



Wie KI Innere Führung lernt

Wertebasierte Technik mit IEEE7000™-2021

Yvonne Hofstetter

DAIO Studie 22|06

Ein Projekt im Rahmen von

dtec.bw
Zentrum für Digitalisierungs- und
Technologieforschung der Bundeswehr

Über das Defense AI Observatory

Das Defense AI Observatory (DAIO) an der Helmut-Schmidt-Universität/Universität der Bundeswehr Hamburg beobachtet und analysiert den Einsatz künstlicher Intelligenz durch Streitkräfte. Im Fokus stehen drei inhaltlich eng miteinander verbundenen Arbeitsschwerpunkte:

- Kultur- und Konzeptentwicklung sowie Organisationstransformation im Kontext der Anforderungen an die militärische Innovation
- Aktuelle und künftige Konfliktbilder, Konfliktdynamik und Einsatzerfahrungen vor dem Hintergrund des Einsatzes neuer Technologien
- Rüstungsindustrielle Dynamik mit besonderem Blick für die Gestaltung industriell-technischer Ökosysteme im Wechselspiel klassischer und moderner Technologien

DAIO ist Teil des Projekts GhostPlay. Dieses entwickelt Fähigkeiten und Technologien der militärischen Entscheidungsunterstützung für Operationen höchster Intensität. GhostPlay wird finanziert durch das Zentrum für Digitalisierungs- und Technologieforschung der Bundeswehr (dtec.bw)

Ein Projekt im Rahmen von

 **dtec.bw**
Zentrum für Digitalisierungs- und
Technologieforschung der Bundeswehr

Wie KI Innere Führung lernt

Wertbasierte Technik mit IEEE7000™-2021

Yvonne Hofstetter

DAIO Studie 22|06

Ein Projekt im Rahmen von

 **dtec.bw**
Zentrum für Digitalisierungs- und
Technologieforschung der Bundeswehr

Zur Autorin

Prof. Yvonne Hofstetter ist IEEE700™ Value Lead für GhostPlay, Gründerin und Co-Geschäftsführerin der 21strategies GmbH sowie Honorarprofessorin für Digitalisierung und Gesellschaft am Zentrum für Ethik und Verantwortung der Hochschule Bonn-Rhein-Sieg. Sie arbeitet seit 1999 an der Nahtstelle von Produkt- und Technologieentwicklung sowie Ethik- und Rechtsfragen. Prof. Hofstetter hat mehrere Sachbuch-Bestseller verfasst, darunter Sie wissen alles (München: C. Bertelsmann Verlag, 2014), Das Ende der Demokratie (München: C. Bertelsmann Verlag, 2016) sowie Der unsichtbare Krieg (München: Droemer Knaur, 2019).

Design

Almasy Information Design Thinking

Imprint

Yvonne Hofstetter, Wie KI Innere Führung lernt. Wertebasierte Technik mit IEEE7000™-2021.DAIO Studie 22/06 (Hamburg: Defense AI Observatory, 2022)

Defense AI Observatory | Chair of Political Theory | Helmut Schmidt University
Holstenhofweg 85 | 22043 Hamburg | T +49 40 6541 2776
www.defense.ai | contact@defenseai.eu | @Defense_AIO

ISSN (online): 2749-5337

ISSN (print): 2749-5345

Übersicht

1 Zusammenfassung.....	6
2 Innere Führung: Eine Skizze.....	7
3 Kann KI Innere Führung? Kritik der taktischen KI für die Bundeswehr.....	12
3.1 Werte? Welche Werte denn?	14
3.2 Ein Leben in Freiheit auch für die KI?	15
3.3 Wer trägt die Verantwortung?.....	17
3.4 Eine Frage der Zeit?	19
4 Wie die Innere Führung in die Maschine kommt: Wertebasierte Technik mit IEEE7000™-2021	22
4.1 Von Interesse: Operational Concept and Context Exploration.....	25
4.2 Die Werteerkundung	26
4.3 Von der Theorie in die Praxis: Ethical Value Requirements	30
5 Fazit	32
Literatur.....	34

1 Zusammenfassung

Dass sich künstliche Intelligenz (KI) weltweit ausgebreitet hat, ist eine Binsenwahrheit. Die rasche und unaufhaltsame Proliferation von KI der letzten zehn Jahre spricht für sich, und längst ziehen auch Gesetzgeber und Regulierungsbehörden nach, um KI und ihre Technikfolgen einzuhegen. Für Deutschland relevante Gestaltungsanforderungen haben die High-Level Expert Group on Artificial Intelligence der Europäischen Kommission (HLEG AI) und auf nationaler Ebene die Datenethikkommission der Bundesregierung (DEK) und die Enquete-Kommission Künstliche Intelligenz des Deutschen Bundestags (EKKI) geäußert. A priori verbindlich sind die von Kommissionen oder Arbeitsgruppen aufgestellten Forderungen nicht. Wo sie nicht in hoheitlich gesetztes Recht transformieren – etwa als EU AI Act, dessen Verabschiedung und Inkrafttreten um das Jahr 2025 erwartet wird –, können sie allenfalls zu Soft Law werden, das auf vertraglicher Übereinkunft von Sozialpartnern oder in Wertschöpfungsketten auf der (Selbst-)Verpflichtung zu einer Corporate Social Responsibility (CSR) beruht. Wird dieses Recht „zweiter Klasse“ verletzt, drohen vornehmlich finanzielle Konsequenzen. Deshalb rückt das Anliegen, „Prinzipien des Sein Sollens, die unserem Handeln Orientierung geben“¹ – die Werte –, in technische Funktionen einer KI zu übersetzen, stärker in den Fokus.

Dieser Beitrag fasst die Werte der Bundeswehr, die ihren Ausdruck in der Führungsphilosophie der Inneren Führung finden, grosso modo zusammen und stellt sie überblickartig in den Kontext von KI. Für die Be-Wertung von KI und Innerer Führung kommen Offiziere der Bundeswehr höchstpersönlich zu Wort. Ihren Äußerungen schließt sich ein Rundgang durch den im September 2021 in Kraft getretenen Standard für wertebasierte Technik – Value-based Engineering (VBE) –, IEEE 7000™-2021, an, der sich eine große Aufgabe zum Ziel gemacht hat: die Hervorbringung von Technik, die systematisch durchdachte Systemanforderungen des Rechts und der Ethik erfüllt.

¹ Spiekermann, „Die Ethik in der Künstlichen Intelligenz“.

2 Innere Führung

Eine Skizze

Das nachfolgende kolportierte Verständnis von Innerer Führung lehnt sich an eine nicht repräsentative Interviewreihe mit Offizieren der Bundeswehr aus dem Jahr 2022 an,² der folgende Leitfrage zugrunde lag: „Bitte helfen Sie uns, das Spannungsverhältnis von Innerer Führung und dem Einsatz künstlicher Intelligenz (Defense AI) bei der Bundeswehr zu verstehen, damit aus taktischer KI ‚moralisch gute Smart Comrades‘ hervorgehen, die mit dem Verständnis von Innerer Führung in Einklang stehen, von der Bundeswehr gerne eingesetzt werden, die Verteidigungs- und Reaktionsfähigkeit verbessern und Technologieführerschaft herstellen.“ Beispielhaft werden einige Stimmen aus den geführten Interviews herausgegriffen, um die angesprochenen Themen zu stützen.

Innere Führung ist das grundlegende Führungsprinzip der Bundeswehr:

(Innere Führung ist) (e)in sehr weites Feld, ein absolutes Alleinstellungsmerkmal der Bundeswehr. Ich habe das sehr spezifisch erlebt, dann im Kontext mit anderen Armeen“³

Als Quellen Innerer Führung werden an erster Stelle das geltende Recht, besonders das Grundgesetz und das Völkerrecht, genannt. Mit dem Rückgriff auf das Grundgesetz legt sich die Bundeswehr auf ein bestimmtes Menschenbild fest: Es ist das Menschenbild der „Person (‚per se existere‘), die individuell von allen anderen unterschieden (‚per se unum‘) und frei [ist], weil aus sich heraus handelnd (‚per se agere‘)“.⁴ Diese Definition des Thomas von Aquin ist zugleich Würdebegründung und erhebt die Person im Gegensatz zur Sache, dem Objekt, zum Rechtssubjekt und Träger von Rechten und Rechtspflichten, das sowohl in seinen Entscheidungen als auch Handlungen stets souverän, das heißt frei, bleibt. Daran hat sich in juristischer Hinsicht bis heute nichts geändert. Die Wahlfreiheit der Person macht ihre Würde aus und bleibt normierend auch für das Bild der einzelnen Angehörigen der Bundeswehr.

Als Personen sind Angehörige der Bundeswehr aber nicht nur Träger ihrer eigenen Würde. Im Konfliktfall muss auch die Würde des Gegners geachtet und gewahrt werden. Dafür können die Streitkräfte neben ihrer Bindung an das Grundgesetz einerseits auf das humanitäre Völkerrecht, andererseits auf die eigene sittliche Haltung einer Menschlichkeit zurückgreifen, wie sie historisch und kulturell in Europa gewachsen ist: „Auch wenn etwas nicht vorgeschrieben oder verboten ist, gibt es trotzdem das Prinzip der Humanität – dass man sich (durch Humanität) leiten lässt (und einschätzt), ob es richtig ist, nach dem moralischen Kompass, den

² Die Sample Size beträgt 10. In der Folge beziehen sich die Zitate auf die anonymisierten Gesprächspartner (Großbuchstabe) und die jeweilige Seitennummer des Protokolls.

³ 21strategies, E-2.

⁴ Thomas von Aquin, zit. nach: Precht/Burkard, Metzler Lexikon Philosophie, S. 444.

wir als Staatsbürger in Uniform haben sollten, etwas zu tun oder (besser) nicht zu tun“.⁵

Das Zitat deutet an, dass auch „Verfahrensethik ohne Verfahren“, also eine Ethik ohne die inhaltlichen Festlegungen etwa einer Kant'sche Sittenlehre mit ihrer „Gesolltheit von Handlungsweisen“,⁶ legitime Quelle Innerer Führung ist. Gemeint sind damit das Wissen (um Tatsachen) und das Gewissen von Bundeswehrangehörigen:

(Innere Führung) sagt zum einen: Jeder Soldat ist auch ein denkender, fühlender, ethisch bewertender Mensch. Und das zuerst, bevor er oder sie Soldat ist. Und (er oder sie) darf auch, muss auch nach ethischen Grundprinzipien handeln.⁷

Präskriptiv für das Soldatenbild der Bundeswehr sind neben dem geltenden Recht also auch Ethik und Moral, wobei Moral, da sie auch das „Empfinden der eigenen Würde und Fragen, wodurch unser Leben einen Sinn erhält oder Erfüllung findet“ betrifft,⁸ nach herrschender juristischer Meinung der Gefühlswelt zuzuordnen ist.⁹ Auch Bundeswehroffiziere interpretieren Ethik und Moral als subjektives Gefühl oder Intuition:

Dann gibt es ja das Bauchgefühl, nachdem jemand entscheidet... der am Ende auch entweder wertebasiert entscheidet oder eine rein statistische Entscheidung (trifft), weil: Ich habe das schon zehn Mal gesehen. Dann wird es beim elften Mal auch richtig sein.¹⁰

Das Gewissen als Intuition, die geprägt ist von „Weltbild, (...) Religion, (...) Erziehung, (...) dem elterlichen Vorbild, (...) der Sozialisierung in der Schule, in der Gemeinschaft“ eines Individuums¹¹, birgt auch für die Bundeswehr Risiken:

Im Prinzip können Sie (Werte) nur vorleben und hoffen, dass [die Formung] funktioniert, aber Sie wissen (es) nicht.¹²

5 21strategies, D-34.

6 Vöneky, „Ethische Experten und moralischer Autoritarismus“, S. 88.

7 21strategies, E-14.

8 Precht/Burkard, Metzler Lexikon Philosophie, S. 391.

9 „Dagegen, dass eine solche rationale Rechtfertigung normativer Aussagen möglich ist, haben sich viele gewandt (...). Sie alle sind mit unterschiedlichen Gründen davon ausgegangen, dass ethische Äußerungen keine begründungs- oder wahrheitsfähigen Aussagen sind, sondern entweder bloße Scheinaussagen, oder nur Gefühlsäußerungen, Empfehlungen oder Imperative.“ Vöneky, „Ethische Experten und moralischer Autoritarismus“, S. 89.

10 21strategies, A-63.

11 Ders., G-14.

12 Ders., I-50.

Dem Soldaten sein Gewissen zu geben – als mächtige Waffe und als Einschätzungstool –, ist nicht einfach. Deswegen tut das Militär das normalerweise nicht; und deswegen ist es in Deutschland auf dramatische Weise schief gegangen. (Trotzdem) sagt die Bundeswehr: Wir akzeptieren die Probleme, die wir damit haben.¹³

Die Probleme, die in den Zitaten implizit anklingen, sind jedoch nicht primär durch die Person des einzelnen Soldaten verursacht, sondern durch den Mangel objektiv gültiger Normen für moralisches Handeln. „Was früher als grundlegende ethische Lehrsätze der westlichen Zivilisation (...) betrachtet wurde, [war] im Laufe der Zeit bis auf die Ebene reiner Konventionen (die so leicht austauschbar sind wie ein Code von Tischmanieren) abgewertet worden“.¹⁴ 1965 ist es Hannah Arendt, die den Mangel an Objektivität in der Moral beklagt und wenigstens auf einen gesellschaftsübergreifenden „moralischen Geschmack“¹⁵ hofft: „Es widerspricht eigentlich dem Menschlichen, der menschlichen Natur, wenn man Sadist ist oder Mörder.“¹⁶

Neben der normierenden, präskriptiven Sicht auf die Angehörigen der Bundeswehr hat die Innere Führung noch einen weiteren Aspekt. Es ist die Perspektive der militärischen Führung bei der Auftragserteilung an unterstellte Kräfte und betrifft damit nicht den Inhalt, sondern die hierarchische Ordnung und die Kommunikation auf diesem Linienweg. „(Innere Führung stellt) aber auch sehr hohe Ansprüche an die Führung, denn ich muss auch so führen, um meinen Soldaten diesen Raum zu geben, dass sie selber als Menschen mit Emotionen, Ethik, Moral nach moralischen Aspekten handeln können und dürfen“.¹⁷

In diesem Zusammenhang wird das Konzept „Führen mit Auftrag“ – oft synonym mit dem Begriff der „Auftragstaktik“ verwendet – der Befehlstaktik gegenübergestellt:

Das sind für mich die beiden Hauptpunkte (bei der Inneren Führung), dass man (i) eine gewisse Freiheit bzw. Flexibilität bei der Auftragserfüllung hat und (ii) gleichzeitig halt nicht jeden Auftrag oder Befehl umsetzen muss, sondern eher dazu angehalten ist, auch innezuhalten und aufzuzeigen, wenn offensichtlich was Verbrecherisches oder Selbstmörderisches befohlen wird.¹⁸

13 Ders., E-74.

14 Arendt, Das Urteilen, S. 166.

15 Ebd., S. 167.

16 21strategies, G-11.

17 Ders., E-7.

18 Ders., D-20.

Beim Führen mit Auftrag vertraut die übergeordnete militärische Führung darauf, dass Untergebene den ihnen erteilten Auftrag in ihrem Sinne ausführen. Im Sinne dieses intersubjektiven Vorgangs gilt Innere Führung dann als gelungen, wenn die Untergebenen ihre Gewissens- und Handlungsfreiheit behalten, sich aber gleichzeitig auch die Führung selbst vorbildlich verhält.

3 Kann KI Innere Führung?

Kritik der taktischen KI für die Bundeswehr

Ich behaupte, wir gewinnen dann, wenn wir at machine speed arbeiten. Und das heißt, wir müssen eigentlich den Menschen rausnehmen. Großes Entsetzen immer.¹⁹

Mit ihren Prinzipien Rechtsstaatlichkeit und Ethik und Moral trifft die Bundeswehr erstmals auf das volle Potenzial von KI der Dritten Welle („taktische KI“). Aufklärung, Auswertung, Lageanalyse und -darstellung, Warnung bis hin zur Informationsverdichtung sind die Anwendungsfelder der Zweiten Welle KI für das Militär seit den 90er-Jahren des letzten Jahrhunderts. Doch für die moderne Verteidigung reichen diese Fähigkeiten allein nicht mehr aus. Heute verlangt die Bundeswehr von Maschinen strategische und taktische Fähigkeiten, die sich in Maschinengeschwindigkeit vollziehen, eben at machine speed. Taktische KI ist die Basis jener Maschinen, die „den Menschen rausnehmen“. Gefechtsfeldsimulatoren und Trainingssysteme, (teil-)autonome Drohnen in der Luft und zur See sowie an Land, smarte Munition, cyber-physische Systeme oder domänenübergreifende KI-gesteuerte Netzwerke von Verteidigern – die Liste der denkbaren Anwendungsfelder für taktische KI ist lang. Auf strategischer Ebene macht sie sich zum Ziel, Planungsaufgaben wie etwa eine Air Tasking Order auf ein Drittel oder Viertel des heutigen Zeitaufwands zu reduzieren,²⁰ auf operativer Ebene kann sie den Kampfwert alter Systeme steigern, sie in technisch (teil-)autonome Waffensysteme transformieren,²¹ höheren Eigenschutz herstellen und Munitionsverbrauch und -logistik optimieren oder auch die algorithmische taktische Führung von Truppenteilen oder Einheiten während eines Gefechts übernehmen. In Zukunft benutzt der Kommandant keine KI mehr. Die KI ist der Kommandant.

An taktischer KI für die Bundeswehr entzündet sich indes eine Debatte: „Großes Entsetzen immer.“ Denn taktische KI für das Militär bewegt sich in einem Extrembereich: Sie dient der staatlichen Gewaltanwendung; und als taktische KI im Dienste des Staates muss sie eingebettet sein in den Rechts- und Werterahmen des Grundgesetzes für die Bundesrepublik Deutschland, relevanter EU-Vorschriften und des (humanitären) Völkerrechts, genauer: in das Führungsprinzip der Bundeswehr, die „Innere Führung“. Als große Herausforderung erweist sich dabei, rechtliche, ethische oder soziale Normen in die Funktionen eines Softwaresystems zu gießen.

19 21strategies, F-1.

20 Diesen Aspekt greift das Forschungsprojekt Strategic Chaos Engine for Planning, Tactics, Experimentation and Resiliency (SCEPTER) der Defense Advanced Research Projects Agency auf. Siehe hierzu weiterführend: „DARPA SCEPTER program will explore machine-generated strategies for warfare“.

21 Stichwort: Smart Gepard. Siehe hierzu Kapitel 4.1.

3.1 Werte? Welche Werte denn?

Die positiven Bedeutungen, die Bundeswehrangehörige der Inneren Führung zumessen, lesen sich wie ein bunter Strauß guter Motivationen und Qualitäten von Menschen und von Dingen. De facto äußern die Offiziere im Zusammenhang mit Innerer Führung zig unterschiedliche Phänomene, die in ihnen als „fühlende() und begehrende() Subjekte (...) gewisse Gefühlszustände“²² auslösen. Was ihre Seelen bewegt, reicht von A wie Abenteuerlust bis Z wie Zurückhaltung.

Ein Grund für die hohe Anzahl der genannten Phänomene, die nicht – wie in anderen Fällen der Erhebung qualitativer Forschungsdaten einer Zielgruppe häufig beobachtbar – schon bei einer Samplegröße von 10-12 gegen dieselben Begriffe konvergieren, wurde bereits genannt: der Mangel der moralischen Objektivität. Dass Wertanschauungen subjektiv beeinflusst sind²³ und sich vor dem Hintergrund bestimmter Ideologien, Weltbilder oder Paradigmen, der „kulturellen Hintergrundstrahlung“²⁴ ausbilden, spitzt ein General wie folgt zu: „Aber was wollen wir? Es wird andere geben, die sagen: Was ist das, ethisch? Ethisch ist, wenn wir gewinnen und die anderen verlieren.“²⁵

Neben die subjektive Erlebniswelt tritt ein weiteres Defizit: Es fehlt eine „philosophisch fundierte Begriffswelt“²⁶ wenn Menschen über Werte sprechen. Die strukturellen Eigenschaften von Werten finden keine Abbildung in der Sprache; die Sprache von den Werten ist nicht exakt genug. Dabei ist die „Sprachkritik (...) ein konstitutiver Bestandteil methodischer Philosophie und Ethik“.²⁷ Diese Unzulänglichkeit insbesondere ist der Grund für die späteren Schwierigkeiten der Entwickler und Ingenieure, taktische KI zu implementieren, die selbst Träger der gewünschten positiven Phänomene oder sie zu unterstützen in der Lage ist.

„Technische Robustheit und Sicherheit“ oder „Transparenz“, drei Bedeutungen, wie sie die Ethics Checklist der HLEG AI für KI definiert,²⁸ sind generell Qualitätsattribute technischer Systeme,²⁹ wie sie per default in Softwarespezifikationen beschrieben wurden, bevor sich die agile Softwareentwicklung mit ihren anderen Dokumentationsanforderungen durchzusetzen begann. Sie sind „Hygiene-

22 Scheler, Der Formalismus in der Ethik und die Materiale Wertethik, S. 10.

23 „Max Scheler war der Prophet eines neuen Objektivismus. Es sollte sein radikaler Abbruch des alten kantischen sichernden Subjektivismus: (...) des ethischen Sicherns der Welt der Ideale und Werte durch die subjektive Autonomie (...)“ Zit. gemäß: Przywara, Humanitas: Der Mensch gestern und morgen, S. 30-31.

24 Funk, Roboter- und KI-Ethik, S. 17.

25 21strategies, F-48.

26 Spiekermann „Die Ethik in der Künstlichen Intelligenz“.

27 Funk, Roboter und KI-Ethik, S. 36.

28 Wahl, „High-Level Expert Group Publishes Ethics Checklist (2020)“.

29 Spiekermann, „Value-based Engineering: Prinzipien und Motivation für bessere IT-Systeme“. Siehe auch: https://en.wikipedia.org/wiki/List_of_system_quality_attributes (letzter Zugriff: 17. Oktober 2022)

anforderungen“ an Softwaresysteme,³⁰ schärfer ausgedrückt: Eigentlich sind sie selbstverständlich. Doch seit über zwei Dekaden haben sich stark verkürzte Softwareentwicklungszyklen, Kostendruck, der Fokus auf den Shareholdervalue, die Kommodifizierung von Programmierleistung und die dadurch schnellere Inverkehrbringung von Software zwar zugunsten von Umsatz und Gewinn eines Herstellers oder Betreibers ausgewirkt; gleichzeitig aber hat die Qualität selbst von missionskritischen Systemen stark gelitten. Den tödlichen Beweis dafür haben die auf mangelhafte Softwarekonzeption, -qualität und physikalische Expertise der beteiligten Programmierer zurückzuführenden Abstürze zweier 737-MAX-8-Maschinen der Firma Boeing geführt.³¹

Die selbstverständlich zu erwartenden Qualitätsattribute von Softwaresystemen reichen längst nicht aus, um Kernwerte wie Menschenwürde und Freiheit, Friede, Gerechtigkeit oder tugendhafte Charaktereigenschaften wie Heimatliebe, Wahrhaftigkeit oder Mut in einer taktischen KI zu verwirklichen. Wertebasierter KI muss daher Bedeutungsarbeit und eine eindeutige Begrifflichkeit vorausgehen, eine Aufgabe, die der Standard IEEE 7000™-2021 für wertebasierte Technik den Value Leads, einem neuen Berufsstand, auferlegt. Auf dessen Rolle und Aufgabenumfang im Entwicklungsprozess kommt Kapitel 4 zu sprechen.

3.2 Ein Leben in Freiheit auch für die KI?

Intuitiv übertragen die Offiziere die Idee vom Menschsein gedanklich auch auf taktische KI. Denn es wird in Erwägung gezogen, dass taktische KI moralisches Verhalten bewerten und selbst ethisch urteilen oder sich an Menschengesetze halten könne:

Ich gebe dem Smart Gepard die Regeln vor, an die wir uns noch moralisch und ethisch gebunden fühlen.³²

Wenn wir sagen, Innere Führung in der KI... wie macht man das? Indem man sie so trainiert, dass sie die Dinge, die sie dann auslösen soll, an den Maßstäben orientiert, die man als Mensch vielleicht anlegt.³³

³⁰ Sarah Spiekermann, persönliche Kommunikation, 12. November 2021.

³¹ Boeing hat in den letzten zwanzig Jahren erfahrene Flugzeugingenieure verloren und die Softwareentwicklung nach Indien verlagert; manch Ingenieur hat das Unternehmen verlassen, weil er von den Unzulänglichkeiten und den Umgang des Managements mit den Fehlern wusste. Beispielhaft zum Vorgang: Spaeth, „Grundlegender Softwarefehler in der Boeing 737 Max gefunden“.

³² 21strategies, E-60.

³³ Ders., F-3.

Aber das ist so interessant an der KI. (KI) ist vielleicht sogar noch menschlicher als der Mensch selbst und in gewissen Situationen auch sicherlich berechenbarer als ein Soldat.³⁴

Die Tendenz, KI-Systeme nicht als Sache, als Objekt, sondern als „Subjekte ethischen Urteilens“ zu betrachten,³⁵ ist so fragwürdig wie folgenreich. Betrachtet man KI als Sache, als Ding, dann ist unter KI-Ethik zu verstehen, dass Menschen „den zweckgerichteten Einsatz (von KI) moralisch auf den Prüfstand stellen“.³⁶ Das ist ein anderer Fall als zwei andere Definitionen von KI-Ethik, wonach (i) KI selbst ethische Erwägungen anstellen könnte und (ii) KI Regeln und Vorschriften folgen müsse.³⁷ In den beiden letztgenannten Fällen ginge es um KI als ein Subjekt, das selbst moralisch handeln könnte und einem Kodex – einem geschriebenen oder ungeschriebenen Ethos – unterworfen wäre.

Bemerkenswert an den beiden letzten herangezogenen Zitaten ist die Ausdrucksweise in Bezug auf KI. „(KI) ist vielleicht sogar noch menschlicher“ ist nicht dasselbe wie „(KI) handelt vielleicht sogar noch menschlicher“ oder „(KI) verhält sich vielleicht sogar noch menschlicher“. Personen handeln in Handlungsfreiheit, der „Fähigkeit und (dem) Vermögen zum bewussten und freiwilligen Tun“.³⁸ In diesem Sinne handelt jedoch taktische KI niemals, schlicht weil ihr die innere Fähigkeit dazu fehlt. Ungeachtet ihrer erstaunlichen technischen Fertigkeiten hat sie „keinen personalen Charakter, weil sie nicht in der Lage ist, sich selbst Gesetze zu geben, sondern Menschengesetzen“ gehorchen muss.³⁹ Als Maschine fehlt es ihr an Bewusstsein, an Freiwilligkeit, an eigener Motivation. Ignoriert man den Mangel theoretisch exakter Formulierungen in den genannten Zitaten, zögern die drei beispielhaft herangezogenen Interviewpartner also instinktiv, KI dieselbe Freiheit zuzugestehen wie dem Menschen. „KI ist menschlicher“, bezeichnet nichts mehr als ein Sein im Sinne einer Möglichkeit, dass Menschlichkeit dem materiellen Objekt namens KI immanent sein könnte. „Dinge, die sie dann auslösen soll“, lässt offen, ob KI entweder handelt oder sich ohne mentale Prozesse verhält oder schlicht nur funktioniert, weil die Äußerung nur auf die Wirkung von KI abstellt. „Handeln, das die Maschine am Ende ausgeführt hat“,⁴⁰ weist die eigentliche Handlung – den Einsatz von KI – der Person des Soldaten zu; es ist die Handlung des Soldaten kausal für die Wirkung der KI.

34 Ders., I-57.

35 Funk, *Roboter und KI-Ethik*, S. 40.

36 Ders., S. 25.

37 Ders., S. 69-88.

38 Precht/Burkard, *Metzler Lexikon Philosophie*, S. 188.

39 „Person ist dasjenige Subjekt, dessen Handlungen einer Zurechnung fähig sind. Die moralische Persönlichkeit ist also nichts anderes als die Freiheit eines vernünftigen Wesens unter moralischen Gesetzen..., woraus dann folgt, dass eine Person keinen anderen Gesetzen, als denen, die sie (entweder allein, oder wenigstens zugleich mit anderen) sich selbst gibt, unterworfen ist“.
Kants gesammelte Schriften, AA22.

40 21strategies, E-24.

Die Objekteigenschaft von KI stünde auch dann vor der Auflösung, wenn sich KI an einen Ethos, an einen Kodex, halten sollte. Sowohl hoheitlich gesetztes Recht wie das Grundgesetz als auch Einsatzregeln der Bundeswehr (Rules of Engagement) sind solche Codices. Mit Blick auf die Wertediskussion normieren sie im besten Fall das „ethische Minimum“.⁴¹ Recht aber ist von Menschen für Menschen geschaffen – und nicht von Menschen für Maschinen. Recht macht frei, weil Menschen die Freiheit überlassen wird, gegen Recht zu verstoßen. Die Durchsetzung von Regeln durch Maschinen – das smarte Auto, das nicht anspringt, weil sich seine Insassen noch nicht angeschnallt haben – bedeuten Zwang und Moralpaternalismus. Würde eine KI menschliche Codices selbst beachten können oder müssen, würde sie also auch in diesem Fall zum „Subjekt moralischen Handelns“ werden.⁴² Hätte die Unterwerfung von KI unter Menschengesetze dann zur Folge, dass KI auch Verantwortung tragen könnte oder müsste?

3.3 Wer trägt die Verantwortung?

Mit der imaginierten Subjektivierung und Individualisierung von KI würde die Mensch-Maschine-Unschärfe des digitalen Zeitalters weiter voranschreiten. Noch aber hat das Konzept der Person juristischen Bestand. Nur das Subjekt, nicht eine Sache wie eine KI, kann freie ethische Urteile fällen und in Freiheit handeln. Für frei getroffene Werturteile und sein freies Handeln trägt der Mensch – und nur der Mensch – dann auch die Verantwortung:

Lenk Waffen kann ich, nachdem ich sie abgeschossen habe, (auch) nicht mehr sehen. Ich kann sie nicht mehr steuern. Ich kann nicht mehr eingreifen, und ich bin trotzdem dafür verantwortlich, wenn die eben kein Kampfflugzeug treffen, sondern einen zivilen Jet. Und dann bin ich wieder dran, und das auch zu Recht.⁴³

Verantwortung für ihr Handeln ist für Angehörige der Bundeswehr besonders dann ein zentrales Anliegen, wenn sie zum Tod von Kameraden oder auch des Gegners geführt haben. Den Offizieren ist wichtig, Erklärungen zu geben und die Gründe und Intentionen ihrer Handlung darzulegen:

Wo Menschen arbeiten, passieren Fehler. Wenn das im militärischen Einsatz schlimm wird, muss man auch zu seiner Verantwortung dort stehen, solange

41 „Das Recht ist nichts anderes als das ethische Minimum.“ Zugeschrieben Georg Jellinek (1851-1911).

42 Vgl. dazu: Funk, Roboter- und KI-Ethik, S. 69.

43 21strategies, E-30.

man sagt: Ich habe das aus diesen und diesen Gründen gemacht. Ich bin in der Conclusio zu dem Ergebnis gekommen: Ich kann das verantworten.⁴⁴

Verantwortung ist etwas, das ein intersubjektives Geschehen beschreibt. Es geht um das Reden, um das Verarbeiten von Geschehnissen von Person zu Person. Die Befragten stellen dabei ausdrücklich die Unterschiede zwischen Mensch und Maschine heraus, besonders die Fähigkeit des Menschen zu Empathie, Mitleid, Stärke, Intuition, Barmherzigkeit. Verantwortung ist auch der Grund dafür, dass die Offiziere menschliche Entscheidungen den (präzisen und akkuraten) maschinellen Berechnungen vorziehen, selbst dann, wenn sich eine menschliche Entscheidung als schlecht erwiesen oder ihre Wirkung ganz verfehlt hat:

Die Entscheidung der KI könnte ja sogar besser sein im Sinne von: Sie hat vielleicht zu weniger Toten geführt auf der eigenen Seite. Aber die Angehörigen der gestorbenen Kameraden haben niemanden, mit dem sie darüber reden können. Und ist das nicht am Ende furchtbarer als andere Konsequenzen, die man aber im Nachhinein bearbeiten kann?⁴⁵

Aber einem autonomen System verzeihen Sie keine Fehler. Einem Menschen verzeihen Sie immer Fehler oder eher Fehler.⁴⁶

Ohne Ausnahme befürworten alle Befragten den Einsatz von KI; interviewübergreifend herrscht auch Einigkeit darüber, dass nur der militärische Führer für den Einsatz von KI verantwortlich ist:

(D)as einzig Realistische, Pragmatische, was wir beim Militär auch noch immer brauchen, ist, dass jener, der dieses Produkt am Ende einsetzt, derjenige ist, der die Verantwortung trägt. Man wird (die Verantwortung für den Einsatz von KI) nicht auf den Entwickler, der sie fünf Jahre vorher programmiert hat, zurückführen können.⁴⁷

Doch die Bereitschaft zur Übernahme von Verantwortung für den Einsatz einer Maschine ist an Voraussetzungen geknüpft: „(W)ie können [Operateure] dieser Verantwortung gerecht werden für das Handeln, das die Maschine am Ende ausgeführt hat?“⁴⁸ Die Soldaten wollen „ihre“ KI in- und auswendig kennen und wissen, worauf sie sich einlassen, bevor sie KI in den realen Einsatz bringen: „Dann brauche ich wirklich gutes Training, und zwar sehr realitätsnahes Training“.⁴⁹

44 Ders., C-118.

45 Ders., B-72.

46 Ders., I-54.

47 Ders., E-25.

48 Ders., E-24.

49 Ders., E-40.

Neben intensivem Training des Menschen mit einer KI wird zudem gefordert, dass die Funktionsweise der KI durch einen Qualitätssicherungsprozess und Standards sichergestellt ist:

Da wird man wahrscheinlich Standards schaffen müssen. Wie oft und mit welcher Fehlerwahrscheinlichkeiten muss in einem gewissen Szenario die KI zu folgendem Ergebnis kommen?⁵⁰

Als zweite Voraussetzung schlagen die Befragten ganz im Sinne Innerer Führung vor, dass ihnen Wahlfreiheit zugestanden wird. Sie wollen selbst entscheiden, ob sie taktische KI einsetzen wollen oder nicht:

Die KI darf nicht Unsicherheiten in die Situation bringen. Deshalb würde ich vielleicht sagen, ein militärischer Führer, der (KI) einsetzt, sollte sich dazu vielleicht selbst entscheiden. Also nicht so, dass ein anderer ihm sagt: Du musst diese KI nutzen.⁵¹

Während sich die Befragten also sämtlich dafür aussprechen, KI – selbst mit technisch-autonomen Fähigkeiten – einzusetzen, wenn die beiden Voraussetzungen Qualitätssicherung und Training eingehalten und geschaffen werden, und dann auch bereit sind, für das Funktionieren der KI die volle Verantwortung zu tragen, brauchen sie dennoch innere Stärke, wenn es um das zeitliche Auseinanderfallen von Einsatzentscheidung und Wirkung geht.

3.4 Eine Frage der Zeit?

Wie kommt es, dass Menschen annehmen, KI träge freie Entscheidungen? Die Frage wird beim militärischen Einsatz von KI, bei dem es um eine Bekämpfung- oder Tötungsentscheidung geht, besonders prekär. Sie ist es auch, um die sich die Debatte zur Ächtung von „Killerrobotern“ bei der Ständigen Abrüstungskonferenz dreht und mit dem Schlagwort „letale autonome Waffensysteme“ zusammengefasst wird.

Ein Offizier ist auf der richtigen Spur und bringt es auf den Punkt: Das Problem liegt bei der Frage nach der Zeit. Beim Einsatz von taktischer KI tritt die Wirkung entweder stark verzögert ein oder aber in Maschinengeschwindigkeit und so schnell, dass dem Menschen die Kontrolle über die KI entzogen zu sein scheint. Hinzukommt, dass die soldatische Handlung, KI einzusetzen, im komplexen

⁵⁰ Ders., F-34.

⁵¹ Ders., B-95.

dynamischen System des Gefechtsfelds getroffen wird. Im dynamischen System ist Zeit auch deshalb relevant, weil sich der Systemzustand ständig verändert. Das eigene Handeln des Einsetzers und das Funktionieren der KI stehen also zweifach „in der Zeit“. Sie sind mit dem Medium Zeit eng verknüpft:

Der Schuss muss durch das Individuum(abgelegt) werden. Jetzt übertragen wir das auf den Leopard. Er markiert das Ziel. Da ist es eigentlich kein Problem. Klar (schießt der Leopard) auf das Ziel, aber da ist schon vorher alles gelaufen. Es ist interessant – der Faktor Zeit.⁵²

Was hier zum Ausdruck kommt, ist eine Theorie von Zeit, die Zeit als Abfolge kausaler Zusammenhänge, als Fluss von Ursache und Wirkung, definiert.⁵³ Während bei der schnell zu treffenden Handlungsentscheidung das Wissen als Zustand einer Erkenntnis die größte Rolle spielt, entscheidet bei der verzögert eintretenden Wirkung einer Handlung der menschliche Einsetzer jetzt über den Einsatz taktischer KI. Seine Handlung jetzt ist kausal für die spätere Wirkung der KI, die möglicherweise erst mit (langer) zeitlicher Verzögerung eintritt. Denn Einsetzer und KI können zeitlich – vielleicht auch örtlich – voneinander „abgetrennt“ sein. Beim Reinforcement Learning, einer KI-Lernmethode, ist ein solch zeitlicher Effekt bekannt und wird als Temporal Difference bezeichnet. Eine Backgammon-KI muss jetzt einen Spielzug machen, „weiß“ aber erst viel später, ob dieser Zug jetzt dazu beigetragen wird, das Brettspiel am Ende des Spiels zu gewinnen. Die Auswirkung eines Spielzugs, der jetzt vollzogen wird, tritt also erst 1.500, 1.800, 2.300 usw. Zeitschritte später ein. Ganz ähnlich verhält es sich beim Einsatz einer taktischen KI. Durch die zeitlich verzögerte Wirkung wirkt es so, als habe die KI eine eigene Bekämpfungs- oder Tötungsentscheidung getroffen („autonom“) und nicht bereits der militärische Führer, der längst vorher über ihren Einsatz entschieden hat. Er trifft die Handlungsentscheidung jetzt, die Wirkung seiner Handlungsentscheidung liegt aber in der Zukunft und möglicherweise auch örtlich getrennt von ihm.

Manchen Menschen gefällt diese zeitliche und auch räumliche Trennung von Entscheidung und dem Eintritt der Entscheidungsfolgen nicht

(Autonomie, BVR⁵⁴ ist,) was dem typischen Panzersoldaten nicht gefällt. Der möchte gerne aus seiner Luke alles sehen. Obwohl er weiß, dass er da extrem gefährdet ist, wollen trotzdem alle oben rausschauen und wollen dann die Entscheidung treffen, wer wohin fährt.⁵⁵

52 Ders., I-98.

53 Vgl. dazu: Reichenbach, The Direction of Time.

54 BVR: Beyond Visual Range.

55 21strategies, E-27

Das Auseinanderfallen von Entscheidung und Wirkung ist also nicht neu:

(Der Einsatz von KI) ist am Ende des Tages nicht viel anderes als das, was der Offizier – wenn man über einen klassischen Krieg redet – irgendwo in seinem Zelt [tut], das auf dem Gefechtsfeld steht, aber halt zehn Kilometer hinter der Front, und der seinen Truppen gesagt hat: Bitte diese Stellung da vorne einnehmen. Und der dafür verantwortlich ist, dass die [Truppen] es richtig tun und nicht das Dorf und die Zivilisten daneben attackieren.⁵⁶

Das lässt einen Oberst darauf schließen, dass nachgerade die Auftragstaktik „die richtige Führungsphilosophie (für KI ist), weil die Auftragstaktik nur das „Was“, aber nicht das „Wie“, vorgibt und der Rahmen der KI schon etabliert worden ist.“⁵⁷

Zum Problem der Zeit zwischen Handlung und Wirkung sei im Übrigen auf die neurowissenschaftliche Literatur verwiesen, die sich mit der Belohnung für eine Entscheidung befasst. Grundsätzlich bevorzugen Menschen Entscheidungen, deren Effekte sich in zeitlicher Nähe zur Entscheidung einstellen, gegenüber denen, bei denen sie auf den Wirkungseintritt länger warten müssen.⁵⁸ Das erklärt das Unwohlsein von Menschen beim Einsatz technisch-autonomer Systeme. Es beruht nicht auf der Auflösung des Objektcharakters der KI oder der Infragestellung, dass der Mensch handelt und KI nur funktioniert. Vielmehr sind es psychologische Gründe, die zu Missbehagen führen. So setzen „Verzögerungsdiskontierungen“ bei Entscheidungen mit aufgeschobener Wirkung ein,⁵⁹ die auf Motiven wie Zweifel, Bedenken, Versuchungen usw. beruhen. Demgegenüber braucht der Einsetzer von KI Willensstärke, um den Einsatz taktischer KI gegen die dann folgende Unsicherheit, Unruhe usw. abzuschotten, bis die KI endlich Wirkung zeigt. An dieser Stelle sei darauf hingewiesen, dass die genannten Seelenkräfte als Un-Werte ebenfalls in eine Wertebetrachtung der Inneren Führung einfließen müssen. Denn Werte verfügen nicht nur über positive, sondern auch negative Bedeutungen. Als Untugenden, die den Tugenden ihnen gegenübergestellt sind, kann man einige von ihnen etwa anhand der Aristotelischen Nikomach'schen Ethik studieren.⁶⁰

56 Ders., E-27.

57 Persönliche Kommunikation, 13. April 2022.

58 Klapproth, „Time and decision making in humans“.

59 Ebd.

60 Birnbacher, Analytische Einführung in die Ethik, S. 68.

4 Wie die Innere Führung in die Maschine kommt

Wertebasierte Technik mit IEEE7000™-2021

Erst im September 2021 hat das global renommierte Institute of Electrical and Electronics Engineers (IEEE) den Standard IEEE 7000™-2021 für VBE publiziert, der eine schwierige Aufgabe meistert: die Fusion von Geistes- und Naturwissenschaften. Der Standard, der sowohl einen bestimmten Prozess als auch Arbeitsergebnisse (Outcomes) fordert, um volle Konformität von Technik mit geltendem Recht und anderen gesellschaftlichen Forderungen nachzuweisen, hat einen wohl-durchdachten philosophischen Kern voll innerer Schönheit. Diese ist weniger dem Standard selbst geschuldet als dem ihm zugrundeliegenden Ethikpluralismus, der auf den Theorien großer (westlicher) Philosophen besonders des 18. und 19. Jahrhunderts beruht. Dabei schließen potenzielle zukünftige Versionen des Standards nicht aus, auch Axiologien aus anderen Kulturregionen der Erde – denn das IEEE ist weltweit repräsentiert – in Betracht zu ziehen.

IEEE 7000™-2021 ist ein globaler Wurf, der auf Systemdesigns aus allen Industrien aller Länder anwendbar ist. Insbesondere ist IEEE 7000™-2021 auch für militärische Systeme und Defense AI geeignet. Das ist nicht selbstverständlich, denn eine Abstimmung unter den Mitgliedern des Standardisierungsgremiums zur Anwendbarkeit des Standards auf Systeme der Verteidigung ging denkbar knapp aus. Schließlich war eine Intention, „Start-ups die Zähne zu ziehen“.⁶¹ Dabei steht fest, dass der Staat auf ganz andere Weise an verfassungsmäßig garantierte Rechte und den Rechtsstaat gebunden ist als private Unternehmen. Die Not, rechtlich und ethisch konforme Technik zu bauen und einzusetzen, ist beim Staat ungleich größer als bei privaten Organisationen. Dementsprechend hoch ist das (supra-)staatliche Interesse an IEEE 7000™-2021, von dem auch NATO oder ISO erwägen, ihn einzuführen.⁶²

IEEE 7000™-2021 gehört zu einer Normenreihe, von denen sich die auf Norm 7000 nachfolgenden Standards mit technischen Details befassen: mit der Transparenz autonomer Systeme (7001), dem Umgang mit persönlichen Daten (7002) oder algorithmischen Vorurteilen (7003). Norm 7000 hingegen erfordert exaktes philosophisches Arbeiten. Das ist eine große Herausforderung für Technologen und Ingenieure, denen häufig nichts ferner liegt, als ihre Technik auf interpretationsbedürftige nicht-technische Bedürfnisse abzustimmen. Das ist übrigens auch die Herausforderung, auf die andere internationale Organisationen für technische Normung stoßen. Technisch versiert, sind sie sprachlos, wenn es um Recht und Ethik geht.

Zur sozialen Innovation von VBE gehören deshalb sogenannte Value Leads. Sie bilden einen neuen Berufsstand und müssen in Fragen der Wertetheorie geschult

61 Sarah Spiekermann, persönliche Kommunikation, 17. Februar 2022.

62 Zu ISO/IEC/IEEE 24748-7000 siehe auch: <https://standards.iteh.ai/catalog/standards/iso/c6ac9a0c-e70d-4acf-9e44-2b730be96522/iso-iec-ieee-fdis-24748-7000> (letzter Zugriff: 17. Oktober 2022).

sein. Bei der Systementwicklung fällt ihnen die Begriffsarbeit, wie wir sie in Kapitel 3 angedeutet haben, und die ethische Anforderungsanalyse zu. Dabei wirken sie in gleicher Weise als Sammler, Strukturierer und Moderatoren, die sich – analog zum Produktmanagement – in jede gängige Vorgehensweise für die Softwareentwicklung nahtlos einfügen können.⁶³ Value Leads, die für die Outcomes, die der Standard fordert, wesentlich mit verantwortlich zeichnen, verfügen sowohl über Kontext- und Domänenwissen, das die Wertqualitäten einer Domäne kennt, als auch über eine ethische Bildung, mit dem sie (objektive) ethische Anforderungen an eine taktische KI sowohl mit deren Stakeholdern als auch mit der Literatur und dem geltenden Recht und Gesetzesvorhaben abgleichen. Sie selbst verfügen über die Tugend der Apatheia, der Zurückhaltung der eigenen Meinung, und sollten nicht ihre eigenen Vorstellungen, sondern müssen insbesondere die Forderungen der Stakeholder einer Maschine berücksichtigen.⁶⁴

VBE mit IEEE 7000TM-2021 stellt sich der Unordnung entgegen, die in einen relevanten Bereich der Ethik Einzug gehalten hat und von der schon Hannah Arendt sagte: „Wir würden uns um all dies nicht zu kümmern brauchen, wenn die Metaphysik und dieses ganze Wertegeschäft nicht so heruntergekommen wären“.⁶⁵ Mit der Werteverwirrung, die schon Hannah Arendt bemängelte, können Menschen vermutlich umgehen, Maschinen ganz sicher nicht. Die besondere Anziehungskraft von IEEE 7000TM-2021 liegt daher in der Logik und Ausgestaltung des Standards, dort, wo es um Werte, ihre Sammlung, Organisation und Übersetzung in technischen Anforderungen geht. Ihm liegt Max Schelers Axiologie zugrunde. Während Immanuel Kant die Welt ethisch durch Ideale und Werte, die der subjektiven Vernunft zugänglich sind, „alles an die Enge des Menschen gebunden“ hat,⁶⁶ bricht Max Scheler mit dem Subjektiven und führt mit seiner Materialen Wertethik einen Objektivismus ein, der zur Unendlichkeit hin offensteht.⁶⁷ Schelers Wertetheorie objektiviert die Wertedefinition, indem er feststellt: Werte liegen a priori ohne subjektive Erkenntnis und unabhängig vom Kontext vor.⁶⁸ Damit wird der Standard kultur- und nationenübergreifend anwendbar und dies unabhängig davon, ob es sich bei der in Frage stehenden Technik um ein kommerzielles System oder eines der Verteidigung handelt. Für die Verteidigung ist VBE mit IEEE 7000TM-2021 ohne Abstriche, aber auch ohne Spezialisierung vollumfänglich anwendbar.⁶⁹

63 Dazu gehören das Wasserfallmodell, das V-Modell XT oder auch die agile Softwareentwicklung bis hin zur agilen Gesamtorganisation nach dem SAFe® Vorgehensmodell.

64 Sarah Spiekermann, persönliche Kommunikation, 17. Februar 2022.

65 Arendt, Das Urteilen, S. 172.

66 Przywara, Humanitas: Der Mensch gestern und morgen, S. 30-31.

67 Ebd.

68 Zur „Unabhängigkeit des Seins der Werte von Dingen, Gütern, Sachverhalten“, vgl. Scheler, Der Formalismus in der Ethik und die Materiale Wertethik, S. 7.

69 Einen sinnvollen Vorschlag zur tieferen Ausgestaltung des Standards hat die Innovation Unit der NATO gemacht. Dort wird die Frage geprüft, ob nicht bestehende Tools des Standards einzelne Prozessschritte des Standards detaillieren können.

4.1 Von Interesse: Operational Concept and Context Exploration

Doch zurück zur taktischen KI. Schelers offener Blick in die Unendlichkeit ist hier dringend geboten, denn die eingangs erwähnten Listen von Qualitäten „robuster“ oder „vertrauenswürdiger“ oder „ethischer“ KI, wie sie etwa von der HLEG AI aufgestellt wurden, führen in eine vordefinierte Enge und genügen nicht, wertebasierte KI zu implementieren. Sprachlich ungenau und unsystematisch, legen sie eine Vorauswahl von Phänomenen fest, die einen Ingenieur beeinflussen und seine Kreativität behindern, anstatt ihn zu motivieren, Werte selbst zu explorieren, die für ein definiertes KI-System tatsächlich relevant sind. Demgegenüber beurteilt VBE jeweils nur ein einzelnes System of Interest (SOI) in seinem ganz eigenen Kontext. Jede einzelne KI verfügt über ihre eigenen Wertqualitäten, die sehr kontextspezifisch sind – so wie Obst den Wert des Wohlgeschmacks trägt, aber die Qualität des Wohlgeschmacks eine andere ist, je nachdem, ob es sich beim Obst um eine Erdbeere oder eine Avocado handelt – und definitiv über die genannten Listen hinausgehen werden. Deshalb kann nach IEEE 7000™-2021 auch nur ein einzelnes KI-System, nicht aber das produzierende Unternehmen insgesamt zertifiziert werden.

Zur Veranschaulichung eines SOI stelle man sich vor, eine taktische KI werde im ausrangierten FlaK-Panzer Gepard dazu eingesetzt, das Gerät mit einer Besatzung von ehemals drei Personen in einen technisch autonomen Roboter zu transformieren. Aus dem Gepard soll mit Hilfe von taktischer KI ein Smart Gepard werden⁷⁰ – ohne jede Besatzung. Welchen Umfang hätte dann das SOI? Ginge es ausschließlich um die KI-Software oder sogar nur eine ihrer Komponenten? Dann gehörten wahrscheinlich auch die Daten dazu, aufgrund derer das SOI aktiv würde, also die Sensorik. Hinge der Smart Gepard von der eigenen Sensorik ab, müsste auch seine Sensorik in die VBE-Wertanalyse einbezogen werden. Vielleicht würde seine überholte Sensorik aber gar nicht mehr gebraucht. Bezüge der Smart Gepard stattdessen seine Daten aus einem Netzwerk, müsste analysiert werden, wer die Daten bereitstellt, die ihn zu einer Aktion veranlasste. Hier zeigt sich bereits, wie wichtig die Auswahl der Zulieferer eines SOI ist. Sind die Zulieferer aufrichtig und machen transparent, wie sie etwa mit Daten umgehen?

Zur Analyse des SOI gehört auch, in welchem Kontext ein Smart Gepard eingesetzt würde. Handelte es sich dabei nur um einen Demonstrator, an dem gezeigt werden sollte, wie taktische KI Investitionsschutz und Lebenszyklusver-

⁷⁰ Vieregge, „Fähigkeitslücke der Bundeswehr schließen“.

längerung von Altgeräten erlaubt? Oder würde der Smart Gepard in der Ukraine in den Kampfeinsatz gehen? Vom Einsatzkonzept und vom Einsatzumfeld hängt ab, wessen und welche Werte betroffen sind. Am ersten Fall sind vielleicht nur die staatliche Rechnungsprüfung, der Bundestag oder die Beschaffungsbehörden, vielleicht auch die KI-Forschung selbst, interessiert. Im zweiten Einsatzszenario geht es hingegen um Leben und Tod auf dem Gefechtsfeld, um die Interessen von Militärangehörigen, Zivilisten und gegnerischen Regierungen. Vom Einsatzkontext also hängt ab, wer die Stakeholder des Smart Gepard sind und welche ihrer Werte vom SOI betroffen sind.

4.2 Die Werteerkundung

Kernwerte, Wertqualitäten und Wertdispositionen eines SOI

Wenn Werte ganz allgemein als (gute oder schlechte) „Eigenschaften“ definiert sind, sei an dieser Stelle kurz darauf hingewiesen, dass mehrere Dimensionen von Werten existieren. Werte können moralischer oder nicht-moralischer Natur sein, es gibt intrinsische und extrinsische Werte und je nach Wertetheorie eine oder keine Werthierarchie.⁷¹

Um herauszufinden, wie Technik das Potenzial guter Eigenschaften verwirklichen kann, welches ein Einsatzkontext bereithält, werden gleich im Frühstadium des VBE die Stakeholder eines SOI involviert. Ihre Überlegungen werden mit Hilfe von philosophischen Gerüsten der Ethik erfasst. Es ist dieser philosophische Teil von VBE mit IEEE 7000™-2021, der besondere sprachliche Präzision erfordert.

Ein hoher, ein Kernwert (Core Value) für die Bundeswehr ist die Menschenwürde. In westlichen Demokratien mit ihrem Menschenbild des freien Menschen ist die Menschenwürde ein intrinsischer Wert. Ein intrinsischer Wert ist ein an sich wertvolles Phänomen, welches um seiner selbst willen wünschenswert ist. Der intrinsische Wert der Menschenwürde ist für eine Demokratie so wichtig, dass er nicht nur durch Gesetz, sondern auch durch die Technik, nicht nur durch den Staat, sondern auch durch private Dritte gewahrt werden muss.

Die Menschenwürde kommt durch viele Wertqualitäten (Value Demonstrators) zum Ausdruck, die alle auf die Menschenwürde einzahlen. Die Wertqualitäten bedingen den Kernwert; sie gehen in den Kernwert ein.⁷² So setzt sich ein Kern-

⁷¹ Im Detail dazu: Birnbacher, Analytische Einführung in die Ethik, S. 242 ff.

⁷² Scheler, Der Formalismus in der Ethik und die Materiale Wertethik, S. 15.

wert aus vielen Wertqualitäten zusammen. Beistandspflicht, Barmherzigkeit, Diskriminierungsfähigkeit, Eigenschutz der Panzerbesatzung im Gefecht oder Entscheidungsfähigkeit sind nur einige der Wertqualitäten, die Angehörige der Bundeswehr nennen und die die Qualität der Menschenwürde ausmachen.

In den Wertdispositionen (Value Dispositions) eines SOI schließlich verwirklichen sich die Wertqualitäten im SOI. Sie können sowohl technischer als auch organisatorischer Natur sein. Wird beim Smart Gepard etwa „der Mensch rausgenommen“⁷³ und der Gepard durch taktische KI zum selbststeuernden Roboter transformiert, um den Eigenschutz zu erhöhen, fallen technische und organisatorische Vorkehrung zusammen: technisch, weil die taktische KI eine Quasi-Automatisierung der Panzerbesatzung ist, und organisatorisch, weil die Besatzung den Panzer nicht mehr selbst bedienen muss, sondern sicher in einiger Entfernung vom Gefechtsfeld agiert.

Wert-Arbeit: Das SOI im Lichte dreier Ethiktheorien

Wurde das SOI hinreichend bestimmt, erforschen seine Stakeholder sein Werteverseum und bedienen sich dafür der drei Ethiktheorien des Utilitarismus, der Pflicht- und der Tugendethik. Im Lichte dieses Ethikpluralismus beantworten sie die Frage, welche Maßnahmen positive Wertqualitäten stärken und negative vermeiden können. Das Vorgehen vervollständigt nicht nur die Werteanalyse, sondern bewirkt auch, dass das Einsatzkonzept eines SOI weiter Gestalt annimmt.

Der Utilitarismus gehört zur konsequenzialistischen Ethik, die grundsätzlich der deontologischen Ethik entgegengesetzt wird. Stark vereinfacht gesagt: Bei der konsequenzialistischen Ethik kommt es auf die guten Folgen einer Handlung an. Bei der deontologischen Ethik ist es gerade umgekehrt: Was zählt, ist die moralische gute Handlung gerade ohne Ansehen ihrer Folgen, wie sie etwa die Kant'sche Pflichtethik fordert. Es gibt Mischformen; darüber hinaus bestehen Unterschiede beider Theorien auch darin, dass die deontologische Ethik moralische Werte verwirklicht sehen will, der Utilitarismus hingegen nur einen einzigen nicht-moralischen Wert kennt: den Nutzen. Der Utilitarismus gilt deshalb auch als monistisches System. Hinzu kommt, dass beim Utilitarismus Wertekonflikte entfallen, während die deontologische Ethik die Abwägung kennt. Ergänzt wird das Duo von der Tugendethik, die keine weitere Ethiktheorie einführt, sondern die Motivation, Charakterzüge, Charismen eines Individuums betrachtet, die zu moralischem Handeln führen.

⁷³ 21strategies, F-1.

Beginnend mit der utilitaristischen Ethik, im 18. Jahrhundert begründet von Jeremy Bentham und John Stuart Mill, reflektieren die Stakeholder unter Anleitung des Value Lead den Smart Geparad im Lichte seines Nutzens: „Bei mir geht es letztendlich nur um Wirkung“.⁷⁴ Für den Utilitaristen zählt nur dieser einzige Nutzen als intrinsischer Wert. Objektive Werte wie Freiheit, Gerechtigkeit oder Gemeinwohl gelten dem Utilitarismus als extrinsische Werte, „ihr Wert ist abhängig von dem Wert der von ihnen bewirkten subjektiven Zustände“.⁷⁵ Hieraus ergibt sich eine Art Staffelung von Werten, die höher gewichtet, was dem Nutzen zuträglicher ist. Anders als bei der deontologischen Ethik entfällt beim Utilitarismus damit die Abwägung von Werten gegeneinander.

Für die Prüfung des SOI im Sinne der Pflichtethik ist Immanuel Kants Kategorischer Imperativ Lehrbuchbeispiel: „Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.“⁷⁶ Eisern gilt: „Ich möchte, dass x sichergestellt ist.“ Der Kategorische Imperativ ist unbedingter Sollens-Satz: „Es widerspricht eigentlich dem Menschlichen, der menschlichen Natur, wenn man Sadist ist oder Mörder.“⁷⁷ Du sollst nicht töten; du sollst nicht quälen, dies sind allgemein gültige Maximen, aus denen moralisches Handeln folgt.

Beim Einsatz des Smart Geparad stehen gleich zwei menschliche Leben auf dem Spiel: das eigene und das des Gegners. Käme der Smart Geparad insgesamt als SOI – und nicht nur sein Teilsystem taktische KI – in Betracht, würde der Eigenschutz der eigenen Truppe maximal gefördert, weil der Smart Geparad gänzlich ohne menschliche Besatzung auskäme. Für das Leben der gegnerischen Soldaten hingegen gelten Sonderregeln. Vom allgemeinen Grundsatz: „Du sollst nicht töten“, wird beim Einsatz auf dem Gefechtsfeld eine gesetzliche Ausnahme gemacht, denn das Töten gegnerischer Kräfte in Uniform ist grundsätzlich erlaubt, obwohl die Praxis nahelegt: „So richtig schießen wollte keiner (im Einsatz), weil jeder Angst hatte, sofort hinterher vom Staatsanwalt (strafrechtlich verfolgt zu werden).“⁷⁸

Zuletzt wird die Frage, wie das SOI den Charakter seiner Stakeholder untergräbt, im Lichte der Tugendethik erwogen. Wie würde sich der Umgang mit dem Smart Geparad langfristig negativ auf den Charakter der betroffenen Stakeholder auswirken, träfe man nicht die entsprechenden Vorkehrungen? Der ständige Umgang mit Technik verändert uns Menschen. Welche Tugenden untergräbt das SOI? Welche Laster fördert es? Wird ein Smart Geparad den Einsetzer unvor-

74 21strategies, I-19.

75 Birnbacher, Analytische Einführung in die Ethik, S. 218.

76 Birnbacher, Analytische Einführung in die Ethik, S. 141.

77 21strategies, G-11.

78 21strategies, G-34.

sichtiger machen? Wird die Tötung des Gegners banalisiert, wenn man ein Gerät auf größere Entfernung einsetzen kann, ohne selbst physisch am Gefecht teilzunehmen? Ohne die soldatischen Tugenden zu bewerten oder in eine Rangfolge zu bringen, fragen sich die Stakeholder unter der Anleitung des Value Lead hier ein Smart Gepard dazu beitragen kann, dass den Operateuren bei Nutzung des Geräts kein Schaden – physisch oder psychisch – entsteht:

Ich finde den Punkt sehr interessant, dass Anwendungen oft das Potenzial haben, Menschen so zu unterstützen, dass sie keine oder weniger Fehler machen, (die Menschen ein Leben lang belasten würden), weil die ein Gewissen haben. Also (quasi) als Hilfe, ein sauberes Gewissen zu behalten. Das finde ich einen sehr wichtigen Aspekt“.⁷⁹

Die Wertebereinigung

Um die Wertebereinigung abzuschließen, konsolidiert und clustert der Value Lead Kernwerte zusammen mit ihren Wertqualitäten.

Ein Stakeholder hat vielleicht geäußert, dass die taktische KI eines Smart Gepard ihre Berechnungen nicht unautorisiert in die Tat umsetzen darf. Er wünscht sich Man in the Loop. In der Sprache des Standards entfaltet sich der Kernwert der Menschenwürde dann in der Wertqualität der Autorisierung des Zielbeschusses: „Human dignity demonstrates itself in human authorization of fire.“ Die Autorisierung drückt eine Wertqualität des SOI aus, die die Menschenwürde Dritter, aber auch die eigene Würde schützen soll; sie muss technisch verankert werden. Wie die Autorisierung erfolgt, steht an dieser Stelle noch nicht fest. Vielmehr bestehen potenziell mehrere Möglichkeiten der Autorisierung, insbesondere auch im Lichte des zeitlichen Aspekts, wie oben angedeutet.

Das Ergebnis der Wertebereinigung sind Wertclusterdarstellungen, ergänzt um kuratierte Tabellen, in denen die Aussagen von Stakeholdern jeweils intrinsischen Kernwerten zugeordnet sind. Die Tabellen sind essenziell für die Nachverfolgbarkeit und die Zuordnung von Stakeholder-Äußerungen zu später abgeleiteten Systemfunktionen. Sie werden von den Stakeholdern des SOI abgezeichnet und im Anschluss daran priorisiert.

Bei der dann folgenden Priorisierung stellt der Standard einmal mehr unter Beweis, dass er keinen Moralpaternalismus betreibt. Denn jede Priorisierung von

⁷⁹ 21strategies, I-105.

Wertclustern muss insbesondere auch die Corporate Story des Herstellers des SOI berücksichtigen. Für die Prioritätensetzung kommt es also wesentlich darauf an, welche Mission und Vision der Hersteller des SOI verfolgt. Seine berechtigten wirtschaftlichen Interessen sollen und müssen in Betracht gezogen werden.

Prioritäten werden, zweiter Punkt, auch von der aktuellen und zukünftigen Rechtslage, die das SOI betrifft, beeinflusst. Zur Expertise des Value Lead gehören deshalb auch Kenntnisse des geltenden Rechts, aktueller Vorhaben der Gesetzgebung, der Rechtsprechung, der Philosophie und der (wissenschaftlichen) Literatur. Schon Gesetzesvorhaben wie der EU AI Act wirken sich auf eine taktische KI aus, die, wenn vorhanden, auch das Soft Law eines Herstellers würdigen muss, da bei dessen Verletzung empfindliche finanzielle Strafen drohen können.

4.3 Von der Theorie in die Praxis: Ethical Value Requirements

Von Wertqualitäten wie „Autorisierung des Zielbeschusses“ kann man nicht direkt zur Implementierung übergehen. Denn Wertqualitäten beschreiben noch keine Wertdispositionen. Diese Rolle übernehmen die Ethical Value Requirements, kurz: EVR, die nicht nur technischer, sondern auch organisatorischer oder sozialer Natur sein können.

Für die Autorisierung des Zielbeschusses beim Smart Gepard müsste eine taktische KI Informationen für ihren Einsetzer bereitstellen. So könnte eine taktische KI beispielsweise (i) den Grad ihrer Überzeugung kommunizieren, wonach es ein legitimer Gegner ist, der bekämpft werden soll („98,9 Prozent Wahrscheinlichkeit“) und (ii) einen Countdown („n Sekunden“) anzeigen, wie lange ein Zielbeschuss noch sinnvoll und wirksam wäre und (iii) einen Bedienhebel bereitstellen, bei dem der Mensch die Bekämpfung eines Ziels per Knopfdruck an die taktische KI übergibt.

Die Definition von EVR ist der wichtigste Arbeitsschritt von VBE mit IEEE 7000™-2021. Dabei ist darauf zu achten, dass EVRs so formuliert werden, dass sie, wenn sie einmal als Systemfunktionalität implementiert wurden, test- und bewertbar sind. Das Ergebnis: Die Ethik einer KI wird messbar! „Ein Countdown zeigt an, bis zu welchem Zeitpunkt ein Zielbeschuss sinnvoll wäre.“ Bediener und Bedienerinnen wollen sich auf den angezeigten Wert verlassen können – sie wollen sicher sein, dass das Berechnungsverfahren korrekt implementiert ist, dass Unsicherheiten der Rohdaten in die Berechnung mit einfließen oder dass die Berechnung statistisch getestet wurde.

Auf die EVRs folgt ein letzter wichtiger Schritt. In der Interaktion mit den Stakeholders führt der Value Lead eine Risikobetrachtung durch. Er sucht Antwort auf die Frage, wie gefährdet die spezifizierten EVRs sind. Wodurch ist ein EVR bedroht, und welche Vorkehrungen müssen getroffen werden, um ein EVR sicherzustellen? Von Systementwicklern kann etwa verlangt werden, dass der eben erwähnte Countdown nicht nur auf bekannten, sondern auch auf unbekanntem Datensätzen richtig berechnet oder der Testdatensatz zusammen mit dem System veröffentlicht wird.

Standardrisiken, zu denen beispielsweise eine unzureichende Testabdeckung gehört, trägt der Value Lead auf eine Risikomatrix auf und misst ihnen eine Wahrscheinlichkeit zu. Denn nicht alle identifizierten Risiken treten mit derselben Wahrscheinlichkeit ein. Es gilt die Faustregel: Je größer das Risiko einer Bedrohung, desto höher muss auch eine Systemanforderung (Control) gewichtet werden, die später auch von der Softwareentwicklung bei der Implementierung des SOI priorisiert werden muss. Der Kontrollmechanismus ist so zu spezifizieren, dass er in der Lage ist, das EVR sicherzustellen.

Als Standardrisiko gilt aber nicht, wenn intrinsische Werte und ihre Wertqualitäten bedroht sind: „Mit 98,9 Prozent Wahrscheinlichkeit ist das Ziel ein legitimer Gegner.“ Droht Hazard, verbietet sich jedes Ranking. Für dessen Risikobestimmung nimmt der Value Lead deshalb eine Technikfolgenabschätzung vor, während das Entwicklerteam zum frühestmöglichen Zeitpunkt Vorkehrungen zur Bewältigung und Minderung von Risiken mit sehr hohem Gefährdungspotenzial treffen muss.

5 Fazit

Genau genommen ist das, was Sie schaffen, quasi die Innere Führung für KI. Man könnte dieses Wertekonstrukt der Inneren Führung übertragen. Es schließt sich nicht aus.⁸⁰

Dass auch KI das Werteuniversum, das Angehörige der Bundeswehr der Inneren Führung zumessen, berücksichtigt, ist ein dringender und rechtlich gebotener Wunsch der Bundeswehr. Mit wertebasierter Technik nach IEEE 7000™-2021, einem globalen Standard, der für Systeme der Verteidigung nicht adaptiert werden muss, aber gegebenenfalls durch bestehende Standards wie das NATO Risk Assessment Tool detailliert werden könnte, kann er sich erfüllen, doch dafür ist weit mehr gefragt als die begrenzten Listen aus Politik und Industrie mit vordefinierten Anforderungen an die Softwarequalität einer KI. Inzwischen rollt sich der Standard aus, und Ausbildungszentren für Value Leads, die über die Gabe und die Ausbildung sowohl eines philosophischen als auch eines technischen Verständnisses verfügen, breiten sich aus. Nichts weniger als das wird einem Standard gerecht, dessen innere Schönheit in Logik und Struktur seiner Wert-Arbeit liegt, die nur eines zum Ziel hat: dass bessere Technik dem Wohl der Menschen dient.

⁸⁰ 21strategies, A-173.

Literatur

„DARPA SCEPTER program will explore machine-generated strategies for warfare“, Defense Techconnect, 24. Januar 2022, <https://defensetechconnect.com/2022/01/24/darpa-scepter-program-will-explore-machine-generated-strategies-for-warfare/> (letzter Zugriff: 17. Oktober 2022):

Arendt, Hannah, *Das Urteilen* (München: Piper, 2020).

Birnbacher, Dieter, *Analytische Einführung in die Ethik*. 3. durchgesehene Edition (Berlin: De Gruyter, 2013).

Funk, Michael, *Roboter- und KI-Ethik: Eine methodische Einführung*. Band 1. (Wiesbaden: Springer Vieweg, 2022).

Kants gesammelte Schriften. Herausgegeben von der Königlich preußischen Akademie der Wissenschaften (Berlin: G. Reimer, 1900 ff.).

Klapproth, Florian „Time and decision making in humans“, *Cognitive, Affective & Behavioral Neuroscience*, 8:4 (December 2008), S. 509-524 <https://link.springer.com/article/10.3758/CABN.8.4.509> (letzter Zugriff: 17. Oktober 2022).

Precht, Peter und Burkard, Franz-Peter (Hrsg.), *Metzler Lexikon Philosophie. Begriffe und Definitionen*. 3., erweiterte und aktualisierte Auflage (Heidelberg und Berlin: J.B. Metzler'sche Verlagsbuchhandlung, 2008).

Przywara, Erich, *Humanitas: Der Mensch gestern und morgen* (Nürnberg: Glock und Lutz Verlag, 1952).

Reichenbach, Hans, *The Direction of Time* (Mineola, N. Y.: Dover Publications, 1956).

Scheler, Max *Der Formalismus in der Ethik und die Materiale Wertethik. Neuer Versuch der Grundlegung eines ethischen Personalismus* (Halle an der Saale: Verlag von Max Niemeyer, 1916).

Spaeth, Andreas, „Grundlegender Softwarefehler in der Boeing 737 Max gefunden“, *Neue Zürcher Zeitung*, 5. August 2019, <https://www.nzz.ch/mobilitaet/luftfahrt/boeing-737-max-grundlegender-softwarefehler-gefunden-ld.1499955> (letzter Zugriff: 17. Oktober 2022).

Spiekermann, Sarah „Value-based Engineering: Prinzipien und Motivation für bessere IT-Systeme“, *Informatik Spektrum*, 44:4 (August 2021), S. 247-256 (2021), <https://link.springer.com/article/10.1007/s00287-021-01378-4> (letzter Zugriff: 17. Oktober 2022)

Spiekermann, Sarah, „Die Ethik in der Künstlichen Intelligenz“, in: Klaus Mainzer (Hrsg.): *Philosophisches Handbuch der Künstlichen Intelligenz* (München: Springer Reference Geisteswissenschaften, im Druck).

Vieregge, Kerstin, „Fähigkeitslücke der Bundeswehr schließen“, Pressemitteilung vom 15. März 2022, <https://www.kerstin-vieregge.de/artikel/faehigkeitsluecke-der-bundeswehr-schliessen> (letzter Zugriff: 17. Oktober 2022).

Vöneky, Silja, „Ethische Experten und moralischer Autoritarismus“, in: Silja Vöneky et al. (Hrsg.) *Legitimation ethischer Entscheidungen im Recht: Interdisziplinäre Untersuchungen* (Berlin/Heidelberg: Springer, 2009), S. 85-97.

Wahl, Thomas, „High-Level Expert Group Publishes Ethics Checklist (2020)“, <https://eucrim.eu/news/ai-high-level-expert-group-publishes-ethics-checklist/> (letzter Zugriff: 17. Oktober 2022).

Defense AI Observatory Studies

- 22|06 Yvonne Hofstetter, *Wie KI Innere Führung lernt. Wertebasierte Technik mit IEEE7000™-2021*
- 22|05 Andrea Gilli, Mauro Gilli, and Ivan Zaccagnini, *Exploring the Benefits of a New Force Enabler: Defense AI in Italy*
- 22|04 Kenneth Payne, *Bright Prospects – Big Challenges. Defense AI in the United Kingdom*
- 22|03 Heiko Borchert, Christian Brandlhuber, Armin Brandstetter, and Gary S. Schaal, *Free Jazz on the Battlefield. How GhostPlay's AI Approach Enhances Air Defense*
- 22|02 Peter Layton, *Evolution not Revolution. Australia's Defence AI Pathway*
- 21|01 Heiko Borchert, Torben Schütz, Joseph Verbovsky, *Beware the Hype. What Military Conflicts in Ukraine, Syria, Libya, and Nagorno-Karabakh (Don't) Tell Us About the Future of War*

