NJW Neue Juristische Wochenschrift

djt Seit 1860. 2022

Eine Sonderveröffentlichung von NJW und DJT

September 2022

Künstlich-intelligente Maschinen? Fragen an Techniker und Juristen

In der Biosphäre sind Intelligenz und Autonomie allgegenwärtig; muss gehandelt werden, verknüpfen Lebewesen Sinneseindrücke mit erlernter und mitgeteilter Information. In der komplexen Technosphäre können selbstlernende Maschinen menschliche Intelligenz und Autonomieleistungen ergänzen oder sogar ersetzen. Diese selbstlernenden Systeme sind auf dem Weg, alle Lebensbereiche zu durchdringen und damit auch die menschliche Zukunft von Intelligenz und Autonomie neu zu vermessen. Der 73. Deutsche Juristentag beschäftigt sich in diesem Jahr unter anderem mit der Frage der Haftung autonomer Systeme. Die Frage der Verantwortung und Haftung für Entscheidungen digitaler autonomer Systeme wird in einer Vorfeldveranstaltung des djt am 12. September 2022 im OLG Köln interdisziplinär beleuchtet und diskutiert. Diese Sonderveröffentlichung versammelt die Beiträge der Referenten.

Verlag C.H.Beck München und Frankfurt a.M.

Inhalt

Generalmajor Dr.-Ing. Michael Färber*

Digitalisierung der Bundeswehr

Digitalisierung optimiert die Durchsetzungsfähigkeit der Streitkräfte, erhöht die Einsatzfähigkeit der Bundeswehr als Ganzes sowie auf dem digitalisierten Gefechtsfeld und unterstützt das Verwaltungshandeln. Damit trägt Digitalisierung entscheidend zur Auftragserfüllung der Bundeswehr bei.

[1] Für die Streitkräfte ist Digitalisierung ein Schlüsselelement zur Informations-, Führungs- und Wirkungsüberlegenheit, aber auch zur Verbesserung des Schutzes sowie der Durchsetzungs- und Reaktionsfähigkeit.

[2] Die Digitalisierung der Bundeswehr und der Aufbau eines durchgängigen, leistungsfähige Informations- und Kommunikationsverbundes werden gemäß der Strategischen Leitlinie Digitalisierung[1] auf drei aufeinander aufbauenden Ebenen aktiv gestaltet:

- IT-Standardisierungschafft modular aufgebaute, skalierbare und wiederverwendbare IT-Strukturen im Geschäftsbereich (GB) des BMVg, in denen Prozesse flexibel, umfassend und auch in der Mobilität digital unterstützt werden können.
- IT-Evolutionbaut die bestehenden IT-Services weiter aus mit dem Ziel, technologische Fortschritte zu nutzen, die Interoperabilität zu erhöhen und die IT-Steuerung effizient auszugestalten.
- IT-Innovationerschließt Neuerungen und Schlüsseltechnologien, um dem GB BMVg mit der zeitnahen Einführung digitaler Technologien zu einem Vorsprung im Einsatz und letztendlich auch im Grundbetrieb zu verhelfen und diesen zu erhalten.

[3] Um der stetig wachsenden Bedeutung der Informationstechnik, in einem umfassenderen Ansatz aber auch des Cyber- und Informationsraums insgesamt, Rechnung zu tragen, hat die Leitung BMVg entschieden, zum 1. Oktober 2016 eine neue ministerielle Abteilung Cyber/Informationstechnik (CIT) aufzustellen, und zum 1. April 2017 den neuen Organisationsbereich Cyber- und Informationsraum (CIR).

I. Der Organisationsbereich Cyber- und Informationsraum (CIR)

1. Aufgabenwahrnehmung in der Dimension Cyber- und Informationsraum (CIR)

[4] Einsatzorientierte Leistungserbringung der Bundeswehr („Markenkern Einsatz“) findet in unterschiedlichen physikalischen Dimensionen statt. Die Planungssystematik der Bundeswehr kennt hierfür den Begriff der so genannten Erbringungsdimensionen, in denen Leistungen erbracht werden. Dies sind:

– Land	landbasierte Operationsführung
– Luft/Weltraum	luft-/ weltraumgestützte Operationsführung
– See	seegestützte Operationsführung
– Cyber- und Informationsraum (CIR)	Wirken im Cyber- und Informationsraum

[5] Diese Einteilung basiert auf der Annahme, dass die für die Erbringung von Leistungen notwendigen Systeme mit Blick auf ähnliche Aufgabenstellungen organisatorisch zusammen geführt werden sollten (… alles, was auf dem Boden agiert, sich in der Luft/ im Weltraum befindet oder in diesen/ aus diesem wirkt, sich über oder unter Wasser bewegt, im Cyber- und Informationsraum agiert …).

[6] Hinter den erstgenannten drei Erbringungsdimensionen stehen die „klassischen“ Teilstreitkräfte Heer, Luftwaffe und Marine, die vierte der genannten Dimensionen wurde mit der Neugründung des Organisationsbereiches CIR eingerichtet.

[7] Mit Aufstellung des Organisationsbereiches wurde dieser zunächst als sechster Organisationsbereich der Bundeswehr aufgestellt. Die Logik der Erbringungsdimensionen militärischer Leistungen – oder einfach der Dimensionen – hat sich jedoch mittlerweile festetabliert und ist auch durch die klassischen Bereiche Heer, Luftwaffe und Marine akzeptiert. Damit ist CIR neben Heer, Luftwaffe und Marine der vierte „Dimensionsverantwortliche“.

[8] Deutschland ist weltweit bislang das einzige Land, das diesen Schritt so konsequent gegangen ist. In gewisser Weise war die Aufstellung des Organisationsbereiches eine „Wette auf die Zukunft“, die auf eine weiter zunehmende Bedeutung militärischer Leistungserbringung in dieser Dimension setzt. Dass diese Entscheidung richtig war, ist mittlerweile an vielen Stellen deutlich geworden.

[9] In einem ersten Schritt wurden im Rahmen der Aufstellung drei größere Bereiche aus der Streitkräftebasis herausgelöst und unter einem neu aufzustellenden Kommando – dem Kommando CIR – zusammengeführt. Dies waren das Kommando Informationstechnik der Bundeswehr (KdoITBw), das Kommando Strategische Aufklärung (KdoStratAufkl), sowie das Zentrum für Geoinformationswesen der Bundeswehr (ZGeoBw).

[10] Zurzeit durchläuft der Organisationsbereich im sechsten Jahr seines Bestehens eine grundlegende Anpassung seiner organisatorischen Gliederung, die unter dem Begriff CIR 2.0 firmiert. Im Rahmen dieses Projektes werden im Kern zwei Ziele verfolgt: den Organisationsbereich CIR in seiner Rolle als Dimensionsverantwortlicher zu stärken und zum andern als Treiber der Digitalisierung der Bundewehr zu etablieren. Diese beiden Standbeine bilden der Kern der zukünftigen Aufgabenwahrnehmung des Organisationsbereiches. Mit dieser Ausrichtung hat auch die Digitalisierung der Bundeswehr eine feste Heimat gefunden.

[11] Zu den Aufgaben des Organisationsbereiches gehören das Wirken im CIR, das Militärische Nachrichtenwesen, das

* Kommandeur des Kommandos Informationstechnik der Bundeswehr.

1 BMVg CIT vom 1.4.2017.

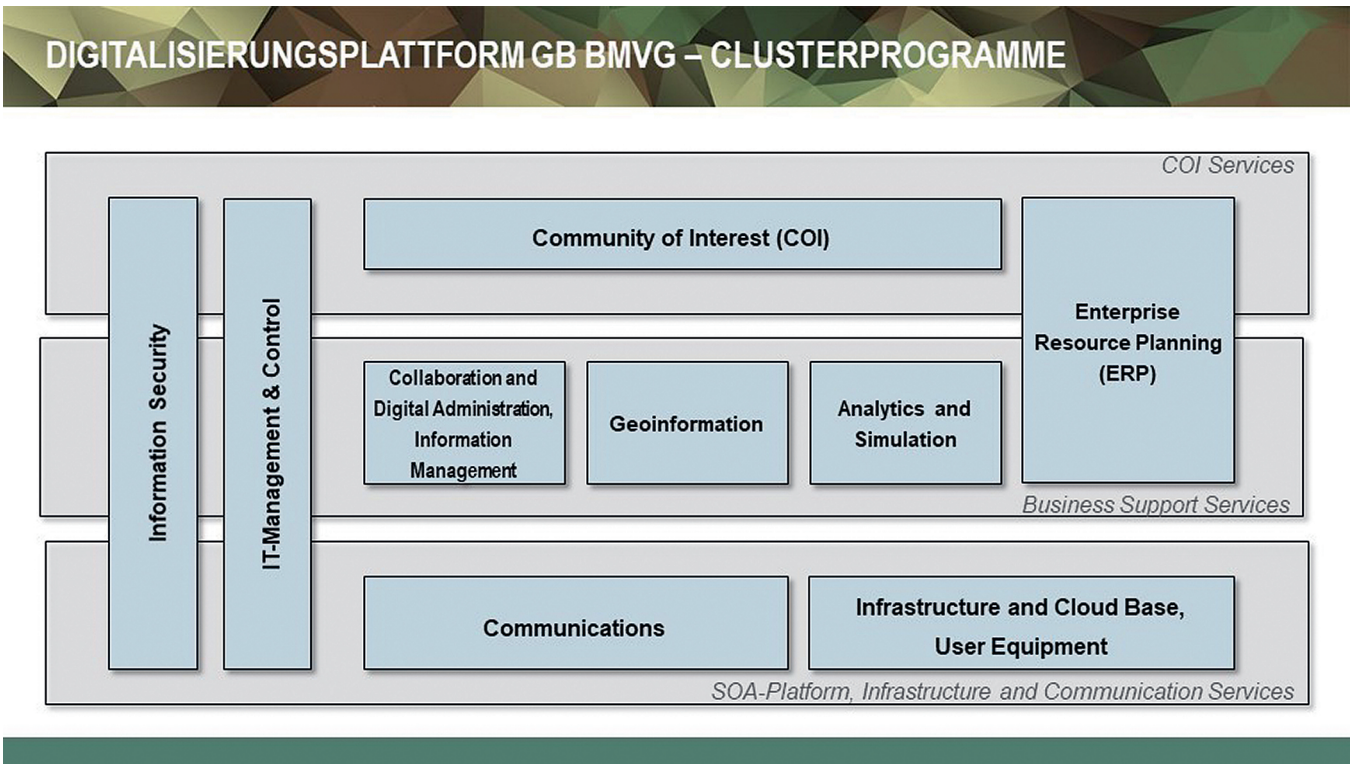

Abbildung 1

Bereitstellen von IT-Services, die Sicherstellung der Informationssicherheit und das Geoinformationswesen – also „CIR-Operationen aus einer Hand“ sowie die Wahrnehmung der Rolle als „Treiber der Digitalisierung der Bundeswehr“.

2. Die militärische Wertschöpfungskette: Führung, Aufklärung, Wirkung und Unterstützung

[12] Den Erbringungsdimensionen oder einfach Dimensionen stehen die so genannten Fähigkeitsdomänen gegenüber, die den Markenkern Einsatz der Bundeswehr in funktionalen Kategorien beschreiben. Die Festlegung dieser Domänen basiert auf der Annahme, dass für eine erfolgreiche Operationsführung eine Funktionskette durchlaufen werden muss, die vom Sensor bis zum Effektor reicht (Sensor to Shooter). Diese Funktionskette im Verbund Aufklärung – Führung – Wirkung, ergänzt um Unterstützung, stellt damit den Wesenskern einer Befähigung der Bundeswehr zur Vernetzten Operationsführung dar. Sie beschreibt das für die Einsatzführung in den einzelnen Erbringungsdimensionen – also Heer, Luft/Weltraum, See und CIR – notwendige funktionale Profil. In der Konzeption der Bundeswehr (KdB) ist diese Wertschöpfungskette als Verbund FAWU hinterlegt.

[13] Die vorgenannten Dimensionen zeichnen sich dadurch aus – dies ist gleichzeitig ihr Alleinstellungsmerkmal – dass sie in der Lage sind, diese Wertschöpfungskette vollständig zu durchlaufen, als Voraussetzung für eine erfolgreiche Operationsführung.

II. Die Digitalisierungsplattform des Geschäftsbereichs (GB) BMVg

[14] Die Digitalisierungsplattform GB BMVg soll zukünftig modular aufgebaute, wiederverwendbare, skalierbare, leicht und schnell adaptierbare und nach einem einheitlichen Regelwerk aufgebaute IT-Services zur Verfügung stellen, die nach Servicegruppen in neun Clustern zusammengefasst sind.

1. Rational

[15] Die IT der Bundeswehr ist bisher in mehr als 500 Einzelprojekten organisiert, deren Entwicklung, Beschaffung und Einsatz vorwiegend mit Fokus auf die verschiedenen Organisationsbereiche erfolgen. Diese Projekte sind überwiegend als Stove Pipes monolithisch aufgebaut und erbringen zwar jeweils für sich genommen die geforderte Funktionalität, folgen aber nur unzureichend einer gemeinsamen Architektur und damit einem verbindenden Systemgedanken.

[16] Dies führte in der Vergangenheit zu eingeschränkter Interoperabilität, Medienbrüchen, geringen Datenübertragungsraten, bedingten Netzwerkfähigkeiten, einer Vielfalt von Altsystemen, hoher Produktvielfalt, langen Analyse- und Entwicklungszeiten bis hin zu Projektabbrüchen.

[17] Die derzeit im Aufbau befindliche Digitalisierungsplattform soll die Vielfalt von IT-Systemen verringern, Finanzbedarfe durch Skaleneffekte reduzieren, die Projektrealisierung beschleunigen und den Betrieb der Informationstechnik vereinfachen.

[18] Hierzu werden die Einzelprojekte künftig nach Gesichtspunkten einer strengen Serviceorientierung betrachtet und mit ihren konstituierenden Elementen den in der Abbildung 1 dargestellten funktionalen Clustern zugeordnet werden. Damit entsteht zunehmend ein „IT-Baukasten“, welcher dem GB BMVg zur Wahrnehmung seines Auftrags zur Verfügung steht und gleichsam eine Angebotssituation schafft, aus der sich die verschiedenen Bereiche wie in einem Warenhaus bedienen können. Die für den „Kauf“ der erfor-

derlichen Services notwendigen Finanzmittel müssen allerdings nach wie vor selbst „mitgebracht“ werden.

[19] Durch Bündelung aller Maßnahmen eines Clusters in einem Clusterprogramm werden Portfolios harmonisiert und Beschaffungen beschleunigt. Ziel ist die möglichst umfassende Wiederverwendbarkeit und damit Skalierbarkeit der einzelnen Produkte, um in der Beschaffung, aber insbesondere auch in der Nutzung effizienter und damit wirtschaftlicher als bisher agieren zu können.

2. Cluster

[20] Cluster entwickeln proaktiv „schlüsselfertige“, wiederverwendbare und skalierbare IT-Lösungen. Das bedeutet: Alles, was Angehörige des GB BMVg in ihrer täglichen IT-Arbeit benötigen oder in naher Zukunft benötigen werden, wird im Idealfall bereits heute vorgedacht. So stehen die erforderlichen Lösungen, wie zum Beispiel die Hardware, Software und Rechenzentrumsleistungen zeitgerecht bereit und können bei Bedarf quasi „aus einem Regal“ zügig abgerufen werden.

[21] Die Cluster bündeln ihre Angebote in Clusterprogrammen, die die verfügbaren Leistungen dokumentieren. Jedes Programm umfasst eine Vielzahl inhaltlich verwandter IT-Services. Dabei sind die IT-Lösungen nicht nur standardisiert und wiederverwendbar, auch Integrationsfähigkeit und Interoperabilität werden gewährleistet, so dass sich Neuerungen stets nahtlos in das Gesamtsystem der IT der Bundeswehr einfügen.

[22] Wird ein neues IT-Projekt initiiert, so kann sich dieses aus den modular aufgebauten Clustern flexibel bedienen. Aus den vordefinierten IT-Lösungen suchen sich die Projekte diejenigen heraus, die sie für die Umsetzung ihrer Anforderungen benötigen. Wie aus einem Katalog lassen sich so die erforderlichen Lösungskomponenten schnell auswählen und zuverlässig kombinieren.

[23] Jedes Cluster deckt ein bestimmtes Themengebiet ab. So verantwortet beispielsweise das Cluster „Infrastructure, Cloud Base, User Equipment“ Themenfelder wie eine stationäre Cloud Infrastruktur, verlegefähige Rechenzentren sowie Endgeräte in Form von Laptops oder Smartphones. Das Cluster „Collaboration & Digital Administration, Info Management“ wiederum stellt beispielsweise Kollaborationswerkzeuge bereit, die den gezielten Austausch von Informationen und die effiziente Zusammenarbeit über die Grenzen von Abteilungen, Organisationseinheiten und Arbeitsorten hinweg ermöglichen.

[24] Der große Bereich der Künstlichen Intelligenz wird in einem Cluster gebündelt, das die Bezeichnung „Analytics and Simulation“ trägt. Hier sind diejenigen Anwendungsbereiche zusammengeführt, die üblicherweise mit dem Bereich der künstlichen Intelligenz in Verbindung gebracht werden. Dazu gehören unter anderem Verfahren zur Automatisierung standardisierter oder standardisierbarer Prozesse, Mustererkennung, Entscheidungsunterstützung, maschinelles Lernen, aber auch der Bereich der Simulation, hier im Besonderen Anwendungen zur Erweiterung der real wahrgenommenen Umgebung (Virtual Reality).

3. Clusterlogik und Markenkern Einsatz

[25] Der Verbund Aufklärung – Führung – Wirkung – Unterstützung stellt in generischer Form die für eine erfolgreiche Operationsführung zu durchlaufenden Prozessschritte dar. Die Digitalisierung dieser Prozesskette greift – so die Idee – zunehmend auf normierte IT-Services der Digitalisierungsplattform zurück.

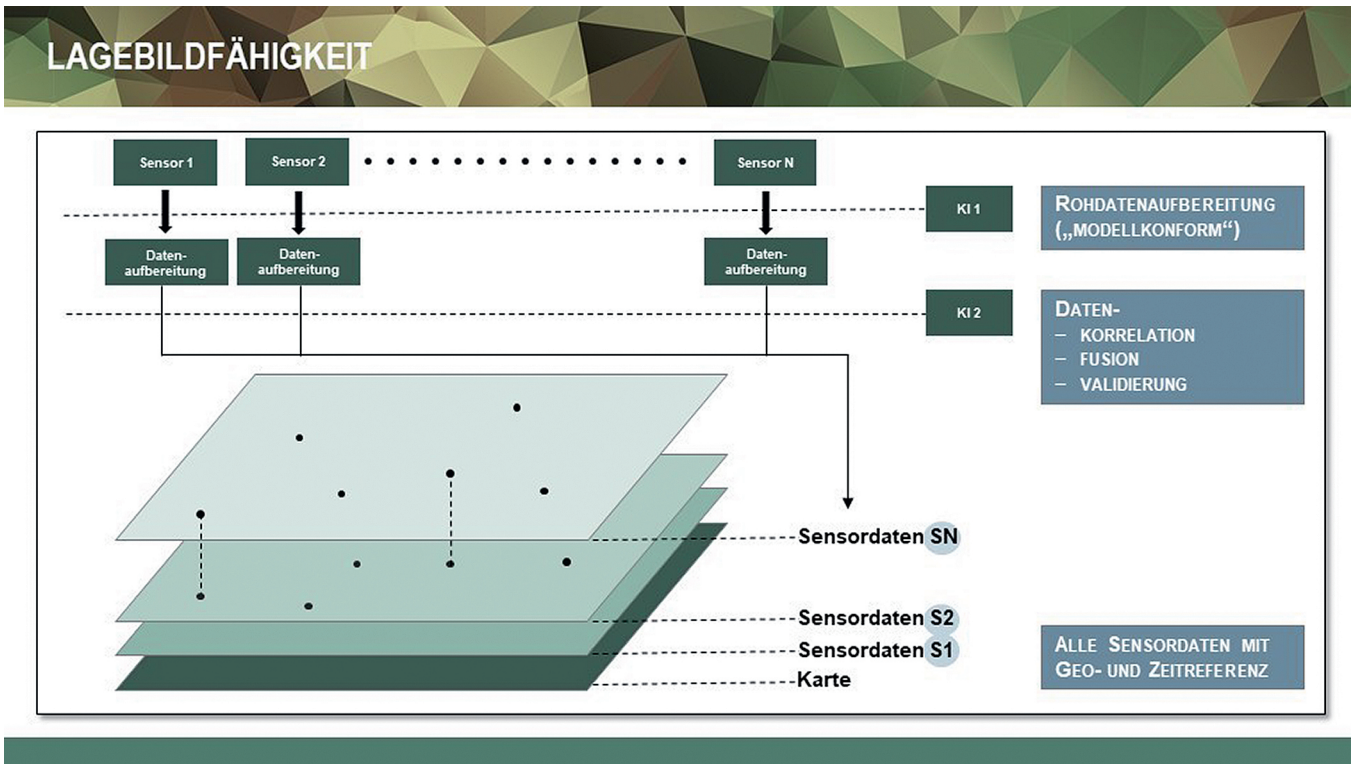

Abbildung 2

[26] Als Beispiel sei hier das Projekt „Digitalisierung Landbasierter Operationen (D-LBO) genannt, in dessen Rahmen IT-Services zur Unterstützung eines beweglich geführten Gefechtes zur Verfügung gestellt werden sollen. Schwerpunkt bildet hier zunächst die Domäne Führung: Durch querschnittliche Ausstattung mit softwarebasierten Funkgeräten sowie die Ausstattung mit einem Battle Management System wird die Möglichkeit geschaffen, breitbandig Informationen auszutauschen, Führungsstrukturen in einem logischen Verbund zu ordnen und relevante Informationen über einen Führungsprozess entscheidungsreif aufzubereiten.

[27] Daneben steht die Domäne Aufklärung, in der relevante Informationen aus unterschiedlichen Quellen zunächst erfasst werden müssen. Dazu gehören Sensoren über alle Bereiche des elektromagnetischen Spektrums, aber auch der Bereich der Human Intelligence.

[28] Der Aspekt der Künstlichen Intelligenz spielt hier mit einigen Aspekten bereits im Rahmen der Einsteuerung von Sensordaten in das zur Verfügung stehende IT-System eine wichtige Rolle: Die Daten müssen automatisiert so aufbereitet werden, dass sie der Darstellung auf einem normierten, verlässlichen Kartenuntergrund zugänglich sind. Dazu muss Information von Rauschen getrennt werden, die Formatierung der Daten muss so angepasst werden, dass sie den Regeln des zu Grunde liegenden Informationssystems gehorcht. Schließlich müssen die weiter zu verarbeitenden Daten mit einem Zeitstempel und einer Geo-Referenz versehen werden.

[29] Mit Blick auf ein in der Domäne Führung daraus herzustellendes gemeinsames, relevantes Lagebild[2] ist dies eine erste KI-unterstützte Stufe der Aufbereitung bzw. Auswertung. Die weitere Aufbereitung der Daten durchläuft Schritte wie Korrelation, Verifikation und Fusion, diese zweite Stufe der Auswertung erfolgt ebenfalls KI-basiert.

[30] Die so gewonnenen Daten sind zu bewerten und in einem Battle Management System darzustellen. Das damit erstellte Lagebild kann mit Verfahren der Entscheidungsunterstützung so aufbereitet werden, dass dem militärischen Führer anhand nachvollziehbarer Kriterien Handlungsempfehlungen zur Verfügung gestellt werden, die den jeweils gültigen Rules of Engagement entsprechen und sowohl an das jeweilige Szenar als auch an die aktuelle Gefährdungslage angepasst werden müssen.

[31] Das schließlich ausgewählte Wirkmittel muss in der Lage sein, die jeweils übermittelten Zielkoordinaten mit der Zielsignatur zur Deckung zu bringen. Hier sind Abbruchkriterien festzulegen, die potenzielle Nebenwirkungen oder unerwünschte Effekte minimieren.

[32] Die in der Prozesskette Aufklärung – Führung – Wirkung ausschnittsweise dargestellten Schritte zeigen die potenzielle Allgegenwart künstlicher Intelligenz.

[33] Ziel der Digitalisierungsplattform ist es, die zur Unterstützung von Prozessabläufen notwendigen IT-Services nach einem einheitlichen, bewertbaren Regelwerk zur Verfügung zu stellen, mit einem besonderen Augenmerk auf dem Cluster Analytics and Simulation, dass Anwendungen der Künstlichen Intelligenz ebenfalls aus einem konsolidierten Baukasten zur Verfügung stellt und insofern einer Normierung der für operationelle Verfahren verwendeten Anwendungen zugänglich ist.

4. Beherrschbarkeit Künstlicher Intelligenz

[34] Die Beherrschbarkeit künstlicher Intelligenz steht und fällt mit der Bewertbarkeit, Vorhersagbarkeit und Nachvollziehbarkeit der in ihr verwendeten Algorithmen. Mit besonderem Augenmerk sind daher Verfahren des maschinellen Lernens zu betrachten. Hier sind Testszenare zu entwickeln, die die Spannweite möglicher Einsatzoptionen so gut wie möglich abdecken und die ein Training der eingesetzten Verfahren ermöglichen. Auf dieser Basis sind standardisierte Bibliotheken anzulegen, die den „Lernerfolg“ sichern und abrufbar vorhalten.

[35] Die weiter oben bereits genannten Verfahren zur Automatisierung standardisierter oder standardisierbarer Prozesse und zur Mustererkennung, aber auch der Bereich der Simulation, hier im Besonderen Anwendungen zur Erweiterung der real wahrgenommenen Umgebung (Virtual Reality) scheinen mit Blick auf die Beherrschbarkeit der eingesetzten Algorithmen grundsätzlich weniger risikobehaftet.

III. Fazit

[36] Das Konzept der Inneren Führung hat sich in der Bundeswehr über viele Jahre bewährt. Es stellt den mündigen Staatsbürger in den Mittelpunkt, den auch in der Wahrnehmung seines militärischen Auftrages wertegebundenes Handeln leitet. Damit einher geht ein Regelwerk normativer Vorgaben, die auf unserer demokratischen Grundordnung fußen. Daraus abgeleitet ist das Konzept des „Führens mit Auftrag“, das nach dem Prinzip „Aufgabe – Kompetenz – Verantwortung“ die Entscheidungsbefugnis in der Durchführung eines Auftrages auf die dazu am besten geeignete Ebene legt.

[37] Die vorgestellte, im Aufbau befindliche Digitalisierungsplattform der Bundeswehr macht mit ihrem Prinzip der Modularität, Skalierbarkeit und Wiederverwendbarkeit den Aufbau von Wirkungsketten so transparent wie möglich. Anpassungen dieses „Baukastens“ an einer Stelle werden nach dieser Logik sofort im gesamten System umgesetzt. Damit wird ein hohes Maß an Steuerungsfähigkeit erreicht, die regelbasiertes Arbeiten überhaupt erst ermöglicht.

[38] Die Normierung der Fähigkeiten zur Künstlichen Intelligenz in einem Cluster „Analytics and Simulation“ schafft auch in diesem Bereich die Voraussetzungen für einen verantwortlichen Einsatz der dort vorgehaltenen Verfahren. Das dazu etablierte Regelwerk kann – und muss – sich an dem in der Bundeswehr etablierten Wertekanon ausrichten. ■

2 Common Relevant Operational Picture (CROP).

Dr. iur. Tassilo Singer*

KI im operationellen Kontext: Zur technischen Übertragbarkeit von Regeln des humanitären Völkerrechts auf Künstliche Intelligenz

I. Die Frage nach der Fähigkeit zur Regelkonformität von Künstlicher Intelligenz (KI)

[1] Bei der vielschichtigen, disziplinübergreifenden Debatte über eine militärische Nutzung von Künstlicher Intelligenz wird meist eine ganz zentrale Frage vernachlässigt: Welche technologischen Möglichkeiten bestehen überhaupt, Regeln in KI zu übertragen und diese KI in Systeme – auch militärische Systeme – zu integrieren?

[2] Dabei ist die Frage einzuschränken von *Dem, was technisch möglich ist* auf *Das, was rechtlich zulässig ist*. Denn nur für das technisch Mögliche und rechtlich zulässige KI-System, das, das also eingesetzt werden könnte und dürfte, lohnt sich ein Diskurs. Also muss es um die Frage gehen, **wie gut oder schlecht eine KI von uns bestimmte Regeln anwenden kann**.

[3] Digitalisierung optimiert die Durchsetzungsfähigkeit der Streitkräfte, erhöht die Einsatzfähigkeit der Bundeswehr als Ganzes sowie auf dem digitalisierten Gefechtsfeld und unterstützt das Verwaltungshandeln. Damit trägt Digitalisierung entscheidend zur Auftragserfüllung der Bundeswehr bei.

II. Definition: Autonomie, Vollautonomie; Funktionalität

[4] Zunächst muss klargestellt werden, dass es nach heutigem Stand der Technik immer nur um autonome Funktionalitäten bzw. Anwendungen und nicht um ein vollständig autonom „orchestrierendes" und sich selbstständig steuerndes System geht. Ein System, das vollständig autonom durch eine KI gesteuert wird und das selbstständig, ohne menschliches Zutun und ohne menschliche Einflussnahme-Möglichkeit operieren kann, ist vollautonom. Ein derart weit entwickeltes, unkontrollierbares System gibt es heute nicht (**man out of the loop**).

[5] Wenn das System zwar in engen Grenzen autonom handeln kann, aber menschliches Eingreifen in die rechtlich relevanten Entscheidungsprozesse immer möglich ist, ist es kontrollierbar (man on oder in the loop). Die menschlichen Steuerpersonen haben mit Ihrer Steuerungsmöglichkeit die Einhaltung des Rechts zu gewährleisten. Diese Menschen sind dadurch auch für alle damit verbundenen Handlungen verantwortlich.

[6] Für sich betrachtet handelt es sich damit auch bei einem Programm (oder einer Applikation), das selbstständig ein Ergebnis auf ein Problem findet oder Aufgaben löst, um eine autonome Funktionalität. Die Funktion ist determinierend für den anzuwendenden Rechtsrahmen und die damit verbundene Überprüfung.

III. Humanitär Völkerrechtliche Grundlagen und Legal Review nach Art. 36 ZP

[7] Den Ausgangspunkt für die Frage der Rechtskonformität von allen Waffensystemen bildet die Kombination von Art. 35 (1) und Art. 36 Zusatzprotokoll I zu den Genfer Konventionen.

[8] Danach ist eine Konfliktpartei durch die Regeln des humanitären Völkerrechts in Form von Waffenrecht und Einsatzrecht beschränkt in der Wahl von Mitteln und Methoden der Kriegführung. Man darf nicht einfach alle technisch möglichen Waffen und Waffensysteme einsetzen.[1] Die Überprüfung eines Waffensystems durch eine Konfliktpartei ist nach Art. 36 ZP I verpflichtend. Dies beinhaltet, ob eine Waffe oder ein Waffensystem *stets* oder *unter bestimmten* Umständen in einem bewaffneten Konflikt verwendet werden darf oder nicht.

[9] Daraus ergibt sich die Frage, ob ein System mit KI Funktion für sich gesehen und dessen Einsatz in einem bewaffneten Konflikt im Einklang mit dem humanitären Völkerrecht wäre.

IV. Analyse und Erkenntnisse zur Regelkonformität von KI in Waffensystemen nach Humanitärem Völkerrecht

[10] Im Detail wurde diese Frage durch den Autor bereits in einer umfangreichen Publikation untersucht und beantwortet.[2] Darin wurden der normative Inhalt und die sich daraus ergebenden Handlungsschritte aus der operationellen Praxis ermittelt und daraus die notwendigen Fähigkeiten objektiv herausgefiltert. Um die Erreichbarkeit der rechtlichen Anforderungen durch KI in Waffensystemen bewerten zu können, wurden so mögliche technische Herausforderungen herausgearbeitet und Lösungsansätze dafür entwickelt.

[11] Im Ergebnis lassen sich zwei Typen von Regeln unterscheiden.

- Regeln die aufgrund ihres normativen Inhalts Fähigkeiten erfordern, die durch ein vollautonomes System unter bestimmten technischen Anforderungen ausgeführt werden

* Capgemini Deutschland GmbH, Chapter Lead Data & AI @ Defense

1 Weapons and the Law of Armed Conflict, Boothby, Oxford University Press, 2nd Edition 2016, S. 2 f., 20 ff.

2 Dehumanisierung der Kriegführung, Springer International 2018, ISBN 978-3-662-57855-1, https://doi.org/10.1007/978-3-662-57856-8.

können, d. h. Regelkonformität eines autonomen Waffensystems (= *Verwendung unter bestimmten Umständen,* Art. 36 ZP I).

- Regeln die nicht technisch übersetzt und damit aufgrund ihres normativen Inhalts Fähigkeiten erfordern, die nicht in ein vollautonomes System integriert werden können (Regelwidrigkeit). Dabei handelt es sich insbesondere um Normen mit komplexem, situationsbedingtem Inhalt, die sich kaum in Heuristik übersetzen lassen. Dies sind Normen, die eine Wertentscheidung durch die KI verlangen, im Sinne einer ethischen, moralischen und rechtlichen Abwägung.[3]

[12] In der Konsequenz ergibt sich, dass eine Verwendung von vollautonomen Systemen ohne menschliche Kontrollmöglichkeit aus derzeitiger Sicht *stets* und in allen Einsatzszenarien *(vgl. Art. 36 ZP I)* rechtlich nicht zulässig ist.

[13] Daraus folgt, dass der Mensch immer eine Einflussmöglichkeit auf das Waffensystem oder die Waffe behalten muss.

[14] Es sind also nur Systeme mit KI Funktionalität möglich, die einer menschlichen Kontrolle unterliegen (**man on the loop**). Zudem müssen bei einer Realisierung ethische, politische und operationelle Bedenken berücksichtigt werden, was zu zusätzlichen Sicherheitsanforderungen führen dürfte (**Sandbox Tests, Auditierung, flexible Einsatzparameter und Anti-Tampering Maßnahmen**).

[15] Sollten KI-Fähigkeiten entwickelt werden können, die nachweislich zu einer effizienteren Zielidentifikation und/oder effektiveren Schonung von geschützten Objekten und Personen führen, kann sich sogar unter dem Vorbehalt der Zumutbarkeit eine Erfordernis eines Einsatzes (siehe dazu Art. 57 (2) a) ii) ZP I) ergeben. Dies wäre beispielsweise der Fall, wenn durch eine KI noch schneller ein Angriff abgebrochen werden kann (**Intervenierbarkeit**).

V. Anforderungen am Beispiel des Unterscheidungsgrundsatzes

[16] Von diesen Erkenntnissen ausgehend, stellt sich die Frage nach den konkreten Anforderungen, um diese Erkenntnisse technisch zu realisieren. Diese werden mit Blick auf den **Unterscheidungsgrundsatz** (Art. 48, 51, 52, 57, 35 (2) ZP I) untersucht.

[17] Diese Norm eignet sich besonders für eine wissenschaftlich-technische Untersuchung aufgrund ihrer Eigenschaft als relativ klar objektivierbare Regel und damit als Paradigma.

[18] Der Unterscheidungsgrundsatz enthält die **Verpflichtung, zwischen zivilen Objekten und Personen und militärischen Zielen zu unterscheiden,** mit der Folge dass rechtlich geschützte Objekte in Abgrenzung zu militärischen Zielen nicht Ziel von Angriffen sein dürfen. In operationeller Sicht ist der Unterscheidungsgrundsatz in den Ablauf des **NATO Joint Targeting Cycle** (NATO AJP-3.9) verpflichtend integriert.

[19] Damit man Einzelbestandteile des Grundsatzes auf ein System mit KI Funktionalität übertragen kann, muss zunächst der Inhalt des Grundsatzes präzise festgestellt werden und ein herauslösbarer Handlungsablauf herausgearbeitet werden.

[20] In diesem Zusammenhang bedeutet der Begriff „unterscheiden" dass man eine Person oder Sache in irgendeiner Form differenziert betrachtet und anders behandelt als eine weitere Person oder Sache. Dazu muss im Vorfeld ein Kriterium für eine Unterscheidung festgelegt werden.

[21] In einem ersten Schritt muss man die entsprechenden Personen oder Sachen anhand dieses Kriteriums wahrnehmen, d. h. auf das Vorliegen bzw. Nichtvorliegen des Kriteriums überprüfen (vgl. Detektion bei der KI Objekterkennung) und darauf folgend entsprechend differenzieren und in Kategorien einordnen (vgl. Klassifizierung bei der KI Objekterkennung).

[22] In einem zweiten Schritt wählt man für die derart getrennten Gruppen eine andere Art und Weise bzw. Form der Behandlung. Im Fall des Art. 48, 51, 52 ZP I bedeutet der zweite Schritt, dass Angriffe nur auf militärische Ziele erfolgen dürfen und zivile sowie schutzwürdige und schützenswerte Objekte und Personen kein Ziel eines Angriffs sein dürfen.

[23] Die Regel enthält also **keine zwingende Anknüpfung daran,** dass ein Mensch die Unterscheidung vornehmen müsste, sondern verpflichtet die Konfliktparteien dazu, dass sie diese Unterscheidung gewährleisten. Wie dies zu geschehen hat, wird nicht festgelegt.

[24] Damit verlangt die Regel objektiv, dass – soweit festgelegt ist was zu unterscheiden ist – ein Mensch oder System die verschiedenen Kategorien **detektieren** und zuverlässig **klassifizieren** kann, d. h. bspw. einen Panzer als militärisches Ziel und einen Zivilisten als geschütztes Objekt zu klassifizieren und im **Führungsinformationssystem** bzw. **Human-Machine Interface** als solche unterscheidend darzustellen.

VI. Übertragung auf ein System mit KI Funktionalität

[25] In einem weiteren Schritt müssen diese Anforderungen mit den heutigen Fähigkeiten von KI Modellen abgeglichen und adaptiert werden.

[26] Gerade in Bezug auf **Objekterkennung, Detektion und Klassifizierung** gab es in den letzten Jahren signifikante Fortschritte. KI Modelle können bei entsprechender Ausgestaltung und der Verfügbarkeit von tauglichen **Trainingsdaten** bereits auf die Detektion und Klassifizierung unterschiedlicher Fahrzeuge (ziviler PKW oder Panzer) und teilweise sogar Fahrzeugtypen trainiert werden.[4]

[27] Allerdings gelingt dies nicht für alle Situationen des Unterscheidungsgrundsatzes. So kann **situationsbedingte bzw. zweckorientierte Klassifizierung,** die nach Art. 52 (2) ZP I verlangt würden („Zweckbestimmung"; „Verwendung"), aus heutiger Sicht ein KI Modell noch vor zu große Schwierigkeiten stellen. Für diese Bestimmung ist ein vertieftes **Kontextverständnis,** wenn nicht sogar limitiertes Rechtsverständnis notwendig, über das derzeitige Modelle nicht verfügen. Zudem stehen entsprechende Trainingsdaten

3 Eine zahlenbasierte Übersetzung verbietet sich bei der Bewertung von Menschleben, s. BVerfGE 115, 118-166.

4 Vgl. auch: Multidimensionalen Operationen: Ableitungen für die Kernfähigkeiten zukünftiger Landstreitkräfte, Doll/Uysal, Wehrtechnischer Report 2/2022, Mai 2022, S. 13 ff., 15, 16.

nicht zur Verfügung, um ein Modell so zu trainieren, dass eine ausreichend effektive Präzision der Unterscheidung und Zuordnung erfolgt.

[28] Aus rechtlicher Sicht entspricht eine solche selektive, **use case-basierte Einsatzform** dem Gedanken des Art. 36 ZP I. Die Erkennung und Klassifizierung von bestimmten Objekten in Multi-Domain-Operations stellen ein bestimmtes Einsatzszenario als Teil des Unterscheidungsgrundsatzes dar. Dies ist eine Verwendung „unter bestimmten Umständen" nach Art. 36 ZP I (Gefechtsfahrzeug/nicht; Kriegsschiff/nicht; Kampfjet/nicht).

[29] Zusammenfassend ist eine Einbindung von einer auf Objektdetektion geschulten KI Funktionalität nicht für alle Situationen des Unterscheidungsgrundsatzes angezeigt, sondern nur für bestimmte Elemente und Funktionen des Unterscheidungsgrundsatzes, d. h. für bestimmte Einsatzszenarien, zu empfehlen.

Einsatz eines Systems mit KI Funktionalität in der Praxis

[30] Bei Identifikation geeigneter Einsatzszenarien und einer entsprechender Ausgestaltung kann ein System mit KI Funktionalität zur Objekterkennung einen erheblichen operationellen Mehrwert bieten.

Praxisbeispiel: ***Einsatz im Rahmen einer Combat Cloud***

Die **Multi-Domain Combat Cloud**, die im Rahmen des transnationalen **FCAS** Projekts entwickelt wird, ist ein möglicher Anwendungsfall für eine solche KI Funktionalität. Grundsätzlich sollen alle Informationen in der Multi-Domain Combat Cloud zusammenlaufen, analysiert und den beteiligten Akteuren rollen- und ebenengerecht in Echtzeit intelligent bereitgestellt werden. Durch die geplante offene Architektur und die offenen Application-Programing-Interfaces wäre die Einbindung einer KI Applikation zur intelligenten Erkennung von bspw. Gefechtsfahrzeugen möglich. Als Datenquellen kämen zunächst die optischen Sensordaten von unbemannten Begleitsystemen in Frage. Diese würden durch die KI-Applikation ausgewertet und klassifiziert in potentiell rechtmäßige und rechtswidrige Ziele in die Combat Cloud eingespeist.

[31] Durch eine Vor-Selektion von Bedrohungen wie bspw durch optische Hervorhebung im Führungsinformationssystem wird der menschliche Entscheidungsträger schneller und deutlicher auf mögliche „rechtmäßige" Ziele gelenkt und kann seine Entscheidung qualifizierter, aber weiterhin selbstbestimmt und eigenverantwortlich treffen.

[32] Schon heute ist dies als zusätzlicher Input für Befehlshaber zur Entscheidungsunterstützung im **Targeting Cycle** technisch machbar in Form von ausgewählten, gekennzeichneten und mit entsprechend hoher Wahrscheinlichkeit klassifizierten Objekten. Dazu würden Sensorinformationen von einem KI System verarbeitet und darauf basierend Objekte im Einsatzgebiet des Sensors identifiziert und klassifiziert. Diese Klassifizierung im Rahmen einer Applikation enthält eine Zuordnung eines Objekts mit einem entsprechenden Prozentsatz (bspw. „*zu 84 % ist das Objekt ein Kriegsschiff Typ XY*"). Über ein **Human-Machine Interface** erhält ein Mensch in einem Führungsinformationssystem Zugang zu den Informationen und kann möglicherweise über Bedrohungen durch optische Hervorhebung sogar explizit alarmiert werden. Dadurch kann der Entscheider auf ein „Mehr" an Informationen zurückgreifen und so schneller auf Bedrohungen reagieren. Durch die KI Funktionalität wird also zusätzliche „**situational awareness**" gewonnen und das Gesamtbild des „**theatre of operations**" vervollständigt bzw. präzisiert.

VII. Anforderungen und Ausblick

[33] Insgesamt enthält das humanitäre Völkerrecht also Regeln, die objektivierbar und übertragbar auf KI Systeme wären.

[34] Voraussetzung für eine entsprechend leistungsfähige KI sind dabei **ausreichend und hochwertige Trainingsdaten**, eine **spezifisch entwickelte KI Architektur** bzw. **Modelle**, ein sorgfältiger **Leistungsreview**, **Robustheitschecks** (Täuschmöglichkeiten und Fehleranfälligkeit der KI) und ein sog. **Äquivalenztest** (als *lex ferrenda*). Dieser Äquivalenztest meint eine rechtliche Überprüfung des jeweiligen KI Models mit simulierten Tests. Kernfrage ist, inwieweit eine mit KI angereicherte (enhanced) Entscheidungsfindung zu **gleichwertigen Entscheidungen** wie mit rein menschlichem Input käme. Um eine ausgewogene Entscheidungs-Evaluation zu bewirken, müssten in den dazu nötigen Referenzrahmen Grenzfälle und klare Verstöße integriert werden.

[35] Begleitet werden muss diese Analyse von einem breiten gesellschaftlichen, ethischen, politischen, rechtlichen und technischen Diskurs. Im Ergebnis wäre durch diesen holistischen Ansatz die Möglichkeit eröffnet, eine objektivierte und abgesicherte Herangehensweise der Integration von KI in operationellem Kontext zu entwickeln und dadurch im Ergebnis zivile Objekte und Zivilpersonen vor den Folgen der Kriegsführung besser zu schützen. ■

Prof. Dr.-Ing. Verena Nitsch/Prof. Dr.-Ing. Frank Flemisch*

Kooperative Systeme und Hybride Intelligenz – Plädoyer für ganzheitliche Human Systems Integration

Um Chancen und Risiken „künstlich intelligenter Maschinen" einzuschätzen, hilft das Konzept der Disruptiven Technologie. Ursprünglich im zivilen Umfeld entwickelt, wird dieses Konzept zunehmend auch auf Anwendungen in der Verteidigung übertragen, auf das „scharfe Ende der Digitalisierung". Der Blick auf die Verteidigung soll uns als Beispiel dienen, aus dem sich viel ableiten lässt.

[1] In Rahmen der NATO Science and Technology Organization (STO) werden disruptive militärische Technologien diskutiert. Die Situation zu Beginn des 21sten Jahrhunderts gibt deutliche Hinweise, dass es z. B. im Ukraine-Krieg die Kombination aus neuer Technologie (schultergestützte Waffen sowie Drohnen) in Verbindung mit agilen Gestaltungs- und Einsatzverfahren ist, die es der Ukraine ermöglichte, sich gegen einen zahlenmäßig weit überlegenen Gegner zu behaupten. Für die Zunahme der autonomen Fähigkeiten von Maschinen bis hin zu autonomen Systemen, insbesondere die Zunahme der kognitiven Fähigkeiten von Maschinen (Künstliche Intelligenz) wird bereits seit einiger Zeit als disruptive Technologie eingeschätzt und innerhalb der NATO manchmal als „Autonomous Systems", besser als Human Autonomy Teaming sowie als Cognitive Warfare intensiv diskutiert.[1]

[2] Zu bedenken ist ferner Henry Kissingers Beobachtung, der 2021 von einem zunehmenden Wettlauf nicht nur zwischen Mensch und Technik sprach, sondern auch zwischen Nationen wie USA und China. In diesen Wettlauf ist die Bundeswehr über ihre Bündnisverpflichtungen eingebunden. Ihre Konzepte der „Inneren Führung" und des „Staatsbürgers in Uniform" sind dabei eine besondere Herausforderung, aber auch Chance.

[3] Eine wesentliche Erkenntnis aus mehreren Jahrzehnten Forschung zu „künstlich intelligenten Maschinen" zeigt, dass noch nicht alle, aber wesentliche Probleme zu autonomen Fähigkeiten nun so ausreichend gelöst sind, dass sie bereits in die Anwendung gebracht werden können. Die nächsten Schritte, nämlich die Integration dieser Fähigkeiten in bestehende Systeme, insbesondere die Integration zwischen Menschen, Technik und Organisationen, sind jedoch zwar gut erforscht, müssen aber noch deutlich mehr auch in die Anwendung gebracht werden, um diese Systeme auch sicher einsetzen zu können. Insbesondere steht die systematische, ganzheitliche Berücksichtigung der menschlichen Faktoren und ihre Verbindung mit technischen und organisatorischen Faktoren über alle Systemschichten hinweg noch aus, und soll in diesem Beitrag als holistisches Modell, in einem Wechselspiel aus Theorie und anschaulichen Anwendungsbeispiele, überblicksartig skizziert werden.

I. System-Analyse 1: Intelligente Systeme aus Human Factors Sicht

[4] Um allgemein zu verstehen, wie KI und Autonomie zukünftige Verteidigungssysteme beeinflussen könnte und eine Mensch-System Integration erreicht werden kann, ist es hilfreich, zunächst Konzepte der menschlichen Kognition aus Human Factors Sicht, insbesondere vor dem Hintergrund psychologischer Forschung, zu beleuchten. Dabei ist menschliche Kognition nicht nur fokussiert auf ein Individuum zu erforschen, sondern zunehmend auf mehrere Menschen bis hin zu Gruppen, Organisationen und Gesellschaften. Verbundene Kognitive Systeme von Menschen und Organisationen, z. B. in Form von Kooperation, Führung oder Management, werden u. a. in der Sozial- sowie der Organisationspsychologie erforscht.

[5] Der Einsatz fortschrittlicher Informationstechnologien kann Führungsprozesse auf unterschiedliche Weise unterstützen und wird so Teil des Führungsprozesses. So können z. B. KI-gestützte Systeme Führungskräfte bei der Personaleinsatzplanung, der Personalentwicklung und der Leistungskontrolle unterstützen. Indem sie insbesondere aufgabenbezogenes Verhalten unterstützt, kann die Führungskraft mehr Kapazitäten für mitarbeiterbezogene Führung und die aktive Gestaltung zwischenmenschlicher Interaktion übernehmen. Es wird jedoch auch an künstlichen Systemen mit emotionalen Fertigkeiten geforscht, die eines Tages routinemäßig auch mitarbeiterbezogene Aufgaben unterstützen oder gar übernehmen könnten.[2]

II. System-Analyse 2: Kooperative Systeme (Joint Cognitive Systems)

[6] Ein entscheidender Schritt für unsere Diskussion hin zu autonomen und/oder KI-basierten Systemen ist, auch sie wie vorher bereits menschliche Systeme nicht isoliert, sondern als Teil eines kooperativen Systems mit einer kooperativen Kognition zu sehen. Ausgehend von der Idee der Mensch-Computer Symbiose wird Kognition zunehmend als Kooperative Kognition bzw. kooperative Automation erforscht und gestaltet. Ein Blick auf die Fortschritte der KI in Form von Deep Neural Networks offenbart einen technischen Fortschritt im Sinne einer verbesserten Lernfähigkeit aus Datensätzen, der jedoch zunächst mit einer im Vergleich zu

* Lehrstuhl und Institut für Arbeitswissenschaft der RWTH Aachen, Fraunhofer FKIE, Wachtberg

1 Flemisch, Towards a Holistic Understanding of Cognitive Warfare, including Human Factors, Human Systems Integration, Human-Machine Teaming and Human-AI Cooperation. Report of NATO-STO-RTG 356 "Cognitive Warfare" (2022 b, in press).

2 Nitsch/Popp, Emotions in robot psychology. Biological Cybernetics, 108 (5), 621-629 (2014).

klassischen Verfahren wie Zustandsautomaten oder Petri-Netzen verminderten Transparenz für den Menschen und Replizierbarkeit einhergeht. Insbesondere die Verifikation und Validierung dieser quasi-nichtdeterministischen Systeme ist eine Herausforderung und Gegenstand intensiver Forschung. Kommt der Mensch ins Spiel, sind Systemtransparenz, Erklärbarkeit und Kalibrierung des Vertrauens von entscheidender Bedeutung.

[7] Deutlicher Forschungsbedarf besteht in der direkten Beeinflussung der Nutzer durch KI, und insbesondere die Handhabung von Konflikten. Zwar ist aus der Forschung an Navigationssystemen bekannt, dass Menschen je nach Automatisierungsgrad eine unterschiedliche Kritik- und Restfähigkeit erhalten, weiterhin sind erste Ansätze zur systematischen Konfliktaushandlung oder Arbitrierung in der Grundlagenforschung erforscht.[3] Von einer wirklichen Durchdringung und Beherrschung solcher komplexen Führungs- und Entscheidungssituationen sind wir noch weit entfernt, insbesondere wenn mehrere Menschen und Rechner involviert sind.

III. System-Analyse 3: System of Systems, Organisation, Gesellschaft und globales System

[8] Bereits Clausewitz beschreibt den Kampf als ein „vielfach gegliedertes Ganzes." KI-basierte Waffensysteme agieren nicht im luftleeren oder rechtsfreien Raum, sondern sind in eine Vielzahl von Um- und Übersystemen eingebettet. Funktional sofort einsichtig ist die Einbettung der Mensch-Maschine-Systeme in ein System of Systems, zum Beispiel in einem komplexen Luftverteidigungssystem. Weiterhin klar wäre in dem Fall die Einbindung in Organisationen, wie z. B. die NATO und die Bundeswehr.

[9] Etwas unschärfer ist die Einbindung in ein politisches System und in die Gesellschaft. Die Soldaten sind zwar in dem Moment der Entscheidung vor allem in der aktuellen Situation Teil des dortigen Systems, sind aber auch Teil der Organisation(en) und, als entscheidender Punkt der modernen deutschen Streitkräfte, auch Staatsbürger in Uniform. Die verschiedenen Schichten des komplexen Verteidigungssystems sind einerseits physisch z. B. über Energie-, Material- und Nahrungsmittel-Nachschub, kognitiv z. B. durch Informationsaustausch, aber auch ethisch verbunden, indem die Staatsbürger in Uniform ethisches Denken nicht nur anwenden dürfen, sondern im Rahmen des zeitlich Möglichen auch sollen.

IV. System-Analyse 4: Verbindungen und Ketten

[10] Entscheidend sind weitere Denkschritte, die Verbindungen und Ketten der unterschiedlichen Schichten untereinander aufzuzeigen und damit erst versteh- und beeinflussbar zu machen. Im Falle der Bundeswehr unterscheiden sich diese Ketten von denen anderer Armeen. Die Prinzipien „Innere Führung“ und „Staatsbürger in Uniform“ haben immer auch die Integrität dieser Ketten als Ziel und sind auf die Integrität dieser Ketten angewiesen. So wäre es ethisch unangemessen und für das Vertrauen in diesen fragilen Ketten extrem schädlich, wenn wir unsere Entscheider in Situationen bringen würden, in denen sie die Verantwortung tragen, aber von den einbettenden Systemen wie Organisation und Gesellschaft nicht mit ausreichend Autonomie, Fähigkeiten oder Autorität ausgestattet wurden. Diese Art von Situationen wird auch als „Unsicheres Tal der Automation/AI“[4] oder als Moral Crumble Zone[5] bezeichnet.

[11] Eine intensive Erforschung, Diskussion und bewusste Gestaltung dieser Ketten wird umso kritischer, je mehr sie in Vor-Kriegs- und Kriegszeiten von einem Systemrivalen oder Gegner bewusst angegriffen oder erodiert werden. So diskutiert die NATO gerade unter „Cognitive Warfare“ mögliche Angriffs- und Verteidigungsvektoren bzgl. eigener und gegnerischer kognitiver Fähigkeiten[6], die auch auf eine Bundeswehr der Zukunft, erst recht mit dem Anspruch von innerer Führung und von Staatsbürgern in Uniform, erhebliche Auswirkungen haben wird.

V. Zusammenfassung und Ausblick

[12] Der Beitrag baute ausgehend von einem anschaulichen Beispiel schrittweise ein kybernetisches, ganzheitliches Modell von Menschen und KI in Verteidigungssystemen der Zukunft auf. Zentral dabei sind Rückkopplungsschleifen / Feedback-Loops, die auch in Dilemma-Situationen im „Nebel des Krieges“ sichere Entscheidungen und Handlungen erlauben. Die entscheidenden Ketten sind dabei aus Vertrauen, Autorität, Fähigkeit, Autonomie, Kontrollierbarkeit und Verantwortung gewoben über alle Schichten unserer Gesellschaft und Organisation wie der Bundeswehr, eingebettet in eine NATO, bis hin zum einzelnen Waffensystem und zum Individuum, hier ein Staatsbürger in Uniform.

[13] Spätestens mit diesem Modell wird klar, dass es keine wirklich autonomen Waffensysteme geben darf, sondern „nur“ Waffensysteme mit autonomen Fähigkeiten, die sicher in das jeweilige System-of-Systems, die Organisation und die Gesellschaft eingebettet werden müssen. Weiterhin ist klar, dass es keine Systeme geben sollte, die nur auf KI beruhen, sondern immer die gemeinsame Kognition (Joint Cognition, manchmal auch hybride Intelligenz genannt) von Menschen, KI und Organisationen. Neben physikalischen und kognitiven Sachverhalten müssen gerade in einem modernen Verteidigungssystem wie in Deutschland immer auch ethische Sachverhalte ausreichend mitgedacht werden. Nur ein Teil der Ketten ist bereits so ausführlich und über alle notwendigen Schichten wie Bundeswehr, Politik und Gesellschaft diskutiert und verstanden, dass KI bereits flächendeckend und sicher ein-

3 Flemisch et al., Conflicts in Human-Machine Systems as an Intersection of Bio- and Technosphere: AiAIDo: ICHMS International Conference on Human-Machine Systems (2020).

4 Flemisch et al., Uncanny and Unsafe Valley of assistance and automation: First Sketch and application to Vehicle Automation. Advances in Ergonomic Design of Systems, Products and Processes, 319-334 (2016). doi:10.1007/978-3-662-53305-5_23.

5 Elish, Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. Engaging Science, Technology, and Society (pre-print), SSRN: https://ssrn.com/abstract=2757236 or http://dx. doi.org/10.2139/ssrn.2757236.

6 Cao et al., Countering cognitive warfare: Awareness and resilience 2021, https://www.nato.int/docu/review/articles/2021/05/20/countering-cognitive-warfare-awareness-and-resilience/index.html (abgerufen am 10.8.2022); Flemisch, Towards a Holistic Understanding of Cognitive Warfare, including Human Factors, Human Systems Integration, Human-Machine Teaming and Human-AI Cooperation. Report of NATO-STO-RTG 356 “Cognitive Warfare” (2022 b, in press).

gesetzt werden kann. Gleichzeitig haben wir eine gute Vorstellung davon, wie wir in einem gemeinsamen Prozess aus Forschung, gesellschaftlichem Diskurs und Entwicklung, KI und autonome Funktionen als kooperativen Systeme und als hybride, kooperative Intelligenz so in einen schrittweise und überwachten Einsatz bringen, dass unsere Bundeswehr eingebettet in die NATO auch in Zukunft eine realistische Chance hat, Systemrivalen und Gegner erfolgreich abzuschrecken oder uns dagegen zu verteidigen. ■

Jun.-Prof. Dr. iur. Katharina Kaesling*

Verantwortung und Haftung für „Künstliche Intelligenz" zwischen Recht und Technik

I. Verantwortung durch Recht und die Verantwortung des Rechts

[1] Rechtswissenschaft kann als Wissenschaft von der Verantwortung beschrieben werden.[1] Wird eine Rechtsnorm gebrochen, so können hieran verschiedene Folgen geknüpft werden, wie etwa strafrechtliche Sanktionen oder zivilrechtliche Ansprüche gegen den Verantwortlichen. Der Begriff der Verantwortung findet sich dabei in der Rechtssprache eher selten.[2] Er ist kein konturierter Fachbegriff, gleichwohl aber Schlüsselbegriff der Rechtswissenschaft.[3] Sachverhalte werden nicht nur unter Rechtsnormen subsumiert und damit rechtlich bewertet, sondern das Recht wirkt gerade durch die Begründung von Verantwortungsbereichen auch auf künftige Sachverhalte und das Handeln der Akteure ein. Dass es ohne Verantwortung kein Recht gibt,[4] bedeutet aber nicht, dass es keine Verantwortung ohne Recht gäbe. Das Recht muss sich nicht nur die Frage gefallen lassen, inwieweit die rechtlich geschaffene Verantwortungsordnung außerrechtlichen Konzepten von Verantwortung entspricht, sondern Antworten auf diese geradezu einfordern. Neben die Frage, wer wie haftet, tritt die Frage, wer wie haften sollte.

II. Haftungsrecht als Technikrecht

[2] Im zivilrechtlichen Haftungsrecht finden sich bereits eine Reihe von Haftungskonzepten, mit denen Personen Verantwortung zugeordnet wird. So finden sich Tatbestände der Verschuldenshaftung, von denen einige das Verschulden vermuten.[5] Die Gefährdungshaftung knüpft die Haftung demgegenüber nicht an Verschulden, sondern an die Schaffung bzw. Aufrechterhaltung einer Gefahrenquelle.[6] So haftet etwa der Tierhalter, dessen Tier eine Sache beschädigt, grundsätzlich finanziell für den hieraus entstehenden Schaden, der Grundstücksbesitzer für Schäden durch den Einsturz eines Gebäudes und Eltern für Schäden, die ihr Kind einer anderen Person widerrechtlich zufügt. Das Haftungsrecht steuert Risiken auf der Grundlage von Risikobewertungen. Erlaubte Risiken werden haftungsrechtlich kompensiert. Haftungsrecht trägt so zum Rechtsfrieden bei und kann bei neuen Technologien Vertrauen in und Akzeptanz für ihre Nutzung fördern. Haftungsrecht ist daher (auch) als Technikrecht zu denken, nämlich als Teil der Rechtsnormen, die Innovation, Diffusion und Nutzung oder die Folgen dieser Nutzung i. S. v. Chancen und Risiken,[7] regeln.

[3] Dieses Verständnis zugrundelegend, muss auch bei Haftungsfragen ein sog. *more technological approach*, wie auch für das Immaterialgüterrecht gefordert,[8] angewandt werden. Nach diesem ist insbesondere die Funktion des Rechts als „Einrichtung der Gesellschaft, die über den Erfolg von Technologien und Geschäftsmodellen entscheidet", zu schärfen.[9] Rechtsdogmatische und rechtspolitische Analysen sind einerseits an den Bedürfnissen der technischen Umwelt auszurichten. Andererseits sind diesen Bedürfnissen die Bedürfnisse der Gesellschaft im jeweiligen Kontext gegenüberzustellen, deren Festlegung und Konkretisierung gerade bei Fragen hoher Technizität Herausforderungen darstellen.

III. „Künstliche Intelligenz" als Diskursobjekt

[4] Der Begriff der Künstlichen Intelligenz ist, trotz durchgreifender Bedenken, eine vielgenutzte Umschreibung für einen Kreis neuartiger Technologien, die (scheinbar) ohne menschliches Eingreifen Lösungen für komplexe Problemstellungen erarbeiten, wie sie bislang als nur dem menschlichen Intellekt zugänglich erschienen.[10] Die in diesem Kontext aufkommenden Fragen werden gesamtgesellschaftlich mit Bezug zu diesem KI-Begriff verhandelt. Künstliche Intelligenz besteht also zumindest als Objekt von Diskursen über

* Juniorprofessorin für Bürgerliches Recht, Recht des Geistigen Eigentums und Rechtsfragen der KI an der TU Dresden.

1 *Klement*, Rechtliche Verantwortung, in: Heidbrink et al. (Hrsg.), Handbuch Verantwortung, Wiesbaden 2016, 1.
2 *Scheuner*, Staatszielbestimmungen, in: Schnur (Hrsg.), Festschrift Forsthoff, München 1970, 330, 379; *Wilke*, DÖV 1975, 509.
3 *Dreier*, Verantwortung im demokratischen Verfassungsstaat. Archiv für Rechts- und Sozialphilosophie, Beiheft 74 (2000), S. 9, S. 10; *Steiger*, Verantwortung vor Gott und den Menschen…"?, in: Raffelt (Hrsg.), In Weg und Weite, 2. Aufl., Freiburg im Breisgau 2001, S. 663, S. 675; zur Ubiquität des Begriffs s. *Höfling*, Verantwortung im Umweltrecht – Eine grundrechtsdogmatische Problemskizze, in: Lange (Hrsg.), Gesamtverantwortung statt Verantwortungsparzellierung im Umweltrecht, Baden-Baden 1997, S. 155.
4 *Klement*, S. 2.
5 §§ 823–826, 830 Abs. 1 S. 2, §§ 831, 832, 833 S. 2, §§ 834, 836–838 BGB.
6 Wie § 833 S. 1 BGB; § 7 StVG.
7 *Zech*, Life Sciences and Intellectual Property: Technology Law Put to the Test, ZGE / IPJ 2015, 1, 3.
8 *Grünberger/Podszun*, Ein more technological approach für das Immaterialgüterrecht?, ZGE / IPJ 2014, 269 f.
9 *Grünberger/Podszun*, ZGE / IPJ 2014, 269, 270.
10 Vgl. *Djeffal*, Künstliche Intelligenz, Wiesbaden 2019, 1.

digitale autonome Systeme.[11] Umfragen zu Bekanntheit, Nutzung und Akzeptanz von KI in verschiedenen Lebensbereichen zeigen, dass die Bekanntheit des Begriffs sowie die Selbsteinschätzung zum Wissen hierüber von 2017 bis 2021 deutlich zugenommen hat.[12] Gaben 2017 noch mehr als ein Fünftel der Befragten an, den Begriff noch nie gehört zu haben, waren es 2020 und 2021 nur noch 5 % der Befragten. Nicht nur im öffentlichen Diskursraum, insbesondere in den Massenmedien,[13] werden Aussagen zur KI als Diskursobjekt getätigt. Ein rechtswissenschaftlicher Diskurs zu KI wurde spätestens mit dem Vorschlag der EU-Kommission für einen sog. *Artificial Intelligence Act* angestoßen.[14]

[5] Akzeptiert man nach alledem Künstliche Intelligenz als Objekt der aktuellen Diskurse, so gilt es bei Diskussionen um Regulierung von Verantwortung und Haftung den mit dem Begriff einhergehenden Ungenauigkeiten und Gefahren entgegenzuwirken. Der Begriff vermag sowohl übersteigerte Hoffnungen als auch Ängste zu schüren.[15] Es ist daher von grundlegender Bedeutung, zu verdeutlichen, dass KI-Anwendungen nicht menschliche Intelligenz nachbilden, sondern ihren eigenen Gesetzmäßigkeiten folgen.[16] Der Bedeutungsgehalt der Intelligenz beschränkt sich insofern auf eine komplexe Informations- oder Datenverarbeitung.[17] Ein *humanwashing*[18] von KI-Anwendungen ist gerade für den rechtlichen Umgang mit ihnen nicht zielführend.[19]

IV. Autonomie statt Anthropomorphismus?

[6] Entgegen eines solchen Anthropomorphismus wird verstärkt auf Autonomie abgestellt. Damit gehen aber nicht unähnliche Probleme einher: Auch Konzepte von Autonomie knüpfen herkömmlich an den Menschen an, der insofern anderen Wesen – wie etwa Sachen und Tieren[20] – gegenübergestellt wird. Das Recht kennt den Begriff der Privatautonomie, der das Recht der einzelnen Person, seine Rechtsverhältnisse nach dem eigenen Willen zu gestalten, bezeichnet.[21] Die Reichweite der gewährleisteten Privatautonomie, abgeleitet vom Menschenbild der natürlichen Freiheit, ist aber nicht absolut bestimmbar, sondern muss insbesondere mit der Wahrnehmung ebendieser durch andere Personen in Einklang gebracht werden. Autonomie im Recht muss daher immer neu gedacht werden,[22] und zwar unter Berücksichtigung der jeweiligen außerrechtlichen Bedingungen,[23] wie insbesondere technischer Entwicklungen.

[7] Es kommen auch nicht alle Menschen in den vollen Genuss der Privatautonomie. Viele Personengruppen werden vielmehr vor ihren eigenen Entscheidungen geschützt, etwa Verbraucher, Kinder und Betreute. Das Abstellen auf Autonomie führt auch nicht eindeutig zum Ausschluss von Tieren oder Robotern mit KI. Gerade die Tierhalterhaftung findet ihren Grund in dem spezifischen, für den Menschen insoweit unkontrollierbaren, „selbsttätigen willkürlichen Verhalten des Tieres".[24] Das Reichsgericht beschrieb diese Tiergefahr als „gefährlichen Ausbruche der tierischen Natur, in der von keinem vernünftigen Wollen geleiteten Entfaltung der tierischen organischen Kraft, in der selbständigen Entwicklung einer nach Wirkung und Richtung unberechenbaren tierischen Energie".[25] Wohnt dem Einsatz von KI eine vergleichbare Gefahr inne? Oder ist der Einsatz dieser eher dem Einsatz eines Tieres unter menschlicher Leitung gleichzusetzen, für den dann die Verwirklichung einer spezifischen Tiergefahr verneint würde? Die Beantwortung dieser Fragen kann nur mit einem *more technological approach* sinnvoll gelingen, der Verantwortung und Haftung für KI an den tatsächlichen Fähigkeiten und Risiken von KI-Anwendungen in verschiedenen Formen und Kontexten ausrichtet. Außerrechtliche Bewertungen von Chancen und Risiken sind dann durch rechtliche Normierung von Verantwortung und Haftung in einer Art und Weise „scharf zu stellen",[26] die ihren Folgen für die Innovation, Diffusion und Nutzung von KI Rechnung trägt. ■

11 S. zu Diskursen als Praktiken, „die systematisch die Gegenstände bilden, von denen sie sprechen" *Foucault*, Archäologie des Wissens. Frankfurt am Main 1981, insbes. 74.
12 Repräsentative Umfragen von *Bitkom Research* im Auftrag von Bitkom – Bundesverband Informationswirtschaft, Telekommunikation und neue Medien seit 2017.
13 Vgl. zu durch die Massenmedien geschaffenen Kommunikationsraum zwischen allen Bürgern *Martinsen*, Öffentlichkeit als „Mediendemokratie" aus der Perspektive konkurrierender Demokratietheorien, in: Marrcinkowski et al. (Hrsg.), Politik in der Mediendemokratie, Wiesbaden 2009, 37 f.
14 COM (2021) 206 final.
15 *Herberger*, NJW 2018, 2825-2829; *Jandach*, Juristische Expertensysteme, 1993, 5.
16 Vgl. *Salles/Evers/Farisco*, Anthropomorphism in AI, AJOB Neuroscience, 88, 93.
17 *Spiekermann*, SZ 24.11.2018, 15.
18 S. *Scorici et al.*, Anthropomorphization and beyond: conceptualizing humanwashing of AI-enabled machines. AI & Soc (2022), https://doi.org/10.1007/s00146-022-01492-1.
19 *Hilgendorf*, Robotik, Künstliche Intelligenz, Ethik und Recht. Neue Grundlagenfragen des Technikrechts, in: *Hetnschel* et al, Mensch – Technik – Umwelt, Baden-Baden 2020, 545, 549.
20 Denen mit § 90 a BGB und Art. 20 a GG eine Sonderstellung gegenüber Sachen zukommt.
21 S. nur *BVerfG* NJW 1994, 36, 38; für das Bürgerliche Recht s. *Flume*, Allgemeiner Teil des Bürgerlichen Rechts: Das Rechtsgeschäft, Berlin/Heidelberg 1979, 1; für das Europäische Privatrecht s. *von Bar et al*, Study Group on a European Civil Code and Research Group on EC Private Law: Principles, Definitions and Model Rules of European Private Law – Draft Common Frame of References, 2009, 13, 60 ff.
22 S. *Bumke*, Privatautonomie, in: Bumke/Röthel (Hrsg.), *Autonomie im Recht*, Tübingen 2017, 53; Hilgendorf, 549.
23 *Riesenhuber*, *ZfPW* 2018, 352, 356 (Privatautonomie als „Rechtsprinzip oder ‚mystifizierendes Leuchtfeuer'"); *Röthel*, Forschungsgespräche über Autonomie im Recht, in: Bumke/Röthel (Hrsg.), *Autonomie im Recht*, Tübingen 2017, 51 f.
24 RGZ 80, 237, 238 f.; RGZ 141, 406 (407); BGH NJW 1971, 509.
25 RGZ 80, 237, 238 f.; RGZ 141, 406 (407).
26 Klement, Jan Henrik, Rechtliche Verantwortung, L. Heidbrink et al. (Hrsg.), Handbuch Verantwortung, Springer Reference Sozialwissenschaften, DOI 10.1007/978-3-658-06175-3_30-1, Springer Fachmedien Wiesbaden 2016, S. 2.

Prof. Dr. rer. nat. Wolfgang Koch*

Verantwortbarkeit als technisches Designprinzip künstlich intelligenter Maschinen

Überraschenderweise wurzelt die gedankliche Präzisierung des Begriffes der „Unsicherheit von Daten“ und die „algorithmische“ Unterstützung im Umgang mit ihr im Recht. „Subjektive Wahrscheinlichkeiten“ und ihre Verknüpfung, Bayesian Reasoning, kennt bereits die rabbinische Rechtsprechung des 12. und 13. Jahrhunderts.[1] Wie geht man mit unsicheren Zeugenaussagen methodisch sauber um? Wie akkumuliert man Wahrscheinlichkeiten, um ein sichereres Urteil zu gewinnen, die Grundlage für eine angemessene Entscheidung? Diese Fragen der rabbinischen Juristen, lange vor dem presbyterianischen Pfarrer Thomas Bayes (1702-1761) gestellt, bleiben zeitlos.

[1] Im Zeitalter der Digitalisierung wächst aus diesen Wurzeln ein mächtiger Baum, der nicht nur, aber auch die Gewährleistung der äußeren und inneren Sicherheit revolutioniert. Künstlich intelligente Maschinen werden die Aufklärungs- und Waffensysteme der Bundeswehr prägen, die Deutschlands Souveränität und die der Bündnispartner verteidigen, ihre Bürger und die Staatsgebiete schützen sowie Staaten und Gesellschaften widerstandsfähig gegen äußere Bedrohungen halten. Überwachungstechnik öffnet jedoch auch für Sicherheitsbehörden neue Wege zur Gefahrenabwehr im Inneren.

[2] Gerade nach dem 24. Februar 2022 muss auch die deutsche „Zivilgesellschaft“ die militärischen Möglichkeiten, aber auch die ethischen und rechtlichen Risiken künstlich intelligenter Maschinen verstehen und in ihre politische Meinungsbildung einbeziehen – und dies nicht nur, um drohende Gefahren abwehren zu können, sondern überdies, um Einfluss auf die weitere Entwicklung zu behalten, gerade auch in den zahlreichen zivilen Anwendungen künstlich intelligenter Maschinen.[2] Denn am „scharfen Ende“ der Digitalisierung zeigen sich ihre generellen Probleme wie in einem Brennglas.

I. Gesteigerte Wahrnehmung, vergrößerte Wirkungskreise

[3] Mit derartiger Technik tut sich unsere Gesellschaft schwer. Wer nach China blickt, erschrickt; denn bald ist jeder Chinese sensoriell „getrackt“, datentechnisch erfasst und durch *social credits* maschinell bewertet. Natürlich führt dies zu mehr Sicherheit, aber was geschieht dann mit unseren bürgerlichen Freiheiten? Zwar scheinen die Debatten um bewaffnete Drohnen beendet. Aber zeigten sie nicht, wie wenig der Widerstand gegen die Wiederbewaffnung der 1950er Jahre gesellschaftspolitisch nach so vielen Jahrzehnten überwunden war? Ist er es jetzt? Was geschieht, wenn sich die Schockstarre des 24. Februar 2022 löst?

[4] Künstlich intelligente Maschinen gestalten nicht nur die Gefahrenabwehr tiefgreifend um, sondern bergen eigene Gefahren und wecken Ängste, die nicht nur unbegründet sind. Es bedarf nüchterner Debatten über die Allgegenwart vernetzter Sensoren und der Welt der Algorithmen, die aus Datenströmen maschinen- und menschenverwertbare Information „fusionieren“ und automatisch durchgeführte Aktionen auslösen.

[5] „Kognitive“ und „volitive“ Maschinen steigern bewusstes menschliches Wahrnehmen weit über natürliches Maß hinaus und transformieren absichtlich getroffene Entscheidungen menschlichen Wollens in automatisierte Befehlsketten, ohne die wir in der komplexen Technosphäre, die uns nicht nur im Bereich der Verteidigung immer dichter umschließt, kaum noch verantwortlich wirken können.

II. Ziele künstlich intelligent unterstützten rechten Handelns

[6] Wer über Ethik und Recht im Kontext künstlich intelligenter Maschinen nachdenkt, muss sich der Ziele rechten Handelns bewusstwerden, das solcherart unterstützt wird. Für den Philosophen Robert Spaemann (1927-2018) bestehen ‚künstliche Dinge', also insbesondere künstlich intelligente Maschinen, „aus einem Woraus und einem Was. Ihr Wie und Wozu liegt nicht in ihnen, sondern im Menschen“. „Naturdinge“ tragen ihr inneres Ziel in sich; in ihnen „fallen ihr Was und Wozu in ihnen selbst in eins“.[3]

[7] Nur wer weiß, wo er herkommt, und sein Ziel kennt, kann rechte Wege wählen. Was leitete den bundesdeutschen Kanzler, der Westdeutschland in die NATO führte? Seine Gründe sind zeitlos und prägen auch „das Wie und Wozu“ künstlich intelligenter Waffensysteme. Die NATO sei eine Gemeinschaft freier Nationen, entschlossen, „das gemeinsame Erbe der abendländischen Kultur, die persönliche Freiheit und die Herrschaft des Rechts zu verteidigen“, unterstrich Konrad Adenauer (1876-1967) an jenem 9. Mai 1955. Daher entsprächen ihre Ziele „angesichts der politischen Spannungen in der Welt vollständig den natürlichen

* Fellow IEEE, Universität Bonn, Fraunhofer FKIE, Wachtberg.

1 N. Rabinovitch, *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*. University of Toronto Press, 1973.

2 N. Bossong, A. Rieks, W. Koch, *Künstliche Intelligenz für die Landesverteidigung*, Frankfurter Allgemeine Zeitung, 31.1.2022, Nr. 25, S. 18.

3 R. Spaemann u. a., *Natürliche Ziele. Geschichte und Wiederentdeckung teleologischen Denkens*, Stuttgart: Klett-Cotta (2005), 51 f.

Interessen des deutschen Volkes, das sich […] wie kaum ein anderes Volk nach Sicherheit und Frieden sehnt". Gemeinsame Verteidigung müsse jedoch eingebettet sein in „die Förderung der allgemeinen Wohlfahrt der Völker und zur Bewahrung ihres gemeinsamen Kulturerbes zu einer Zusammenarbeit in wirtschaftlichen und kulturellen Fragen". Deutschland werde „alle seine Kräfte darauf verwenden, dass die menschliche Freiheit und die menschliche Würde erhalten bleiben".[4]

III. Verantwortbarkeit als technisches Designprinzip

[8] Als Konzeptionär zog Wolf von Baudissin (1907-1993) für „Adenauers Armee" in diesem Sinne nicht nur Lehren aus einem Verbrecherstaat, sondern einem „totalen Krieg", der von Hochtechnologie geprägt war: „Das aufs höchste technisierte Gefecht verlangt, dass die Verantwortung an sehr vielen unteren Stellen gesehen und getragen wird", formuliert er schon 1954, also zwei Jahre vor der Dartmouth Summer School, die den Begriff ‚künstliche Intelligenz' prägte. „Daher muss alles getan werden", setzt er fort, „um den Menschen vor Situationen zu stellen, die seine Verantwortung herausfordern und ihn die Folgen von Tun und Unterlassen erleben lassen."[5]

[9] Verantwortbarkeit ist fundamentaler als etwa *Human-in-* oder *-on-the-Loop*; denn auch Automation von Waffensystemen kann verantwortbar sein, wenn Reaktionszeiten für Menschen zu kurz oder die Datenfülle zu groß sind. Daher müssen künstlich intelligente Maschinen gerade dann technische Beherrschbarkeit und verantwortungsvollen Einsatz gewährleisten. Auch dort muss der Mensch eingebunden sein, nicht nur durch die Entscheidung, derartige Maschinen zu nutzen, sondern sie so zu entwerfen, dass ihr Einsatz verantwortbar bleibt.

[10] Zugleich ist im Einsatz Wissen zu vermitteln, durch das Menschen einerseits automatisierter Entscheidungsunterstützung vertrauen können und andererseits die Grenzen kennen, die jede derartige Assistenz besitzt. Insbesondere müssen auf künstlich intelligenten Maschinen basierende Erkenntnisse erklärbar sein. Zugleich ist zu verhindern, dass menschliche Operateure kritiklos Handlungsempfehlungen ohne eigenes Abwägen bestätigen.

IV. Ethical AI Demonstrator für das Future Combat Air System

[11] Erstmals in Deutschland begleitet gedankliches Ringen um die technische Umsetzung ethischer und rechtlicher Prinzipien ein militärisches Großprojekt von Beginn an. Ziel der Arbeitsgemeinschaft „Technikverantwortung für ein Future Combat Air System (FCAS)" ist es, Ethik, Recht und politisches Wollen technisch zu operationalisieren (www.fcas-forum.eu).

[12] Um den Einsatz künstlich intelligenter Maschinen gerade in Stresssituationen realitätsnah zu erfahren, wird zurzeit ein *FCAS Ethical AI-Demonstrator* entwickelt. Dieser wird beispielhaft und für konkrete Szenarien ermöglichen, mit künstlich intelligenten Maschinen zu interagieren, die für den militärischen Einsatz entwickelt wurden. So kann eine realitätsnahe Vorstellung von den Möglichkeiten, Grenzen und ethisch-rechtlichen Implikationen dieser Technologie wachsen. Mit Unterstützung der Luftwaffe entstanden die zu Grunde liegenden Szenarien. Da dieses Vorgehen ein Novum ist, besitzt es naturgemäß experimentellen Charakter. Ohne ein Menschenbild jedoch, das verantwortlichen Technikgebrauch ermöglicht, den „Staatsbürger in Uniform", ist jede digitale Assistenz für moralisch akzeptable Entscheidungen fraglich.

[13] Der *FCAS Ethical AI-Demonstrator* wird bei der weiteren Fähigkeitsentwicklung der Luftwaffe Aspekte klären, um verantwortete Beherrschbarkeit durch ein entsprechend ausgelegtes Systemdesign zu garantieren, die beispielhaft genannt seien:

1. Der oft genannte Begriff *meaningful human control* ist weiter zu fassen als es die gängigen Konzepte nahelegen. Fundamentaler ist der Begriff „Verantwortbarkeit". Denn der Einsatz vollautomatischer Wirksysteme kann durchaus in bestimmten Situationen verantwortbar, ja geboten sein.
2. Zulassung und Qualifizierung sind Schlüsselthemen für FCAS. Robuste KI-basierte Systeme werden sowohl datengetriebene als auch modellbasierte Algorithmen umfassen. Systemisch wären datengetriebene Algorithmen durch modellbasiertes *reasoning* „einzuhegen" – *AI in the Box*.
3. Nachzuweisen sind vorhersagbare Systemeigenschaften, Insensibilität gegenüber unbekannten Einflüssen, Adaptivität gegenüber variablem Einsatzkontext und *graceful degradation*. Statistische Test- und Charakterisierbarkeit sind ebenso wie *explainability* bei kritischen Komponenten Voraussetzungen für Zulassung und Qualifizierung. Hinzukommt systemimmanente *Compliance to a code of conduct*.

[14] Umfassende Analysen zur technischen Beherrschbarkeit und persönlichen Verantwortlichkeit sind bei digitalisierungsdominierten Rüstungsprojekten zwingend und müssen soweit wie möglich öffentlich sichtbar, transparent und überprüfbar durchgeführt werden. Andernfalls wären die gravierenden Paradigmenwechsel und materiellen Aufwendungen politisch, gesellschaftlich und finanziell kaum durchsetzbar.

V. Fazit

[15] Die wehrtechnische Digitalisierung wird die Verteidigung Europas prägen, da die militärische Technosphäre ohne die Welt der Algorithmen unbeherrschbar ist. Ganz im Sinne ihrer konzeptionellen Identität liegt die Bedeutung der Künstlichen Intelligenz für die Bundeswehr auch gemäß ministeriellen Texten „nicht in der Entscheidung Mensch oder KI, sondern in einer effektiven und skalierbaren Kombination von Mensch und KI, um eine bestmögliche Aufgabenerfüllung zu gewährleisten".[6]

[16] Bei der Nutzung künstlich intelligenter Maschinen verdichten sich technische, ethische und rechtliche Heraus-

4 K. Adenauer, *Aufnahme der Bundesrepublik Deutschland in die NATO*, Paris: Palais de Chaillot, 9.5.1955.
5 W. v. Baudissin, *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr*, München: Piper Verlag (1969), S. 234.
6 *Erster Bericht zur Digitalen Transformation*, Bonn: BMVg (2019), S. 27.

forderungen der KI, die sich generell stellen. Daher ergeben sich auch für die wehrtechnische Forschung neuartige Aufgaben.[7] Das technische Design derartiger Maschinen muss dabei eine militärische Kernforderung erfüllen: „Kennzeichnende Merkmale militärischer Führung sind die persönliche Verantwortung militärischer Entscheider und die Durchsetzung ihres Willens in jeder Lage", wie die „Konzeption der Bundeswehr" 2018 bekräftigt.

[17] Verantworteter Waffeneinsatz und effektive „Wirkung im Ziel" schließen sich nicht aus. Entscheidend ist die Frage: Welcher normative Rahmen trägt den sicherheitspolitischen und operativen Realitäten des 21. Jahrhunderts Rechnung und ermöglicht zugleich die Nutzung künstlich intelligenter Maschinen gemäß dem „gemeinsamen Erbe der abendländischen Kultur"? ■

7 W. Koch, *Zur Ethik der wehrtechnischen Digitalisierung – Informations- und ingenieurwissenschaftliche Aspekte*, in M. Rogg u. a. (Hrsg.), *Ethische Herausforderungen digitalen Wandels in bewaffneten Konflikten*, Hamburg: GIDS (2020), S. 17 ff.