Value-based Engineering (VBE) with IEEE 7000™-2021

Yvonne Hofstetter, Professor h.c. for Digitization and Society
Chief Executive Officer
21strategies GmbH
Lilienthalstrasse 27, 85399 Hallbergmoos GERMANY
yvonne.hofstetter@21strategies.com

ABSTRACT

This article provides a high-level tour of Value-based Engineering (VBE) with IEEE 7000™-2021, a standard which was put into effect in September 2021 bringing together natural sciences and the humanities, a task long overdue, to establish balance between technology and non-technical needs of post-modern digitized democracies. VBE fosters system engineers of being open-minded for providing digitized societies with systems not only providing state-of-the-art system quality attributes, but also observing legal and ethical codes. To give an overview how VBE can be applied to Defense AI, this article relies on an empirical research of AI in the context of military use and voices of the German Bundeswehr specifying their understanding for the legal and ethical compliance of Defense AI.

1. INTRODUCTION

It is a truism that artificial intelligence (AI) has spread worldwide. The rapid and unstoppable proliferation of AI over the past decade speaks for itself, and regulators have long since followed suit to contain AI and its societal consequences. For example, AI requirements relevant for Germany have been listed by the European Commission's High-Level Expert Group on Artificial Intelligence (HLEG AI) and, at national level, by the German government's Data Ethics Commission (DEK) and the German Parliaments Commission of Inquiry on Artificial Intelligence (EKKI). A priori, requirement lists issued by governmental commissions, or NGO, or commercial firms or whatever working groups are not legally binding. Where they are not being transformed into sovereign law – for example, into the EU AI Act, which is expected to be passed by the European Parliament and enter into force in the year 2025 – they can at best become second-class law based on contractual agreement between social partners in AI manufacturing firms on the (self-) obligation to a corporate social responsibility (CSR). If such "soft law" is violated, at least financial consequences – the "hard facts" – are threatening. For this reason, the concern to translate irreducible principles of one's feeling perception which provides orientation for one's actions – the *values* – into the technical functions of an AI is coming into sharper focus.

Third Wave AI is becoming more human-like

Despite regulators' acts of catching up the technological progress, AI is developing more and more capabilities. Machines which are human-like is an old dream. But the AI of a Siri, an Alexa, a recommendation engine for products or streaming data, and even "autonomous driving" as practiced in the early 20s of the 21st century is classified as Second Wave AI – a term coined by DARPA – which discovers correlations in mass data and is nothing else but input-output mapping. In fact, Second Wave AI is still far from producing machines that are the spitting image of a human being. AI does computing; on the contrary, human beings have cognition, which is the ability to process information, however: not only by a mere data or information-processing procedure, but by an evaluating and judging process in search of the "view of the whole". Whoever wants to go so far as to apply the concept of cognition to the functions of an AI must then specify its semantic meaning in the context of computing machines – and AI is no more than that – which have an insight into the world that remains reduced to the "mathematical structure of being": "Can a mathematician who looks at the world in a mathematical view find anything but mathematics in the universe?” “[The highly superfluous miracle of the

beautiful] will never be discovered by physics, and cannot, because in its questioning physics abstracts by its very nature from aesthetic feeling and moral attitude (...).”

Third Wave AI, less based on mass data, but rather on contextual models, aims at approaching this view of the whole. To do so, Third Wave AI cannot be reduced to artificial neural networks or machine learning, but must make use of the entire toolbox of mathematical theories and information techniques. Nevertheless, it too will not transcend the merely scientific reason. Its models are a mathematical representation of a complex reality, reducing it to the observable and measurable, to data and information. Nevertheless, Third Wave AI promises enormous epistemological advances of humans, for example in medicine, climate research and biology. Its ability to simulate and understand larger contexts over time, to plan and act with foresight, to make optimal sequential decisions under uncertainty, and to self-correct its decisions as circumstances change also makes it the core of complete technical autonomy. Efforts to integrate it more closely with robotics are also already in full swing in research landscape all over the globe.

While Defense AI’s potential is huge, so is the fear

With its principles of the rule of law and ethics and morality, democracies and their military forces are encountering the full potential of Third Wave AI for the first time. Reconnaissance, data fusion, situation analysis have been the fields of application of Second Wave AI for the military since the 1990s. But for modern complex engagement, these capabilities alone are no longer sufficient. Today, military forces demand strategic and tactical capabilities from machines that take place at machine speed. Third Wave AI is the basis to such machines which potentially “take man out of the loop”. Battlefield simulators and training systems, (partially) autonomous drones, both airborne and naval, smart ammunition, cyber-physical systems or cross-domain AIcontrolled defense networks – the list of conceivable fields of application for Third Wave AI is long. At the strategic level, it aims to reduce the planning time for tasks such as an air tasking order to a third or a quarter of the time required nowadays, and at the operational level, Third Wave AI can increase the combat value of legacy systems, transform them into technically (partially) autonomous robots, establish a higher selfprotection, optimize ammunition consumption and logistics, or even amend C2 systems, supplementing them with a more emergent behaviour of assets on the battlefield. In the future, the commander will not make use of AI. The AI *is* the commander.

Meanwhile, Third Wave AI for defense (hereinafter: “Defense AI”) is igniting a debate about “great horror”, since Defense AI acts in an extreme area: It serves a state’s use of force; deployed for the defense of democracies, it must be mandatorily embedded into the legal and value national frameworks, relevant supranational rules, e.g., of the European Union, and international law, as well. In practice, however, it proves to be a great challenge to cast legal, ethical or social norms into the functions of a Defense AI.

2. CRITIQUE OF DEFENSE AI FOR THE GERMAN ARMY

Values? Which values, please?

According to a survey undertaken among leaders of the German Bundeswehr in 2022 which principles of their core leadership doctrine called “Innere Führung” shall be observed not only by soldiers, but by Defense AI, too, the positive meanings that the leaders attribute to this doctrine read like a colourful bouquet of good motivations and qualities of people and things. In fact, the leaders express umpteen different phenomena related to “Innere Führung”, which trigger certain emotional states in themselves as “feeling [] and desiring [] subjects (...)”. What moves their souls ranges from A for adventurousness to Z for zealousness.

One reason for the high number of phenomena required from Defense AI is the lack of moral objectivity.

How value views are subjective and are formed against the background of certain ideologies, world views or paradigms – the “cultural background radiation” –, was pointed out by a German general as follows: “But what do *we* want? There will be others who say: What does that mean, ethical? Ethical means we win and the others lose.”

In addition to the subjective world of experience, there is another deficit when we speak about values: a “philosophically founded conceptual world” is missing when people talk about values. The structural properties of values do not find a proper mapping in language; our language of values is simply not precise enough. Thereby, the “critique of language (...) is a constitutive part of methodological philosophy and ethics”. The inadequacy mentioned is in particular the reason for the difficulties engineers face when being asked to implement Defense AI that shall carry the desired positive phenomena or shall be capable of supporting them.

“Technical robustness and security” or “transparency”, three meanings as defined by the HLEG AI Ethics Checklist for AI, are nothing but general quality attributes of many technical systems, as they were described by default in software specifications before agile software development began to take hold with its very different documentation requirements. Quality attributes of technical systems are “hygienic requirements” of software systems (S. Spiekermann, personal communication, November 12, 2021), or to put it more sharply: Actually, they are self-evident. But for over two decades now, greatly shortened software development cycles, cost pressures, focus on shareholder value, commoditization of programming power, and the faster time to market of software have worked in favour of a software vendor’s revenue and profit; but at the same time, the quality of even mission-critical systems has suffered greatly. Fatal proof of this has been provided by the crashes of two Boeing 737 MAX-8 aircrafts attributable to poor software design, quality, and lack of physics expertise of the programmers involved.

Quality attributes of software systems that are to be expected as a matter of course are far from being sufficient to realize or foster core values such as human dignity and freedom, peace, and justice or virtuous character traits such as love of one’s homeland, truthfulness or courage. Value-based AI must therefore be preceded by unambiguous conceptualization, a task that the IEEE 7000™-2021 standard for VBE imposes on Value Leads, a new profession and social innovation triggered by the standard. The Value Leads’ role and scope of tasks in the development process will be addressed in the below figure 3.

A life in freedom for Defense AI?

Intuitively, leaders of the German Bundeswehr mentally transfer the idea of being human also to Defense AI. It is considered that Defense AI could evaluate moral behaviour and judge ethically itself or abide by human laws:

“I give a ‘smart tank’ the rules to which we still feel morally and ethically bound.”

“If we said that Defense AI implemented the leadership doctrine of the German Bundeswehr..., how would you do that? By training it to trigger the actions that it is supposed to trigger based on the standards that you might apply as a human.”

“But that's what's so interesting about Defense AI. It is maybe even more human than humans themselves, and certainly more predictable than a soldier in certain situations.”

The tendency to view Defense AI not as a thing, as an object, but as “subjects of ethical judgment”, is momentous. If one views Defense AI as a thing, then “AI ethics” can be defined as humans “morally scrutinizing the purposeful use of Defense AI. 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 must follow (human) rules and regulations. The latter two cases involve AI as a subject that can itself act morally and is subject to written or unwritten norms.

What is remarkable about the last two quotes drawn upon is the language used in reference to Defense AI. “AI is perhaps even more human” is not the same statement as “AI perhaps acts even more human”. Persons *act* in freedom, with the “ability and [the] capacity to act consciously and voluntarily”. In this sense, Defense AI never acts either, simply because it lacks the 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.” As a machine, it lacks consciousness, voluntariness, its own motivation. Ignoring the lack of theoretically exact formulations in the aforementioned quotes, the three leaders of the German Bundeswehr used as examples are thus instinctively hesitant to grant to Defense AI the same freedom as to 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. “Actions that it is supposed to trigger”, leaves open whether AI either acts or behaves without mental processes, or simply just *functions*, because the utterance only focuses on the effect of AI. “Acts, which the machine executes in the end”, assigns the actual action – namely the use of AI – to the person of the soldier; it is the soldier’s action that is causal for the effects of Defense AI.

The object property of Defense AI would also face dissolution if Defense AI was to adhere to an ethos, a “book of norms” or a code. E.g., both sovereign law and Rules of Engagement are such codes. With regard to the discussion of values, they standardize the “ethical minimum” in the best case. Law, however, is created by people for people – and not by people for machines. Law frees people, because people are given the freedom to violate the law. Rules for machines or enforced by machines – the smart car that won’t start because its occupants haven’t fasted seat belt yet, and even the cookies on your website – are morally paternalistic. Thus, if an AI was able or required to observe human codes itself, it would become the “subject of moral action” in this case as well. Would the subjection of AI to human laws then have bear the consequence that AI could or would also have to carry responsibility?

Who bears responsibility?

With the subjectification and individualization of AI, the man-machine-blur of the digital age would continue to advance. However, in democracies, the concept of the “person” still has legal standing. Only a legal subject, not an object like AI, which is software, can make free ethical judgments and act in freedom. For freely made value judgments and free actions, the human being – and only the human being – then also bears responsibility: “Guided missiles, after I have shot them down, I can no longer see them. I can no longer control them. I can no longer intervene, and I am still responsible, if they don’t hit a fighter plane, but a civilian jet. And then it’s my turn again, and rightfully so.”

Responsibility for their actions is a central concern for members of the German Bundeswehr, especially when certain actions have led to the death of comrades or even of the enemy. It is important for the leaders to explain the reasons and intentions of their actions: “Where people work, mistakes happen. If that turns out to be bad in military operations, one must also stand by one’s responsibility, as long as one can say: I did this for these and those reasons. I came to the conclusion: I am responsible for this.”

Responsibility is something that describes an intersubjective process. It is about talking, about processing what happened from person to person. Leaders of the German Bundeswehr explicitly highlight the differences between humans and machines, especially humans’ capacity for empathy, compassion, strength, intuition, mercy. Responsibility is also the reason the leaders prefer human decisions to (precise and accurate) machine computations, even when a human decision has proven poor or missed its mark altogether:

“After all, the AI decision might even be better in the sense of: It may have led to fewer blue casualties. But the relatives of a dead comrade then have no one to talk to about it. And in the end, isn’t that more terrible than other consequences, but which can be dealt with after the fact?"

"But you don't forgive an autonomous system for mistakes. You always or more likely forgive a human for mistakes."

Leaders of the German Bundeswehr favour the use of Defense AI; across the board, there seems to be consent that nobody else than the the military leader is responsible for the use of Defense AI: "The only realistic, pragmatic thing that we still need in the military, too, is that the one who ends up using this product is the one who bears the responsibility. You're not going to be able to put the responsibility for deploying AI back on the developer who programmed it five years before."

But the willingness to accept responsibility for the use of Defense AI is conditional: "What can operators do to live up to that responsibility for the action that the machine ended up performing?" Soldiers want to know "their" AI inside out and what they're getting into before putting AI into real-world use: "Then I need really good training, and very realistic training." In addition to intensive training of humans with an AI, there is also a requirement that the functioning of the AI be ensured through a quality assurance process and standards: "That's where you're probably going to have to create standards. How often and with what probability of error does the AI have to come to the following conclusion in a certain scenario?"

As a second requirement which leaders of the German Bundeswehr suggest in keeping with their leadership doctrine is that they be given freedom of choice. They want to decide for themselves whether they want to use Defense AI or not: "The AI must not bring uncertainty into the situation. So, I would say maybe a military leader who uses AI should decide himself whether or not to use it. So, it's not like someone else is telling him, you have to use this AI."

So, while leaders of the German Bundeswehr are in favour of using Defense AI (even with technically autonomous capabilities) if the two prerequisites of quality assurance and exhaustive training are met, they are then willing to take full responsibility for the functioning of Defense AI.

3. VALUE-BASED ENGINEERING WITH IEEE 7000™-2021

There are technical innovations and non-technical social innovations. The latter are designed to promote good in a society. VBE with IEEE 7000™-2021 falls into the category of an institutional social innovation: a global association such as the IEEE creates a new institution, the IEEE 7000™-2021 standard, as a "collective response of all members of a community to a particular situation". Such a situation had given rise to numerous software systems in the first two decades of the 21st century, whose deleterious effects on democratic societies or even individuals have now become clear. Building better technology was therefore the motivation for the IEEE 7000™-2021 standard, for which it is the first standard worldwide to accomplish the difficult fusion of technology and philosophy, of natural sciences and humanities, and to enable the consideration not only of technical, but also of ethical, legal and social requirements of all stakeholders of a new technical system.

VBE's social innovation also includes the presentation of *Value Leads*. They form a new profession and must be trained in ethical theories and axiology. In system development, the conceptual work, as we indicated in the previous section, and the ethical requirements analysis falls to them. In doing so, Value Leads act in equal measure as collectors, structurers, and facilitators who can fit seamlessly into any common approach to system development, analogous to product managers. Value Leads themselves possess the virtue of *apatheia*, the withholding of one's own opinion, and must not only consider their own ideas, but especially the demands of a technical system's stakeholders (S. Spiekermann, personal communication, February 17, 2022).

VBE with IEEE 7000™-2021 confronts the mess that has entered a relevant area of ethics: "We wouldn't have to worry about all this if metaphysics and this whole values business hadn't gotten so down." Humans can probably deal with the confusion of values that Hannah Arendt complains about back in the 60s, but machines certainly cannot. The great inner beauty of IEEE 7000™-2021 therefore lies in the logic and the design of the standard, where it deals with values, with their collection, organization, and translation into technical requirements. Underlying to it is Max Scheler's axiology. While Immanuel Kant has ethically bound the world by ideals and values accessible to subjective reason and "everything to the narrowness of man", Max Scheler breaks with the subjective and introduces with his Material Value Ethics an objectivism literally open to infinity. Scheler's theory of values objectifies the definition of values by stating that values exist a priori without subjective cognition and independent of context. This makes the IEEE 7000™-2021 applicable across cultures and nations, and regardless of whether the system in question is a commercial system or one of defense. For defense, IEEE 7000™-2021 is fully applicable without trade-offs, and also without specialization, while there might be room to more detail, e.g., through the use of NATO's risk framework or lawfulness checks.

System of Interest: Operational Concept And Context Exploration (clause 7)

But back to Defense AI. Scheler's open look into "infinity" is of urgent need here, because the lists of AI qualities of "robust" or "trustworthy" or "ethical" AI mentioned at the beginning lead into a predefined bias and are insufficient to implement value-based Defense AI. Linguistically imprecise and unsystematic, they pre-select phenomena that influence an engineer and hamper his creativity instead of motivating him to explore values himself that are actually relevant for a defined AI system. In contrast, VBE assesses only a single System of Interest (SOI) at a time in its own and unique context. Each individual AI has its own value qualities that are very 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 will definitely go beyond the lists mentioned above. This is also why, according to IEEE 7000™-2021, only a single AI system can be certified, not the manufacturing firm as a whole.

To be certified, a certain Defense AI SOI must meet the normative part of the standard (clauses 7 through 10). Unlike many other standards, for full compliance with the standard it is not mere the adherence to a process that is expected; instead, the results of a process (outcomes) shall be documented and traceable as well. To do this, the Value Leads responsible for the outcomes required by the standard have both contextual and domain knowledge, knowing the value qualities of the defense domain, as well as an ethical education, with which they align (objective) ethical requirements for a Defense AI with both its stakeholders and the literature and applicable law and legislative intent.

Figure: FlakPz Gepard. © Hans-Hermann Bühling from https://commons.wikimedia.org/wiki/File:Gepard_1a2_overview.jpg under CC BY-SA 3.0

To illustrate an SOI, imagine that a Defense AI is deployed in the decommissioned German FlaK tank Gepard – which was delivered to Ukraine despite being taken out of service in Germany – to transform the tank with a crew of formerly three into a technically autonomous robot. With the help of Defense AI, we imagine FlaK tank Gepard to become *Smart Gepard* – without any crew. 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., any sensor technology. If 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 long-outdated sensors would no longer be required. If the Smart Gepard instead obtained its data from a network, it would have to be analysed who provided the data that caused it to take an action. This already shows how important the selection of a SOI's suppliers is. Are the suppliers to a SOI honest and transparent about how they handle data, for example?

The analysis of the SOI also includes the context in which a Smart Gepard would be used. Was it just a demonstrator to show how Defense AI allows for investment protection and lifecycle extension of legacy equipment? Or would Smart Gepard enter combat operations in Ukraine? It depends on the concept of operations whose and what assets are affected. In the first case, perhaps only government auditing, the military procurement agencies, and perhaps AI research itself, are interested. The second operational scenario, on the other hand, involves life and death on the battlefield, the interests of military staff, civilians, and adversarial governments. The operational context, then, determines who the Smart Gepard's stakeholders are and which of their values are affected by the SOI.

Value exploration (clause 8)

Core values, value qualities and value dispositions of an SOI

If values are defined quite generally as "qualities" (good or bad), it should be briefly noted at this point that several dimensions of values exist. Values can be moral or non-moral in nature, there are intrinsic and extrinsic values, and depending on the value theory, there may or may not be a hierarchy of values.

To find out how technology can realize the potential of good qualities that a SOI's concept of operations holds, the stakeholders of a SOI are involved right in the early stages of VBE. Their considerations are captured and measured against three value theories. It is this philosophical part of VBE with IEEE 7000™-2021 that requires particular linguistic precision.

Workmanship: The SOI in the light of three ethical theories

After eliciting relevant values from stakeholders, the SOI runs through the three ethical theories of utilitarianism, duty ethics, and virtue ethics. In light of this ethical pluralism, the Value Lead answers the question of which measures can strengthen positive value qualities while avoiding negative ones. The approach not only completes the value analysis, but also causes the concept of operations of an SOI to take further shape.

Utilitarianism belongs to consequentialist ethics, which is basically opposed to deontological ethics. In very simplified terms, consequentialist ethics is concerned with the good consequences of an action. 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 duty ethics. 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, while deontological ethics knows weighing. The duo is complemented by virtue ethics, which does not introduce another theory of ethics, but looks at the motivation, character traits, charisms 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 Value Lead reflects on the Smart Gepard in light of its utility: "I'm all about effect, ultimately." For the utilitarian, only this single utility 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". This results in a kind of gradation of values that gives higher weight to what is more conducive to utility. In contrast to deontological ethics, utilitarianism does not weigh values against each other.

For the test of SOI in terms of duty ethics, Immanuel Kant's Categorical Imperative is textbook example: "Act only according to that maxim by which you want at the same time that it become a general law." Ironically, "I want x to be ensured," the Categorical Imperative is unconditional ought-sentence: "It is actually contrary to what is human, to human nature, to be a sadist or a murderer." Thou shalt not kill; thou shalt not torture, these are universally valid maxims from which moral action follows.

When using the Smart Gepard, two human lives are at stake: one's own and that of the enemy – in military terms, both blue and red lives. If the Smart Gepard as a whole was to be considered as an SOI – and not just its Defense 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 wearing a uniform is generally permitted, although practice suggests: "No one really wanted to shoot [in action] because everyone was afraid of being [prosecuted] by the prosecutor immediately afterwards."

Last, the question of how a SOI reinforces positive value qualities is posed in light of virtue ethics. How would dealing with the Smart Gepard affect the character of the stakeholders involved in the long run? Constant interaction with technology changes humans. What virtues does the SOI undermine? 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? Or does the deployment have a psychological effect like that of American drone pilots, who are known to suffer trauma because they are under constant observation while deployed and the meaning of a mission – standing by other people, wanting to help – is no longer immediately apparent from a distance?

Without evaluating or ranking the soldierly virtues, a Value Lead asks itself here: How can a Smart Gepard help make stakeholders better people? “I think the point is very interesting that applications often have the potential to help people in a way that they don't make mistakes or they make fewer mistakes [that would burden people for life] because those have a conscience. So [sort of] helping them keep a clean conscience. I think that's a very important aspect.”

Value Clean-up

To complete the value clean-up, the Value Lead consolidates and clusters core values along with their value qualities.

A stakeholder may have expressed that a Smart Gepard’s Defense AI must not execute its computations without authorization. He wishes for “man in the loop”. In the language of the standard, the core value of human dignity then unfolds in the value quality of authorizing target fire: “Human dignity demonstrates itself in human authorization of fire.” Authorization expresses a value quality of the SOI that is intended to protect the human dignity of third parties, but also one's own dignity; it must be technically anchored. At this point, it is not yet clear how authorization will be implemented. Rather, there are potentially several possibilities for authorization, especially in light of the temporal aspect, as indicated above.

The result of the value clean-up is curated tables in which stakeholder statements are each assigned to intrinsic core values. The tables are essential for traceability and mapping stakeholder statements to system functions. They are signed off by the SOI stakeholders and subsequently prioritized.

In the prioritization that then follows, the standard once again demonstrates that it does not engage in moral paternalism. This is because any prioritization of value clusters must also take into account, in particular, the corporate story of the manufacturer of the SOI. In setting priorities, it is therefore essential to know what

mission and vision the manufacturer of the SOI is pursuing. Its legitimate economic interests should and must be taken into account.

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

From Theory to Practice: Ethical Value Requirements (clause 9)

From value qualities like “authorization of target firing” one cannot go directly to implementation. This is 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 firing in the Smart Gepard, a Defense AI would need to provide information to its operator. For example, a Defense AI could (i) communicate its level of belief that it is a legitimate adversary to engage (“98.9 percent probability”) and (ii) provide a countdown (in seconds) of how much longer target engagement would be useful and effective and (iii) provide an operating lever where the human hands over engagement of a target to the Defense AI with the push of a button.

Defining EVR is the most important step in VBE’s work with IEEE 7000™-2021, taking care to formulate EVR so that once implemented as a system functionality, it is testable and assessable. The result: The ethics of a Defense AI become measurable! “A countdown indicates the time to which target firing would be useful.” Operators want to be able to rely on the displayed value – they want to be sure that the AI’s calculation procedure is implemented correctly, that uncertainties of the raw data are included in the calculation, or that the calculation has been statistically tested.

Risk-based engineering of EVRs (clause 10)

The EVR is followed by a final important step. In interaction with the stakeholders, the Value Lead performs a risk assessment. He seeks answers to the question of how at risk are the specified EVRs. What threatens an EVR, and what precautions must be taken to ensure an EVR? Development may be required to correctly calculate the countdown just mentioned not only on known data sets, but also on unknown data sets, or to publish the test data set with the system.

Standard risks such as inadequate testing 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 (control), which must also be prioritized later by software development when implementing the SOI. The control mechanism must be specified in such a way that it is capable of ensuring EVR.

However, intrinsic values and their value qualities are not considered a standard risk if they are threatened: “With 98.9 percent probability, the target is a legitimate adversary.” If hazard threatens, any ranking is prohibited. To determine such risk, the Value Lead therefore undertakes a technology impact assessment, while the development team must take precautions at the earliest possible stage to manage and mitigate risks with very high hazard potential.

4. CONCLUSION

“Strictly speaking, what you are creating is the Bundeswehr’s leadership doctrine for AI. It’s not exclusive to humans.” That Defense AI provided to the German Bundeswehr also takes into account the universe of values that members of the German Bundeswehr assign to their leadership doctrine is an urgent desire of the Bundeswehr. It can be fulfilled with VBE according to IEEE 7000™-2021, a global standard that does not need to be adapted for Defense AI but could be detailed, if necessary, by existing frameworks such as the NATO Risk Assessment Tool, however, the standard requires far more than the limited lists of predefined software quality attributes provided by policymakers or the AI industry. In the meantime, the standard is rolling out, and training centres for Value Leads with the gift and training of both a philosophical and technical understanding are proliferating. Nothing less than this does justice to a standard whose intrinsic beauty lies in the logic and structure of its philosophical part, which pursues only one goal: better technology serving the wellbeing of people.

5. RECOMMENDED LITERATURE

[1] **Birnbacher, Dieter.** 2013. *Analytische Einführung in die Ethik.* 3. durchgesehene Edition. Berlin: De Gruyter.

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

[3] **Spiekermann, Sarah.** 2015. *Ethical IT Innovation: A Value-Based System Design Approach.* Boca Raton, Fla: Auerbach Publications.

[4] **Spiekermann, Sarah and Winkler, Till.** 2022. *Value-based Engineering with IEEE 7000^{TM}*. Pre-print version. DOI: 10.48550/arXiv.2207.07599.

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

[6] **Tzimas, Themistoklis.** 2022. *Legal and Ethical Challenges of Artificial Intelligence from an International Law Perspective*. Law, Governance and Technology Series, Vol. 46. Heidelberg: Springer.