

On Digital Ethics for Artificial Intelligence in Hybrid Military Operations

Wolfgang Koch
Fraunhofer FKIE
GERMANY

wolfgang.koch@fkie.fraunhofer.de

ABSTRACT

For being able to defend their way and values of life in an increasingly fragile world, western democracies united within the framework of NATO must be capable “to fight at machine speed” if necessary. For this reason, digitization in defense cannot not be confined to logistics, maintenance, intelligence, surveillance, and reconnaissance, but must equally enable responsible weapon engagement. With a focus on a European Future Combat Air System (FCAS), we address aspects of ethically aligned systems engineering for AI-based weapon systems that might find a broader consent within the international community [1]. In the FCAS program, the largest European armament effort since WW II, manned jets are elements of a networked system of systems, where unmanned ‘remote carriers’ protect the pilots and assist them on combat missions. In view of ongoing debates, the German Minister of Defense has emphasized: “The idea of strategic autonomy for Europe goes too far if it is taken to mean that we could guarantee security, stability and prosperity in Europe without NATO and without the US. That is an illusion [2].” In this sense, FCAS is aligned to NATO’s goals.

INTRODUCTION

“The more lethal and far-reaching the effect of weapons are, the more necessary it is that people behind the weapons know what they are doing,” observes General Wolf von Baudissin (1907-1993), the visionary architect of the Adenauerian *Bundeswehr*, the German post-WWII armed forces, founded in 1955 (see Fig.1). “Without the commitment to the moral realms, the soldier threatens to become a mere functionary of violence and a manager.” Thoughtfully, he adds: “If this is only seen from a functional point of view, i.e. if the goal to be achieved is in any case above human beings, armed forces will become a danger [3].”



Fig. 1: “The most highly mechanized combat requires [...] making soldiers aware of their responsibility and making them experience the consequences of their actions and omissions.”
Wolf von Baudissin (1954). © Bundeswehr

Francis Bacon's (1561-1626) statement on achieving power as the meaning of all knowledge marks the very beginning of the modern project [4]. Since the advent of Artificial Intelligence (AI) in the defense domain, however, technology meant for the benefit of humanity may turn against it. This type of instrumental knowledge makes the modern crisis as visible as in spotlight. Ethical knowledge of man, of his nature and his ends, must complement Baconian-type knowledge. There is an "ecology of man", reminds a German pope the German parliamentarians: "He does not make himself; he is responsible for himself and others [5]." Any ethically aligned engineering must thus be anthropocentric. This is most pressingly true for AI in Defense. Digital ethics and a corresponding ethos and morality are thus essential skills to be built up systematically in parallel to technical excellence. Leadership philosophies and personality development plans should therefore encourage ethical competence for designing and using AI-based defense systems.

How can the science and technology community of NATO STO *technically* support responsible use of the great power we are harvesting from AI? To argue more specifically, we let us guide by documents of the German *Bundeswehr* from its foundation in the 1950s, when the term AI was actually coined, to most recent statements. Since these armed forces have learned lessons from tyranny and "total war" characterized by HighTech of this time, they seem conceptually be prepared for mastering the digital challenge. This is even more true since the *Bundeswehr* is a parliamentary army enshrined in the German *Grundgesetz*, which acts exclusively in accordance with specific mandates from the *Bundestag*, i.e. in the name of the German people.

AI in defense intends to unburden military decision-makers from routine or mass tasks and 'to tame' complexity to let them do what only persons can do, i.e. to perceive a situation intelligently and to act responsibly. The importance of automation for the *Bundeswehr* was recognised early: "Then, human intelligence and manpower will once again be able to be deployed in the area that is appropriate human beings," von Baudissin formulated in 1957 [6]. From this point of view, armed forces are not facing fundamentally new challenges as users of AI-based systems, since the technological development has always extended the range of perception and action.

TECHNICAL ASSISTANCE FOR MINDS AND WILLS

'Intelligence' and 'autonomy' are omnipresent phenomena in the *biosphere*. Before any scientific reflection or technical realization, all living creatures fuse sensory perceptions with information they have learned themselves and received from other creatures. This gives them a model of their environment, the basis to act appropriately for reaching their goals and avoiding harm. In the *technosphere*, Artificial Intelligence (AI) and Information Fusion (IF) combined with comprehensive automation provides technical tools that enhance the perceptive mind and active will of persons who alone are capable to perceive consciously and to act responsibly. We thus deliberately leave the term 'AI' imprecisely defined here. For us, it comprises not only Machine or Deep Learning, e.g., but a whole world of algorithms, including approaches to Bayesian Learning.

As illustrated in Fig. 2, algorithms, realized by the craft and art of programming and enabled by qualitatively and quantitatively appropriate testing and training data, drive a data processing cycle that starts from elementary real-time sensor signals and observer reports collected from multiple and heterogeneous sources. Information Fusion combines these streams of mass data with context knowledge and provides pieces of mission-relevant information at several levels that are integrated into comprehensive and near real-time situational awareness pictures. On their basis, decision makers become aware of the current situation and decide to act according to the ends of their mission in a challenging environment. Algorithms transform their acts of will into partially or fully automated command sequences for controlling networking platforms, sensors, and effectors. Among the AI algorithms are Neural Networks and Machine Learning.

What ends guided Konrad Adenauer (1876-1967) who led Germany into NATO? They seem to timeless and can shape the ‘how’ and ‘to-what-end’ of technically designing an FCAS. For the first post-war Chancellor, NATO was a community of free nations determined “to defend the common heritage of Western culture, personal freedom and the rule of law”, he emphasised on May 9, 1955. Therefore and “in view of the political tensions in the world” its ends “correspond completely to the natural interests of the German people, who [...] long for security and peace like hardly any other people.” Common defense, however, must be embedded in “the promotion of the general welfare of the peoples and, in order to preserve their common cultural heritage, cooperation in economic and cultural matters”. Germany would “devote all its energies to ensuring that human freedom and human dignity are preserved [11].”

INNERE FÜHRUNG AS A GUIDING PRINCIPLE

The ‘Concept of the *Bundeswehr*’, updated in 2018, stimulates a philosophical discussion, when emphasizing the principles of *Staatsbürger in Uniform* [citizen in uniform] and *Innere Führung*, as “the underlying philosophy of leadership valid for the German soldiers [12].” The term *Innere Führung* [leadership from within] is not easily translated. According to von Baudissin’s notion, it comprises all aspects of military leadership with special consideration of the individual and social aspects of the person as such. Its overall goal is to reconcile the functional conditions of operational armed forces with the principles of a democratic constitutional state [13]. This leadership philosophy closely links the personality development of German soldiers with the Common Good, deeply rooted in Western moral thinking.

Essential is the role of the ends to be reached, which superior leaders specify, usually combined with a time frame and the required forces and means. Within this framework, subordinate leaders pursue and reach the ends independently, i.e. they are largely free to fulfill their mission, but inform the higher leadership levels about the status of the mission in order to enable corrections if necessary. Own responsibility as well as judgement and decisiveness are important characteristics required of the soldiers.

Therefore, digitalization in defense can only be successful if it adapts *Innere Führung*. Its ‘digital update’ leads to a timeless question: How can ethical, legal and social compliance be guaranteed? Any answer leads to two distinct research questions for defense scientists and engineers:

1. How can we intellectually and morally continue to be the masters of our tools?
2. Which design principles of systems engineering facilitate responsible use of AI?

FCAS ETHICAL AI DEMONSTRATOR

Algorithms drive an information cycle by processing massive amounts of data that can no longer be handled by humans in the complex technosphere of Manned-unManned Teaming (MuM-T, see Fig. 3). In this way, they *cognitively* assist the minds of military decision-makers in understanding complex, spatially distributed and variable situations and *volitively* support “the enforcement of their will in every situation”[14] in terms of appropriate and responsive action.

The concepts of mind and will and, therefore, of consciousness and responsibility bring natural beings into view that are ‘somebody’ and not ‘something’, i.e., persons. Cognitive and volitive assistance systems on the other hand, whatever degree of technical sophistication they will reach, are and will always be ‘something’. It seems important to stress the dichotomy ‘something vs. somebody’ to counter both overexpectation and overfear that apparently characterize the public apprehension. The pop culture reveals a psychogram of modern man with pseudoreligious hopes and gloomy foreboding, while the Image of Man is increasingly shaped according to the model of machines, while human, even superhuman qualities are attributed to machines.



Fig. 3: AI und IF driven “system of systems” with Manned-Unmanned Teaming (MuM-T) for combat and reconnaissance missions. © Fraunhofer FKIE

In Germany, but also in many NATO partner nations, there is a basic consensus that the final decision on the use of armed force must be reserved for a human decision-maker. In view of automation and the use of AI procedures in warfare – especially by potential opponents – it must be clarified on which technically realizable basis a human operator can ultimately make balanced, consciously considered decisions regarding the use of armed force (meaningful authorisation). This is particularly true in case of AI algorithms such as Deep Learning (DL) that have the character of a ‘black box’ for the user.

For this reason, it is important to make the AI-based findings comprehensible and explainable to human decision makers, and, on the other hand, to prevent soldiers from confirming recommendations for action without weighing them up themselves, simply on the basis of some kind of “trust” in the AI-based system. Especially for FCAS, engineers must develop comprehensible and explainable methods. The ‘FCAS Ethical AI Demonstrator’ will let soldiers experience the use of AI in a military scenario with all associated aspects of psychological stress as realistically as possible. Examples such as automated target recognition for decision making in air combat enable interaction with a real AI developed for military use in order to enable a realistic view of the possibilities, limits, ethical implications, and engineering demands of this technology in practice.

FIRST STEPS TOWARDS REALISATION

An exploratory study by industry, a start-up company, and Fraunhofer currently paves ways towards such a demonstrator. Discussions with the German Air Force clarified the scenarios considered. This way ahead is a novelty within the framework of large European armament projects and therefore still has an experimental character. Nevertheless, it is a ‘quantum leap’ in the sense that it is a tiny and actually the smallest possible step. However, a real and unprecedented leap it is.

One of the missions envisaged for FCAS is the elimination of enemy air defense, where it is conceivable that remote carriers with electro-optical and signal intelligence sensors collect data on positions of supporting equipment of enemy air defense systems. This use case comprises the following (simplified) steps:

1. Exploiting control of multiple sensor systems on a remote carrier, the user will detect, identify, and track enemy vehicles in different scenarios with AI support.

2. The AI system will graphically highlight relevant objects accordingly and enrich them with basic context information (e.g. type of detected vehicle, confidence).
3. The user, who is virtually in the role of the payload operator of the remote carrier flying ahead, has the task of recognising and identifying all relevant objects.
4. To this end, optional confirmation dialogues provide information for all individual objects recognised or preselected by the AI system that is much more detailed.

This dialogue offers the possibility

- to take a magnified image of the object in question to confirm the target by visual address, to understand in the magnified section by means of appropriate highlighting of the Explainable AI (XAI) which has recognised elements of the object under track;
- to exploit sensor data fusion with additional data sources, to understand which sensor technology, if any, has "tipped the scales" for classification as hostile object, as well as to view corresponding levels of confidence for the respective sensor category;
- to check compliance with the rules of engagement for the object in question, insofar as a deterministic algorithm can provide support here; confirm compliance with the rules of engagement as checked.

Apparently, this dialogue may provide a more unambiguous identification of an object as hostile.

RESPONSIBLE CONTROLLABILITY

Any responsible use of technology requires continuous controllability. In some applications, occasional malfunction of AI and automation may have no consequences. In its military use, however, rigorous requirements must be guaranteed with all legal consequences. The use of technically uncontrollable technology is immoral *per se*. We stress selected aspects to be considered in designing AI-based weapon systems responsibly.

1. The notion of ‘meaningful human control’ needs to be interpreted more broadly than the concept of ‘human-in/on-the-loop’ suggests. Formulations such as “For unmanned aerial vehicles, the principle of human-in-the-loop and thus the immediate possibility of operator intervention must be ensured at all times” in official documents should be reconsidered [15]. More fundamental is ‘accountable responsibility’. The use of fully automated effectors on unmanned platforms may well be justifiable, even necessary in certain situations, if appropriately designed.
2. Certification and qualification of AI-based automation are key issues. Robust military systems will comprise both data-driven and model-based algorithms, where data-driven AI could be confined by model-based reasoning—‘AI in the Box’. Predictable system properties, insensitivity to unknown effects, adaptivity to variable usage contexts, and graceful degradation must be verified. Statistical testability as well as explainability are essential prerequisites for critical components. Finally, compliance to a code of conduct must be guaranteed by design.
3. With a view on the working group ‘Responsible Technologies for an FCAS’, we finally suggest that comprehensive analyses of technical controllability and personal accountability should typically accompany digitization projects in a publicly visible, transparent, and verifiable manner. Otherwise, the grave paradigm shifts and large material efforts associated with AI in defense would hardly be politically, societally, and financially enforceable.

SAPERE AUDE AS THE KEY

Situation pictures are ‘fused’ from sensor and context data that never meet ideal expectations. They are always imperfect, inaccurate, ambiguous, unresolved, corrupted or deceptive, difficult to be formalized, or even contradictory. Probabilistic algorithms, however, enable responsible action even on an imperfect data basis. In many cases, reliable situational pictures can be inferred in a much more precise, complete, and much faster way than humans could ever have hoped to obtain. Nevertheless, also these methods have their limitations, which decision-makers must be made aware of.



Fig. 4: “Without the commitment to the moral realms, the soldier threatens to become a mere functionary of violence and manager, [...] degraded to weapons without human cohesion and conscience; with them every act of violence becomes possible.” Wolf von Baudissin (1967). © NATO

Moreover, data integrity is fundamental to any use of AI: Are valid sensor and context data available at all? Are they produced reliably and do the deficits correspond to the assumptions made? In naive systems, violated integrity easily turns data fusion into *confusion*. Moreover, algorithms always generate artifacts that do not exist in reality, or have ‘blind spots’, i.e., do not show what is actually there. In case of cyber attacks, enemies may take over sensors or subsystems, which then produce deceptive data or unwanted action.

Mature AI comprises detection of such deficits, the basis for making them resilient toward hostile interference. The capability of ‘artificial self-criticism’ in this sense requires naturally intelligent critical capabilities on the part of decision-makers vis-à-vis AI. Otherwise, there is a danger of uncritical acceptance of what is offered, and ultimately a refusal to actually bear responsibility. AI-based systems must therefore be technically designed in such a way that they train the vigilance of their users and convey to them how the machine solutions were created – AI must not ‘dumb down’ its users.

Only alert natural intelligence is able to assess plausibility, develop understanding and ensure control. “The uncontrolled pleasure in functioning, which today is almost synonymous with resignation before the technical automatism, is no less alarming [than the dashing, pre-technical feudal traditions] because it suggests the unscrupulous, maximum use of power and force,” von Baudissin observes [16, see also Fig.4].” This addresses not only the soldierly ethos. Don’t all of us need a new enlightenment for dealing with AI in a mature and ethical way, “man’s emergence from his self-imposed nonage”? *Sapere aude*—Have the courage to use your own intellect [17]!“

HIPPOCRATIC OATH – AN ANALOGY?

Only if based on Image of Man that is compatible with responsible use of technology, digital assistance can support morally acceptable decisions. “It is the responsibility of our generation, possibly the last to look back to pre-digital ages and into a world driven by artificial intelligence, to answer the question of whether we continue to recognize the integrity of the human person as a normative basis,” thoughtfully observes Ellen Ueberschär (b. 1967), an influential German political thinker [18].

Reminding of such an Image of Man and *Innere Führung* towards this image is a task of military pastoral care. Since the Hippocratic Oath is regarded as a symbol of a professional ethic that is committed to responsibility, it would be worth considering whether the swearing-in ceremony, which was considered indispensable when the *Bundeswehr* was founded, should be viewed with a fresh eye. For von Baudissin it is “one of the essential tasks of the military clergy to point out the sanctity of the oath, as well as the vow, to show the recruit the seriousness of the assumption of his official duties on his own conscience, but at the same time also the limits set by God for everyone and also for this obligation [19].”

In this spirit, Konrad Adenauer said farewell to the *Bundeswehr* in 1963 “as the most visible expression of the reconstruction of Germany, as the restoration of order, as proof of the integration into the front of free nations.” Adenauer’s motives may also apply to the ‘why’ and ‘how’ of an FCAS and the use of AI in defense in general in NATO’s future missions: “Soldiers, if we had not created our armed forces, we would have lost freedom and peace long ago. So you, soldiers, through the work you have done, have in truth given and preserved peace for the German people [20].”

CONCLUDING RECOMMENDATIONS

The Digital Defense Council to the German Minister of Defense, states that “Digitization affects more than the aspect of technical innovation. It influences the entire way of thinking and acting of the *Bundeswehr* at all levels in the sense of a ‘digital self-image’ of the *Bundeswehr* [21].”

Since we feel encouraged to assume that there might be a broader consent within the community of NATO STO with these considerations, we are closing with some recommendations.

1. Digital ethics and a corresponding ethos and morality are to be part of the skills that need to be built up systematically for responsibly using AI in Defense without serious harm for humanity. In particular, such skills enable military decision makers “to assess the potential and impact of digital technologies and to manage and to lead in a digitized environment [22].” In particular, *consideration should be given to leadership philosophies and personality development instruments* such as *Innere Führung* as a guiding principle for the development of ethical competence and to encourage its systematic development with regard to AI in Defense.
2. In addition to the operational benefit of defense digitization in closing capability gaps, expanding the range of capabilities, and developing corresponding concepts, operational procedures, and organizational measures, *ethical competence in dealing with digital technologies and their ethical acceptance need to be achieved*. Only then, AI in Defense will become acceptable before the conscience of the individual soldiers, but also in the broader view of the Common Good of the society as such. Success in both aspects will indicate a real innovation.
3. Digitization projects should be accompanied by comprehensive analyses of *technical controllability and personal accountability in a publicly visible, transparent, and verifiable manner*. Otherwise, the paradigm shifts and material efforts associated with artificial intelligence and automation would hardly be politically, societally, and financially enforceable. Of course, there will be more problematic and less problematic projects, implying that an exemplary approach would be appropriate.

4. Although the International Law requires human responsibility, it defines an “ethical minimum” only. To achieve progress in the legal development, it seem promising to consider Corporate Social Responsibility. The principles anchored in this way aim to *create binding standards for the responsible design and leverage of cognitive and volitive systems, including the entire supply chain*. Misconduct is to be tangibly sanctioned – soft law and hard sanctions, ranging from extraordinary contract termination to contractual penalties. Although soft law is not enacted by the legislature, it may become as a source of legal knowledge.

The USAF General John E. Hyten (b. 1959), Vice Chairman of the Joint Chiefs of Staff, has characterized the situation of including AI in defense by re-phrasing a passage from John F. Kennedy’s (1917-1963) famous Moon Speech [23]:

“We set sail on this new sea because there is new knowledge to be gained, and new rights to be won, and they must be won and used for the progress of all people. For [artificial intelligence], like nuclear science and technology, has no conscience of its own. Whether it will become a force for good or ill depends on man, [...] whether this new ocean will be a sea of peace or a new terrifying theater of war.”

REFERENCES

- [1] See also: W. Koch, “On Digital Ethics for Artificial Intelligence and Information Fusion in the Defense Domain,” in *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 7, pp. 94-111, July 1, 2021, doi: 10.1109/MAES.2021.3066841.
- [2] A. Kramp-Karrenbauer, *Second Keynote Speech by German Federal Minister of Defense*. Hamburg: Helmut Schmidt University, Nov. 19, 2020. [Online]. Available: <https://www.bmvg.de/en/news/second-keynote-speech-german-minister-of-defence-akk-4503976>
- [3] W. von Baudissin, *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr* [Soldier for Peace. Drafts for a Contemporary *Bundeswehr*]. München: Pieper, 1969, p. 205.
- [4] The phrase *ipsa scientia potestas est* was made popular by Thomas Hobbes (1588–1679), who was a secretary to Bacon.
- [5] Benedict XVI, *Address to the German Parliament*. Berlin: Bundestag, Sep. 22, 2011. [Online]. Available: <https://www.bundestag.de/parlament/geschichte/gastredner/benedict/speech>
- [6] Baudissin, p. 174.
- [7] *Erster Bericht zur Digitalen Transformation* [First Report on Digital Transformation]. Berlin: MoD, 10/2019. [Online]. Available: <https://www.bmvg.de/resource/blob/143248/7add8013a0617d0c6a8f4ff969dc0184/20191029-download-erster-digitalbericht-data.pdf>
- [8] *Konzeption der Bundeswehr* [Concept of the *Bundeswehr*]. Berlin: MoD, 2018, p. 83. [Online]. Available: <https://www.bmvg.de/resource/blob/26544/9ceddf6df2f48ca87aa0e3ce2826348d/20180731-konzeption-der-bundeswehr-data.pdf>
- [9] The responsible use of new technologies in a Future Combat Air System, www.fcas-forum.eu

- [10] R. Spaemann et al., *Natürliche Ziele* [Natural Ends]. Stuttgart: Klett-Cotta, 2005, p. 51.
- [11] K. Adenauer, *Aufnahme der Bundesrepublik Deutschland in die NATO* [Admission of the Federal Republic of Germany to NATO]. Paris: Palais de Chaillot, May 9, 1955. [Online]. Available: <https://www.konrad-adenauer.de/quellen/reden/1955-05-09-rede-paris>
- [12] *Konzeption der Bundeswehr*, p. 50.
- [13] H. Dierkes, Ed., “Global warriors? German soldiers and the value of *Innere Führung*,” *Ethics Armed Forces*, no. 2016/1, Jan. 2016, p. 46. [Online]. Available: <http://www.ethikundmilitaer.de/en/full-issues/2016-innere-fuehrung/>
- [14] *Konzeption der Bundeswehr*, p. 84.
- [15] *Militärische Luftfahrtstrategie 2016* [Military Aviation Strategy]. Berlin: MoD, 2016, p. 23. [Online]. Available: <https://www.bmvg.de/resource/blob/11504/3e76c83b114f3d151393f115e88f1ffb/c-19-01-16-download-verteidigungsministerium-veroeffentlichtmilitaerische-luftfahrtstrategie-data.pdf>
- [16] Baudissin, p. 180.
- [17] I. Kant, *An Answer to the Question: What is Enlightenment?* (1784, transl.: M. C. Smith)
- [18] E. Ueberschär, *Opening Speech*. Bad Aying: Responsible technology for an FCAS, Sep. 27, 2019. [Online]. Available: www.fcas-forum.eu
- [19] Baudissin, p. 181.
- [20] K. Adenauer, *Ansprache*. Wunstorf, Germany: Parade of the *Bundeswehr*, Oct. 12, 1963. [Online]. Available: <https://www.konrad-adenauer.de/quellen/reden/1963-10-12-ansprache-wunstorf>
- [21] German MoD (2019). *Umsetzungsstrategie Digitale Bundeswehr* [Implementation Strategy Digital Bundeswehr]. 14.06.2019, Nr. 102, p. 4. <https://www.bmvg.de/de/themen/ruestung/digitalisierung/umsetzungsstrategie-digitale-bundeswehr>.
- [22] *Umsetzungsstrategie Digitale Bundeswehr*, Nr. 209, p. 8.
- [23] John E. Hyten (2020), “Remarks to the Joint Artificial Intelligence Symposium,” September 9, 2020, <https://www.defense.gov/Newsroom/Transcripts/Transcript/Article/2344135/remarks-by-general-john-e-hyten-to-the-joint-artificial-intelligence-symposium/>. See also: John F. Kennedy (1962), “Address at Rice University, Houston, Texas, 12 September 1962,” John F. Kennedy Presidential Library and Museum, <https://www.jfklibrary.org/asset-viewer/archives/JFKPOF/040/JFKPOF-040-001>.