernwartung des Systems, zum fortgesetzten Teaching, zum regelmäßigen Update o. ä. verpflichtet sein und so u. U. zum Verantwortlichen werden. Diese Tätigkeiten können je nach Ausgestaltung als Verantwortlicher oder Auftragsverarbeiter erfolgen.*

5.3.2.5 Haftungsrecht

Die **Vernetzung** nach außen kann im Haftungsfall für den Geschädigten zu Beweisproblemen führen, wenn nicht erkennbar ist, von welchem System die

---

[307] *Ernst* in: Paal/Pauly, DS-GVO, Art. 26 Rn. 8.
[308] Vergleiche die Darstellung bei *Gierschmann*, Gemeinsame Verantwortlichkeit in der Praxis, ZD 2020, 69, 70 f. zu EuGH, Urteil vom 05.06.2018 - C-210/16 „ULD Schleswig-Holstein/Wirtschaftsakademie Schleswig-Holstein" („Facebook-Fanpage"), EuGH, Urteil vom 10.07.2018 – C-25/17 „Zeugen Jehovas", EuGH, Urteil vom 29.07.2019 – C-40/17 „Fashion ID".

Schädigung ausgeht. Denn grundsätzlich muss der Geschädigte als Gläubiger bzw. Kläger darlegen und beweisen, wer durch welche Verletzungshandlung den Schaden (schuldhaft) verursacht hat. Hier kann bei der deliktischen Haftung die Kausalitätsvermutung des § 830 Abs.1 S. 1 BGB helfen. Allerdings bleibt im Falle der verschuldensabhängigen Haftung das Problem, dass bei mangelnder **Kontrollierbarkeit durch Involviertheit des Menschen** ein Verschulden des für das schädigenden System Verantwortlichen schwer nachweisbar ist.

5.3.2.6 Zwischenergebnis Vernetzung

- *Externe unabgesprochene Vernetzung führt dazu, dass ggf. eine neue Gesamtheit von Maschinen entsteht und die Herstellerverantwortung dafür schwierig zu klären ist.*
- *Bei externer Vernetzung sind die Risikobeurteilung für den Hersteller, die Gefährdungsbeurteilung für den Arbeitgeber und die Risikoermittlung für den Anlagenbetreiber erschwert, da Daten von außen berücksichtigt werden müssen.*
- *Die Bestimmung der Verantwortlichkeit nach DSGVO kann bei externer Vernetzung erschwert sein.*
- *Entsprechendes gilt für die Ermittlung haftungsrechtlicher Verantwortlichkeit in einem umfangreich extern vernetzten System.*

Die Vernetzung stellt für die hier untersuchten Rechtsgebiete eine äußerst relevante Dimension dar. Die externe Vernetzung verschiedener Systeme, die für sich jeweils technikrechtlichen Anforderungen unterliegen, führt dazu, dass die isolierte Betrachtung eines (Teil-) Systems durch das Recht nicht mehr ausreicht, angemessene Regelungen zur Gewährleistung technischer Sicherheit zu treffen. Das (Teil-) System steht nicht mehr als einheitliches und geschlossenes Objekt der Regulierung zur Verfügung, sondern ist eingebettet in einen Zusammenhang, der durch eine sinnvolle holistische Regulierung eingefangen werden muss. Es reicht nicht mehr, dem Hersteller eines vernetzten Produkts aufzugeben, dass sein System für sich genommen den technikrechtlichen Anforderungen entsprechen muss. Sicherheitsrelevante Aspekte finden sich auf einer hohen Stufe der Vernetzung (auch) außerhalb des untersuchten Produkts. Es ist dann zu ermitteln, wer für diese Aspekte verantwortlich ist. Die Frage nach der sinnvollen Zuordnung dieser Verantwortung findet sich in allen hier untersuchten Rechtsgebieten.

5.3.3 Kontrollierbarkeit

Die Dimension der Kontrollierbarkeit beschreibt einerseits, zu welchem Maß und mit welcher Granularität das Systemverhalten beim Entwicklungsprozess und im Betrieb gesteuert werden kann. Linearer Softwarecode ist bezüglich der Wirkungen in der Regel in hohem (bzw. ausrechend hohem) Maße kontrollierbar. Demgegenüber stehen Systeme mit hoher Emergenz bzw. Autonomie, deren Verhalten sich aus Kriterien oder vorgegeben Zielen ergibt und nur bezüglich dieser Kriterien, aber nicht mehr bezüglich des Einzelverhaltens kontrollierbar sind. Die Dimension der

Kontrollierbarkeit beschreibt andererseits mit welchen Maßnahmen, die (sicherheitsrelevanten) Wirkungen hoch emergenter bzw. hoch autonomer Systeme beschränkt und dadurch kontrolliert werden können, indem der Spielraum der Systeme beschränkt wird. Dies kann durch eine *supervised architecture* oder durch Betrieb in einer *sand box*, über die das System nicht „hinauswachsen“ kann, oder durch die Beschränkung des Einsatzbereichs erfolgen.
Weiter kann zwischen der **Datenqualität oder -validität** und der **Datenverfügbarkeit** unterschieden werden, wozu die **Übertragungsqualität und -frequenz** gezählt werden kann. Insbesondere bei vernetzten Systemen ist zu unterscheiden, ob und inwiefern die von außen kommenden Daten vom System selbst oder dem Betreiber kontrolliert werden können.
Für den Hersteller kommt es schon bei der Fertigung eines Produkts darauf an, dass es sich nur so verhält, wie es die Sicherheitsanforderungen des Produktsicherheitsrechts verlangen.

> *Insofern kann die Problematik weiterlernender Systeme auch als Ausdruck einer Kombination aus hoher Veränderbarkeit (weiterlernend) und geringer Kontrollierbarkeit verstanden werden.*

Durch entsprechende Beschränkungen kann der Hersteller das erforderte Maß an Sicherheit garantieren. Die Kontrollierbarkeit durch Beschränkungen wird bei der Minderung der identifizierten Risiken im Rahmen der Risikobeurteilung ausdrücklich in 1.1.2 lit. b) Anhang I der Maschinen-RL gefordert. Demnach soll die Risikominderung zuerst durch inhärente sichere Konstruktion, sofern dies nicht möglich ist durch technische und ergänzende Schutzmaßnahmen und erst zuletzt durch Benutzerinformation erfolgen.
Entsprechende Anforderungen finden sich für die Kontrollierbarkeit durch den Arbeitgeber in der TRBS 1111.
Der Anlagenbetreiber hat durch entsprechende Beschränkungen vornehmlich dafür zu sorgen, dass die Immissionsgrenzwerte nicht überschritten werden und im Rahmen der Störfall-VO die ermittelten Risiken durch notwendige Maßnahmen vermieden oder gemindert werden.
In der DSGVO hat der Verantwortliche jedenfalls dafür zu sorgen, dass durch das datenverarbeitende System die Grundsätze der DSGVO gewahrt bleiben.
Im Haftungsrecht ist die Kontrollierbarkeit Voraussetzung des haftungsbegründenden Tatbestands. So wird für die Begründung der Haftung grundsätzlich gefordert, dass der Haftende durch Inverkehrbringen des Systems die Quelle einer später eingetretenen Schädigung in Verkehr gebracht hat (ProdHaftG, Produzentenhaftung) oder weil er die unmittelbare Kontrolle darüber hat (Haftung aufgrund Verletzung einer Garantenpflicht). Andererseits kann ein Mangel an *individueller* Kontrollierbarkeit im Einzelfall den Schuldvorwurf ausräumen.
Im Ergebnis ist die Kontrollierbarkeit Voraussetzung für die Zulässigkeit der Herstellung bzw. Verwendung eines Systems, wobei sich die Anforderungen an die Kontrollierbarkeit zwischen den Rechtsgebieten unterscheiden.

5.3.4 Transparenz

Der Begriff der Transparenz beschreibt die Transparenz des Systems aus Sicht von Experten einerseits und aus Sicht anderer relevanter Beteiligter andererseits.
Transparenz aus Expertensicht umfasst die Erklärbarkeit, die Beschreibbarkeit der Funktionsgrenzen, die Spezifizierbarkeit der Systemfunktionen sowie die Nachvollziehbarkeit des Systemverhaltens.

Die zu beantwortende Frage lautet, ob ein Experte in der Lage ist, Systemverhalten genau vorzugeben und zu beschreiben und nachzuvollziehen, wie das System in allen erdenklich möglichen Situationen agieren wird. Die Nachvollziehbarkeit hängt davon ab, wie komplex das System, seine Aufgaben und seine Umwelt ist. Sie wird qualitativ verändert, sobald ein System nicht mehr auf einem vorab explizierten, sondern einem implizit erlernten Regelwerk beruht. Diese Expertenrolle kommt im Zuge einer kritischen Beurteilung der Sicherheit eines Systems zu Tragen und nicht im (un-) mittelbaren Umgang mit dem System während des Betriebs. Doch auch während des Betriebs erscheint die Transparenz des Systemverhaltens – je nach Aufgabenfeld und Umwelt – als notwendige Voraussetzung eines sicheren Umgangs mit dem System.
Aus Beteiligtensicht geht es ebenfalls um die für einen sicheren Umgang erforderlichen Informationen. Dazu gehören das Wissen um die Funktionen des Systems, um den Einsatzbereich und die Grenzen des Systems sowie die Vorhersehbarkeit der Dynamik des Systems und damit die Vorhersehbarkeit zukünftiger Systemzustände.

5.3.4.1 Transparenz und Produktsicherheitsrecht

Die Transparenz des Systems findet sich auch als Voraussetzung für ein Inverkehrbringen nach dem **ProdSG** wieder. So muss der Hersteller dafür sorgen, dass nach § 3 Abs. 2 Nr. 3 der 9. ProdSV die erforderlichen Informationen, insbesondere die **Betriebsanleitung** beigefügt sind. Es muss für die Verwender erkennbar sein, wo die Grenzen der Maschine liegen, wie sie sicher zu installieren, warten, bedienen und zu demontieren ist.
Durch den Konformitätsnachweis und die technischen Unterlagen wird zudem dokumentiert, wie das Produkt die jeweiligen Sicherheitsanforderungen erfüllt. Dadurch wird eine **Transparenz des Systems gegenüber den Marktüberwachungsbehörden** hergestellt.

5.3.4.1.1 Marktüberwachung

Marktüberwachung meint im Anwendungsbereich der 9. ProdSV die Kontrolle des Produkts nach dessen Bereitstellung am Markt. Es geht also um ein dem Markt nachgelagertes hoheitliche **Produktrisikomanagement durch die zuständigen Behörden**.
Der § 7 Abs. 1 S. 1 der 9. ProdSV sieht die Generalklausel für diese hoheitliche Marktüberwachung vor. Konkretisiert wird diese spezialgesetzliche Ermächtigung durch den § 26 ProdSG, der die Generalklausel in der 9. ProdSV ergänzt. Es gilt jedoch weiterhin der Grundsatz des § 1 Abs. 4 ProdSG, wonach bei anderslautenden Regelungen in der 9. ProdSV diese Vorrang haben.
Diese Marktüberwachung kann dabei in **drei Stufen** unterteilt werden: Der **Informationsgewinnungung** folgt eine **Risikobewertung**, die ggf. zur **notwendigen hoheitlichen Maßnahme** führt.[309] Die hoheitliche Maßnahme kann zudem die Weitergabe von gewonnenen Informationen an andere Stellen beinhalten, z. B. an Marktüberwachungsbehörden anderer Mitgliedstaaten oder die Kommission, sodass dort wiederum die zweite Stufe, also die Risikobewertung ausgelöst werden kann.

5.3.4.1.1.1 Informationsgewinnung

[309] *Schmidt am Busch* in: Eifert, Produktbeobachtung durch Private, 149, 151.

Zur Vorbereitung einer hoheitlichen Maßnahme muss die zuständige Behörde gemäß dem **Untersuchungsgrundsatz des § 24 VwVfG**[310] den Sachverhalt selbst ermitteln, den sie zur Entscheidungsfindung benötigt. Hier betreffen die Ermittlungen dann die Vorbereitung von Maßnahmen gemäß § 7 der 9. ProdSV und § 26 Abs. 2 ProdSG.
Betrachtet man zunächst den corpus delicti, anlässlich dessen Maßnahmen ergriffen werden sollen, also die Maschine, so springt zunächst ein wesentliches Element des Produktsicherheitsrechts ins Auge: **Die Konformitätsvermutung**, deren Bedeutung für die Überwachung des Marktes für Maschinen aus § 7 Abs. 1 S. 2 der 9. ProdSV folgt. Voraussetzung für die Konformitätsvermutung ist die angebrachte CE-Kennzeichnung sowie das Vorliegen der EG-Konformitätserklärung entsprechend den Vorgaben der Maschinen-RL. Zudem können wesentliche Informationen aus den **technischen Unterlagen** und insbesondere der **Betriebsanleitung** entnommen werden. Die dort aufgeführten bestimmungsgemäßen Verwendungen sind letztlich auch maßgeblich für die durch die Marktüberwachungsbehörde durchzuführende Risikobeurteilung.
In § 26 Abs. 1 ProdSG wird auf die **Informationsgewinnung** durch die zuständige Behörde eingegangen. Hierfür sind gemäß § 26 Abs. 1 S. 1 ProdSG zunächst angemessene Stichproben vorgesehen, um zu überprüfen, ob ein Produkt die Anforderungen des Abschnitt 2 des ProdSG erfüllt, also insbesondere auch die Anforderungen des § 3 Abs. 1 ProdSG und damit bei Maschinen der 9. ProdSV und der Maschinen-RL. Mit § 26 Abs. 1 ProdSG wird inhaltsgleich die europarechtliche Regelung der Marktüberwachung in Art. 19 Verordnung 765/2008 übernommen. Die eigentlichen hoheitlichen Befugnisse zur Informationsgewinnung finden sich in § 28 Abs. 1 S. 1 bis 3, Abs. 2 und 3 S. 1 ProdSG.[311] Dort sind u. a. die Befugnisse zum Betreten von Geschäftsräumen und Betriebsgrundstücken, zur Entnahme von Proben und zur Anforderung von Proben, Mustern, Unterlagen und Informationen geregelt.
Die Marktüberwachungsbehörden haben also zunächst **umfangreiche Befugnisse**, um sich die relevanten Informationen zur Sachverhaltsermittlung zu beschaffen. Diese sind durch die Behörden **proaktiv zu nutzen**. Dies folgt aus dem unmittelbar anwendbaren Art. 16 Abs. 3 Verordnung 765/2008, wonach durch Strukturen und Programme für die Marktüberwachung sichergestellt werden muss, dass wirksame Maßnahmen ergriffen werden können. Dies erfordert eine hinreichend aktuelle Erkenntnislage auf Seiten der Behörde. Die Behörde kann sich also nicht in die Passivität zurückziehen und auf die Anzeige möglicherweise regulierungsbedürftiger Sachverhalte warten.[312] Das ergibt auch der Blick auf § 25 Abs. 1 S. 2 Nr. 2 ProdSG, der zusammen mit § 26 Abs. 1 ProdSG zu lesen ist.[313] Demnach sind Marktüberwachungsprogramme zu erstellen, die das Vorgehen der Behörde zur regelmäßigen Überprüfung der Produkte regeln. Die Behörden regeln dabei insbesondere, welche Produktgruppen und dort welche Aspekte der materiellen und formellen Anforderungen an die Produkte sie über einen bestimmten Zeitraum prüfen wollen.[314] Das Marktüberwachungsprogramm ist Teil des **Überwachungskonzepts**,

---

[310] Verwaltungsverfahrensgesetz (VwVfG), neugefasst durch Beschluss vom 23.01.2003, BGBl. I S. 102; zuletzt geändert durch Artikel 5 Abs. 25 des Gesetzes vom 21.06.2019, BGBl. I S. 846.
[311] *Klindt* in: Klindt, Produktsicherheitsgesetz, § 26 Rn. 6.
[312] *Schucht* in: Klindt, Produktsicherheitsgesetz, § 26 Rn. 8.
[313] *Klindt* in: Klindt, Produktsicherheitsgesetz, § 26 Rn. 9.
[314] Vgl. z.B. das Marktüberwachungsprogramm des Landes Baden-Württemberg für den Zeitraum 2018 – 2021, abrufbar unter https://um.baden-wuerttemberg.de/fileadmin/redaktion/m-

auf das die Marktüberwachungsbehörden gemäß § 26 Abs. 1 ProdSG ihre Tätigkeiten stützen müssen.[315] Das Überwachungskonzept umfasst gemäß § 26 Abs. 1 S. 2 Nr. 1 ProdSG auch die Erhebung und Auswertung von Mängelschwerpunkten und Warenströmen. Das dient einer vom konkreten Fall unabhängigen Informationsgewinnung. Zudem dient die hoheitliche **Marktüberwachung als Korrektiv** zum marktliberalen Ansatz des Produktsicherheitsrechts, das Bereitstellen von Produkten grundsätzlich nicht unter einen Erlaubnisvorbehalt zu stellen und somit kein hoheitliches Zulassungsverfahren durchzuführen. Nur eine informiert und proaktiv handelnde Marktüberwachungsbehörde ist in der Lage, diese Rolle effektiv wahrzunehmen.[316]
Neben eigenen Informationsgewinnungsmaßnahmen sollen die Marktüberwachungsbehörden auch über ein **Beschwerdesystem** und den **Informationsaustausch zwischen den Marktüberwachungsbehörden der Mitgliedstaaten** Informationen erlangen. Das folgt aus § 26 Abs. 1 S. 4 ProdSG, wonach eingegangene Beschwerden und sonstige Informationen zu berücksichtigen sind.
Hier sei noch einmal darauf hingewiesen, dass die Marktüberwachungsbehörden die Einhaltung der Anforderungen des 2. Abschnitts, also insbesondere des § 3 ProdSG überwachen. Dazu gehört also auch die **Prüfung der Risikobeurteilung** nach Anhang I der Maschinen-RL. Der **maßgebliche Zeitpunkt** für das Vorliegen der Voraussetzungen des § 3 Abs. 1 ProdSG bestimmt sich für Maschinen gemäß § 3 Abs. 1 der 9. ProdSV. Das ist spätestens die Inbetriebnahme, also die erstmalige bestimmungsgemäße Verwendung.

5.3.4.1.1.2 Marktüberwachungsmaßnahmen

Auf Grundlage des so ermittelten Sachverhalts trifft die Marktüberwachungsbehörde die Entscheidung darüber, ob und wenn ja welche Maßnahme sie ergreift, um die Einhaltung der Anforderungen an das betroffene Produkt zu gewährleisten. Die Maßnahmen reichen dabei bis zur Anordnung von Rücknahmen und Rückrufen oder Verboten der Bereitstellung des Produkts. Trotz der teilweise sehr weitreichenden Maßnahmen, die den Behörden zur Marktüberwachung zur Verfügung stehen, gilt auch hier der Grundsatz, dass vorrangig der Hersteller und die anderen Produktverantwortlichen die Erfüllung der Anforderungen an die Produkte sicherzustellen haben. Es gilt auch hier der **Verhältnismäßigkeitsgrundsatz, normiert in § 40 VwVfG**.[317]
Liegen jedoch die Voraussetzungen für eine Ordnungsverfügung vor, besteht **kein Entschließungsermessen**. Die Marktüberwachungsbehörde muss also die angemessene Maßnahme zur Beseitigung des Verstoßes ergreifen.[318] Hinsichtlich der Auswahl der Maßnahme hat sie indes Ermessen.

---

um/intern/Dateien/Dokumente/6_Wirtschaft/Marktüberwachung/Vorgesehene_Produktkontrollen_2018-2021.pdf (zuletzt abgerufen am 02.03.2020).
[315] Die Marktüberwachungsbehörden werden nach § 32 Abs.4 S. 2 ProdSG von der BAuA bei der Entwicklung und Durchführung des Überwachungskonzepts unterstützt.
[316] *Schucht* in: Klindt, Produktsicherheitsgesetz, § 25 Rn. 6.
[317] *Länderausschuss für Arbeitsschutz und Sicherheitstechnik*, Leitfaden zum Produktsicherheitsgesetz, 26/3, S. 30.
[318] *Schucht* in: Klindt, Produktsicherheitsgesetz, § 26 Rn. 33.

In § 26 Abs. 2 S. 2 ProdSG werden die möglichen Maßnahmen der Marktüberwachungsbehörden aufgezählt. Die Aufzählung ist nicht abschließend.

5.3.4.1.2 Problem: Intransparenz und Marktüberwachung

Bis zu einem gewissen Grad kann Intransparenz durch technische Normen und Prüfverfahren beherrschbar sein. Je intransparenter ein Produkt ist, desto schwieriger werden jedoch der Konformitätsnachweis und die technische Dokumentation, die dazu dienen, Dritten die zur sicheren Verwendung und zur Überprüfung der Konformität notwendigen Informationen zu verschaffen.
Bei „herkömmlichen" Produkten wird die Marktüberwachungsbehörde bei der Untersuchung eines Produkts darauf schließen können, ob die Voraussetzungen im maßgeblichen Zeitpunkt vorlagen und ob im Zeitpunkt der behördlichen Untersuchung dem Produkt ein Risiko innewohnt.

> *Dies kann bei KI-Systemen, unabhängig von ihrer Veränderbarkeit, zu Schwierigkeiten führen, wenn sich diese durch eine ausgeprägte Opazität und Intransparenz auszeichnen und möglicherweise als „Black Box" für die zuständigen Behörden nicht nachvollziehbar sind.[319] Außerdem wird ein solches System die Frage aufwerfen, auf welcher Grundlage die Risikobeurteilung durch den Hersteller erfolgte und die Konformitätsbewertung ausgestellt wurde. Die Marktüberwachungsbehörde wird dann eine genaue Untersuchung des Systems durchführen müssen, bei Zweifeln an der Konformität können auch weitergehende Eingriffe nötig werden. Denn die Marktüberwachungsbehörden sind verpflichtet einzugreifen, wenn der begründete Verdacht besteht, dass ein Produkt die produktsicherheitsrechtlichen Anforderungen nicht erfüllt.*

5.3.4.1.3 Die neue Marktüberwachungsverordnung

Die Marktüberwachungsbehörden werden mit der neuen ab dem 16.07.2021[320] unmittelbar in den Mitgliedstaaten anwendbaren **Marktüberwachungsverordnung (Marktüberwachungs-VO)[321]** gestärkt, indem insbesondere im Online-Handel die Pflichten von Händlern konkretisiert und erweitert werden.
Die Marktüberwachungsmaßnahmen, zu denen die nationalen Marktüberwachungsbehörden durch die Mitgliedstaaten befugt werden müssen, werden in Art. 14 Abs. 4 und 5 Marktüberwachungs-VO geregelt.
Dazu gehören nun ausdrücklich auch das Recht auf **Zugang zu eingebetteter Software** (Art. 14 Abs. 4 lit. a) Marktüberwachungs-VO) und das Recht, die Konformität im Wege von Laboruntersuchung und dabei auch unter **Verwendung von *reverse engineering*** (Art. 14 Abs. 4 lit. j) Marktüberwachungs-VO).
Den Marktüberwachungsbehörden wird nun also ein wirksames Instrumentarium an die Hand gegeben, ggf. die zur Prüfung der Konformität erforderliche Transparenz selbst herzustellen.

---

[319] *Europäische Kommission* COM(2020) 65 final, Weißbuch – Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, S. 14.
[320] Die Regelungen zum neu zu errichtenden Unionsnetzwerk für Produktkonformität gelten bereits ab dem 01.01.2021.
[321] Verordnung (EU) 2019/1020 des Europäischen Parlaments und des Rates vom 20. Juni 2019 über Marktüberwachung und die Konformität von Produkten sowie zur Änderung der Richtlinie 2004/42/EG und der Verordnungen (EG) Nr. 765/2008 und (EU) Nr. 305/2011.

5.3.4.2 Transparenz und Recht des technischen Arbeitsschutzes

Für die ordnungsgemäße Durchführung der Gefährdungsbeurteilung benötigt der Arbeitgeber die erforderlichen Informationen.

5.3.4.2.1 Problem: Gefährdungsbeurteilung bei geringer Transparenz

Für die Gefährdungsbeurteilung hat der Arbeitgeber nach § 3 Abs. 4 S. 1 BetrSichV die erforderlichen Informationen zu beschaffen.

> *Systeme, die schon von vornherein auch für Experten nicht transparent sind, können also auch keiner ordnungsgemäßen Gefährdungsbeurteilung unterzogen werden. Es bedarf also eines bestimmten Grades an Transparenz, der vom Einzelfall abhängt.*

Den Arbeitgeber treffen zudem entsprechende Informationspflichten gegenüber den Beschäftigten, wenn er in Wahrnehmung seiner Pflichten aus § 4 Abs. 1 BetrSichV die nach der Gefährdungsbeurteilung erforderlichen Schutzmaßnahmen ergreift. Das erforderliche Maß an Transparenz des Systems, das durch entsprechende Hinweise, Betriebsanleitung, Schulungen o. ä. erreicht wird, ist im Einzelfall zu ermitteln.

5.3.4.2.2 Überwachung des betrieblichen Arbeitsschutzes

Die Überwachung des betrieblichen Arbeitsschutzes nach der BetrSichV obliegt gemäß § 21 Abs. 1 S. 2 ArbSchG den zuständigen Landesbehörden. Sie teilen sich diese hoheitliche Aufgabe mit den Unfallversicherungsträgern, die gemäß § 17 SGB VII diese Aufgabe wahrnehmen. Länder und Unfallversicherungsträger koordinieren ihre Überwachungstätigkeit in der **Gemeinsamen deutschen Arbeitsschutzstrategie (GDA)**. Die GDA erlässt auf Grundlage des § 20a Abs. 2 Nr. 4 ArbSchG Leitlinien für die Überwachung. An diesen wiederum orientieren sich auch die TRBS. Auf diese Weise können die Leitlinien auch auf die einzelne Gefährdungsbeurteilung einwirken.
Vorrang vor der Überwachung hat die **Beratung der Arbeitgeber** bei der Wahrnehmung ihrer arbeitsschutzrechtlichen Pflichten nach § 21 Abs. 1 S. 2 ArbSchG. Hierzu gehört insbesondere/ unter anderem die **technische Aufklärung**, die auch bei der Einführung neuer Technologien erfolgen kann.[322] Denn die Pflicht des Arbeitgebers zur Gefährdungsbeurteilung beginnt nach § 3 Abs. 3 S. 1 BetrSichV bereits vor der Beschaffung neuer Arbeitsmittel.
Die **Überwachung** meint nach § 21 Abs. 1 S. 2 ArbSchG die Überprüfung, der Einhaltung der Vorgaben des ArbSchG und der Arbeitsschutzverordnungen. Im Zentrum steht dabei die Überprüfung, ob die **Gefährdungsbeurteilung** durch den Arbeitgeber **ordnungsgemäß durchgeführt** wurde.[323] Dementsprechend sind die Unterlagen über die Gefährdungsbeurteilung eine wesentliche Informationsquelle.
Die Aufsichtsbehörden haben hinsichtlich der Überwachung ein **Entschließungs- und Auswahlermessen**. Sie können sowohl die Frequenz der durchgeführten Überwachungsmaßnahmen, als auch die Art der vorgenommenen Kontrollmaßnahme auswählen.[324] Dabei ist der **Verhältnismäßigkeitsgrundsatz** aus § 40 VwVfG zu beachten. Die Überwachung muss geeignet, erforderlich und angemessen sein. Die Aufsichtsbehörde muss durch die Überwachungsmaßnahme insbesondere alle

---

[322] *Schucht* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, ArbSchG § 21 Rn. 19.
[323] *Schucht* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, ArbSchG § 21 Rn. 11.
[324] *Schucht* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, ArbSchG § 21 Rn. 13.

erforderlichen Informationen erlangen können, um die ordnungsgemäße Durchführung der Gefährdungsbeurteilung prüfen zu können.
Die Ermächtigungsgrundlage für Aufsichtsmaßnahmen findet sich in § 22 Abs. 1 und 2 ArbSchG. Dort sind umfangreiche Befugnisse zur Informationsbeschaffung geregelt. Stellt sich aufgrund des ermittelten Sachverhalts heraus, dass Anforderungen des ArbSchG oder der Verordnungen nach §§ 18 und 19 ArbSchG nicht erfüllt sind, können nach § 22 Abs. 3 ArbSchG auch verpflichtende Verwaltungsakte ergehen. Diese können sowohl gegenüber dem Arbeitgeber, als auch gegenüber den sonst gemäß § 13 ArbSchG verantwortlichen Personen erlassen werden. Regelmäßig wird jedoch ein sogenanntes Revisionsschreiben verfasst, in dem der Arbeitgeber auf die, aus Sicht der Aufsichtsbehörde bestehenden, Mängel aufmerksam gemacht wird und er unter Fristsetzung zur Beseitigung dieser aufgerufen wird. Dieses Vorgehen entspricht dem hohen Grad an Eigenverantwortung des Arbeitgebers für die Gewährleistung der Arbeitssicherheit und dem kooperativen Vorgehen der Aufsichtsbehörden. Es handelt sich im Ergebnis eher um eine Beratungsmaßnahme als um eine Maßnahme der Überwachung.[325]
Als **Überwachungsanordnung** kommt insbesondere die Anordnung der Durchführung einer durch die Aufsichtsbehörde konkretisierte Gefährdungsbeurteilung in Betracht. Es kann auch im Wege einer **Ermittlungsanordnung** die Ermittlung von erforderlichen Schutzmaßnahmen zur Beseitigung einer erkannten Gefährdung angeordnet werden.[326] Die **Anordnung konkreter Schutzmaßnahmen** kann sowohl mit Blick auf den **Verhältnismäßigkeitsgrundsatz** als auch bezogen auf das **Bestimmtheitsgebots** aus § 37 Abs. 1 VwVfG problematisch sein. Denn der Arbeitgeber hat einen weiten Ermessenspielraum, wie er die Sicherheit im konkreten Fall durch Schutzmaßnahmen gewährleistet. Zudem ist oft unklar, wie das erforderliche Maß an Sicherheit erreicht werden soll, sodass die schlichte Anordnung von „Maßnahmen“ zur Beseitigung konkreter Gefährdungen unbestimmt ist.

5.3.4.2.3 Problem: Hoheitliche Überwachung bei geringer Transparenz

So wie der Arbeitgeber bei der Gefährdungsbeurteilung ist die Aufsichtsbehörde auf eine gewisse Transparenz des Systems angewiesen, um ihre hoheitliche Aufgabe wahrnehmen zu können.

> *Bei Gefährdungsbeurteilungen zu Arbeitsmitteln mit KI-Komponenten kann sich die Informationsbeschaffung als schwierig erweisen, wenn ein intransparentes KI-System zum Einsatz kommt. Dann kann der Schluss naheliegen, dass die dokumentierte Gefährdungsbeurteilung per se schon nicht mehr aktuell ist. Dadurch wird die Aufsichtsbehörde zu eingehender Prüfung veranlasst sein. Das ist insbesondere dann der Fall, wenn aufgrund möglicher Gefahren, die von dem Arbeitsmittel für die Beschäftigten ausgehen können, wegen des **Untermaßverbots** und den **Schutzpflichten des Staates** eine strengere Überwachung angezeigt ist.*[327]

Anders als bei der Marktüberwachung im Produktsicherheitsrecht besteht jedoch keine Pflicht zum Eingriff bei Verdacht der Verletzung arbeitsschutzrechtlicher Pflichten. Insofern besteht eine größere Flexibilität bei der Überwachung, sodass durch Intransparenz hervorgerufene Unklarheiten ohne strenge Maßnahmen behördlicherseits ausgeräumt werden können.

---

[325] *Wiebauer*, Behördliche Anordnung im Arbeitsschutz, NVwZ 2017, 1653, 1654.
[326] *Wiebauer*, Behördliche Anordnung im Arbeitsschutz, NVwZ 2017, 1653, 1655.
[327] *Schucht* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, ArbSchG § 21 Rn. 13.

5.3.4.3 Transparenz und Immissionsschutzrecht

Der **Anlagenbetreiber** muss im Genehmigungsverfahren alle Informationen beibringen, die die Behörde zur Prüfung der Genehmigungsvoraussetzungen benötigt, das eingesetzte System muss also entsprechend transparent sein. Hier gilt das bereits zum Recht des betrieblichen Arbeitsschutzes Ausgeführte.

5.3.4.4 Transparenz und DSGVO

Das Transparenzgebot der DSGVO stellt ebenfalls Mindestanforderungen an die Transparenz von Systemen, die personenbezogene Daten verarbeiten. Hier interessiert insbesondere der **Auskunftsanspruch** der betroffenen Person.

5.3.4.4.1 Auskunftsanspruch der betroffenen Personen

Die von der Datenverarbeitung betroffenen Personen haben nach Art. 15 DSGVO gegenüber den Verantwortlichen einen Anspruch auf Information über die Datenverarbeitung. Dazu gehört nach **Art. 15 Abs. 1 lit. h) DSGVO** der Anspruch auf Auskunft über die involvierte Logik bei **automatisierter Entscheidungsfindung einschließlich Profiling** nach Art. 22 Abs. 1 DSGVO.
Nach Art. 4 Nr. 4 DSGVO liegt **Profiling** vor, wenn personenbezogene Daten in automatisierter Verarbeitung verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten. In der dann folgenden beispielhaften Aufzählung werden u. a. die Arbeitsleistung und das Verhalten genannt. Ein **Profiling nach Art. 22 Abs. 1 DSGVO** liegt wiederum vor, wenn die **Entscheidung gegenüber** der betroffenen Person **rechtliche Wirkung entfaltet** oder sie **sonst in erheblicher Weise beeinträchtigt**. Es handelt sich also um eine qualifizierte Form des Profilings. Unter die zweite Variante fallen solche Entscheidungen, die die betroffene Person in ihrer Persönlichkeitsentfaltung beeinträchtigen. Dabei ist ein objektiver Maßstab anzulegen. Als Beispiele werden in der Literatur der verweigerte Kreditvertrag oder die verweigerte Zahlungsart beim Onlinekauf genannt.[328] Die Beeinträchtigung muss unmittelbar aus dem Profiling folgen. Ist das Profiling lediglich entscheidungsvorbereitend und folgt die Beeinträchtigung aus einer menschlichen Entscheidung, sind die Voraussetzungen des Art. 22 Abs. 1 DSGVO nicht erfüllt.[329]
Die Norm versucht einen Ausgleich zu schaffen zwischen Schutz der Privatsphäre und der Ermöglichung technischer Innovationen mit erheblichem Wertschöpfungspotenzial.[330] Nach Art. 1 Abs. 2 DSGVO ist jedoch neben dem Schutz personenbezogener Daten auch der Schutz der anderen Grundrechte und Grundfreiheiten Ziel der DSGVO. Jede erhebliche Beeinträchtigung von Grundrechten und Grundfreiheiten durch Profiling eröffnet damit den Anwendungsbereich des Art. 15 Abs. 1 lit. h) DSGVO und begründet damit einen Auskunftsanspruch der betroffenen Person.

---

[328] *Martini* in: Paal/Pauly, DS-GVO, Art. 22 Rn. 27.
[329] *Martini* in: Paal/Pauly, DS-GVO, Art. 22 Rn. 8.
[330] *Martini* in: Paal/Pauly, DS-GVO, Art. 22 Rn. 8.

5.3.4.4.2 Problem: Transparenz und Auskunftsanspruch des Betroffenen

Dies vorausgesetzt, sind viele Konstellationen denkbar, in denen KI-Systeme ein Profiling im Sinne des Art. 22 Abs. 1 DSGVO durchführen. Bei interagierenden oder kollaborierenden KI-Systemen kann das Arbeitstempo für die Kalibrierung der Aktorik maßgeblich sein. Auch wenn dies nicht zu einer *unmittelbaren* Beeinträchtigung führt, weil sich das Arbeitstempo nicht unmittelbar auf das Arbeitsverhältnis auswirkt, so sind auch weitergehende Anwendungen denkbar. Wird das kollaborierende System mit einem Personalmanagement-Tool verknüpft, das Dispositionen innerhalb des Arbeitsverhältnisses treffen kann und dafür das übliche Arbeitstempo eines Beschäftigten zur Grundlage nimmt, so ist der Anwendungsbereich des Art 22 Abs. 1 DSGVO eröffnet.

> *Mag dies auch nur spezielle Fälle betreffen, so ist der Auskunftsanspruch des Art. 15 Abs. 1 lit. h) DSGVO zu berücksichtigen, wenn KI-Systeme personenbezogene Daten verarbeiten. Jedenfalls können intransparente KI-Systeme, deren Logik sich als „black box" darstellt, insofern mit der DSGVO in Konflikt geraten.*

5.3.4.4.3 Problem: Transparenz und Datenschutz-Folgenabschätzung

Nach Art. 36 Abs. 1 DSGVO hat der Verantwortliche die Aufsichtsbehörde zu konsultieren, wenn er zu dem Ergebnis kommt, dass er zur Beseitigung des identifizierten Risikos erforderliche Maßnahmen ergreifen muss. Die Aufsichtsbehörde bewertet die Verarbeitungsvorgänge und kann nach Art. 36 Abs. 2 DSGVO Empfehlungen über Maßnahmen zur Reduzierung des Risikos abgeben und ihre Befugnisse aus Art. 58 DSGVO ausüben.

> *Im Ergebnis werden Verantwortliche, die KI-Systeme einsetzen, welche personenbezogene Daten verarbeiten, nicht umhinkommen, die Funktionsweise der KI-Systeme in einer Datenschutz-Folgenabschätzung zu berücksichtigen und ggf. auch zu dokumentieren.*

5.3.4.5 Zwischenergebnis Transparenz

Die Transparenz des Systems ist in allen Rechtsbereichen relevant. Wenn ein System für Experten nicht mehr transparent ist, dann können die jeweils Verantwortlichen auch nicht ihren gesetzlichen Pflichten nachkommen bzw. können die Behörden die Einhaltung der Sicherheitsanforderungen nicht kontrollieren.

5.3.5 Widerstandsfähigkeit

Die Widerstandsfähigkeit beschreibt das Vermögen von Systemen, trotz innerer oder äußerer Störungen fehlerfrei 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 unter dem Begriff Resilienz gefasst. Die Robustheit eines Systems bildet dessen Fähigkeit ab, einerseits bei nichtvorhergesehenen oder nichtvorhersehbaren Einflüssen und andererseits bei Störungen dennoch bestimmungsgemäß oder zumindest sicher agieren zu können. Die Robustheit ist insbesondere bei Deep-Learning-Algorithmen ein Problem. Mitunter können für den Menschen geringfügig erscheinende oder gar nicht erkennbare Abweichungen von gelernten und bewältigten Situationen zu einem fehlerhaften Systemverhalten in

diesen Situationen führen. Dies wird oftmals am Beispiel der Bilderkennung auf Basis künstlicher neuronaler Netze demonstriert. Von besonderer Sicherheitsrelevanz sind künstliche neuronale Netze, die für die Planung eingesetzt werden, da es dort zu großen Verhaltenssprüngen kommen kann. Bei datenverarbeitenden Programmen, zu denen die hier untersuchten KI-Systeme zählen, ist damit insbesondere die Zuverlässigkeit der **Datenverarbeitung durch das System** gemeint.[331]
Zur Robustheit zählt schließlich die Security bzw. Cybersicherheit, also die **Sicherheit des Systems vor Missbrauch oder böswilligen Eingriffen** von außen.
Die Resilienz des Systems beschreibt die Widerstandsfähigkeit gegen die Folgen eines Systemversagens bzw. ungewollten Systemverhaltens. Dazu können systembezogene Einrichtungen zählen, wie z. B. ein KI-Sicherheitssystem, oder auf das Umfeld bezogene Einrichtungen, z. B. Schutzeinrichtungen für Verwender oder Not-Halt-Schalter.

5.3.5.1 Widerstandsfähigkeit und Produktsicherheitsrecht

Diese Varianten der Widerstandsfähigkeit werden rechtlich bedeutsam für die Risikobeurteilung nach Anhang I der Maschinen-RL. Je geringer die Widerstandsfähigkeit des Systems ist, desto strengere Maßnahmen müssen zur Eindämmung der damit verbundenen Risiken getroffen werden.

5.3.5.2 Widerstandsfähigkeit und Recht des technischen Arbeitsschutzes

Auch das Arbeitsrecht sieht in der BetrSichV verschiedene Pflichten des Arbeitgebers als Verwender eines Arbeitsmittels vor, die ein hohes Maß an Sicherheit gewährleisten sollen. Die Widerstandsfähigkeit des Arbeitsmittels ist hierbei ein wesentlicher Aspekt. Im Rahmen der Gefährdungsbeurteilung nach § 3 Abs. 1 BetrSichV muss der Arbeitgeber alle Maßnahmen ermitteln und ergreifen, die ein sicheres Arbeiten mit dem Arbeitsmittel ermöglichen. Dazu gehört auch zu verhindern, dass durch missbräuchliche Eingriffe oder Verwendungen von dem Arbeitsmittel Gefahren ausgehen. Er kann nur solche Arbeitsmittel mit sicherheitsrelevanter Software einsetzen, deren sichere Datenverarbeitung auch bei versehentlich oder böswillig falschem Dateninput noch gewährleistet ist. Die Arbeitsmittel müssen außerdem auch über die Zeit die erforderliche Robustheit aufweisen.
Um die erforderliche Widerstandsfähigkeit sicherzustellen, besteht z. B. die Pflicht zu wiederkehrenden Prüfungen nach § 14 Abs. 2 BetrSichV: Arbeitsmittel, die schädigenden Einflüssen ausgesetzt sind, die zu Gefährdungen der Beschäftigten führen können, müssen regelmäßig geprüft werden. Damit soll die Robustheit über die Zeit garantiert werden, auch wenn das Arbeitsmittel über die Zeit aufgrund der schädigenden Einflüsse an Zuverlässigkeit einbüßen kann. Die Prüfintervalle werden gemäß § 3 Abs. 6 BetrSichV aufgrund der Gefährdungsbeurteilung festgelegt. Es kommt dabei ganz auf das konkrete Arbeitsmittel an.
Für überwachungsbedürftige Anlagen sieht das Produktsicherheitsrecht für den Betreiber dieser Anlagen Überwachungspflichten vor. So finden sich in §§ 15 – 18 BetrSichV die Pflichten des Betreibers zur Prüfung der Anlage vor Inbetriebnahme, wiederkehrende Prüfungen sowie für bestimmte Anlagen eine Erlaubnispflicht, bei der ein Betrieb erst nach vorheriger Erlaubnis durch die zuständige Behörde zulässig ist. Diese Regelungen sollen gewährleisten, dass bei besonders sicherheitsrelevanten

[331] Bei „herkömmlichen“ Systemen, Produkten und Materialien wird hierzu typischerweise der Aspekt der Materialermüdung zu zählen sein, also die physische Zuverlässigkeit über die Zeit.

Anlagen die erforderliche Robustheit über die Zeit gegeben ist. Im Fahrzeugbereich findet sich eine entsprechende Pflicht zur wiederkehrenden Untersuchung eines Kraftfahrzeugs (Hauptuntersuchung) in § 29 Abs. 1 StVZO[332].
Die zeitpunktbezogenen Pflichten des Herstellers werden hier also wieder durch zeitraumbezogene Pflichten des Arbeitgebers kompensiert.

5.3.5.3 Widerstandsfähigkeit und Immissionsschutzrecht

Die Störfall-VO schreibt in § 8 die Erstellung eines **Sicherheitskonzeptes** und für Betriebsbereiche der oberen Klasse[333] gem. § 9 eines **Sicherheitsberichts** vor. Diese müssen z. B. Maßnahmen zur Gewährleistung eines hohen Schutzniveaus für Mensch und Umwelt vor Störfällen enthalten.
Im Genehmigungsverfahren nach dem BImSchG wird die Widerstandsfähigkeit explizit in der Störfall-VO thematisiert, wo das Erfordernis des Schutzes vor Eingriffen durch Dritte ausdrücklich in § 3 Abs. 2 Nr. 3 und § 4 Nr. 4 Störfall-VO vorgesehen ist. Im BImSchG selbst wird die Widerstandsfähigkeit im Sinne der Robustheit gegen Missbrauch nicht vorausgesetzt. Die Robustheit als Zuverlässigkeit über die Zeit spielt im Immissionsschutzrecht insofern eine Rolle, als dass die Grenzwerte über den Lebenszyklus der Anlage nicht überschritten werden dürfen. Der Betreiber muss also dafür sorgen, dass die Anlage entsprechend robust und damit widerstandsfähig bleibt.

5.3.5.4 Widerstandsfähigkeit und DSGVO

Ist die DSGVO einschlägig, folgen aus ihr sowohl konkrete Pflichten der Verantwortlichen die Datenverarbeitung betreffend. Dazu gehört auch die **Pflicht zur Gewährleistung der Sicherheit der Verarbeitung**. Damit werden ausdrücklich Anforderungen zur Widerstandsfähigkeit des Systems formuliert.

5.3.5.4.1 Sicherheit der Verarbeitung

Nach **Art. 24 Abs. 1 DSGVO** müssen die **Verantwortlichen** die geeigneten technischen und organisatorischen Maßnahmen ergreifen, um eine Verarbeitung gemäß der DSGVO sicherzustellen und einen entsprechenden Nachweis erbringen zu können. Hierbei handelt es sich um die **Grundnorm der technischen Sicherheit im Zeichen des Datenschutzes** nach DSGVO. Die Art. 25 und 32 DSGVO konkretisieren sie.[334]
Unter **technischen Maßnahmen** werden u. a. Anpassungen der Software verstanden, zu den **organisatorischen Maßnahmen** gehören sämtliche, das Umfeld betreffende Maßnahmen, wie beispielsweise Zugriffsprotokolle.[335] Art. 25 Abs. 1 DSGVO nennt zudem beispielhaft die Pseudonymisierung der Daten. Ansonsten belässt es die Vorschrift bei einer Zielbestimmung. Die dementsprechend zu ergreifenden Maßnahmen müssen die Verantwortlichen bestimmen, wobei sie sich am Stand der Technik, den Implementierungskosten, Art, Umfang, Umständen und Zwecken der Verarbeitung sowie der Eintrittswahrscheinlichkeit und der Schwere des Risikos für die

---

[332] Straßenverkehrs-Zulassungs-Ordnung (StVZO), Artikel 1 der Verordnung vom 26.04.2012 BGBl. I S. 679 (Nr. 18); zuletzt geändert durch Artikel 1 der Verordnung vom 26.11.2019 BGBl. I S. 2015.
[333] Dies sind solche, in denen gefährliche Stoffe in größeren Mengen vorhanden sind, § 2 Nr. 2 Störfall-VO.
[334] *Martini* in: Paal/Pauly, DS-GVO, Art. 32 Rn. 7 f.
[335] *Martini* in: Paal/Pauly, DS-GVO, Art. 25 Rn. 28.

Rechte der betroffenen Personen orientieren müssen. Mit **Maßnahmen entsprechend dem Stand der Technik** sind solche Verfahren gemeint, die einem fortgeschrittenen Anspruch gerecht werden, der Begriff geht also über den der „allgemein anerkannten Regeln der Technik“ hinaus. Dabei ist er jedoch weniger streng als der Begriff des „Standes der Wissenschaft und Technik“, wonach Maßstab das aufgrund gesicherter technischer und wissenschaftlicher Erkenntnisse aktuell Mögliche ist.[336] Auch hier kann zur Ermittlung des Standes der Technik auf **technische Normen** zurückgegriffen werden. Ein solches Normenwerk findet sich im **IT-Grundschutz des Bundesamtes für Sicherheit in der Informationstechnik (BSI)[337]**. Hält sich der Verantwortliche an die Vorgaben des IT-Grundschutzes, kann er dies als Nachweis für die Erfüllung der Pflicht aus Art. 25 Abs. 1 DSGVO anführen.[338] Eine Vermutungswirkung begründet der Nachweis jedoch nicht. Darüber hinaus können zur Bestimmung des Standes der Technik die **Technischen Regeln des BSI (BSI-TR)** herangezogen werden.

In **zeitlicher Hinsicht** muss der Verantwortliche nach Art. 25 Abs. 1 DSGVO bereits bei Auswahl der Mittel der Datenverarbeitung dafür sorgen, dass im Zeitpunkt der Verarbeitung durch diese kein Verstoß gegen die Grundsätze der DSGVO vorliegt.[339] **Art. 25 Abs. 1 DSGVO** sieht in technischer Hinsicht vor, dass die Verarbeitungssysteme von vornherein technisch und organisatorisch den Anforderungen der DSGVO entsprechen müssen **(„privacy by design“)**. Es muss bereits bei Programmierung der Verarbeitungssysteme die Konformität mit den gesetzlichen Regelungen beachtet werden. Die **Grundsätze des Datenschutzes**, die stets beachtet werden müssen und durch die einzelnen Rechte der betroffenen Personen und Pflichten der Verantwortlichen konkretisiert werden, finden sich in **Art. 5 DSGVO**. Zu diesen Grundsätzen gehört nach Art. 5 Abs. 1 lit. a) DSGVO auch die **Transparenz der Datenverarbeitung**. Durch das datenverarbeitende Programm soll Datenschutz unter Beachtung dieser Grundsätze ermöglicht werden.[340] **Normadressaten** des Art. 25 Abs. DSGVO sind die Verantwortlichen. Der Hersteller ist nur Adressat, wenn er als (gemeinsam) Verantwortlicher über die Verarbeitung mitentscheidet. Da dies in vielen Anwendungen nicht der Fall ist, weil der Hersteller nach Inverkehrbringen eines Produkts keinen Einfluss mehr auf die Datenverarbeitung nimmt, betrifft ihn die Regelung lediglich indirekt. Es bleibt nach diesem Regulierungsansatz Aufgabe der für die Datenverarbeitung Verantwortlichen, auf dem Markt eine Nachfrage nach entsprechend programmierten Verarbeitungssystemen zu schaffen und so auf den Herstellungsprozess einzuwirken. Zugleich kann von den Verantwortlichen regelmäßig nicht verlangt werden, die ihnen am Markt zur Verfügung

---

[336] *Martini* in: Paal/Pauly, DS-GVO, Art. 25 Rn. 39b ff.
[337] BSI, IT-Grundschutz-Kompendium 2020 (Stand: Februar 2020).
[338] *Djeffal*, IT-Sicherheit 3.0: Der neue IT-Grundschutz, MMR 2019, 289, 292.
[339] *Martini* in: Paal/Pauly, DS-GVO, Art. 25 Rn. 43.
[340] *Martini* in: Paal/Pauly, DS-GVO, Art. 25 Rn. 10.

stehenden Verarbeitungssysteme vor ihrem Einsatz DSGVO-konform umzuprogrammieren.

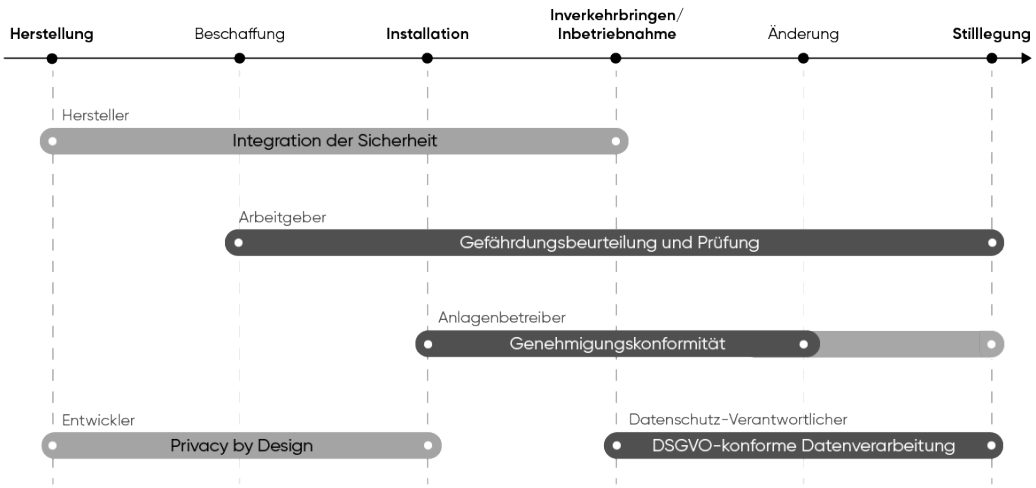


Abb. 5.4 Verantwortlichkeit nach DSGVO (unten). Der Verantwortliche darf nur solche Datenverarbeitungssysteme verwenden, die von vornherein technisch du organisatorisch den Anforderungen der DSGVO entsprechen.

Neben den technischen Datenschutz aus Art. 25 Abs. 1 DSGVO tritt die Pflicht zur datenschutzfreundlichen Voreinstellung **(„privacy by default“)** des **Art. 25 Abs. 2 DSGVO**. Es wird von den Verantwortlichen verlangt, von vornherein den **Grundsatz der Datenminimierung** Art. 5 Abs. 1 lit. c) DSGVO zu beachten. Demnach sollen Daten nur soweit verarbeitet werden, wie für den Zweck der Verarbeitung erforderlich. Eine Datenerhebung außerhalb des eigentlichen Zwecks ist vor dem Hintergrund des großen wirtschaftlichen Wertes personenbezogener Daten verlockend, was durch diese Regelung unterbunden werden soll.[341]

> *Dieser Aspekt kann für KI-Systeme relevant werden, die regelmäßig große Datensätze benötigen. Sammelt das System auch personenbezogene Daten, kann die Frage nach der Erforderlichkeit der Erfassung aufkommen. Diese Frage wird sich wiederum nur dann abschließend klären lassen, wenn erkennbar ist, welches Datum das System warum benötigt. Auch wenn die DSGVO nicht fordert, zu diesem Zweck den Code eines Datenverarbeitungsprogramms offenzulegen (sog. Open-Source), so wird sich der Verantwortliche doch über die Frage der Erbringung eines Nachweises der Erforderlichkeit Gedanken machen müssen. Dieser Nachweis kann insbesondere gegenüber der betroffenen Person relevant werden, wenn sie ihre Auskunftsansprüche geltend macht (siehe dazu die Ausführungen oben zur Transparenz).*

[341] *Martini* in: Paal/Pauly, DS-GVO, Art. 25 Rn. 45.

Den **Nachweis** der Einhaltung der Pflichten aus Art. 25 Abs. 1 und 2 DSGVO kann nach Art. 25 Abs. 3 DSGVO durch Verweis auf die Einhaltung eines genehmigten Zertifizierungsverfahrens nach Art. 42 DSGVO geführt werden. Einen zwingenden Schluss auf die Einhaltung der Pflichten des Art. 25 Abs. 1 DSGVO, vergleichbar mit einer Konformitätsvermutung, sieht das Gesetz jedoch nicht vor.
Nach **Art. 32 Abs. 1 DSGVO** müssen die (ggf. gemeinsamen) **Verantwortlichen und der Auftragsverarbeiter** die Sicherheit der Datenverarbeitung gewährleisten. Damit wird bezweckt, einen **unzulässigen Umgang mit Daten zu verhindern** und die Integrität und Zuverlässigkeit der Daten mittels technischer und organisatorischer Maßnahmen zu gewährleisten.[342] Das Schutzniveau muss sich, wie auch die Pflicht zum technischen Datenschutz nach Art. 25 Abs. 1 DSGVO, an technischen, organisatorischen und wirtschaftlichen Aspekten sowie der Eintrittswahrscheinlichkeit und der Schwere des Risikos für die Rechte der betroffenen Personen orientieren.
Art. 32 Abs. 1 lit. d) DSGVO nennt als mögliche Maßnahme die Einführung **regelmäßiger Überprüfungen, Bewertungen und Evaluierungen** der anderen technischen und organisatorischen Maßnahmen zur Gewährleistung des ermittelten Schutzniveaus. Das können z. B. regelmäßige Penetrationstests sein. In welchem Abstand die Evaluierungen erfolgen müssen, hängt von dem ermittelten Risiko ab. Wie die Maßnahmen konkret ausgestaltet sein müssen, hat der Gesetzgeber wiederum offengelassen.[343]
Wie zuvor bei Art. 25 DSGVO können die Pflichten des Art. 32 DSGVO auch den Hersteller treffen. Da außerdem der Auftragsverarbeiter adressiert ist, kann der Hersteller bereits bei engen vertraglichen Pflichten zur Datenverarbeitung z. B. im Rahmen von Update-Pflichten, bei denen personenbezogene Daten eingesetzt werden, aus Art. 32 DSGVO verpflichtet sein.
Der **Nachweis** der Einhaltung des ermittelten Schutzniveaus kann gemäß Art. 32 Abs. 3 DSGVO durch Verweis auf genehmigte Verhaltensregeln nach Art. 40 DSGVO oder auf die Durchführung eines Zertifizierungsverfahrens nach Art. 42 DSGVO geführt werden.

5.3.5.5 Robustheitund BSI-G

Besondere Anforderungen an die Robustheit im Sinne der Security stellt das Gesetz über das Bundesamt für Sicherheit in der Informationstechnik (BSI-Gesetz)[344]. Für bestimmte Systeme sind dort Vorgaben getroffen, wie sie im Hinblick auf die Security ausgestaltet und betrieben werden müssen. Daher wird das BSI hier im Überblick dargestellt.
Mit dem BSI werden die Aufgaben des Bundesamtes für Sicherheit in der Informationstechnik (BSI) geregelt. Mit der Novellierung von 2017 wurde das BSI-Gesetz an die neue Richtlinie über Maßnahmen zur Gewährleistung eines hohen Sicherheitsniveaus von Netz- und Informationssystemen (NIS-RL)[345] angepasst,

---

[342] *Martini* in: Paal/Pauly, DS-GVO, Art. 32 Rn. 1.
[343] *Martini* in: Paal/Pauly, DS-GVO, Art. 32 Rn. 44 f.
[344] Gesetz über das Bundesamt für Sicherheit in der Informationstechnik (BSI-Gesetz – BSIG) Artikel 1 des Gesetzes vom 14.08.2009 BGBl. I S. 2821 (Nr. 54); zuletzt geändert durch Artikel 13 des Gesetzes vom 20.11.2019 BGBl. I S. 1626.
[345] Richtlinie (EU) 2016/1148 des Europäischen Parlaments und des Rates vom 6. Juli 2016 über Maßnahmen zur Gewährleistung eines hohen gemeinsamen Sicherheitsniveaus von Netz- und Informationssystemen in der Union (NIS-RL), ABl. L 194 vom 19.7.2016, S. 1.

wobei es außer im Bereich der Anbieter digitaler Dienste und einiger kleinerer Änderungen keiner großen Eingriffe in den Inhalt und die Struktur des Gesetzes bedurfte, um die NIS-RL in deutsches Recht umzusetzen.[346]
Das BSI fördert nach § 3 Abs. 1 S. 1 BSI-G die Sicherheit in der Informationstechnik und hat dafür die in § 3 Abs. 1 S. 2, Abs. 3 und 4 BSI-G gelisteten Aufgaben. **Informationstechnik** meint nach § 2 Abs. 1 BSI-G alle technischen Mittel zur Verarbeitung von Informationen. Sicherheit in der Informationstechnik meint nach § 2 Abs. 2 BSI-G die Einhaltung bestimmter Sicherheitsstandards, die die Verfügbarkeit, Unversehrtheit oder Vertraulichkeit von Informationen betreffen, durch Sicherheitsvorkehrungen in informationstechnischen Systemen, Komponenten oder Prozessen oder bei der Anwendung von informationstechnischen Systemen, Komponenten oder Prozessen.

> *Damit sind die hier untersuchten KI-Systeme als technische Mittel zur Verarbeitung von Informationen grundsätzlich vom sachlichen Anwendungsbereich des BSI-G erfasst. Dieser zunächst sehr weite sachliche Anwendungsbereich wird bei den konkreten Befugnissen weiter eingegrenzt und durch den persönlichen Anwendungsbereich jeweils ergänzt.*

Das BSI ist nach § 1 S. 1 BSI-G eine Bundesoberbehörde und untersteht nach § 1 S. 3 BSI-G dem Bundesministerium des Innern.
Um die Relevanz des BSI-G für die hier untersuchten KI-Systeme zu bestimmen, müssen also die einzelnen Aufgaben und Befugnisse des BSI in den Blick genommen werden. Nicht alle sind hier von Bedeutung, weshalb sich die weitere Darstellung auf die Warnungen nach § 7 BSI-G, die Untersuchung der Sicherheit der Informationstechnik nach § 7a BSI-G, die Regelungen zu Betreibern kritischer Infrastrukturen in § 8a BSI-G sowie die Zertifizierung nach § 9 BSI-G beschränkt.

5.3.5.5.1 Warnungen durch das BSI

Das BSI kann die Öffentlichkeit oder interessierte Kreise nach § 7 Abs. 1 BSI-G vor Sicherheitslücken in informationstechnischen Produkten, Schadprogrammen oder im Falle des Verlustes von Daten oder eines unerlaubten Zugriffs darauf warnen. Es kann zudem Sicherheitsmaßnahmen und den Einsatz bestimmter Sicherheitsprodukte empfehlen. Die Warnung der Öffentlichkeit unter Nennung der Bezeichnung und des Herstellers des betroffenen Produkts sowie die Empfehlung von Sicherheitsmaßnahmen und bestimmten Sicherheitsprodukten ist nach § 7 Abs. 3 S. 1 BSI-G nur möglich, wenn hinreichende Anhaltspunkte dafür vorliegen, dass Gefahren für die Sicherheit in der Informationstechnik hiervon ausgehen. Denn eine solche Warnung ist regelmäßig ein schwerer Eingriff gegenüber dem betroffenen Hersteller.

> *KI-Systeme können als informationstechnische Produkte also Gegenstand von Warnungen durch das BSI sein. Wenig widerstandsfähige Systeme können das BSI zu Warnungen veranlassen.*

Dem BSI kommt insofern entsprechend § 3 Abs. 1 S. 1 Nr. 14 BSI-G eine Überwachungsfunktion zu. Eine ähnliche Befugnis findet sich auch in § 26 Abs. 2 S. 2 Nr. 9 ProdSG für die Marktüberwachungsbehörde. Auch sie kann Warnungen aussprechen, allerdings nur, wenn der Hersteller hierzu nicht bereit oder in der Lage ist, in erforderlicher Weise vor seinem Produkt zu warnen.

---

[346] *Kipker*, Umsetzungsgesetz zur NIS-RL nur mit geringen Anpassungen gegenüber der bisherigen Rechtslage beschlossen, MMR-Aktuell 2017, 389121.

5.3.5.5.2 Untersuchung der Sicherheitstechnik durch das BSI

Der § 7a Abs. 1 BSI-G regelt die Befugnis des BSI, auf dem Markt bereitgestellte oder zur Bereitstellung auf dem Markt vorgesehene informationstechnische Produkte und Systeme zu untersuchen. Dazu kann es sich auch der Hilfe Dritter bedienen. Damit soll das BSI in die Lage versetzt werden, seinen Aufgaben als zuständige Behörde für die Abwehr von Gefahren für die die Sicherheit der Informationstechnik des Bundes, für die Beratung und Warnung der Stellen des Bundes, der Länder sowie der Hersteller, Vertreiber und Anwender in Fragen der Sicherheit in der Informationstechnik nachkommen. Auch der Aufgabe der Unterstützung bei der Wiederherstellung der Sicherheit oder Funktionsfähigkeit informationstechnischer Systeme in herausgehobenen Fällen und als zentrale Stelle für die Sicherheit in der Informationstechnik kritischer Infrastrukturen und der Anbieter digitaler Dienste soll das BSI sachgerecht nachzukommen.
Der § 7a Abs. 1 BSI-G erlaubt damit ein Reverse Engineering durch das BSI. Nach dem Willen des Gesetzgebers soll damit das sonst nach § 202a Abs. 1 StGB[347] strafbare Ausspähen von Daten beim Reverse Engineering durch das BSI nicht einschlägig sein.[348] Damit ist dem BSI im Ergebnis wohl auch die sonst nach § 69c Nr. 1 und 2 UrhG[349] ohne Zustimmung des Rechtsinhabers verbotene Dekompilierung erlaubt. Nach 7a Abs. 2 S. 2 BSI-G darf das BSI die dabei gewonnenen Erkenntnisse weitergeben oder veröffentlichen, wenn dies der Erfüllung seiner Aufgaben dient.

> *Das BSI kann damit KI-Systeme auch gegen den Willen des Herstellers eingehender Untersuchungen unterziehen. Die Befugnis greift jedoch erst mit Bereitstellung auf dem Markt bzw. kurz zuvor, wobei jedenfalls aus dem Gesetz kein konkreter Zeitpunkthervorgeht. Eine Art Zulassung oder Prüfung durch das BSI als Voraussetzung für die Bereitstellung auf dem Markt wird damit jedoch nicht begründet.*

5.3.5.5.3 Anforderungen an Kritische Infrastrukturen

Die §§ 8a – 8e BSI-G wurden 2015 im Zuge der Novellierung des BSI-G eingeführt. Sie wurden im Jahr 2017 zur Umsetzung der NIS-RL in Teilen angepasst und ergänzt. In ihrer heutigen Form dienen sie also (auch) der Umsetzung der NIS-RL im Hinblick auf die dort formulierten Anforderungen an die Betreiber wesentlicher Dienste, die im BSI Betreiber kritischer Infrastrukturen genannt werden.[350]
Was zu den kritischen Infrastrukturen zählt, wird nach § 2 Abs. 10 S. 2 BSI-G durch die Verordnung zur Bestimmung kritischer Infrastrukturen nach dem BSI-G (BSI-KritisV)[351] festgelegt. Dort finden sich Kategorien von Anlagen und Schwellenwerte zur Bestimmung ihrer Bedeutung für die Sektoren Energie, Wasser, Ernährung, Informationstechnik und Telekommunikation, Gesundheit, Finanz- und Versicherungswesen sowie Transport und Verkehr. Betreiber von dort aufgeführten

---

[347] Strafgesetzbuch (StGB), neugefasst durch Beschluss vom 13.11.1998 BGBl. I S. 3322; zuletzt geändert durch Artikel 1 des Gesetzes vom 03.03.2020 BGBl. I S. 431.
[348] Vgl. die Gesetzesbegründung zum IT-Sicherheitsgesetz 2015, BT-Drucksache 18/4096, S. 25 und zum Umsetzungsgesetz zur NIS-RL, BT-Drucksache 18/11242, S. 45.
[349] Urheberrechtsgesetz (UrhG), Gesetz vom 09.09.1965 BGBl. I S. 1273; zuletzt geändert durch Artikel 1 des Gesetzes vom 28.11.2018 BGBl. I S. 2014.
[350] BT-Drucksache 18/11242, S. 45.
[351] Verordnung zur Bestimmung Kritischer Infrastrukturen nach dem BSI-Gesetz (BSI-Kritisverordnung - BSI-KritisV), Verordnung vom 22.04.2016 BGBl. I S. 958 (Nr. 20); zuletzt geändert durch Artikel 1 der Verordnung vom 21.06.2017 BGBl. I S. 1903.

Anlagen, die die Schwellenwerte im jeweiligen Anhang erreichen oder überschreiten, unterliegen den Pflichten der §§ 8d – 8e BSI-G.
Zu den Anforderungen gehört zunächst nach § 8a Abs. 1 S. 1 BSI-G, dass sie angemessene organisatorische und technische Vorkehrungen zur Vermeidung von Störungen der Verfügbarkeit, Integrität, Authentizität und Vertraulichkeit ihrer informationstechnischen Systeme, Komponenten oder Prozesse treffen, die für die Funktionsfähigkeit der von ihnen betriebenen kritischen Infrastrukturen maßgeblich sind. Sie müssen dabei nach § 8a Abs. 1 S. 2 BSI-G den Stand der Technik umsetzen. Was der Stand der Technik ist, kann sich nach § 8a Abs. 2 S. 1 BSI-G aus von den Betreibern und ihren Branchenverbänden erarbeiteten branchenspezifischen Sicherheitsstandards (B3S) ergeben. Diese können dem BSI vorgelegt und von ihm nach § 8a Abs. 2 S. 2 BSI-G als geeignet für die Gewährleistung der Anforderungen des § 8a Abs. 1 BSI-G anerkannt werden. Damit können die Branchen also selbst zur Definition des Standes der Technik beitragen.
Die Erfüllung dieser Anforderungen muss dem BSI durch die Betreiber nach § 8a Abs. 3 S. 1 BSI-G alle zwei Jahre nachgewiesen werden, was nach § 8a Abs. 3 S. 2 BSI-G durch Sicherheitsaudit, Prüfungen und Zertifizierungen erfolgen kann. Das BSI kann zudem nach § 8a Abs. 4 S. 1 BSI-G die Einhaltung der Anforderungen des § 8a Abs. 1 BSI-G überprüfen und dafür auch die Geschäfts- und Betriebsräume des Betreibers betreten und von ihm die erforderlichen Informationen verlangen.

> *Dort, wo KI-Systeme in Anlagen der kritischen Infrastruktur zum Einsatz kommen, werden sie als informationstechnische Systeme auch Gegenstand der Pflicht des Betreibers nach § 8a Abs. 1 BSI-G sein. Da das Gesetz jedoch keine konkreten materiellen Anforderungen formuliert, sondern sich auf den Stand der Technik zurückzieht, wird es hier an den Betreibern und Branchenverbänden sowie dem BSI liegen, die Anforderungen an die angemessenen organisatorischen und technischen Vorkehrungen zu definieren. Allein der Turnus von zwei Jahren zum neuerlichen Nachweis könnte bei hochgradig veränderbaren Systemen zu kurz greifen.*

5.3.5.5.4 Zertifizierung durch das BSI

In § 9 Abs. 2 BSI-G wird das BSI ermächtigt, für bestimmte Produkte und Leistungen Sicherheits- oder Personenzertifizierungen sowie Zertifizierungen von IT-Sicherheitsdienstleistern vorzunehmen. Diese Zertifizierungen können nach § 9 Abs. 3 BSI-G durch zuvor durch das BSI anerkannte sachverständige Stellen erfolgen. Die Zertifizierungen und Anerkennungen sind genauer in der BSI-Zertifizierungs- und -Anerkennungsverordnung (BSIZertV)[352] geregelt. Die Zertifizierung kann freiwillig auf Antrag der Hersteller oder Anbieter erfolgen. Sie kann jedoch auch gesetzlich vorgeschrieben sein, so beispielsweise nach § 24 Abs. 1 S. 1, Abs. 4 S. 1 Messstellenbetriebsgesetz[353] für Smart-Meter-Gateways, die als Bestandteil eines intelligenten Messsystems verwendet werden sollen.

---

[352] Verordnung über das Verfahren der Erteilung von Sicherheitszertifikaten und Anerkennungen durch das Bundesamt für Sicherheit in der Informationstechnik (BSIZertV), Verordnung vom 17.12.2014 BGBl. I S. 2231 (Nr. 61); zuletzt geändert durch Artikel 40 des Gesetzes vom 29.03.2017 BGBl. I S. 626.
353 Messstellenbetriebsgesetz (MsbG), Artikel 1 des Gesetzes vom 29.08.2016 BGBl. I S. 2034 (Nr. 43); zuletzt geändert durch Artikel 90 des Gesetzes vom 20.11.2019 BGBl. I S. 1626.

5.3.5.6 Robustheit (Security) und Haftungsrecht

Neben der Produktbeobachtungspflicht des Herstellers nach den Grundsätzen der Produzentenhaftung kann auch der Verkäufer eines Softwareprodukts durch Vertrag zur Gewährleistung dauerhafter Widerstandsfähigkeit des Produkts im Sinne der Gewährleistung der Security verpflichtet sein.
Hervorgehoben werden soll hier eine zukünftig relevante Regelung der neuen europäischen **Richtlinie über den Warenkauf**[354]. Diese Richtlinie gilt für den Verbraucherkaufvertrag, sodass sie nicht für den B2B-Bereich gilt. **Art. 7 Abs. 3 Warenhandels-RL** sieht vor, dass der Verkäufer sicherstellen muss, dass der Käufer von Waren mit digitalen Inhalten Aktualisierungen und Sicherheitsaktualisierungen erhält, die für den Erhalt der Vertragsmäßigkeit der Ware erforderlich sind.

> *Den Verkäufer trifft also über den üblicherweise maßgeblichen Zeitpunkt, die Übergabe der Kaufsache, hinweg eine Leistungspflicht zur Erhaltung der Kaufsache. Er muss die digitalen Inhalte also laufend aktualisieren, wenn dies für die vereinbarte oder vorausgesetzte Funktion erforderlich ist.*

Die Verletzung von Vertragspflichten kann nach § 280 Abs. 1 S. 1 BGB zur Haftung für Schäden führen, die durch die Pflichtverletzung hervorgerufen wurden. Das gilt nach § 437 Nr. 3 BGB auch für Verletzungen von kaufvertraglichen Pflichten. Im Vertragsrecht hilft dem Gläubiger, also in Falle der Aktualisierungspflicht dem Käufer, nach § 280 Abs. 1 S. 2 BGB die Umkehr der Darlegungs- und Beweislast hinsichtlich des Verschuldens der Pflichtverletzung. Wenn der Käufer darlegen und beweisen kann, dass der Verkäufer seiner Aktualisierungspflicht nicht nachgekommen ist, muss der Verkäufer darlegen und beweisen, dass ihn diesbezüglich kein Verschulden trifft. Diese Beweislastumkehr hilft dem Käufer, indem sie ihn von der sonst aufwendigen Beweisführung befreit, die mit dem Beweis des Verschuldens seitens des Verkäufers einhergeht.
Nach Art. 24 Abs. 1 Warenhandels-RL ist die Richtlinie bis zum 01.07.2021 umzusetzen.
Bei der **gesetzlichen Haftung** des Herstellers (also Produkthaftung und Produzentenhaftung) und der Verwender (also insbesondere des Arbeitgebers und des Anlagenbetreibers) ist die Widerstandsfähigkeit im Sinne Security ein wesentlicher Aspekt. Ein Mangel an Widerstandsfähigkeit kann im Einzelfall die Haftung begründen, da er als Produktfehler oder als Verletzung der Sorgfaltspflicht anzusehen ist. Hier geht es jedoch – wie stets im Haftungsrecht – um Einzelfallgerechtigkeit, aus dem Haftungsrecht selbst lassen sich keine allgemeingültigen Aussagen zur Widerstandsfähigkeit einzelner Systeme oder gar Systemarten ableiten.

5.3.5.7 Zwischenergebnis Widerstandsfähigkeit

Die Widerstandsfähigkeit ist in den untersuchten Rechtsgebieten unterschiedlich relevant. Wie die Anforderungen im Einzelnen aussehen, ergibt sich regelmäßig aus technischen Normen. Arbeitgeber, Anlagenbetreiber und der Verantwortliche nach DSGVO haben durch ihre zeitraumbezogenen Pflichten über den gesamten Lebenszyklus dafür zu sorgen, dass auch bei unvorhergesehenen Ereignissen die Widerstandsfähigkeit gewahrt bleibt. Der Hersteller hat die Widerstandsfähigkeit im

---

[354] Richtlinie (EU) 2019/771 des Europäischen Parlaments und des Rates vom 20. Mai 2019 über bestimmte vertragsrechtliche Aspekte der des Warenkaufs, zur Änderung der Verordnung (EU) 2017/2394 und der Richtlinie 2009/22/EG sowie zur Aufhebung der Richtlinie 1999/44/EG, ABl. L 136 vom 22.5.2019, S. 28 (Warenkaufs-RL).

maßgeblichen Zeitpunkt des Inverkehrbringens bzw. der Inbetriebnahme auf den bestimmungsgemäßen Zweck und die erwartbare Fehlanwendung auszurichten.

5.3.6 Involviertheit des Menschen

Involviertheit des Menschen meint die Einbindung des Menschen in das System, der entweder Einfluss auf die Sicherheit des Systems hat oder dessen Sicherheit durch das System beeinflusst wird.
Beide Aspekte finden sich im hier untersuchten Recht wieder. Zunächst ist es stets der Mensch[355], der Adressat von Recht ist. Letztlich ist ein Mensch für die Sicherheit des Systems verantwortlich, sei es als Hersteller oder Verwender (Arbeitgeber, Anlagenbetreiber, Verantwortlicher nach DSGVO oder – retrospektiv – im Haftungsfall der Schuldner). Zudem dient das hier untersuchte Recht grundsätzlich dem Schutz von Menschen bzw. (letztlich Menschen zugeordneten) Sachen.[356]
Die Involviertheit des Menschen ist also für das Recht notwendige Voraussetzung. Ein System, dass weder durch Menschen gesteuert wird oder geschaffen wurde und auch keinen sicherheitsrelevanten Einfluss auf Menschen hat, ist einer Regulierung schon nicht zugänglich.
Die Taxonomie unterscheidet jedoch weiter nach Art und Ausmaß der Involviertheit des Menschen. Diese Unterscheidung ist auch für die Bestimmung des geeigneten Adressaten eines Gesetzes und der Bestimmung des Schutzzwecks maßgeblich. So kann der Mensch sicherheitserhöhend oder sicherheitsmindernd auftreten, er kann als Betroffener über Einfluss auf das System und Expertise im Umgang mit dem System verfügen oder ihm ausgeliefert sein.
Wie die Art der Involviertheit die Verantwortungsverteilung bestimmt, zeigt sich bei allen zuvor diskutierten Dimensionen, für die sich in allen Rechtsgebieten die Frage stellt, ob der Adressat des Rechts bei bestimmten Ausprägungen noch geeigneter Adressat ist, um die Zwecke des Gesetzes zu erreichen. Besonders bei der Veränderbarkeit ist die Verantwortungsverteilung zwischen Hersteller und Verwender zweifelhaft, wenn es sich um ein weiterlernendes System handelt.
Denn der Hersteller ist wegen des zeitpunktbezogenen Ansatzes des ProdSG nicht verpflichtet, bei späteren Änderungen für die Sicherheit des Produkts zu sorgen. Die Verwender hingegen sind in der Pflicht, obwohl sie – anders als bei „herkömmlichen“ Produkten – die Änderungen des Systems nicht in der Hand haben. Der Hersteller hat wiederum durch die Programmierung des Systems die Änderung ihrem Grunde nach verursacht.
Im Bereich des technischen Arbeitsschutzes bestehen zudem besondere Regelungen, nämlich das ArbSchG und die BetrSichV, die der besonderen Rolle des Arbeitnehmers und seinem Verhältnis zum Arbeitgeber Rechnung tragen. Er ist jedoch auch geprägt durch die unterschiedlichen Akteure in dem Bereich (Sozialpartner, Unfallversicherungsträger), die bei der Regulierung beteiligt sind.[357] Die Involviertheit von Arbeitnehmern geht also mit anderen Regulierungsanforderungen einher, als z. B. die Involviertheit von Verbrauchern. Bei Letzteren werden wiederum ihre besonderen Interessen und Bedürfnisse bei der Regulierung berücksichtigt, wie sich z. B. aus Erwägungsgrund 7 der Produktsicherheits-RL ergibt.

---

[355] Gleichwohl können auch juristische Personen Adressaten des hier untersuchten Rechts sein.
[356] Beachte jedoch, dass das BImSchG die Umwelt schützt, von der Menschen mit erfasst sind.
[357] *Wink* in: Kollmer/Klindt/Schucht, Arbeitsschutzgesetz, ArbSchG Vor § 1 Rn. 18 und 21.

Dagegen schützt das Immissionsschutzrecht den unbeteiligten Dritten als involvierten Menschen, sofern er Bewohner der Umgebung einer unter das Immissionsschutzrecht fallenden Anlage ist.
Wenn der Mensch das System nur noch in Gang setzt und sonst nur als Betroffener involviert ist, kann im **Haftungsrecht** eine Haftung dieses Menschen über Garantenpflichten begründet werden. Er haftet dann für das Unterlassen der Abwendung von Gefahren, die von der von ihm geschaffenen Gefahrenquelle ausgehen. In Kombination mit geringer Kontrollierbarkeit kann jedoch fraglich sein, ob dann eine Haftung desjenigen, der das System in Gang setzt, noch sachgerecht erscheint.

5.3.7 Schadensfolgen

Die Bewertung der Schadensfolgen des Systems erfolgt anhand einer aufgaben- und umweltspezifischen Betrachtung des Systems und der von ihm ausgehenden möglichen Gefährdungen. Die Gefährdungen können sich auf die unterschiedlichen geschützten Rechtsgüter beziehen, also auf Leben und Gesundheit von Menschen, auf Sachen, auf die Umwelt oder immaterielle Güter wie Daten. Zudem kann nach dem Ausmaß des jeweils möglichen Schadens unterschieden werden.
Die Bewertung der Schadensfolgen ist rechtlich für die Bestimmung der Verhältnismäßigkeit staatlicher Regulierung relevant.

5.3.8 Kritische Kombinationen von Taxonomiedimensionen

Neben der isolierten Betrachtung der Taxonomiedimensionen soll auch eine Bewertung derjenigen Kombinationen von Taxonomiedimensionen erfolgen, deren Umsetzung bei dem gegebenen Rechtsrahmen problematisch ist.
Die Darstellung soll nicht alle möglichen Kombinationen sämtlicher Taxonomiedimensionen in ihren unterschiedlichen Ausprägungen erfassen. Sie konzentriert sich auf die „Extremfälle“, bei denen stark ausgeprägte Taxonomiedimensionen zusammenkommen.

5.3.8.1 Hohe Veränderbarkeit und geringe Transparenz

Bei KI-Systemen mit hoher Veränderbarkeit stellt sich die Frage, ob der aktuelle Zustand des Systems Rückschlüsse auf die Beschaffenheit des Systems im maßgeblichen Zeitpunkt (der Inbetriebnahme bzw. des Inverkehrbringens) zulässt. Bei geringer Transparenz ist dies nicht möglich. Die Marktüberwachungsbehörden können dann ihrer Aufgabe nicht mehr erfüllen, nämlich zu prüfen, ob nur Produkte in Verkehr sind, die den produktsicherheitsrechtlichen Anforderungen entsprechen gem. § 26 ProdSG und (für Maschinen) § 7 Abs. 1 S. 2 der 9. ProdSV. Denn sie müssen in der Lage sein zu bewerten, wie ein Produkt beschaffen ist und ob es die technischen Anforderungen des Produktsicherheitsrechts und auch der zum Nachweis der Konformität herangezogenen technischen Normen erfüllt. Dabei stehen ihnen umfangreiche Befugnisse zu, die sich aus § 28 Abs. 1 S. 1 bis 3, Abs. 2 und 3 S. 1 ProdSG ergeben. Besteht der Verdacht, dass ein Produkt den produktsicherheitsrechtlichen Anforderungen nicht entspricht, sind die erforderlichen Maßnahmen zur Ermittlung aller relevanten Umstände zu ergreifen. Erweist sich der Verdacht als begründet, muss die zuständige Marktüberwachungsbehörde die zur Beseitigung der dem Produkt innewohnenden Risiken erforderlichen Verfügungen

erlassen. Bei intransparenten veränderbaren Systemen kann bei einer Abweichung der Ist-Beschaffenheit von dem bestimmungsgemäßen Zustand laut Konformitätserklärung im Moment der Untersuchung durch die Marktüberwachungsbehörde der Verdacht entstehen, dass auch bei Inverkehrbringen keine Konformität vorlag. Dann müssen ggf. Maßnahmen folgen, die das mutmaßlich gefährliche und nicht konforme Produkt beherrschbar machen oder gar aus dem Verkehr ziehen. Es muss also eine gewisse Transparenz nach außen hergestellt werden, um die veränderbaren Systeme durch die Marktüberwachungsbehörden überprüfbar zu machen.
Im betrieblichen Bereich ist der Arbeitgeber darauf angewiesen, die für die Durchführung der Gefährdungsbeurteilung erforderlichen Informationen zu erhalten. Er muss diese Informationen gemäß § 3 Abs. 4 S. 1 BetrSichV beschaffen. Dabei kann er auch auf die Bedienungsanleitung zurückgreifen und darauf vertrauen, dass die vom Hersteller mitgelieferten Informationen stimmen. Wenn das Arbeitsmittel aber nach Inbetriebnahme, durch Updates von außen oder, bei einem weiterlernenden System, „von innen“, veränderbar ist, muss diese Veränderung für den Arbeitgeber erkennbar sein, es muss also ein entsprechendes Maß an Transparenz herrschen. Hohe Veränderbarkeit kann also nicht mir geringer Transparenz einhergehen, wenn dem Arbeitgeber dadurch die Durchführung der Gefährdungsbeurteilung unmöglich wird.
Gleiches gilt für den Anlagenbetreiber, der im Genehmigungsverfahren immissionsschutzrechtlich relevante Veränderbarkeiten der Anlage darlegen muss. Auch hier erfordert die Veränderbarkeit einen gewissen Grad an Transparenz, gänzlich intransparente und dabei veränderbare Systeme sind nicht denkbar.
Der nach DSGV Verantwortliche wird ebenfalls keine Datenschutz-Folgenabschätzung anstellen bzw. seinen Auskunftspflichten nachkommen können.

5.3.8.2 Hohe Veränderbarkeit und geringe Widerstandsfähigkeit

Im **Produktsicherheitsrecht** nach ProdSG muss der Hersteller anhand der Widerstandsfähigkeit (Robustheit und Resilienz) des Produkts einerseits die erforderlichen Maßnahmen bestimmen und ergreifen, um das Produkt bei bestimmungsgemäßer Verwendung und erwartbarer Fehlanwendung sicher zu machen. Andererseits gibt die Widerstandsfähigkeit die möglichen Verwendungszwecke vor. Sicherheitsrelevante Veränderungen des Produkts im Betrieb, unabhängig davon, ob sie von außen durch Updates oder „von innen“ durch ein Weiterlernen des Systems initiiert werden, werfen die Frage auf, ob das Produkt dann noch die ursprünglich vorausgesetzte Widerstandsfähigkeit aufweist.
Bereits in der Risikobeurteilung lässt sich so keine abschließende Bewertung der Risiken nach einer solchen Änderung treffen. Im Konzept des ProdSG führt eine sicherheitsrelevante Änderung zu einem neuen Produkt, das dann ggf. einem neuen Konformitätsbewertungsverfahren unterzogen werden muss. Es ist jedoch nicht vorgesehen, dass eine Veränderbarkeit, die Einfluss auf die Widerstandsfähigkeit und damit auf den (eigentlich von Anfang an) anzulegenden Sicherheitsmaßstab hat, bereits bestimmungsgemäß im Produkt angelegt ist. Es handelt sich in der Kombination aus Veränderbarkeit und Widerstandsfähigkeit also um ein anderes Konzept als das, welches z. B. bei Druckgeräten zu der Pflicht regelmäßiger Kontrollen durch den Hersteller führt.[358] Dort wird die Robustheit des Produkts über die Zeit

[358] So z. B. im Konformitätsbewertungsverfahren nach Modul A2 Anhang III der Druckgeräte-Richtlinie.

berücksichtigt, also insbesondere der Verschleiß. Hier hingegen findet eine gezielte Veränderung statt, die sich auf die Widerstandsfähigkeit auswirkt. So können durch ein Software-Update alte Schwachstellen beseitigt werden, aber (ungewollt) neue entstehen. Oder eine durch das System selbst initiierte Änderung des Betriebs führt zu einem schnelleren Materialverschleiß oder zu einer Gewichtung bestimmter Sensorinputs oder Daten, die das System anfälliger für böswillige Manipulation und damit ebenfalls weniger widerstandsfähig macht.
Hier ist also wieder der zeitpunkbezogene Ansatz des produktsicherheitsrechtlichen Konformitätsbewertungsverfahrens problematisch. Im Gegensatz zum Hersteller haben der Arbeitgeber gemäß BetrSichV und der Anlagenbetreiber gemäß BImSchG die Möglichkeit, über die Zeit auch bei veränderbaren Anlagen geringer Widerstandsfähigkeit die Gefährdungsbeurteilung anzupassen bzw. die Änderungen der Immissionsschutzbehörde anzuzeigen und ggf. eine Neugenehmigung zu beantragen. In der Praxis mag dies bei hoher und ständiger Veränderbarkeit wiederum dazu führen, dass eine Kombination hoch ausgeprägter Veränderbarkeit und geringer Widerstandsfähigkeit nicht umsetzbar ist.

5.3.8.3 Hohe Veränderbarkeit und geringe Kontrollierbarkeit durch Involviertheit des Menschen

Im **Haftungsrecht** kann neben der Haftung des Herstellers eine Haftung des Verwenders eines KI-Systems in Betracht kommen. Wer eine Gefahrenquelle schafft, haftet als Garant für das Unterlassen der zur Abwendung eines drohenden Schadens erforderlichen und ihm möglichen Maßnahmen. Er hat entsprechend seiner Verkehrssicherungspflichten und im Rahmen seiner individuellen Möglichkeiten zu handeln. Was die Verkehrssicherungspflichten bei veränderbaren Systemen sind, ist nicht nur, wie gezeigt, für den Hersteller schwierig zu bestimmen. Auch für den Verwender stellt sich die Frage, wie er sich angesichts des von ihm eingesetzten hochgradig veränderbaren Systems verhalten muss. Grundsätzlich wird er sich an den vom Hersteller bereitgestellten Informationen orientieren müssen und das System nur bestimmungsgemäß verwenden dürfen. Insofern setzen sich hier die Schwierigkeiten des Herstellers bei der haftungsrechtlich unbedenklichen Konstruktion des hochgradig veränderbaren Produkts und Instruktion der Verwender hier fort. Bei hoher Veränderbarkeit, auf die der Verwender keinen Einfluss hat, er also nur noch als Betroffener involviert ist, kann ihm jedoch möglicherweise kein Verschuldensvorwurf bei Unterlassen objektiv erforderlicher Gefahrenabwehrmaßnahmen nachgewiesen werden. Hier können unbillige Haftungslücken entstehen, wenn der Verwender zwar die Vorteile aus dem Einsatz der Technik zieht, allerdings mangels Kontrollierbarkeit nicht mehr für die Schäden aufkommen muss.

5.3.8.4 Kontrollierbarkeit und Vernetzung

Im **Produktsicherheitsrecht** sind nach außen vernetzte Systeme nicht vorgesehen. Das führt dazu, dass der Hersteller eines Produkts alle externen Einflüsse, also auch die Produkte fremder Hersteller und andere Datenquellen, bei der Integration der Sicherheit in sein Produkt berücksichtigen muss. Der Vernetzung werden also hier schon praktische Grenzen gesetzt: Der Hersteller wird festlegen, welche äußeren (Teil-) Produkte und Datenquellen er in seine Risikobeurteilung einbeziehen kann. Mit solchen wird eine Vernetzung möglich sein.

Wenn der Hersteller jedoch die Kontrollierbarkeit dieser externen Bestandteile des vernetzten Gesamtsystems nicht abschließend bewerten kann, kann er auch keine abschließende Risikobeurteilung für sein Produkt vornehmen.
Entsprechendes gilt für die **Gefährdungsbeurteilung des Arbeitgebers** und für die **immissionsschutzrechtliche Genehmigungsverfahren**, bei denen durch die Vernetzung der Prüfungsumfang erheblich erweitert wird.
Der bestehende Rechtsrahmen schlägt die Verantwortung allein einer Person in der Kette der Vernetzungen zu. Der Gesamthersteller, der jeweilige Arbeitgeber, der Anlagenbetreiber oder der Verantwortliche nach DSGVO trägt die Verantwortung für die Kontrollierbarkeit und damit Sicherheit des Gesamtsystems. Andererseits kann es zu Doppelprüfungen kommen, wenn das Gesamtsystem wiederum als externes Element Prüfungsgegenstand im Rahmen z. B. eines Genehmigungsverfahrens einer anderen Anlage nach BImSchG ist.
Gegenwärtig sieht sich eine ausgeprägte Vernetzung nach außen also vor allem mit praktischen Umsetzungsproblemen konfrontiert. Im Produktsicherheitsrecht ist die Kontrollierbarkeit von nach außen vernetzten Produkte zudem nicht gesetzlich geregelt.

5.3.8.5 Zwischenergebnis Kombinationen

- *Hochgradig veränderbare Systeme bringen hohe Anforderungen an die Transparenz mit sich, damit die Marktüberwachungsbehörden im Zweifelsfall die Konformität überprüfen und die Verwender ihren gesetzlichen Pflichten zur Gewährleistung eines hohen Sicherheitsniveaus in ihrem jeweiligen Zuständigkeitsbereich nachkommen können.*
- *Der Hersteller hat wegen seiner zeitpunktbezogenen Pflicht Schwierigkeiten, bei hochgradig veränderbaren Produkten die Anforderungen an die Widerstandsfähigkeit sowie die über die bestimmungsgemäße Lebensdauer erforderlichen Maßnahmen zum Erhalt der Widerstandsfähigkeit zu bestimmen*
- *Die Haftung des Verwenders kann bei hochgradig veränderbaren Systemen, die er nicht kontrollieren kann, unbillig erscheinen bzw. mangels Verschuldens im Einzelfall zu Haftungslücken führen.*
- *Für extern vernetzte Systeme mangelt es an einem gesetzlichen Rahmen insbesondere hinsichtlich der erforderlichen Kontrollierbarkeit.*

Es sind daneben noch viele andere Kombinationen und Ausprägungen denkbar, die jeweils unterschiedliche rechtliche Relevanz besitzen. Hier wurde die Darstellung jedoch auf die herausragenden Kombinationen beschränkt.

5.4 Anwendungsbeispiele

Im Folgenden werden Anwendungsbeispiele anhand des vorgestellten Rechtsrahmens begutachtet, denen allesamt eine Verwendung von KI-Systemen in dem jeweiligen Produkt gemein ist. Die Beispiele dienen der Veranschaulichung der in der Taxonomie aufgezeigten rechtlichen Probleme bei der Anwendung des

dargestellten Rechtsrahmens. Dabei handelt es sich um fiktive und verkürzte Sachverhalte.
Es wird geprüft, welche Regelungen problematisch sind, weil sie der Bereitstellung des jeweiligen Produkts entgegenstehen, den Hersteller mit hohen Unsicherheiten hinsichtlich der potentiell einschlägigen nachmarktlichen Pflichten konfrontieren, den Arbeitgeber vor eine komplexe Gefährdungsbeurteilung stellen etc. Die Prüfung ist dabei auf das Aufzeigen der wesentlichen Probleme beschränkt, eine abschließende Prüfung der einzelnen Beispiele erfolgt nicht. Denn Zweck der Begutachtung der Beispiele ist die Ermittlung derjenigen Normen, die bei KI-Systemen Anwendungsschwierigkeiten bereiten, also Rechtsunsicherheit bergen, oder diesen Systemen in der jeweiligen Ausprägung ganz entgegenstehen. Die Prüfung lässt zudem die hoheitliche Überwachung und das Haftungsrecht außer Betracht. Auf sie wird in den Schlussfolgerungen kurz eingegangen.
Dem kursiv dargestellten Sachverhalt des jeweiligen Beispiels folgt vor dem eigentlichen Gutachten eine kurze Hervorhebung der taxonomischen Besonderheiten.

5.4.1 Beispiel 1 – Schweißroboter

Veränderbarkeit – Vernetzung
Ein Roboter schweißt in einem Automobilwerk Karosserieteile nach einem immer gleichen Muster. Rückt das Fließband die Werkstücke in eine vorgegebene Position, wird das von entsprechenden Sensoren des Roboters erfasst, durch Bildverarbeitungsprogramme identifiziert und löst folgenden Reaktionsalgorithmus beim Roboter aus: Der Roboterarm bewegt den Schweißstab an die richtige Position und verschweißt zwei Blechteile der Karosserie. Er steht neben vielen anderen Robotern, die ebenfalls schweißen. Menschen sind während des Schweißens nicht in der Nähe. Ein Zaun mit Sicherheitsfunktionen und ggf. entsprechende Warnhinweise sorgen dafür, dass kein Mensch in den Reaktionsraum des Schweißarms gelangen kann, ohne dass ihm die Missbräuchlichkeit seines Verhaltens bewusst werden muss (z. B. über den Zaun steigen unter Missachtung der Warnhinweise). Im Greifer des Roboters ist eine Kamera verbaut. Mit Hilfe von künstlichen neuronalen Netzen wird der variierende Spalt zwischen den zu verschweißenden Blechteilen vermessen und der zugrundeliegende Algorithmus verändert unter Einbeziehung verschiedener variabler Parameter des Spaltes und des Materials die Dicke der Schweißraupe. Erfasst der Roboter, dass der Spalt zwischen den Blechen größer als zulässig ist, ist eine Reaktionsoption des Algorithmus eine Meldung an die Zentrale zu schicken. Die Meldung geht auch an die anderen Roboter, die daraufhin die Karosserie als Ausschuss „brandmarken“ und nicht weiter bearbeiten.

5.4.1.1 Taxonomie

Das System ist insofern **veränderbar im Betrieb**, als dass es sich der Spaltbreite anpasst und ggf. auf eine unzulässige Spaltbreite reagiert. Die Reaktion des Systems ist gleichwohl von vornherein vorgegeben. Es lernt nicht während des Betriebs, wo die Grenze der zulässigen Spaltbreite erreicht ist. Das System mag vor Inbetriebnahme geteacht worden sein, es ist jedoch im Betrieb „ausgelernt“. Es handelt sich nicht um ein im Betrieb weiterlernendes System.
Es besteht eine **interne Vernetzung**, bei der die Informationen eines Roboters handlungsleitend für die anderen Teilsysteme sind, da der einzelne Roboter mit anderen Robotern und dem Steuerungssystem der Zentrale der Fertigungsstraße

kommuniziert und diese ihre Arbeitsweise entsprechend anpassen. Menschen sind in den vom KI-System ausgeführten Arbeitsbereich **nicht involviert**.

5.4.1.2 Produktsicherheitsrecht

Der Roboter fällt als Maschine unter den Anwendungsbereich der 9. ProdSV. Der Hersteller und sein Bevollmächtigter mussten also insbesondere nach § 3 Abs. 2 Nr. 1 der 9. ProdSV die **Risikobeurteilung** nach Anhang I der Maschinen-RL durchführen, bevor der Roboter in der Fertigungshalle in Betrieb genommen wird. Dafür werden sie mit den anderen Maschinen an der Fertigungsstraße und dem Steuerungssystem der Zentrale in Zusammenhang gesetzt. Durch den Zaun werden im Betrieb jegliche Risiken für Menschen ausgeschlossen, sodass insofern die Risikobeurteilung zu einem zweckmäßigen Ergebnis gekommen ist. Für die bestimmungsgemäße Verwendung wurde eine Schutzeinrichtung vorgesehen, um einen Kontakt der Maschine mit Menschen während des Betriebes zu verhindern.
Hier kann die **Vernetzung** unterschiedlicher Maschinen bei der **Generalklausel des § 3 Abs. 1 der 9. ProdSV** problematisch sein. Von der 9. ProdSV sind nach dessen § 3 Abs. 1 auch Güter geschützt, also auch die anderen Maschinen in der Fertigungsstraße. Es muss also im Rahmen der Prüfung der Anforderungen der Generalklausel ausgeschlossen werden, dass durch den Schweißroboter andere Teile der Fertigungsstraße beschädigt werden. Sofern der Schweißroboter jedoch wie hier nur einen Fehler am Werkstück zum Anlass nehmen kann, dieses aus der Produktion auszuscheiden, bestehen keine Gefährdungen für die geschützten Rechtsgüter. Auch hinsichtlich der Generalklausel bestehen also keine Bedenken.
Hinsichtlich der **Vernetzung** ist auch zu beachten, dass bei einer **Gesamtheit von Maschinen** deren Zusammenwirken bei den Sicherheits- und Gesundheitsschutzanforderungen nach Anhang I der Maschinen-RL, also auch in der Risikobeurteilung, zu berücksichtigen ist.
Sofern der Hersteller die formellen Anforderungen erfüllt, also ein vorgeschriebenes Konformitätsbewertungsverfahren durchgeführt hat, bestehen keine produktsicherheitsrechtlichen Probleme.

5.4.1.3 Betrieblicher Arbeitsschutz

Bei dem Schweißroboter handelt es sich um ein Arbeitsmittel im Sinne des § 2 Abs. 1 BetrSichV, sodass für dessen Verwendung die Voraussetzungen des § 4 BetrSichV erfüllt sein müssen. Sofern der Zaun nicht Bestandteil der Maschine war und mitgeliefert wurde, wird der Arbeitgeber regelmäßig zur Verhinderung von Verletzungen den Zaun installieren.

5.4.1.4 Ergebnis

Der Schweißroboter steht in der konkreten Ausgestaltung rechtlich vor keinen relevanten Problemen.
Es ist jedoch schon hier erkennbar, dass die zunehmende **Vernetzung** von Systemen im Hinblick auf die Generalklausel des § 3 Abs. 1 der 9. ProdSV den Hersteller vor Herausforderungen stellen kann. Das kann vor allem dann der Fall sein, wenn der Schweißroboter mit anderen KI-Systemen zusammenarbeitet und in Kontakt kommt. Dann ist durch den Hersteller zu ermitteln, welche Risiken im konkreten Fall bestehen, wo die Bagatellgrenze zu ziehen ist und wie darüber liegenden Risiken zu begegnen ist. Bei einer weiteren Vernetzung kommt es dann darauf an, wer Hersteller der Gesamtheit von Maschinen ist.

5.4.2 Beispiel 2 - KI-Steuerungssystem

Veränderbarkeit –Widerstandsfähigkeit – Vernetzung – Schadensfolgen
In einer hochautomatisierten Produktion des Maschinenbaus werden verschiedene Produkte an einer Fertigungsstraße produziert, die dem konkreten Produkt entsprechend modifizierbar ist. Einige Arbeiten werden von Menschen ausgeführt, andere vollautomatisiert durch Roboter. Die Steuerung der Geschwindigkeit der in den Boden der Fertigungsstraße eingelassenen Förderbänder, die Auswahl der Reihenfolge der in der Schicht zu bearbeitenden Werkstücke und die entsprechende Modifizierung der einzelnen Stationen der Fertigungsstraße funktioniert wie folgt: Es werden an vielen Stellen der Fertigungsstraße Kameras und andere Sensoren eingebaut. Die Auswertung der damit gesammelten Informationen erfolgt mit Deep-Learning-Algorithmen. Diese erkennen Arbeitsfortschritte an den einzelnen Stationen, eventuelle Staus, Unregelmäßigkeiten bei den Werkstücken oder Werkstücke mit besonderen Montageanforderungen, die an einzelnen Stationen mehr Zeit beanspruchen etc. Das Auswertungssystem lernt während des Betriebes weiter, es erkennt selbstständig die Anforderungen an die Fördergeschwindigkeit und an die Planung der Produktionsreihenfolge der verschiedenen Produkte bei der Vielzahl an möglichen Anforderungen entlang der Fertigungsstraße. Daraus abgeleitet werden die steuerbaren Prozessglieder, wie beispielsweise Motorgeschwindigkeiten des Förderbandes, durch das Auswertungssystem angepasst. Das Auswertungssystem ist dabei frei, mit seinen Steuerungsmöglichkeiten lösbare Probleme zu identifizieren und zu lösen, ihm wurde kein starres Ziel vorgegeben, wie z. B. „Ermögliche eine möglichst hohe Stückzahl pro Tag bei gleichbleibender Fördergeschwindigkeit“. Es kennt vielmehr die Leistungspotenziale der steuerbaren Elemente der Fertigungsstraße sowie die von ihm nicht beeinflussbaren Voraussetzungen, wie die vorgegebene Leistungsfähigkeit der Beschäftigten, die aktuellen Kapazitäten des Produktlagers und der aktuelle Stand des Lagers für Zuliefererteile und Grundstoffe. Daraus ermittelt es die möglichen Optimierungspotenziale. Die Beschäftigten an den manuellen Stationen erfahren über entsprechende Signale, wann welches Produkt durch die Fertigungsstraße läuft und stellen sich darauf ein. Im Lauf des Prozesses identifiziert das Auswertungssystem ein „Optimierungspotenzial“, wenn es ein etwas leichteres Werkstück besonders schnell über die Förderbänder laufen lässt. Nach einigen Durchgängen, die reibungslos verlaufen, kommt es an einer der manuell bedienten Stationen zum Unfall: Der Beschäftige dort war aufgrund einer Unachtsamkeit bei dem vorangegangenen Werkstück noch nicht auf das etwas schneller herangeförderte Werkstück eingestellt und deshalb noch nicht auf entsprechender Position, was die Software über die Kameras zwar registriert, aber falsch interpretiert. Der Prozess läuft daher weiter und das Werkstück fährt dem Beschäftigten von hinten in die Beine. Der Beschäftigte kommt zu Fall und verstaucht sich das Handgelenk. Das Auswertungssystem erkennt den Sturz und stoppt sofort den Produktionsprozess.

5.4.2.1 Taxonomie

Gegenstand der Untersuchung soll hier das Auswertungssystem sein. Es ist Teil eines Gesamtsystems, weist jedoch die hier untersuchten Merkmale der Taxonomie auf und gibt die maßgeblichen Befehle an die anderen Teilsysteme der Fertigungsstraße.
Dieses Auswertungssystem kann **während des Betriebes weiterlernen**, zeichnet sich also durch einen **hohen Grad an Veränderbarkeit** aus.
Die **Widerstandsfähigkeit** und hier speziell die **Robustheit** des Systems kann eingeschränkt sein. Denn die hohe Komplexität des Gesamtsystems, bedingt durch

die unterschiedlichen Informationen und möglichen Lösungswege für die durch das System identifizierten Probleme, bergen auch die Gefahr, dass das System durch unvorhergesehene Ereignisse den auszuwertenden Informationen falsche Werte zuordnet und so zu falschen Annahmen für die Steuerung der Fertigungsstraße kommt.
Das Auswertungssystem ist als Teil eines Gesamtsystems aus Kameras, Steuerungseinheit und Motoren für die Auswertung und Umsetzung der durch die Kameras aufgenommenen Bilder zuständig. Durch die Auswertung werden die Befehle der Steuerungseinheit und damit die Motorengeschwindigkeit determiniert. Das Auswertungssystem ist also **hochgradig vernetzt**. Es handelt sich zudem nicht um eine zentrale Instanz, die letztlich die Motoren steuert, sondern um eine Arbeitsteilung zwischen Bildverarbeitung und Sensorik. Die Vernetzung führt also zu einem **dezentralen Gesamtsystem**. Die Vernetzung ist nicht nur rein informativer Natur, sondern **sicherheitsrelevant**. Denn das Auswertungssystem trifft seine Entscheidungen auf Grundlage der verschiedenen Sensoren.
Der **Mensch ist nicht involviert**, sofern es um die Bewertung der von den Sensoren aufgenommenen Informationen und die Ableitung von Steuerungsbefehlen geht, kann aber geschädigt werden.
Das Auswertungssystem kann bei Fehlfunktionen relevante Schäden hervorrufen, da sie vitale Funktionen der Fertigungsanlage beeinflusst, die, fehlerhaft beeinflusst, ein Risiko darstellen. Insbesondere sind Menschen dem Gesamtsystem ausgesetzt, da sie passiv mit diesem interagieren.

5.4.2.2 Produktsicherheitsrecht

Das Auswertungssystem ist durch Einbettung in die Steuerungseinheit und die Verbindung mit den Sensoren und der Fertigungsanlage eine Komponente einer **Maschine** im Sinne des § 2 Nr. 2 lit. a) der 9. ProdSV. Auch wenn das Auswertungssystem als Software keine Maschine ist, stellt sie doch ein sicherheitsrelevantes Element des Gesamtsystems dar. Zur Erfüllung der Sicherheits- und Gesundheitsschutzanforderungen nach der Maschinen-RL muss der Hersteller dieser Maschine bei der Konstruktion sicherstellen, dass nur Komponenten eingesetzt werden, die diesen Anforderungen und der Funktionalität entsprechen. Unabhängig davon, ob sie als Produkt selbst Gegenstand einer Konformitätsbewertung waren oder lediglich Einzelteile sind, die in einer bestimmten Maschine aufgehen sollen, sind alle Komponenten des Endprodukts Maschine vom Hersteller so auszuwählen, dass die Maschine die an sie zu stellenden Anforderungen erfüllt.
Die Auswahl der Komponenten der Maschine ist also Teil der **Risikobeurteilung** nach Anhang I der Maschinen-RL. Das Auswertungssystem ist den möglichen **Schadensfolgen** entsprechend zu beurteilen.
Bei der **Bestimmung der Grenzen der Maschine** im Rahmen der Risikobeurteilung sind durch die Deep-Learning-Algorithmen weite Grenzen zu setzen. Sie können anhand der bestimmungsgemäßen Verwendung und vernünftigerweise erwartbaren Fehlanwendung identifiziert werden. Weite Grenzen bedeuten eine Vielzahl von möglichen Risiken, die wiederum schwer vor Inbetriebnahme bewertet werden können. Zudem handelt es sich um ein weiterlernendes System, sodass kaum ermittelt werden kann, mit welcher Wahrscheinlichkeit welches risikobehaftete Bildverarbeitungsergebnis aus der Menge möglicher Ergebnisse (innerhalb der Maschinengrenzen) vorkommt. Das gilt besonders dann, wenn das System **nicht sehr robust** ist. Dem kann **dank der Vernetzung** entgegengewirkt werden, indem **Redundanzen** vorgesehen werden. Das Auswertungssystem kann dann anhand von

Plausibilitätsprüfungen eine relativ zuverlässige Informationsgrundlage schaffen. Die **Kombination aus Veränderbarkeit und Vernetzung** kann hier dazu führen, dass die rechtlichen Anforderungen erfüllt werden und das System eine **Widerstandsfähigkeit** aufweist, die den bei der Risikobeurteilung ermittelten Anforderungen gerecht wird. Die Vernetzung gleicht hier einen hohen Grad der Veränderbarkeit aus.

Der **maßgebliche Zeitpunkt** der Risikobeurteilung durch den Hersteller kann sich jedoch als problematisch erweisen. Das weiterlernende System entzieht sich im maßgeblichen Zeitpunkt für die Risikobeurteilung, also spätestens bei der Inbetriebnahme, einer abschließenden Bewertung. Wie gezeigt, ist bereits die Eintrittswahrscheinlichkeit einzelner Gefährdungssituationen bei Betrieb des Systems im Auslieferungszustand schwierig. Lernt das System wie hier weiter, schließt es also auf Grundlage der gesammelten Informationen bestimmte Lösungen aus, rückt dafür andere in den Fokus, stellt sich die Frage, wie mit hinreichender Genauigkeit im maßgeblichen Zeitpunkt der Inbetriebnahme die eine Lösung in Zukunft vom System eher als nicht zu bevorzugen kategorisiert wird und die andere dafür verstärkt gewählt wird. Eine **Verifizierung** der für die Risikobeurteilung relevanten Eigenschaften wird damit erschwert. Der **Stand der Technik** ist der Maßstab für die Risikobeurteilung und die Konstruktion der Maschine. Kann den Sicherheits- und Gesundheitsschutzanforderungen nach Anhang I der Maschinen-RL nach dem Stand der Technik nicht nachgekommen werden, so ist die Maschine gemäß **Nr. 3 Allgemeine Grundsätze Anhang I der Maschinen-RL** so weit wie möglich auf diese Ziele hin zu konstruieren und zu bauen. Der Stand der Technik findet sich nicht nur in den technischen Normen. Besteht keine technische Norm für eine der Sicherheits- und Gesundheitsschutzanforderung, so ist der Stand der Technik aus anderen Quellen zu ermitteln. Das können auch Erfahrungen der Praxis mit bereits in Betrieb befindlichen Auswertungssystemen des in der Maschine installierten Typs. Insoweit kann den Hersteller eine **indirekte Produktbeobachtungsobliegenheit** treffen, seine in Betrieb befindlichen Maschinen zu einem bestimmten Grad zu beobachten. Wie weit diese ausgestaltet sein muss, hängt jedoch stark vom Einzelfall ab. Die daraus gezogenen Schlüsse können gleichwohl nicht ohne weiteres auf die konkrete Maschine in der konkreten Umgebung übertragen werden, wenn es sich um ein weiterlernendes System handelt.

Wegen der **möglichen Schadensfolgen** sind allerdings strenge Anforderungen an die Ermittlung und Bewertung der möglichen Risiken zu stellen. Die in dem Beispiel realisierten Schäden lassen das dort verwirklichte Risiko besonders schwer erscheinen, mag die Eintrittswahrscheinlichkeit (die wiederum zu ermitteln wäre) auch gering sein.

Der Hersteller der Maschine muss ebenso den Grad der **Vernetzung** der KI-Komponenten mit den anderen Teilen der Maschine beachten. Da das Auswertungssystem als Software Teil der Gesamtheit ist, die als Maschine den Gegenstand der Risikobeurteilung bildet, stellt dies den Hersteller rechtlich vor keine Probleme: Wer die einzelnen Komponenten Sensorik, Sensordatenverarbeitungsprogramm, Steuerungseinheit und Antriebsmotoren zusammenfügt, ist Hersteller im Sinne des § 2 Nr. 10 der 9. ProdSV.

Wenn der Hersteller die Herausforderungen der Risikobeurteilung überwinden und die Konformität der Maschine im Zeitpunkt der Inbetriebnahme nachweisen und die formellen Anforderungen der 9. ProdSV erfüllen kann, stellt sich die Frage, ob eine **Konformität später noch gegeben ist**, wenn das System sich in sicherheitsrelevanter Weise verändert. Mag es für die Inbetriebnahme genügen, für

diesen maßgeblichen Zeitpunkt die Konformität nachzuweisen, kann die Bewertung zu einem späteren Zeitpunkt durch die **Marktüberwachungsbehörde** anders ausfallen, da das System dann faktisch bei Untersuchungen der Marktüberwachungsbehörde als **nicht mehr konform** erscheinen kann. Denn die Marktüberwachungsbehörde muss mit allen ihr nach §§ 26 Abs. 1, 28 ProdSG zustehenden Mitteln den Sachverhalt ermitteln, um feststellen zu können, ob den Anforderungen nach Abschnitt 2 des ProdSG, also über § 3 Abs. 1 Nr. 1 ProdSG auch den Anforderungen der 9. ProdSV **im maßgeblichen Zeitpunkt** entsprochen wurde. Die Marktüberwachungsbehörde muss den Sachverhalt wegen des Amtsermittlungsgrundsatzes des § 24 Abs. 1 S. 1 und 2 VwVfG selbst ermitteln, sie führt also ggf. auch eine eigene Risikobeurteilung durch. Da ihr kein Entschließungsermessen hinsichtlich der Durchführung von Ermittlungen zusteht, muss sie bereits bei einem Anfangsverdacht entsprechend ermitteln. Bei **entsprechend hohen potenziellen Schäden** kann sich die Pflicht zur Ermittlung derart verdichten, dass bei Unterlassen der erforderlichen und angemessenen Ermittlungsmaßnahmen ggf. auch Amtshaftung drohen kann, wenn durch das rechtswidrige Unterlassen der Ermittlungen Schäden verursacht werden.
Die Marktüberwachungsbehörde hat zur adäquaten Wahrnehmung ihrer Aufgaben einen **hohen Informationsbedarf**. Sie wird daher gemäß § 26 Abs. 1 S. 2 ProdSG auf die technischen Unterlagen zurückgreifen und ggf. Laborprüfungen durchführen. Der Hersteller wird daher insbesondere bei weiterlernenden komplexen und vernetzten Systemen wie dem hier vorliegenden über die formellen Dokumentationspflichten des § 3 Abs. 2 der 9. ProdSV hinaus ein Interesse haben, das System erklärbar und damit **transparent** zu halten. Denn nur so kann er verhindern, dass die Marktüberwachungsbehörde durch die nach Inbetriebnahme eingetretenen sicherheitsrelevanten Veränderungen im System zu dem Schluss kommt, dass ein Verstoß gegen § 3 Abs. 1 und 2 Nr. 1 der 9. ProdSV vorliegt.

5.4.2.3 Betrieblicher Arbeitsschutz

Der Arbeitgeber muss auch hier, wie in Beispiel 1, eine **Gefährdungsbeurteilung** durchführen. Ist er durch die Konstruktion der Gesamtheit als Maschine zur geschäftsmäßigen Nutzung in seinem eigenen Betrieb Hersteller im Sinne der 9. ProdSV, liegt die erste Gefährdungsbeurteilung als Voraussetzung für die Verwendung des Arbeitsmittels nach § 4 Abs. 1 Nr. 1 BetrSichV vor.
Die Gewährleistung der Sicherheit am Arbeitsplatz ist jedoch eine dauernde Pflicht des Arbeitgebers, sodass er auch zu einer **Überprüfung** und ggf. **Aktualisierung der Gefährdungsbeurteilung** nach § 3 Abs. 7 BetrSichV verpflichtet ist. Er hat dafür auch zu ermitteln, in welchem Turnus die Überprüfung zu erfolgen hat. Bei dem hier vorliegenden **veränderbaren** System wird eine solche Überprüfung entsprechend häufig erfolgen müssen. Gleiches gilt für die **wiederkehrenden Prüfungen**, die nach der Gefährdungsbeurteilung erforderlich werden.
Dies gilt hier umso mehr, da das Auswertungssystem **relevante Schäden** hervorrufen kann.
Vorausgesetzt, die Gefährdungsbeurteilung überwindet die für die Risikobeurteilung des Herstellers festgestellten Schwierigkeiten, stellt sich sogar vielmehr die Frage, ob das System nicht **dauerhaft überwacht werden muss.** Denn die noch in der letzten Gefährdungsbeurteilung gewonnenen Erkenntnisse und insbesondere die getroffenen Schutzmaßnahmen können durch die Veränderbarkeit quasi jederzeit hinfällig sein. Jedenfalls die Pflicht zur wiederkehrenden Prüfung wird damit zur Pflicht zur dauerhaften Prüfung.

Damit geht ein **hoher Bedarf an Informationen** über die Funktionsweise des Systems einher. Der Arbeitgeber muss sich gemäß § 3 Abs. 4 S. 1 BetrSichV die erforderlichen Informationen beschaffen. Wenn nötig, muss er sich nach § 3 Abs. 3 S. 4 BetrSichV fachkundig beraten lassen. Mit hochgradig veränderbaren und wenig widerstandsfähigen Systemen kann er keine zweckmäßige Gefährdungsbeurteilung durchführen, wenn er nicht die erforderlichen Informationen über die Logik des Systems hat. Die **Transparenz** des Systems ist für den Arbeitgeber von hoher Bedeutung. Das ist nicht zuletzt auch Voraussetzung für die Unterweisung der Beschäftigten gemäß § 12 BetrSichV. Damit ein sicherer Betrieb gewährleistet ist, muss entsprechend transparent sein, wann welches Teil über die Fertigungsstraße läuft. Das ist Gegenstand der Unterweisung, aber auch der zu auf Grundlage der Gefährdungsbeurteilung zu ergreifenden Maßnahmen zur Gewährleistung der Sicherheit.
Für die **Aufsichtsbehörden** gilt das zu den Marktüberwachungsbehörden Ausgeführte. Sie können durch den kooperativen Ansatz des Arbeitsschutzrechts jedoch flexibler auf den Verdacht der Gefährdung von Beschäftigten durch Arbeitsmittel reagieren. Letztlich stellen sie aber den Arbeitgeber als aus dem ArbSchG und der BetrSichV Verpflichteten vor die gleichen Herausforderungen, wie den Hersteller: Er muss das System de facto im Zweifel erklären können, um umfangreicheren Maßnahmen der Aufsichtsbehörden zu begegnen.

5.4.2.4 Ergebnis

Dieses Szenario wirft für den Hersteller und den Arbeitgeber rechtliche Fragen auf, da beide für eine Erfüllung ihrer jeweiligen Pflichten hinsichtlich des Risikomanagements letztlich zur dauerhaften Überwachung der KI-Komponente angehalten sind.
Der Hersteller muss sich formell am maßgeblichen Zeitpunkt der Inbetriebnahme orientieren, wird aber tatsächlich eine belastbare Aussage über die möglichen Zustände der Maschine und der damit einhergehenden Gefährdungssituationen nur bei einer dauernden Beobachtung im laufenden Betrieb am konkreten Einsatzort treffen können. Die Anforderungen des § 3 Abs. 1 der 9. ProdSV versuchen dies für individuell konstruierte Maschinen insofern zu entschärfen, indem sie Konformität erst mit Inbetriebnahme verlangen, also nach der Montage und etwaigen Testläufen. Diese Regel geht jedoch von Systemen aus, die sich ab Inbetriebnahme nicht mehr ändern. Der hieran orientierte maßgebliche Zeitpunkt ist damit für das vorliegende Szenario hinderlich und führt dazu, dass der Hersteller die erforderlichen Nachweise nicht erbringen kann. Auch wenn er sie im maßgeblichen Zeitpunkt erbringen kann, kann die Marktüberwachungsbehörde zu einem späteren Zeitpunkt zu einem anderen Ergebnis kommen. Der Hersteller wird also auch über den Zeitpunkt der Inbetriebnahme hinaus sein System so transparent gestalten müssen, dass keine hoheitlichen Maßnahmen wegen der Annahme erfolgen, dass System sei nicht konform gewesen.
Eine direkte Pflicht zur Beobachtung ergibt sich aus dem Recht derzeit für den Arbeitgeber. Sie stellt ihn bei einem veränderbaren System wie dem hier vorliegenden vor erhebliche Herausforderungen. Weiter muss er zur Vermeidung der im Beispiel realisierten Gefährdungen letztlich bei der hier eingesetzten KI-Komponente einen technikbezogenen Ansatz wählen und den „Entscheidungsspielraum“ der KI-Komponente auf ein beherrschbares Niveau begrenzen oder eine menschliche Kontrollinstanz integrieren.

5.4.3 Beispiel 3 – Kooperierender (kollaborierender[359]) Roboter

Transparenz – Schadensfolgen
Die Geschäftsleitung eines Elektronikherstellers entscheidet sich, zur Optimierung der Arbeitsabläufe einen kooperierenden (kollaborierenden) Roboter einzusetzen. Er packt selbstständig die Artikel ein, seine menschlichen Kollegen stehen nur noch in unmittelbarer Nähe daneben und kontrollieren die Arbeit des Roboters. Sie können den Arbeitsprozess unterbrechen, den Roboter wenn nötig manuell einstellen und dann wieder in Gang setzen. Der Roboter ist mit einer Software ausgestattet, die es ihm erlaubt, mithilfe seiner Sensorik unterschiedliche Artikelchargen zu erkennen und seine Bewegungen entsprechend adaptiv anzupassen. Er bewegt sich zudem langsam, um eine Verletzung der Beschäftigten zu vermeiden. Der Beschäftigte S hat sich an den Roboter gewöhnt und überwacht sechs Monate lang ohne Zwischenfälle dessen Arbeit. Dann kommt eine andere Charge in die Packabteilung. Der S weiß nichts von der Adaptivität des Roboters und bewegt sich daher noch so in dessen Umfeld, als verarbeitete er noch die alte Charge. Daher stößt er mit dem Arm des Roboters zusammen und verletzt sich.

5.4.3.1 Taxonomie

Der hier eingesetzte Roboter verfügt über eine adaptive Steuerungssoftware, die die Bewegungen an die Artikel anpasst. Der Roboter ist in seiner **Veränderbarkeit** mit dem aus Beispiel 1 vergleichbar. Die höchste Stufe der Veränderbarkeit, bei der ein Weiterlernen im Betrieb erfolgt, ist hier nicht erreicht.
Die **Transparenz** des Systems ist nicht sehr ausgeprägt. Es warnt die Beschäftigten nicht vor Veränderungen der Bewegungen.
In das System sind **Menschen involviert**. Es handelt sich um einen kooperierenden Roboter. Die Kooperation beschränkt sich auf eine Kontrolle durch den Menschen. Trotzdem sind physische Kontakte möglich.
Aus dieser systemimmanenten[360] Involviertheit folgt hier eine hohes **Schädigungspotenzial**. Durch physischen Kontakt mit Menschen besteht das Risiko einer Körperverletzung.

5.4.3.2 Produktsicherheitsrecht

Im Wesentlichen kann auf die Ausführungen zu Beispiel 1 verwiesen werden: Der Hersteller muss gemäß der 9. ProdSV die Sicherheit bereits „mitliefern“.
Da hier **Menschen in das System involviert** sind, ist dieser Umstand bei der **Risikobeurteilung** besonders zu berücksichtigen. Aus den ermittelten Risiken, wie hier das Risiko, bei bestimmungsgemäßer Verwendung mit dem Roboter zusammenzustoßen, müssen entsprechende Konsequenzen für die Konstruktion der Maschine gezogen werden. Nach Nr. 1.1.2 lit. b) Anhang I der Maschinen-RL muss die Maschine zunächst so konstruiert werden, dass es keine Verletzungen durch Zusammenstöße geben kann. Auch unterhalb der Schwelle der Verletzungsgefahr

---

[359] Gemäß Onnasch et al., 2016 handelt es sich im folgenden Fall um einen kooperierenden Roboter, da keine direkte mechanische Interaktion zwischen Mensch und Roboter stattfindet. In vielen Veröffentlichungen wird dieser Differenzierung nicht vorgenommen und immer von kollaborierenden Robotern gesprochen.

[360] Im Gegensatz dazu war der verunglückte Beschäftigte in der Abwandlung zum Beispiel 1 nicht in den Teil des Systems involviert, in dem die für die Taxonomie relevanten Kriterien der Veränderbarkeit und der Vernetzung zum Ausdruck kommen.

muss der Roboter gemäß Nr. 1.1.6 Anhang I der Maschinen-RL unter Berücksichtigung ergonomischer Prinzipien gestaltet werden. Ist dies nicht möglich, müssen Schutzmaßnahmen ergriffen werden, wie physische Abschirmungen. Zuletzt ist auf die Restrisiken hinzuweisen. Bleibt in letzter Konsequenz nur ein Hinweis auf das Restrisiko des Zusammenpralls, so ist der Roboter entsprechend **transparent** zu gestalten. Transparenz kann durch Warnhinweise, Personalschulungen und Warnsignale (z. B. Tonsignal bei Annährung oder durch den Roboter identifizierter wiederholter Bewegungen in seinem möglichen Arbeitsradius) geschaffen werden.
Eine entsprechende Voraussetzung lässt sich auch aus der **Generalklausel des § 3 Abs. 1 der 9. ProdSV** ableiten. Der Roboter muss in seinen Bewegungen und seiner Adaptivität so transparent sein, wie dies bei der vorausgesetzten oder erwartbaren Verwendergruppe erforderlich ist, um keine Gefährdung für die Sicherheit und die Gesundheit von Personen darzustellen.

5.4.3.3 Betrieblicher Arbeitsschutz

Auch hier kann auf die Ausführungen zu Beispiel 1 verwiesen werden. Zu beachten ist aber auch hier, dass die **Involviertheit des Menschen** in das System eine **umfangreiche Gefährdungsbeurteilung** nötig macht. Der Arbeitgeber kann grundsätzlich nach § 3 Abs. 4 BetrSichV auf die mitgelieferten Informationen vertrauen. Weist der Hersteller auf das bestehende Restrisiko hin, muss der Arbeitgeber dies auch in seine Gefährdungsbeurteilung und die Bestimmung der erforderlichen Schutzmaßnahmen einbeziehen. Er hat also die Beschäftigten entsprechend zu schulen.
Ist das Risiko nicht vom Hersteller benannt worden, weil es erst durch die besonderen Bedingungen im Betrieb des Arbeitgebers entsteht, so muss der Arbeitgeber hier entsprechend tätig werden. Hier gibt der § 3 Abs. 2 S. 2 Nr. 1 und 2 BetrSichV dem Arbeitgeber insbesondere auf, die ergonomische Gestaltung und Zusammenhänge zu beachten.

5.4.3.4 Ergebnis

Die Involviertheit des Menschen und die damit verbundene Sicherheitsrelevanz führt nach dem geltenden Recht dazu, dass der Roboter mit KI-Komponente entsprechend transparent ausgestaltet sein muss. Diese Pflicht trifft in erster Linie den Hersteller. Der Arbeitgeber muss die Risiken beseitigen oder minimieren, die den besonderen Umständen in seinem Betrieb geschuldet sind.
Insofern kann das bestehende Recht kollaborierende KI-Systeme in der hier dargestellten Ausprägung erfassen und die Pflichten des Risikomanagements zweckmäßig auf die Beteiligten aufteilen.

5.4.4 Abwandlung Beispiel 3 – Adaptiv-kollaborierender Roboter

Schadensfolgen (im Sinne der DSGVO)
Der Roboter aus Beispiel 3 packt die Artikel in Zusammenarbeit mit den Beschäftigten. Sie befüllen die vom Roboter vorgefaltete Verpackung je nach Kundenwunsch und reichen dem Roboter die Packung, der sie verschließt und verklebt. Der Roboter passt sich den Bewegungen der kollaborierenden Beschäftigten an. Da einige Beschäftigte in Teilzeit arbeiten, teilen sich mehrere einen Roboter. Der Roboter speichert also von verschiedenen Beschäftigten jeweils ihre Vorlieben. Die Beschäftigten wissen davon, da sie in ihrem Arbeitsvertrag und einer entsprechenden Schulung darauf hingewiesen

wurden. Der S, diesmal besser vorbereitet, loggt sich an seinem Arbeitsplatz ein und der Roboter bewegt sich seinem Arbeitsrhythmus entsprechend.

5.4.4.1 Taxonomie

Im Gegensatz zu Beispiel 3 sind hier die möglichen **Schadensfolgen** weiter ausgeprägt, da der Roboter zur Anpassung seiner Bewegungen Informationen über den jeweils kollaborierenden Beschäftigten sammelt, auswertet und speichert. Jedenfalls bei einem **weiten Sicherheitsbegriff**, unter den auch andere geschützte Rechtsgüter fallen, wie das Recht auf informationelle Selbstbestimmung, sind hier personenbezogene Daten in relevanter Weise von dem System betroffen.

5.4.4.2 DSGVO

Der Schutz personenbezogener Daten ist nicht Zweck des ProdSG und der BetrSichV. Die Informationen über die Bewegungsabläufe der Beschäftigten sind als personenbezogene Daten im Sinne der Art. 2 Abs. 1 und Art. 4 Nr. 1 DSGVO von dieser erfasst. Der Roboter erfasst diese Daten, wertet sie aus und speichert sie, sodass eine automatisierte Datenverarbeitung im Sinne der Art. 2 Abs. 1 und Art. 4 Nr. 2 DSGVO vorliegt.
Daraus folgt, dass die **Verantwortlichen** verschiedene **Pflichten zur technischen Ausgestaltung** des Roboters treffen. Verantwortlich ist nach Art. 4 Nr. 7 DSGVO derjenige, der über die Zwecke und Mittel der Verarbeitung entscheidet. Das ist im vorliegenden Beispiel zunächst der Arbeitgeber, der den Roboter in seinem Betrieb einsetzt. Es kann auch mehrere Verantwortliche geben, die nebeneinander über unterschiedliche Aspekte der Zwecke und Mittel der Verarbeitung entscheiden. So kann der Hersteller neben den Arbeitgeber treten, wenn er z. B. über vertragliche Pflichten zur Wartung des Roboters oder der Software Zugang zu den Daten hat und diese z. B. zwecks weiteren Teachings organisiert und damit im Sinne des Art. 4 Nr. 2 DSGVO verarbeitet.
Der Art. 25 Abs. 1 DSGVO verlangt **privacy by design**, der Roboter muss also von vornherein so konstruiert sein, dass er die Grundsätze des Art. 5 DSGVO wahrt. Dazu gehört auch, dass die Verarbeitung transparent sein muss. Der Verantwortliche muss also gegenüber den betroffenen Personen, hier den Beschäftigten, darüber Auskunft geben können, welche Daten über sie wie verarbeitet werden, damit sie im Sinne der informationellen Selbstbestimmung Kontrolle über ihre Daten haben und ggf. Rechte gegenüber dem Verantwortlichen geltend machen können. Das verlangt eine diesbezügliche **Transparenz** des KI-Systems. Der Verantwortliche muss selbst ermitteln, welches Maß an Transparenz erforderlich ist und wie er das technisch und organisatorisch erreichen kann. Das kann den Arbeitgeber, der nicht Hersteller der KI-Komponente ist, vor große Herausforderungen stellen.
So können die Pflichten des verantwortlichen Arbeitgebers aus der DSGVO auch auf den Hersteller durchschlagen, der der Nachfrage nach DSGVO-konformen Maschinen entsprechend nur solche anbietet.

5.4.4.3 Ergebnis

Die **Transparenz** der KI-Komponente wird dann rechtlich besonders relevant, wenn es sich um personenbezogene Daten handelt, mit denen sie arbeitet. Die DSGVO stellt hier technische und organisatorische Anforderungen, die bei dem Einsatz von Systemen mit KI-Komponente jedenfalls durch den für die Datenverarbeitung Verantwortlichen beachtet werden müssen, um den Schutzzweck der DSGVO zu

erfüllen Diese Anforderungen laufen parallel zu den anderen Sicherheitsanforderungen aus dem Produktsicherheitsrecht und dem betrieblichen Arbeitsschutz.

5.4.5 Beispiel 4 – Transportroboter

Veränderbarkeit – Widerstandsfähigkeit – Kontrollierbarkeit – Involviertheit des Menschen

Eine selbstfahrende mobile Plattform in einem Hotel transportiert als Serviceroboter die Koffer der älteren und der geh- oder trageeingeschränkten Gäste von der Rezeption zum Zimmer der Gäste und entlastet dadurch das Hotelpersonal. Dieses muss dem Roboter nur die Zimmernummer sprachlich mitteilen, den Koffer auf den Transportroboter stellen und der Roboter fährt mit dem Koffer selbstständig auf Basis einer internen selbstgelernten Karte des Gebäudekomplexes bis vor das Zimmer. Die Gäste folgen ihm und müssen am Ende der Reise den Koffer nur ins Zimmer tragen. Der Roboter erfasst während der Fahrt das Gesicht des Gastes und lernt ihn wiederzuerkennen. Daher weiß der Roboter, ob „sein" Gast im noch folgt. Während der Fahrt begegnet der Serviceroboter anderen Personen, die ihm auf den Gängen des Hotels entgegenkommen. Mit Hilfe einer Kamera und Algorithmen der KI erkennt der Roboter, ob die Objekte in Fahrtrichtung Menschen oder andere Objekte (z. B. entgegenkommende Serviceroboter) sind und verändert selbstständig seine Geschwindigkeit und sein Verhalten. Entgegenkommende Personen muss der Roboter nicht individuell identifizieren. Er entscheidet selbstständig, ob er links oder rechts an der ihm entgegenkommenden Person vorbeifährt.

5.4.5.1 Taxonomie

Der Roboter verfügt über eine **Veränderbarkeit** im Betrieb, da er Personengesichter lernt und sich an die jeweilige Geometrie des Gebäudes anpassen kann. Er lernt die Umgebung selbst kennen, er ist mit Inbetriebnahme nicht „ausgelernt", sondern **lernt weiter**. Er sammelt die erforderlichen Daten **während des Betriebs, also online**. Die Ziele sind verhältnismäßig explizit geregelt („Finde das Zimmer" und „Verursache keinen Unfall durch Zusammenstoßen").
Die Bildererkennung in einer hochkomplexen Umwelt kann zudem zu einer **geringen Widerstandsfähigkeit** des KI-Systems führen. Das weiterlernende System kann bei der Vielzahl an Daten, die es durch die Sensorik sammelt und dann verarbeitet, zu unkalkulierbaren Lösungen kommen, die mit dem Ziel nicht vereinbar sind.
Es sind im gesamten System **Menschen involviert**, sowohl aktiv als Befehlsinstanz als auch passiv als Gäste oder Personen, denen der Roboter auf dem Weg begegnet.

5.4.5.2 Produktsicherheitsrecht

Der Hersteller des Roboters, bei dem es sicher wieder um eine Maschine im Sinne des § 1 Abs. 1 Nr. 1 der 9. ProdSV handelt, hat in seiner **Risikobeurteilung** zunächst die möglichen **Schadensfolgen** bei Einsatz des Roboters in Rechnung zu stellen. Durch die **Involviertheit von Menschen** an quasi jeder Stelle des Systems sind Gefährdungen zu berücksichtigen, die mitunter ein hohes Risiko begründen können. Gleichzeitig ist die **hohe Veränderbarkeit** des Systems im Betrieb ein Faktor, der die Bestimmung der Grenzen der Maschine, jedenfalls aber der Ermittlung der Gefährdungen und die Bewertung der Risiken, erschwert. Denn wie schon in Beispiel 2 ist nicht abzusehen, welche Lösungen zur Erreichung der Ziele das KI-System wählt.

Die **mangelnde Widerstandsfähigkeit** aufgrund der komplexen Umwelt führt zudem zu der Erkenntnis, dass die ermittelten Gefährdungssituationen als wahrscheinlich zu gelten haben, also ein hohes Risiko verbleibt. Dieses ist dann im Wege der Integration der Sicherheit entsprechend Nr. 1.1.2 Anhang I der Maschinen-RL zu minimieren und dabei aber die Funktion der Maschine noch zu erhalten. Dem kann außer durch konservative technische Sicherheitsmaßnahmen[361] auch durch **Redundanzen**[362] vorgebeugt werden, auch die Vernetzung mit außerhalb des Transportroboters liegenden Sensoren kommt in Betracht. Das Recht steht einer solchen Vernetzung nicht entgegen, wie Beispiel 1 gezeigt hat, erfordert jedoch eine umfangreichere Risikobeurteilung und als Vorfrage eine klare Zuordnung der Herstellereigenschaft des Gesamtsystems.
Zu diesen hohen Anforderungen hinzu kommt die Problematik, die sich durch die Zäsur der **Inbetriebnahme als maßgeblichen Zeitpunkt** für das Vorliegen der materiellen und formellen Voraussetzungen nach der 9. ProdSV ergibt. Denn de jure ist der Hersteller damit aus seinen Pflichten entlassen. De facto musste er aber bereits in der Risikobeurteilung über diesen Zeitpunkt hinweg den Roboter *wie er sein wird* als Gegenstand seiner Risikobeurteilung heranziehen.

5.4.5.3 Betrieblicher Arbeitsschutz

Der Arbeitgeber steht, wie auch in Beispiel 2, vor der Herausforderung, dass die **Pflichten zur Gefährdungsbeurteilung**, zu **deren Aktualisierung** und zur **wiederkehrenden Prüfung** des Roboters sich derart verdichten müssen, dass von einer dauerhaften Beobachtung ausgegangen werden müsste. Das wäre auch hier mit einem nicht unerheblichen Bedarf an **Informationen über das System** verbunden, sodass sich die Frage stellt, wie **transparent** dieses ausgestaltet sein müsste, damit der Arbeitgeber es hinreichend beobachten kann.
Er hat zudem nach § 4 Abs. 1 Nr. 2 BetrSichV durch geeignete und dem Stand der Technik entsprechende Schutzmaßnahmen dafür zu sorgen, dass die bestehenden Risiken minimiert werden. Bei einer Vielzahl an in das System **involvierten Menschen** in Kombination mit der **Veränderbarkeit** des Roboters bei der Wahl seine Route sind hier hohe Anforderungen an die geeigneten Schutzmaßnahmen zu stellen. Dies folgt auch aus der deshalb sehr ausgeprägten Dimension der **Schadensfolgen**.

5.4.5.4 Ergebnis

Hier zeigt sich eine juristisch anspruchsvolle Kombination an Dimensionsausprägungen der Taxonomie. Die hohe Veränderbarkeit in einem komplexen Szenariofeld, die dadurch möglicherweise bedingte mangelnde Widerstandsfähigkeit, die Involviertheit von Menschen und letztlich die hohe Sicherheitsrelevanz führen dazu, dass Hersteller und Arbeitgeber zur Erfüllung ihrer Pflichten zur Gewährleistung des technischen Arbeitsschutzes besonders gefordert sind.
Während der Hersteller mit der Inbetriebnahme zwar aus der produktsicherheitsrechtlichen Pflicht entlassen ist, spielt die Unvorhersehbarkeit des späteren Lebenszyklus‘ eine kritische Rolle bei der Erfüllung seiner Pflicht zur Risikobeurteilung. Der in diesem Teil des Lebenszyklus‘ nähere Akteur, der

---

[361] Siehe Unterkategorie **Beschränkungen** in der Dimension **Kontrollierbarkeit.**
[362] **Aktive Wirkungsminderung nach Fehler** in der Unterkategorie **Resilienz** der Dimension **Widerstandsfähigkeit.**

Arbeitgeber, sieht sich hohen Anforderungen an die Gefährdungsbeurteilung ausgesetzt.

5.4.6 Abwandlung Beispiel 4 – Transportroboter

Widerstandsfähigkeit – sicherheitsrelevante und unabgesprochene Vernetzung – Involviertheit des Menschen

Der Transportroboter in Beispiel 4 ist nicht der einzige Roboter, der im Hotel eingesetzt wird. Der Hotelbetreiber setzt zur Unterstützung des Personals einen Roomservice-Roboter ein, der die vom Reinigungspersonal vor die Zimmertüren gelegte schmutzige Wäsche einsammelt und in die Wäscherei fährt. Die Reinigung der Flure des Hotels übernimmt eine Drittfirma, die ihrerseits Putzroboter einsetzt. Die Putzroboter werden von der Drittfirma in das Hotel verbracht und dort „angelernt", sodass sie danach den Dienst selbstständige ausführen können. Sie fahren die Flure bei Bedarf ab, saugen dabei Staub und können Flecken mithilfe eines Reinigungsschaums beseitigen. Alle drei Roboter sind von unterschiedlichen Herstellern. Die in den Robotern eingesetzte Steuerungstechnik entspricht grundsätzlich der des im Ausgangsbeispiel beschriebenen Transportroboters: Sie orientieren sich anhand einer selbsterlernten Karte und erkennen Menschen.

Die drei Robotertypen können sich untereinander über WLAN vernetzen. Sie tun dies mit Robotern in ihrer Nähe, die ihnen z. B. entgegenkommen. Dabei kommunizieren sie ihren Betriebsstatus, ihre aktuelle Geschwindigkeit, ihren aktuellen Bremsweg, ihr Ziel etc. Auf Grundlage dieser Informationen kann der empfangende Roboter seinen Fahrtweg anpassen. Die Anpassung kommuniziert er wiederum an den entgegenkommenden Roboter und über akustische oder optische Signale an Menschen in der Umgebung (z. B. „Blinker").

Aufgrund einer bekannten Schwachstelle in der Software der Reinigungsroboter der Drittfirma, die noch nicht durch ein zur Verfügung stehendes Update beseitigt wurde, kann einer der Hotelgäste mit seinem Laptop die Steuerung eines der gerade aktiven Reinigungsroboter übernehmen. Er veranlasst absichtlich einen Zusammenstoß mit einem der Transportroboter, der dabei das von ihm transportierte Gepäck eines anderen Gastes verliert und das dabei Schaden nimmt. Der Reinigungsroboter sandte aufgrund der Manipulation falsche Informationen an den Transportroboter, sodass ein Ausweichen für den Transportroboter nicht möglich war.

5.4.6.1 Taxonomie

Die hier untersuchten Roboter sind zusätzlich zu den im Beispiel 4 gegebenen Taxonomiedimensionen extern und handlungsbeeinflussend miteinander vernetzt. Der Reinigungsroboter der Drittfirma weist zudem eine geringere Widerstandsfähigkeit auf, da seine Software nicht in gleicher Weise gegen unerlaubte Eingriffe von außen geschützt ist, wie die der anderen Robotertypen. Es handelt sich damit auch um ein Beispiel für die Involviertheit des Menschen als sicherheitsmindernder Faktor. Diese Taxonomiedimension zeigt sich hier als eng verknüpft mit der Widerstandsfähigkeit.

5.4.6.2 Produktsicherheitsrecht

Der Hersteller bringt eine handlungsbeeinflussend und offen vernetzte Maschine in Verkehr. Diese Art der Vernetzung entspricht der bestimmungsgemäßen Verwendung der Maschine. Eine Gesamtheit von Maschinen im Sinne des § 2 Nr. 2 der 9. ProdSV

entsteht durch die spontane Vernetzung nicht, da kein produktionstechnischer Zusammenhang zwischen den einzelnen Robotern entsteht. Trotzdem muss der Hersteller des jeweiligen Roboters die Relevanz der von den anderen Robotern kommunizierten Daten in seiner Risikobeurteilung berücksichtigen. Er muss dafür sorgen, dass durch sie kein unsicherer Zustand seines Roboters hervorgerufen wird. Hierzu gibt es jedoch keine gesetzlichen Vorgaben. Der Hersteller des hier verunglückten Transportroboters muss also bestimmen, wie gewährleistet werden kann, dass nur Informationen einer bestimmten Qualität und z. B. keine verfälschten Informationen seinen Roboter beeinflussen. Der Hersteller des korrumpierten Reinigungsroboters ist indes produktsicherheitsrechtlich nicht verpflichtet, das System selbstständig zu aktualisieren oder nach unentdeckten Fehlern in der Software zu suchen (sofern nicht die von ihm angewendeten technischen Normen dies vorsehen).

5.4.6.3 Anforderungen an die IT-Sicherheit

Es bestehen für die Verwender des Transportroboters und des Reinigungsroboters keine ordnungsrechtlichen Pflichten, die die IT-Sicherheit unmittelbar adressieren. Der Hotelbetreiber als Arbeitgeber wird im Rahmen der Gefährdungsbeurteilung und der zur Beseitigung etwaiger Risiken erforderlichen Maßnahmen auch die IT-Sicherheit, also die Widerstandsfähigkeit der bei ihm eingesetzten Roboter beachten müssen. Sofern der Reinigungsroboter auch als Arbeitsmittel des Hotelbetreibers betrachtet werden kann, erstreckt sich dessen Pflicht auch auf den Reinigungsroboter. Die Drittfirma wiederum wird dem Hotelbetreiber gegenüber vertraglich verpflichtet sein, das erforderliche Maß an Sicherheit bzw. Widerstandsfähigkeit seiner Reinigungsroboter zu garantieren. Eine darüber hinaus gehende Pflicht zur Gewährleistung eines bestimmten Grades an Widerstandsfähigkeit besteht jedoch für die Drittfirma nicht.

5.4.6.4 Ergebnis

Die Widerstandsfähigkeit jeder der hier beteiligten Maschinen im Sinne der Missbrauchsrobustheit ist für die anderen Maschinen sicherheitsrelevant. Die mangelnde Robustheit des Reinigungsroboters konnte nicht nur zur Steuerung des Roboters genutzt werden, sondern auch dazu, die mit dem Transportroboter geteilten Informationen zu manipulieren, sodass der Zusammenstoß nicht mehr vermeidbar war. Diese sicherheitsrelevante Vernetzung wird jedoch weder im Produktsicherheitsrecht noch im Recht des betrieblichen Arbeitsschutzes adressiert.

5.5 Überblick zur rechtlichen Bewertung der Taxonomie

Die **einzelnen Dimensionen** und die unterschiedlichen **Kombinationen** werden von dem hier untersuchten Recht unterschiedlich bewertet. Die folgende Tabelle gibt einen Überblick darüber, welche Aspekte in den einzelnen Rechtsgebieten als problematisch bezeichnet werden können bzw. inwiefern sie dort eine Rolle spielen.

Tab. 5.1 Relevanz der Dimensionen der Taxonomie auf verschiedene Rechtsgebiete

	ProdSG / 9. ProdSV	BetrSichV	BImSchG	DSGVO	Haftung
Veränderbarkeit	Risikobeurteilung im maßgeblichen Zeitpunkt nicht abschließend möglich	• Gefährdungsbeurteilung nicht abschließend möglich, daher häufige Aktualisierung + Prüfung nötig • Ggf. Herstellerpflichten bei neuem Produkt	Änderungsgenehmigung bei hoch veränderbaren Anlagen häufig nötig	Datenschutz-Folgenabschätzung nicht abschließend möglich	Mangels produktsicherheitsrechtlicher Vorgaben hohe Relevanz der Produzentenhaftung (Produktbeobachtungspflicht)
Transparenz	Marktüberwachung erschwert durch Intransparenz	• Gefährdungsbeurteilung bei Intransparenz erschwert, wenn nicht die notwendigen Informationen vorliegen • Überwachung bei Intransparenz erschwert	Genehmigungsverfahren bei Intransparenz erschwert	Auskunftsanspruch Betroffener und Datenschutz-Folgenabschätzung verlangt Transparenz	/
Kontrollierbarkeit	Grundsätzlich in allen Rechtsgebieten relevant aber isoliert unproblematisch.				
Widerstandsfähigkeit	(Gegenstand der Sicherheitsanforderungen)	(Maßgeblich für Prüfungsintervalle und -umfang)	(Robustheit gegen Missbrauch Gegenstand in Störfall-VO)	(Gegenstand der Sicherheitsanforderungen, „security by design“)	(Maßgeblich für Verkehrssicherungspflichten)
Vernetzung	• Verantwortung für Gesamtheit bei externer unabgesprochener Vernetzung unklar • Risikobeurteilung bei externer Vernetzung umfangreich	Gefährdungsbeurteilung bei externer Vernetzung umfangreich	Genehmigungsverfahren / Risikoermittlung bei externer Vernetzung umfangreich	Verantwortlichkeit bei externer Vernetzung ggf. unklar	Beweisschwierigkeiten bei umfangreicher externer Vernetzung
Involviertheit d. Menschen	Grundsätzlich in allen Rechtsgebieten relevant aber isoliert unproblematisch.				
Schadensfolgen	Grundsätzlich in allen Rechtsgebieten relevant, Maßstab für Regulierungsart und -intensität.				
Hohe Veränderbarkeit + geringe Transparenz	Veränderbarkeit bei Intransparenz kann Verdacht der Nonkonformität wecken, sodass Eingriffspflicht besteht.	Gefährdungsbeurteilung nicht abschließend möglich	Anlage ggf. nicht genehmigungsfähig (insbesondere nach Störfall-VO)	• Datenschutz-Folgenabschätzung nicht möglich • Auskunftsanspruch nicht erfüllbar	/
Hohe Veränderbarkeit + geringe Widerstandsfähigkeit	Veränderbarkeit mit Einfluss auf Widerstandsfähigkeit widerspricht zeitpunktbezogenem Ansatz des ProdSG	(Zeitraumbezogene Pflicht des Arbeitgebers zur Gefährdungsbeurteilung kann Veränderbarkeit ausgleichen, siehe aber oben „Veränderbarkeit“)	(Zeitraumbezogene Pflicht des Arbeitgebers zum genehmigungskonformen Betrieb kann Veränderbarkeit ausgleichen, siehe aber oben „Veränderbarkeit“)	Wie zuvor.	/
Hohe Veränderbarkeit + geringe Kontrollierbarkeit	/	/	/	/	Ggf. Haftungslücken
Widerstandsfähigkeit + Vernetzung	Security nicht Teil der Sicherheitsanforderungen	Security nicht Teil der Sicherheitsanforderungen	Security als Schutz vor Missbrauch nur in Störfall-VO vorgesehen	/	/

5.6 Rechtliche Handlungsbedarfe

Sofern die als problematisch identifizierten Systeme rechtssicher umgesetzt werden sollen, wären in den untersuchten Rechtsgebieten die folgenden Aspekte anzugehen.

5.6.1 Produktsicherheitsrecht

Im **Ergebnis** lassen sich für das Produktsicherheitsrecht zusammenfassend folgende Bedarfe identifizieren:

1. **Hoher Informationsbedarf nach Inbetriebnahme:** Weiterlernende Systeme begründen einen hohen Informationsbedarf über ihr Verhalten bzw. ihren Zustand (Transparenz) auch nach Inverkehrbringen bzw. Inbetriebnahme für den Hersteller, um die Anforderungen an das Produkt (Risikobeurteilung, Konformitätsnachweis) zu erbringen und für die Marktüberwachungsbehörde, um die materielle Konformität überwachen zu können.
2. **Schutz vor Missbrauch:** Widerstandsfähigkeit im Sinne von Missbrauchsresilienz wird nicht durch die Sicherheitsanforderungen nach Anhang I der Maschinen-RL adressiert. Ebenso wenig finden sich Vorgaben zur Robustheit bei offener und handlungsbeeinflussender Vernetzung, was neben der Widerstandsfähigkeit auch die Kontrollierbarkeit betrifft.
3. **Externe unabgesprochene Vernetzung:** Diese wird vom Maschinenbegriff der 9. ProdSV nicht berücksichtigt. Gleiches gilt auf Ebene des ProdSG für den Produktbegriff Zudem ist die produktsicherheitsrechtlich konforme Vernetzung des Produkts für den Hersteller bei unabgesprochener Vernetzung mit vielen externen Systemen rechtlich kaum umsetzbar.

5.6.2 Recht des technischen Arbeitsschutzes

Schon bei Beschaffung von Arbeitsmitteln muss der Arbeitgeber im Rahmen der Gefährdungsbeurteilung dafür sorgen, dass nur sichere Arbeitsmittel in seinem Betrieb verwendet werden. Er kann sich grundsätzlich darauf verlassen, dass Arbeitsmittel, die die produktsicherheitsrechtlichen Anforderungen erfüllen, auch sicher sind, sofern ihm nicht andere Informationen vorliegen oder die konkreten Umstände im Betrieb eine andere Gefährdungsbeurteilung nahelegen. Insofern entlastet ihn das Produktsicherheitsrecht, entbindet ihn jedoch nicht von einer Gefährdungsbeurteilung. Daher sind bestimmte Taxonomiedimensionen rechtlich problematisch, sodass jeweils Anpassungen zu erwägen sind.
Im **Ergebnis** kann auf die für das Produktsicherheitsrecht identifizierten Bedarfe verwiesen werden.

5.6.3 Immissionsschutzrecht

Das Immissionsschutzrecht nach BImSchG und Störfall-VO ist für die vorliegende Untersuchung insbesondere hinsichtlich der Pflichten des Betreibers von genehmigungspflichtigen Anlagen von Bedeutung. Denn die Genehmigung erfordert einen feststehenden Sachverhalt, der Gegentand der Prüfung ist, ähnlich dem Konformitätsbewertungsverfahren im Produktsicherheitsrecht. Das produktbezogene Immissionsschutzrecht ergänzt das Produktsicherheitsrecht. Der wesentliche Unterschied ist der klar definierte Grenzwert, der nicht überschritten werden darf.
Im **Ergebnis** lassen sich im Immissionsschutzrecht folgende Handlungsbedarfe ausmachen:

4. **Überwachung von Anlagen:** Im Anwendungsbereich der Störfall-VO bedarf es einer Konkretisierung der Anforderungen an die Ermittlung und Durchführung der erforderlichen Vorkehrungen und der Bestimmung von Art und Maß möglicher Gefahren durch Anlagen, die mit weiterlernenden KI-Komponenten ausgestattet sind.

5. **Schutz vor Missbrauch:** Außerhalb des Anwendungsbereichs der Störfall-VO und der §§ 8a – 8e BSI-G ist der unerlaubte Eingriff Dritter nicht durch die gesetzlichen Sicherheitsanforderung im BImSchG abgedeckt.
6. **Externe unabgesprochene Vernetzung:** Diese Art der Vernetzung ist nicht Gegenstand des Genehmigungsverfahrens, sie wird nach der jetzigen Rechtslage nicht berücksichtigt.

5.6.4 Haftungsrecht

Da das Haftungsrecht auf den Ausgleich im Einzelfall gerichtet ist und sich an den vertraglichen Vereinbarungen, den Verkehrssicherungspflichten und gesetzlichen Verboten orientiert, stellt es selbst auch keine Anforderungen hinsichtlich der Ausgestaltung der KI-Systeme bzw. der Ausprägungen und Kombinationen der Taxonomiedimensionen.
Es ergeben sich im Haftungsrecht folgende Handlungsbedarfe:

1. **Unentdeckte Risiken nicht durch Produkthaftungsrecht gedeckt:** Die Möglichkeit unentdeckter Risiken weiterlernender Produkte wird nicht durch das Produkthaftungsrecht gedeckt. Dadurch bleibt nur die Produzentenhaftung, die für den Geschädigten und den Hersteller mit Unsicherheiten verbunden ist.
2. **Schwierige Verantwortungszuordnung bei vernetzten Systemen:** Extern vernetzte Systeme ohne aktiv involvierte Menschen führen zu Beweisproblemen für den Geschädigten.

5.7 Lösungsansätze

Die anhand der Taxonomie aufgezeigten rechtlichen Handlungsbedarfe sollen im folgenden Teil Grundlage für Überlegungen zur organischen Weiterentwicklung des Rechtsrahmens sein. Diese Überlegungen sollen sich an der Taxonomie und der bestehenden Regelungssystematik orientieren.
Zum Thema KI wurden auf internationaler[363], europäischer[364] und nationaler[365] Ebene bereits verschiedene Vorschläge erörtert und formuliert, wie mit dieser Technologie bzw. den verschiedenen Anwendungen, die die sog. KI erfahren kann, gesellschaftlich umzugehen sei. Ihnen ist gemein, dass grundlegende Forderungen an eine künftige Regulierung der Technologie gestellt werden mit dem Ziel, Vertrauen in die Technologie zu ermöglichen und gleichzeitig möglichst technologieneutral ein Ökosystem zu schaffen, in dem Innovationen möglich sind. Die Wahrung der Grundrechte soll im Vordergrund stehen, es soll z. B. sichergestellt werden, dass es durch den Einsatz der Technologie zu keiner verbotenen Diskriminierung kommt. Außerdem soll ein hohes Maß an Transparenz gewährleistet werden, damit Verantwortlichkeiten klar zugeordnet werden können.
Die vorliegende Untersuchung geht dagegen von dem bestehenden Rechtsrahmen und, vermittelt durch die Taxonomie, von einem KI-Begriff aus, der sich aus distinkten Merkmalen bilden lässt, durch die Vielzahl der möglichen Kombinationen an

---

[363] Z. B. durch die *OECD*, Recommendation of the Council on Artificial Intelligence, OECD/Legal/0449.
[364] *Hochrangigen Expertengruppe für künstliche Intelligenz* (High-Level Expert Group on AI – AI HLEG), Ethik-Leitlinien für eine vertrauenswürdige KI, sowie *Kommission*, Weißbuch – Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, COM(2020) 65 final, sowie der parallel veröffentlichte Bericht der *Kommission* über die Auswirkungen künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung, COM(2020) 64, final.
[365] Veröffentlichungen der Enquete-Kommission “Künstliche Intelligenz” des Bundestages, *Bundesregierung*, Strategie Künstliche Intelligenz der Bundesregierung (Stand: November 2018).

Merkmalen aber dennoch offen ist. Durch die abstrakte und beispielhafte juristische Erörterung der einzelnen Taxonomiedimensionen und einiger Kombinationen von ihnen konnten in den hier untersuchten Rechtsbereichen diejenigen Regelungen benannt werden, die bei der Realisierung der durch die Taxonomie definierten KI-Systeme, z. B. als Produkt, rechtliche Unsicherheiten mit sich bringen. Mit dieser Erkenntnis lassen sich Handlungsbedarfe bestimmen, für die dann Lösungsansätze erörtert werden können.
Damit unterscheidet sich der hier verfolgte Ansatz von den genannten Vorschlägen insofern, als dass er nicht ganzheitliche Regulierungsziele aufstellt, sondern von den einzelnen Regelungen ausgeht und versucht, diese technologieoffen und an der Taxonomie orientiert weiterzuentwickeln. Es werden also für die verschiedenen Regelungsbereiche jeweils eigene Vorschläge erörtert. D
Untersucht werden sollen hier jedoch nur die Regelungsbereiche, die (auch) den Schutz des Lebens und der Gesundheit von Menschen bezwecken, also das Produktsicherheitsrecht, das Recht des technischen Arbeitsschutzes, das Immissionsschutzrecht und das Haftungsrecht. Die DSGVO soll hier außen vor bleiben.

5.7.1 Organische Weiterentwicklung des Rechts

Organische Weiterentwicklung des Rechts meint in diesem Zusammenhang, die bestehenden rechtlichen Strukturen wo nötig und möglich anzupassen und als Ausgangspunkt für neue Ansätze zu nutzen. Das Recht, insbesondere das Technikrecht, definiert die Rahmenbedingungen für unser Zusammenleben, also hier für die Entwicklung und Verwendung von Technik. Es reagiert auf die Entwicklung neuer Technologien und vollzieht sie nach, indem es Anforderungen stellt, Verantwortlichkeiten regelt und ggf. auch Verbote ausspricht. Das bestehende Recht ist damit auch Ausdruck technologischen Fortschritts: In ihm manifestieren sich die Reaktionen unserer Gesellschaft auf die sozialen, gesundheitlichen, ökonomischen und ökologischen Auswirkungen von Technik. Es ist darüber hinaus in unserem demokratisch verfassten Gemeinwesen Ergebnis eines formalisierten gesellschaftlichen Aushandlungsprozesses und nicht bloßer Ausdruck technologischer, sozialer oder gesundheitlicher Notwendigkeiten. Indem das Recht also die Rahmenbedingungen setzt, reagiert es nicht nur auf bereits bestehende Entwicklungen. Es greift auch kommenden Entwicklungen voraus, da sich die Forschung und Entwicklung von Technologien im Rahmen des Rechts bewegen muss. Da das (formelle) Recht sich jedoch verhältnismäßig langsam ändert, bleiben einmal getroffene Entscheidungen des Gesetzgebers für längere Zeit maßgeblich. Der Gesetzgeber muss also drei verschiedenen Ansprüchen gerecht werden: Der Wahrnehmung seiner grundrechtlichen Schutzpflichten, der Schaffung eines klaren Rahmens für rechtssicheren und technologischen Fortschritt sowie der Wahrung der grundrechtlich gewährleisteten (unternehmerischen) Freiheiten und dem daraus folgenden Anspruch größtmöglicher Technologieoffenheit.
Das Produktsicherheitsrecht nach dem ProdSG versucht diesen Spagat dadurch zu bewerkstelligen, dass es nur sehr allgemeine Sicherheitsanforderungen an das Produkt und die Pflichten für den Hersteller formuliert. Die Konkretisierung dieser allgemeinen Regelungen erfolgt dann im Einzelfall entweder durch den Hersteller selbst, der auf Grundlage eine Risikobeurteilung für das jeweilige Produkt bestimmt, wie dieses ausgestaltet sein muss, um den allgemeinen gesetzlichen Anforderungen

zu entsprechen. Oder es werden für die einzelne Produktgruppen und Sicherheitsanforderungen technische Normen durch private Normungsgremien erstellt, die den aktuellen Stand der Technik widerspiegeln und an denen sich der Hersteller orientieren kann. Der Gesetzgeber gibt also nur den groben Rahmen vor, die konkrete Ausgestaltung der Sicherheitsanforderungen liegt hingegen in privater Hand.

5.7.2 Das Verhältnis zwischen ProdSG und BetrSichV und BImSchG

Während das ProdSG die an der Fertigung und dem Vertrieb des Produkts beteiligten Wirtschaftsakteure in den Blick nimmt, adressieren das ArbSchG mit der BetrSichV und das BImSchG die Verwender der Technologie. Wie bereits zur Entwicklung des Produktsicherheitsrechts dargestellt, ist das ProdSG aus dem nutzerorientierten Arbeitsschutzrecht heraus entstanden. Die öffentliche-rechtliche Verpflichtung des Herstellers zur Gewährleistung der Sicherheit und Zuverlässigkeit der von ihm in Verkehr gebrachten technischen Produkte hatte sich als erforderlich erwiesen. So wird der Arbeitgeber als Verantwortlicher für die Sicherheit im Betrieb entlastet, ohne dass er von seinen Pflichten entbunden wird: Er darf nicht blind auf die Herstellerangaben vertrauen, sondern muss die konkreten Gegebenheiten im Betrieb beachten und entsprechende Maßnahmen ergreifen.
Auch der Anlagenbetreiber kann Produkte im Sinne des ProdSG als Anlagen einsetzen. Für die Einhaltung der Pflichten des BImSchG ist er jedoch allein verantwortlich. Auch wenn immissionsschutzrechtliche Anforderungen in ähnlicher Form in den speziellen Sicherheitsanforderungen an das als Anlage eingesetzte Produkt bereits Niederschlag gefunden haben können, werden sie trotzdem erneut Gegenstand des Genehmigungsverfahrens.
Diese ordnungsrechtlichen Verantwortlichkeiten sind in der folgenden Tabelle zusammengefasst. Dabei wird nach der Zeit bis zum Inverkehrbringen des Produkts und der darauffolgenden Zeit unterschieden und ergänzend nach der Marktüberwachung sowie dem Verwender, beispielsweise Verbrauchern, denen keine ordnungsrechtlichen Pflichten obliegen.

	Hersteller	Marktüberwachung	Verwender	Arbeitgeber	Anlagenbetreiber
Vor Inverkehrbringen / Inbetriebnahme	• Risikobeurteilung • Konformitätsbewertung • CE-Kennzeichnung	• Einfuhrkontrolle	/	• Informationsbeschaffung • Gefährdungsbeur-teilung • ggf. Prüfung vor Inbetriebnahme	• Genehmigung
Nach Inverkehrbringen / Inbetriebnahme	• Produktbeobachtung • Notwendige Maßnahmen zur Risikobeherrschung • Unterrichtung der Marktüberwachung	• Sachverhaltsermittlung • Bei Gesetzesverstoß: Ergreifen erforderlicher Maßnahmen	• *Bestimmungs-gemäße Verwendung* • *ggf Beschwerde* • ggf. Herstellerpflichten	• Prüfung Arbeitsmittel • Überprüfung Gefährdungsbeurtei-lung • Ggf. Aktualisierung Gefährdungsbeurteilung	• Betrieb gemäß Genehmigung • Anzeige von Änderung • ggf. Neugenehmigung

Abb. 5.5 Pflichten zur Gewährleistung der Sicherheit eines Produkts/Arbeitsmittels/ Anlage. In der Darstellung kursiv sind die Obliegenheiten, deren Nichteinhaltung

höchstens zu einem Verlust von Vorteilen führt: So hat die nicht bestimmungsgemäße Verwendung des Produkts durch den Verwender keine unmittelbaren rechtlichen Nachteile, kann aber z. B. zu einer Beschädigung des Produkts führten.

Diese scheinbare Doppelung der Pflichten ist auf die unterschiedlichen Regelungszwecke zurückzuführen. Während das Produktsicherheitsrecht die von Produkten ausgehenden Gefahren bereits im Herstellungsprozess vermeiden soll, zielen die anderen vorgestellten Gesetze auf deren konkrete Nutzung ab. Mit Inverkehrbringen bzw. Inbetriebnahme hat der Hersteller jedoch keinen direkten Einfluss mehr auf die Nutzung seines Produkts. Ab diesem Zeitpunkt kann er das Produkt höchstens noch beobachten, wozu er für bestimmte Produkte wie Verbraucherprodukte oder Spielzeug auch verpflichtet ist. So werden ab diesem Zeitpunkt andere Adressaten in den Blick genommen, um einen möglichst lückenlosen Schutz gefährdeter Rechtsgüter und Interessen zu gewährleisten. Es kommen auch neue Schutzziele hinzu. Das BImSchG erweitert den Kreis der geschützten Interessen um die der Umwelt der Anlage. Dementsprechend sind die von den verschiedenen Gesetzen an das Produkt (bzw. Arbeitsmittel/Datenverarbeitung/Anlage) gestellten Anforderungen nicht immer deckungsgleich.
Außerdem variieren die Rollen des Staates, die er in den jeweiligen Regelungsbereichen spielt. Am deutlichsten wird dies, wenn man die Rolle der Marktaufsichtsbehörde nach ProdSG und die der Aufsichtsbehörde nach ArbSchG betrachtet. Die Marktaufsichtsbehörde hat kein Entschließungsermessen, wenn aufgrund der ihr vorliegenden Informationen ein Verdacht der Pflichtverletzung durch die Produktverantwortlichen begründet ist.[366] Die Aufsichtsbehörden dagegen haben einen weiten Spielraum bei Verdacht einer Pflichtverletzung durch den Arbeitgeber.

> *Ein und dasselbe KI-System kann so Gegenstand unterschiedlicher Gesetze sein bzw. im Laufe seines „Lebens“ werden. Unterschiedliche Personen können für seine Beschaffenheit und die von ihm ausgehenden Gefahren verantwortlich sein, mitunter auch nebeneinander. Auch die technischen Anforderungen können variieren, da jedes Gesetz unterschiedlichen Zwecken dient und zu deren Erreichung eigene Mittel vorhält.*

Zeitlicher Dreh- und Angelpunkt der Verantwortungsallokation zwischen der Herstellerseite und den Verwendern ist dabei das Inverkehrbringen bzw. die Inbetriebnahme. Er muss die bestimmungsgemäße und vorhersehbare Verwendung im Blick haben, wenn er die Risikobeurteilung vornimmt, ist jedoch nicht verantwortlich, wenn sich die Verwendung außerhalb dessen bewegt, was er seiner Risikobeurteilung zugrunde gelegt hat bzw. zugrunde legen konnte. Die Pflichten des Arbeitgebers greifen bereits früh, da er schon bei Beschaffung des Arbeitsmittels die Sicherheit im Betrieb im Blick haben muss. Auch der Betreiber einer Anlage im Sinne des BImSchG wird bereits früher verantwortlich, nämlich mit Errichtung der Anlage. Entscheidend bleibt jedoch spätestens der Zeitpunkt der Inbetriebnahme, da der Hersteller damit aus der Pflicht entlassen ist, vorbehaltlich spezieller Produktbeobachtungspflichten wie z. B. bei Verbraucherprodukten.

[366] Siehe dazu oben 5.3.4.1.1.2 – Marktüberwachungsmaßnahmen.

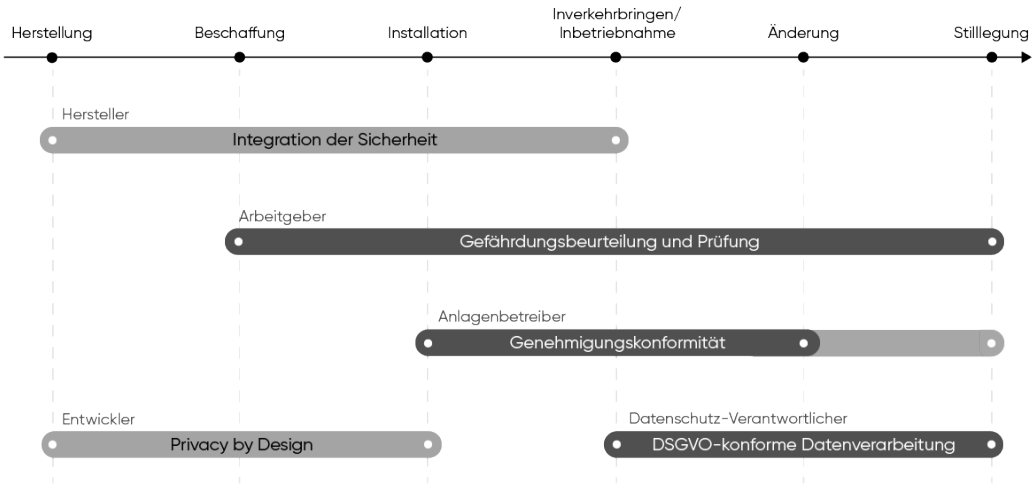


Abb. 5.6 Die ordnungsrechtlichen Verantwortlichkeiten den Lebenszyklus eines Produkts im Vergleich.

Der hier vorgeschlagenen Regulierungsansätze versuchen diese Struktur weiterzuentwickeln. Sie werden in aufsteigender Reihenfolge nach dem Umfang der dort vorgeschlagenen Rechtsänderungen sortiert dargestellt.

5.7.3 Weiterentwicklung des Produktsicherheitsrechts

Anpassungen im Produktsicherheitsrecht können damit auch in den zeitlich nachfolgend einschlägigen Rechtsbereichen Einfluss haben, da die Verwender regelmäßig auf die Konformität des Produkts abstellen.
Die Untersuchung der Taxonomiedimensionen hat gezeigt, dass die zeitpunktbezogene Herstellerverantwortung dem Wissensvorsprung des Herstellers und seiner Fähigkeit, insbesondere bei für den Verwender intransparenten Systemen am ehesten neu auftretenden Risiken durch Änderungen an der Systemarchitektur begegnen zu können, nicht gerecht wird bzw. nicht angemessen rechtlich würdigt.
Die hier untersuchten Lösungsansätze konzentrieren sich daher auf den Hersteller.

5.7.3.1 Weiterentwicklung zur Regulierung von Veränderbarkeit und Transparenz

Die beiden Taxonomiedimensionen der Veränderbarkeit und Transparenz führen in hoher Ausprägung bei der Risikobeurteilung durch den Hersteller zu Problemen und werfen die Frage auf, ob durch eine Veränderung ein neues Produkt entsteht. Es sollen hier drei Ansätze erörtert werden, wie diese Taxonomiedimensionen im ProdSG reguliert werden könnten, um die Umsetzung möglichst hoher Ausprägungen zu ermöglichen.

5.7.3.1.1 **Ansatz 1**: Anpassung des Produktbegriffs

Der Produktbegriff des § 2 Nr. 22 ProdSG ist zusammen mit den tatbestandsmäßigen Handlungen des Inverkehrbringens bzw. der Bereitstellung für die Bestimmung des Pflichtenprogramms des Herstellers und der anderen Marktakteure nach ProdSG maßgeblich.
Der Produktbegriff könnte daher ergänzt werden, um klarzustellen, dass eine Veränderung des Produkts im Rahmen der bestimmungsgemäßen Veränderbarkeit nicht zur Herstellung eines neuen Produkts führt.
Eine Regelung zur Ergänzung des § 2 Nr. 22 ProdSG könnte wie folgt lauten[367]:

> *„[...]; die Veränderung [1] eines Produkts [2] im Rahmen der bestimmungsgemäßen Veränderbarkeit [3] stellt keinen Fertigungsprozess [4] dar.“*

Zu [1]: Die Veränderung umfasst jede sicherheitsrelevante Änderung, die zu einem neuen Produkt führen würde. Es ist dabei offen, wer diese Veränderung anstößt. Wie diese Veränderung konkret aussieht, ist durch den Hersteller zu bestimmen.
Zu [2]: Dieser zweite Halbsatz konkretisiert die Begriffsdefinition in § 2 Nr. 22 ProdSG. Die Produkteigenschaft selbst wird also nicht berührt.
Zu [3]: Die Veränderung muss im Rahmen der bestimmungsgemäßen Veränderbarkeit erfolgen. Damit wird die Veränderbarkeit durch den Hersteller definierbar. Er kann so einen klaren Rahmen setzen für alle nach Bereitstellung möglichen Veränderungen. Dadurch wird sein Pflichtenprogramm bei der Integration der Sicherheit klar, denn er hat seine Risikobeurteilung auch auf die von ihm vorausgesetzten Veränderungen auszurichten. Für die Verwender wird dadurch erkennbar, mit welchen sicherheitsrelevanten Veränderungen sie rechnen können. Initiieren sie darüber hinausgehende Veränderungen, können wie gehabt Herstellerpflichten entstehen.
Zu [4]: Der Halbsatz grenzt den Tatbestand des § 2 Nr. 22 ProdSG ein, indem das Tatbestandsmerkmal des Fertigungsprozesses konkretisiert wird. Bestimmungsgemäße Veränderungen sind nunmehr explizit nicht als Fertigungsprozess anzusehen. Damit soll anerkannt werden, dass es neben den herkömmlichen „unfertigen“ Produkten, wie den unvollständigen Maschinen, veränderbare Produkte gibt, für deren Veränderung sich der Hersteller durch eine entsprechende Zweckbestimmung verantwortlich erklärt. Er hat diese Veränderbarkeit bei der Integration der Sicherheit mitberücksichtigt, sodass es bei einer entsprechenden Veränderung nicht eines Auflebens der produktsicherheitsrechtlichen Pflichten bedarf, um den Gesetzeszweck zu erfüllen, also ein möglichst hohes Sicherheitsniveau zu gewährleisten.
Insgesamt hat diese verhältnismäßig kleine Anpassung vor allem klarstellende Funktion, indem sie die bereits bestehenden Herstellerpflichten unberührt lässt, sie jedoch für die bestimmungsgemäß veränderbaren Produkte klar dem Hersteller zuordnet.
Davon nicht erfasst werden jedoch solche Veränderungen, die bei Bereitstellung des Produkts noch nicht erkennbar waren, da es sich um ein hochgradig veränderbares Produkt handelt, das sich zudem durch große Intransparenz auszeichnet. Diese Kombination versucht der folgende Ansatz zu regulieren.

---

[367] Die einzelnen Elemente bzw. Tatbestandsmerkmale der Norm sind durch die Zahlen in eckigen Klammern zur anschließenden Kommentierung markiert.

5.7.3.1.2 **Ansatz 2**: Zeitraumbezogene Pflicht des Herstellers zur Erhaltung der Sicherheit mit engem Begriff des „veränderlichen“ Produkts

Ein anderer Ansatz nimmt nicht allein den Produktbegriff in den Blick, sondern die Herstellerpflichten.
Der Hersteller soll nach diesem Ansatz die Möglichkeit bekommen, 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. Er kann dann z. B. Updates zur Verfügung stellen oder seine Instruktionen oder Warnhinweise zu dem Produkt aktualisieren. Er soll also sein Produkt nach Inbetriebnahme begleiten, weshalb hier von *Produktbegleitung* die Rede sein kann. Durch Erarbeitung eines entsprechenden *Produktbegleitungskonzepts* legt er für sein konkretes Produkt die jeweils geeigneten Verfahren dazu fest.
Tatsächlich besteht schon jetzt im Produktsicherheitsrecht bei bestimmten Produkten die Pflicht zur Produktbeobachtung nach Inverkehrbringen, um unerkannten Risiken begegnen zu können, so z. B. in § 6 Abs. 3 ProdSG bei Verbraucherprodukten. Bei dieser herkömmlichen Art der Produktbeobachtung geht es darum, ein zunächst konformes und damit legal auf den Markt gekommenes Produkt entgegen der abgeschlossenen und ordnungsgemäß durchgeführten Risikobeurteilung als gefährlich zu erkennen, es zurückzurufen, davor zu warnen oder ähnliche Maßnahmen zu ergreifen. Bei den hier problematisierten veränderbaren Produkten ist das unbekannte Risiko dagegen nicht die Ausnahme, sondern wird als Teil der bestimmungsgemäßen Verwendung vom Hersteller vorausgesetzt und vom Markt erwartet. Es handelt sich nicht um die tolerierbare und hinreichend geringe Wahrscheinlichkeit, dass noch unerkannte Risiken bestehen, der damit begegnet wird, dass der Hersteller eine Produktbeobachtungspflicht hat. Vielmehr ist das unbekannte Risiko aufgrund einer hohen Veränderbarkeit wesentliche Eigenschaft des Produkts. Denn die hohe Veränderbarkeit, eventuell in Verbindung mit niedriger Transparenz und geringer Kontrollierbarkeit zeichnen das Produkt als „KI-Produkt“ aus und sind damit wertbildende Faktoren. Wenn das Produkt also im Rahmen der bestimmungsgemäßen Verwendung ein im maßgeblichen Zeitpunkt der Bereitstellung am Markt noch unbekanntes „Verhalten“ an den Tag legt, so ist das keine Folge unvorhergesehener und unerwünschter Produkteigenschaften, sondern wird vom Hersteller und dem Markt vorausgesetzt oder zumindest erwartet.
Es geht also nicht mehr um die Gewährleistung der Sicherheit in dem Sinne, dass der Hersteller alles ihm Mögliche dafür tut, damit sich nur solche Produkte in Verkehr befinden, die während ihrer bestimmungsgemäßen Lebensdauer den Zustand aufweisen, der im Zeitpunkt der Bereitstellung am Markt für diese Lebensdauer vorausgesetzt wurde. Die Gewährleistung der Sicherheit meint für die „*veränderlichen*“ Produkte, dass das für eine Bereitstellung am Markt vorausgesetzte Sicherheitsniveau erhalten bleibt, obwohl sicherheitsrelevante Eigenschaften erst später abschließend bestimmt werden können oder sich ändern.
Das herkömmliche Prinzip der Produktbeobachtung zur Beherrschung unerkannter und unvertretbarer Risiken sollte daher weiterentwickelt werden zu einer Produktbeobachtung zur Ermöglichung der Bereitstellung von Produkten, bei denen vorausgesetzt wird, dass sie mit den herkömmlichen Verfahren zur Risikobeurteilung oder neuen Simulationsverfahren noch nicht abschließend beurteilt werden können

und ggf. weitere Maßnahmen durch den Hersteller nötig sind, um die Sicherheit trotz Veränderbarkeit oder Komplexität zu erhalten.
Der Hersteller wird im Fertigungsprozess zunächst darüber entscheiden, ob er ein nicht abschließend beurteilbares Produkt herstellt. Dabei legt er einerseits den bestimmungsgemäßen Verwendungszeitraum fest. Anderseits bestimmt er, wie veränderbar und intransparent er sein Produkt gestaltet und inwiefern es sich damit von einem „herkömmlichen“ Produkt unterscheidet. Daraus folgt wiederum, ob und inwiefern eine abschließende Risikobeurteilung nicht mehr möglich ist, also ob und inwiefern eine neue Form der Produktbeobachtung erforderlich ist.
Im Folgenden werden die für die Einführung einer Pflicht zur Erstellung eines Produktbegleitungskonzepts denkbaren Regelungen vorgestellt, also zunächst die Definition des „veränderlichen“ Produkts und dann die entsprechende Rechtsfolge.

5.7.3.1.2.1 Definition des „wandelbaren“ Produkts

Die Definition des „wandelbaren“ Produkts bestimmt den Anwendungsbereich neu zu schaffender Herstellerpflichten. Wenn ein Produkt so beschaffen ist, dass im maßgeblichen Zeitpunkt für das Produkt oder für spezifische Risiken eine abschließende Risikobeurteilung nicht mehr möglich ist, dann handelt es sich um ein „wandelbares“ Produkt, für das spezielle Pflichten gelten.
Unabhängig davon, wie der Begriff des wandelbaren Produkts eingeführt wird, könnte die Definition wie folgt lauten:
Zu den Elementen der Norm im Einzelnen:

> *„Wandelbare [1] Produkte [2]sind solche Produkte, für die insgesamt oder hinsichtlich spezifischer [3] Risiken [4] eine den allgemeinen Sicherheitsanforderungen entsprechende [5] abschließende Risikobeurteilung [6] aufgrund der bestimmungsgemäßen Veränderbarkeit [7] im Zeitpunkt der Bereitstellung oder dem nach der für das Produkt geltenden Verordnung nach § 8 Abs. 1 ProdSG maßgeblichen Zeitpunkt [8] nicht möglich ist.“*

Zu [1]: Das Attribut „*wandelbar*“ bezeichnet die Tatbestandsvoraussetzung der Pflicht zur Erstellung eines Produktbegleitungskonzepts, die in dieser Definition festgelegt wird. Aus Sicht der Taxonomie wäre der Begriff „veränderbar“ verkürzend, da nicht nur die Taxonomiedimension der Veränderbarkeit erfasst sein soll, sondern auch Ausprägungen der Transparenz und der Kontrollierbarkeit. Daher wurde hier der Begriff „*wandelbar*“ gewählt.
Allerdings ist die Veränderbarkeit bei diesem Ansatz prägendes Merkmal der von der vorgeschlagenen Definition erfassten Produkte. Intransparente, aber nicht veränderbare – oder nur zu einem geringen Maße veränderbare - Produkte sollen nach diesem Ansatz nicht unter den Begriff des „*wandelbaren*“ Produkts fallen, wenn eine abschließende Risikobeurteilung nur deshalb nicht möglich ist, weil noch keine geeigneten Verfahren zur Verfügung stehen. Für die nach diesem Ansatz erfassten „*wandelbaren*“ Produkte ist die Unmöglichkeit der abschließenden Risikobeurteilung nicht Folge verfahrenstechnischer Unzulänglichkeiten, sondern Voraussetzung der Erfüllung des bestimmungsgemäßen Zwecks (dazu mehr bei der Kommentierung zu [7]).

Zu [2]: Mit der Definition wird eine Teilmenge der bereits in § 2 Nr. 22 ProdSG definierten Produkte bestimmt. Jedes „*wandelbare*“ Produkt ist also auch Produkt im Sinne des § 2 Nr. 22 ProdSG.
Gleichzeitig wird damit für den hergebrachten Produktbegriff des § 2 Nr. 22 ProdSG zweierlei klargestellt:
Erstens, dass die bestimmungsgemäße Veränderbarkeit kein negatives Tatbestandsmerkmal des Produkts gemäß § 2 Nr. 22 ProdSG ist. Die Veränderbarkeit einer Ware, einer Zubereitung oder eines Stoffes schließt also nicht die Produkteigenschaft aus. Wenn an einem Produkt nach § 2 Nr. 22 ProdSG also sicherheitsrelevante Veränderungen vorgenommen werden können, die nicht Teil der bestimmungsgemäßen Verwendung sind und damit nicht Gegenstand der Risikobeurteilung waren, dann kann ein neues Produkt entstehen, womit auch entsprechenden Herstellerpflichten etc. einhergehen. Diese Möglichkeit steht aber nicht der Produkteigenschaft des veränderten Produkts entgegen. Gleichzeitig kann der Hersteller eine Veränderbarkeit seines Produkts gemäß § 2 Nr. 22 ProdSG vorsehen und sie zum Gegenstand der Risikobeurteilung machen. Dann entsteht bei Eintritt einer solchen Veränderung kein neues Produkt. Im Unterschied dazu handelt es sich um ein „*wandelbares*“ Produkt mit den entsprechenden Rechtsfolgen zur Beherrschung der damit einhergehenden Unwägbarkeiten wenn diese Veränderung nicht mit der Risikobeurteilung abschließend beherrschbar ist.
Zweitens wird damit für die Verwender des Produkts klar, dass bei bestimmungsgemäßer Veränderung kein neues Produkt entstehen kann, sie also auch keine Herstellerpflichten treffen können.
Die Definition des „*wandelbaren*“ Produkts übernimmt, sofern sie Teil des ProdSG wird, damit die Funktion der oben in Ansatz 1 diskutierten Klarstellung in § 2 Nr. 22 ProdSG.
Zu [3]: Bei der Integration der Sicherheit und der dafür erforderlichen Risikobeurteilung wird das Produkt als Ganzes betrachtet. Es kann daher für das Produkt insgesamt unmöglich sein, wegen seiner Veränderbarkeit eine abschließende Risikobeurteilung anzustellen. Die Unterscheidung nach einer „*insgesamt*“ nicht abschließenden Risikobeurteilung und der nur „*für spezifische Risiken*“ nicht abschließende Risikobeurteilung ist jedoch nötig, da sich die Rechtsfolge der Pflicht zur Erstellung eines Produktbegleitungskonzepts nur auf die Risiken bezieht, die nicht abschließend beurteilt werden können. Es kann also das Produkt „*insgesamt* wandelbar“ sein, oder nur im Hinblick auf „*spezifische Risiken*“.
Zu [4]: Mit den „*Risiken*“ sind alle Risiken für die geschützten Rechtsgüter gemeint, also im Anwendungsbereich des ProdSG gemäß § 3 Abs. 2 S. 1 ProdSG für die Sicherheit und Gesundheit von Personen.
Sofern gemäß § 1 Abs. 4 ProdSG spezielles Produktsicherheitsrecht mit entsprechenden oder weitergehenden Vorschriften Vorrang vor denen des ProdSG hat, werden auch die dort geschützten Rechtsgüter erfasst. Auch für sie können also nicht abschließend beurteilbare Risiken bestehen, die den Tatbestand des „*wandelbaren*“ Produkts ausmachen können. Hier kommt die Auffangfunktion des ProdSG zum Tragen. Auch wenn das spezielle Produktsicherheitsrecht keine Regelungen zu „*wandelbaren*“ Produkten trifft, gelten die in Ansatz 2 diskutierten Neuregelungen im ProdSG trotzdem. Wenn die Regelungen zum „*wandelbaren*“ Produkt vertikal erfolgen, also nur für bestimmte Produktgruppen im speziellen Produktsicherheitsrecht, dann gelten sie nur für diese Produktgruppe.

Zu [5]: Die Risikobeurteilung ist Grundlage für die Integration der Sicherheit. In § 3 Abs. 2 S. 3 ProdSG finden sich Anhaltspunkte für die Durchführung einer Risikobeurteilung, soweit im nichtharmonisierten Bereich der § 3 Abs. 2 ProdSG Anwendung findet. Ansonsten finden sich die Anforderungen an die Risikobeurteilung[368] in der für die jeweilige Produktgruppe einschlägigen Produktsicherheitsverordnung nach § 8 Abs. 1 ProdSG bzw. in der unmittelbar anwendbaren europäischen Verordnung.
Die so ermittelten Anforderungen an die Risikobeurteilung sind für den Tatbestand des „*wandelbaren*“ Produkts entscheidend, da sie den Maßstab bilden dafür, wie genau und abschließend die Risikobeurteilung im konkreten Fall sein muss. Letztlich werden sich die Details hierzu in den jeweils einschlägigen technischen Normen finden, die durch die Normungsgremien erarbeitet werden.
Den technischen Normen kommt so eine Schlüsselrolle bei der Bestimmung des „*wandelbaren*“ Produkts zu. Dort wird festgelegt, welche Verfahren zu Risikobeurteilung und zum Nachweis der Sicherheit dem Stand der Technik entsprechen. Trotzdem treffen sie keine bindenden oder abschließenden Regelungen. Dem Hersteller ist es unbenommen, auch für noch nicht normierte Produkte eine eigene Risikobeurteilung anzustellen. Er kann dann ohne Rückgriff auf technische Normen die Risikobeurteilung abschließen oder feststellen, dass dies nicht möglich ist und damit ein „*wandelbares*“ Produkt vorliegt.
Zu [6]: Die „*abschließende Risikobeurteilung*“ ist Voraussetzung für die Integration der Sicherheit. Ist sie nicht möglich, liegt soweit ein „*wandelbares*“ Produkt vor.
Bei dem vorliegenden Produkt werden also sowohl beurteilbare Risiken ermittelt als auch solche, die mit den hergebrachten Methoden entsprechend den jeweils einschlägigen Sicherheitsanforderungen nicht beurteilt werden können. Es kann aber auch ein Rest an Zuständen bleiben, die als möglich erkannt werden können, jedoch weder hinreichend bestimmt und ermittelt noch beurteilt werden können. Dabei handelt es sich nicht um (tolerierbare) Restrisiken. Denn Restrisiken sind bekannt, können also ermittelt und beurteilt werden. Hingegen gehen die hier erfassten Risiken von Zuständen aus, die schon nicht ermittelt werden können oder für die jedenfalls keine Aussage über ihre Eintrittswahrscheinlichkeit möglich ist.
Der verbleibende unbestimmbare Rest der dem Produkt innewohnenden Risiken markiert zudem den Bezugspunkt für die aus der „*Wandelbarkeit*“ des Produkts folgenden Pflichten des Herstellers.
Durch diese Definition wird das wesentliche Erschwernis im Produktsicherheitsrecht für die Umsetzung der hier untersuchten Systeme benannt und zum Tatbestandsmerkmal erhoben: die unvollständige Risikobeurteilung. Damit soll ein gewisses Maß an Technologieneutralität gewahrt bleiben. „KI-Produkte“ oder „weiterlernende“ Produkte werden nicht definiert.
Zu [7]: Die „*bestimmungsgemäße Veränderbarkeit*“ des Produkts muss der Grund dafür sein, dass die Risikobeurteilung nicht abgeschlossen werden kann.
Damit soll der Begriff des „*wandelbaren*“ Produkts auf die Taxonomiedimension der Veränderbarkeit hin verengt werden. Zwar können auch ausgeprägt schwierig zu kontrollierende oder intransparente Produkte eine abschließende Risikobeurteilung unmöglich machen, weil z. B. noch keine geeigneten Test- oder Simulationsverfahren existieren. In diesem Fall soll jedoch kein „*wandelbares*“ Produkt vorliegen. Der

[368] Die Bezeichnung variiert in den unterschiedlichen harmonisierten Bereichen.

Hersteller soll nicht die legalisierte Möglichkeit bekommen, nur eine „unfertige“ Risikobeurteilung über die hier vorgeschlagenen neuen Sonderregeln zum Produktbegleitungskonzept zu erstellen. Durch diesen engen Begriff des „*wandelbaren*“ Produkts soll verhindert werden, dass unter Berufung auf die Unmöglichkeit der abschließenden Risikobeurteilung „unreife“ Produkte auf den Markt kommen und der Hersteller dann im Betrieb weiter Informationen sammelt, um sein Produkt anzupassen. Mithin soll diese Möglichkeit nur für solche Produkte bestehen, bei denen die hohe Veränderbarkeit wesentliche Eigenschaft ist.
Der Tatbestand des „*wandelbaren*“ Produkts wird damit stark verengt und es stellt sich die Frage, für welche Produkte dieser Begriff überhaupt anwendbar ist. Wann ist eine Risikobeurteilung „nur“ praktisch unmöglich, weil sie entweder derart aufwendig ist, dass sie wirtschaftlich nicht abschließbar ist oder weil nach dem aktuellen oder sogar neuesten Stand der Technik noch keine belastbaren Verfahren zur Risikobeurteilung zur Verfügung stehen? Und wann ist die Unmöglichkeit der abschließenden Risikobeurteilung der Veränderbarkeit des Produkts immanent und kann auch mit künftigen Verfahren nicht überwunden werden?
Hier liegt nach der Konzeption des Produktsicherheitsrechts die Entscheidung zunächst beim Hersteller. Durch die Wahl der eingesetzten Technik gibt er vor, wie die Risikobeurteilung gestaltet wird. Im Regelfall kommt es anschließend auf die technischen Normen an. Wenn sie ein hinreichendes Verfahren für die konkret eingesetzte Technik zur Verfügung stellen, so ist dieses anzuwenden. Gibt es kein Verfahren für die konkret eingesetzte Technik, ist es wieder am Hersteller, ein eigenes Verfahren zur Risikobeurteilung zu entwickeln. Kommt er dabei zu dem Ergebnis, dass aufgrund der Veränderbarkeit keine abschließende Risikobeurteilung möglich ist, kann er auf das Produktbegleitungskonzept zurückgreifen. Durch die Marktüberwachungsbehörde ist *ex post* zu überprüfen, ob er dies zu Recht getan hat oder ob die Risikobeurteilung „nur“ wegen hoher Intransparenz nicht abgeschlossen werden konnte. Sofern eine notifizierte Stelle in der Fertigungskontrolle mitwirkt, kann sie auch schon *ex ante* eingreifen, wenn die Risikobeurteilung nicht abgeschlossen ist. Trotzdem bleibt die Frage, wie die wesentliche Veränderbarkeit zu bestimmen ist. Diese wird durch den Hersteller im Einzelfall zu klären sein. Damit bleibt dieser Ansatz insofern mit Rechtsunsicherheiten für den Hersteller verbunden.
Die gesetzliche Vorgabe, dass nur bestimmte Veränderbarkeiten den Tatbestand des „*wandelbaren*“ Produkts eröffnen, kann jedoch auch für die Normungsgremien Anlass sein, für eben diese Veränderbarkeiten Standards aufzusetzen, um sie zu definieren und den Herstellern damit einen Maßstab zu bieten. Es könnte auf die Erfahrungen mit der Unterscheidung zwischen unwesentlichen und wesentlichen Änderungen von Anlagen gemäß §§ 15, 16 BImSchG zurückgegriffen werden.
Zu [8]: Die Definition des maßgeblichen Zeitpunkts hängt ebenfalls von der jeweiligen Produktgruppe ab. So findet sich z. B. in der Maschinen-RL neben dem Inverkehrbringen die Inbetriebnahme als maßgeblicher Zeitpunkt, während das ProdSG auf das Inverkehrbringen und die Bereitstellung abstellt.

5.7.3.1.2.2 Die Rechtsfolge: Produktbegleitungskonzept

Sofern der Tatbestand des „wandelbaren“ Produkts erfüllt ist, sollen den Hersteller nach diesem Ansatz Pflichten zur Gewährleistung der Sicherheit treffen, die über den maßgeblichen Zeitpunkt hinaus gehen und den gesamten bestimmungsgemäßen Produktlebenszyklus umfassen.

Soweit er die Risikobeurteilung nicht abschließen kann, soll er ein *Produktbegleitungskonzept* erstellen und umsetzen. Er soll auch nach Bereitstellung des Produkts am Markt diejenigen Informationen aus dem Betrieb des Produkts sammeln, die zur Ergänzung oder Aktualisierung seiner Risikobeurteilung erforderlich sind. Wird durch eine wiederholte Neubeurteilung deutlich, dass Maßnahmen zur Gewährleistung des jeweils gebotenen Sicherheitsniveaus erforderlich werden, so hat er diese zu ergreifen.
Es folgt formuliert werden:

> *„Der Hersteller eines veränderlichen Produkts erstellt ein Produktbegleitungskonzept [1], soweit er die Risikobeurteilung nicht abschließen kann [2]. Er setzt das Produktbegleitungskonzept über den bestimmungsgemäßen Lebenszyklus [3] des Produkts um [4]. Er überprüft das Produktbegleitungskonzept in regelmäßigen Abständen und überarbeitet es, soweit dies nach dem Ergebnis der Überprüfung erforderlich ist [5].“*

Zu den Elementen der Regelung im Einzelnen:
Zu [1]: Das Produktbegleitungskonzept umfasst sowohl das Sammeln von Informationen im Betrieb des Produkts und deren Auswertung als auch das Ergreifen von Maßnahmen. Es geht also um Beobachtung und Reaktion. In Abgrenzung zur herkömmlichen Produktbeobachtung wurde jedoch der Begriff der „*Produktbegleitung*“ gewählt.
Diese *positive* Produktbeobachtungs- und Reaktionspflicht kann ganz unterschiedlich ausgestaltet sein. Es hängt vom konkreten Produkt und seinen Eigenschaften ab, wie engmaschig der Hersteller Informationen sammeln muss und welche Maßnahmen er ergreifen muss. Damit soll eine hohe Technologieoffenheit gewahrt bleiben. Der Vorschlag hier verzichtet daher auf beispielhafte Aufzählungen von Maßnahmen. Es ist Sache der Hersteller, für ihre Produkte die passenden Produktbegleitungskonzepte auszuarbeiten. Dazu kann auch gehören, dass die Betreiber der Produkte als Teil der zu ergreifenden Maßnahmen informiert werden, wenn das Ergebnis der Überprüfung durch den Hersteller ergibt, dass sicherheitsrelevante Änderungen an dem Produkt auftreten. Die technische Normung kann dann entsprechende Standards für Produktbegleitungskonzepte für bestimmte KI-Produktgruppen setzen.

Produktbeobachtung (z.B. § 6 Abs. 3 ProdSG)	**Produktbegleitungskonzept (PBK)**
Erkennung und Beseitigung von Risiken, die erst nach Inverkehrbringen erkannt werden (*„negative Produktbeobachtung“*).	Voraussetzung für sichere Verwendung veränderbarer und damit nicht abschließend vor Inverkehrbringen beurteilbarer Produkte (*„positive Produktbeobachtung“*).
Unterrichtung der Marktüberwachung über erkannte Risiken zur Vorbereitung hoheitlicher Maßnahmen.	Marktüberwachung muss nicht unterrichtet werden, da PBK unterhalb der Schwelle eines relevanten Risikos ansetzen kann (keine Risikobeherrschung, sondern Beherrschung von Veränderbarkeit). Prüfung des Vollzugs des PBK durch Marktüberwachung (ggf. unter Beteiligung der notifizierten Stellen).
Art und Umfang der Produktbeobachtung vom Risikopotenzial abhängig und wird erst nach Inverkehrbringen im Zeitpunkt des Auftretens des Risikos rechtssicher bestimmbar.	Art und Umfang von Veränderbarkeit und weiteren Eigenschaften des Produkts abhängig (Transparenz, Widerstandsfähigkeit, Kontrollierbarkeit, etc.) und ist vor Inverkehrbringen festzulegen.
Aufzählung im Gesetz exemplarisch, Anlehnung an zivilrechtliche Produktbeobachtungspflicht.	Gesetzliche Pflicht zur Erstellung des PBK bleibt abstrakt, Konkretisierung durch Hersteller bzw. technische Normen (im Sinne des New Legislative Framework). Damit hohe Flexibilität und Technologieoffenheit.
Prüfung durch Marktüberwachung oder im Haftungsfall durch Justiz. Nicht Gegenstand des Konformitätsbewertungsverfahrens.	Prüfung PBK ist Gegenstand des Konformitätsbewertungsverfahrens.
Verwaltungsprivatisierung zur Unterstützung der (Nach-)Marktüberwachung durch Informationsgewinnung und -allokation.	Marktzugangsvoraussetzung zur Gewährleistung eines hohen Sicherheitsniveaus.

Abb. 5.7 „Herkömmliche“ produktsicherheitsrechtliche Produktbeobachtung und Produktbegleitungskonzept im Vergleich.

Das „*Produktbegleitungskonzept*“ ist an den möglichen Risiken auszurichten. Es hängt damit insbesondere auch von den jeweils geschützten Rechtsgütern ab. Wie schon bei der Definition des „wandelbaren“ Produkts können diese Rechtsgüter von Produktgruppe zu Produktgruppe unterschiedlich sein. Wenn das „*Produktbegleitungskonzept*“ im ProdSG geregelt wird, kommt wieder die Auffangfunktion des ProdSG zum Tragen, sodass es insoweit das jeweils geltende Produktsicherheitsrecht ergänzt. Wird das „*Produktbegleitungskonzept*“ dagegen nur für eine Produktgruppe im harmonisierten Bereich geregelt, also vertikal, dann sind nur die dort geschützten Rechtsgüter relevant.

Zu [2]: Grundsätzlich hat der Hersteller bei der Integration der Sicherheit alle Risiken zu adressieren. Lediglich für den nicht bestimmbaren Rest soll bei „wandelbaren“ Produkten die Pflicht bestehen, ein Produktbegleitungskonzept zu erstellen. Nur für die Aspekte des Produkts, die bei der herkömmlichen Risikobeurteilung einer Konformität des Produkts entgegenstehen, ist das Produktbegleitungskonzept zu erstellen. Damit wird dem Hersteller ermöglicht, trotz nicht abschließender Risikobeurteilung ein Produkt legal auf dem Markt bereitzustellen. Ihm soll dagegen nicht die Möglichkeit eingeräumt werden, auch für eigentlich abschließend zu beurteilende Risiken das Produktbegleitungskonzept zu wählen und so „unreife“ Produkte auf den Markt zu bringen. Die Pflicht zur Integration der Sicherheit und der Durchführung einer Risikobeurteilung vor der Bereitstellung des Produkts soll nicht

erodiert werden, indem für bestimme Aspekte eine abschließende Risikobeurteilung obsolet wird, wenn nur ein Produktbegleitungskonzept erstellt werden muss.
Zu [3]: Der Hersteller legt fest, wie lange das Produkt bestimmungsgemäß verwendet werden soll. Für diese Zeit legt er ein Produktbegleitungskonzept vor. Darüberhinausgehende Verwendungen unterliegen nicht mehr der Pflicht des Herstellers zur Integration der Sicherheit.
Damit kann für bestimmte Produkte im Rahmen des Produktbegleitungskonzepts auch vorgesehen werden, dass nach Ablauf der bestimmungsgemäßen Lebensdauer bestimmte Funktionen des Produkts nicht mehr zur Verfügung stehen, wenn von ihnen zu große Risiken ausgehen (könnten). Auch hier gilt wieder, dass es dem Hersteller überlassen ist, das geeignete Produktbegleitungskonzept für das jeweilige Produkt zu erstellen.
Zu [4]: Der Hersteller muss das Produktbegleitungskonzept auch „*umsetzen*“. Damit wird die zeitraumbezogene Pflicht des Herstellers überprüfbar. Die Marktüberwachungsbehörde überprüft nicht nur das Produktbegleitungskonzept selbst, sondern überwacht im Rahmen ihrer Kompetenzen auch dessen Umsetzung. Sie kann Maßnahmen ergreifen, um die Umsetzung zu gewährleisten. Ebenso kann eine notifizierte Stelle, sofern sie in die Fertigungskontrolle eingebunden ist, bereits vor Bereitstellung des Produkts das Produktbegleitungskonzept prüfen und dessen Umsetzung begleiten.
Zu [5]: Mit jeder neuen Information, die der Hersteller über das Produkt gewinnt, kann sich die Risikobeurteilung und damit das Produktbegleitungskonzept ändern. Dementsprechend muss der Hersteller beides anpassen, sobald dies erforderlich erscheint. Das Produktbegleitungskonzept ist also eine dynamische Angelegenheit, ähnlich der Gefährdungsbeurteilung des Arbeitgebers, die gemäß § 3 Abs. 7 BetrSichV auch ständig überprüft und ggf. aktualisiert werden muss.

5.7.3.1.3 Ansatz 3: Zeitraumbezogene Pflicht des Herstellers zur Erhaltung der Sicherheit mit weitem Begriff des „wandelbaren“ Produkts

Die Definition des „wandelbaren“ Produkts soll nach Ansatz 2 nur solche Produkte erfassen, deren Risikobeurteilung wegen der hochgradigen Veränderbarkeit des Produkts nicht abgeschlossen werden kann. Das führt zu der Frage, wie solche Produkte von denjenigen unterschieden werden sollen, die „nur“ wegen hoher Intransparenz oder geringer Kontrollierbarkeit nicht abschließend beurteilt werden können. Die in Ansatz 2 gewählte Definition überlässt es im Sinne der Technologieneutralität und der Grundsätze des Produktsicherheitsrechts den Herstellern und Normungsgremien, die konkreten Maßstäbe hierfür zu entwickeln.
Der Anwendungsbereich des Produktbegleitungskonzepts ist damit bewusst eng gehalten. Die nach dem neuesten Stand der Technik möglichen Verfahren zur Risikobeurteilung müssen ausgeschöpft werden, um auch veränderbare, unkontrollierbare und intransparente Produkte abschließend zu beurteilen und damit dem Zweck des Produktsicherheitsrechts entsprechend ein hohes Sicherheitsniveau zu gewährleisten. Reicht das nach dem neuesten Stand der Technik Mögliche nicht aus, obwohl aber neue Verfahren denkbar sind, so kann das Produkt erst in Verkehr gebracht werden, wenn diese zur Verfügung stehen.
Es ist jedoch denkbar, das Produktbegleitungskonzept auch für solche Produkte zu öffnen, die dem Stand der Technik noch insofern voraus sind, als dass sie derzeit nicht abschließend beurteilt werden können. Es könnten damit insbesondere

unkontrollierbare oder intransparente Produkte unter den Begriff des „wandelbaren“ Produkts fallen.
Die Definition des „wandelbaren“ Produkts könnte hierfür wie folgt lauten:

> *„Im Sinne dieses Gesetzes / dieser Verordnung [….] sind veränderliche Produkte solche Produkte, für die für das Produkt insgesamt oder für spezifische Risiken eine den allgemeinen Sicherheitsanforderungen entsprechende abschließende Risikobeurteilung im Zeitpunkt der Bereitstellung oder dem nach der für das Produkt geltenden Verordnung nach § 8 Abs. 1 ProdSG maßgeblichen Zeitpunkt nicht möglich ist.“*

Diese Variante der Definition wurde um das Tatbestandsmerkmal „*aufgrund der bestimmungsgemäßen Veränderbarkeit*“ gekürzt.
Damit ist der Tatbestand weiter als bei der Definition in Ansatz 2.
Durch die Öffnung des Produktbegleitungskonzepts auch für Produkte, die nach dem neuesten Stand der Technik nicht abschließend beurteilt werden können, wird die regelmäßige Überprüfung der Risikobeurteilung nach Bereitstellung des Produkts bedeutsamer. Wenn sich der Stand der Technik weiterentwickelt, können vormals nicht beurteilbare Risiken des Produkts nunmehr beurteilbar werden. Das Produktbegleitungskonzept ist in diesem Fall entsprechend anzupassen und wird dann ggf. weniger umfangreich.

5.7.3.1.4 Gesamtschau

Die drei Ansätze konkretisieren und erweitern im Ergebnis die Pflichten des Herstellers bei hochgradig veränderbaren, wenig kontrollierbaren und intransparenten Produkten. Sie unterscheiden sich in der Eingriffsintensität in die Grundrechte des Herstellers, also insbesondere die Berufsfreiheit aus Art. 12 Abs. 1 S. 1 GG und die Handlungsfreiheit aus Art. 2 Abs. 1 GG. Denn insbesondere die Ansätze 2 und 3 verpflichten den Hersteller über den Zeitpunkt der Bereitstellung am Markt hinaus dazu, die Sicherheit der „wandelbaren“ Produkte zu gewährleisten. Gleichzeitig eröffnen die Ansätze 2 und 3 jedoch dem Hersteller die Möglichkeit, eine neue Produktkategorie auf den Markt zu bringen.
Außerdem bauen die Ansätze aufeinander auf. Die Definition des „wandelbaren“ Produkts in den Ansätzen 2 und 3 nimmt die Konkretisierung aus Ansatz 1 auf, indem der Hersteller auch für die sichere Ausgestaltung der bestimmungsgemäßen Veränderbarkeit des Produkts im Betrieb verantwortlich ist.
Die Ansätze können daher als Alternativvorschläge gesehen werden. Ihre Auswirkung auf die Pflichten der Verwender, insbesondere von Arbeitgeber und Betreiber einer BImSchG-Anlage, werden weiter unten erörtert.

5.7.3.2 Weiterentwicklung zur Regulierung von Vernetzung und Widerstandsfähigkeit

Die externe sicherheitsrelevante Vernetzung als hohe Ausprägung der Taxonomiedimension der Vernetzung wirft für den Hersteller die Frage auf, wo die sachliche Grenze seines Produkts verläuft.
Ausgangspunkt ist auch hier die **bestimmungsgemäße Verwendung**. Wenn sich das Produkt mit anderen Systemen in handlungsbeeinflussender Weise vernetzen kann, dann muss dies bei der Integration der Sicherheit durch den Hersteller beachtet

werden. Bei einer **unabgesprochenen Vernetzung** kann deshalb wieder ein neues Produkt entstehen bzw. eine (neue) Gesamtheit von Maschinen.

5.7.3.2.1 Anpassung des Produktbegriffs

Damit hier Klarheit für Hersteller und Verwender besteht, wann dies der Fall ist bzw. um zu verhindern, dass bei jeder unabgesprochenen Vernetzung neue Herstellerpflichten entstehen, kann der Produktbegriff wieder entsprechend konkretisiert werden.
Auch hier könnte eine Ergänzung des § 2 Nr. 22 ProdSG erfolgen:

> *„[…]; die Veränderung eines Produkts im Rahmen der bestimmungsgemäßen Veränderbarkeit **sowie die bestimmungsgemäße Vernetzung mit externen Systemen** [1] stellen keinen Fertigungsprozess [2] dar."*

Zu [1]: Der Hersteller kann auch hier bestimmen, wie eine abgesprochene oder unabgesprochene Vernetzung mit externen Systemen nach Inbetriebnahme möglich sein soll. Die Möglichkeit der Vernetzung wird damit ausdrücklich von der Pflicht des Herstellers erfasst, ein sicheres Produkt zu schaffen. Er muss also auch dafür sorgen, dass nur solche Vernetzungen möglich sind, durch die das erforderliche Maß an Sicherheit nicht in Frage gestellt wird.
Zu [2]: Auch hier wird klargestellt, dass kein neues Produkt entsteht. Der Verwender, der über das sich vernetzende Produkt verfügt, muss damit keine Herstellerpflichten beachten, sofern die Vernetzung bestimmungsgemäß erfolgt.
Diese Regelung wirkt auch in die ProdSV, sodass die Frage nach dem Entstehen einer neuen Gesamtheit von Maschinen in der 9. ProdSV damit gelöst werden kann.

5.7.3.2.2 Vernetzung und IT-Sicherheitsanforderungen im Produktsicherheitsrecht

Wenn die Sicherheit des Produkts von äußeren Systemen abhängt, dann handelt es sich im Ergebnis wie gezeigt um eine Frage der rechtlichen Bewertung der Kombination aus hoher Vernetzung und Widerstandsfähigkeit, wobei letztere durch die Vernetzung beeinflusst wird. Die Vernetzung wird zur Frage nach der Widerstandsfähigkeit im Sinne der Security / IT-Sicherheit.
Es geht hier also um die Regulierung der Anforderungen an die IT-Sicherheit in vernetzten Systemen. Hierbei handelt es sich um eine Materie, die vom Produktsicherheitsrecht in seiner jetzigen Form nur bedingt beherrscht werden kann. Denn dieses hat stets das individuelle Produkt im Blick. Hier ist eine **Regulierung der IT-Sicherheit** gefragt. Sie ist bereits angelegt, auf europäischer Ebene beispielsweise in der **Cybersecurity-VO** und der **NIS-Richtlinie** und auf deutscher Ebene im **BSI-G**. Dort können Regelungen z. B. für eine **Zertifizierungspflicht für bestimmte sicherheitsrelevante Anwendungen** erfolgen (dazu mehr bei 5.7.6).
Der Hersteller eines vernetzten Produkts kann dann auf eine solche Zertifizierung der äußeren Systeme verweisen, sie in die Risikobeurteilung einstellen und eine bestimmungsgemäße Vernetzung nur bei Nachweis entsprechender Zertifizierung des äußeren Systems vorsehen.
Im Produktsicherheitsrecht selbst kann auch die **IT-Sicherheit** adressiert werden, indem die **allgemeinen Sicherheitsanforderungen** entsprechend erweitert werden. Die Anforderungen können sich an der Ausprägung einzelner Taxonomiedimensionen

orientieren. Bei handlungsbeeinflussender Vernetzung nach außen werden strengere Anforderungen zu stellen sein als bei lediglich informativer Vernetzung. Kann bei dem System die Sicherheit technisch gewährleistet werden, indem z. B. die von außen kommende Information geprüft wird, bevor sie verarbeitet wird, wird an die Qualität der von außen kommenden Informationen ein weniger strenger Maßstab zu stellen sein, als bei Informationen, die das System nicht überprüfen kann. Die Anforderungen sollten zudem für beide Richtungen der Kommunikation zwischen den Systemen aufgestellt werden. Das aktive System sollte nur Daten zur Verfügung stellen, die diesen Anforderungen entsprechen, während das passive System nur solche Daten empfangen und verarbeiten sollte, die den Anforderungen entsprechen.

5.7.3.3 Zusammenfassung der Lösungsansätze im Produktsicherheitsrecht

Im Produktsicherheitsrecht können somit folgende Ansätze in Betracht kommen:

- Anpassung des Produktbegriffs zur Regulierung hochgradig veränderbarer und extern vernetzter Systeme
- Schaffung einer Pflicht zur Erstellung eines Produktbegleitungskonzepts zur Regulierung hochgradig veränderbarer Systeme
- Schaffung von Sicherheitsanforderungen für IT-Sicherheit

5.7.3.4 Regulierungsoptionen Produktsicherheitsrecht

Im Folgenden soll kurz beleuchtet werden, wo die diskutierten Lösungsansätze geregelt werden könnten.
Das Produktsicherheitsrecht ist zu großen Teilen harmonisiert, sodass eine Regelung auf EU-Ebene in Betracht kommt. Voraussetzung ist weiter, dass eine Regulierung den Anforderungen der Subsidiarität und der Verhältnismäßigkeit nach Art. 5 Abs. 1 S. 2 EUV[369] entspricht. Nach Art. 5 Abs. 3 UABs. 1 EUV setzt das Subsidiaritätsprinzip voraus, dass die angestrebte Regelung durch die Mitgliedstaaten nicht in zweckmäßiger Weise getroffen werden kann und von der EU wegen ihrer Wirkung und ihres Umfangs besser verwirklicht werden kann.
Für die hier diskutierten Lösungsansätze kann angenommen werden, dass sie besser auf Ebene der EU verwirklicht werden können. Denn die Gewährleistung eines einheitlich hohen Schutzniveaus, neben der Verwirklichung des Binnenmarktes der Hauptzweck des europäischen Produktsicherheitsrechts, lässt sich bei komplexen Produkten wie den hier untersuchten sinnvoll und nur dann erreichen, wenn in der EU grundlegende Anforderungen einheitlich geregelt sind, um zu verhindern, dass ein *race to the bottom* die Sicherheitsanforderungen sinken lässt.
Innerhalb des Produktsicherheitsrechts stellt sich die Frage, ob eine Regelung der erörterten Lösungsansätze vertikal, also für bestimmte Produktgruppen, oder horizontal, also übergreifend für alle Produktgruppen erfolgen sollte.
Der Anwendungsbereich der aktuell geltenden Maschinen-RL zum Beispiel erstreckt sich gem. Art. 1 Abs. 2lit. e) und lit. k) Maschinen-RL nicht auf Kraftfahrzeuge sowie Haushalts- und übliche Bürogeräte. Auch in diesen Bereichen dürften KI-Systeme aber in den kommenden Jahren eine erhebliche Rolle spielen. Eine solche horizontale Regelung findet sich etwa in Art. 5 Abs. 1 ProdS-RL (umgesetzt durch § 6 ProdSG), der für sämtliche Verbraucherprodukte, unabhängig davon, ob sie sonst in einen vertikal harmonisierten Bereich fallen oder nicht, bestimmte Transparenz- und Produktbeobachtungspflichten normiert. Insofern erscheint eine horizontale

[369] Vertrag über die Europäische Union (EUV), ABl. C 326/13 vom 26.10.2012.

Harmonisierung sinnvoll. Dagegen spricht wiederum, dass Sicherheitsanforderungen bisher der vertikalen Harmonisierung unterliegen und sich in den Anhängen zu den produktspezifischen Richtlinien und Verordnungen finden. Daher sollte für die einzelnen Lösungsansätze unterschieden werden, wo sie zweckmäßigerweise geregelt werden könnten.
Die Lösungsansätze könnten im Produktsicherheitsrecht wie im Folgenden dargestellt umgesetzt werden.

5.7.3.4.1 Anpassung Produktbegriff und Produktbegleitungskonzept

Regelungen zur Anpassung des Produktbegriffs und der Schaffung eines Produktbegleitungskonzepts können für einzelne Produktgruppen im Wege der vertikalen Harmonisierung getroffen werden, bei denen der Einsatz von KI-Komponenten möglich ist.
Statt der so bewirkten vertikalen Harmonisierung könnte eine horizontale Harmonisierung durch eine **KI-Produkte-RL** für alle Produkte mit KI-Komponenten Regelungen schaffen (dies entspricht auch der oben gewählten Darstellung der Lösungsansätze durch neue Regelungen im ProdSG).
Denn die mit der Produktbegleitungspflicht adressierte **Veränderbarkeit, Intransparenz** und **Unkontrollierbarkeit** bestimmter KI-Systeme findet sich nicht nur im Maschinenbereich als typisches Problem. Es kann also anhand der Taxonomie bestimmt werden, welche Systeme einer horizontalen Harmonisierung unterliegen sollen. So kann für Produkte mit weiterlernender KI-Komponente eine Pflicht zur Beobachtung und Aktualisierung statuiert werden. Es können auch Transparenzanforderungen formuliert werden, die für alle Produkte mit KI-Komponenten gelten, wenn es sich um Produkte handelt, bei denen Menschen nur noch passiv involviert sind, bei denen das sicherheitsrelevante „Produktverhalten“ also nicht mehr durch einen Menschen unmittelbar gesteuert wird und bei denen eine technische Gewährleistung der Sicherheit nur bedingt möglich ist.
Gleiches gilt für den Lösungsansatz zur Regulierung einer externen Vernetzung.

5.7.3.4.2 Anforderungen an IT-Sicherheit

Die Anforderungen an die IT-Sicherheit sollten insbesondere die Gefahren abdecken, die durch eine Vernetzung der Produkte nach außen entstehen können. Die nach der Cybersecurity-VO erstellten Schemata und die dort vorgesehene Konformitätsbewertung stellen ein Instrument dar, mit dem solche Anforderungen inhaltlich normiert und überprüft werden können. Es bedarf jedoch der spezialgesetzlichen Pflicht zur Durchführung entsprechender Konformitätsbewertungsverfahren. Eine solche kann auch neben den Pflichten aus der NIS-RL durch den europäischen Gesetzgeber geregelt werden, vgl. Art. 1 Abs. 7 NIS-RL. In den vertikalen Harmonisierungsvorschriften könnte als grundlegende Sicherheitsanforderung an das Produkt also die Pflicht zur Erfüllung bestimmter, in den europäischen Schemata harmonisierter Standards für IT-Sicherheit vorgesehen werden.[370]
Die Anforderungen an die IT-Sicherheit (also Widerstandsfähigkeit) können von Produktgruppe zu Produktgruppe sehr unterschiedlich sein, da die möglichen Schadensfolgen sehr unterschiedlich sein können. Daher erscheint eine vertikale Harmonisierung sinnvoller.

---

[370] *Kommission*, Bericht über die Auswirkungen künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung, COM (2020) 64, final, S. 7 f.

Als grundlegende Sicherheitsanforderung könnte jedoch bei einer handlungsbeeinflussenden und offenen Vernetzung vorgesehen werden, dass sich die Produkte nur mit Systemen vernetzen, die ihrerseits die entsprechenden Sicherheitsanforderungen des einschlägigen Produktsicherheitsrechts und der europäischen Schemata erfüllen (Sicherheitsanforderungen für passive Vernetzung). Zudem kann verlangt werden, dass bei einer Vernetzung, bei der die handlungsbeeinflussenden Informationen von dem zu bewertenden Produkt ausgehen, diese Informationen ihrerseits bestimmten Sicherheitsanforderungen hinsichtlich ihrer Robustheit entsprechen (Sicherheitsanforderungen für aktive Vernetzung).

5.7.4 Weiterentwicklung des Rechts des technischen Arbeitsschutzes

Wie dargestellt, finden sich die Schwierigkeiten des Herstellers mit bestimmten KI-Systemen auch auf Seiten des Arbeitgebers bei Erfüllung seiner Pflichten aus ArbSchG und BetrSichV wieder. Im Gegensatz zum Hersteller trifft den Arbeitgeber jedoch eine dauernde Pflicht zur Gewährleistung der Sicherheit im Betrieb. Zu diesem Zweck bestehen die dargestellten Prüfpflichten und die Pflicht zur Aktualisierung der Gefährdungsbeurteilung. So kann für die oben festgestellten Bedarfe im betrieblichen Arbeitsschutz zwischen Lösungsansätze auf Produktseite und auf Arbeitgeberseite unterschieden werden.

5.7.4.1 Produktsicherheitsrechtliche Ansätze

Wenn die soeben dargestellten Lösungsansätze zur Beherrschung von weiterlernenden Systemen durch den Hersteller umgesetzt werden, kann der Arbeitgeber in der ersten Gefährdungsbeurteilung und bei der sicheren und ergonomischen Gestaltung des Arbeitsplatzes auf die vom Hersteller mitgelieferten Informationen zurückgreifen. Insofern erscheint die historische Aufteilung des technischen Arbeitsschutzes zwischen Hersteller und Arbeitgeber sinnvoll: Schon im Produktsicherheitsrecht finden sich die Sicherheitsanforderungen, die einen gefahrlosen Einsatz des Produkts am Arbeitsplatz ermöglichen sollen.
Gleiches gilt für die Transparenzanforderungen. Die Transparenz des Systems wird bereits durch den Hersteller geschaffen werden müssen.
Auch bei Arbeitsmitteln, die extern vernetzt sind, wird schon der Hersteller nach der hier vorgeschlagenen Lösung dafür sorgen müssen, dass eine Vernetzung nur mit entsprechend sicheren Systemen erfolgt.

5.7.4.2 Arbeitsschutzrechtliche Ansätze

Bei handlungsbeeinflussender Vernetzung, die nicht offen ist, sondern vom Verwender, also hier dem Arbeitgeber, eingerichtet wird, muss die Verantwortung für die IT-Sicherheit jedoch beim Arbeitgeber liegen. Er könnte gesetzlich verpflichtet werden, dafür zu sorgen, dass nur solche externen Daten in sicherheitsrelevanter Weise auf seine Arbeitsmittel einwirken können, die bestimmten Sicherheitsanforderungen entsprechen.
Die Prüfpflichten und die Pflicht zur Aktualisierung der Gefährdungsbeurteilung durch den Arbeitgeber werden hochgradig veränderbaren Systemen nicht gerecht. Es erscheint daher sinnvoll, je nach Grad der Veränderbarkeit enge Prüfintervalle vorzuschreiben.

5.7.4.3 Regulierungsoptionen Recht des technischen Arbeitsschutzes

Auch hier liegt – wie im Produktsicherheitsrecht - eine Harmonisierung durch europäisches Recht vor, sodass eine Regulierung auf Ebene der EU nötig ist. Wie auch beim Produktsicherheitsrecht ist eine unionsrechtliche Regelung besser geeignet im Sinne des Subsidiaritätsprinzips, ein einheitlich hohes Schutzniveau für Arbeitnehmer zu schaffen, als mitgliedstaatliche Regelungen.
Eine Anpassung der Richtlinie 2009/104/EG über Arbeitsmittel erscheint daher sinnvoll. Das betrifft zunächst die Anforderungen an eine handlungsbeeinflussende Vernetzung des Arbeitsmittels mit anderen Systemen. Hier kann wie bereits oben bei den Vorschlägen zu den produktsicherheitsrechtlichen Anforderungen an die IT-Sicherheit an das Konzept der Cybersecurity-VO angeknüpft werden und eine spezialgesetzliche Pflicht zur Vernetzung nur mit solchen Systemen statuiert werden, die entsprechend zertifiziert sind. Es geht hier also um eine passive Vernetzung. Die erforderliche Vertrauenswürdigkeitsstufe kann wiederum an dem Taxonomiedimension der Schadensfolgen des Arbeitsmittels orientiert werden.
Außerdem können die Regelungen zur Gefährdungsbeurteilung (bzw. Gewährleistung der Sicherheit und des Gesundheitsschutzes bei Betrieb der Arbeitsmittel) und der Prüfung von Arbeitsmitteln so angepasst werden, dass eine dauernde Überwachung der weiterlernenden Systeme erforderlich ist.
Alternativ erscheint jedoch auch eine Regelung auf nationaler Ebene über technische Normen möglich. Denn anders als im Produktsicherheitsrecht ist der Arbeitgeber dauerhaft zur Gewährleistung der Sicherheit verpflichtet. Er hat zudem das Arbeitsmittel stets in seiner Verfügungsgewalt, anders als der Hersteller. Um eine angemessene Prüfung von Arbeitsmitteln mit weiterlernenden KI-Komponenten und eine entsprechende Überprüfung und Aktualisierung der Gefährdungsbeurteilung zu bewerkstelligen, können auch Vorgaben in speziellen TRBS zu weiterlernenden Systemen gemacht werden. Diese konkretisieren dann die bereits jetzt bestehenden Pflichten des Arbeitgebers, um sie auf weiterlernende Systeme anwenden zu können

5.7.5 Weiterentwicklung des Immissionsschutzrechts

Im anlagenbezogenen Immissionsschutzrecht der Störfall-VO finden sich ähnliche Bedarfe wie im betrieblichen Arbeitsschutz. Wegen der hohen Sicherheitsrelevanz der dort behandelten Anlagen stellt sich die Frage nach den rechtlichen Anforderungen an die Überwachung der Anlage durch den Betreiber. Hier könnte der Gesetzgeber konkrete Pflichten zur steten Überwachung vorsehen. Es kann auch eine aktive Involviertheit des Menschen (Mensch als sicherheitsgewährender Teil der Prozesskette) vorgesehen werden, um zu verhindern, dass ein veränderbares System ohne menschliche Überwachungsinstanz agieren kann. Bei besonders hochgradig veränderbaren Systemen kann auch ein Verbot in Betracht kommen, wenn in Verbindung mit einer schweren möglichen Schadensfolgen das im Ergebnis möglicherweise bestehende, bei z. B. weiterlernenden Systemen aber nicht mehr mit der erforderlichen Sicherheit voraussehbare Restrisiko nicht tragbar erscheint.
Für Anlagen, die nicht unter die Störfall-VO und das BSI-G fallen, kann der Gesetzgeber Anforderungen an die IT-Sicherheit vorschreiben. Insofern wird auf die Ausführungen oben bei Vernetzung und IT-Sicherheit im Produktsicherheitsrecht verwiesen.
Entsprechendes gilt für die externe Vernetzung mit anderen Systemen.
Zur Möglichkeit der Regulierung gilt Folgendes:

Die Störfall-VO dient der Umsetzung der Seveso-III-RL, sodass eine Anpassung auf EU-Ebene erforderlich wird. Auch hier ist wieder im Sinne des Subsidiaritätsprinzips eine unionsweite Regelung angezeigt.
Dort kann also für solche Anlagen, die hochgradig veränderbare und sicherheitsrelevante KI-Komponenten enthalten, eine Pflicht zur ständigen Überwachung vorgeschrieben werden. Soll ein gesetzliches Verbot von Anlagen mit KI-Komponenten bestimmter Ausprägung erwogen werden, so ist ein strenger Maßstab bei der Verhältnismäßigkeitsprüfung anzulegen.
Die beiden weiteren Lösungsansätze, die Regulierung von Anforderungen an die Widerstandsfähigkeit im Sinne der IT-Sicherheit und an die externe handlungsbeeinflussende Vernetzung, betreffen die Genehmigung von Anlagen und damit einen Bereich, der durch die Industrieemissionsschutz-RL[371] reguliert ist. Entsprechend ist dort eine Anpassung des Rechtsrahmens nötig. Auch hier kann wieder eine verpflichtende Zertifizierung nach europäischen Schemata im Sinne der Cybersecurity-VO vorgesehen werden. Die Vernetzung mit externen Systemen sollte bei handlungsbeeinflussender Vernetzung in sicherheitsrelevanten Bereichen zudem ebenfalls von einer Zertifizierung dieser externen Systeme abhängig gemacht werden. Insofern gelten also die Ausführung zum Produktsicherheitsrecht und dem Recht des betrieblichen Arbeitsschutzes entsprechend (oben 5.7.3.2 und 5.7.4.1).

5.7.6 Neues „Produktsicherheitsrecht für Daten“

Die zunehmende Vernetzung von Systemen könnte durch eine Flankierung des Produktsicherheitsrechts, welches den Hersteller eines individuellen Systems in die Pflicht nimmt, mit einer **Regulierung der Sicherheitsanforderungen an sicherheitsrelevante Daten** durch den Gesetzgeber aufgegriffen werden.
Durch eine **Zertifizierungspflicht bestimmter sicherheitsrelevanter Daten** könnte ein Markt für Daten geschaffen werden, die im Wege der externen Vernetzung zwischen verschiedenen Systemen zirkulieren können. Sicherheitsrelevante Daten können z. B. solche sein, die in einen Trainingsdatensatz eines KI-Systems einfließen, der sicherheitsrelevante Optimierungen ermöglicht. Den Herstellern bzw. Verwendern dieser Systeme wird durch die Zertifizierungspflicht die ausführliche Prüfung der Datenqualität abgenommen, die Prüfung kann sich auf eine Prüfung des Zertifikates beschränken. Ein solcher Ansatz besteht bereits im Verhältnis Hersteller und Arbeitgeber: Der Arbeitgeber kann sich grundsätzlich auf das CE-Kennzeichen und die Angaben in der Konformitätserklärung verlassen.
Dieser Ansatz könnte durch ein neu zu schaffendes „**Produktsicherheitsrecht für Daten“** aufgegriffen werden. Für bestimmte sicherheitsrelevante Datenprodukte können gesetzlich vorgeschrieben und von anerkannten Prüfstellen ausgestellte Zertifikate deren Verkehrsfähigkeit erhöhen und damit eine umfangreiche sicherheitsrelevante externe Vernetzung von Systemen ermöglichen.

[371] Richtlinie 2010/75/EU des Europäischen Parlaments und des Rates vom 24. November 2010 über Industrieemissionen (integrierte Vermeidung und Verminderung der Umweltverschmutzung), ABl. L 334 vom 17.12.2010, S. 17.

Die **Cybersecurity-VO**[372] sieht **Zertifizierungsschemata** vor, die eine solche Zertifizierung ermöglichen. Der Gesetzgeber ist gefordert, hierfür entsprechende Pflichten zur Zertifizierung einzuführen.

5.7.7 IT-Sicherheit nach Cybersecurity-VO

Nach Art. 2 Nr. 12, 13, 14 Cybersecurity-VO sind Produkte, Dienste und Prozesse erfasst, die Teil eines solchen Netz- und Informationssystems sind. Die Cybersecurity-VO betrifft in sachlicher Hinsicht also zunächst Netz- und Informationssysteme, worunter nach Art. 4 Nr. 1 NIS-RL Übertragungssysteme und gegebenenfalls Vermittlungs- und Leitwegeinrichtungen sowie anderweitige Ressourcen zu verstehen sind, die die Übertragung von Signalen über Kabel, Funk, optische oder andere elektromagnetische Einrichtungen ermöglichen, einschließlich Satellitennetze, feste (leitungs- und paketvermittelte, einschließlich Internet) und mobile terrestrische Netze, Stromleitungssysteme, soweit sie zur Signalübertragung genutzt werden, Netze für Hör- und Fernsehfunk sowie Kabelfernsehnetze, unabhängig von der Art der übertragenen Informationen (siehe Art. 2 lit. a) der Richtlinie 2002/21/EG über einen gemeinsamen Rechtsrahmen für elektronische Kommunikationsnetze und -dienste). Weiter gehören dazu Vorrichtungen oder Gruppen miteinander verbundener oder zusammenhängender Vorrichtungen, die einzeln oder zu mehreren auf der Grundlage eines Programms die automatische Verarbeitung digitaler Daten durchführen, oder auch digitale Daten, die von den genannten Übertragungssystemen oder Vorrichtungen zum Zwecke ihres Betriebs, ihrer Nutzung, ihres Schutzes und ihrer Pflege gespeichert, verarbeitet, abgerufen oder übertragen werden.

> *Die Cybersecurity-VO betrifft also höchstens eine Teilmenge der hier untersuchten Systeme, nämlich solche, die Teil eines Netz- und Informationssystems sind. Sie trifft also Sonderregelungen für eine bestimmte Produktgruppe.*

Die am 27.06.2019 in Kraft getretene[373] Cybersecurity-VO regelt einerseits die Aufgaben und Befugnisse der Agentur der Europäischen Union für Cybersicherheit (ENISA). Andererseits – und darauf soll hier der Fokus liegen – regelt sie die Schaffung eines europäischen Zertifizierungsrahmens für Cybersicherheit sowie die Ausarbeitung von europäischen Schemata für Cybersicherheitszertifizierung. Die ENISA soll hier die Rolle eines Informations- und Wissenszentrums in Sachen Cybersicherheit in der EU übernehmen.[374] Diese Schemata sollen nach Maßgabe eines fortlaufenden Arbeitsplans der Kommission nach Art. 47 Abs. 1 Cybersecurity-VO und nach Beauftragung durch die Kommission oder die noch einzurichtende Europäische Gruppe für die Cybersicherheitszertifizierung nach (ECCG) Art. 48 Cybersecurity-VO ausgearbeitet werden.[375] Sie sollen die Grundlage eines EU-weit einheitlichen Zertifizierungssystems für IKT-Produkte, -Dienste und -Prozesse sein. Da

---

[372] Verordnung (EU) 2019/881 des Europäischen Parlaments und des Rates vom 17. April 2019 über die ENISA (Agentur der Europäischen Union für Cybersicherheit) und über die Zertifizierung der Cybersicherheit von Informations- und Kommunikationstechnik und zur Aufhebung der Verordnung (EU) Nr. 526/2013 (Rechtsakt zur Cybersicherheit – Cybersecurity-VO), ABl. L 151 vom 7.6.2019, S. 15.

[373] Die Art. 58, 60, 61, 63, 64 und 65 Cybersecurity-VO, die Regelungen über die Umsetzung der Cybersecurity-VO durch Verwaltung und Justiz der Mitgliedstaaten treffen, gelten nach Art. 69 Abs. 2 Cybersecurtiy-VO erst ab dem 28.06.2021.

[374] Erwägungsgrund 17 der Cybersecurity-VO.

[375] *Kipker/Scholz*, EU Parlament verabschiedet EU Cybersecurity Act, MMR-Aktuell 2019, 414986.

sich die Sicherheitsanforderungen in diesem Bereich bisher allein an mitgliedstaatlichen Regelungen orientieren, liegt insofern ein zersplitterter Markt innerhalb der EU vor. Im Sinne der Cybersicherheit, aber auch um das Vertrauen der Verbraucher in den digitalen Binnenmarkt zu stärken und europäische Anbieter von IKT international wettbewerbsfähiger zu machen, sollen ein EU-weit einheitliches Zertifizierungsverfahren sowie einheitliche Standards in Gestalt der europäischen Schemata für Cybersicherheitszertifizierung geschaffen werden.[376] So soll ein EU-Binnenmarkt für IKT entstehen, auf dem ein hohes Cybersicherheitsniveau herrscht.
Die europäischen Schemata für Cybersicherheitszertifizierung sind nach Art. 2 Nr. 9 Cybersecurity-VO umfassende Pakete von Vorschriften, technischen Anforderungen, Normen und Verfahren, die auf Unionsebene festgelegt werden und die für die Konformitätsbewertung oder Zertifizierung von IKT-Produkten, -Diensten und -Prozessen gelten. Sie werden von der ENISA ausgearbeitet, wobei diese im Laufe des Verfahrens nach Art. 49 Abs. 3 Cybersecurity-VO alle in Betracht kommenden Interessenträger im Wege eines förmlichen, offenen und transparenten Prozesses konsultiert. Sie setzt nach Art. 49 Abs. 4 Cybersecurity-VO zudem für jede Ausarbeitung eine Ad-hoc-Arbeitsgruppe ein, um spezifische Sachkenntnis und Beratung einholen zu können. Außerdem arbeitet die ENISA nach Art. 49 Abs. 5 S. 1 Cybersecurity-VO eng mit der ECCG zusammen und berücksichtigt nach Art. 49 Abs. 6 Cybersecurity-VO deren Stellungnahmen zum jeweils auszuarbeitenden Schema. Die ECCG setzt sich nach Art. 62 Abs. 2 Cybersecurity-VO aus Vertretern der nationalen Behörden für die Cybersicherheitszertifizierung oder Vertretern anderer einschlägiger nationaler Behörden zusammen. Außerdem soll die ENISA bei der Ausarbeitung internationale Normungsgremien konsultieren.[377] Die fertigen Schemata werden durch die Kommission nach Art. 49 Abs. 7 Cybersecurity-VO als Durchführungsrechtsakt erlassen. Es findet zudem nach Art. 49 Abs. 8 und 56 Abs. 3 Cybersecurity-VO eine systematische Bewertung und ggf. Überarbeitung der angenommenen Schemata statt.
Sie sind so zu konzipieren, dass sie mindestens die in Art. 51 Cybersecurity-VO aufgeführten Sicherheitsziele zu verwirklichen helfen. Sie sollen also entsprechende Anforderungen an die zu zertifizierenden IKT-Produkte, -Dienste und -Prozesse stellen. Dazu zählen u. a. auch die Anforderung, dass Sicherheit durch Voreinstellungen und Technikgestaltung gewährleistet wird (Art. 51 lit. i) Cybersecurity-VO) sowie die Anforderung, dass aktuelle Software und Hardware, die keine allgemein bekannten Sicherheitslücken aufweist, bereitgestellt und mit Mechanismen für sichere Updates ausgestattet wird (Art. 51 lit. j) Cybersecurity-VO).
Die Zertifizierung nach den so entwickelten unionsweit einheitlichen Schemata liegt nach Art. 56 Abs. 4 Cybersecurity-VO in der Hand von Konformitätsbewertungsstellen, die gemäß Art. 60 Abs. 1 Cybersecurity-VO durch eine benannte nationale Akkreditierungsstelle akkreditiert und von der Kommission nach Art. 61 Abs. 1 Cybersecurity-VO notifiziert worden sind. Ausnahmsweise darf eine Konformitätsbewertung gemäß Art. 56 Abs. 5 Cybersecurity-VO nur durch eine nationale Behörde durchgeführt werden, wenn das anzuwendende Schema dies vorsieht.
Die Schemata unterscheiden die jeweils zu zertifizierenden IKT-Produkte, -Dienste und -Prozesse gemäß Art. 52 Abs. 1 Cybersecurity-VO nach den Vertrauenswürdigkeitsstufen „hoch“, „mittel“ und „niedrig“. Die Einstufung orientiert sich gemäß Art. 52 Abs. 1 S. 2 Cybersecurity-VO in angemessener Weise an dem mit

[376] Erwägungsgründe 65 f. der Cybersecurity-VO.
[377] Erwägungsgrund 53 der Cybersecurity-VO.

der beabsichtigten Verwendung eines IKT-Produkts, -Dienstes oder -Prozesses verbundenen Risiko im Hinblick auf die Wahrscheinlichkeit und die Auswirkungen eines Sicherheitsvorfalls. Je nach Einstufung sind in formeller Hinsicht im Zertifizierungs- bzw. Konformitätsbewertungsprozess unterschiedlich strenge Prüfungen vorzunehmen und in materieller Hinsicht unterschiedlich strenge Maßstäbe an die Prüfung anzulegen. Die formellen Anforderungen ergeben sich aus Art. 52 Abs. 5, 6 und 7 Cybersecurity-VO, wobei die Schemata auch abweichende Regelungen treffen können. Verkürzt dargestellt entspricht demnach ein Produkt der **Vertrauenswürdigkeitsstufe „niedrig“** solchen Sicherheitsanforderungen, die darauf gerichtet sind, die bekannten grundlegenden Risiken für Sicherheitsvorfälle und Cyberangriffe möglichst gering zu halten. Dabei reicht für die Bewertung die Überprüfung der technischen Dokumentation aus. Dagegen muss bei der **Vertrauenswürdigkeitsstufe „mittel“** Anforderungen entsprochen werden, die darauf gerichtet sind, bekannte Cybersicherheitsrisiken und das Risiko von Cybersicherheitsvorfällen und Cyberangriffen seitens Akteuren mit begrenzten Fähigkeiten und Ressourcen möglichst gering zu halten. Die Bewertung ist dabei in formeller Hinsicht schon strenger. Eine Überprüfung der technischen Dokumentation reicht nicht mehr, es muss das zu zertifizierende Produkt auf bekannte Sicherheitslücken untersucht und die erforderlichen Sicherheitsfunktionen überprüft werden. Für die **Vertrauenswürdigkeitsstufe „hoch“** wiederum sind solche Sicherheitsanforderungen zu stellen, die darauf gerichtet sind, das Risiko von dem neuesten Stand der Technik entsprechenden Cyberangriffen durch Akteure mit umfangreichen Ressourcen möglichst gering zu halten. Die Bewertung muss hier zusätzlich zur Vertrauenswürdigkeitsstufe „mittel“ prüfen, ob die Sicherheitsfunktionen dem neuesten Stand der Technik entsprechen und ordnungsgemäß funktionieren und zudem Penetrationstests durchführen. Die Unterscheidung der Vertrauenswürdigkeitsstufen geht also vom Bedrohungsszenario aus. Je höher die Vertrauenswürdigkeitsstufe, desto geringer muss das Risiko eines erfolgreichen Angriffs auch durch versierte Angreifer sein.

Wie die Anforderungen im Detail jedoch aussehen, ist dem jeweiligen Schema überlassen. Der Art. 54 Cybersecurity-VO gibt lediglich Elemente der Schemata vor, also Regelungsinhalte, die jedes Schema aufweisen muss. Hervorzuheben ist hier, dass jedes Schema gemäß Art. 54 Abs. 1 lit. j) Cybersecurity-VO Vorschriften über Mechanismen zum Nachweis der beständigen Einhaltung der konkreten Sicherheitsanforderungen aufweisen muss. Weiter müssen nach Art. 54 Abs.1 lit. m) Cybersecurity-VO Vorschiften über die Meldung und Behandlung bislang nicht erkannter Cybersicherheitslücken von IKT-Produkten und -Diensten und -Prozessen enthalten sein. Der Verordnungsgeber geht also davon aus, dass die hier geregelten Produkte auch nach der Zertifizierung Änderungen unterliegen können oder sogar müssen und dass auch später bisher noch unerkannte Sicherheitslücken entdeckt werden können.

> *Das kann für KI-Systeme relevant sein, wenn sie als IKT-Produkt oder als Teil dessen nach einem Schema zertifiziert werden sollen. Dann bestehen über einen längeren Zeitraum, also auch noch nach der maschinensicherheitsrechtlich maßgeblichen Inbetriebnahme, Informations-, Update- und Prüf-Pflichten des Herstellers des Produkts.*

In Art. 55 Cybersecurity-VO werden dem Hersteller oder Anbieter des IKT-Produkts, -Dienstes oder -Prozesses zudem **Informationspflichten** auferlegt. Er muss der Öffentlichkeit insbesondere Informationen in Form von Leitlinien und Empfehlungen über die sichere Konfiguration, Installation, Bereitstellung, den sicheren Betrieb sowie

der sicheren Wartung des IKT-Produkts oder -Dienstes zur Verfügung stellen. Außerdem muss er über den Zeitraum der Sicherheitsunterstützung für die Endnutzer, über Kontaktmöglichkeiten zum Hersteller oder Anbieter zur Mitteilung auftretender Sicherheitslücken und über Online-Register zur Veröffentlichung bekanntgewordener Sicherheitslücken informieren. Hier liegt eine **Instruktionspflicht** des Herstellers begründet. Außerdem wird eine Sicherheitsunterstützung, ein „Beschwerdemanagement“ und eine Online-Register über Sicherheitslücken vorausgesetzt. Hierin kann eine gewisse **Produktbeobachtungspflicht** erkannt werden.
Die Zertifizierung nach einem Schema begründet gemäß Art. 56 Abs. 1 Cybersecurity-VO die Vermutung, dass das zertifizierte Produkt den Anforderungen des konkreten Schemas entspricht. Bei der Vertrauenswürdigkeitsstufe „hoch“ erfolgt die Zertifizierung gemäß Art. 56 Abs. 6 Cybersecurity-VO regelmäßig durch eine Behörde. Dagegen kann bei der Vertrauenswürdigkeitsstufe „niedrig“ die Konformitätsbewertung gemäß Art. 53 Abs. 1 Cybersecurity-VO durch den Hersteller oder den Anbieter erfolgen. In Art. 56 Abs. 2 Cybersecurity-VO ist außerdem festgelegt, dass die Zertifizierung freiwillig ist, sofern nicht das Unionsrecht oder das Recht der Mitgliedstaaten etwas anderes vorsieht. Dort wo also eine Zertifizierung nach einem europäischen Schema für die Cybersicherheitszertifizierung in einem anderen Rechtsakt der Union oder im mitgliedstaatlichen Recht vorgeschrieben ist, besteht die Vermutungswirkung des Art. 56 Abs. 1 Cybersecurity-VO. Gemäß Art. 56 Abs. 3 Abs. 1 Cybersecurity-VO soll die Kommission ermitteln, wo eine Pflicht zur Zertifizierung nach Unionsrecht eingeführt werden soll. Die Prüfung des Regulierungsbedarfs durch die Kommission soll nach Art. 56 Abs. 3 UAbs. 3 Cybersecurity-VO vorrangig in den in Anhang II der NIS-RL aufgeführten Bereichen erfolgen.
Im Ergebnis ähnelt das Zertifizierungsverfahren nach der Cybersecurity-VO dem Ansatz im Produktsicherheitsrecht nach dem New Legislative Framework. Es gibt jedoch – anders als in den Produktsicherheitsverordnungen – keine allgemeinen Sicherheitsanforderungen, die durch den Verordnungsgeber vorgeschrieben sind. Die Cybersecurity-VO beschränkt sich auf die Vorgabe allgemeiner Sicherheitsziele und formeller Anforderungen an das Zertifizierungsverfahren. Die Formulierung konkreter Anforderungen obliegt der ENISA, die in den Schemata auch technische Normen in Bezug nehmen kann. Aus der Cybersecurity-VO geht zudem keine Pflicht zur Zertifizierung bestimmte Produkte hervor. Diese kann sich aus anderem Unionsrecht oder dem Recht der Mitgliedstaaten ergeben. Es besteht zudem die Vermutungswirkung des Art. 56 Abs. 1 Cybersecurity-VO, wenn eine Zertifizierung anhand eines europäischen Schemas für die Cybersicherheitszertifizierung erfolgt ist. In diesem Punkt ähnelt das System der Zertifizierung nach Cybersecurity-VO wieder dem europäischen Produktsicherheitsrecht.

5.7.8 Haftungsrecht

Der Bedarf an einer Haftungsregelung, die dem Umstand gerecht wird, dass das Produkt mit hochgradig veränderbarer KI-Komponente im Zeitpunkt des Inverkehrbringens noch Risiken bergen kann, die in der Konstruktion noch nicht erkennbar waren, entspricht den Schwierigkeiten des Herstellers, eine produktsicherheitsrechtlich ordnungsgemäße Risikobeurteilung schon im Zeitpunkt der Inbetriebnahme der Maschine abschließend vornehmen zu können.

Die Gefährdungshaftung des Herstellers nach ProdHaftG könnte daher auch auf Fehler bei der Produktbeobachtung ausgedehnt werden. Damit würden Entwicklungsrisiken vom Hersteller zu tragen sein. Sofern es sich um weiterlernende Systeme handelt, bei denen die Änderungen also nicht verwenderinitiiert sind, kann dies angemessen sein. Wenn die Änderung jedoch vom Zutun des Verwenders abhängt, muss sich dies auch in der Haftung widerspiegeln: Wenn der Verwender zur Verfügung gestellte Updates nicht aufspielt, sollte es auch nur eine eingeschränkte oder gar keine Haftung des Herstellers geben.[378]
Die durch die Veränderbarkeit, Vernetzung und Intransparenz entstehenden Beweisschwierigkeiten legen zudem eine Gefährdungshaftung desjenigen nahe, der ein KI-System verwendet, also die Gefahrenquelle schafft. Wer für die Schaffung einer Gefahrenquelle haftet, wie z. B. nach dem HaftPflG, sieht sich ggf. hohen finanziellen Risiken ausgesetzt. Entsprechend eng sollte der Anwendungsbereich sein.[379]
Eine mögliche Regulierung muss auch hier die Kompetenzverteilung zwischen Mitgliedstaaten und EU beachten.
Eine Regelung zur Haftung des Herstellers muss wegen der durch die ProdHaft-RL bestehenden Vollharmonisierung auch in der ProdHaft-RL erfolgen.

5.7.8.1 Laufende Überarbeitung der ProdHaft-RL

Unter der Europäischen Kommission arbeitet derzeit die **Expert Group on Liabilty and New Technologies** an möglichen Änderungen der ProdHaft-RL. Sie soll den rechtlichen Herausforderungen mit der Softwareprodukten und insbesondere auch KI-Anwendungen gerecht werden können. Dazu wird unter anderem diskutiert, für Hersteller eine **Produktbeobachtungs- und Reaktionspflicht** einzuführen. Solange der Hersteller das Produkt noch aktualisieren muss oder eine Aktualisierung erforderlich gewesen wäre, um das Produkt fehlerfrei zu halten, müsse er verschuldensunabhängig für Schäden haften, die aus der Verletzung dieser Pflicht resultieren.[380]
Dieser Regulierungsvorschlag entspricht dem oben erörterten Lösungsvorschlag. Es ist hier anhand der Taxonomiedimensionen der Veränderbarkeit, der Transparenz, der Kontrollierbarkeit und der Involviertheit des Menschen zu bestimmen, für welche Systeme eine solche verschuldensunabhängige Haftung des Herstellers für die Verletzung von Produktbeobachtungs- und Reaktionspflichten verhältnismäßig wäre. Hochgradig veränderbare, intransparente und schlecht kontrollierbare Systeme sollten hierunter fallen. Dagegen kann eine aktive Involviertheit des Verwenders dergestalt, dass er für die Veränderung des Systems verantwortlich ist, z. B. durch das Aufspielen der Updates oder durch Beeinflussung der Trainingsdaten, dazu führen, dass eine Haftung des Herstellers entsprechend gemindert oder ausgeschlossen wird. Zudem ist zu berücksichtigen, dass eine verschuldensunabhängige Haftung nur von wirtschaftlich potenten Herstellern riskiert werden kann, die sich entsprechende

[378] *Kommission*, Bericht über die Auswirkungen künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung, COM(2020) 64, final, S.18.
[379] Wie sich eine solche Gefährdungshaftung auf die Entwicklung entsprechender Technologien auswirkt, ist eine Frage, die wirtschaftswissenschaftlich zu untersuchen wäre.
[380] *Expert Group on Liability and New Technologies*, Liability for Artificial Intelligence and other emerging digital technologies, (Stand: 2019) S. 42.

Versicherungen leisten können.[381] Der Anwendungsbereich der verschuldensunabhängigen Haftung ist insofern auf besonders sicherheitsrelevante KI-Produkte zu beschränken.
Weitaus schwieriger gestaltet sich die Frage der zweckmäßigen Haftung bei weiterlernenden, handlungsbeeinflussend und ggf. sogar offen vernetzten, wenig kontrollierbaren und intransparenten Systemen. Eine einheitliche Regelung erscheint jedoch wegen der Vielzahl möglicher Anwendungen der hier untersuchten KI-Systeme und der dort jeweils vorherrschenden Umstände wenig zielführend. Sinnvoller erscheint ein produkt- bzw. anwendungsbezogener Ansatz, der für den jeweiligen Lebensbereich eine zweckmäßige Haftungsregelung findet.[382] So gelten heute schon im Arzneimittelrecht eigene Haftungsregelungen, die vom Produkthaftungsrecht abweichen. Der pharmazeutische Unternehmer, der ein Arzneimittel in Verkehr bringt, haftet nach § 84 Abs. 1 S. 1 AMG verschuldensunabhängig für Schäden, die ein Mensch durch das in Verkehr gebrachte Arzneimittel erleidet. Hier wird die Gefährdungshaftung zudem ergänzt durch die Vermutung der Kausalität des Arzneimittels für den Schaden gemäß § 84 Abs. 2 S. 1 AMG. Ebenso zeichnet sich das Straßenverkehrsrecht durch ein eigenes Haftungsregime aus.
Insofern sollten Überlegungen zu speziellen Haftungsregelungen von Verwendern der hier untersuchten KI-Systemen, die über das verschuldensabhängige Deliktsrecht in den Mitgliedstaaten hinausgeht, dort angestellt werden, wo konkrete Regelungen für bestimmte Produkte erörtert werden. Es kann zudem bei Einführung von Gefährdungshaftungstatbeständen eine entsprechende Versicherungspflicht erörtert werden.

5.7.9 Gesamtschau der Lösungsansätze und ihr Verhältnis untereinander

Die hier diskutierten Regulierungsoptionen im **Ordnungsrecht** betreffen also zunächst das Produktsicherheitsrecht. Dort könnte für neu zu definierende „wandelbare“ Produkte ein Produktbegleitungskonzept eingeführt werden. Die Einführung von gesetzlichen Sicherheitsanforderungen für die Widerstandsfähigkeit des KI-Systems betrifft ebenfalls zunächst den Hersteller. Außerdem könnte für die externe unabgesprochene Vernetzung klar geregelt werden, wo die Grenzen des konkreten Produkts verlaufen, also für welches „Gesamtprodukt“ der Hersteller ein Konformitätsbewertungsverfahren durchführen muss.

---

[381] *Wissenschaftliche Dienste des Bundestages*, Gefährdungshaftung für Emittenten und verfassungsrechtliche Aspekte eines „Finanz-TÜV“, WD 4 – 3000 – 058/16, WD 7 – 3000 – 080/16, S. 9.
[382] *Lohmann*, Ein europäisches Roboterrecht – überfällig oder überflüssig?, ZRP 2017, 169, 171.

		Hersteller	Marktüberwachung	Verwender	Arbeitgeber	Anlagenbetreiber
Herkömmliches Produkt	Vor Inverkehrbringen / Inbetriebnahme	• Risikobeurteilung • Konformitätsbewertung • CE-Kennzeichnung	• Einfuhrkontrolle	/	• Informationsbeschaffung • Gefährdungsbeurteilung • ggf. Prüfung vor Inbetriebnahme	• Genehmigung
	Nach Inverkehrbringen / Inbetriebnahme	• Produktbeobachtung • Notwendige Maßnahmen • Unterrichtung der Marktüberwachung	• Sachverhaltsermittlung • Bei Gesetzesverstoß: Erforderliche Maßnahmen	• *Bestimmungs-gemäße Verwendung* • *ggf Beschwerde* • ggf. Herstellerpflichten	• Prüfung Arbeitsmittel • Überprüfung Gefährdungsbeurteilung	• Betrieb gemäß Genehmigung • Anzeige von Änderung • ggf. Änderungsgenehmigung
wandelbares Produkt	Vor Inverkehrbringen / Inbetriebnahme	• Risikobeurteilung • **Produktbegleitungs-konzept (PBK)** • Konformitätsbewertung • CE-Kennzeichnung	s.o.	/	• s.o. • **Gefährdungsbeurteilung berücksichtigt PBK**	• s.o. • **Genehmigung berücksichtigt PBK**
	Nach Inverkehrbringen / Inbetriebnahme	• **Umsetzung PBK** • Notwendige Maßnahmen zur Risikobeherrschung • Unterrichtung der Marktüberwachung	• s.o. • *Verstoß gegen. PBK ist Gesetzesverstoß*	• s.o. • keine Herstellerpflichten bei bestimmungsge-mäßer Änderung	• s.o. • keine neue Überprüfung Gefährdungsbeurteilung / Prüfung Arbeitsmittel bei bestimmungsge-mäßer Veränderung	• s.o. • keine Änderungsgenehmigung bei bestimmungsgemäßer Veränderung

Abb. 5.8 Die Verantwortlichkeiten der Beteiligten und ihre Zusammenhänge bei herkömmlichen Produkten und den neu zu definierenden „wandelbaren“ Produkten im Überblick.

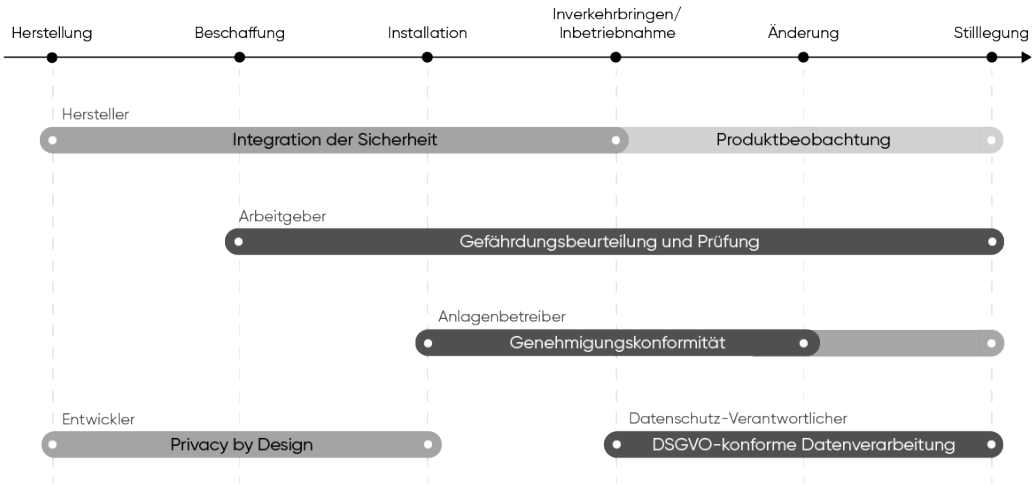


Abb. 5.9 Ordnungsrechtliche Verantwortlichkeiten über die Lebensdauer des Produkts im Überblick. Die Pflichten des Herstellers werden ergänzt um eine Produktbeobachtungs- und Reaktionspflicht nach dem Produktbegleitungskonzept, sodass er auch über den Zeitraum der Inbetriebnahme für die Sicherheit des KI-Produkts verantwortlich bleibt.

Bei Einführung des Produktbegleitungskonzepts stellen sich die Verantwortlichkeiten im Vergleich zur bisherigen Rechtslage wie folgt dar:

Die Neuregelungen im Produktsicherheitsrecht haben Einfluss auf den Pflichtenkatalog des Arbeitgebers und des Anlagenbetreibers, da sie alle sich grundsätzlich auf die Konformitätsbewertung und damit bei „wandelbaren“ Produkten auf das Produktbegleitungskonzept des Herstellers berufen können.
Das **Haftungsrecht** kann für die Haftung des Herstellers für weiterlernende Systeme um eine verschuldensunabhängige Herstellerhaftung auch nach Inverkehrbringen des Produkts ergänzt werden. Für besonders sicherheitsrelevante Anwendungen weiterlernender, intransparenter, komplexer und vernetzter Systeme kann zudem eine Gefährdungshaftung der Verwender der Technik kommen.
Zwischen dem präventiven Technikrecht und dem repressiven Haftungsrecht wird eine Balance zu finden sein, die bei Gewährleistung eines hohen Maßes an Sicherheit zugleich innovationsfördernd ist.
Rechtlicher Maßstab ist hierbei der **Verhältnismäßigkeitsgrundsatz**. Denn jede Regelung stellt, wie gezeigt, einen Eingriff in die Grundrechte der Regelungsadressaten dar.
Hinsichtlich der Verhältnismäßigkeit der Regelungen kann hier abschließend keine Aussage getroffen werden, da dies nur möglich ist, wenn sie ausformuliert und im Zusammenhang mit anderen Regelungen untersucht werden können. Zur Bestimmung der **Verhältnismäßigkeit** ist zudem die Taxonomiedimension der **Schadensfolgen** von zentraler Bedeutung. Je größer die durch die Regulierung

abzuwendende Gefahr, desto strenger kann die Regulierung ausgestaltet werden. Das gilt auch in den anderen zu untersuchenden Rechtsbereichen.[383]

5.7.10 Ausblick

Neben dem bereits angesprochenen Verfahren zur Überarbeitung der ProdHaft-RL, um sie auf die Herausforderungen mit KI-Systemen einzustellen, gibt es in anderen Bereichen ebenfalls Regelungsaktivitäten, von denen zwei hier kurz dargestellt werden sollen.
Im Bereich der Maschinen-RL gibt es bereits ein Verfahren zur Überarbeitung des Produktsicherheitsrechts. Zudem wird im Fahrzeugbereich auf Ebene der United Nations Economic Commission for Europe (UNECE) an Regelungen zur Ermöglichung von automatisierten Fahrfunktionen gearbeitet.

5.7.10.1 Roadmap zur Überarbeitung der Maschinen-RL

Die Europäische Kommission hat zur Änderung der Maschinen-RL eine Roadmap erstellt, die Gegenstand eines Konsultationsverfahrens war. Dort konnte die interessierte Öffentlichkeit Stellungnahmen zur vorgestellten Roadmap abgeben. Nach dem Zeitplan der Kommission ist die nächste Stufe des Überarbeitungsverfahrens, die Vorstellung eines Kommissionsentwurfs zur erneuten öffentlichen Konsultation, für die zweiten Jahreshälfte 2021 geplant.[384]
Die Roadmap nennt folgende Änderungsoptionen:

1. Keine Änderung;
2. Anpassung der Maschinen-RL an den „New Legislative Framework“ ohne Änderungen;
3. Anpassung der Maschinen-RL an den „New Legislative Framework“ mit Änderungen:
    - im Anwendungsbereich (insbesondere in der Liste der vom Anwendungsbereich ausgenommenen Niederspannungsprodukte und der Überarbeitung der Definition von unvollständigen Maschinen);
    - in den grundlegenden Anforderungen (insbesondere bezüglich digitaler Dokumentation sowie Erfordernissen aus neuen Technologien, z. B. IT-Sicherheit, Internet of Things und KI, insbesondere in Bezug auf autonome Roboter);
4. Inhaltliche Änderungen wie bei 3., ohne Anpassung an den „New Legislative Framework“;
5. Überführung der Maschinen-RL in eine Maschinenverordnung, in Kombination mit einer der oberen Optionen.

---

[383] Diesen Ansatz betont die *Kommission* in ihrem Weißbuch zur künstliche Intelligenz, 5. C, wo sie die Eingriffsintensität der Regulierung davon abhängig macht, ob die KI in einem Sektor eingesetzt, in dem mit hohen Risiken zu rechnen ist und auch so eingesetzt wird, dass mit hohen Risiken zu rechnen ist.
[384] Zum Stand und zur Roadmap siehe https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/2019-Revision-of-the-Machinery-Directive (zuletzt abgerufen am 22.06.2020).

Im Rahmen dieser Überarbeitung werden die Weichen gestellt für den Einsatz von KI-Systemen im Maschinenbereich. Damit wird ein wesentlicher Produktbereich für KI-Systeme erschlossen, dem sicher weitere Produktbereiche folgen werden[385].

5.7.10.2 UNECE-Regeln für Cybersecurity und Software-Updates

Im Automobilbereich hat die UNECE auf Grundlage des Genfer Übereinkommens für Fahrzeugteile von 1958[386] zwei neue Regelungen für die Realisierung hochautomatisierter Fahrzeuge erlassen, die Cybersecurity[387] und die Durchführung von Software-Updates[388] betreffen.
Demnach haben die Autohersteller zur Gewährleistung eines hohen Maßes an Cybersecurity für die Lebensdauer des Fahrzeugs ein *Cybersecurity Management System (CSMS)* einzuführen und umzusetzen. Um Software-Updates aufspielen zu können, ohne dass ein Nachtragverfahren zum Typengenehmigungsverfahren erforderlich wird, ist zudem ein *Software Update Management System (SUMS)* einzuführen und umzusetzen.
Beides wird durch die nationalen Genehmigungsbehörden oder beauftragten Prüfstellen auditiert und wird Gegenstand der Typengenehmigung.
Damit wird ein ähnlicher Weg beschritten, wie hier er hier für das Produktsicherheitsrecht vorgeschlagen wird: Der Hersteller wird über den Lebenszyklus des Fahrzeugs für die Cybersecurity verantwortlich und muss dafür ein Managementsystem einführen, was die Beobachtung des Fahrzeugs im Betrieb umfasst. Zudem wird die Möglichkeit eingeräumt, auch nach Inverkehrbringen Software-Updates vorzunehmen, sofern das Verfahren hierzu dem Stand der Technik entspricht.

5.7.10.3 Verlagerung verhaltensbezogener Regelungen ins Technikrecht

Die hier vorgeschlagenen Lösungsansätze bewirken eine Verlagerung der Pflichten zur Gewährleistung eines hohen Sicherheitsniveaus bei Verwendung von KI-Systemen vom Kreis der Verwender hin zum Hersteller.
Sofern dabei vormals verhaltensbezogene Regelungen, die sich an die Verwender richteten, obsolet werden, weil das dort geregelte Verhalten nicht mehr für eine sichere Verwendung des KI-Systems relevant ist, werden die **technischen Normen zur sicheren Gestaltung des KI-Systems** wichtiger. War es vorher noch ein Mensch, der sicherheitsrelevante Entscheidungen traf und dessen Entscheidung man entsprechend dem gewünschten Sicherheitsniveau regulierte, so ist diese Entscheidung bei Rückgang der Kontrollierbarkeit und der aktiven Involviertheit des Menschen in den Herstellungs- bzw. Programmierungsprozess vorverlagert.[389]
Dieser Aspekt soll hier auch mit der Ausweitung der Herstellerpflichten über den gesamten Lebenszyklus des Produkts adressiert werden.

---

[385] Die Fortsetzung des Verfahrens ist für die zweite Jahreshälfte 2021 geplant.
[386] Agreement Concerning the Adoption of Harmonized Technical United Nations Regulations for Wheeled Vehicles, Equipment and Parts which can be Fitted and/or be Used on Wheeled Vehicles and the Conditions for Reciprocal Recognition of Approvals Granted on the Basis of these United Nations Regulations (Revision 3).
[387] UNECE-R155, ECE/TRANS/WP.29/2020/79 Revised (konsolidierte Fassung).
[388] UNECE-R156, ECE/TRANS/WP.29/2020/80 (Beschlussvorlage mit Regelungstext).
[389] *Europäische Kommission* COM(2020) 65 final, Weissbuch – Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, S. 11

Auch hier werden technische Normen den Herstellern die Ausgestaltung ihres Produktbegleitungskonzepts und der Integration der Sicherheit bei zunehmend veränderbaren Systemen erleichtern.
Bei der Ausfüllung der „Regelungslücke“, die das Verhaltensrecht hinterlässt, sollte jedoch beachtet werden, dass es nunmehr die Maschine ist, die die vormals durch das Verhaltensrecht angestrebte Sicherheit gewährleisten muss. Eine schlichte Übersetzung des Verhaltensrecht in technische Normen oder gar die Forderung, die Maschine müsse schlicht das Verhaltensrecht einhalten, kann nicht zielführend sein. Es bedarf vielmehr einer genauen Prüfung, welche Ziele des vormals relevanten Verhaltensrechts durch die Maschine zu erreichen sind und wie die Gestaltung der Maschine an diesen Zielen auszurichten ist. Insofern sollte keine „Vermenschlichung“ der Maschine durch Übertragung menschenbezogener Verhaltenspflichten auf die Maschine erfolgen. Stattdessen sollte die Gewährleistung eines hohen Sicherheitsniveaus durch technische Normen bewerkstelligt werden, die die Vor- und Nachteile der jeweiligen Maschine berücksichtigen. Dabei können auch weitere Aspekte in die Konstruktion der Maschine fließen, wie ethische Überlegungen oder der Schutz vor Diskriminierungen.[390]

5.8 Ergebnis

Die Untersuchung der Taxonomie hat ergeben, dass hochgradig veränderbare, intransparente, reduziert kontrollierbare und extern vernetzte Systeme (Klasse 4a oder 5a) mit rechtlichen Problemen verbunden sein können. Diese fangen auf Seiten des Herstellers an und setzen sich konsequent bei den Verwendern fort.
Zusammenfassend gilt für alle ordnungsrechtlich Verantwortlichen, dass bei einer hohen Veränderbarkeit die Bestimmung der erforderlichen Sicherheitsmaßnahmen nicht abschließend erfolgen kann. Bei Eintritt einer sicherheitsrelevanten Veränderung kann zudem eine neue Beurteilung der Risiken erforderlich werden. Da bei KI-Produkten, die als weiterlernend bezeichnet werden können, solche Veränderungen produktimmanent sind, erscheint es fragwürdig, wenn die Verwender dadurch zu Herstellern werden können. Bei entsprechend hoher Intransparenz ist die durch sie durchzuführende Gefährdungsbeurteilung bzw. Risikoermittlung nicht abschließend möglich. Bei hochgradig nach außen vernetzten Produkten ist wiederum unklar, ob und wie weit die für die Teilprodukte durchzuführende Risikobeurteilung andere Teilprodukte erfassen soll. Auch bestehen im Produktsicherheitsrecht keine Regelungen zur Widerstandsfähigkeit bei hochgradiger Vernetzung.
All diese Probleme setzen sich im Haftungsrecht fort. Mangels ordnungsrechtlicher Vorgaben für die hier untersuchten Systeme bleibt unklar, wie die Verkehrssicherungspflichten der Hersteller und Verwender ausgestaltet sind. Hier kann also eine hohe Rechtsunsicherheit bestehen. Eine hohe Veränderbarkeit wird zudem nicht durch das streng auf den Zeitpunkt des Inverkehrbringens fokussierte ProdHaftG angemessen berücksichtigt.
Die diskutierten Lösungsansätze im Ordnungsrecht betreffen vor allem das Produktsicherheitsrecht. Denn es ist der Hersteller, der durch die Gestaltung und

[390] Vgl. dazu ausführlicher *Remmers/Hartwig/Stegmaier* in: BAuA - Autonome Roboter für Assistenzfunktionen: Interaktive Grundfertigkeiten – Ergebnisse und Forschungsperspektiven des Förderprogramms ARA1, S. 28 ff.

Konstruktion des Produkts bestimmt, wie die einzelnen Taxonomiedimensionen in seinem Produkt ausgeprägt sind und über die notwendigen Informationen zur Risikobeurteilung des Produkts verfügt. Dementsprechend setzen auch die **Anpassung des Produktbegriffs** und die **Einführung eines Produktbeobachtungskonzepts** bei den Herstellerpflichten an.

5.8.1 Produktsicherheitsrechtlicher Ansatz 1 – Kein neues Produkt bei bestimmungsgemäßer Veränderbarkeit

Für die bestimmungsgemäße Veränderbarkeit kann ausdrücklich angeordnet werden, dass eine dementsprechend eintretende Veränderung nicht zu einem neuen Produkt führt. Dadurch wird für die Verwender des Produkts vermieden, dass sie unerwartet Herstellerpflichten treffen können.

5.8.2 Produktsicherheitsrechtlicher Ansatz 2 – Produktbegleitung für „wandelbare“ Produkte im engeren Sinne

Fertigt der Hersteller ein hochgradig veränderbares Produkt, für das er keine abschließende Risikobeurteilung anstellen kann, so sollte er auch für die dauernde Aktualisierung der Risikobeurteilung durch eine Produktbegleitung zuständig sein.
Im besonders sicherheitsrelevanten Bereich des Arbeitsschutzes und der Anlagensicherheit treten daneben der Arbeitgeber und der Anlagenbetreiber als Verantwortliche auf. Sie können sich, wie bisher, auf die Konformitätsbewertung durch den Hersteller verlassen, sofern ihnen aufgrund der konkreten Umstände im Betrieb bzw. der Anlage andere Informationen vorliegen, die zu differenzierten Gefährdungsbeurteilungen bzw. Risikoermittlungen führen. Wenn der Hersteller also mit seinem Produktbegleitungskonzept die Veränderbarkeit, Intransparenz und möglicherweise reduzierte Kontrollierbarkeit des Produkts beherrschen kann, sind sie insofern entlastet.
Diese Möglichkeit soll nur für solche Produkte bestehen, denen eine Unbestimmbarkeit ihrer künftigen Zustände derart immanent ist, dass auch mit neuen Simulationsverfahren keine hinreichende Risikobeurteilung denkbar ist.

5.8.3 Produktsicherheitsrechtlicher Ansatz 3 – Produktbegleitung für „wandelbare“ Produkte im weiteren Sinne

Hier sollen auch solche Produkte erfasst sein, bei denen mit den nach dem neuesten Stand der Technik verfügbaren Verfahren eine Risikobeurteilung nicht möglich ist, aber in Zukunft zu erwarten ist, dass geeignete Verfahren vorliegen.
Dieser Ansatz ermöglicht eine breitere Anwendung des Produktbegleitungskonzepts. Neuartige Produkte können somit in Verkehr kommen, die erst mit der Zeit und durch ein ständiges Dazulernen sowohl der Hersteller als auch der Prüfinstitutionen und der Marktüberwachung abschließend beurteilt werden können.

5.8.4 Ansatz zur Ermöglichung externer Vernetzung bei hoher Widerstandsfähigkeit, „Produktsicherheitsrecht für Daten“

Durch die Anerkennung der bestimmungsgemäßen Vernetzung im Produktbegriff des Produktsicherheitsrechts wird dem Hersteller die Möglichkeit gegeben, die Grenzen

seines Produkts hinsichtlich der Vernetzung zu bestimmen und damit seine Risikobeurteilung einzugrenzen.
Da die Widerstandsfähigkeit bei vernetzten Produkten von den externen Systemen abhängt, könnte durch die Einführung einer Zertifizierungspflicht für sicherheitsrelevante externe Daten hinsichtlich Qualität, Verfügbarkeit, Validität und weiteren Aspekten die dahingehende Risikobeurteilung für den Hersteller erleichtert werden.
Das so entstehende „Produktsicherheitsrecht für Daten“ kann sich auf alle Rechtsbereiche auswirken, bei denen der jeweils für die Sicherheit des extern vernetzten Systems auf entsprechende Informationen über die von außen kommenden sicherheitsrelevanten Daten angewiesen ist.
Ein regulativer Ansatz findet sich in der Cybersecurity-VO, deren Zertifizierungsschemata zur Anwendung kommen könnten.

5.8.5 Produktbeobachtungspflicht im Produkthaftungsrecht

Da sich die Veränderbarkeit von Systemen erst im Betrieb zeigt, ist der zeitpunktbezogenen Ansatz des Produkthaftungsrechts in ProdHaft-RL und ProdHaftG nicht mehr ausreichend. Dem hier diskutierten Ansatz des Produktbegleitungskonzepts folgend, kann auch hier eine Produktbeobachtungspflicht des Herstellers eingeführt werden. Eine besondere Gefährdungshaftung auch für den Verwender eines hochgradig veränderbaren und wenig kontrollierbaren KI-Systems kann erwogen werden.

6 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 dem direkten menschlichen 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“ (ePerson oder E-Person) als neue Rechtspersönlichkeit. Obwohl schon seit über 20 Jahre diskutiert[391], ist die Diskussion darüber im Zuge des technologischen Fortschritts – insbesondere bei mobilen Robotern oder hochautomatisierten Fahrzeugen – in den öffentlichen Fokus gekommen.

6.1 Einführung

Autonome und KI-Systeme agieren auf Grundlage komplexer Programme und teilweise sehr großer Datenmengen. Die Programme können zudem im Betrieb veränderbar sein, indem sie dem System z. B. erlauben, die Entscheidungsparameter anzupassen. Ihr Agieren kann damit selbst für den Programmierer im Einzelfall unvorhersehbar werden – das System kann intransparent und unkontrollierbar werden und sein Agieren Herstellern, Betreibern und Nutzern willkürlich erscheinen. Diese scheinbare Willkür drängt den Schluss auf, dass das System tatsächlich einen eigenen Willen entwickelt und zum Ausdruck bringt. Aus dieser Schlussfolgerung lassen sich unterschiedliche rechtliche Detailfragen ableiten (z. B. zur Zulassung von autonomen Fahrzeugen), es können aber auch Fragen zur Rechtssystematik und zu Grundannahmen unseres Normengefüges abgeleitet werden, so etwa zur Haftung bei Schäden, die solch ein intransparentes und unkontrollierbares Produkt mit KI-Anteilen verursacht. Dessen scheinbare Willkür des Agierens kann wiederum den Schluss nahelegen, dieses selbst könne zweckmäßiger Adressat von Haftungsregelungen sein.[392]

Für die Haftung von Robotern hat die EU bereits Erwägungen angestellt. Das EU-Parlament erklärte mit seiner Entschließung vom 16.02.2017[393], dass in den Fällen, in denen ein Roboter eigenständig Entscheidungen treffen kann, die herkömmlichen Regeln nicht mehr ausreichen würden, um eine Haftung für verursachte Schäden auszulösen. Sie seien nicht ausreichend, um den Beteiligten zu ermitteln, der für den Ausgleich des Schadens verantwortlich ist, sowie diesem Beteiligten dann vorzuschreiben, den von ihm verursachten Schaden zu ersetzen. Daher sei wegen der Autonomie der Roboter, vor dem Hintergrund der bestehenden rechtlichen Kategorien, ihre Rechtsnatur nicht eindeutig zu klären. Das EU-Parlament regte deshalb an, dass ein spezieller rechtlicher Status für autonome Roboter als elektronische Person

[391] Schweighofer, E., Menzel, T., & Kreuzbauer, H. M. (2001). Auf dem Weg zur ePerson: Aktuelle Fragestellungen der Rechtsinformatik, Schriftenreihe: Schweighofer, Erich/Lachmayer, Friedrich (Hrsg der Schriftenreihe): Schriftenreihe Rechtsinformatik.

[392] Vgl. *Borges*, Rechtliche Rahmenbedingungen für autonome Systeme, NJW 2018, 977.

[393] Entschließung des Europäischen Parlaments vom 16.2.2017 mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik.

geschaffen werden sollte, womit sie für alle von ihnen verursachten Schäden verantwortlich zu machen wären.[394] Der autonome Roboter oder vielleicht auch autonom agierende oder entscheidende „KI-Systeme“ sollen also nicht nur als Rechtsobjekte Gegenstände rechtlicher Zuordnung werden, sondern auch rechtsfähig, d. h. als Rechtssubjekte Träger von Rechten und Pflichten sein.
Flankierend zur Schaffung eines neuen Rechtsubjekts sollte ein Versicherungssystem für Robotik eingeführt werden, durch das der Hersteller, Eigentümer oder Benutzer des Roboters verpflichtet wird, für jeden autonomen Roboter eine Versicherung abzuschließen. Es sollen alle potenziellen Verantwortlichen in der Kette abgedeckt werden und eben nicht nur menschliche Handlungen und Unterlassungen. Dieses Versicherungssystem könnte durch einen Fond ergänzt werden.[395]
Dieser Ansatz kann letztlich als Ausdruck und zugleich als Zuspitzung der eingangs erwähnten rechtlichen Fragestellungen gedeutet werden, läuft er doch auf eine umfassende Reform hinaus, bei der neben die natürliche und die juristische Person eine neue Art von Rechtssubjekten gestellt wird.[396] Er stellt damit den weitreichendsten Reformansatz dar.[397]
Zugleich kommt hierin eine Vermenschlichungstendenz[398] zum Ausdruck, bei der davon ausgegangen wird, dass an eine Maschine grundsätzlich dieselben Sicherheits- und „Verhaltensanforderungen“ zu stellen seien wie an einen Menschen oder ihr sogar menschliche Attribute zugeschrieben werden.[399]

6.2 Haftungsrechtliche Problemstellung

Rechtlicher Hintergrund dieses Ansatzes ist die Annahme, dass mit der Entwicklung von KI-Systemen das bestehende Haftungsrecht an seine Grenzen stößt. Für den Geschädigten liegt bei Schädigung durch ein KI-System eine schwierige Beweislage vor. Während ihm gegenüber dem Hersteller, im Rahmen der Produzentenhaftung, von der Rechtsprechung entwickelte Beweiserleichterungen beim Nachweis der Schuld helfen, trägt er gegenüber dem Betreiber eines KI-Systems grundsätzlich die volle Darlegungs- und Beweislast. Gleichzeitig trägt der Betreiber eines KI-Systems grundsätzlich das Haftungsrisiko für Schäden, die das KI-System verursacht. Denn der Hersteller mag potenziell neben ihm haften, bei einem hochgradig veränderbaren System kann der Nachweis eines Produktfehlers jedoch schwierig sein. Dieser ist jedoch Voraussetzung für eine Haftung nach Produkthaftungsgesetz. Im Rahmen der Produzentenhaftung nach BGB ist zudem unklar, wie die Sorgfaltspflichten des

---

[394] Entschließung des EU-Parlaments vom 16.2.2017 (2015/2103 [INL]) – P8_TA(2017)0051, Ziff. 59 f.
[395] Entschließung des EU-Parlaments vom 16.2.2017 (2015/2103 [INL]) – P8_TA(2017)0051, Ziff. 57, 58.
[396] *Borges*, Rechtliche Rahmenbedingungen für autonome Systeme, NJW 2018, 977, 979.
[397] *Eichelberger* in: Ebers/Heinzle/Krügel/Steinrötter, Künstliche Intelligenz und Robotik, § 5 Rn. 71.
[398] Vgl. dazu *Remmer/Hartwig/Stegmaier* in: Bundesanstalt für Arbeitsschutz- und Arbeitsmedizin, Autonome Roboter für Assistenzfunktionen – Interaktive Grundfertigkeiten - Ergebnisse und Forschungsperspektiven des Förderprogramms ARA1, S. 29.
[399] Z.B. wenn das EU-Parlament davon ausgeht, dass „die Menschheit mittlerweile an der Schwelle einer Ära steht, in der immer ausgeklügeltere Roboter, Bots, Androiden und sonstige Manifestationen Künstlicher Intelligenz („KI“) anscheinend nur darauf warten, eine neue industrielle Revolution zu entfesseln“, Entschließung des EU-Parlaments vom 16.2.2017 (2015/2103 [INL]) – P8_TA(2017)0051, lit. B.

Herstellers ausgestaltet sind. Bei unvorhersehbarem Verhalten eines hochgradig veränderbaren KI-Systems kann es dabei unbillig erscheinen, dem Betreiber die Haftung aufzuerlegen.[400]
Doch auch der Rückgriff auf den Betreiber kann für den Geschädigten schwierig sein. Grundsätzlich drohen Verantwortungslücken im Deliktsrecht gem. §§ 823 ff. BGB und dem dort verankerten Verschuldensprinzip, wenn man von der Annahme ausgeht, dass einem Betreiber eines autonomen KI-Systems, welches eigenständige Entscheidungen trifft, ein unvorhersehbares Fehlverhalten der KI nicht mehr als Verletzung seiner eigenen Verkehrssicherungspflicht angerechnet werden kann.[401]

6.3 ePerson als Lösung?

Als Lösung für dieses haftungsrechtliche Problem wird die Schaffung einer sog. ePerson als neue Art eines Rechtssubjekts diskutiert, die damit auch als Haftungssubjekt fungieren soll.[402] Das würde ihre Rechtsfähigkeit und die Existenz einer Haftungsmasse voraussetzen. Der Lösungsansatz der ePerson gleicht einer Revolution des Rechtssystems, sowohl in zivil- als auch strafrechtlicher Hinsicht, indem zum ersten Mal ein Rechtssubjekt geschaffen werden soll, bei dem jegliche Verbindung zu einem Menschen im Hintergrund fehlt.
Autonome KI-Systeme sollen als digitales Rechtsubjekt am Rechtverkehr teilnehmen und damit die befürchtete Verantwortungslücke schließen, die mit einer gleichzeitig bezweckten Entlastung der Hersteller einhergeht. Mit der Rechtsfähigkeit würde die ePerson in die Lage versetzt, grundsätzlich Träger aller Rechte und Pflichten zu sein und es wäre folglich begründungsbedürftig, wenn sie vom Vermögen der Trägerschaft bestimmter Rechte und Pflichten ausgenommen würde.[403] Damit wäre die ePerson auch Zuordnungssubjekt eines eigenen Vermögens. Da das BGB die Geschäftsfähigkeit jedes Rechtssubjekts vermutet[404] und diesbezügliche Einschränkungen im Recht bisher nur für natürliche Personen vorgesehen sind (vgl. §§ 104ff. BGB) wäre eine ePerson auch geschäftsfähig, also in der Lage Rechtsgeschäfte selbst vollwirksam vornehmen zu können, so dass sie über eigenes Vermögen auch verfügen könnte und rechtliche Pflichten erfüllen müsste. Als solche wäre sie auch Adresssatin von etwaigen Haftungsansprüchen oder Ansprüchen aus Vertragsrecht. Dies soll zu einer Entlastung der Betreiber und Hersteller solcher autonomen Roboter und KI-Systeme führen, da nun die ePerson für ihr Verhalten direkt verantwortlich gemacht werden könnte. Zugleich stellt sie dem geschädigten Haftungsgläubiger einen entsprechenden Schuldner gegenüber.[405]
Durch die Anerkennung der Rechtsfähigkeit eines Systems kann der Betreiber von der Haftung für das System befreit werden – zulasten der haftenden ePerson Die hierdurch einhergehende Haftungsbeschränkung zugunsten des Betreibers soll das Eingehen

---

[400] *Riehm* in: Kaulartz/Braegelmann, Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 6.1 Rn. 7.
[401] *Riehm*: Nein zur ePerson!, RDi 2020, 44.
[402] *Zech*, Gutachten A zum 73. Deutschen Juristentag, S. A 96.
[403] Vgl. *Behme*, in Hager beck-online.GROSSKOMMENTAR zum BGB, § 1, Rn. 4, Stand: 1. 9. 2020.
[404] *Schneider*, in Hager beck-online.GROSSKOMMENTAR zum BGB, § 104, Rn. 7, Stand: 15. 10. 2020.
[405] *Wettig/Zehender*, The Electronic Agent: A Legal Personality under German Law?, 2003, sub IV.2.

gesamtwirtschaftlich erwünschter Risiken fördern.[406] Dies greift natürlich nur, wenn die Haftung wie bei rechtsfähigen Personengesellschaften grundsätzlich auf die ePerson beschränkt ist und ein Durchgriff auf die Hinterleute nur mit Ausnahmetatbeständen möglich ist.[407]
Gleichzeitig wird damit eine „Verhaltenssteuerung“ bezweckt, denn wenn alle Chancen und Risiken bei ein und demselben Subjekt zusammenkommen, erhält es den Anreiz, sein Verhalten auf die jeweiligen Chancen und Risiken anzupassen und demnach nur solche Risiken einzugehen, bei dem die Chancen zumindest gleichwertig erscheinen. Erst die Verbindung der Konsequenzen etwaiger Verstöße durch die Haftungsübernahme von rechtswidrig verursachten Schäden beim tatsächlichen Verursacher macht das Rechtssubjekt für die Rechtsordnung steuerbar.[408] Die Konsequenzen können sich außerhalb des Zivilrechts auch durch staatliche Sanktionen, in Form von Bußgeldern oder Geldstrafen, niederschlagen.

6.4 Kritik

Gegen diesen Ansatz kann vorgebracht werden, dass seine Umsetzung mit Schwierigkeiten bei der Abgrenzung des Systems mit Rechtssubjektqualität verbunden wäre, dass er das auf Verhaltenssteuerung ausgelegte Haftungsrecht auf Maschinen anwenden will, obwohl es für menschliche Adressaten konzipiert ist, dass er das Ausfallrisiko allein dem Gläubiger der ePerson auferlegt, dass Probleme von juristischen Personen auch für ePersonen gelöst werden müssen, dass Herstellern ein Anreiz zum sicheren Konstruieren genommen wird, dass er dem strafrechtlichen Schuldprinzip widerspricht und er den Blick auf andere Lösungsansätze versperrt, die die identifizierten Probleme besser zu lösen vermögen.

6.5 Das „System“ als Rechtssubjekt

Die gängige Definition des Rechtssubjekts als Träger von Rechten und Pflichten setzt eine zuverlässige Zuordnung oder einen eindeutig abgrenzbaren „Träger“ also Zuordnungssubjekt voraus. Bei natürlichen Personen wird ihre rein physische Fassbarkeit unausgesprochen vorausgesetzt. Auffällig ist jedoch, dass über das Namens-, Personenstands- und Melderecht einiger Aufwand getrieben wird, um die Fassbarkeit, Unverwechselbarkeit, Erreichbarkeit bis hin zur Greifbarkeit jeder natürlichen Person im Rechtsverkehr sicherzustellen. Wer ein Recht geltend machen möchte, muss im Zweifel seine Identität nachweisen und es besteht ein hohes rechtliches Interesse, dass eine Person nicht verschwindet oder eine andere Identität annimmt.
Bei juristischen Personen dienen zur Sicherung der Fassbarkeit, Unverwechselbarkeit, Erreichbarkeit, Greifbarkeit und Identität einerseits der Firmenname mit Firmenzusatz, andererseits die Eintragung der Person und ihrer wichtigsten Attribute (z. B. GmbH: Firmenname, Sitz der Gesellschaft, Gegenstand des Unternehmens, Höhe des Stammkapitals, Personen der Geschäftsführer und das Ausmaß ihrer jeweiligen

[406] *Riehm*: Nein zur ePerson!, RDi 2020, Rn. 45.
[407] *Spindler* CR 2015, 766 (776f.).
[408] *Riehm*: Nein zur ePerson!, RDi 2020, Rn. 45.

Vertretungsbefugnisse nebst Anlagen wie Gesellschaftsvertrag, Liste der Gesellschafter mit Namen, Geburtsdatum und Anschrift sowie der Höhe der jeweils eingebrachten Stammeinlage u. a.) in der Anmeldung in einem öffentlichen Register. Alle Änderungen in diesen Attributen haben i. d. R. erst dann Wirkung, wenn sie selbst im Register eingetragen sind. Die zunächst rein juristische Person wird damit physisch greifbar, mit klarer Zuordnung zu – wiederum identifizierten und greifbare - natürlichen Personen und Vermögenswerten. Regeln zum Durchgriff auf die Hinterleute der juristischen Person und deren Verantwortlichkeit bei diesbezüglichen Mängeln stellen sicher, dass die Gläubiger oder der Staat, der die juristische Person in die rechtliche Pflicht nehmen möchte, nicht ins Leere greifen. Entsprechende Abgrenzbarkeit und Erkennbarkeit ist also zwingend erforderlich, um die ePerson als Rechtssubjekt zu konstituieren. Bereits diese Qualität steht aber hier in Frage.

Es wäre nämlich zunächst zu klären, aufgrund welcher Merkmale ein „System“ als Rechtssubjekt in Betracht gezogen werden soll, um daran eine Registereintragung zu knüpfen.[409] Die Problembeschreibung der Befürworter einer ePerson knüpfen hier offenbar an Merkmale, die sich an den hier beschriebenen Taxonomiedimensionen der Veränderbarkeit und der Transparenz festmachen. Die Veränderbarkeit und Intransparenz des „KI-Systems“ rühren aber nicht aus dessen physischen Komponenten, denen ein menschlicher Designer aus rein ästhetischen Gründen eine entfernt menschliche Anatomie gegeben haben mag („autonomer Roboter“), sondern nur aus dessen Programmcode, der die KI ausmacht. Dieser ist nicht physisch fassbar, dafür aber unbeschränkt replizierbar und nachträglich (auch durch Fremdeinwirkung) veränderbar. Ein „KI-System“ ist also bereits nicht hinreichend beständig, um selbst Rechtssubjekt zu sein. Insoweit ist es nicht verwunderlich, dass auf einen physischen (und damit in seiner äußeren Substanz greifbaren) „autonomen Roboter“ abgestellt wurde. Eine permanente Verbindung zu einem festen Medium ist wohl das Mindestmaß, um einen „Träger“ von Rechten zu konstituieren. Wie herausgearbeitet wurde, muss dieses qua Definition veränderbare System aber ein Mindestmaß an „Identität“ aufweisen und die Zurechnung von Rechtssubjektqualität an einen Programmcode bleibt fraglich.

Denkt man an Produkte wie ein autonomes Fahrzeug, dass seine Betriebserlaubnis nur aufgrund bestimmter geprüfter technischer Kriterien im Erlaubniszeitpunkt erhält und bei erheblichen technischen Veränderungen diese wieder verliert, ist eine solche „Identität“ fassbar. Es zeigt sich daran aber, dass bereits durch dieses grundlegende Kriterium der „Fassbarkeit“ eine Reihe von KI-Systemen als Rechtsubjekte ausscheiden, die sich ganz überwiegend als Programm darstellen. Auch muss dem Veränderbaren gleichzeitig eine Grenze der Veränderbarkeit gesetzt werden, um ihm eine „Identität“ zusprechen zu können. Noch klarer muss dies für Veränderungen von außen durch den Hersteller oder durch Dritte gelten, denn sie ermöglichen gleichsam den kompletten Austausch der ePerson. Wo ist also die Grenze zu ziehen, an der eine veränderbare ePerson ihre Rechtssubjektqualität wieder verliert? Für aus einem dieser Gründe nicht hinreichend fassbaren „KI-Systeme“ bestehen somit die beschriebenen rechtlichen Probleme weiterhin und müssten also ohne Rückgriff auf die Regeln für die ePerson gelöst werden. Es müssten Regeln entwickelt werden für „KI-Systeme“ ohne Rechtssubjektqualität und für ePersonen, was das Unterfangen wiederum erschwert. Die Befürworter der ePerson müssten sich also mit den genauen Taxonomie-

[409] *Riehm* in: Kaulartz/Braegelmann, Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 6.1 Rn. 27.

Anforderungen an eine solche ePerson erheblich tiefer auseinandersetzen, um all diese Abgrenzungen treffen zu können.
Insbesondere bei hochgradig vernetzten Systemen wäre bei Einführung einer ePerson zusätzlich zu klären, wie das KI-System von anderen Systemen abzugrenzen ist, die eine Rechtspersönlichkeit haben sollen, denn letztlich produziert hier das eine Programm den Input des nächsten, so dass sie häufig durchaus als Gesamtsystem aufgefasst werden könnten, bei dem beispielsweise die Zuordnung der Ursache eines Schadens und damit des Haftungsschuldners schwerfallen wird. Mit der vermeintlichen Problemlösung durch ePersonen können also neue Zuordnungs- und Beweisprobleme auftreten.
Es wird angeführt, dass bei Systemen, die sich spontan vernetzen und als Schwarm agieren (z. B. autonome Kleinroboter), eine Publizität entsprechend spontan hergestellt werden müsse, was zumindest technisch denkbar sei. So könne mithilfe einer Block-Chain die Vernetzung bestimmter Teilsysteme in einem bestimmten Zeitraum dokumentiert und publik gemacht werden. Im Schadensfall könne auf diese Dokumentation zurückgegriffen werden und das verantwortliche System definiert werden.[410] Auch hier ist aber bei qua Definition veränderbaren Systemen, die in ihrer Veränderbarkeit auch noch zusammenwirken, ein Fragezeichen zu setzen. Es werden wohl kaum erhebliche Teile des Programmcodes zu jedem Zeitpunkt in die Block-Chain aufgenommen, so dass der jeweilige Systemzustand unter Umständen schwer nachvollziehbar sein wird. Technisch mag das nicht ausgeschlossen sein, setzt aber wiederum konkrete Regelungen für das Maß der Veränderbarkeit, des Zusammenwirkens und der erforderlichen Dokumentation voraus, die bisher nicht vorgeschlagen wurden.
Materiell fehlt es also noch in vielerlei Hinsicht an konkreten Kriterien, um zumindest ein hinreichend konkretes Zurechnungssubjekt für Rechtsfähigkeit zu definieren.[411] Eine Bestimmung anhand der im Projekt entwickelten Taxonomie würde ihr sicherlich weiterhelfen. Erst dann könnte durch Eintragung in einem Register die nach außen wirksame Fiktion einer Rechtsperson geschaffen werden, sodass bei hinreichender Identität und Abgrenzbarkeit ein Mindestmaß an Rechtssicherheit nicht von vornherein ausgeschlossen ist. Nicht zuletzt aufgrund der Erforderlichkeit von eigenen Regeln für KI-Systeme ohne Rechtssubjektqualität stellt sich jedoch die Frage nach dem Verhältnis von Kosten und Nutzen der Einführung einer ePerson. Das für die Herstellung der Publizität erforderliche digitale Register wäre zunächst mit einem hohen technischen Aufwand staatlicherseits verbunden. Wird der materielle Anwendungsbereich der ePerson weit gezogen, umfasst er also eine Vielzahl möglicher technischer Ausgestaltungen der Systeme mit Rechtspersönlichkeit – gerade im Hinblick auf eine offene und spontane Vernetzung – wächst der technische Aufwand. Bleibt der Anwendungsbereich beschränkt, stellt sich die Frage nach der Erforderlichkeit, insbesondere wenn man die Alternativen zur ePerson berücksichtigt, wie eine Gefährdungshaftung oder eine Versicherungspflicht.[412]

---

[410] *Riehm* in: Kaulartz/Braegelmann, Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 6.1 Rn. 21 f.
[411] *Riehm* in: Kaulartz/Braegelmann, Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 6.1 Rn. 27.
[412] *Zech*, Gutachten A zum 73. Deutschen Juristentag, S. A 97.

6.6 Haftungsrecht zur Verhaltenssteuerung

Die Schaffung einer rechtsfähigen ePerson soll der Schließung von Haftungslücken dienen. Die ePerson soll für von ihr zu verantwortende Schäden aufkommen. Sie haftet mit ihrem Vermögen. Durch den Schadensersatz erfolgt ein vermögenswerter Ausgleich (unter Beachtung des Vorrangs der Naturalrestitution) für verursachte Schäden. Neben diesem Vermögensausgleich bezweckt das Haftungsrecht jedoch auch eine Verhaltenssteuerung. Deutlich wird dies bei der strafrechtlichen Haftung, bei der die Sanktion im Vordergrund steht bzw. die Kompensation für verursachte Schäden lediglich zivilrechtlicher Annex des Strafverfahrens ist.[413]
Dieses Prinzip auf Maschinen anzuwenden, ist aus zweierlei Gründen problematisch.

6.6.1 Keine Reflektion des Systems über vermögensrechtliche Konsequenzen

Erstens setzt eine Verhaltenssteuerung durch Haftungsrecht voraus, dass das haftende Rechtssubjekt ein Interesse daran hat, das eigene Vermögen zu erhalten, worin letztlich ein Überlebenswille zum Ausdruck kommt. Eine juristische Person handelt und entscheidet niemals selbst, vielmehr handeln für sie Organwalter, also letztlich natürliche Personen, die die juristische Person als Medium ihre Interessendurchsetzung nutzen.[414] Wie bereits im Kapitel über den Begriff „autonom“[415] erläutert, bezeichnet Kant Autonomie als die Eigengesetzgebung des freien Willens, die unabhängig aller empirischen Bedingungen ist und damit erst ethisches Handeln ermöglicht. Dieser Autonomiebegriff liegt auch dem zivilrechtlichen Begriff der Privatautonomie zugrunde. Natürliche Personen haben diese Privatautonomie aus eigener Kraft, juristische über ihre Organwalter. Maschinen wird sie lediglich zugeschrieben. Ihnen fehlt der freie Wille (zumindest bis in nicht absehbare Zukunft) und damit bereits ein konstituierendes Merkmal aller Rechtsubjekte, welches Voraussetzung für die bezweckte Verhaltenssteuerung ist.
Auch gegenüber Vermögensverlusten ist eine Maschine indifferent und damit nicht durch rechtliche Steuerungsmechanismen adressierbar.[416] Maschinen werden daher auch nicht durch (rechtliche) Verhaltensnomen, sondern durch technische Normen, gerichtet an deren Hersteller und Betreiber, reguliert. Es wäre zwar bis zur Erschöpfung des Vermögens der ePerson für Haftung gesorgt, eine Verhaltenssteuerung durch Haftungsrecht ist jedoch ausgeschlossen. Es bestünde die Gefahr, dass die ePerson den langfristigen Vermögensverlust für die Erreichung kurzfristiger Ziele hinnimmt, die von der Rechtsordnung missbilligt werden, also mit entsprechender Haftung verbunden sind.
Einem derartigen Agieren des Systems kann zwar durch entsprechende Programmierung begegnet werden, dann stellt sich jedoch die Frage, ob für eine solchermaßen eingehegte KI noch die Notwendigkeit der Schaffung einer eigenen Rechtspersönlichkeit besteht. Sachgerecht sind auch hier technische Normen, die den Herstellern und Betreibern entsprechende Vorgaben für die Konstruktion und

---

[413] *Simmler/Markwalder*, ZStW 2017, 20, 36.
[414] Vgl. *Riehm* in: Kaulartz/Braegelmann, Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 6.1 Rn. 29 f.
[415] Siehe Kapitel 4.2.3
[416] *Riehm* in: Kaulartz/Braegelmann, Rechtshandbuch Artificial Intelligence und Machine Learning, Kap. 6.1 Rn. 27.

Programmierung der Maschinen vorgeben, damit ein rechtlich unerwünschtes Verhalten in ihren Reaktionen von vornherein ausgeschlossen ist. So fordert die Ethik-Kommission in einem Bericht zum automatisierten und vernetzten Fahren, dass „*die Technik nach ihrem jeweiligen Stand so ausgelegt sein muss, dass kritische Situationen gar nicht erst entstehen, dazu gehören auch Dilemma-Situationen, also eine Lage, in der ein automatisiertes Fahrzeug vor der ‚Entscheidung' steht, eines von zwei nicht abwägungsfähigen Übeln notwendig verwirklichen zu müssen*."[417]

6.6.2 Menschliches Verhalten als Ursache für das Agieren des Systems

Zweitens verlangt haftungsrechtliche Verantwortlichkeit ein willensgetragenes Verhalten als Anknüpfungspunkt. Trotz der Vielzahl an Theorien[418], die sich in der Rechtswissenschaft unter der Überschrift des Handlungsbegriffs gebildet haben, lässt sich festhalten, dass jeder Handlung ein menschlicher Wille des Handelnden zugrunde liegen muss, um als Handlung im rechtlichen Sinne zu gelten. Nicht-Handlungen sind dagegen Geschehensabläufe, die zwar unter Mitwirkung der physischen Kräfte eines Menschen ablaufen, sich aber ohne die Mitwirkung seiner geistigen Kräfte vollziehen.[419] Wie eingangs geschildert, können besonders intransparente, wenig kontrollierbare und veränderbare Systeme den Eindruck erwecken, sie verfügten über einen solchen Willen. Letztlich ist jedoch das Agieren des Systems auf die Programmierung zurückzuführen. Zwar ist das konkrete Agieren im Zeitpunkt des Schadenseintritts und erst recht ex ante für den Programmierer nicht in jedem Fall erklärbar oder vorhersehbar. Es basiert jedoch auf einem Kausalverlauf, der willentlich durch einen Menschen in Gang gesetzt wurde und der bis zum Eintritt des Schadens nicht mehr durch ein eigenverantwortliches Dazwischentreten eines anderen Menschen unterbrochen wird.[420] Jede Entscheidung des Systems ist daher allein Ausdruck der Programmierung, die letztlich einem Menschen zugeordnet werden kann. Jede Reflektion über die Vor- und Nachteile eines Agierens, auch im Hinblick auf dessen Konsequenzen für das Vermögen der ePerson, ist in der Programmierung angelegt. In ihr angelegte Gefahren müssen von vornherein minimiert werden. Adressat einer entsprechenden Verhaltenssteuerung muss also immer der entwickelnde/programmierende Mensch und nicht die Maschine sein.
Diesem Argument liegen folgende technische Überlegungen zugrunde:
Zum einen trifft eine künstliche Intelligenz ihre Entscheidungen immer daten- und programmbasiert. Dadurch können Schäden, die durch das KI-System entstanden sind, auf mangelhaften Daten oder mangelhafte Programmierung eines Menschen beruhen. Insbesondere können die sogenannten Entwicklungs- oder Trainingsdaten für ein künstliches neuronales Netz fehlerhafte Entscheidungen nach sich ziehen, wenn sie fehlerhaft bearbeitet wurden oder Lücken aufweisen. Außerdem drohen Eingriffen von unberechtigten Dritten, wodurch die Datenqualität beeinträchtigt werden könnte (*Data Poisoning*).
Zum anderen kann die Zielfunktion, die Grundlage aller Entscheidungen einer KI ist, durch Menschen fehlerhaft programmiert oder vorsätzlich mit einer falschen Motivation

---

[417] Ethik-Kommission - Automatisiertes und vernetztes Fahren, Bericht Juni 2017, S. 10.
[418] *Puppe* in: Kindhäuser/Neumann/Paeffgen, StGB, Vor §§ 13ff Rn. 41 ff.
[419] *Kühl*, Strafrecht AT, § 2 Rn. 4.
[420] *Kühl*, Strafrecht AT, § 4 Rn. 68.

versehen werden, wodurch auch hier im Haftungsfall die Ursache menschliches Verhalten darstellt.
Die letzte Fehlerquelle ist schadhafte Hardware, beispielsweise ein nicht funktionierender Sensor. Auch das hat der Mensch zu vertreten.[421] Mit einer korrekten Datenbestückung der KI kann auch sichergestellt werden, dass Rechtsbegriffe, die eine moralische Wertung erfordern (etwa Arglist, Kollusion) von der KI erkannt werden und sie in der Lage ist, eine nach menschlichen Maßstäben moralisch richtige Entscheidung zu fällen.[422]
Zudem zeigt der Blick auf die juristischen Personen, dass der Gesetzgeber auch dort immer auf menschliches Verhalten zurückgreift, um dem Agieren der juristischen Person Rechtwirkung zu verschaffen:
Bei der juristischen Person sind es die Organe, also die Menschen als Organwalter, die die Handlungen für die juristische Person vornehmen, vgl. § 31 BGB für die Haftung des Vereins für Handlungen seiner Organe oder §§ 35 Abs. 1 S. 1, 6 Abs. 2 S. 1 GmbHG. Die GmbH wird durch den Geschäftsführer vertreten und Geschäftsführer kann nur eine natürliche, unbeschränkt geschäftsfähige Person sein. Erst der Mensch kann also nach außen wirksam handeln, jedoch wird dessen Handlung der juristischen Person zugerechnet.

6.7 Geschädigter trägt Ausfallrisiko

Im Vertragsrecht könnte eine ePerson als geschäftsfähiges Rechtssubjekt Erklärung im eigenen Namen und auf eigene Rechnung (d. h. mit Wirkung für und gegen das eigene Vermögen) abgeben. Damit haftete sie auch mit ihrem eigenen Vermögen und könnte verklagt werden (Parteifähigkeit im Zivilprozess). Derzeit würde das „KI-System“ als juristisches Konstrukt aber noch über keinerlei Vermögenswerte verfügen, aus denen Schadensersatzansprüche Drittbetroffener befriedigt werden könnten. Die Anerkennung des „KI-Systems“ als Haftungssubjekt, etwa im Straßenverkehr beim autonomen Fahren[423], würde zurzeit zu einer vollständigen Risikoverlagerung auf den Geschädigten führen, der keinen Schadensausgleich erhalten würde, wenn hinter der primär verantwortlichen KI keine Sicherungsmechanismen stehen.
Vor allem fehlt so der Anreiz, künftige Schäden zu vermeiden. Die KI kann den Schaden ohne Eigenkapital nicht zahlen, sodass gleichgültig ist, ob ein Schaden entsteht oder nicht. Hersteller und Betreiber eines autonomen Systems könnten so die Haftung der autonomen KI vorschieben. Lösbar erscheint dieses Problem nur unter Rückgriff auf die bereits für juristische Personen bestehenden Regelungen.

6.8 Probleme von juristischen Personen müssen auch für ePerson gelöst werden und machen sie entbehrlich

Die etwa zweihundertjährige Rechtsgeschichte von wirtschaftlich tätigen juristischen Personen kann teilweise als Prozess beschrieben werden, den Nachteilen ihrer

---

[421] *Scheufen*: Künstliche Intelligenz und Haftungsrecht: die e-Person aus ökonomischer Sicht, Wirtschaftsdienst, 2019, S. 411–414.
[422] *Wendehorst*, Ist ein Roboter haftbar?, Forschung&Lehre vom 13.1.2020.
[423] Vgl. dazu *Armbrüster*, ZRP 2017, 83, 85.

mangelnden Fassbarkeit zu begegnen, die Möglichkeiten zur Verantwortungsdiffusion und Aushöhlung der Haftungsmasse einhergehenden Nachteile für Gläubiger und Gesellschafter/Hinterleute zu begrenzen. Dazu zählen Regelungen durch den Gesetzgeber und die Rechtsprechung bezüglich der Durchbrechung der Haftungsbegrenzung und dem Durchgriff auf die Hinterleute, insbesondere bei Verletzung von Anforderungen an die Publizität und der Vermögenserhaltung. Hinter diesen Standards bezüglich der Garantie und Kontrolle einer angemessenen Vermögensausstattung eines fiktiven Rechtssubjekts, Publizität und vergleichbaren Sicherungen, kann auch für die ePerson kaum zurückgegangen werden. Wenn eine ePerson als rechtliche Lösung vorgeschlagen wird, muss entweder an die ohnehin vorhandenen Regelungen, etwa für die GmbH, angeknüpft werden. Soll hinter dem dort entwickelten Schutzstandart zurückgeblieben werden, wäre jeweils zu begründen, weshalb der Schutzbedarf von Gläubigern (insbesondere im Haftungsfall) und Hinterleuten in Bezug auf die ePerson geringer ist. Anhaltspunkte für einen geringeren Schutzbedarf liegen derzeit nicht vor. Wenn hier darüber hinaus festgestellt wurde, dass ein sogenanntes „KI-System“ als Träger der ePerson, gemessen an rechtlich erforderlichen Standards, auch nicht eine den Anforderungen der Privatautonomie genügende Willensentscheidung treffen kann, die rechtserhebliche Handlungen und Erklärungen tragen könnte, kann auch hier wie bei den juristischen Personen eine Anleihe genommen werden. Denkbar wäre es, auch für eine ePerson menschliche Organe vorzusehen. Organwalter müssten dann die erforderlichen Willensentscheidungen für die ePerson treffen oder sich deren Reaktionsvorschläge und vorab durch den Organwalter freigegebenen Reaktionen zu Eigen machen. In der Gesamtschau wird dann jedoch fraglich, weshalb nicht bestehende juristische Personen für diesen Zweck genutzt werden und weshalb es dann einer ePerson bedarf. Schon jetzt ist es jedem Nutzer eines KI-Systems freigestellt, bei Erfüllung der diesbezüglichen Anforderungen eine GmbH zu gründen, die als wesentliche Einlage das KI-System enthält und zum Gesellschaftszweck dessen Betrieb zu wirtschaftlichen Zwecken hat. Die GmbH wäre dann eine rechtsfähige Hülle, um das KI-System zu betreiben. Diese Hülle hat, da sie alle Anforderungen der GmbH erfüllt, auch den Vorteil, dass sie physisch schwerer fassbare KI-Systeme aufnehmen kann, soweit dabei die Anforderungen an eine Sacheinlage erfüllt werden (insbesondere ein Sachgründungsbericht, der für die Angemessenheit der Leistungen für Sacheinlagen einschließlich seiner etablierten Regeln folgenden Bewertung der Sacheinlage). Wenn einerseits eine „Roboter GmbH“ bereits gegründet werden kann, andererseits die rechtstheoretischen Anforderungen an eine ePerson nicht hinter der einer juristischen Person zurückbleiben können, bleibt jedoch fraglich, weshalb die Einführung einer ePerson erforderlich sein soll.

6.9 Haftung als Anreiz für Konstruktion sicherer Produkte fällt weg

Ferner tritt bei einer direkten Haftung der autonomen KI das Problem auf, dass die Motivation, immer sicherere Roboter zu bauen, für die Hersteller entfiele, weil sie nicht mehr für ihre eigenen Produktionsfehler haften müssten und nicht mehr dazu gezwungen wären, durch bessere Qualität haftungsauslösende Schadensereignisse

zu vermeiden. Das verringert die Qualität der KI an sich.[424] Hier kommt der Ausfall der Steuerungswirkung des Haftungsrechts zum Tragen. Falls es eine „ePerson“ mit der gerade geforderten Einlage und Vermögenssicherung gäbe, wäre diese wahrscheinlich in den meisten Fällen eine Art „Tochter“ des Unternehmens, dass das KI-System nutzt. Ein Vermögensverlust infolge eines Haftungsfalles, würde sich dann in der Bewertung der „Tochter“ bemerkbar machen, das Mutterunternehmen im selbem Umfang treffen und die verhaltenssteuernde Wirkung des Haftungsrecht wäre weitestgehend wiederhergestellt. Nur Schäden, die die Haftungsmasse der ePerson übersteigen, könnten durch die Haftungsbegrenzung das Mutterunternehmen nicht treffen. Um einen angemessenen Opferschutz gerade bei gefahrträchtigen KI-Systemen zu gewährleisten, müssten jedoch rechtliche Vorkehrungen getroffen werden, dass die ePerson ein angemessenes Vermögen im Verhältnis zu ihrem Einsatzbereich haben muss. Das würde aber wie im vorhergehenden Punkt den beabsichtigten Nutzen der ePerson in Frage stellen.
Um diesem Problem zu begegnen, schlägt die EU ein Versicherungssystem für ePersonen vor, die durch einen Fond ergänzt werden.[425] Auch Versicherungen kann man jedoch bereits heute für „KI-Systeme“ abschließen, um einerseits Haftungsrisiken beherrschbar zu machen und (im Fall von Haftpflichtversicherungen) potenziellen Opfern von Maschinen mit entsprechendem Gefahrenpotenzial in jedem Fall einen Zahlungsfähigen Schuldner gegenüber zu stellen. So müssen Halter autonomer Fahrzeuge bereits heute (wie für alle anderen Kraftfahrzeuge) „eine Haftpflichtversicherung zur Deckung der durch den Gebrauch des Fahrzeugs verursachten Personenschäden, Sachschäden und sonstigen Vermögensschäden“ abschließen und aufrechterhalten (vgl. § 1 Pflichtversicherungsgesetz). Für diese Art der Absicherung ist jedoch eine ePerson nicht erforderlich. Die Fondlösung ist eine im Bereich des Haftungsrechts schon lange diskutierte Figur, die mit der Einführung vieler Technologien mit Gefahrpotenzial immer mal wieder diskutiert und verworfen wurde. Mit der damit einhergehenden Vergemeinschaftung von Haftung wurde nämlich auch bisher befürchtet, dass die verhaltenssteuernde Wirkung des Haftungsrechts verloren geht, die im Pflichtversicherungsfall zumindest noch indirekt über die Gefahr der Hochstufung des Versicherungsnehmers nach einem Schadensfall Wirkung entfaltet.

6.10 Strafrechtliche Verantwortung setzt Schuldfähigkeit voraus

Betrachtet man den Bereich des Strafrechts, ist das Konstrukt einer eigenen Rechtspersönlichkeit für KI nicht zielführend. Entscheidend ist hier, dass sich der strafrechtliche Fähigkeits-, Personen- und damit auch Schuldbegriff an der Funktion der Strafe und des Strafrechts orientiert. Es geht anders als im Zivilrecht nicht um bloß finanziellen Schadensausgleich oder eine sichergestellte Zahlungsfähigkeit[426], sondern eine nach der individuellen Schuld des Täters bemessene Sanktion, die einen

[424] *Scheufen*, Künstliche Intelligenz und Haftungsrecht: die e-Person aus ökonomischer Sicht, Wirtschaftsdienst, 2019, S. 411–414.
[425] Entschließung des EU-Parlaments vom 16.2.2017 (2015/2103 [INL]) – P8_TA(2017)0051, Ziff. 57, 58.
[426] *Simmler/Markwalder*, ZStW 2017, 20, 36.

Lern- und Korrektureffekt beim Täter zur Folge haben soll, in Zukunft kein Unrecht mehr zu begehen. Schuld ist aber die persönliche Vorwerfbarkeit der Tat. Dieses Verschuldensprinzip und der Gedanke der Gefahrschaffung setzen aber menschliche Eigenschaften voraus.[427] Eine Personenqualität in diesem Sinn hat das KI-System nicht. Die meisten Strafzwecke haben darüber hinaus eine verhaltenssteuernde Dimension, die für Maschinen nicht greift. Schadensfälle wirken sich hier dagegen in aller Regel wiederum in einer Nachbesserung in technischen Normen aus und ggf. einer entsprechenden Nachrüstungspflicht der betroffenen Maschinen auf den darin zum Ausdruck kommenden neuen Stand der Technik. Für die Beschränkung und Untersagung des Betriebs von Fahrzeugen und Maschinen, von denen Gefahren ausgehen und im äußersten Fall deren Sicherstellung, stellt das Polizei-, Produktsicherheits- und Verwaltungsvollstreckungsrecht den zuständigen Behörden umfangreiche Möglichkeiten zur Verfügung, so dass auch alle Strafzwecke der präventiven Sicherung für Maschinen entbehrlich sind.
In der Frage nach der vermögensrechtlichen Straffolgen für ePersonen kann auf die Diskussion zur Einführung eines Unternehmensstrafrechts verwiesen werden, also insbesondere der Einführung einer Strafbarkeit juristischer Personen. Das in Art. 1 und 20 GG verankerte Schuldprinzip knüpft an die Eigenschaft des Täters als Mensch an und setzt voraus, dass er mit sozialethischem Tadel, also der Strafe, überhaupt ansprechbar ist. Damit scheinen eine Schuldfähigkeit und die Strafbarkeit von Unternehmen nicht ohne weiteres vereinbar zu sein,[428] was auf die ePerson in Bezug auf strafrechtliche Vermögensfolgen übertragbar wäre. Auch hier würde die ePerson also wie ein Unternehmen behandelt, soweit man sie für erforderlich hält.

6.11 Alternative Lösungen

Somit lässt sich festhalten, dass grundlegende Anforderungen für die Schaffung einer neuen Rechtspersönlichkeit für autonome KI-Systeme nicht einzuhalten sind. Auch die Datenethikkommission lehnt folglich mit ihrem Gutachten vom 23.10.2019 die Rechtsfähigkeit einer ePerson ab, da eine Analogie zwischen Mensch und Maschine bereits ethisch nicht vertretbar wäre. [429] Eine Rechtsfähigkeit für solche KI-Systeme ist nach geltendem Recht nicht vorgesehen und erscheint auch deshalb nicht erstrebenswert, da sie die aufgeworfenen Probleme kaum befriedigend und insbesondere nicht einheitlich für alle KI-System zu lösen vermag.[430] Eher könnte sich die Verantwortungslücke mit bereits bestehendem Recht oder durch analoge Rechtsanwendung bzw. durch kleine Modifikationen bestehender Regelungen zufriedenstellend füllen lassen.
Teilweise werden neue Zurechnungsregeln der vertraglichen Schadensersatzhaftung in Bezug auf KI-Systeme bestimmter Qualität für erforderlich gehalten (auch hier müsste weiter konkretisiert werden, welche Eigenschaften ein solches System haben soll). Es wird vorgeschlagen, das „Verhalten“ von KI-Systemen unbeschränkt dem

[427] *Denga*, CR 2018, 69, 77.
[428] *Kühl*, Strafrecht AT, § 10 Rn. 2 und 7.
[429] Gutachten der Datenethikkommission vom 23.10.2019, S. 31, S. 219.
[430] Vgl. auch *Zech*, Gutachten A zum 73. Deutschen Juristentag, S. A 97; *Spindler*, CR 2015, 766 ff.

Betreiber über § 278 BGB analog zuzurechnen.[431] Ebenso wie für die Verantwortung für das Verhalten Dritter gem. § 278 BGB könne ein Betreiber für Verletzungen durch ein KI-System einstehen. Die charakteristisch hohe Flexibilität der §§ 276 ff. BGB würde dabei helfen. Damit könnten im Einzelfall passendere Lösungen gefunden werden als durch die Schaffung einer komplett neuen Rechtspersönlichkeit.[432] Jedoch ist bereits die Forderung zur Anwendung des § 278 BGB nicht frei von den oben beschriebenen Vermenschlichungstendenzen, die auch der Forderung nach einer ePerson innewohnen. Es geht dort um die Verantwortungszurechnung der Handlungen von anderen Rechtssubjekten (Dritten), denen sich ein „Schuldner […] zur Erfüllung seiner Verbindlichkeit bedient" (so § 278 S. 1 BGB). KI-Systeme haben aber keine Rechtssubjektqualität und auch nicht die für die Konstituierung eines solchen erforderlichen Attributs. Für die Zurechnung der Reaktionen eines Werkzeuges (und das sind KI-Systeme) bedurfte es bisher keiner Zurechnungsregelung und solche werden auch in absehbarer Zukunft nicht benötigt, da die haftungsverursachenden Kausalketten nicht durch das Dazwischentreten anderen Rechtssubjekte unterbrochen werden.
Auch im Vertragsrecht können „Erklärungen" von KI-Systemen unmittelbar als eigene Erklärungen des Betreibers eingestuft werden, ohne dass dabei Analogien zum Stellvertreterrecht angewendet werden müssten.[433] Unabhängig davon, wie ein KI-System programmiert wurde, sind seine Entscheidungen jedes Mal wieder auf ihren Betreiber zurückzuführen. Er hat das Risiko gesetzt, dass ein KI-System Reaktionen tätigt und bekommt diese als seine eigenen rechtlich relevante Handlungen zugerechnet, wenn ein genereller Wille bei ihm vorlag, dass er mittels des KI-Systems handeln wollte. Davon wird rechtlich solange ausgegangen, wie kein Dazwischentreten Dritter (z. B. die ungewollte Ingangsetzung durch unberechtigte Dritte) die Zurechnung unterbricht. Neben den Chancen, die mit der Nutzung eines KI-Systems einhergehen, muss der Betreiber damit auch für unvorhersehbaren Risiken haften. Ebenfalls werden ihm Schäden durch nicht vorhergesehene Reaktionen der Maschine zugerechnet, soweit er die Gefahr einer Rechtsgutverletzung durch ihre Bereitstellung, ihre fehlende ausreichende Sicherung oder dergleichen geschaffen hat (Ingerenz). Somit entsteht gleichzeitig eine gesteigerte Motivation beim Betreiber dafür Sorge zu tragen, dass die Systeme immer weiter optimiert werden und das Risiko für unvorhersehbare Erklärungen sinkt. Gleiches gilt für den Programmierer, auch wenn hier durch haftungsrechtliche Regelungen eine sachgerechte Haftungsverteilung zwischen den Nutznießern der Technologie zu finden ist.[434] Der Bertreiber hat die üblichen Regressmöglichkeiten gegenüber dem Hersteller über das Vertragsrecht sowie über Produkt- und Produzentenhaftung, falls haftungsbegründende Tatsachen auf eine fehlerhafte Konstruktion des KI-Systems zurückzuführen sind. Sowohl Betreiber als auch Hersteller haben weitreichende Möglichkeiten, sich gegen entsprechende Risiken zu versichern. Für Kraftfahrzeuge muss ohnehin eine Pflichtversicherung abgeschlossen und erhalten werden, die auch dort eingesetzte KI-Systeme erfasst. Versicherungsgesellschaften bringen lange Erfahrung darin mit, Risiken abzuschätzen und (solange es sich um vertretbare Risiken handelt) mit einem

---

[431] Datenethikkommission, Gutachten 219 (8.2.2); *Hilgendorf/Hanisch*, Robotik im Kontext von Recht und Moral, 27 (59 f.).
[432] *Riehm*, Nein zur ePerson!, RDi 2020, 49.
[433] *Gitter*, Softwareagenten im elektronischen Geschäftsverkehr, 181 f.
[434] *Riehm*, Nein zur ePerson!, RDi 2020, 49.

risikoadäquaten Preis zu belegen. In Forschungsprojekten rund um autonome Fahrzeuge sind solche Versicherungen bereits heute teilweise beteiligt, um mit diesen neuen Risiken Erfahrungen zu sammeln. Maschinen im betrieblichen Bereich werden häufig von spezifischen Versicherungssystemen abgedeckt, wie Betriebshaftpflichtversicherungen, Unfallversicherungen etc. Sollte sich hier widererwarten eine Versicherungslücke auftun, könnte sie *de lege ferenda* durch eine weitere Pflichtversicherung nach bekanntem Muster leicht geschlossen werden.
Zudem lässt sich im Deliktsrecht über die bereits bestehenden Gefährdungshaftungstatbestände drohende Haftungslücken schließen.[435] Etwaige Regelungen könnten durch Anpassungen für KI-Systeme dazu führen, dass der Betreiber einer KI verschuldensunabhängig haftbar gemacht werden kann. Dieses Haftungsregime besteht bereits u. a. in der Straßenverkehrsordnung (§ 7 StVG) und ist auf automatisierte und autonome Fahrzeuge anwendbar. Für deren Einführung wurden mit der Novelle zum Straßenverkehrsgesetz von 2017 folgerichtig die Haftungsgrenze erhöht, um Befürchtungen bzgl. mangelndem Opferschutz in diesem Bereich vorzubeugen. Für besonders gefahrträchtige Unternehmungen, insbesondere den Betrieb von Eisenbahnen, sieht das Haftpflichtgesetz eine Gefährdungshaftung des Betreibers vor und trägt dem Umstand Rechnung, dass für einzelne Schädigungen bei Betrieb kein Verschulden nachweisbar ist, aber trotzdem derjenige haften soll, der aus dem Betrieb Nutzen zieht. Hier schlägt einerseits wieder der Ingerenzgedanke durch. Kraftfahrzeuge und Eisenbahnen bringen erhebliche Gefahren mit sich, sollen aber aufgrund ihres hohen gesellschaftlichen Nutzens gleichwohl genutzt werden dürfen. Wer sie nutzt, muss für diese Gefahren, die er bewusst in die Welt bringt, jedoch einstehen. Letztlich ist das „schuldhafte“ Verhalten, an das die Haftung angeknüpft wird, also das Inverkehrbringen selbst, an das die Haftung allerdings nur indirekt angeknüpft werden kann, da es sich um erlaubtes Verhalten handelt. Andererseits soll mit der Gefährdungshaftung verhindert werden, dass die durch die typische Betriebsgefahr entstehenden Kosten externalisiert werden.[436] Über die Pflichtversicherungsbeiträge sind sie beispielsweise dem Automobilbetrieb unmittelbar aufgelastet. Man hat für diese Gefahr im gewissen Sinne schon bezahlt. Des Weiteren wird diskutiert, dass für Betreiber eine Übertragung der Beweislastverteilung gem. § 831 Abs. 1 BGB angewendet werden könnte, wodurch sich die Betreiber für Rechtsgutsverletzungen ihrer KI-Systeme exkulpieren müssten.[437] Auch diese Analogie resultiert jedoch aus der bereits beschriebenen Vermenschlichung der KI-Systeme und ist voraussichtlich nicht erforderlich. Eine Exkulpation ist sogar abzulehnen, da sie mit einer Haftungslücke einhergeht. Vielmehr ist auch hier die direkte Zurechnung der Reaktionen eines Werkzeugs nur in den oben zu § 278 BGB beschriebenen Fällen denkbar.
Wesentlich besser werden die bestehenden Probleme bereits von der Produzentenhaftung nach § 823 Abs.1 BGB erfasst, nämlich durch eine richterrechtlich anerkannte Beweislastumkehr: Wenn feststeht, dass ein Produktfehler vorliegt, muss der Hersteller in Umkehrung der Beweislast beweisen, dass keine objektive Sorgfaltspflichtverletzung und kein Verschulden vorliegen. Das hat den Grund, dass der klagende Geschädigte keinen Einblick in die Produktion hat, also auch nicht

[435] *Zech*, ZfPW 2019, 198, 214 f.
[436] Vgl. dazu Zwischenbericht AP 6, 3.1.5.2.4.
[437] *Hilgendorf/Hanisch*, Robotik im Kontext von Recht und Moral, 27, 59.

beweisen kann, dass der Hersteller bei der Produktion sorgfaltswidrig gehandelt hat. Der Hersteller wiederum hat diese Informationen, sodass es ihm obliegt, sich zu entlasten.[438] Im Arzthaftungsrecht hat sich eine detaillierte Kasuistik zur Verteilung der Beweislast zwischen Behandelndem und Patient etabliert, die mittlerweile in § 630h BGB auch im Gesetz Ausdruck gefunden hat. Sie trägt u. a. den Informationsasymmetrien bei Beweis eines Behandlungsfehlers Rechnung.[439] Hinter beiden Ansätzen der Beweislastumkehr steht auch der Gedanke, dass sich sowohl Hersteller als auch Ärzte durch entsprechende Preisgestaltung ihrer Produkte und Dienste absehbare Haftungssituationen bereits einpreisen und sich zusätzlich versichern können. Für Opfer von fehlerhaften Produkten und Kunstfehlern bestehen diese Möglichkeiten jedoch nicht und sie sind, in einem für sie häufig undurchschaubaren System, ohne eine vereinfachte Haftung häufig schutzlos ausgeliefert. Die Versicherungen dieser Systeme durchschauen diese dagegen aufgrund ihrer Erfahrungswerte häufig gut, können „schwarze Schafe“ durch Kündigung oder Erhöhung der Prämien von zukünftigen Fehlern häufig abhalten und auf Verbesserungen in Bereichen drängen, in denen Schäden häufig auftreten und vermeidbar erscheinen. Hier realisiert sich die verhaltenssteuernde Wirkung des Haftungsrechts über die Versicherungen.
Ähnliche Ansätze der Beweislastumkehr sollten auch in Bezug auf KI-Systeme erwogen werden, soweit sie über die Produzentenhaftung nicht ohnehin bereits bestehen. Eine Ausweitung der Anwendungsfälle der Beweislastumkehr sollte insbesondere in Bezug auf die Betreiber der Systeme erwogen werden und ein für verschiedene Anwendungsfälle jeweils passendes Haftungskonzept entwickelt werden. Es muss jedoch auch berücksichtigt werden, dass der Halter hier mit einem durch die Beweislastumkehr verstärkten Haftungsrisiko für ein System belastet wird, dass er unter Umständen selbst nicht durchschaut. Insoweit müssten also auch seine Regressmöglichkeiten gegenüber dem Hersteller bei der Anpassung des Rechts geprüft werden. Hier könnten nicht zuletzt rechtliche Anforderungen an die Vertragsgestaltung zwischen Halter und Hersteller nützlich sein, die den Hersteller gleichzeitig verpflichten, den Halter ggf. in Haftpflichtprozessen mit ihrer Kenntnis des Produkts zu unterstützen.

6.12 Fazit

Das Konzept der ePerson kann die komplexen haftungsrechtlichen Probleme beim Einsatz von KI-Systemen nicht befriedigend lösen. 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. Da es für viele KI-Systeme ohne hinreichend physische Verfestigung (wie in einem einmal zugelassenen und insoweit begrenzt wandelbaren Fahrzeug) unmöglich sein wird, hinreichende Identität herzustellen, wäre für die KI-Systeme ohnehin eine alternative rechtliche Lösung zur ePerson zu suchen. Sinnvoller

[438] *Jäckel*, Das Beweisrecht der ZPO, Rn. 901.
[439] *Wagner* in: Münchener Kommentar BGB, § 630h Rn. 1 und 3.

erscheint es daher, sektorspezifisch für KI-Systeme das Recht dort weiterzuentwickeln, wo die Technik tatsächlich auf eine Lösung drängt. So scheint der Straßenverkehr eher mit den „KI-typischen“ Fragestellungen konfrontiert zu werden als der Industriebereich. Beide Bereiche zeichnen sich jedoch durch differenzierte Haftungsregime aus. Mit einem sektorspezifischen Ansatz lassen sich gezieltere und sachgerechtere Lösungen entwickeln, die auf bestehende Haftungskonzepte der Beweislastumkehr und Gefährdungshaftung aufsetzen können. Durch maßgeschneiderte Versicherungen und erforderlichenfalls Versicherungspflichten können so KI-Systeme in ihrem sektorspezifischen Risiko für Hersteller und Betreiber beherrschbar gemacht werden. Dabei ist auch das komplexe Zusammenspiel aus einerseits vorbeugendem Ordnungsrecht und Marktüberwachung sowie andererseits nachsorgendem, repressiven Haftungsrecht zu beachten. Dem Gesetzgeber bietet sich damit ein breites Instrumentarium zur Regulierung der KI-Technologie an, wobei viele vermeintlichen Probleme bereits mit dem bestehenden Recht ausgezeichnet gelöst werden können.

7 Projektzusammenfassung

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, Anlass zur Notwendigkeit von Änderungen an bestehenden Rechtsverordnungen wie beispielsweise Produktsicherheits- und Betriebssicherheitsrecht sein kann. Diese potentiell gefährdenden physischen Systeme, deren Verhalten durch Software bestimmt ist, die zu einem Teil aus KI-Algorithmen besteht, werden im Forschungsvorhaben 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 zusätzlichen 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 Sicherstellung der Sicherheit garantieren lassen. 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.
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 war in den unterschiedlichsten Normungsgremien mit der Thematik Safety in Verbindung mit KI tätig. 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. Deshalb hatten nur sehr wenige der Experten eine Doppelexpertise sowohl im Bereich KI als auch im Bereich Sicherheitsbewertung. Von einem etablierten Wissensgebiet KI-Safety kann bisher noch nicht gesprochen werden.
In der Folge der Befragungen stellte sich heraus, dass die Safety-Experten nur eine vage Vorstellung über Grenzen und Fähigkeiten von KI und die dort verwendeten Begrifflichkeiten und die KI-Experten kaum Vorstellung über Vorgehen, Denkweisen und regulatorische Rahmenbedingungen im Themenfeld Produktsicherheit hatten. So

wurde beispielsweise einerseits von einer KI-Expertin beklagt, dass übertriebene, KI als „menschgleich“ auffassende Vorstellungen über die Fähigkeiten der KI-Systeme vorherrschen. Andererseits war vielen Safety-Fachleuten nicht klar, was KI eigentlich sei, wodurch sich das Verhalten von KI-Software von normaler Software unterschied und welche Bedeutung Begriffe wie autonom, selbstlernend oder intelligent im KI-Kontext haben. 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.

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 erst die ersten Versuche gestartet werden. Eine Abschätzung über Zeithorizonte konnte von keinen der Experten gegeben werden. Als übereinstimmender Hinderungsgrund für eine schnelle Einführung wurden 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.

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 dem maschinellen Lernverfahren zu rechnen sind – ein Teilgebiet der KI. Das Besondere der hier eingesetzten KNN sind eine sehr große Zahl von verhaltensdeterminierenden Variablen, deren Werte durch Daten bestimmt werden, die in einem Lernprozess für das Eingangs-Ausgangsverhalten determinierend sind. 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. 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 Mächtigkeit werden eingesetzt, weil mit Ihnen Systemverhalten erzeugt werden kann, dass 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, dass man bei klassischen Algorithmen nicht oder nur in wesentlich kleinerem Maße kennt, und das dazu führt, dass das Systemverhalten in einem gewissen Maße unvorhersehbar 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. Das 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.

Für die rechtlichen Betrachtungen besonders relevant ist die Tatsache, dass für Systemverhalten, dass sich aus der Verwertung vieler Daten in einem Trainingsprozess ergibt, die Möglichkeit besteht, dieses durch Erweiterung des Trainingsvorgangs in die Zeit des betrieblichen Einsatzes optimal an die spezifischen Einsatzverhältnisse anzupassen. 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.

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. Insbesondere für letztere erwachsen mögliche Probleme des Sicherheitsnachweises, wenn das interne Systemverhalten auch von äußeren digitalen Daten determiniert ist. Die Taxonomie wird durch sieben Dimensionen mit jeweils zwei Unterkategorien konstituiert, die sicherheitsbezogene Merkmale des Systems auch unter Berücksichtigung seines Einsatzgebiets beschreiben. Dabei handelt es sich vorwiegend, aber nicht ausschließlich um Merkmale, die den Unterschied zwischen konventionellen Systemen und KI-Systemen beschreiben. Das Hauptziel der Taxonomie ist, einen Ordnungsrahmen für die rechtliche Kernfrage des Forschungsprojektes zu bilden – eine über diesen Zweck hinausgehende erweiterte Nutzung erscheint indes aber möglich.

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. Im 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 (die Inbetriebnahme) die Veränderungen des Systems nach der Inbetriebnahme ausblendet.

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. Für Maschinen kann dies entsprechend klargestellt werden. 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.

Im Forschungsprojekt wurden als Konsequenz der eben kurz angerissenen Problematik zwei Lösungsansätze formuliert:

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

Im Rahmen des Projekts wurden verschiedene Vorschläge zur Weiterentwicklung des Produktsicherheitsrechts ausgearbeitet. 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. 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. Im Zuge der Ausarbeitung alternativer Rechtsvorschriften wird der Begriff „wandelbar“ als eine Kombination der Taxonomiedimensionen Veränderbarkeit, Transparenz und Kontrollierbarkeit definiert.
Als Rechtsfolge eines „wandelbaren Produkts“ wird ein angepasstes „Produktbegleitungskonzept“ erarbeitet. Den Hersteller treffen nach diesem Ansatz Pflichten zur Gewährleistung der Sicherheit über den gesamten bestimmungsgemäßen Produktlebenszyklus. Das Produkt*begleitung*skonzept 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 – in Abgrenzung zur herkömmlichen Produkt*beobachtung*.
Um möglichen Problemen bei der rechtlichen Bewertung bei bestimmten externen Vernetzungen zu begegnen, wurde im Projekt ein Vorschlag für ein

- „Produktsicherheitsrecht für Daten“

entwickelt. 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. So kann ein regulierter Markt für solche sicherheitsrelevanten Daten geschaffen werden, in dem ein eigenes Produktsicherheitsrecht gilt, das für Rechtssicherheit sorgt.
Die rechtlichen Erörterungen werden ergänzt durch eine Erörterung 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 natürlichen und juristischen 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 aufsetzen können. Dem Gesetzgeber bietet sich damit ein breites Instrumentarium zur Regulierung der KI-Technologie an, wobei viele vermeintliche Probleme bereits mit dem bestehenden Recht ausgezeichnet gelöst werden können.

8 Glossar

Begriff	Definition	Erläuterung
Adaptivität	Das Systemverhalten adaptiert sich entweder während der Inbetriebnahme an vorhersehbare Randbedingungen oder während des Betriebs an Änderungen weniger Bedingungen (bspw. infolge von Temperaturänderungen, Abnutzung, Fabrikationsschwankungen).	Adaptivität ist die automatische Anpassung auf (in Bezug auf die Stärke der Änderungen) unterster Stufe.
Autonomie	Durch die Vorgabe von Gütekriterien, Zielen oder Zielhierarchien induzierte Eigenschaft der relativen Unabhängigkeit im Systemverhalten.	Die Verwendung des Begriffs Autonomie ist uneinheitlich. Generell sind Systeme bezüglich anderer Systeme autonom, wenn sie unabhängig (selbstständig) von diesen sind. Die Unabhängigkeit bezieht sich in diesem Gebrauch des Begriffes auf die Unabhängigkeit des Systemverhaltens vom Entwicklereinfluss.
autonom	Autonom bezeichnet das unabhängig/selbstständig sein von etwas und bezüglich etwas.	Diese allgemeine Definition kann je nach Kontext unterschiedlich spezifisch besetzt sein.
autonomes System	Ein autonomes System bezeichnet ein System, das selbstständig von etwas und bezüglich etwas ist. Es agiert in einem definierten Sinne unabhängig.	Das heute gängige Sprachverständnis versteht unter „autonomes System“ ein selbstfahrendes Fahrzeug.
autonom (FTS)	Mobilität außerhalb einer festen Spur.	Zweidimensionales Fahren, „unabhängig“ von der Spur.
autonom (Kraftfahrzeug)	Unabhängig von menschlichem Eingriff in die Aufgabe im Verkehr zu fahren.	U.U nicht unabhängig: andere Einflüsse des Menschen, wie Zielwahl, Entscheidung um Tanken, Entscheidung zur Reparatur, etc.
autonom (Software)	Softwareteil hängt von mehr Einflüssen als einer einzigen kontrollierenden Software ab.	Ein nicht autonomes Stück Software ist bspw. ein Unterprogramm ohne separaten Input.

Begriff	Definition	Erläuterung
autonom (KI)	Attribut für Software, die Züge autonomer Software und weitere Eigenschaften hat.	Weiter Eigenschaften sind beispielsweise: System plant, nimmt wahr.
Beschaffenheit	Eigenart oder Zustand einer Sache.	
Beschaffenheitsanforderung	Anforderungen an Eigenart oder Zustand einer Sache, meist aus dem Aspekt der Sicherheit heraus.	
Beschränkungen	Überwachen des externen Datenzugangs sowie schwer kontrollierbarer Teilsysteme oder Beschneidung deren Einsatzbereichs oder deren Funktionsumfangs.	
determinieren	Exaktes Bestimmen in allen Einzelschritten bzw. kausalen Abfolgen.	Handlung (z.B. Entwicklung), die gewolltes Systemverhalten zur Folge hat.
deterministisches System	Gleiche Eingangsdaten erzeugen gleiche Ausgangsdaten und gleiche interne Zustände.	Es gibt noch die Unterscheidung zwischen determiniert und deterministisch. Determinierte Systeme sind deterministische Systeme, bei denen die Forderungen nach jeweils gleichen internen Zuständen nicht erfüllt sein müssen. Die Unterscheidung spielt in diesem Projekt keine Rolle.
Emergenz	Verhalten aus sich heraus, nicht durch eine externe Instanz vorgegeben.	Aus Leeb. Vgl. Sedlacek, K. D. (2010), S. 44 f. Der Begriff der Emergenz stammt ursprünglich aus der Philosophie und bedeutet im logistischen Kontext die Erreichung einer höheren Entwicklungsstufe durch die Kommunikation niedriger Entwicklungsstufen. Vgl. Heidenblut, V., Hompel, M. ten. 2006), S.63.
Intelligenz	Messbares Merkmal einer speziellen psychischen Leistungsfähigkeit des Menschen.	Mögliches Unterscheidungsmerkmal von Menschen. Die Messbarkeit kann nicht sinnvoll auf Maschinen übertragen werden.

Begriff	Definition	Erläuterung
KI, Künstliche Intelligenz	Heterogenes Teilgebiet der Informatik, das aus dem ursprünglichen Ziel, menschliches Denken nachzubilden, entstanden ist.	In anderen veröffentlichten Definitionen wird auf den Begriff „Intelligenz“ Bezug genommen. Siehe dazu Diskussion des Intelligenzbegriffs (Kap. 2.2).
KI-Intelligenz	Attribut von KI-Systemen mit bestimmten Eigenschaften oder Attribut, das KI-Algorithmen zugestanden wird.	Wird in der Regel als Label und nicht als Messgröße verwendet. Teilweise wird Intelligenzrad mit „Fortschritt der technischen Entwicklung“ gleichgesetzt.
Marketing-Intelligenz	System der neuesten Entwicklungsklasse mit neuen, besonderen Eigenschaften.	
KI-System	Physisches System, das Softwarekomponenten auf Basis von Algorithmen der KI enthält.	
konnektionistisches Systeme	Eigenschaften entstehen aus der Interaktion vieler Komponenten, die sich wechselseitig beeinflussen.	
Kontrollierbarkeit	Eigenschaften von Systemen in Bezug auf das Vermögen des Herstellers oder Betreibers, das Systemverhalten determinieren zu können, sowie Maßnahmen, um die Beherrschbarkeit sicherstellen zu können.	
Lernen	Veränderung von Systemen nach Kriterien.	
Nachvollziehbarkeit	Maß für die Möglichkeit, Gründe für Systemverhalten geben zu können.	Ein Algorithmus ist nachvollziehbar, wenn ein Mensch die Frage beantworten kann „Warum verhält sich das System so?“. Wegen des expliziten Bezugs auf einen menschlichen Nachvollziehenden hängt das Maß der Nachvollziehbarkeit von dem Maß der Verstehbarkeit ab, ist also an Wissen und Fähigkeiten des Analysierenden gebunden.

Begriff	Definition	Erläuterung
Quasi-deterministisches System	Ein deterministisches System, dessen Determinismus wegen der Komplexität des Systemverhaltens nicht sichtbar ist.	Ein Beispiel quasi-deterministischer Systeme sind physikalische Mehrkörpersysteme, die zwar deterministisch sind, deren sichtbares Verhalten aber chaotisch ist. Große künstliche neuronale Netze sind quasi-deterministisch.
Quasi-Indeterminiertheit	De facto deterministische Systeme die aber aus Komplexitätsgründen praktisch nicht mehr vollständig berechenbar sind.	
quasi-nichtdeterministisches System	Tatsächlich deterministisches System, wegen der Komplexität aber in der Wirkung nicht deterministisch wirkend.	Nicht deterministisch wirkend heißt, dass das Merkmal deterministischer Systeme, die Berechenbarkeit von Zuständen, praktisch nicht erfüllt ist.
Risiko	Kombination aus Schadenshöhe und Schadensmöglichkeit	Üblich ist Risiko als (vermehrend) verknüpfte Kombination aus Eintrittswahrscheinlichkeit und Schadenshöhe zu definieren. Das setzt aber ein stochastisch denkbares Modell von Wirklichkeiten voraus. Wahrscheinlichkeiten sind an das Vorhandensein von Ereignissen gebunden. Der Begriff der „Möglichkeit“, wie er in der Fuzzy-Logik entwickelt wurde, ist weicher gefasst.
Robustheit	Widerstandsfähigkeit gegenüber unbekannten, untypischen oder bekannten, aber ungewollten Gegebenheiten und Einflüssen.	
Security	Fähigkeit eines Systems, Angriffen von außen zu widerstehen bzw. trotz Angriffen fehlerfrei (oder fehlerminimiert) zu agieren.	

Begriff	Definition	Erläuterung
selbst	„selbst“ bezieht sich auf etwas, das normalerweise von etwas (jemand) anderem durchgeführt oder gesteuert wird.	
Selbstorganisation	Erzeugung von Ordnung und Verhalten von Systemen aus dem nicht orchestrierten (von außen oder zentral) verändernden Wechselspiel der Komponenten des Systems.	
Sicherheit	Freiheit vor unvertretbarem Schaden	
Sicherheitsbezogene Systeme	Systeme, deren Verhalten einen Einfluss auf die Gefährdungslage haben können.	
Spezifizierbarkeit	Spezifizierbarkeit bezieht sich auf den Grad und das Abstraktionsniveau der Möglichkeit, Funktionen, Einsatzbereich, Umgebung, etc. für Systeme festlegen zu können.	
Taxonomie	Hierarchisches Klassifikationsschema	
Transparenz	Zugänglichkeit, Erklärbarkeit, Nachvollziehbarkeit und Vorhersehbarkeit aller relevanten Informationen über Eigenschaften und Verhalten eines Systems für die relevanten Empfänger.	
Veränderbarkeit	Veränderbarkeit beschreibt Änderungen der Eigenschaften bzw. des Verhaltens eines Systems oder seines Umfelds im Betrieb.	
Vernetzung	Beschreibt Eigenschaften von technischen, software-physischen Systemen in Bezug auf interne Prozesse des Austauschs von digitalen Daten und bezüglich des Einflusses systeminternen Verhaltens von äußeren digitalen Daten.	

Begriff	Definition	Erläuterung
Vorhersehbarkeit	Maß für die Möglichkeit, Systemverhalten durch Analyse, Modellierung oder Simulation vorhersagen zu können.	
Wandelbarkeit	Eine Kombination aus Veränderbarkeit, Transparenz und Kontrollierbarkeit.	Wird im Rechtsbegriff „wandelbares Produkt“ eingesetzt.
Weiterlernen	Lernen von Software nach Auslieferung des Systems an den Kunden.	
Weiterlernende Systeme	Systeme, die sich im Betrieb durch Lernen auf Basis von Daten verändern.	Insbesondere gilt das für Systeme, die vor Auslieferung durch Lernprozesse entstanden sind. Auch System, die nach Inbetriebnahme zu ersten Mal lernen, sollen dazugehören. Wesentlich ist das Merkmal des Lernens im Betrieb.
Widerstandsfähigkeit	das Vermögen von Systemen, trotz Störungen frei von sicherheitsrelevanten Fehlern zu agieren und etwaige sicherheitswirksame Fehlfunktionen abzuwenden oder zumindest deren Folgen abzuschwächen.	
Zielhierarchien	Hierarchisches Netz von Ober- und Unterzielen	Ein Beispiel ist das aus fünf Ebenen bestehende Handlungsfeld der Handlungsausführung von Menschen nach Oesterreich (1981) [440]. Ziel der obersten Ebene ist das „Kontrollstreben“, eine „höhere Form des Strebens nach Überleben“.

[440] Oesterreich, R. (1981). *Handlungsregulation und Kontrolle*. Urban & Schwarzenberg.

9 Literaturverzeichnis

ARMBRÜSTER, Christian: Automatisiertes Fahren – Paradigmenwechsel im Straßenverkehrsrecht?. In: *ZPR* 2017, 83-86.

BACKMANN, Ben; WAGNER, Christoph: Die „berechtigten Sicherheitserwartungen“ an Herzschrittmacher und Defibrillatoren im Sinne von § 3 Abs. 1 ProdHaftG. In: *MPR*, 2008, 29.

BAUA: *Autonome Roboter für Assistenzfunktionen: Interaktive Grundfertigkeiten – Ergebnisse und Forschungsperspektiven des Förderprogramms ARA1*, 2020.

BEHME, GROSSKOMMENTAR zum BGB, § 1, Rn. 4, Stand: 1. 9. 2020. In: *beck-online*, 2020.

BMJV: *Gutachten der Datenethikkommission vom 23.10.2019*. https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_DE.pdf;jsessionid=BBDF9C5D07D52CDECAE7A3FE376320A5.1_cid297?__blob=publicationFile&v=5 - Aktualisierungsdatum: 27.01.2021.

BMVI: *Ethik-Kommission - Automatisiertes und vernetztes Fahren. Bericht Juni 2017*: https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile

BOEHME-NEßLER, Volker: Die Macht der Algorithmen und die Ohnmacht des Rechts – Wie die Digitalisierung das Recht relativiert. In: *NJW* 42/2017, 3031.

CONRAD, Sebastian Conrad: Künstliche Intelligenz – Die Risiken für den Datenschutz. In: *DuD*, 2017, 740.

DENGA, Michael: Deliktische Haftung für künstliche Intelligenz. In: *CR*, 2018, Heft 2, 2018, 69-78.

DETTLING, Heinz-Uwe; KRÜGER, Stefan: Erste Schritte im Recht der Künstlichen Intelligenz – Entwurf der “Ethikleitlinien für eine vertrauenswürdige KI”. In: *MMR*, 2019, 211.

DJEFFAL, Christian: IT-Sicherheit 3.0: Der neue IT-Grundschutz - Grundlagen und Neuerungen unter Berücksichtigung des Internets der Dinge und Künstlicher Intelligenz. In: *MMR*, 2019, 289.

EBERS, Martin; HEINZE, Christian; KRÜGEL, Tina; STEINRÖTTER, Björn (Hrsg.): *Künstliche Intelligenz und Robotik Rechtshandbuch*, München: Beck C. H., 1. Auflage 2020.

EIFERT, Martin (Hrsg.): *Produktbeobachtung durch Private*. Berlin: Duncker & Humblot 2015.

EUROPÄISCHES PARLAMENT: *Entschließung des Europäischen Parlaments vom 16.02.2017 mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik*. https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_DE.html - Aktualisierungsdatum: 27.01.2021.

GAUGER, Dörte: *Produktsicherheit und staatliche Verantwortung*. Berlin: Duncker & Humblot, 2015.

GIERSCHMANN, Sibylle: Gemeinsame Verantwortlichkeit in der Praxis – Systematische Vorgehensweise zur Bewertung und Festlegung. In: *ZD* 2020, 69.

GITTER, Rotraud: *Softwareagenten im elektronischen Geschäftsverkehr – Rechtliche Vorgaben und Gestaltungsvorschläge*. 1. Auflage 2007.

GSELL, Beate; KRÜGER, Wolfgang; LORENZ, Stephan; REYMANN, Christoph: GROSSKOMMENTAR zum Zivilrecht, Stand: 01.03.2020. In: *beck-online*, 2020.

HILGENDORF, Eric (Hrsg.): *Robotik im Kontext von Recht und Moral*. 1. Auflage 2013.
JÄCKEL, Holger: *Das Beweisrecht der ZPO*. 3. Auflage, 2021.

JARASS, Hans D.: *Bundesimmissionsschutzgesetz – Kommentar*. 12. Auflage, 2017.

JARASS, Hans D.: Grundstrukturen des Immissionsschutzrechts. In: *JuS*, 2009, 608.
KAULARTZ, Markus; BRAEGELMANN, Tom (Hrsg.): *Rechtshandbuch Artificial Intelligence und Machine Learning*. 1. Auflage 2020.

KINDHÄUSER, Urs; NEUMANN, Ulfrid; PAEFFGEN, Hans-Ullrich (Hrsg.): *Strafgesetzbuch*. 5. Auflage 2017.

KIPKER, Dennis-Kenji: Umsetzungsgesetz zur NIS-RL nur mit geringen Anpassungen gegenüber der bisherigen Rechtslage beschlossen. In: *MMR-Aktuell*, 2017, 389121.

KIPKER, Dennis-Kenji / SCHOLZ, Dario: EU Parlament verabschiedet EU Cybersecurity Act. In: *MMR-Aktuell*, 2019, 414986.

KOHTE, Wolfhard: Arbeitsschutz in der digitalen Arbeitswelt. In: *NZA*, 2015, 1417.

KOHTE, Wolfhard; FABER, Ulrich; FELDHOFF, Kerstin: *Gesamtes Arbeitschutzrecht*. 2. Auflage 2018 (Handkommentar Arbeitsschutzrecht).

KOLLMER, Norbert; KLINDT, Thomas; SCHUCHT, Carsten: *Arbeitsschutzgesetz*. 3. Auflage, 2016.

KÜHL, Kristian: *Strafrecht Allgemeiner Teil*. 6. Auflage, 2008.

LANGHEID, Theo; WANDT, Manfred: *Münchener Kommentar zum Versicherungsvertragsgesetz*. 2. Auflage, 2017.

LOHMANN, Melissa F.: Ein europäisches Roboterrecht – überfällig oder überflüssig?. In: *ZRP*, 2017, 168.

OPPERMANN, Bernd; STENDER-VORWACHS, Jutta: *Autonomes Fahren – Rechtsprobleme, Rechtsfolgen, technische Grundlagen*. 2. Auflage, 2020.

PAAL, Boris. O.; PAULY, Daniel A.: *Datenschutz-Grundverordnung Bundesdatenschutzgesetz*. 2. Auflage, 2018.

RIEHM, Thomas: Nein zur ePerson!. In: *Recht Digital RDi*, 2020, 42-49.

SÄCKER, Franz Jürgen; RIXECKER, Roland; OETKER, Hartmut; LIMPERG, Bettina (Hrsg.): *Münchener Kommentar zum Bürgerlichen Gesetzbuch.* 7. Auflage 2017.

SCHEUFEN, Marc: Künstliche Intelligenz und Haftungsrecht: die ePerson aus ökonomischer Sicht. In: *Wirtschaftsdienst*, 2019, Heft 6, 411-414.

SCHNEIDER, GROSSKOMMENTAR zum BGB, § 104, Rn. 7, Stand: 15.10.2020. In: *beck-online*, 2020.

SIMMLER, Monika; MARKWALDER, Nora: Roboter in der Verantwortung?. In: *ZSTW*, 2017, Band 129, Heft 1, 20-47.

SPINDLER, Gerald: Roboter, Automation, künstliche Intelligenz, selbst-steuernde Kfz – Braucht das Recht neue Haftungskategorien?. In: *CR*, 2015, 766-776.

WAGNER, Gerhard: Produkthaftung für autonome Systeme. In: *Archiv für die civilistische Praxis*, 2017, 709.

WAHLSTER, Wolfgang; WINTERHALTER, Christoph: *Deutsche Normungsroadmap Künstliche Intelligenz*. Berlin: DIN/DKE, November 2020, 232 pp.

WENDEHORST, Christiane: Ist ein Roboter haftbar?. In: *Forschung & Lehre* vom 13.1.2020. https://www.forschung-und-lehre.de/recht/ist-ein-roboter-haftbar-2415/ - Aktualisierungsdatum: 21.01.2021.

WETTIG Steffen; ZEHENDER, Eberhard: *The Electronic Agent: A Legal Personality under German Law?*. 2003. https://beck-link.de/kc3tt - Aktualisierungsdatum: 27.01.2021.
WIEBAUER, Bernd: Behördliche Anordnung im Arbeitsschutz. In: *NVwZ* 2017, 1653.

WILRICH, Thomas: Verantwortlichkeit und Pflichtenverteilung gemäß Betriebssicherheitsverordnung. In: *NZA* 2015, 1433.

ZECH, Herbert: Entscheidungen digitaler autonomer Systeme: Empfehlen sich Regelungen zu Verantwortung und Haftung?. In: *Gutachten A zum 73. Deutschen Juristentag*, Hamburg 2020/Bonn 2020, 1. Auflage, 2020.

ZECH, Herbert: Künstliche Intelligenz und Haftungsfragen. In: *ZfPW*, 2019.

10 Abbildungsverzeichnis

11 Tabellenverzeichnis