



# How AI Learns the Bundeswehr's “Innere Führung”

**Value-Based Engineering with IEEE7000™-2021**

### **About the Defense AI Observatory**

The Defense AI Observatory (DAIO) at the Helmut Schmidt University in Hamburg monitors and analyzes the use of artificial intelligence by armed forces. DAIO comprises three interrelated work streams:

- Culture, concept development, and organizational transformation in the context of military innovation
- Current and future conflict pictures, conflict dynamics, and operational experience, especially related to the use of emerging technologies
- Defense industrial dynamics with a particular focus on the impact of emerging technologies on the nature and character of techno-industrial ecosystems

DAIO is an integral element of GhostPlay, a capability and technology development project for concept-driven and AI-enhanced defense decision-making in support of fast-paced defense operations. GhostPlay is funded by the Center for Digital and Technology Research of the German Bundeswehr (dtec.bw).

Ein Projekt im Rahmen von

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

# How AI Learns the Bundeswehr's “Innere Führung”

**Value-Based Engineering with  
IEEE7000™-2021**

Yvonne Hofstetter and Joseph Verbovsky

DAIO Study 23|10

Ein Projekt im Rahmen von

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

## About the Authors

Prof. Yvonne Hofstetter is IEEE700™ Value Lead for GhostPlay, Founder and Co-Managing Director of 21strategies GmbH and Honorary Professor for Digitalization and Society at the Center for Ethics and Responsibility at Bonn-Rhein-Sieg University of Applied Sciences. She is the German representative in the new NATO Data and Artificial Intelligence Review Board (NATO DARB) and has been working at the interface of product and technology development as well as ethics and legal issues since 1999. Prof. Hofstetter has written several bestselling nonfiction books, including *Sie wissen alles* (Munich: C. Bertelsmann Verlag, 2014), *Das Ende der Demokratie* (Munich: C. Bertelsmann Verlag, 2016), and *Der unsichtbare Krieg* (Munich: Droemer Knaur, 2019).

Joseph Verbovzsky is DAIO's Research Fellow for defense technology-industrial affairs. Before joining DAIO, he worked in the areas of international coordination and strategic analysis at different defense companies. He holds a master's degree in international relations and economics from the Johns Hopkins School of Advanced International Studies (SAIS). He graduated with a PhD on German Structural Pacifism from the University of the Bundeswehr München. His Twitter account is @warkhorse.

## Design

Almasy Information Design Thinking

## Imprint

Yvonne Hofstetter and Joseph Verbovzsky, *How AI Learns the Bundeswehr's "Innere Führung."* Value-Based Engineering with IEEE7000™-2021, DAIO Study 23/10 (Hamburg: Defense AI Observatory, 2023).

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

ISSN (online): 2749-5337

ISSN (print): 2749-5345

# Content

1 Summary .....	6
2 “Innere Führung”: A Sketch .....	7
3 Can AI Do “Innere Führung?” A Critique of Tactical AI for the Bundeswehr.....	12
3.1 Values? What Values then?.....	14
3.2 A Life in Freedom for AI?.....	15
3.3 Who Bears the Responsibility? .....	17
3.4 A Question of Time?.....	19
4 How “Innere Führung” Comes to the Machine: Value-Based Engineering with IEEE7000™-2021.....	22
4.1 Of Interest: Operational Concept and Context Exploration.....	25
4.2 The Value Exploration.....	26
4.3 From Theory to Practice: Ethical Value Requirements.....	29
5 Conclusion.....	31
Literature.....	33

# 1 Summary

The increasing ubiquity of Artificial Intelligence (AI) poses significant political consequences. The rapid proliferation of AI over the past decade has prompted legislators and regulators to attempt to contain AI's technological consequences. For Germany, relevant design requirements have been expressed by the European Commission's High-Level Expert Group on Artificial Intelligence (HLEG AI), and, at the national level, by the German government's Data Ethics Commission (DEK) as well as the German Bundestag's Commission of Inquiry on Artificial Intelligence (EKKI). These requirements set by commissions or working groups are not binding. Where they are not transformed into sovereign law – such as the EU AI Act, which is expected to be passed and enter into force around 2025 – they can still become “soft law” based on contractual agreement between social partners or, in value chains, on the (self-) commitment to corporate social responsibility (CSR). Violation of such soft law still bears significant financial consequences. As a result, translation of the “principles of the ought-to-be, which provide orientation for our actions”<sup>1</sup> i.e., values, into technical functions of AI, becomes a priority.

This article walks through exemplary values of the Bundeswehr, which are expressed in the leadership philosophy of “Innere Führung,” and places them in the context of AI. For the elicitation of values from “Innere Führung,” officers of the Bundeswehr themselves were interviewed. Their remarks are followed by a quick tour of the standard for Value-based Engineering (VBE), ISO/IEC/IEEE 24748-7000:2022 Systems and software engineering — Life cycle management — Part 7000: Standard model process for addressing ethical concerns during system design (in short: IEEE 7000™-2021), which was at first published by the IEEE in September 2021. This new standard aspires to achieve a daunting task: to produce technology that meets the rigorous and complex system requirements of law and ethics.

---

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

# 2 “Innere Führung”

A Sketch

This study explores the application of “Innere Führung” (see Box 1) to defense Artificial Intelligence (AI) through a non-representative series of interviews with officers of the Bundeswehr in 2022.<sup>2</sup> The interview series was based on the following guiding question: “Please help us to understand the tension between ‘Innere Führung’ and the use of AI in the Bundeswehr, so that

- 1) a ‘morally good Smart Comrade’ emerges from tactical AI,
- 2) which is in line with the understanding of ‘Innere Führung,’
- 3) is willingly used by the Bundeswehr, to
- 4) improve defense and response capabilities and
- 5) establish technology leadership.”

### **Box 1: “Innere Führung” – the Moral Code of the Bundeswehr**

“Innere Führung” refers to the moral code of the Bundeswehr. This code sees German soldiers as “Citizens in Uniform,” who remain committed the principles of the German constitution, both in peace and in war. Among the most prominent principles are the following:

- Independent judgment on the morality and legality of orders. German soldiers are forbidden to issue or follow orders which are contrary to the norms and principles of the German Basic Law, and even contrary to their conscience
- The primacy of civilian political control of the armed forces through the German Bundestag (Parliament).
- A clear rejection of the Wehrmacht as a source of military tradition (except for those military officers who formed part of the German resistance to Hitler).

Source: “Was bedeutet Innere Führung.”

---

<sup>2</sup> The sample size is 10. In the following, the citations refer to the anonymized interviewees (capital letter) and the respective page number of the protocol. In order to adequately capture the meaning of the participants’ statements, most are provided in the original German. A courtesy approximate translation is offered in the footnotes.



In the following examples, a few voices from the interviews highlight the issues raised.

“Innere Führung” is unique to the 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.<sup>3</sup>*

The sources of “Innere Führung” are first and foremost the applicable law, especially the Basic Law and international law. With its recourse to the Basic Law, the Bundeswehr commits itself to a certain conception of man: it is the conception of the “person (‘per se existere’), who is individually distinct from all others (‘per se unum’) and free [because] acting out of himself (‘per se agere’).”<sup>4</sup> This definition from Thomas Aquinas elevates the person through dignity (in contrast to objects) to a legal subject and bearer of rights and legal obligations. Moreover, a person always remains sovereign, i.e., free both in its decisions and actions. This freedom of choice of the person constitutes his dignity and remains the normative foundation for individual members of the Bundeswehr.

As sovereign individuals (persons), members of the Bundeswehr are not only bearers of their own dignity. In the event of conflict, the dignity of the enemy must also be respected and upheld. For this purpose – in addition to their commitment to the Basic Law – the Bundeswehr can draw on international humanitarian law, as well as their own moral understanding of humanity – as constructed historically and culturally in Europe:

*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 wir als Staatsbürger in Uniform haben sollten, etwas zu tun oder (besser) nicht zu tun.<sup>5</sup>*

The quotation suggests that “procedural ethics without procedure,” i.e., ethics without context-related specifications such as the Kantian moral doctrine with its “intended nature of actions”, is also<sup>6</sup> a legitimate source of “Innere Führung.” This specifically refers to the knowledge (of facts) and the conscience of members of the Bundeswehr:

---

<sup>3</sup> “(‘Innere Führung’ is) a broad topic, an absolutely unique to the Bundeswehr. I have experienced this first hand, and then in the context of other militaries.” 21strategies, E-2.

<sup>4</sup> Thomas Aquinas, quoted from: Precht/Burkard, Metzler Lexikon Philosophie, p. 444.

<sup>5</sup> “Even if something is not prescribed or forbidden, there is nevertheless the principle of humanity – that one is guided (by humanity) and assesses whether it is right to do something or (better) not to do something according to the moral compass that we should have as citizens in uniform.” 21strategies, D-34.

<sup>6</sup> Vöneky, “Ethical Experts and Moral Authoritarianism,” p. 88.

*(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.<sup>7</sup>*

Ethics and morality, in addition to applicable law, are therefore prescriptive for the self-image of the Bundeswehr.

Morality, since it also concerns the “feeling of one’s own dignity and questions as to whereby our life acquires meaning or finds fulfillment,”<sup>8</sup> is to be assigned to the world of feelings according to the prevailing legal opinion.<sup>9</sup> Federal military officers likewise interpret ethics and morals as subjective feelings or intuitions:

*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.<sup>10</sup>*

Conscience as an intuition, which is shaped by an individual’s “world view, (...) religion, (...) upbringing, (...) parental example, (...) socialization at school, in the community,”<sup>11</sup> also poses risks for the Bundeswehr:

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

*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.<sup>13</sup>*

---

7 “(‘Innere Führung’) says on the one hand: Every soldier is also a thinking, feeling, ethically evaluating human being. And that first, before he or she is a soldier. And (he or she) may also, must also act according to basic ethical principles.” 21strategies, E-14.

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

9 “Against the fact that such a rational justification of normative statements is possible, many have objected (...). They have all assumed, with different reasons, that ethical statements are not statements capable of justification or truth, but either mere sham statements, or only expressions of feeling, recommendations, or imperatives.” Vöneky, “Ethische Experten und moralischer Autortarismus,” p. 89.

10 “Then there is the gut feeling, according to which someone decides... who in the end also either decides based on values or (makes) a purely statistical decision, because (they think to themselves): I have already seen this ten times. Then it will also be right the eleventh time.” 21strategies, A-63.

11 Ibid., G-14.

12 “Basically, you can only model (values) and hope that [shaping] works, but you don’t know (it).” Ibid., I-50.

13 “Giving the soldier his conscience – as a powerful weapon and assessment tool – is not easy. That’s why the military usually doesn’t do it; and that’s why it went dramatically wrong in Germany. (Nevertheless) the Bundeswehr says: We accept the problems we have with it.” Ibid., E-74.

The problems implicit in the quotations are not primarily caused by the individual soldier, but by the lack of objectively valid norms for moral action. "What used to be regarded as the basic ethical tenets of Western civilization (...) [had] been devalued over time to the level of mere conventions (as easily replaceable as a code of table manners)."<sup>14</sup> In 1965, Hannah Arendt lamented the lack of objectivity in morality and hoped at least for a "moral taste" that transcends society:<sup>15</sup> "*Es widerspricht eigentlich dem Menschlichen, der menschlichen Natur, wenn man Sadist ist oder Mörder.*"<sup>16</sup>

In addition to the normative, prescriptive power for members of the Bundeswehr as a whole, "Innere Führung" influences the understanding of leadership and hierarchy with the military organization:

*(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.*<sup>17</sup>

In this context, the concept of "Führen mit Auftrag" – often used synonymously with the term "Auftragstaktik"<sup>18</sup> – is juxtaposed to command tactics:

*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.*<sup>19</sup>

When leading with a mission, the superior military leadership entrusts subordinates to carry out the mission they have been given according to the commander's intent without the need for explicit instruction. In the sense of this intersubjective process, "Innere Führung" is considered successful when the subordinates retain their freedom of conscience and action, but at the same time the leadership itself also behaves in an exemplary manner.

---

14 Arendt, Das Urteilen, p. 166.

15 Ibid., p. 167.

16 "It actually contradicts human nature, if one is a sadist or a murderer." 21strategies, G-11.

17 "(Innere Führung' also puts) very high demands on the leadership, because I must also lead in such a way to give my soldiers this space that they themselves can and may act as people with emotions, ethics, morals according to moral aspects." Ibid., E-7.

18 Führen mit Auftrag is generally understood as leading with the mission in mind while Auftragstaktik can be translated as mission tactics.

19 "For me, these are the two main points (in 'Innere Führung'), that (i) one has a certain freedom or flexibility in the execution of orders and (ii) at the same time one does not have to implement every order but is rather required to pause and point out when something obviously criminal or suicidal is ordered." Ibid., D-20.

# **3 Can AI Do “Innere Führung?”**

**A Critique of Tactical AI for the Bundeswehr**

*Ich behaupte, wir gewinnen dann, wenn wir at machine speed arbeiten. Und das heißt, wir müssen eigentlich den Menschen rausnehmen. Großes Entsetzen immer.*<sup>20</sup>

With its principles of the rule of law and ethics and morality, the Bundeswehr engages the full potential of Third Wave AI ("tactical AI"). Reconnaissance, evaluation, situational analysis, warning, and even information aggregation have been traditional fields for the application of Second Wave AI for the military since the 1990s. But for modern defense, these capabilities alone are no longer sufficient. Today, the Bundeswehr demands strategic and tactical capabilities from machines that take place at machine speed. Tactical AI is the basis of those machines that "take the man out." Battlefield simulators and training systems, (partially) autonomous drones in the air and at sea as well as on land, smart munitions, cyber-physical systems, or cross-domain AI-controlled networks of defenders – the list of conceivable fields of application for tactical AI is extensive. At the strategic level, it aims to compress planning cycles – such as an air tasking orders – to a third or a quarter of the time required at present.<sup>21</sup> At the operational level, it can increase the combat value of legacy systems, transform them into technically (partially) autonomous weapon systems,<sup>22</sup> establish higher self-protection, optimize ammunition consumption and logistics, or even provide algorithmic tactical command and control of units during battle. In the future, the commander no longer uses AI. The AI is the commander.

Tactical AI for the Bundeswehr, however, is the subject of a debate: "There is always great horror." This is because tactical AI for the military operates in an extreme environment: it serves the state's use of force; and as tactical AI in the service of the state, it must be embedded in the legal and value framework of the Basic Law for the Federal Republic of Germany, relevant EU regulations and (humanitarian) international law, or more precisely: in the Bundeswehr's leadership principle, "Innere Führung." It proves to be a great challenge to transform legal, ethical, or social norms into the functions of a software system.

---

20 "I maintain that we win when we work at machine speed. And that means we actually have to take the human out of it. Great horror always." Ibid., F-1.

21 This aspect is addressed by the Defense Advanced Research Projects Agency's Strategic Chaos Engine for Planning, Tactics, Experimentation and Resiliency (SCEPTER) research project. See further: "DARPA SCEPTER program will explore machine-generated strategies for warfare".

22 Vieregge, „Fähigkeitslücke der Bundeswehr schließen.“

## 3.1 Values? What Values then?

The positive meanings that members of the Bundeswehr attribute to “Innere Führung” read like a colorful bouquet of good intentions and positive qualities. De facto, the officers express a plethora of ideas in connection with “Innere Führung” that trigger certain emotional states in them as “feeling and desiring subjects (...).”<sup>23</sup> What moves their souls ranges from A for adventurousness to Z for Zeitgeist.

One main reason for the high number of differing ideas expressed by the study participants is the lack of moral objectivity. That value views are subjectively influenced<sup>24</sup> and are formed against the background of certain ideologies, world views or paradigms i.e., “cultural background radiation,”<sup>25</sup> is described by a general as follows: *“Aber was wollen wir? Es wird andere geben, die sagen: Was ist das, ethisch? Ethisch ist, wenn wir gewinnen und die anderen verlieren.”*<sup>26</sup>

In addition to the subjective world of experience, there is another deficit: A “philosophically founded conceptual world” is missing,<sup>27</sup> when people talk about values. The structural properties of values are not reflected in language as the language of values is not precise enough. At the same time, the “critique of language (...) is a constitutive part of methodological philosophy and ethics.”<sup>28</sup> This inadequacy is a crucial reason for the difficulties of developers and engineers to implement tactical AI as a bearer of the desired positive ideas or even capable of supporting them.

“Technical robustness and security” or “transparency,” three meanings as defined by the HLEG AI Ethics Checklist for AI,<sup>29</sup> are generally quality attributes of technical systems,<sup>30</sup> as they were described by default in software specifications before agile software development began to prevail. They are “hygiene requirements” for software systems,<sup>31</sup> or, more pointedly expressed: They are self-evident. But for more than two decades, greatly shortened software development cycles, cost pressures, the focus on shareholder value, the commoditization of programming power, and the resulting faster time to market of software have worked in favor of a manufacturer’s or operator’s sales and profits; but at the same time, the quality

---

23 Scheler, *Der Formalismus in der Ethik und die Materiale Wertethik*, p. 10.

24 “Max Scheler was the prophet of a new objectivism. It was to be his radical demolition of the old Kantian securing subjectivism: (...) of the ethical securing of the world of ideals and values by subjective autonomy (...).” Cited according to: Przywara, *Humanitas: Der Mensch gestern und morgen*, pp. 30-31.

25 Funk, *Roboter- und KI-Ethik*, p. 17.

26 “But what do we want? There will be others who say, what is that, ethical? Ethical is when we win and they lose.” 21strategies, F-48.

27 Spiekermann „Die Ethik in der Künstlichen Intelligenz.“

28 Funk, *Roboter- und KI-Ethik*, p. 36.

29 Election, “High-Level Expert Group Publishes Ethics Checklist (2020).”

30 Spiekermann, “Value-based Engineering: Prinzipien und Motivation für bessere IT-Systeme.” See also: [https://en.wikipedia.org/wiki/List\\_of\\_system\\_quality\\_attributes](https://en.wikipedia.org/wiki/List_of_system_quality_attributes) (last accessed: October 17, 2022)

31 Sarah Spiekermann, personal communications, November 12, 2021.

of even mission-critical systems has suffered greatly. Fatal evidence is provided by the crashes of two Boeing 737 MAX-8 aircraft which were attributable to poor software design, as well as to the lack of the aerodynamics expertise of the programmers involved.<sup>32</sup>

The self-evident expected quality attributes of software systems are far from sufficient to realize core values such as human dignity and freedom, peace, justice, or virtuous character traits such as love of one's homeland, truthfulness, or courage in tactical AI. Value-based AI must therefore be preceded by meaning-making and rigorous conceptualization, a task that the IEEE 7000™-2021 standard for values-based technology imposes on value leads – a wholly new employment category in software development. Their role and scope of responsibilities in the development process will be addressed in chapter 4.

## 3.2 A Life in Freedom for AI?

Intuitively, officers mentally transfer the idea of being human to tactical AI as well. Especially, since it is currently being considered, that tactical AI could evaluate moral behavior, judge ethically or abide by human laws on its own:

*Ich gebe dem Smart Gepard die Regeln vor, an die wir uns noch moralisch und ethisch gebunden fühlen.*<sup>33</sup>

*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.*<sup>34</sup>

*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.*<sup>35</sup>

The tendency to view AI systems not as things, as objects, but as "subjects of ethical judgment"<sup>36</sup> is as questionable as it is profound. If one considers AI as

---

32 Boeing has lost experienced aircraft engineers over the last twenty years and moved software development to India; many an engineer has left the company because he knew about the shortcomings and how management dealt with the errors. For an example of the process, see: Spaeth, "Fundamental Software Flaw Found in Boeing 737 Max."

33 "I give the Smart Gepard the rules to which we still feel morally and ethically bound." 21 strategies, E-60.

34 "When we say 'Innere Führung' in AI... how do you do that? By training it in such a way that it then orients the things it is supposed to trigger to the standards that you might apply as a human being." Ibid., F-3.

35 "But that's what's so interesting about AI. (AI) is perhaps even more human than humans themselves and certainly more predictable than a soldier in certain situations." Ibid., I-57.

36 Funk, Roboter- und KI-Ethik, p. 40.

an object, then AI ethics is to be understood as people “morally scrutinizing the purposeful use (of AI).”<sup>37</sup> This is a different case from two other definitions of AI ethics, according to which (i) AI could itself make ethical considerations and (ii) AI would have to follow rules and regulations.<sup>38</sup> In the latter two cases, we would be talking about AI as a subject that could itself act morally and be subject to a code – a written or unwritten ethos.

What is remarkable about the last two quotes is the way they are expressed in relation to AI. “(AI) is perhaps even more human” is not the same as “(AI) acts perhaps even more human” or “(AI) behaves perhaps even more human.” Persons act in freedom of action, the “capacity and (the) ability to act consciously and voluntarily.”<sup>39</sup> However, tactical AI never acts in this sense, simply because it lacks the intrinsic capacity to do so. Regardless of its amazing technical skills, it has “no personal character because it is unable to give itself laws but must obey human laws.”<sup>40</sup> As a machine, it lacks consciousness, voluntariness, its own motivation. Ignoring the lack of theoretically exact formulations in the aforementioned quotes, the three interviewees used as examples are instinctively hesitant to grant AI the same freedom as humans. “AI is more human,” denotes nothing more than being in the sense of a possibility that humanity might be intrinsic to the material object called AI. “Things it is then supposed to trigger,” leaves open whether AI either acts or behaves without mental processes, or simply functions, because the utterance focuses only on the effect of AI. “Actions that the machine ended up performing,”<sup>41</sup> indicates that the actual action – the use of AI – is assigned to the soldier; it is the soldier’s action that is causal for the effect of the AI.

The object property of AI would also be on the verge of dissolution if AI were to adhere to an ethos, a code. Both sovereign law such as the Basic Law and the Rules of Engagement of the German armed forces are such codes. With a view to the discussion of values, they standardize the “ethical minimum” in the best case.<sup>42</sup> Law, however, is created by people for people – and not by people for machines. Law makes people free, because people are given the freedom to violate the law. The enforcement of rules by machines – the smart car that won’t start because its occupants haven’t buckled up yet – means coercion and moral paternalism. Thus, if an AI were able or required to observe human codes itself, it would become the “subject of moral action” in this case as well.<sup>43</sup> Would the subjection

---

37 Ibid., p. 25.

38 Ibid., pp. 69-88.

39 Precht/Burkard, Metzler Lexikon Philosophie, p. 188.

40 “Person is that subject whose actions are capable of imputation. Moral personality is therefore nothing other than the freedom of a rational being under moral laws..., from which it then follows that a person is subject to no other laws than those which he (either alone, or at least simultaneously with others) gives to himself.” Kants gesammelte Schriften, AA22.

41 21strategies, E-24.

42 “The right is nothing other than the ethical minimum.” Attributed to Georg Jellinek (1851-1911).

43 Cf: Funk, Roboter- und KI-Ethik, p. 69.



of AI to human laws then have the consequence that AI could or would also have to bear responsibility?

### 3.3 Who Bears the Responsibility?

With the imagined subjectification and individualization of AI, the human-machine blurring of the digital age would advance further. However, the concept of the person still has legal validity. Only the subject, not an object like AI, can make free ethical judgments and act freely. The human being – and only the human being – therefore bears the responsibility for freely made value judgments and his or her free actions:

*Lenkwaffen 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.<sup>44</sup>*

Responsibility for their actions is a central concern for members of the Bundeswehr, especially when they have led to the death of both comrades and enemies. It is important for officers to explain the reasons and intentions of their actions:

*Wo Menschen arbeiten, passieren Fehler. Wenn das im militärischen Einsatz schlimm wird, muss man auch zu seiner Verantwortung dort stehen, solange man sagt: Ich habe das aus diesen und diesen Gründen gemacht. Ich bin in der Conclusio zu dem Ergebnis gekommen: Ich kann das verantworten.<sup>45</sup>*

Responsibility is something that describes an intersubjective happening. It involves the discussing and the processing of events from person to person. The interviewees explicitly emphasize the differences between humans and machines, especially the human capacity for empathy, compassion, strength, intuition, mercy. Responsibility is also the reason why officers prefer human decisions to more precise and accurate machine calculations, even when a human decision has turned out completely wrong:

---

44 "I can no longer see guided missiles after I have shot them (either). I can no longer control them. I can no longer intervene, and I'm still responsible if they don't hit a fighter plane, but a civilian jet. And then it's my turn again, and rightly so." 21strategies, E-30.

45 "Wherever people work, mistakes happen. If things go wrong in a military operation, you have to accept your responsibility, as long as you say: I did this for such and such reasons. In the end, I have come to the conclusion: I can take responsibility for that." Ibid., C-118.

*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?<sup>46</sup>*

*Aber einem autonomen System verzeihen Sie keine Fehler. Einem Menschen verzeihen Sie immer Fehler oder eher Fehler.<sup>47</sup>*

Without exception, all interviewees support the use of AI; likewise across interviews, there is agreement that only the military leader is responsible for the use of AI:

*(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.<sup>48</sup>*

But the willingness to take responsibility for the use of a machine is conditional: “(W)hat can [operators] do to live up to this responsibility for the actions that the machine has carried out?”<sup>49</sup> Soldiers want to know “their” AI inside and out and what they are getting into before they put AI into real-world use: “Then I need really good training, and very realistic training.”<sup>50</sup> In addition to intensive human training with an AI, there is also a requirement that the functioning of the AI is ensured through a quality assurance process and standards:

*Da wird man wahrscheinlich Standards schaffen müssen. Wie oft und mit welcher Fehlerwahrscheinlichkeiten muss in einem gewissen Szenario die KI zu folgendem Ergebnis kommen?<sup>51</sup>*

As a second prerequisite, the respondents report – keeping with the spirit of “Innere Führung” – is that they be granted freedom of choice. They want to decide for themselves whether they want to use tactical AI or not:

---

46 “The decision of the AI could even be better in the sense of: It may have led to fewer deaths on the own side. But the relatives of the dead comrades have no one to talk to about it with. And isn’t that in the end more terrible than other consequences that can be dealt with afterwards?” Ibid., B-72.

47 “But you do not forgive an autonomous system for mistakes. A human being you always forgive mistakes or rather errors.” Ibid., I-54.

48 “(T)he only realistic, pragmatic thing we still need in the military is that the person who uses this product in the end is the one who bears the responsibility. You won’t be able to put (the responsibility for using AI) back on the developer who programmed it five years before.” Ibid., E-25.

49 Ibid., E-24.

50 Ibid., E-40.

51 “Standards will probably have to be created. How often and with what error probabilities must the AI arrive at the following result in a certain scenario?” Ibid., F-34.

*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.<sup>52</sup>*

So while the interviewees are all in favor of using AI – even with technical autonomous capabilities and even if the two prerequisites of quality assurance and training are met, and soldiers are then also willing to take full responsibility for the functioning of the AI – they still need inner strength when it comes to the time gap between the deployment decision and the effect.

### 3.4 A Question of Time?

How is it that people assume AI makes free decisions? The question becomes particularly precarious in the military use of AI, where a fight-or-kill decision is at stake. It is also the issue at stake in the debate on the outlawing of “killer robots” or “lethal autonomous weapon systems” at the Permanent Conference on Disarmament.

One officer sums it up: The problem lies with the question of time. When tactical AI is used, the effect is either severely delayed or occurs at machine speed i.e., so quickly that the human seems to be no longer in control of the AI. In addition, the soldierly action of employing AI is taken in the complex dynamic environment of the battlefield. In this dynamic environment, time is also relevant because the environment is constantly evolving. The deployer’s own actions and the functioning of the AI are thus doubly “in time.” They are closely linked to the medium of time:

*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.<sup>53</sup>*

What is expressed here is a theory of time that defines time as a sequence of causal relationships, as a flow of cause and effect.<sup>54</sup> While knowledge as a state of cognition plays the greatest role in the decision to act quickly, in the delayed

---

52 “AI should not bring uncertainty into the situation. So maybe I would say a military leader who uses (AI) should decide on that himself. So it’s not like someone else is telling him: you have to use this AI.” Ibid., B-95.

53 “The shot must be taken by the individual. Now we transfer this to the leopard. He marks the target. There it is actually no problem. Sure (shoots the leopard) on the target, but there everything has already run before. It is interesting – the factor time.” Ibid., I-98.

54 Cf. Reichenbach, The Direction of Time.

effect of an action, the human deployer decides now whether to deploy tactical AI. His action is now causal for the later effect of the AI, which may only occur with a (long) time delay. This is because the deployer and the AI may be “disconnected” from each other in time – perhaps even in location. In AI reinforcement learning, such a temporal effect is known as Temporal Difference. A backgammon AI must make a move now but does not “know” until much later whether that move now will help win the board game at the end of the game. Thus, the effect of a move made now will not occur until 1,500, 1,800, 2,300, or x time steps later. The situation is very similar when using a tactical AI. Due to the time-delayed effect, it appears as if the AI has made its own combat or kill decision (“autonomously”) and not the military leader, who already decided long before on the deployment. He makes the decision to act now, but the effect of his decision to act is both temporally and spatially distanced from him.

Some people do not like this temporal and spatial distance between decision and consequence:

*(Autonomie, BVR55 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.<sup>56</sup>*

The disconnect between decision and effect is not new:

*(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.<sup>57</sup>*

This leads a colonel to conclude that, after all, Auftragstaktik “is the right leadership philosophy (for AI) because Auftragstaktik only provides the “what,” but not the “how,” and the framework of AI has already been established.”<sup>58</sup>

---

55 BVR: Beyond Visual Range.

56 “(Autonomy, BVR), which is not to the liking of the typical tank soldier. He would like to see everything from his hatch. Although he knows that he is extremely endangered, everyone still wants to look out from above and then decide who goes where.”  
21strategies, E-27

57 “(The use of AI) is, at the end of the day, not much different than what the officer – if you’re talking about a classic war – [does] somewhere in his tent, which is on the battlefield, but just ten kilometers behind the front, and who has told his troops: Please take this position in front. And he is responsible for making sure that the [troops] do it right and don’t attack the village and the civilians next to it.” Ibid., E-27.

58 Personal communication, April 13, 2022.

The problem of time between action and effect, can be elucidated by reference to neuroscientific literature, which deals with reward for decisions. People prefer decisions whose effects occur in temporal proximity to the decision over those for which they have to wait longer for the onset of the effect.<sup>59</sup> This explains people's discomfort with the use of autonomous systems. It is not based on the dissolution of the object character of AI or the questioning that humans act and AI only functions. Rather, it is the psychological reasons that lead to discomfort. Thus, "delay discounting" sets in for decisions with deferred effects,<sup>60</sup> based on motives such as doubts, misgivings, temptations, and so on. In contrast, the deployer of AI needs willpower to compartmentalize the use of tactical AI against the uncertainty or unease that then follow until the AI finally takes effect. At this point, it should be noted that the misgivings as "un-values" must also be included in a value consideration of "Innere Führung." This is because values not only have positive meanings, but negative ones as well. As un-virtues – or vices –, some of them can be studied, for example, on the basis of Aristotle's *Nicomachean Ethics*.<sup>61</sup>

---

59 Klapproth, "Time and decision making in humans."

60 *Ibid.*

61 Birnbacher, *Analytische Einführung in die Ethik*, p. 68.

# 4 How “Innere Führung” Comes to the Machine

**Value-Based Engineering with  
IEEE7000™-2021**

As recently as September 2021, the globally renowned Institute of Electrical and Electronics Engineers (IEEE) published the IEEE 7000™-2021 standard for VBE, which addresses a difficult task: merging the humanities and the sciences. The standard, which requires both a specific process and work products (outcomes) to demonstrate full conformity of technology to applicable law and other societal demands, has a well-thought-out philosophical core full of intrinsic beauty. This is owed less to the standard itself than to its underlying ethical pluralism, based on the theories of great (Western) philosophers, especially of the 18th and 19th centuries. At the same time, potential future versions of the standard do not exclude considering axiologies from other cultural regions of the world – because the IEEE is represented worldwide.

IEEE 7000™-2021 is a global blueprint that is applicable to system designs from all industries in all countries. In particular, IEEE 7000™-2021 is also applicable to military systems and Defense AI. This is not a foregone conclusion, as a vote among members of the standards body on the applicability of the standard to defense systems came up very close. After all, one intention of the standardization body was to “pull the teeth out of startups,” with a focus to tame digital businesses and software systems that neglect law and ethics for the sake of corporate growth and revenue-making, harming democracy (e.g. Twitter), physical health (e.g. Instagram) or truth (e.g. Facebook).<sup>62</sup> It is clear, however, that it is the state which is bound by constitutionally guaranteed rights and the rule of law in a very different way than private companies are. The need to build and deploy legally and ethically compliant technology is disproportionately greater for the state than for private organizations. Accordingly, the (supra-)state interest in IEEE 7000™-2021 is high. Also, in 2022 the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) have adopted the standard,<sup>63</sup> and NATO considers to re-use or amend it for its own standardization efforts.

IEEE 7000™-2021 belongs to a series of standards, of which those following standard 7000 deal with technical details: with the transparency of autonomous systems (7001), the governability of personal data (7002), or algorithmic biases (7003). Standard 7000 requires precise philosophical work. This is a major challenge for technologists and engineers, who often find nothing further from their minds than to tailor their technology to non-technical needs that require interpretation. Incidentally, this is also the challenge encountered by other international technical standardization organizations. Technically adept, standards bodies are often speechless when it comes to law and ethics.

---

<sup>62</sup> Sarah Spiekermann, personal communication, February 17, 2022.

<sup>63</sup> For ISO/IEC/IEEE 24748-7000, see also: <https://standards.iteh.ai/catalog/standards/iso/c6ac9a0c-e70d-4acf-9e44-2b730be96522/iso-iec-ieee-fdis-24748-7000> (last accessed October 17, 2022).

VBE's social innovation therefore includes so-called value leads. They form a new profession and must be trained in value axiology and ethical theories. In system development, they are responsible for conceptual work, as we indicated in chapter 3, and the ethical analysis of requirements. In doing so, they act in equal measure as collectors, structurers, and facilitators who can fit seamlessly into any common approach to software development, analogous to product management.<sup>64</sup> Value leads are responsible for the outcomes that the standard calls for, have both contextual and domain knowledge, knowing the value qualities of a domain, and ethical education, with which they align (objective) ethical requirements for tactical AI with its stakeholders, the literature, applicable law and legislative intent. Value leads possess the virtue of *apatheia*, the restraint of one's own opinion, and should not consider their own ideas, but instead the demands of a system's stakeholders.<sup>65</sup>

VBE with IEEE 7000™-2021 confronts the disorder that has entered the domain of ethics as Hannah Arendt described: "We would not need to worry about all this if metaphysics and this whole value business were not so dilapidated."<sup>66</sup> People can probably deal with the confusion of values that Hannah Arendt already criticized, but machines certainly cannot. The special appeal of IEEE 7000™-2021 therefore lies in the logic and design of the standard and where it deals with values, their elicitation, prioritization, and translation into technical requirements. It is based on Max Scheler's axiology. Whereas Immanuel Kant ethically bound the world by ideals and values accessible to subjective reason, "everything to the narrowness of man,"<sup>67</sup> Max Scheler breaks with the subjective and with his Material Ethics of Values introduces an objectivism open to infinity.<sup>68</sup> Scheler's theory of values objectifies the definition of values by stating: Values exist a priori without subjective cognition and independent of context.<sup>69</sup> This makes the standard applicable across cultures and nations, and regardless of whether the technology in question is a commercial system or one of defense. For defense, VBE is fully applicable with IEEE 7000™-2021 in its entirety.<sup>70</sup>

---

64 These include the waterfall model, the V-Modell XT or agile software development up to the agile overall organization according to the SAFe® process model.

65 Sarah Spiekermann, personal communication, February 17, 2022.

66 Arendt, *Das Urteilen*, p. 172.

67 Przywara, *Humanitas: Der Mensch gestern und morgen*, pp. 30-31.

68 *Ibid.*

69 On the "independence of the being of values from things, goods, states of affairs," see: Scheler, *Der Formalismus in der Ethik und die Materiale Wertethik*, p. 7.

70 NATO's Innovation Unit has made a sensible proposal for a more detailed design of the standard. There, the question is being examined whether existing tools of the standard can detail individual process steps of the standard.



## 4.1 Of Interest: Operational Concept and Context Exploration

Scheler's open view into infinity is urgently needed in tactical AI, because the lists of qualities of "robust" or "trustworthy" or "ethical" AI mentioned at the beginning – such as those drawn up by HLEG AI – lead into a predefined narrowness and are not sufficient to implement value-based AI. Linguistically imprecise and unsystematic, they establish a pre-selection of phenomena that influence an engineer and hinder his creativity, rather than motivating him to explore values that are relevant to a defined AI system. In contrast, VBE assesses only a single System of Interest (SOI) at a time in its own unique context. Each individual AI has its own value qualities that are highly context-specific – just as fruit carries the value of good taste, but the quality of good taste is different depending on whether the fruit is a strawberry or an avocado – and extends beyond the aforementioned lists. This is also why, according to IEEE 7000™-2021, only a single AI system can be certified, not the manufacturing company as a whole.

To illustrate an SOI, imagine that tactical AI is used in the decommissioned air defense tank Gepard to transform the unit that normally has a crew of three into a technically autonomous robot. With the help of tactical AI, the Gepard can become a Smart Gepard<sup>71</sup> – completely uncrewed. What would be the scope of the SOI then? Would it be solely about the AI software or even just one of its components? Then it would probably also include the data on the basis of which the SOI would become active, i.e., the sensor technology. If the smart Gepard depended on its own sensor technology, its sensor technology would also have to be included in the VBE value analysis. Perhaps, however, its outdated sensors would no longer be needed. If the Smart Gepard instead obtained its data from a network, it would have to be analyzed who provided the data that caused it to take certain actions. This already shows how important the selection of a SOI's suppliers is. Are the suppliers honest and transparent about how they govern data?

The analysis of the SOI also includes the context in which a Smart Gepard would be deployed. Was it just a demonstrator to show how tactical AI allows a lifecycle extension of legacy equipment? Or would the Smart Gepard enter combat operations in Ukraine? It depends on the concept of operations and the operational environment whose and which values are affected. In the first case, perhaps only government auditing, the Bundestag or procurement agencies, and perhaps AI research itself, are interested. The second operational scenario involves life and death on the battlefield, the interests of military personnel, civilians, and opposing

---

71 Vieregge, „Fähigkeitslücke der Bundeswehr schließen.“

governments. The operational context, then, determines who the Smart Geparde's stakeholders are and which of their values are affected by the SOI.

## 4.2 The Value Exploration

### Core Values, Value Qualities and Value Dispositions of an SOI

If values are defined quite generally as (good or bad) "properties" of a subject, object, action or objective, it should be briefly noted at this point that several dimensions of values exist. Values can be moral or non-moral in nature, are intrinsic or extrinsic and, depending on the theory of values, within a hierarchy.<sup>72</sup>

To find out how technology can realize the potential of the good that a specific context holds, the stakeholders of an SOI are involved right in the early stages of VBE. Their considerations are captured using philosophical frameworks of ethics. It is this philosophical part of VBE with IEEE 7000™-2021 that requires extreme linguistic precision.

A core value for the Bundeswehr is human dignity. In Western democracies with their image of the free man, human dignity is an intrinsic value. An intrinsic value is a valuable phenomenon that is desirable for its own sake. The intrinsic value of human dignity is so important for a democracy that it must be safeguarded not only by law, but also by technology, not only by the state, but also by private third parties.

Human dignity is expressed through many value qualities, all of which feed into human dignity. The value qualities condition the core value; they go into the core value.<sup>73</sup> Thus, a core value is composed of many value qualities. Duty of assistance, mercy, ability to discriminate, self-protection of tank crews in combat, or decision-making ability are just a few of the value qualities that members of the Bundeswehr cite and that make up the quality of human dignity.

Finally, the value dispositions (the "Value Demonstrators", as they are called by the standard) of an SOI are the realization of the value qualities in the SOI. They can be both technical and organizational in nature. If, for example, the Smart Geparde "removes the human"<sup>74</sup> and is transformed into a self-controlling robot by

---

72 In detail: Birnbacher, *Analytische Einführung in die Ethik*, p. 242 ff.

73 Scheler, *Der Formalismus in der Ethik und die Materiale Wertethik*, p. 15.

74 21strategies, F-1.

tactical AI in order to increase self-protection, technical and organizational precautions coincide technically, because the tactical AI is a quasi-automation of the tank crew, and organizationally, because the crew no longer has to operate the tank itself, but acts safely at some distance from the battlefield.

## Ethical Value Elicitation: The SOI in the Light of Three Ethical Theories

Once the SOI has been sufficiently defined, its stakeholders explore its universe of values and make use of the ethical theories of utilitarianism, duty ethics and virtue ethics. In light of this ethical pluralism, the stakeholders answer the question of which measures can strengthen positive value qualities and avoid negative ones. This approach not only completes the value analysis, but also shapes the operational concept of an SOI further.

Utilitarianism belongs to consequentialist ethics, which is essentially opposed to deontological ethics. To put it simply: In consequentialist ethics it is the good consequences of an action that count. In deontological ethics it is just the other way round: What counts is the morally good action without regard to its consequences, as demanded, for example, by Kant's ethics of duty. There are mixed forms; moreover, differences of both theories also exist in the fact that deontological ethics wants to see moral values realized, whereas utilitarianism knows only one non-moral value: utility. Utilitarianism is therefore also considered a monistic system. In addition, utilitarianism does away with value conflicts, whereas deontological ethics engages in weighing values against each other. The duo is complemented by virtue ethics, which does not introduce another theory, but instead examines motivation, character traits of an individual that lead to moral action.

Beginning with utilitarian ethics, founded in the 18th century by Jeremy Bentham and John Stuart Mill, the stakeholders, under the guidance of the value lead, reflect on the Smart Gepard in light of its utility: "*Bei mir geht es letztendlich nur um Wirkung.*"<sup>75</sup> For the utilitarian, only this single benefit counts as an intrinsic value. Objective values such as freedom, justice, or the common good are considered extrinsic values by utilitarianism, "their value depends on the value of the subjective states they bring about."<sup>76</sup> This results in a kind of gradation of values that gives higher weight to what is more conducive to utility.

Immanuel Kant's Categorical Imperative is a textbook example for the examination of the SOI in the sense of obligatory ethics: "Act only according to that maxim

---

<sup>75</sup> "For me, ultimately, it's only about the effect." 21strategies, I-19.

<sup>76</sup> Birnbacher, *Analytische Einführung in die Ethik*, p. 218.

by which you simultaneously want to become a general law. „<sup>77</sup> The Categorical Imperative is an unconditional ought-sentence: „*Es widerspricht eigentlich dem Menschlichen, der menschlichen Natur, wenn man Sadist ist oder Mörder.*“<sup>78</sup> Thou shalt not kill; thou shalt not torture, these are generally valid maxims from which moral action follows.

When the Smart Gepard is deployed, two human lives are at stake: one's own and that of the enemy. If the Smart Gepard as a whole were to be considered as a SOI – and not just its tactical AI subsystem – the self-protection of one's own troops would be maximally promoted because the Smart Gepard would manage entirely without a human crew. On the other hand, special rules apply to the lives of enemy soldiers. A legal exception is made to the general principle: “Thou shalt not kill” when deployed on the battlefield, because killing enemy forces in uniform is permitted in principle, although practice suggests:

*So richtig schießen wollte keiner (im Einsatz), weil jeder Angst hatte, sofort hinterher vom Staatsanwalt (strafrechtlich verfolgt zu werden).<sup>79</sup>*

Last, the question of how the SOI subverts the character of its stakeholders is considered in light of virtue ethics. How would dealing with the Smart Gepard negatively impact the character of the affected stakeholders in the long run if the proper precautions were not taken? Constant interaction with technology changes us humans. What virtues does the SOI subvert? What vices does it promote? Will a Smart Gepard make the deployer more careless? Will killing the enemy be trivialized if a device can be deployed at a greater distance without physically participating in the battle itself? Without evaluating or ranking the soldierly virtues, stakeholders under the guidance of the Value Lead ask themselves here whether a Smart Gepard can help ensure that operators are not harmed – physically or psychologically – when using the device:

*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.<sup>80</sup>*

---

77 Ibid., p. 141.

78 “It actually contradicts human nature, if one is sadist or murderer” 21strategies, G-11.

79 “No one really wanted to shoot (in action) because everyone was afraid of being (prosecuted) by the law immediately afterwards”. Ibid, G-34.

80 “I find the point very interesting that applications often have the potential to help people not to make mistakes, or to make fewer mistakes (that would burden people for a lifetime), because they have a conscience. So (sort of) helping them keep a clean conscience. I think that's a very important aspect.” Ibid., I-105.

## Value Clustering and Prioritization

To complete the value adjustment, the Value Lead consolidates and clusters core values together with their value qualities.

A stakeholder may have voiced that the tactical AI of a Smart Geparard must not put its calculations into action without authorization. He wants to see the Man in the Loop. In the language of the standard, the core value of human dignity then unfolds in the value quality of the target fire authorization: "Human dignity demonstrates itself in human authorization of fire." Authorization expresses a value quality of the SOI that is intended to protect not only the human dignity of third parties, but also one's own. At this point, it is not yet clear how authorization is accomplished. Rather, there are several potential possibilities for authorization, especially considering the temporal aspect, as indicated above.

The results of the value analysis are value cluster representations, supplemented by curated tables in which the statements of stakeholders are assigned to intrinsic core values in each case. The tables are essential for tracing and mapping of stakeholder statements to derived system functions. They are approved by the SOI stakeholders and then prioritized.

In the prioritization that follows any prioritization of value clusters must also consider the corporate story of the SOI manufacturer. In setting priorities, it is essential to know what mission and vision the manufacturer of the SOI is pursuing. Its justified economic interests should and must be included.

Priorities are also influenced by the current and future legal situation facing the SOI. The expertise of the Value Lead therefore also includes knowledge of the applicable law, current legislative projects, case law, philosophy and (scientific) literature. Even legislative projects such as the EU AI Act have an impact on a tactical AI, which, if present, must also appreciate the soft law of a manufacturer, as its violation can result in severe financial penalties.

## 4.3 From Theory to Practice: Ethical Value Requirements

From value qualities like "target firing authorization" one cannot go directly to implementation because value qualities do not yet describe value dispositions. This role is taken over by the Ethical Value Requirements, in short: EVR, which can be not only technical, but also organizational or social in nature.

To authorize target fire in the Smart Gepard, a tactical AI would need to provide information to its deployer. For example, a tactical AI could (i) communicate the probability that there is a legitimate adversary to engage (“98.9 percent probability”) and (ii) display a countdown (“*n* seconds”) of how much longer target fire would be useful and/or effective and (iii) provide an operating interface where the human hands over engagement of a target to the tactical AI with the push of a button.

Defining EVR is the most important step of VBE with IEEE 7000™-2021, making sure that EVRs are formulated in such a way that, once implemented as system functionality, they are testable and assessable. The result: the ethics of an AI become measurable. “A countdown indicates the time in which target firing would be useful.” Operators want to be able to rely on the displayed value – they want to be sure that the calculation procedure is implemented correctly, that uncertainties of the raw data are considered in the calculation or that the calculation has been tested statistically.

The EVRs are followed by an important final step. In interaction with the stakeholders, the value lead performs a risk assessment. He seeks to answer the question of how at risk the specified EVRs are. What threatens an EVR, and what controls (system requirements) must be taken to ensure an EVR? System developers may be required to ensure that the countdown is calculated correctly not only on known data sets but also on out-of-sample data, or that the test data set is published along with the system.

Standard risks, which include insufficient test coverage, for example, are applied by the value lead to a risk matrix and assigned a probability. This is because not all identified risks have the same probability of occurring. The rule of thumb is the greater the risk of a threat, the higher the weighting of a system requirement, which must also be prioritized by software development when implementing the SOI. The control mechanism must be specified in such a way that it can ensure EVRs.

However, it is not considered a standard risk when intrinsic values and their value qualities are threatened: “With 98.9 percent probability, the target is a legitimate adversary.” If there is the threat of a hazard, any ranking is out of the question. To determine risk under these circumstances, the value lead undertakes a technology assessment, while the development team takes precautions at the earliest possible stage to manage and mitigate risks with very high hazard potential.

# 5 Conclusion

*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.*<sup>81</sup>

That AI considers the universe of values of “Innere Führung” is an urgent and legally required wish of the Bundeswehr. It can be fulfilled with value-based technology according to IEEE 7000™-2021, a global standard that is already ready for defense systems (it could also be extended by existing standards such as the NATO Risk Assessment Tool). In the meantime, the standard is being rolled out, and training centers for value leads with the skills and training of both a philosophical and technical understanding are proliferating. This begins the realization of a standard whose intrinsic beauty lies in the logic and structure of its value work: that better technology serves the good of people.

---

<sup>81</sup> “Strictly speaking, what you are creating is, as it were, ‘Innere Führung’ for AI. You could transfer this value construct to ‘Innere Führung’. It’s not mutually exclusive.” Ibid., A-173.



# Literature

"DARPA SCEPTER program will explore machine-generated strategies for warfare," Defense Techconnect, January 24, 2022, <https://defensetechconnect.com/2022/01/24/darpa-scepter-program-will-explore-machine-generated-strategies-for-warfare/> (last accessed October 17, 2022).

"Was bedeutet Innere Führung," *Bundeswehr.de*, <https://www.bundeswehr.de/de/ueber-die-bundeswehr/selbstverstaendnis-bundeswehr/innere-fuehrung>, (last accessed 11 February 2023)

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

Birnbacher, Dieter, *Analytische Einführung in die Ethik*. 3. durchgesehene Auflage (Analytical Introduction into Ethics. 3rd Revised Edition) (Berlin: De Gruyter, 2013).

Funk, Michael, *Roboter- und KI-Ethik: Eine methodische Einführung* (Robotics and AI ethics: a methodological introduction). Volume 1. (Wiesbaden: Springer Vieweg, 2022).

*Kants gesammelte Schriften. Herausgegeben von der Königlich preußischen Akademie der Wissenschaften* (Kant's Collected Writings. Published by the Royal Prussian Academy of Sciences) (Berlin: G. Reimer, 1900 ff.).

Klapproth, Florian "Time and decision making in humans," *Cognitive, Affective & Behavioral Neuroscience*, 8:4 (December 2008), pp. 509-524 <https://link.springer.com/article/10.3758/CABN.8.4.509> (last accessed 11 February 2023).

Prechtl, Peter and Burkard, Franz-Peter (eds.), *Metzler Lexikon Philosophie. Begriffe und Definitionen*. 3., erweiterte und aktualisierte Auflage (Metzler Lexikon Philosophie. Terms and Definitions. 3rd, expanded and updated edition) (Heidelberg and Berlin: J.B. Metzler'sche Verlagsbuchhandlung, 2008).

Przywara, Erich, *Humanitas: Der Mensch gestern und morgen* (*Humanitas: Men Past and Future*) (Nuremberg: 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* (Formalism in Ethics and Material Value Ethics. Renewed Attempt at Establishing Ethical Personalism) (Halle an der Saale: Verlag von Max Niemeyer, 1916).

Spaeth, Andreas, "Grundlegender Softwarefehler in der Boeing 737 Max gefunden (Fundamental Software Flaw Found in Boeing 737 Max)," *Neue Zürcher Zeitung*, August 5, 2019, <https://www.nzz.ch/mobilitaet/luftfahrt/boeing-737-max-grundlegender-softwarefehler-gefunden-ld.1499955> (last accessed 11 February 2023).

Spiekermann, Sarah "Value-based Engineering: Prinzipien und Motivation für bessere IT-Systeme (Value-based engineering: principles and motivation for better IT systems)," *Informatik Spektrum*, 44:4 (August 2021), pp. 247-256 (2021, <https://link.springer.com/article/10.1007/s00287-021-01378-4> (last accessed October 17, 2022).

Spiekermann, Sarah, "Die Ethik in der Künstlichen Intelligenz (Ethics in Artificial Intelligence)," in Klaus Mainzer (ed.): *Philosophisches Handbuch der Künstlichen Intelligenz* (*Philosophical Handbook of Artificial Intelligence*) (Munich: Springer Reference Geisteswissenschaften, in press).

Vieregge, Kerstin, "Fähigkeitslücke der Bundeswehr schließen (Closing the Bundeswehr Capability Gap)," press release, March 15, 2022, <https://www.kerstin-vieregge.de/artikel/faehigkeitsluecke-der-bundeswehr-schliessen> (last accessed 11 February 2023).

Vöneky, Silja, "Ethische Experten und moralischer Autoritarismus" (Ethical Experts and Moral Authoritarianism)," in Silja Vöneky et al. (eds.) *Legitimation ethischer Entscheidungen im Recht: Interdisziplinäre Untersuchungen* (Legitimation of Ethical Decisions in Law: Interdisciplinary Studies) (Berlin/Heidelberg: Springer, 2009), pp. 85-97.

Wahl, Thomas, "High-Level Expert Group Publishes Ethics Checklist (2020)," <https://eucrim.eu/news/ai-high-level-expert-group-publishes-ethics-checklist/> (last accessed 11 February 2023).



## Defense AI Observatory Studies

- 23|10** Yvonne Hofstetter and Joseph Verbovzsky, How AI Learns the Bundeswehr's "Innere Führung." Value-Based Engineering with IEEE7000™-2021
- 23|09** Robert C Engen, When the Teeth Eat the Tail: A Review of Canada's Defence Artificial Intelligence
- 23|08** Çağlar Kurç, Enabling Technology of Future Warfare. Defense AI in Turkey
- 23|07** Lauren A. Kahn, Risky Incrementalism. Defense AI in the United States
- 22|06** Yvonne Hofstetter, Wie KI Innere Führung lernt. Wertbasierte 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 Verbovzsky, Beware the Hype. What Military Conflicts in Ukraine, Syria, Libya, and Nagorno-Karabakh (Don't) Tell Us About the Future of War

