Elements of an Ethical AI Demonstrator for Responsibly Designing Defence Systems

Wolfgang Koch
Fellow, IEEE
Fraunhofer FKIE
Wachtberg, Germany
w.koch@ieee.org

***Abstract*—In order to protect their common heritage of culture, personal freedom, and the rule of law in an increasingly fragile world, democracies must be able to defend themselves "at machine speed" if necessary. The use of AI in defense therefore comprises responsible weapons engagement as well as military use cases such as logistics, predictive maintenance, intelligence, surveillance or reconnaissance. This poses a timeless question: How to decide *well* according to what is recognized as *true*? For approaching towards an answer, responsible controllability needs to be turned into three tasks of systems engineering: (1) Design artificially intelligent automation in a way that human beings are mentally and emotionally able to master each situation. (2) Identify technical design principles to facilitate the responsible use of AI in defence. (3) Guarantee that human decision makers always have full superiority of information, decision-making, and options of action. The Ethical AI Demonstrator (E-AID) proposed here for air defence is paving the way by letting soldiers experience the use of AI in the targeting cycle along with associated aspects of stress as realistically as possible.**

Keywords—ethically-aligned engineering, Artificial Intelligence (AI), automation, cognitive and volitive assistance, Future Combat Air System (FCAS), targeting cycle, target designation.

I. INTRODUCTION

Artificially intelligent automation provides new types of machines that greatly enhance the perceptive mind and the active will of persons, who alone are capable to perceive intelligently and to act autonomously in a proper sense.

1. *Cognitive machines* fuse massive streams of sensor, observer, context, and mission data for producing comprehensive situation pictures, the basis for conscious human cognition to plan, perceive, act, and assess effects appropriately.
2. *Volitive machines* transform deliberately taken overall decisions of responsible human volition into complex chains of automatically executed commands for data acquisition, sub-system control, and achieving effects on objects of interest.

Such machines will become key elements of the Future Combat Air System (FCAS), the largest European armament effort since WW II for protecting European sovereignty. In this program, manned jets are only elements of a larger networked system of systems, where unmanned 'remote carriers' protect the pilots as 'loyal wingmen' and support them on reconnaissance and combat missions. By technically assisting their minds and wills cognitively and volitively, air commanders and staffs will remain capable of appropriately acting even on short time scales in the complex 'technosphere' of modern warfare with spatially distributed and highly agile assets. This is particularly true, when targeting cycles are vastly accelerated and to be executed 'at machine speed' in a network-centric and collaborative way.

"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 Wolf von Baudissin (1907-1993), the visionary architect of Germany's post-WW II armed forces, the *Bundeswehr*. "Without the commitment to the moral realms, the soldier threatens to become a mere functionary of violence and a manager," he continues and thoughtfully adds: "If this is only seen from a functional point of view, i.e., if the goal to be achieved is in any case put above human beings, armed forces will become a danger" [1, p. 205]. It is in this sense that we consider aspects of AI-driven targeting cycles and their responsible design.

The more general key question this paper is intending to help answering therefore reads: How can the information fusion community *technically* support responsible use of the great power we are harvesting from artificially intelligent automation? While facing soberly the risks of digitalization in defence, we nevertheless beware exaggerating them, which may become a risk in itself and prevent innovation in defence. Despite of our clear military focus, we hope that our considerations below might enjoy a broader consent also in civil decision-making.

As will become visible, the use of AI in defense systems of systems such as FCAS intends to unburden military decision-makers from routine or mass tasks. We in particular need to tame technical complexity in such a way that commanders, staffs, and soldiers will be able to focus on doing what only persons can do, i.e., to consciously perceive a situation intelligently and act responsibly. The importance of automation for armed forces was recognized as early as 1957, when von Baudissin wrote that thanks to automation, "human intelligence and manpower will once again be able to be deployed in the area that is appropriate to human beings."[1, p. 174] Seen from this perspective, armed forces do not face fundamentally new challenges as users of artificially intelligent automation, since the technological development has long extended the range of perception and action.

In order to be able to argue in a more focussed way in the sense of a use case, we will examine conceptual documents of the German *Bundeswehr* in our approach that span the period from its founding in the 1950's, when the term 'AI' was actually coined, to its most recent statements on the matter. Since theses armed forces have learned lessons from the totalitarian

Preprint, to appear in Proc. of IEEE International Conference on Information Fusion, July 4-7, 2022, Linköping, Sweden

tyranny in Germany from 1933 to 1945 and the horrors of "total war" [2], characterized by high technology of this time, they are presumably in a conceptual way well prepared for mastering the digital challenge we are confronted with today. This is even more the case, since the *Bundeswehr* is a parliamentary army enshrined in the German Constitution, *Grundgesetz*, which acts exclusively in accordance with specific mandates from the *Bundestag*, i.e., on behalf of the German people.

With a focus on 'combat clouds', we introduce in section II the notions of reflective and normative assistance in military decision-making. Here, ethically relevant implications demanded by official documents are considered that shape the ethics, ethos, and morality of dealing with AI-based weaponry. Based on the fundamental notion of 'responsibility' and its relation to systems engineering aspects, section III discusses design principles of the FCAS Ethical AI Demonstrator, the core contribution of this paper. Considerations towards normative assistance close this section. The problem of transparent criteria development is addressed in section IV, which has implications on acquiring 'digital virtues' in dealing with AI in defence and might establish an analogy between the Hippocratic Oath and soldierly ethos. Finally, we try to draw conclusions in a more generalizing sense.

This paper is harvesting fruits of ongoing discussions in the working group *Responsible Technology for an FCAS* [3] and evolves insights published earlier [4-7]. Our considerations correspond to some extent to the *IEEE P7000 Model Process for Addressing Ethical Concerns During System Design* [8]. Actually, a large community of engineers and technologists is addressing ethical problems and technical realizations to mitigate them throughout the various stages of system initiation, analysis and design for particular use cases, for example, Fair, Accountable or Transparent AI.

Figure 1. Air Combat Cloud enable artificially intelligent automation for military Manned-unManned Teaming. © Fraunhofer FKIE.

II. Artificially Intelligent Combat Clounds

From a digitalization perspective, the core infrastructure for future air defence and combat systems are *air combat clouds*, symbolically visualized in Figure 1. While sensors are collecting data, combat clouds distribute, verify, validate, organize, evaluate, process, and fuse data to enable adaptive management of sensors, platforms, communication links, and effectors such as weapons 'at machine speed'. In the digital age, information superiority in complex situations and decision dominance even at very short time scales decide between success and failure of a mission. According to the introductory remarks, the architecture of a combat cloud, i.e. of the informational backbone for military air operations, has to facilitate the responsible use of weapon systems by human decision makers. Artificially intelligent automation is crucial here, since it enables complexity management and responsible action by providing cognitive and volitive assistance. In parallel, 'digital twins' accompanying the technological development from the very beginning have to ensure that comprehensive ethical and legal compliance is not at the expense of effectiveness in air defence and combat.

We here use the term 'Artificial Intelligence' in a sense that does not only comprise machine or deep learning, for example, but a whole 'cloud' of data-driven and model-based algorithms, including approaches to Bayesian learning, game theory, and adaptive resources management. It seems worthwhile to consider 'Artificial Instinct' as a more appropriate of the acronym 'AI' that was proposed by the Polish science fiction author, philosopher, and futurologist Stanisław Lem (1921-2006) nearly 40 years ago [9].

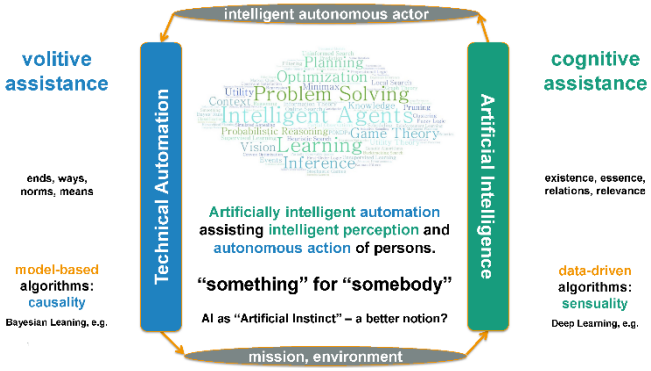

Figure 2. Cognitive and volitive assistance for the intelligent mind and autonomous will of responsibly acting commanders. © Fraunhofer FKIE

As illustrated in Figure 2, a 'cloud of algorithms', realized by the art and craft of programming and enabled by qualitatively and quantitatively appropriate testing and training data, drive a data processing cycle that starts from elementary signals, measurements, and observer reports collected from multiple and heterogeneous sources. For us, 'AI' denotes the process that fuses such streams of mass data and context knowledge, which provide pieces of mission-relevant information at several levels, for producing comprehensive and near real-time situation pictures. On their basis, air commanders and staffs become aware of the current situation in a challenging environment and the status of the mission. Human decision-making for acting according to the ends of the mission to be achieved are made at different levels of abstraction and degrees of detail. Technical Automation transforms deliberate acts of will into complex command sequences to control networking platforms, multifunctional sensors, and effectors.

Algorithms for comprehensively harvesting information by data fusion and adaptively managing the various processes of data collection as well as weapon engagement and effect assessment belong to the methodological core of cognitive and volitive machines that assist the intelligent mind and autonomous will of decision makers. They exploit sophisticated methods of applied mathematics and run on powerful computing devices, where quantum computing may become a game changer [10, 11]. The concepts of mind and will and therefore of consciousness and responsibility bring human beings as persons into view that are "somebody" and not "something."

A. Reflective and Normative Assistance

While artificially intelligent automation is indispensable for achieving situational awareness, a prerequisite of reducing collateral damage, for example, as well as of commanding resources, it also implies specific vulnerabilities such as

1. *loss of data integrity* causing invalid situation pictures and improper decisions due to unintended malfunction of sensors, programming errors, misuse of training data, or data incest,
2. *artifacts generated by AI* algorithms from sensor and context data that do not exist in reality, or blind spots, that are disabling situation pictures to show what is actually present in reality,
3. *hostile intervention* at various levels to be taken into account, where adversaries take over sensors or subsystems, which then produce deceptive data or initiated unwanted action, and
4. *more general issues* of automated systems such as misuse, disuse, abuse, non-use, and blind or overly trust, which are not specifically AI-related, but must be taken in account as well.

See [5] for a more detailed discussion. In consequence, resilient cognitive and volitive machines for defence systems of systems have to comprise the detection and compensation of such deficits in the sense of 'Artificial Self-criticism'.

Any ethically and legally acceptable use of cognitive and volitive machines relies on 'truth', defined as 'equivalence between awareness and the actual situation' and 'goodness', defined as 'equivalence between the choices made and norms'. Their proper use, however, needs to be supported by 'reflective' and 'normative' assistance functions, seen as part of ethically-aligned cognitive and volitive machines as discussed below. Fusion of sensor data and non-sensor information provides mission-relevant insights. Apparently, comprehensive information fusion is the key to seamlessly integrating also formalized ethical or legal constraints, seen as a particular type of context knowledge, into reconnaissance or combat missions. For the sake of simplicity, we confine the discussion of the normative framework to the Rules of Engagements (RoEs) that have to mirror the risks of artificially intelligent automation and must permeate the technical system design.

B. Ethical Implications of Basic Documents

According to the foundational document of the German *Bundeswehr* [12, p. 83], updated in 2018, artificially intelligent automation is expanding its capability profile by providing

1. perception of a military situation as reliably as possible by "obtaining, processing, and distributing information on and between all command levels, units and services with minimum delay and without interruption or media disruption;"
2. support of "targeted deployment of forces and means according to space, time and information, […] where characteristic of military leadership are the personal responsibility of decision-makers and the implementation of their will at any time."

Readiness to defend ourselves against highly armed opponents must not only be technologically credible, but also correspond to the consciously accepted "responsibility before God and man, inspired by the determination to promote world peace as an equal partner in a united Europe," as the very first sentence of the German Constitution, the *Grundgesetz,* proclaims [13]. Guided by this spirit and for the first time in Germany, an intellectual struggle over the *technical* implementation of ethical and legal principles accompanies a major defense project from the outset. The goal of the working group on *Responsible Use of New Technologies in an FCAS* is to operationalize ethically aligned engineering [3].

Official documents released by the German Ministry of Defence implicitly define elementary requirements that are relevant for ethically-aligned FCAS systems design and have direct implications for the *Ethical AI Demonstrator* (E-AID) to be discussed below. With a focus on ethically critical tasks within the targeting cycle to be executed by FCAS commanders, E-AID demonstrates, in which way cognitive, volitive, reflective, and normative assistance systems should be developed and how they interact with each other. Also in view of the international law, considerations on the ethical implications are encouraged, since Article 36 of the Additional Protocol I of the 1949 Geneva Conventions requires states to conduct legal reviews of all new weapons, means, and methods of warfare in order to determine whether their use is prohibited [14]. To be mentioned the eleven guiding principles affirmed by a group of governmental experts within the framework of the UN Convention on Certain Conventional Weapons (CCW) [15].

C. On Ethics, Ethos, and Morality

For properly designing cognitive and volitive machines in the context of FCAS, ethical implications need to be clarified while avoiding moralizing. The following distinction proves to be helpful in designing reflective and normative assistance.

1. *Digital ethics* denotes theoretical reflections about right decisions in using artificially intelligent automation. Required is an Image of Man that makes notions such as *mind*, *will*, and, therefore, *consciousness* and *responsibility* conceptually possible.
2. *Digital ethos* addresses the attitude of decision makers on all levels. "The more momentous the decisions and actions of individual soldiers are, the more their ethos must be determined by responsibility," as von Baudissin observed.
3. *Digital morality*, finally, comprises the formulation of concrete guidelines for dealing with artificially intelligent automation, not only in the battlefield, but also in research, development, and procurement, planning, and mission preparation.

Along such considerations, the German Ministry of Defense underlines that „the importance of AI does not lie in the choice between human or artificial intelligence, but in an effective and scalable combination of human and artificial intelligence to ensure the best possible performance" [16]. Comprising ergonomic as well as ethical and legal dimensions of AI, this statement implicitly demands responsible systems engineering and as such aims as well at fulfilling the military requirements previously mentioned. In particular, numerous research questions for systems engineering result that aim at a fundamental military requirement: "Characteristic features of military leadership are the personal responsibility of decision-makers and the implementation of their will in every situation," according to the 'Concept of the Bundeswehr' [12, p. 83].

III. Decision Making for Weapon Engagement

A challenge for valid situational awareness and responsible decision-making for weapon engagement in the FCAS domain is the ever-decreasing time available for human involvement in the decision-making process. Further problems are limited explainability and deceivability of both, algorithmically generated information and automated execution of complex command chains. The following issues need to be addressed.

1. While in certain applications, occasional malfunctioning of AI-enabled automation may have no consequences, rigorous safety requirements must be guaranteed for FCAS with all legal consequences. The military use of technically uncontrollable technology is immoral *per se*.
2. The notion of *meaningful human control* needs to be interpreted more broadly than the concept of *human-in / on-the-loop* suggests [17]. A more fundamental notion is "accountable responsibility". Since the use of fully automated effectors on unmanned platforms may well be justifiable, even necessary in certain situations, the overall system design must guarantee that always a distinct „somebody" is responsible.

In view of these considerations, artificially intelligent automation for FCAS poses a timeless question: Which design principles facilitate 'good' decisions according to what is recognized as 'true' according to the previous definitions? Turned into systems engineering, this implied two tasks:

1. Design cognitive assistance in a way that human beings are not only mentally, but also emotionally able to master each situation.
2. Design volitive assistance to guarantee that human decision makers always have full superiority of information and the options of action.

In consequence, digital ethics as well as a corresponding ethos and morality are essential soft skills to be built up systematically in parallel to technical excellence. Personality development plans should encourage ethical competence for responsibly designing *and* using AI-based cognitive and volitive assistance.

A. On the Notion of 'Responsibility'

Literally, the very word 'responsibility' is rooted in the language at courts of justice designating the obligation of being called upon to 'respond' to questions about one's own actions by a judge, a primal situation of human existence as a person. This overall concept has far-reaching implications.

1. To speak of responsibility is only reasonable if it is assumed voluntarily. Responsibility, thus, presupposes the notion of a 'free will' and an Image of Man as a free and 'autonomous' person.
2. The concept of free will as the decisive cause of decisions to action implies the idea of an accountable person, which is legally relevant and an essential criterion in the International Law.
3. Responsibility, as considered here, implies in addition to the legal notion of accountability the ability of a person to act freely and the willingness to act well even in case of absent or contradicting rules. Casuistry, formalization of human action by just following well-defined rules, seems impossible.
4. The will, responsible in freedom, is not absolute, but depends on the understanding mind. The 'True' as the formal object of the mind and the 'Good' as the formal object of the will thus form the intellectual basis of responsible action.

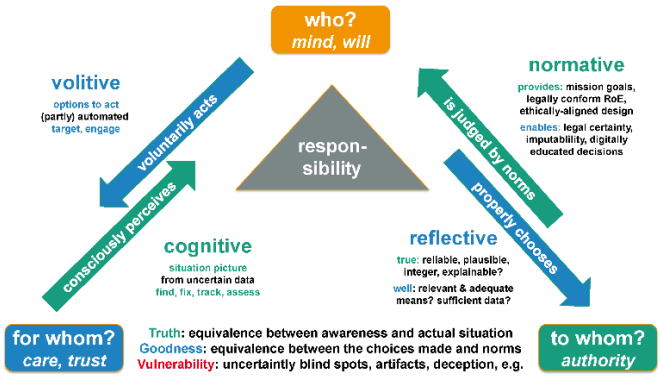

Figure 3. Artificially intelligent automated assistance enabling responsible action for FCAS. © Fraunhofer FKIE

Figure 3 illustrates core elements of the concept of responsibility as a triangle relationship, insofar as it is relevant to the technical systems design. It implies the notion of persons or groups of persons as preciously sketched and establishes characteristic relationships between them. Responsible systems design is, thus, by definition 'anthropocentric'.

1. *Who bears responsibility?* Military capability development takes place at various levels and requires responsible action in research, development, certification, and qualification of military Command & Control, ISR, and weapon systems as well as in the preparation and execution of military operations.
2. *For whom is responsibility borne?* The relationship between responsible persons and those for whom they are responsible is characterized by 'care' and 'trust' and therefore determined by prospective action and reaction. In a proper sense, responsibility can only be assumed by persons for persons. Indirectly, one might speak of responsibility towards animals, cultural heritage, or the environment, for example, insofar as these are related to persons.
3. *To whom is responsibility assumed?* Responsibility implies the notion of a personal authority exercising his or her authority by judgment. The responsible person recognizes this authority by his or her justification. The relationship between responsible persons and a personal authority is retrospective in nature.

Voluntarily assumed responsibility, which shows itself in 'care' and 'trust', as well as in the readiness to justify itself and to choose properly in obedience to norms, keeps military forces stable in combat. It can und should be supported by normative and reflective assistance systems to be specified below. Purely legal constructs, however, such as liability for damage caused by one's actions, are insufficient, especially in military operations.

According to these considerations, no machine can act responsibly or irresponsibly, i.e. in a "good" or "evil" way by responding to moral challenges in one way or another, but persons only. "Good" technical systems encourage the morally acceptable and efficient use of them to achieve military objectives. "Evil" systems facilitate their irresponsible use.

B. FCAS Ethical AI Demonstrator

By the FCAS Ethical AI Demonstrator (E-AID) discussed here we wish to clarify on which technically realizable basis human operators are enabled to make balanced and conscious decisions regarding the use of weaponry based on artificially intelligent automation. One might speak of 'meaningful authorization'. This is particularly pressing in cases where AI algorithms such as Deep Learning (DL) are under consideration, which have the character of a 'black box' for the user.

For approaching a viable solution, it is important to make AI-based findings comprehensible, plausible, or 'explainable' to human decision-makers. On the other hand, soldiers should not confirm recommendations for action without weighing them up themselves, simply based on some kind of 'trust' in the AI-based system. To this end, we introduce the concept of 'reflective' assistance as indicated in Figure 3.

Especially for FCAS, engineers must aim at developing comprehensible, plausible, or 'explainable' methods. With the help of E-AID air commanders and staff experience the use of AI in militarily relevant and close-to-reality scenarios by displaying all associated aspects of psychological stress as realistically as possible. Selected features of the Ethical AI Demonstrator, such as automated target recognition for decision-making in air combat enable interaction with an actual AI developed for military use in order to enable a realistic view of the possibilities, limitations, ethical implications, and engineering demands of this technology in military practice.

Discussions with the officers of the German Air Force have clarified the scenarios to be considered. One of the missions envisaged for FCAS is the elimination of enemy air defense using remote carriers with electro-optical and signal intelligence sensors that collect data on positions of equipment that is supporting enemy air defence. The (much)-simplified steps in such a use case proceed as follows:

- The user will detect, identify, and track enemy vehicles in different scenarios with and without AI support for comparison, by exploiting control of multiple sensor systems on a remote carrier.
- The output of the AI system is used to graphically highlight relevant objects accordingly and enrich them with basic context information (e.g., type of detected vehicle, certainty level).
- The user, who is in the role of a virtual payload operator of the remote carrier flying ahead, has the task of recognizing and identifying all relevant objects.
- Manual target designation needs to be analyzed here as well, which is currently done by AI only and could be seen critical, even unethical.
- To facilitate the user's ability to perform this task, optional confirmation dialogues provide information for all individual objects recognized or preselected by the AI system at a much greater level of detail.

This dialogue will enable the following:

1. To request a magnified image of the object in question to confirm the target by visual address, and to understand in the magnified section by means of appropriate highlighting of Explainable AI (XAI) which has recognized elements of the tracked object;
2. to enhance sensor data fusion with additional data sources, to understand which sensor technology, if any, has "tipped the scales" for classification as a hostile object, and to visualize corresponding levels of confidence for the respective sensor category;
3. to check compliance with the rules of engagement for the object in question, insofar as deterministic algorithms can provide support here; to confirm compliance with the rules of engagement as checked.

Ostensibly, such a dialogue should provide a more unambiguous identification of an object as 'hostile'. In other words, the design needs to allow the operator to query all technical information from the system that is relevant to rationalize the targeting process.

C. Elements of Normative Assistance

As indicated before, the RoEs provide underlying normative framework considered here. In designing a technical system for normative assistance, the possibilities and limitation of implementing legal principles need to be addressed. The following discussion was inspired by comments of the German lawyer Tassilo Singer [18].

1. According to current understanding of the legal state of the art, certain legal principles formulated in the RoEs cannot be translated into an algorithmic form or in such a way that they can make human-type, evaluative decisions (for example, moral or ethical opinions, weighing decisions). An example in the context of the international humanitarian law is the principle of 'proportionality', i.e. prohibition of excess. It will be part of the work with E-AID to identify those legal principles.
2. If it is possible to translate a legal principle, such as "An attack may not be directed against a civilian population. A distinction must be made between civilians and combatants," into an algorithm or an AI-model, certain criteria, threshold values or parameters are decisive prerequisites for the legally compliant behaviour of an AI-controlled system, i.e. the effective restrictability (with probability bordering on certainty) of the actual behaviour. At least on the tactical level for mission execution, a large portion of rule type RoEs should be translatable in algorithmic form. The thresholds mentioned are already present, at least verbally, in military documents such as the procedures of military reporting, and are even assigned to numerical values: 'possibly' (<30%), 'likely' (30-90%), and 'probable' (>90%).
3. This leads to a key thesis: Provided a legal principle can be translated into an AI-model with quantitative criteria being integrated in the previous sense, a legally compliant implementation of legal principles can be achieved through technical system design, supplemented by sandboxing, testing, auditing.
    a. If this is the case, a control mechanism needs to be integrated, either additionally in the AI model or as part of the training, for example, a definable "no fly zone". A threshold value in connection with a rule could be: Only from a certain probability on may a target be classified as a combatant. Below this threshold, the system cannot automatically attack. Nevertheless,

the use of such parameterizations is limited since it might imply attacking two civilians in 100 attacks is acceptable, for example (see the discussion of non-translatable legal statements).

b. Further elements are appropriate safety and security as well as anti-tampering systems that automatically block all automated engagement of effectors in the event of any tampering with the system control and only allow them to be unlocked using special keys, for example.

c. In this way, particular translatable legal principles with additional parameters that enable a certain "fine-tuning," i.e. an individual or subsequent application-related adjustment and the consideration of special reservations, could make a legally compliant autonomous system possible in this respect.

d. In order to achieve operational readiness, test simulations, comprehensive sandboxing with digital twins, real-life tests and objective, third-party audits (possibly by certification authorities) would be necessary in addition to the fulfilment of information and IT security standards yet to be defined. In addition, appropriate operator training and familiarization with the system and its capabilities (trust by understanding the system) is inevitable.

4. Overall, however, it should be pointed out that for the development of a comprehensively legally compliant system, the combination of several individual solutions (legal rates + parameters / thresholds) and the systemic combinability must be given and, thus, building a certain "box" around artificially intelligent automation for weapon engagement.

5. A hurdle that cannot be crossed from today's point of view will remain in the area of decisions on proper values, as a technological solution for support is currently not apparent.

IV. Transparent Criteria Development

In consequence, systems engineering for designing responsible assistance by cognitive and volitive machines, which technically support ethically and legally compliant behavior, has to fulfill four major requirements:

1. situational awareness to enable responsible action;
2. identification of responsible options to act;
3. comprehensive plausibility of propositions;
4. resilience against failure or hostile intervention.

These are basic for ensuring responsible decisions before, during, and after the mission in order to successfully achieve clearly defined ends and intermediate purposes in a given operating theatre. To what extent collateral effects can be tolerated, is part of this decision-making.

A. Realization in the Life Cycle

Figure 4 illustrates how these requirements could be met in the research, development, procurement, deployment, and use phases of assistance systems for responsible action.

1. Transparent criteria development must accompany military capability development from the very outset. Philosophers, lawyers, and the military pastoral care bring in basic insights. Legal standards that apply to defense research, development, and procurement are indispensable. Finally yet importantly, the experience of commanders and soldiers must be taken into account. Analogous to industrial quality assurance and certification processes, these considerations support responsible action not only in battle, but also on all levels of responsibility well before.

2. Evolutionary innovation, on the one hand, replaces outdated technology while letting procedures and processes largely unchanged, whereas disruptive innovation, on the other hand, opens up fundamentally new applications, which require both conceptual and organizational changes. Ultimately, the innovative potential of defense digitization is only realizable if it takes into account the mind set and *esprit de corps* of the armed forces and, last but not least, the maxims of military licensing, certifying, and qualification bodies.

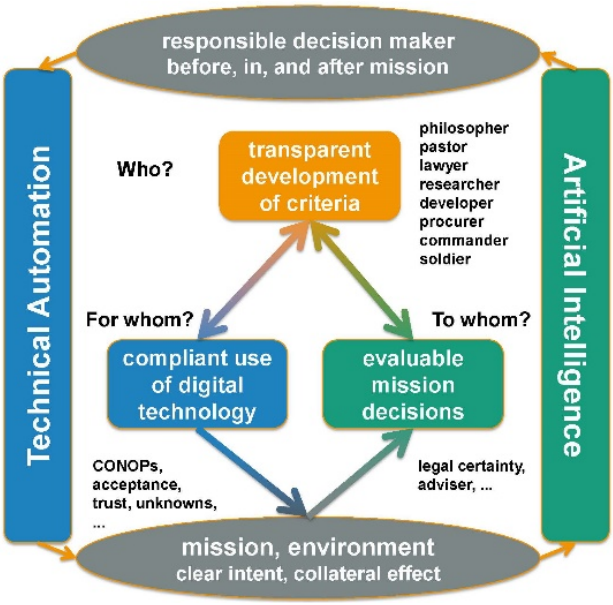

Figure 4: Transparent criteria development for in research, development, procurement, and use. © Fraunhofer FKIE.

3. Mission-relevant decisions can be evaluated and correspond to the mission-specific RoE that define the framework for action in a legally binding manner. RoE, thus, have to have a direct impact on the technical systems design, but can be so complex that computer-aided "synthetic legal advisors" are indispensable for identifying RoE-compliant options for action in battle. This is particularly true in the spatially delimited and accelerated operations "at machine speed", which FCAS is designed for, where ethically relevant knowledge itself must be made electronically accessible.

4. In a first step, RoE assistants would be helpful that at least mechanizes the simple part of the rules, accompanied with the capability to query underlying information in order to validate the underlying rationales. In this phase, the complex part can still remain with the human controller. Over time, more and more aspects might be taken over by the system, alongside

with growing operator trust by understanding the capabilities of the novel AI-enabled supporting functions and, more generally, trust in responsible systems design.

B. Remarks on Soldierly Virtues

Carl von Clausewitz (1780–1831), the Prussian general and military theorist who stressed the moral, psychological, and political aspects of war, speaks of "the courage of responsibility, be it before the judgment seat of some external power or the inner one, namely conscience [19, I.3, p. 71]." It is a "disposition of the mind," which he equates with "courage against personal danger". The Clausewitzian philosophy is rooted in notion of 'virtues', habits of 'good' behaviour, which are acquired by some sort of 'supervised' moral 'training' over time and appear under different names in most cultures. The so-called four 'cardinal virtues', prudence, justice, bravery, temperance, fundamental of Western ethics, are examples with a potential of wider consent.

The willingness to "accept wounds in the struggle for the realisation of the good" [20, p. 118] characterizes bravery as a particularly soldierly virtue, which is closely related to the Clausewitzian "courage of responsibility" previously mentioned. The virtue of justice, on the other hand, is to be seen as the perfection of prudence, which perceives reality, such as a military situation, as it actually is. Bravery can only indirectly complement justice, since it is not directly aiming at the 'good', but rather at the obstacles that arise in the realisation of the 'good'. "Only the prudent can be brave. Bravery without prudence is not bravery [20, p. 119]." The proper meaning on 'temperance', which is also an essential element of the soldierly ethos, "makes a unified whole out of disparate parts", remarks the philosopher Josef Pieper (1904-1997). "This is the first and proper sense of the Latin verb *temperare*; and only on the basis of this broader meaning can *temperare* – negatively – mean 'to restrain'. […] 'Temperance' means: to realize order in oneself [20, p. 140-141]."

Beyond mere 'functioning', but in the sense of acquiring soldierly virtues that are adapted to the requirements of the digital age in combat, E-AID may serve as a simulator for training the responsible execution of the targeting cycles of future combat air systems such as FCAS.

C. Hippocratic Oath – An Analogy?

Only if based on an Image of Man that is compatible with the responsible use of technology along the lines previously discussed, can digital assistance systems support morally acceptable decisions. "It is the responsibility of our generation, possibly the last to look back to a pre-digital age 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 the German political theologian Ellen Ueberschär (b.1967) [21].

Is a task assigned to the military pastoral care to pronounce the necessity of such an Image of Man, especially in the military service, and to provider educational offers towards a realization of this conception. It would be worth considering in this context, whether the swearing-in ceremony, which was considered indispensable when the *Bundeswehr* was founded, shouldn't be reviewed with a fresh eye in the spirit of the Hippocratic Oath, generally regarded as a symbol of another professional ethos that is committed to responsibility for life and death. 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 of 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 therefore for this obligation as well." [1, p. 181]

V. Instead of a Summay

Only alert Natural Intelligence (NI) is able to assess plausibility, develop understanding, and ensure control. "The uncontrolled pleasure in functioning, which today is almost synonymous with resignation to 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 observed in the 1950's [1, p. 180]. These words ring true not only for shaping the soldierly ethos in the digital age. There is a more general need for a new enlightenment in dealing with AI maturely, ethically, and intelligently, i.e., "man's release from his self-imposed immaturity. *Sapere aude*—Have the courage to use your own intellect!" [22] Anthropocentrism in this sense underlines the ethical and legal dimensions of artificially intelligent automation, which characterize the use of AI in defence systems.

Since we feel encouraged to assume that a broader consent among the information fusion community might be achieved, we are closing with some recommendations that address certain blind spots, at least according to the observations of the author.

1. Digital ethics and a corresponding ethos and morality should be built up systematically for responsibly using artificially intelligent automation in the military domains. In particular, such skills enable commanders "to assess the potential and impact of digital technologies and to manage and to lead in a digitized environment," as an official German document states [23]. In particular, leadership philosophies and personality development instruments should encourage such competences.

2. In addition to the operational benefit of artificially intelligent automation in closing capability gaps, expanding the range of capabilities, and developing corresponding concepts, operational procedures, and other organizational measures, ethical and legal compliance needs to be achieved. Only then, cognitive and volitive assistance will become acceptable before the conscience of the individual commanders, 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. Defence projects should be accompanied from their very beginning by comprehensive analyses of technical controllability and personal accountability in a visible, transparent, and verifiable manner. Otherwise, the paradigm shifts and large material efforts associated with artificially intelligent automation would hardly be politically, societally, and financially enforceable. Of course, there will be more and less problematic projects, implying that an exemplary approach according to these lines would be appropriate.

"Firmly confident in his better inner knowledge, the military leader must stand like the rock where the wave breaks," observed Carl von Clausewitz [19, I.6, 96]. Artificially intelligent automation therefore requires the ethos of digitally educated commanders and staffs. They do not

need to know how to design and program AI-based defence systems, but to assess their strengths and weaknesses, risks, and opportunities. The associated digital morality and competence is teachable. It addresses a key question of the soldierly ethos, which is aggravated by artificially intelligent automation but not fundamentally new.

VI. Acknowledgments

We wish to thank the scientists and engineers that are engaged in the working group on *The Responsible Use of New Technologies in a Future Combat Air System*. To be mentioned are in particular the valuable contributions of Florian Keisinger, Bernhard Krach, and Christoph Vernaleken, Airbus Defence and Space, Bert van Heukelom and Martin Lederer, Data Machine Intelligence Solutions, Torsten Fiolka, Felix Govaers, Michael Schleiss, and Martin Ulmke, Fraunhofer FKIE. We also learned a lot from Eleri Lillemae, Estonian Military Academy, Kairi Talves, Estonian Ministry of Defence, and Dierk Spreen, Berlin School of Economics and Law, with whom we discussed the related problem of the ethically and legally aligned use of integrated Modular Unmanned Ground Systems. Moreover, we gained insights from Tassilo Singer, Atos, and the ongoing "von Kármán Horizon Scanning on Artificial Intelligence", organized by the NATO Science & Technology Board and chaired by Maurus Tacke, Fraunhofer IOSB, and Michael Wunder, Fraunhofer FKIE, that have inspired aspects of these reflections.

VII. References

[1] W. von Baudissin, *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr* [Soldier for Peace. Drafts for a Contemporary *Bundeswehr*]. München: Pieper, 1969.

[2] "I ask you: Do you want total war? If necessary, do you want a war more total and radical thananything that we can even imagine today?" In Sportpalast speech of Nazi propaganda ministerJoseph Goebbels (1897-1945) on February 18, 1943.

[3] *The Responsible Use of New Technologies in a Future Combat Air System.* Online: www.fcas-forum.eu.

[4] W. Koch, "On Ethically Aligned Information Fusion for Defence and Security Systems," 2020 IEEE 23rd International Conference on Information Fusion (FUSION), 2020, pp. 1-8, doi: 10.23919/FUSION45008.2020.9190233.

[5] W. Koch, *On Digital Ethics for Artificial Intelligence and Information Fusion in theDefense Domain*, IEEE Aerospace and Electronic Systems Magazine, vol. 36, no. 7, pp. 94-111, July 1,2021, doi: 10.1109/MAES.2021.3066841.

[6] W. Koch, "AI-based Defense Systems – How to Design them Responsibly?," German-Israeli Tech-Policy Dialog, Heinrich Böll Stiftung Tel Aviv, December 30, 2021. Online: https://il.boell.org/en/2021/12/24/ai-based-defense-systems-how-design-them-responsibly.

[7] W. Koch, "What does Artificial Intelligence offer to the Air C2 domain?," NATO Open Perspectives Exchange Network (OPEN) Publication, NATO Allied Command Transformation, July 2022.

[8] S. Spiekermann, "From value-lists to value-based engineering with IEEE 7000™," 2021 IEEE International Symposium on Technology and Society (ISTAS), 2021, pp. 1-6, doi: 10.1109/ISTAS52410.2021.9629134.

[9] Lem anticipated that the metaphor 'instinct control' seems to appropriate for what we call today "autonomous driving", for example. „The wasp probably possesses a sufficient number of nerve cells that it could just as well steer a truck [...] or control a transcontinental missile." In: St. Lem, *Waffensysteme des 21. Jahrhunderts* [Weapon Systems of the 21st Century or the Upside Down Evolution]. Frankfurt am Main: Suhrkamp, 1983, p. 44.

[10] V. Stooß, M. Ulmke and F. Govaers, "Adiabatic Quantum Computing for Solving the Weapon Target Assignment Problem," 2021 IEEE 24th International Conference on Information Fusion (FUSION), 2021, pp. 1-6, doi: 10.23919/FUSION49465.2021.9626902.

[11] F. Govaers, V. Stooß and M. Ulmke, "Adiabatic Quantum Computing for Solving the Multi-Target Data Association Problem," 2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), 2021, pp. 1-7, doi: 10.1109/MFI52462.2021.9591187.

[12] *Konzeption der Bundeswehr* [Concept of the Bundeswehr]. Berlin: MoD, 2018. Online: https://www.bmvg.de/resource/blob/26544/9ceddf6df2f48ca87aa0e3ce2826348d/20180731-konzep-tion-der-bundeswehr-data.pdf

[13] *Basic Law for the Federal Republic of Germany*, Bonn, May 23, 1949. Online: https://www.gesetze-im-internet.de/englisch_gg/

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

[15] Annex III in: Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Geneva, Swizerland, 13 December 2019. Online: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/343/64/PDF/G1934364.pdf?OpenElement

[16] *Erster Bericht zur Digitalen Transformation* [First Report on Digital Transformation]. Berlin:MoD, 10/2019. Online: https://www.bmvg.de/resource/blob/143248/7add8013a0617d0c6a8f4ff969dc0184/20191029-down-load-erster-digitalbericht-data.pdf

[17] Aspects discussed in this context are: (1) Context Control: controlling the space, duration, time and conditions, (2) Understanding the System: functioning, capabilities and limitations in given operational circumstances, (3) Understanding the Environment: situational awareness and understanding of the environment, proper training, (4) Predictability and Reliability: knowledge of the consequences of use and reliability as the likelihood of failure, both in realistic operational environments against adaptive adversaries, (5) Human Supervision and Ability to Intervene, (6) Accountability: certain standard of authority and accountability framework of human operators, teammates and commanders, (7) Ethics and Human Dignity: preserve human agency and uphold moral responsibility in decisions to use force. See [15].

[18] Personal correspondance. See also: T. Singer, *Dehumanisierung der Kriegführung. Herausforderungen für das Völkerrecht und die Frage nach der Notwendigkeit menschlicher Kontrolle* [Dehumanization of Warfare]. Berlin, Heidelberg: Springer, 2019.

[19] C. von Clausewitz, *Vom Kriege* [On War]. 11th ed. Hamburg, Germany: Nikol, 2018, I.6, p. 96.

[20] J. Pieper, *Werkausgabe Letzter Hand* [Last Hand Ed.], vol. IV, *Schriften zur Philosophischen Anthropologie und Ethik: Das Menschenbild der Tugendlehre* [Writings on Philosophical Anthropology and Ethics: The Image of Man in the Doctrine of Virtue], Hamburg, Germany: Felix Meiner, 1996.

[21] E. Ueberschär, *Von Friedensethik, politischen Dilemmata und menschlicher Würde – eine Skizzeaus der Perspektive theologischer Ethik* [Of Peace Ethics, Political Dilemmas and Human Dignity –a Sketch from the Perspective of Theological Ethics], Opening Speech at the first meeting of theworking group *The Responsible Use of New Technologies in a Future Combat Air System*. BadAibling, September 27, 2019. Online: https://www.fcas-forum.eu/publications/Skizze-zur-theologis-chen-Ethik-Ueberschaer.pdf..

[22] I. Kant, *An Answer to the Question: What is Enlightenment?* (1784). Online: http://donelan.faculty.writing.ucsb.edu/enlight.html .

[23] *Umsetzungsstrategie ‚Digitale Bundeswehr'* [Implementation Strategy 'Digital Bundeswehr']. Berlin, Germany: BMVg, Jun 14, 2019, 209, 8. Online: https://www.bmvg.de/de/themen/ruestung/digitalisierung/umsetzungsstrategie-digitale-bundeswehr