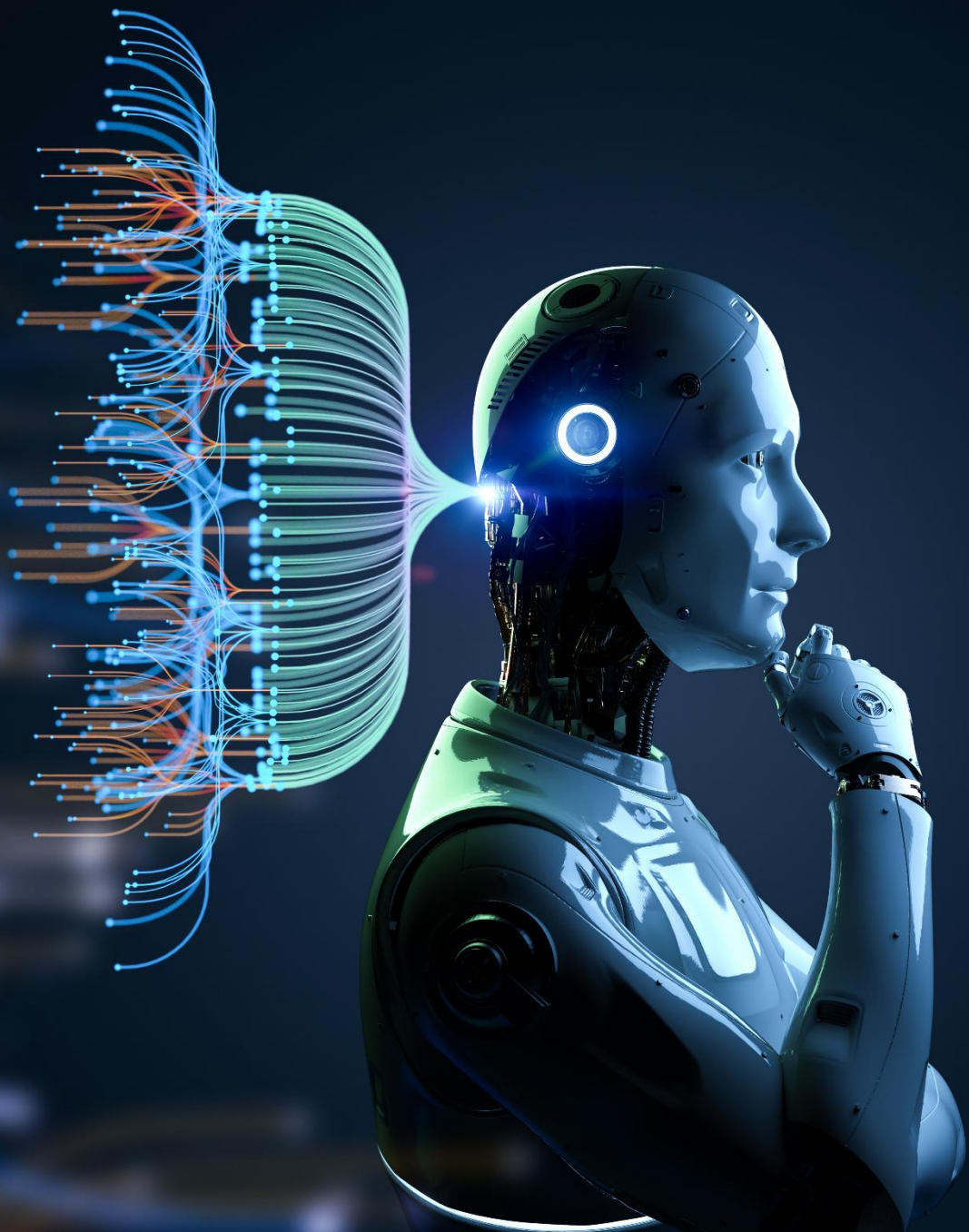




**Beitrag zu Grundlagen und Grenzen  
Künstlicher Intelligenz (KI)  
beim Einsatz in Verteidigungstechnologien**

**2. Impulspapier**

**September 2025**



## Beitrag zu Grundlagen und Grenzen Künstlicher Intelligenz (KI) beim Einsatz in Verteidigungstechnologien – zweites Impulspapier

0. Executive Summary
- I. Einleitung und Einordnung
- II. Besondere Verantwortbarkeits-Herausforderungen eines KI-basierten Einsatzes von Waffensystemen im Rahmen „weitreichender Verteidigung“
- III. Rechtliche und ethische Determinanten militärischer KI-Nutzung
- IV. Aspekte eines ethisch verantwortbaren Umgangs mit militärischer KI im Rahmen von „weitreichender Verteidigung“
- V. Fazit und Empfehlung

---

### 0. Executive Summary

„Gewinnen wollen. Weil wir gewinnen müssen“ lautet das oft wiederholte Credo des Generalinspektors der Bundeswehr, General Carsten Breuer.<sup>1</sup> Vor diesem Hintergrund, dass Verteidigungsfähigkeit Siegfähigkeit voraussetzt, leisten die Autoren mit dem vorliegenden Dokument einen weiteren Debattenbeitrag zu Grundlagen, aber auch den Grenzen - d.h. Rahmenbedingungen - des Einsatzes von Künstlicher Intelligenz (KI) in Waffensystemen „unserer“ Streitkräfte. „Ziel der Bundesregierung ist [...], ein in der Breite unserer Gesellschaft verankertes Verständnis von Integrierter Sicherheit zu entwickeln. Es gilt, gemeinsam zu erfassen, was zum Erhalt und zur Stärkung der Sicherheit und Freiheit Deutschlands getan werden muss, damit unser Land und unsere Politik wehrhaft, resilient und nachhaltig sind.“<sup>2</sup>

Der Impuls richtet sich an Adressaten in Politik und Öffentlichkeit, die sich zu diesen Rahmenbedingungen äußern, den Diskurs prägen und darüber ggfs. zu entscheiden haben. Dabei setzen sich die Autoren für eine Einordnung ein, die es unseren Streitkräften ermöglicht, jeden Angreifer und dessen Waffen wirksam zu bekämpfen, und zwar eingebettet in die Grundsätze der NATO und die ethischen und rechtlichen Maßstäbe, wie wir sie als Gesellschaft zusammen mit unseren Verbündeten anerkennen und umsetzen.

Während das erste Impulspapier des Arbeitskreises „KI & Verteidigung“ im wesentlichen KI-basierten Kombattanten-Schutz zum Gegenstand hatte, konzentriert sich das vorliegende zweite Impulspapier auf die schwieriger zu beurteilenden Fälle des Einsatzes von KI im Rahmen teil- oder vollautomatisierter Waffensysteme, die dem Menschen auf dem Gefechtsfeld mehr als eine bloße Entscheidungshilfe verschaffen können. Ausgangspunkt für die ethische Einordnung ist die Beschreibung des Verhältnisses von Mensch und Maschine in den graduellen

---

<sup>1</sup> Vgl. Grundsatzrede vom 18.07.2023, abgerufen am 14.01.2025, von <https://www.bmvg.de/de/aktuelles/generalinspekteur-beschreibt-bundeswehr-der-zukunft-5652978>

<sup>2</sup> Bundesregierung. (2024). Nationale Sicherheitsstrategie: Wehrhaft. Resilient. Nachhaltig. Integrierte Sicherheit für Deutschland. Berlin: Bundesregierung. Abgerufen am 11.12.2024, von <https://www.nationalesicherheitsstrategie.de/Sicherheitsstrategie-DE.pdf>. S. 73.

Aufgabenverschiebungen durch Teil- oder Vollautomation durch entsprechend programmierte Algorithmen.<sup>3</sup> Hiermit werden drei Fragestellungen aufgeworfen:

- a) In welchen Bereichen ist die Maschine leistungsfähiger als der Mensch?
- b) Welcher menschlichen Kontrolle bedarf die Maschine in dem ihr übertragenen Rahmen („Verantwortbarkeit“ bzw. „Meaningful Human Control“)?
- c) Wie muss die Konstellation der Zusammenarbeit zwischen Mensch und Maschine aussehen?<sup>4</sup>

Vor dem Hintergrund dieser Fragen müssen ethische Kriterien als technische Gestaltungsprinzipien realisiert werden und die KI-nutzende Waffentechnologie von Anfang an prägen. Dabei ist

- erstens die unbedingte Einhaltung von Völkerrecht und Einsatzgrundsätzen auch technisch zu ermöglichen und durch entsprechende Ausbildung und Training aller an der Nutzung beteiligten Akteure zu flankieren,
- zweitens der durch KI ermöglichte und von Menschen tatsächlich beabsichtigte präzise Einsatzerfolg ein positiv zu bewertendes moralisches Gut in konsequentialistischem Sinne. Schließlich realisieren soldatische Tugenden<sup>5</sup> das Konzept des „Staatsbürgers in Uniform“<sup>6</sup> und erhalten durch verantwortlichen Umgang mit KI in Waffensystemen soldatische und damit menschliche Würde.

Klar ist: Menschen tragen Verantwortung (engl. Accountability); nur sie können verantwortlich handeln. Maschinen erzielen Wirkungen. Die uneingeschränkte Handlungsverantwortung dafür (engl. Responsibility) liegt also beim Menschen,<sup>7</sup> der sie im Verteidigungsfall entsprechend den dafür maßgeblichen völkerrechtlichen Rahmenbedingungen auszuüben hat. Daher unterliegt das Wirksamwerden der Maschine der menschlichen Verantwortung.

Die völkerrechtlichen Rahmenbedingungen umfassen sowohl objektiv bestimmbare Restriktionen (wie insbesondere die Unterscheidung zwischen Kombattanten und Zivilisten), als auch wertende Kategorien, wie etwa das Übermaßverbot.<sup>8</sup> Soldatische Ethik und Einsatzrecht im Kontext militärisch genutzter KI müssen insbesondere den jeweiligen Einsatzgrundsätzen genügen, d.h. gewissen „Hard-Constraints“, die zu befolgen sind. Verantwortung für den Einsatz von KI-unterstützten Waffensystemen fängt jedoch nicht beim ausführenden Menschen (in dem Fall Soldatin oder Soldaten)

---

<sup>3</sup> Sauer, Frank: Drei Thesen zur nationalen Regulierung von Autonomie in Waffensystemen, in: Lammert, Norbert, Koch, Wolfgang (Hrsg.), Bundeswehr der Zukunft – Verantwortung und Künstliche Intelligenz, Konrad-Adenauer-Stiftung, Berlin 2023

<sup>4</sup> Quelle: H. W. Meerveld · R. H. A. Lindelauf · E. O. Postma · M. Postma, „The irresponsibility of not using AI in the military“, Ethics and Information Technology (2023) 25:14, abrufbar unter: <https://doi.org/10.1007/s10676-023-09683-0>

<sup>5</sup> Bundeswehr, „Werte und Normen der Bundeswehr“, abrufbar unter: <https://www.bundeswehr.de/resource/blob/5357866/491bd770fb81dd95a23935f52ba83d01/werte-und-normen-web-data.pdf>, zuletzt aufgerufen am 23. Januar 2025.

<sup>6</sup> S. „Das Konzept der Inneren Führung“, abgerufen am 14.01.2025, von <https://www.bmvg.de/de/themen/verteidigung/innere-fuehrung/das-konzept>

<sup>7</sup> Dieser Grundsatz ist in mehreren Bundeswehr-Grundsatz-Dokumenten verankert.

<sup>8</sup> Singer, Tassilo: KI im operativen Kontext: Zur technischen Übertragbarkeit von Regeln des humanitären Völkerrechts auf Künstliche Intelligenz, in: NJW Sonderausgabe 2022: Künstlich-intelligente Maschinen, S. 7 ff

an,<sup>9</sup> sondern hört dort auf. Verschiedenste Bedarfsträger im Planungsprozess - wie Wissenschaftler, Forschende aus der Entwicklung, Planende und Ausführende militärischer Operationen in der militärischen Entscheidungskette - haben jeweils ihren eigenen Anteil an der Verantwortung. Diese nehmen sie wahr, indem sie dafür sorgen, dass der, der die letzte Entscheidung über den Einsatz einer KI-unterstützten Automation trifft, hinreichend ausgebildet ist, um die Effekte und Limitierungen im Kontext der aktuellen Lage zu beurteilen.

Auch wenn zu unterstellen ist, dass mögliche Gegner KI-gestützte Technologie ohne ethische Restriktionen einsetzen, bleibt für uns die Verantwortbarkeit KI-unterstützten Waffeneinsatzes und die Würde des Soldaten oder der Soldatin Eckpfeiler einer rechtlichen und ethischen Beurteilung. Dabei rückt „Verantwortung“ als Schlüsselbegriff bewusstes Wahrnehmen und absichtliches Wollen der handelnden Personen ins Blickfeld.

## I. Einleitung und Einordnung

Spätestens seit dem 24. Februar 2022 muss die deutsche Gesellschaft wieder lernen, wie wenig selbstverständlich Sicherheit ist. Auf ihr fußt u.a. das Gemeinwohl; ohne Sicherheit sind alle anderen individuellen, sozialen, kulturellen, ökonomischen oder ökologischen Güter unerreichbar. Wenn die NATO „bereit, willens und in der Lage“ ist, „jeden Zentimeter des verbündeten Territoriums zu verteidigen“, wie der NATO-Gipfel 2024 in Washington bekräftigte,<sup>10</sup> umfasst „zeitgemäße Landes- und Bündnisverteidigung“ glaubhaft abschreckende und gesellschaftlich mitgetragene Fähigkeiten, sich gegen einen Aggressor verteidigen zu können. Da potentielle Gegner ebenfalls mit Hochdruck an diesen Technologien arbeiten, werden auch unsere Waffensysteme künstlich intelligente Automation nutzen müssen, um taktische und operative Aufträge der Streitkräfte wirksam und präzise erfüllen zu können. Vorliegend geht es um die effektive Abschreckung eines möglichen Aggressors durch das Vorhalten der Fähigkeiten zur Verteidigung im Rahmen eines völkerrechtlich und grundgesetzlich legitimierten Mandats. Ziel ist, dass potentielle Gegner den Frieden wahren und auf Aggression verzichten, da sie in einer solchen keinen Vorteil für sich erkennen.

Mit Blick auf die Würde des Menschen, sowohl der Zivilbevölkerung wie auch der verteidigenden Bürger in Uniform, ist es nicht verantwortbar, einen Abnutzungskrieg führen und erdulden zu müssen, den wir aufgrund der massiven zahlenmäßigen Unterlegenheit auch nur sehr begrenzt durchhalten können.<sup>11</sup> Daraus folgt, dass die zunehmende Ausweitung von Methoden und Anwendungen der KI<sup>12</sup> sowie der Robotik ein wesentlicher Hebel und unerlässlich sind, um eine wirksame Landes- und Bündnisverteidigung (LV/BV) zu ermöglichen. „Landes- und Bündnisverteidigung ist

---

<sup>9</sup> Vgl. Art. 86 (2), 87 (1) ZP I

<sup>10</sup> Bundesministerium der Verteidigung: NATO-Gipfel 2024 in Washington – Auf Deutschland ist Verlass, 12.07.2024. Online: <https://www.bmvg.de/de/themen/dossiers/die-nato-staerke-und-dialog/nato-gipfel-2024-washington>.

<sup>11</sup> Alex Vershinin, The Attritional Art of War: Lessons from the Russian War on Ukraine, RUSI Commentary, 18 March 2024, abgerufen am 14.01.2025, von <https://rusi.org/explore-our-research/publications/commentary/attritional-art-war-lessons-russian-war-ukraine>

<sup>12</sup> Bundesministerium der Verteidigung. (2023). Verteidigungspolitische Richtlinien 2023: Zeitenwende gestalten – Die Bundeswehr der Zukunft. Berlin: BMVg. Abgerufen am 11.12.2024, von <https://www.bmvg.de/resource/blob/5701724/5ba8d8c460d931164c7b00f49994d41d/verteidigungspolitische-richtlinien-2023-data.pdf>. S.11.

Kernauftrag der Bundeswehr; dieser umfasst auch unseren Beitrag zur Abschreckungsfähigkeit der Allianz.“<sup>13</sup> Im Rahmen LV/BV – definiert durch einen Fall nach Art. 5 NATO-Vertrag – sind auch offensive Tätigkeiten des Verteidigers – nach oder bei einem bevorstehenden Angriff – legitimiert. Dabei wird das Gebiet des Angreifers als Handlungsraum nicht ausgenommen. Dies legitimiert Angriffe auf gegnerische militärische Ziele, wie u.a. Kommandostrukturen, Infrastruktur, Waffendepots etc. Entsprechend wird die Nutzung KI-basierter Waffen für charakteristische Fälle auch weitreichender oder fernwirkender militärischer Operationen (im Folgenden vereinfachend als „weitreichende Verteidigung“ bezeichnet) relevant, wie z.B. Loitering Munition, weitreichende Feuer in Form von herkömmlichen Marschflugkörpern aber auch modernerer unbemannter (teil)autonomer Systeme oder Luftkampfsysteme zur Unterdrückung gegnerischer Luftabwehr.

Fundamental ist die Frage, wie Menschen mit Waffensystemen verantwortlich zusammenarbeiten können, die nicht mehr ferngesteuert, sondern durch KI (teil-) automatisiert sind. Nach Aussage des Bundesministeriums der Verteidigung (BMVg) liegt die Bedeutung von KI für die Bundeswehr „nicht in der Wahl zwischen menschlicher oder künstlicher Intelligenz, sondern in einer effektiven und skalierbaren Kombination von menschlicher und künstlicher Intelligenz, um die bestmögliche Leistung zu gewährleisten.“<sup>14</sup> Auch der Europäische Rat forderte schon früh, „die Zuweisung von Funktionen zwischen Menschen und KI-Systemen“ solle „nach menschenzentrierten Entwicklungsgrundsätzen erfolgen“.<sup>15</sup>

Methoden und Anwendungen der Künstlichen Intelligenz sowie der Robotik erstreben operativen und taktischen Mehrwert für Informations-, Führungs-, Entscheidungs- und Wirkungsüberlegenheit. Wie schon im ersten Impulspapier ausgeführt, existiert keine allgemeingültige Begriffsbestimmung von KI.<sup>16</sup> Als Teildisziplin der Informatik befasst sich KI mit Algorithmen, maschinell realisierten Methoden, um Aufgaben auszuführen, die „normalerweise“ menschliche Intelligenz erfordern. Auch jahrzehntealte Technologie ist in diesem Sinne „KI“.<sup>17</sup> Der mediale Hype um KI überzeichnet bisweilen

---

<sup>13</sup> Bundesregierung. (2024). Nationale Sicherheitsstrategie: Wehrhaft. Resilient. Nachhaltig. Integrierte Sicherheit für Deutschland. Berlin: Bundesregierung. Abgerufen am 11.12.2024, von <https://www.nationalesicherheitsstrategie.de/Sicherheitsstrategie-DE.pdf>. S. 13.

<sup>14</sup> Bundesministerium der Verteidigung: Erster Bericht zur Digitalen Transformation, 10/2019, S. 27. Online: <https://www.bmvg.de/resource/blob/143248/7add8013a0617d0c6a8f4ff969dc0184/20191029-download-erster-digitalbericht-data.pdf>.

<sup>15</sup> Europäische Kommission: Leitlinien für eine vertrauenswürdige KI, 08.04.2019, S. 15. Online: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

<sup>16</sup> Vergl.: Europäische Kommission, Eine Definition der KI – Wichtigste Fähigkeiten und Wissenschaftsgebiete, April 2019. Online: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>; hinzuweisen ist ferner auf die Definition für die Zwecke des EU AI Act in Art. 3 der Regulation (EU) 2024/1689 vom 13.06.2024: „KI-System“ ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können“. Online: [https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401689)

<sup>17</sup> „The DoD AI Strategy defines AI as ‘the ability of machines to perform tasks that normally require human intelligence.’ This definition includes decades-old DoD AI.“ Greg, Allan: Understanding AI Technology – A concise, practical, and readable overview of Artificial Intelligence and Machine Learning technology designed for non-technical managers, officers, and executives, Joint Artificial Intelligence Center (JAIC), Department of Defense, USA, April 2020.

Chancen und Risiken gerade auch militärischer KI.<sup>18</sup> Im militärischen Bereich dienen einige Anwendungen der Unterstützung des Menschen bei der Lagebildgewinnung oder der Analyse von Entscheidungsoptionen und Vorbereitung des Einsatzes von Wirkmitteln.<sup>19</sup> Es kommt die KI-unterstützte Umsetzung menschlicher Entscheidungen in automatisiert ausgeführte Befehlsfolgen hinzu. Darüber hinaus wird KI in Mensch-Maschine-Systeme eingebunden und ist deshalb „systemisch“ zu betrachten.

KI-ertüchtigte Unterstützung für Wahrnehmen und Wirken befähigt Soldaten, auch in der Technosphäre künftiger Verteidigung auf kurzen Zeitskalen verantwortlich handlungsfähig zu bleiben. Im Rahmen eines gesellschaftlichen Diskurses wären Fragen zu beantworten wie:

- a) Gemäß welchen Kriterien unterscheiden wir ethische und unethische Handlungen, an denen auch künstlich intelligente Automation in Waffensystemen zu messen wäre?
- b) Wie sind diese Kriterien in geltenden Rechtsnormen abgebildet?
- c) Braucht es die Grundzüge der Inneren Führung ergänzende (ethische) Schulungsmaßnahmen für verantwortliches Handeln in herausfordernden neuen Bedrohungsszenarien wie z.B. in einem beschleunigten *Hyper War*<sup>20</sup>?

Der „Arbeitskreis KI & Verteidigung“, ein von den Verbänden BDI, BDSV, BDLI und Bitkom getragener Kreis von Expertinnen und Experten aus dem Verbands-, Industrie- sowie Forschungs- und Universitätsbereich, hat es sich mit seinen „Impulspapieren“ zur Aufgabe gemacht, Grundlagen und Grenzen von KI beim Einsatz in Verteidigungstechnologien klären zu helfen. Hiermit will er einen praxisbezogenen Beitrag zum Diskurs über Stellenwert und Grenzen des Einsatzes der KI im Bereich militärischer Technologien leisten. Im Oktober 2023 hatte der Arbeitskreis ein erstes Impulspapier (1.0) veröffentlicht<sup>21</sup>, in dem es vor allem um Fragen der Verantwortbarkeit des KI-Einsatzes im militärischen Mensch-Maschine-Kontext beim Eigenschutz von Kombattanten ging.

Das vorliegende zweite Impulspapier konzentriert sich dagegen inhaltlich auf die Verantwortbarkeit des KI-Einsatzes am Beispiel von „weitreichender Verteidigung“, also eines Gegenangriffs des Verteidigers im Rahmen eines nach Art. 51 der UN-Charta gerechtfertigten Verteidigungsfalls gegenüber einem rechtswidrig handelnden Aggressor (entsprechend dem Verteidigungsfall nach Art. 5 des NATO-Vertrages oder nach Art. 42 (7) EUV). Dabei handelt es sich um eine bewusste Zuspitzung im Sinne eines extremen Problems. Sie geschieht in dem Wissen, dass KI im Zusammenhang mit militärisch relevanten Technologien auf operativer und taktischer Ebene in breitem Umfang auch anderweitig zum Einsatz kommt, etwa bei der Analyse- und

---

<sup>18</sup> Sauer, Frank (2022): The Military Rationale for AI, in: Schörnig, Niklas/Reinhold, Thomas (Eds.): Armament, Arms Control and Artificial Intelligence: The impact of software, machine learning and artificial intelligence on armament and arms control, 27-38.

<sup>19</sup> Sauer, Frank 2024: Künstliche in Streitkräften, in: Metis Studien Nr. 40.

<sup>20</sup> Hyperwar – neue Herausforderungen für die Heeresentwicklung, in: Europäische Sicherheit & Technik, 2019, abrufbar unter: <https://esut.de/2019/09/fachbeitraege/streitkraefte-fachbeitraege/14836/hyperwar-neue-herausforderungen-fuer-die-heeresentwicklung/> (letzter Zugriff: 24. Januar 2025).

<sup>21</sup> Handelsblatt vom 16.10.2023: „KI bei der Bundeswehr – Verbände wollen Tabuthema „entdämonisieren“, Online: <https://www.handelsblatt.com/technik/ki/ruestung-ki-bei-der-bundeswehr-verbaende-wollen-tabuthema-entdaemonisieren/29448216.html>

Entscheidungsunterstützung, bei der Datenvernetzung und Lagebilderfassung sowie bei allen Herausforderungen im Zusammenhang mit Umfang, Masse und Geschwindigkeit, von gleichzeitigen Bedrohungen und damit einhergehenden Entscheidungsnotwendigkeiten beim Zusammenwirken von Mensch und Maschine. Hierzu wird – wie schon im ersten Impulspapier – auf die ständig fortschreitenden praktischen Erfahrungen aus dem Ukraine-Krieg<sup>22</sup>, wie auch auf die weitreichende Literatur hierzu<sup>23</sup> verwiesen. Bewusst ausgeklammert ist die strategische Ebene des Einsatzes strategischer Waffen, bei denen es offen ist, inwieweit KI menschliche Entscheidungen sinnvoll ergänzen kann.<sup>24</sup>

Autoren des vorliegenden zweiten Impulspapiers („2.0“) sind die am Schluss aufgeführten Personen, die eine verbandliche, industrielle oder wissenschaftliche Basis haben und sich im Arbeitskreis KI & Verteidigung unter Moderation des Bundesverbands der Deutschen Industrie e.V. (BDI) sowie des Bundesverbands der Deutschen Sicherheits- und Verteidigungsindustrie e.V. (BDSV) zusammengefunden haben.<sup>25</sup> Adressaten des Impulspapiers sind Akteure aus Regierung, Parlament, Bundeswehr, Medien und Interessenvertreter von Nichtregierungsorganisationen. Wie schon das Impulspapier 1.0 vom Oktober 2023 liefert auch der vorliegende zweite Impuls einen konkreten Debattenbeitrag zur Formulierung einer nationalen militärischen KI-Strategie, die angesichts der aktuellen Technologieentwicklung einerseits und der Bedrohungslage für Landes- und Bündnisverteidigung andererseits unerlässlich erscheint.

## II. Besondere Verantwortbarkeits-Herausforderungen eines KI-basierten Einsatzes von Waffensystemen im Rahmen „weitreichender Verteidigung“

Wir nähern uns den Herausforderungen auf zwei Wegen. Erstens durch die Justierung des Blickwinkels und zweitens die Betrachtung der Grenzen und Möglichkeiten von KI. Dabei werden jeweils Lösungsansätze aufgezeigt.

**Blickwinkel:** Eine rein abstrakte Betrachtung der Herausforderungen führt nicht weiter. Was die Einordnung des KI-Einsatzes unter dem Blickwinkel völkerrechtlicher Zulässigkeit und ethischer Verantwortbarkeit überhaupt erst greifbar macht, ist die Betrachtung eines konkreten Kontexts. Hierzu ist die Bildung von Szenarien einer

---

<sup>22</sup> Siehe <https://www.nzz.ch/international/ukraine-krieg-kuenstliche-intelligenz-am-schlachtfeld-id.1695576>; Sandra Lumetsberger, Künstliche Intelligenz auf dem Schlachtfeld: Eine US-Software revolutioniert den Artillerie-Krieg in der Ukraine, Tagesspiegel vom 18.01.2023, abrufbar unter: <https://www.tagesspiegel.de/internationales/kuenstliche-intelligenz-an-der-front-kann-moderne-software-den-krieg-grundlegend-verandern-9160431.html>

<sup>23</sup> Dazu beispielhaft mit weiteren Nachweisen die Beiträge in: Lammert, Norbert, Koch, Wolfgang (Hrsg.): Bundeswehr der Zukunft – Verantwortung und Künstliche Intelligenz, Konrad-Adenauer-Stiftung, Berlin 2023; im Übrigen auch Bundesverband der Deutschen Industrie e.V.: Künstliche Intelligenz in Sicherheit und Verteidigung – Handlungsempfehlungen der deutschen Industrie vom 15.01.2019, abrufbar unter: <https://bdi.eu/publikation/news/kuenstliche-intelligenz-in-sicherheit-und-verteidigung/>

<sup>24</sup> Johnson, James: The AI Commander: Centaur Teaming, Command, and Ethical Dilemmas, Oxford University Press 2024.

<sup>25</sup> Der Arbeitskreis wurde im Jahr 2020 ins Leben gerufen, und zwar unter Beteiligung der Verbände BDI, BDSV, BDLI und Bitkom.

„weitreichenden Verteidigung“ mittels KI-gestützter Waffen als Grundlage zur Betrachtung sinnvoll.

Bezogen auf den jeweiligen Maßnahmen-Typus müssen wiederum typische Szenarien gebildet und nach völkerrechtlichen bzw. ethischen Dimension differenzierend betrachtet werden, wobei sich auch das vorliegende zweite Impulspapier an dem im ersten Impulspapier konstituierten Begriff der „Verantwortbarkeit“ orientiert. Ausschlaggebend ist demzufolge in allen Maßnahmen-Typen erstens der Operationskontext, der darüber entscheidet, wann welche Funktionen an welchen Orten für wie lange von Menschen an Maschinen delegiert werden können. Zweitens wird im vorliegenden zweiten Impulspapier auch reflektiert, welche Grenzen die Einwirkungsmöglichkeiten des Gegners setzen. Denn anders als beim autonomen Fahren besteht in militärischen Kontexten immer das Problem, dass Streitkräfte gerade nicht in einem kooperativen Ökosystem operieren (wie es untereinander und mit ihrer Umgebung vernetzte automatisierte Transportmittel darstellen). Ganz im Gegenteil, sie agieren in einer feindlichen Umgebung, die autonome Systeme mit allen zur Verfügung stehenden Mitteln bekämpfen wird, was „hijacken“, täuschen und zu Fehlern verleiten miteinschließt. Sicherheitsprotokolle und Maßnahmen der Informationssicherheit sind zu implementieren und bereits im Entwicklungsprozess einzubeziehen. Dies verlangt eine noch engere Verzahnung zwischen Bundeswehr, Wirtschaft, Wissenschaft, u.a. in Reallaboren.

Chancen und Möglichkeiten Künstlicher Intelligenz: Wir heben hier auf drei wesentliche Herausforderungen ab: i) den „Black Box“ Charakter der besonders leistungsfähigen daten-getriebenen Modelle, ii) die unvermeidlichen Fehler probabilistischer Methoden und iii) die Risiken durch Tarnung und (Cyber-)angriffe.

Es gibt verschiedene Teilbereiche Künstlicher Intelligenz, die sich hinsichtlich ihrer Grenzen und Möglichkeiten unterscheiden. Es gibt einerseits modellbasierte Ansätze, die A-Priori-Wissen voraussetzen. Ist dieses Wissen „verstanden“, lässt es sich mathematisch sowie durch Taxonomien und Ontologien beschreiben und führt zu modellbasierten Algorithmen. Ist das erforderliche A-Priori-Wissen dagegen zu komplex, um durch derartige Modelle beschrieben zu werden, muss es aus Trainingsdaten „erlernt“ werden. KI-Modelle, die es repräsentieren, entziehen sich vollem menschlichem Verständnis. KI-Systeme, sofern sie nicht rein modellbasiert sind, fußen auf Algorithmen und Trainingsdaten, mit deren Hilfe eine Funktion (das KI-Modell) erzeugt wird.

Ist nicht unmittelbar einsichtig oder nachvollziehbar, wie das KI-Modell zu seinen Ergebnissen kommt, wird landläufig vom „Black Box“-Charakter des KI-Modells gesprochen. Dies betrifft insbesondere Deep Learning Modelle, die sich in vielen Bereichen als besonders leistungsfähig erwiesen haben. Hierbei dennoch Ansätze zu entwickeln, die die Ergebnisse einsichtig oder nachvollziehbar machen, ist Gegenstand der Forschung (Explainable AI) und umfasst bereits ein breites Spektrum an Methodiken und Bibliotheken, die den praktischen Einsatz dieser neuen und besonders leistungsfähigen Modelle ermöglichen.

Eine weitere Herausforderung datengetriebener Algorithmen ist, dass Trainingsdaten nie so vollständig vorliegen, dass sie eine sich verändernde Realität umfassend

beschreiben können. Daher kann es sein, dass das Modell (die Funktion) unerwartete („falsche“) Ergebnisse liefert, wenn sie in der Realität mit Daten konfrontiert wird, die durch die Trainingsdaten nur unzulänglich abgebildet waren. Dieses Problem besteht grundsätzlich in Bezug auf datengetriebene KI. Daraus folgt nicht, dass eine KI aufgrund der Risiken nicht eingesetzt werden darf. Vielmehr gilt es, die Risiken bestmöglich zu mitigieren.

Dazu ist es unerlässlich, die Qualität und Zuverlässigkeit des Modells kontinuierlich zu überwachen und bei Bedarf nachzutrainieren. Dies ist Bestandteil des unerlässlichen Life-cycle Managements von KI.

Darüber hinaus ist es dann aber essentiell, dass der Mensch die Grenzen und Möglichkeiten des ihm an die Hand gegebenen Waffensystems beurteilen und im Kontext einer konkreten Situation bewerten kann. Ein solcher bewusster Umgang mit algorithmisch unterstützter Automation ist in vielen Alltagsbereichen bereits alltäglich. (Beispiel: Ein Anti-Blockiersystem verhindert, dass wir nicht rechtzeitig mit dem Auto zum Stehen kommen. Aber die Straßenverhältnisse (Glätte, Nässe, Aquaplaning) müssen vom Fahrer angemessen berücksichtigt werden, damit er sich und andere Verkehrsteilnehmer nicht gefährdet).

Die verbleibende und (vorläufig) nicht weiter minimierbare Fehlerwahrscheinlichkeit muss der Nutzung derartiger KI-Modelle nicht entgegenstehen (ähnlich der Zulassung von Medikamenten, die auch zu Nebenwirkungen führen können). Wie bei der Qualifizierung von Fluggerät muss das jeweilige KI-unterstützte Gesamtsystem dem Design-Anspruch des „fail-safe/no harm“ genügen. Dabei sind insbesondere alle voraussehbar auftretenden völkerrechtlichen Fragestellungen in korrekter Weise aufzulösen – angefangen beim Design, über das Training bis hin zur Nutzung. Im Ergebnis ist mit größtmöglicher Sicherheit zu erreichen, dass die KI-Systeme in einem definierten Rahmen mit einer mutmaßlich sehr hohen Wahrscheinlichkeit so funktionieren, wie es aufgrund ihrer Programmierung beabsichtigt ist.

Mehr dazu in Abschnitt IV, unten.

Die wesentlichen Herausforderungen im Kontext Cyber-Angriff und Tarnung sind:

- a) KI-Modelle sind angreifbar durch „adversarial attacks“. Kennt ein Gegner das KI-Modell, ist es ihm unter Umständen möglich, durch gezielte Verfälschung von Daten, das KI-Modell in die Irre zu führen, was auf Seiten des Nutzers nicht oder nur sehr schwer zu entdecken ist.
- b) Auch ist es bei der Bilderkennung möglich, z.B. Tarnungen von Objekten durch optische Verfremdungen zu schaffen, durch die sie für eine KI nicht mehr zuverlässig identifizierbar sind.

Dabei ist es plausibel anzunehmen, dass tief in feindlichem Gebiet der Gegner über mehr Möglichkeiten zur Täuschung der KI-Modelle verfügen mag, gegen die das System noch nicht geschützt ist. In vertrautem eigenem Umfeld könnten jedenfalls Daten aus der Vergangenheit wichtige Referenzpunkte liefern, die zu einer Überprüfung und Identifikation von Täuschungen als auch Tarnungen genutzt werden könnten.

Andererseits besteht das Risiko von Adversarial Attacks und anderen Cyber-Security Risiken grundsätzlich für jedes KI-System. Daher gilt es, die erforderlichen Maßnahmen zur Steigerung der Robustheit und Abwehr von Cyber-Angriffen mit aller ohnehin erforderlichen Sorgfalt und Qualität umzusetzen.

Eine wichtige Maßnahme zur Mitigierung von Risiken ist daher der Einsatz mehrerer unabhängiger Sensoren und die Multi-Sensor Datenfusion. Konkret könnten solche zusätzlichen Maßnahmen zur Informationsgewinnung (z.B. durch Drohnenaufklärung im feindlichen Gebiet) und zum Validieren bzw. Nachtrainieren von Modellen dienen. Im Sinne eines „maschinellen Co-Piloten“ dienen demnach KI-basierte Algorithmen insbesondere der Prozessoptimierung und des Risikomanagements, aber auch der Personalisierung auf spezifische Nutzeranforderungen.

### III. Rechtliche und ethische Determinanten militärischer KI-Nutzung

Völkerrecht: Wie schon im ersten Impulspapier sei auch hier betont, dass das humanitäre Völkerrecht insbesondere in den Artikeln 35 und 36 des Zusatzprotokolls I zur Genfer Konvention Regeln zur Beschränkung der Wahl von Mitteln und Methoden der Kriegführung aufstellt.<sup>26</sup> Für die Verpflichtung zur Waffenprüfung nach Art. 36 ZP I gibt es mit Blick auf automatisierte Waffensysteme zwei Kategorien von Regeln. Einerseits werden Regeln umfasst, die – allerdings nur unter der Prämisse einer technisch fehlerfreien Funktionsweise – von automatisierten Systemen unterstützt oder potenziell sogar umgesetzt werden können, wie etwa die Unterscheidung zwischen zulässigen militärischen und nicht zulässigen zivilen Zielobjekten (das Internationale Komitee vom Roten Kreuz spricht in diesem Zusammenhang von einer Beschränkung auf „military objects by nature“).<sup>27</sup> Andererseits verlangt das völkerrechtliche Regelwerk aber auch wertende bzw. abwägende Entscheidungen, wie etwa bei dem in Art. 57 Abs. 2 ZP I enthaltenen Exzessverbot, das die militärischen Entscheidungsträger zum Abbruch eines Angriffs verpflichtet wenn *„damit zu rechnen ist, dass er auch Verluste unter der Zivilbevölkerung, die Verwundung von Zivilpersonen, die Beschädigung ziviler Objekte oder mehrere derartige Folgen zusammen verursacht, die in keinem Verhältnis zum erwarteten konkreten und unmittelbaren militärischen Vorteil stehen“*. Trotz ständig fortschreitender KI-Technikentwicklung gerade auch im Bereich der abwägenden Entscheidungsunterstützung, sind im Rahmen der Entwicklung und Erprobung immer wieder Äquivalenztests i.R. des sog. Weapons Review mit dem Ziel durchzuführen, die entsprechenden KI-Fähigkeiten vor dem Hintergrund der völkerrechtlichen Anforderungen zu validieren.<sup>28</sup> Wegen der Verpflichtung der Staaten nach Art. 36 ZP I, alle neuen Waffen, Mittel und Methoden der Kriegführung rechtlich zu prüfen, um festzustellen, ob ihr Einsatz durch das Völkerrecht verboten ist, müssen daher auch

---

<sup>26</sup> Zusatzprotokoll zu den Genfer Abkommen vom 12. August 1949 über den Schutz der Opfer internationaler bewaffneter Konflikte (Protokoll I) vom 8.6.1977, abrufbar unter: [https://fedlex.data.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/1982/1362\\_1362\\_1362/20180712/de/pdf-a/fedlex-data-admin-ch-eli-cc-1982-1362\\_1362\\_1362-20180712-de-pdf-a.pdf](https://fedlex.data.admin.ch/filestore/fedlex.data.admin.ch/eli/cc/1982/1362_1362_1362/20180712/de/pdf-a/fedlex-data-admin-ch-eli-cc-1982-1362_1362_1362-20180712-de-pdf-a.pdf); im Übrigen Singer, ebda.

<sup>27</sup> Autonomous weapons: ICRC urges states to launch negotiations for new legally binding rules, 05.06.2023, <https://www.icrc.org/en/document/statement-international-committee-red-cross-icrc-following-meeting-group-governmental#:~:text=For%20example%2C%20the%20ICRC%20additionally,human%2Dmachine%20interaction%20to%20ensure>

<sup>28</sup> Singer, ebd., S. 9

Überlegungen zur Technikverantwortung in den Prüfablauf (Weapon Review) miteinbezogen werden.<sup>29</sup> Hingegen hat die von Deutschland vertretene internationale Ächtung sog. autonomer Waffensysteme auf Ebene der Vereinten Nationen bislang keine Aussicht auf Erfolg gezeigt.<sup>30</sup>

Rechtmäßigkeit<sup>31</sup> der „weitreichenden Verteidigung“: Als Randbedingung soll festgehalten werden, dass ethisches militärisches Handeln einem militärischen Auftrag folgt, dessen ethische Qualität hier als solche nicht in Frage stehen soll. Das vorliegende Impulspapier erwägt das *Ius in bello* („Wie“ der Kriegführung), nicht das *Ius ad bellum* („Ob“ der Kriegführung). In einer praktischen Situation gem. Art. 5 NATO-Vertrag gibt es im Einsatz keinen relevanten Unterschied zwischen konventionellen Defensiv- und Offensiv-Waffen. Auch nach Art. 49 Zusatzprotokoll I zu den Genfer Konventionen besteht rechtlich kein Unterschied zwischen offensiver und defensiver Gewaltanwendung. Beides wird als Angriff angesehen. Alle zum Einsatz kommenden Waffen können und werden mit allen denkbaren Anwendungsmöglichkeiten im Einklang mit Art. 51 der UN-Charta und unter Einhaltung der Begrenzungen des humanitären Völkerrechts zum Zweck der Landes- und Bündnisverteidigung eingesetzt.<sup>32</sup> Wie auch bei dem im ersten Impulspapier diskutierten Fall der Überlebensfähigkeit des Nutzers sind auch bei den Szenarien der „weitreichenden Verteidigung“ die konkreten Anwendungsparameter im jeweiligen Einzelfall festgemacht an den Kategorien der rechtlichen und ethischen „Verantwortbarkeit“ bzw. der „Meaningful Human Control“ zu überprüfen.

„Verantwortbarkeit“ und „Meaningful Human Control“ des KI-Einsatzes: Während es sich bei dem Konzept der „Meaningful Human Control“ um ein vor allem im anglo-amerikanischen Raum entwickeltes Gedankenmodell handelt, ist das Konzept der „Verantwortbarkeit“ als Maßstab für das „Einhegen“ von handlungsbezogenen oder technischen Katastrophenrisiken im deutschen und europäischen Ethik-Denken vielfach verankert, u.a. bei der „inneren Führung“ der Bundeswehr<sup>33</sup> oder im Bereich der Umweltethik, bei der Katastrophen-Vermeidung im Falle ziviler Kernkraft-Nutzung.<sup>34</sup> Eine andere, wiederum international gebräuchliche Ausprägung des

---

<sup>29</sup> V. Boulanin, „Implementing Article 36 weapon reviews in the light of increasing autonomy in weapon systems,” in SIPRI Insights on Peace and Security. Solna, Sweden: SIPRI, No. 2015/1, Nov. 2015. [Online]. Available: <https://www.sipri.org/sites/default/files/files/insight/SIPRIInsight1501.pdf>

<sup>30</sup> Human Rights Watch, „Killer Robots: UN Vote Should Spur Action on Treaty“, vom 03.01.2024, Online: <https://www.hrw.org/news/2024/01/03/killer-robots-un-vote-should-spur-action-treaty>

<sup>31</sup> Bonner Völkerrechtsprofessor Matthias Herdegen gegenüber LTO: „Es wäre geradezu absurd, wenn sich ein Aggressorstaat darauf verlassen könnte, ungefährdet aus einer sicheren Zone jenseits der Grenze heraus Angriffe führen zu können und sich immer auf ein sicheres Rückzugsgebiet im eigenen Land stützen zu können. Das widerspräche jeder Logik der Selbstverteidigung.“ Legal Tribune Online, 2023. Ukraine, Russland und die Logik der Selbstverteidigung. Verfügbar unter: <https://www.lto.de/recht/nachrichten/n/ukraine-russland-waffenlieferung-nato-krieg-bundesregierung-selbstverteidigung> (Zugriff am: 11. Dezember 2024).

<sup>32</sup> Wenn eine Gewaltanwendung gegen einen Staat völkerrechtlich durch das kollektive und/oder individuelle Selbstverteidigungsrecht (Art. 51 UN Charta bzw. Völkergewohnheitsrecht) gerechtfertigt ist (ob der Kriegführung) und das humanitäre Völkerrecht eingehalten wird, ist eine Gewaltanwendung nicht rechtswidrig.

<sup>33</sup> Koch, Wolfgang: Verantwortbarkeit als technisches Designprinzip künstlich intelligenter Maschinen, in: NJW Sonderausgabe 2022: Künstlich-intelligente Maschinen, S. 14 ff.

<sup>34</sup> Vgl. u.a. Hülsmann, Heinz, Tschiedel, Robert: Kernenergie und wissenschaftliche Verantwortung, Athenaeum-Verlag, Bodenheim 1977; Altner, Günter: Ist Kernenergie ethisch verantwortbar?,

Konzeptes der Technik-Verantwortbarkeit findet sich in Form des „Fail-safe“-Prinzips im Engineering: *“A design feature or practice that, in the event of a specific type of failure, inherently responds in a way that will cause minimal or no harm to other equipment, to the environment or to people”*.<sup>35</sup>

Zwar verweist auch Meaningful Human Control auf die Menschenwürde als obersten Grundsatz für jegliches – auch militärisches – Handeln. „Verantwortbarkeit“ schließt dies im reflektiven Bereich ein, allerdings erweitert um den Aspekt des gebotenen Selbstschutzes des für die Waffenwirkung verantwortlichen Akteurs. Bezogen auf die Nutzung von KI in Waffen sind sowohl die Denkfigur der „Verantwortbarkeit“ als auch das Konzept der Meaningful Human Control fundamentaler als etwa die Forderung nach einem „Human in the loop“ oder „Human on the loop“. Denn auch Vollautomation – mit dem Menschen „out of the loop“ - kann verantwortlich sein, wenn Reaktionszeiten für Menschen zu kurz oder die Datenfülle zu groß sind, um für sie einen Fail-safe-Zustand zu gewährleisten. Es greift daher zu kurz, immer einen Menschen „in“ oder „on the loop“ halten zu wollen. Selbst bei Vollautomation muss aber der Mensch eingebunden sein, nicht nur durch die Entscheidung, derartige Waffen zu nutzen, sondern sie so zu entwerfen, dass ihr Einsatz bei aller Effektivität verantwortlich bleibt<sup>36</sup>. Mit anderen Worten: Das Gesamtsystem aus Mensch und Maschine kann die Bedingungen für Meaningful Human Control sehr wohl auch dann erfüllen, wenn die Maschine zeitlich und örtlich begrenzt gänzlich ohne menschliches Zutun operiert. Entscheidend ist, dass der Mensch die Aktionen der Maschine nach Aktivierung vorausschauend beurteilen und ihre Operation administrieren kann, so dass im Umkehrschluss Verantwortung zurechenbar ist.<sup>37</sup> Hierfür ist schon im Vorfeld eines Einsatzes dem Soldaten bzw. der Soldatin erstens bei Ausbildung und Training das erforderliche Wissen zu vermitteln und zweitens die automatisierte Entscheidungsunterstützung so für das Waffensystem auszulegen, dass der Mensch beurteilen kann, ob die Automation in der aktuellen Lage verantwortlich ist („control by design“ und „control in use“). Dafür muss er/sie Grenzen der künstlichen Intelligenz in der Automationskette kennen und beurteilen können. Insbesondere müssen KI-basierte Erkenntnisse in diesem Sinne erklärbar sein. Die Komplexität von Prozessen nimmt, u.a. durch immer bessere Sensoren stetig zu. Daten in guter Qualität sind die Grundlage erfolgreicher KI-Anwendungen. Dies erfordert ein intelligentes Datenmanagement, wobei Personalressourcen für Verfahrensbearbeitung jedoch begrenzt sind. Ergebnisse müssen immer schneller bewertet werden und bedürfen daher analytischer und kognitiver Unterstützung beim Prozess der

---

abrufbar unter: [https://link.springer.com/chapter/10.1007/978-3-642-93475-9\\_29](https://link.springer.com/chapter/10.1007/978-3-642-93475-9_29); Zeit-Artikel „Al-Wazir: Weiterbetrieb von

Atomkraftwerken ‚verantwortbar‘“ vom 28.09.2022, abrufbar unter: [https://www.zeit.de/news/2022-09/28/al-wazir-weiterbetrieb-von-atomkraftwerken-verantwortbar?utm\\_referrer=https%3A%2F%2Fwww.google.com](https://www.zeit.de/news/2022-09/28/al-wazir-weiterbetrieb-von-atomkraftwerken-verantwortbar?utm_referrer=https%3A%2F%2Fwww.google.com)

<sup>35</sup> Vgl. Wikipedia zu „Fail-safe“, abrufbar unter: <https://en.wikipedia.org/wiki/Fail-safe>

<sup>36</sup> „KI-Anwendungen können menschliche Intelligenz, Verantwortung und Bewertung nicht ersetzen“, betont Julian Nida-Rümelin, der stellvertretende Vorsitzende des Deutschen Ethikrates lt. einer Pressemitteilung des Deutschen Ethikrates vom 22.03.2023. Online: <https://www.ethikrat.org/presse/mitteilungen/ethikrat-kuenstliche-intelligenz-darf-menschliche-entfaltung-nicht-vermindern/>

<sup>37</sup> Vincent Boulanin, Laura Bruun and Netta Goussac 2021: Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human–Machine Interaction, SIPRI, [https://www.sipri.org/sites/default/files/2021-06/2106\\_aws\\_and\\_ihl\\_0.pdf](https://www.sipri.org/sites/default/files/2021-06/2106_aws_and_ihl_0.pdf)

Entscheidungsfindung. Adäquates Training ist daher eine notwendige Voraussetzung, KI-basierte Waffen verantwortungsvoll einzusetzen.

#### IV. Aspekte eines ethisch verantwortbaren Umgangs mit militärischer KI im Rahmen von „weitreichender Verteidigung“<sup>38</sup>

Im Spannungsfeld der Aspekte eines ethisch verantwortbaren Umgangs mit KI-gestützten Waffen gilt es nun zu betrachten, welchen Einfluss der Grad der Automatisierung auf deren ethisch vertretbaren Einsatz im Rahmen eines Szenarios der weitreichenden Verteidigung hat. Sofern es sich um Aufgaben der Aufklärung handelt, gibt es keine nennenswerten Einschränkungen. Die zentralen Fragen der Verantwortbarkeit beziehen sich daher auf die KI-gestützten Wirksysteme:

- a) Normativ gilt das Völkerrecht<sup>39</sup>, das jeweils bezogen auf den aktuellen Stand der Technik anzuwenden ist. Waffensysteme mit automatischen Funktionen, die bereits im Einsatz waren bzw. es aktuell sind (wie z.B. SAGE, Predator, Patriot oder AEGIS), zeigen auf, wie das Zusammenspiel zwischen menschlicher Entscheidung und Teil-Automation in einem ethisch vertretbaren Rahmen möglich ist. Der Kern dabei ist, dass der Mensch die Entscheidung über den Angriff trifft, indem auf Basis der Situationsbewertung eine Teil-Automation aktiviert wird. Diese Teil-Automation ist insbesondere zeitlich und räumlich begrenzt und die Systeme sind so ausgelegt, dass sie eine klar umrissene Aufgabe automatisiert durchführen. Das konkrete, normativ gerechtfertigte Ziel liegt darin, mit den eigenen Wirkmitteln gegenüber denen des Gegners Überlegenheit zu erreichen. Insofern ist unter normativen Gesichtspunkten, auch nach dem im Völkerrecht geltenden Verhältnismäßigkeitsprinzip, die in den eigenen Wirksystemen zum Einsatz kommende Rolle der Maschine immer abhängig zu sehen von den jeweiligen Mitteln des Gegners.
- b) Kognitiv sind KI-Systeme in der Lage, begrenzte und klar definierte Aufgaben in hoher Zuverlässigkeit zu erfüllen, wie z.B. Objekterkennung. Nach den vorherigen Ausführungen ist im Rahmen der (Äquivalenz-)Prüfung aus der völkerrechtlich gebotenen Ermessens-Abwägung zu entscheiden, ob dem Waffensystem angesichts seiner Trainingsleistungen zusätzliche kognitive Erkenntnisquellen zur Seite gestellt werden müssen, um rechtskonform eingesetzt werden zu können. Dies ist in entsprechenden Einsatzgrundsätzen für das jeweilige Waffensystem festzulegen, um so einen angemessenen hohen Zuverlässigkeitsgrad ihrer Targeting-Fähigkeiten zu gewährleisten. Die militärischen Entscheidungsträger müssen in den Einsatzgrundsätzen, der Verwendung und der Wirkweise entsprechend geschult werden, um zugleich den reflexiven Anforderungen von Verantwortbarkeit zu entsprechen.

---

<sup>38</sup> oberhalb der Betrachtung bestimmter Anwendungsfälle („use cases“)

<sup>39</sup> US-Department of Defence „Law of War Manual“: *„The law of war rules on conducting attacks (such as the rules relating to discrimination and proportionality) impose obligations on persons. These rules do not impose obligations on the weapons themselves; of course, an inanimate object could not assume an “obligation” in any event... The law of war does not require weapons to make legal determinations, even if the weapon (e.g. through computers, software and sensors) may be characterized as capable of making factual determinations, such as whether to fire the weapon or to select and engage a target.“*

- c) Ein weiteres Element der Reflexion bildet die Forderung, dass jeder mit Targeting-Fähigkeiten ausgestattete automatisierte Maschineneinsatz umso mehr in eine übergeordnete Lagebeurteilung durch Menschen eingebettet sein muss, je mehr die beschriebenen Risikofaktoren für den Einsatz eines Waffensystems zutreffen. Raum und Randbedingungen einer dem Waffensystem einzuräumenden begrenzten Autonomie sind also stets einer vorherigen von Menschen überwachten Gesamtbeurteilung der Mission zu unterwerfen. Zum Beispiel kann sich der dem Waffensystem eingeräumte Operationsraum auf einen Militärflughafen beziehen. Ändert sich die Position der Flugzeuge im Lauf des Anflugs des Wirkmittels, dann besteht die Teilautonomie in der Objekterkennung gleichwohl in einem eng begrenzten geografischen Rahmen. Generell gilt: Je enger und eingeschränkter der definierte Operationsraum des Waffensystems, desto weniger konkrete Reflexion; je weiter und volatiler der Operationsraum, umso mehr begleitende Reflexion! Hierbei spielt nicht zuletzt auch eine Rolle wie viel Zeit vergeht, ehe die Maschine das Ziel erreicht. In dieser Zeit könnte sich die Situation verändert haben, siehe dazu auch Art. 57 Abs. 2 b) ZP I zu den Genfer Konventionen. Die Frage ist hier, ob z.B. zum Zeitpunkt, zu dem das teilautomatisierte Wirken stattfindet, sich die Verhältnismäßigkeit anders darstellt als zu Beginn der Mission (weil z.B. in der Zwischenzeit eine Vielzahl von Zivilisten in der Nähe des Ziels aufgetaucht sein könnte).

Folgerungen: Die wichtigste Erkenntnis aus den vorherigen Ausführungen ist, dass der Mensch bei einem ethisch verantwortbaren KI-Einsatz in Waffen den Einsatzrahmen für die Mission und den Operationskontext der Maschine besonders sorgfältig vorab reflektieren muss. Maßgeblich sind hierbei zum einen die normativ durch das Völkerrecht gesetzten Limitierungen des Einsatzes. Zum anderen aber vor allem auch die dem Waffensystem innewohnenden kognitiven Möglichkeiten und Grenzen, die es in Abhängigkeit von den äußeren Missionsbedingungen in einer verantwortlichen Vorab-Abwägung zu bewerten gilt. Sobald dieser Rahmen aus operativen, normativen und maschinellen kognitiven Fähigkeiten - ggfs. ergänzt durch unterstützende zusätzliche kognitive Erkenntnisquellen – vom Prinzip her erfasst und verantwortlich reflektiert ist, erscheint es gerechtfertigt, die entsprechenden volitiv-ausführenden Missionsbefehle zu erteilen.

Intensität und Geschwindigkeit der von dem verantwortlichen menschlichen Operateur ex ante durchzuführenden Abwägung/Ermessensentscheidung müssen dabei immer auch von den äußeren Gegebenheiten der entweder schon einwirkenden oder präventiv abzuwendenden Bedrohung und Gefahr abhängig gemacht werden. Denn das gesamte Entscheidungsszenario unterliegt der Maßgabe, den jeweiligen Gegner wirksam bekämpfen zu müssen, ohne sich hierbei selbst einer unnötigen Gefahr auszusetzen.

## V. Fazit und Empfehlung

Der gesellschaftliche Diskurs über das Mensch-Maschine-Verhältnis bei KI- gestützten letalen Waffen muss dem Umstand Rechnung tragen, dass im Rahmen der Verhältnismäßigkeit volle Unterstützung durch künstlich intelligente Automation rechtlich und ethisch erlaubt ist. Dies gilt aber nicht nur - wie im ersten Impulspapier abgeleitet -, wenn diese für das Überleben der betroffenen Individuen oder Kollektive

in einer Verteidigungssituation erforderlich ist. Vielmehr muss es im Rahmen der Verantwortbarkeit in all ihren Dimensionen auch gelten, wenn es im Rahmen eines Szenarios der weitreichenden Verteidigung darum geht, den Gegner erfolgreich zu bekämpfen.

Auch hier ist die im Bereich von Verhältnismäßigkeitserwägungen relevante Logik eines „Schiebereglers“ anzuwenden. Denn wie im vergleichsweise leichter zu beurteilenden Fall des Selbstschutzes geht es erst recht im Fall der nach Art. 51 der UN-Charta gerechtfertigten offensiven Maßnahmen bei militärischer Verteidigung darum, entsprechend dem anwendbaren humanitären Völkerrecht das Mensch-Maschine-Verhältnis kontextspezifisch in eine verantwortungsbewusste Zusammenarbeit zu bringen. Ein wichtiges Element zur Bestimmung der jeweils angemessenen „Position“ des „Schiebereglers“ ist auch hier die individuelle, wie auch kollektive Bedrohung. Ergänzt um den Anspruch, den jeweiligen Gegner effektiv zu bekämpfen und unter möglichst wenig eigenen Verlusten an Menschen und Material zu bezwingen/besiegen.

Zum weiteren Vorgehen spricht der AK „KI & Verteidigung“ am Ende dieses Impulspapiers 2.0 folgende Empfehlungen an Bundeswehr und Politik aus:

- a.) Verzahnung von Entwicklung und von Betrieb KI durch tiefere Kooperation Bundeswehr-Wirtschaft-Wissenschaft
- b.) Bereitstellung von Mitteln und Strukturen, um theoretische Studien in zeitnahe Prototypen im Einsatzbetrieb zu überführen, um daraus Erfahrungen zu sammeln - Reallabore sind zu etablieren, es gilt testen-testen-testen, um mit gewonnenen Erkenntnissen der Truppe direkt Verbesserungen durch Industrie und Forschung einzuleiten!
- c.) Festlegung von Standards für die KI-Entwickler, die den Zielkonflikt zwischen möglichst geringem Restrisiko und operativ notwendiger Wirksamkeit im Einsatz zu harmonisieren vermögen; hierbei muss klar sein, dass ein zu starkes Ausschließen von Risiken regelmäßig zu wenig wirksamen, dem Gegner unterlegenen – und damit unverantwortbaren – technischen Systemen führt.
- d.) Es bedarf einer weiter ausgeführten KI-Strategie für die Bundeswehr, basierend auf Festsetzung der KI als Schlüsseltechnologie!
- e.) Im Weiteren wären darüber hinaus Entwicklungen von Waffensystemen und neue operative Ansätze im Sinne eines System-of-System- und eines Multi-Domain-Ansatzes zu prüfen und einzubeziehen.

Hierbei wären folgende Überlegungen als Postulate zu verifizieren:

- 1) KI ist für militärische Anwendungen so leistungsfähig wie möglich auszugestalten; zugleich ist sie in ihren Auswirkungen zu begrenzen. Die „Box“, in der sie wirken darf, bzw. welche Ergebnisse nicht akzeptiert werden, ist zu bestimmen.
- 2) Die notwendige Datenfülle und –qualität für das Anlernen der KI ist festzulegen (ergänzt um Simulation).

3) Messbare Standards sind festzulegen, auf deren Basis die KI entwickelt und genutzt wird. Die Verlässlichkeit und damit Nutzbarkeit kann durch die Kombination mehrerer ungleicher KI's erweitert werden.

4) Informatiker bauen und verstehen KI, Ingenieure integrieren sie in Systeme und verstehen In- und Outputs, und Operateure nutzen sie für ihre Tätigkeit. Verantwortung ist dabei nicht übertragbar, das jeweilige Verständnis auch nur bedingt.

5) Das gesamte Mensch-Mandanten-System im Kontext seiner Anwendung ist zu betrachten, bevor Kriterien für verantwortbare KI festgelegt werden. Zuständigkeit/Verantwortung kann situations- und Daten-/Kommunikations-bedingt verschoben werden. Diese Situationen sind im Design mitzudenken und vorzusehen.

6) Systeme laufen grundsätzlich so weit wie möglich automatisiert. Dem Menschen muss ein "Dennoch" und ein "Nein" ermöglicht werden. Menschen sind insbesondere da einzusetzen, wo es Verantwortung zu tragen gilt.

7) Bei der Zusammenarbeit verschiedener Systeme (Multi Domain) ist jedes System für sich funktionsfähig, rechtlich einwandfrei und ethisch akzeptabel zu halten.

8) Systeme sind ausreichend zu testen. Dazu sind akzeptierte Fehlertoleranzen festzulegen. Restrisiken sind im Rahmen einer (ggf. militärischen) Zulassung zu akzeptieren, ggf. durch Versicherungen abzudecken. Dabei können durch Führer auf höherer Ebene oder militärische Zulassungsstellen (aufgrund operativer Notwendigkeiten) Ausnahmen gebilligt werden.

9) Die Beachtung des Völkerrechts mit seinen Prinzipien und Regeln ist technisch zu unterstützen (legal design). Wo dies im Ausnahmefall oder aufgrund operativer Unabdingbarkeiten nicht vollständig möglich ist oder persönlicher Abschätzung bedarf (Proportionalität), ist menschliche Entscheidung einzubauen, d.h. Übersteuerung durch den verantwortlichen Menschen zu ermöglichen.

Autoren:

Dr. Hans Christoph Atzpodien, Bundesverband der Deutschen Sicherheits- und Verteidigungsindustrie e.V. / BDSV

Dr. Jürgen Bestle, HENSOLDT AG / Hensoldt

Jörg Eschweiler, Media Broadcast Satellite GmbH / MBS

Dr. Florian Keisinger, Airbus Defence & Space GmbH / ADS

Prof. Dr. Wolfgang Koch, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie / FKIE

General a.D. Jörg Vollmer, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie / FKIE

Dr. Stefan Mück, IBM Deutschland GmbH / IBM

Wolfgang Niedermark, Bundesverband der Deutschen Industrie e.V. / BDI

Generalleutnant a.D. Dr. Ansgar Rieks,

PD Dr. Frank Sauer, Universität der Bundeswehr München / UniBw M

Dr. Tassilo V. Singer, Zentrale Stelle für Informationstechnik im Sicherheitsbereich / ZITiS

Stephan Ursuleac, Materna Information & Communications SE / Materna

Dr. Christoph Baron, Bitkom e. V. / bitkom

Konstantin Knoll, Bundesverband der Deutschen Luft- und Raumfahrtindustrie e.V. / BDLI