On Digital Ethics for Artificial Intelligence and Information Fusion in the Defense Domain

Wolfgang Koch, *Fellow, IEEE*

***Abstract*—For knowledge itself is power... Francis Bacon's statement on achieving power as the meaning of all knowledge marks the beginning of the modern project. At the latest since the advent of AI in the defense domain, however, technology meant for the benefit of humanity may turn *against* humanity. This specific type of instrumental knowledge makes the modern crisis as visible as in spotlight. Ethical knowledge of man and his nature must complement Bacon's knowledge. There is an 'ecology' of Man: he does not make himself; he is responsible for himself and others. How can the data fusion community *technically* support responsible use of the power we are harvesting from AI? To argue more specifically, we consider documents of the German Bundeswehr, founded in the 1950s when the term AI was coined. Since these Armed Forces have learnt lessons from 'total war' and tyranny, they seem conceptually prepared for mastering the digital challenge. There exist parallels to ongoing IEEE P7000 standardization activities.**

***Index Terms*—digital ethics, artificial intelligence, data fusion, autonomy, ethically aligned design, decision support, responsibility.**

I. Engineering Aspects of Digital Defence Ethics

'Intelligence' and 'autonomy' are omnipresent in the biosphere. Before any scientific reflection or technical realization, all living creatures fuse sensory impressions with information they have learned themselves and received from other creatures. In this way, they create a model of their environment, the basis to act appropriately. In the technosphere, digitization provides tools that powerfully enhance the perceptive mind and active will of those who consciously perceive and responsibly act. The concepts of mind, will, and responsibility bring into view fundamental ideas of human beings as persons that are 'somebody' and not 'something' that imply ethical dimensions.

In the Adenauer Era, the founders of the German Armed Forces, the *Bundeswehr*, wanted to establish structures that prevent barbarism as experienced in WW II. At the same time when the word 'Artificial Intelligence' was coined, its architects have anticipated hyperwar: "The scientificization and mechanization of the military craft will lead to the dissolution of spatial boundaries and acceleration of military action". General Wolf von Baudissin clearly predicted an important consequence of this development: "The most highly mechanized combat requires that responsibility is seen and borne at many lower levels. Therefore, everything must be done to put people in situations that make them aware of their responsibility and make them experience the consequences of their actions and omissions."[1]

What type of digital technology is required to meet these demands? How to design artificially intelligent and technically autonomous systems that enable, facilitate, and encourage their responsible use? Answers to these question are all the more urgent because the "validity of human dignity as an ineluctable basic ethical assumption is by no means unquestioned. [...] In various fields we are confronted with developments in which the boundaries between 'person' and 'thing' are blurred."[2] Ultimately, Artificial Intelligence may become dangerous for humanity because of 'natural stupidity', i.e. the refusal or mental inertia to be a human being in the full sense.

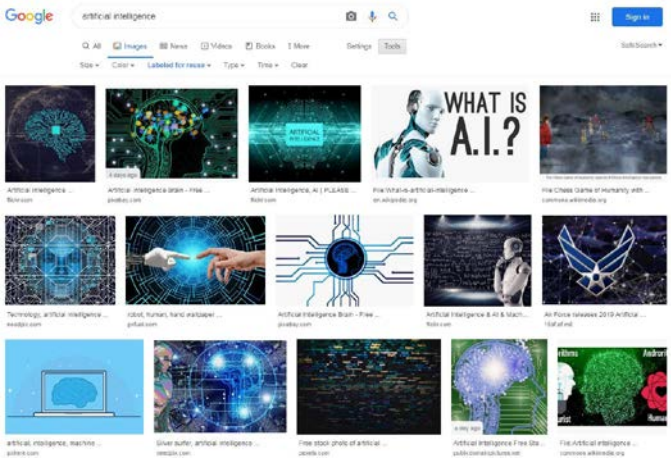

Fig. 1: Result of a Google image search with the keyword *artificial intelligence*: a psychogram of modern networked humanity? (© GIDS).

There is a psychological dimension as well. A Google search for images illustrating AI generates a psychogram of modern humanity with its hopes and fears. In black, turquoise, blue or white, intelligently looking alien beings are rising from circuits and data, cool and superior, quite different from us, yet "in the mage of Man". Something like Michelangelo's "Creation" symbolizes emerging intelligence, robots pose as Rodin's "Thinker". This collective unconscious influences researchers and developers, but also procurement and decision-making. Fig. 1 also reveals an underlying fear: "Mark my words – A.I. is far more dangerous than nukes."[3]

Artificially intelligent and technically autonomous systems are already relevant for all operational dimensions.[4] One could speak of assisted perception and action in the increasingly complex technosphere of military operations. For this reason, the digital transformation in the defense domain is the key to "information, command & control, and engagement superiority" as well as to "improving their ability to assert and respond".[5] Since potential adversaries use or will use digital technologies,

W. Koch is with the University of Bonn and the Fraunhofer Institute FKIE, Fraunhoferstrasse 20, 53343 Wachtberg, Germany (e-mail: w.koch@ieee.org).

dissolution of spatial boundaries and acceleration will characterize digitized combat where cause-and-effect chains run automatically. A European example is Future Combat Air System (FCAS), a system-of-systems consisting of manned and unmanned flying platforms to protect European airspace.[6]

If we accept that armed forces must be able "to fight at machine speed", digitization in the defense domain cannot be limited to surveillance and engagement, but must equally guarantee its responsible use. "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 von Baudissin. "Without the commitment to the moral realms, the soldier threatens to become a mere functionary of violence and a manager."[7] This is all the more valid in sensor-to-shooter loops and multi-domain command & control. The hope for deceleration and limitation by arms control policy to counter military destabilization such as 'flash crashes' on financial markets is honorable,[8] but seems unrealistic in view of the almost unrestricted proliferation of digital technologies and their dual use.

For the sake of clarity, the *Bundeswehr* avoids the term 'autonomy', which obscures gradual differences. 'Automation' seems to be more appropriate as it also includes 'full automation' after stages of partial automation. Absurd is the notion of a technical 'decision of will'. In this sense, also 'artificial intelligence' is a problematic term that is too well established to be avoided.

To ensure that the *Bundeswehr* fulfils their tasks, digitization is expanding its capability profile on two levels:

1. It provides tools for perceiving 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, without interruption or media disruption."[9]
2. It supports the "targeted deployment of forces and means according to space, time and information. […] Characteristic features of military leadership are the personal responsibility of decision-makers and the implementation of their will in every situation."[10]

"Without the commitment to the moral realms, the soldier threatens to become a mere functionary of violence and a manager. Soldiers were then degraded to weapons without human cohesion and conscience; with them every act of violence becomes possible" (General Wolf von Baudissin). © NATO

In this context, it seem to be a duty of defense scientists to clarify the ethical problems of AI and automation, while avoiding 'moralizing'. Let us call thinking about the right decisions in using digital technologies in defense *digital ethics*. Required is an Image of Man that makes 'mind', 'will', and 'responsibility' conceptually possible. *Digital ethos* addresses the attitude of decision-makers. "The more momentous the decisions and actions of individual soldiers are, the more their ethos must be determined by responsibility. If this is only seen from a functional and legal point of view, armed forces become a danger", observes von Bausissin.[11] "In the jungle situation of borderless wars, 'soldiers only' are no longer fit for war in the long run."[12] *Digital morality*, finally, comprises concrete guidelines for dealing with AI and automation, not only in the battlefield, but also in research, development, and procurement.

In this paper, we present selected aspects of sensor data fusion, AI, and resources management from a systems engineering perspective. If were able to solve the ethical problems here, new paths will open up for the civil use of digital technologies as well. After examples for illustration, technical controllability and personal responsibility appear as basic principles of military system design. A discussion of challenges and research questions to be answered during their technically implementation, prepares considerations on how accountability can be supported by digital system technology. This paper leads to the recommendation to explicitly implementing ethical aspects in the various stages strategic planning, defense R&D, procurement, and use from the very beginning.

"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," observes Ellen Ueberschär, a prominent political thinker in Germany.[13]

II. Military Use of Digital Technologies: Examples

Scientific research on 'electronic brains' already began in the 1940s and 1950s with the invention of the computer and the discovery that highly interconnected nerve cells exchange electrical impulses with each other. Norbert Wiener's cybernetics for controlling electrical networks, Claude Shannon's information theory, and Alan Turing's calculability theory have created the intellectual basis. In 1956, the *Dartmouth Summer Research Project on Artificial Intelligence* gave rise to the term 'Artificial Intelligence'. The development of Data Fusion profited from Ronald Reagan's Strategic Defense Initiative SDI. In George Orwell's very year 1984, an advisory board to the US Department of Defense coined the technical term 'Data Fusion'.

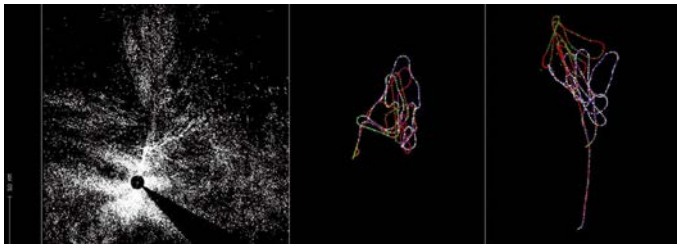

Fig. 2: Trajectories of dogfighting fighter aircraft extracted from poor radar data. © FKIE

As an early example, Fig. 2 illustrates the efficiency of so-called Multiple Hypothesis Tracking (MHT). The left side shows data provided by a long-range radar, accumulated over one hour. Without algorithmic support, it would be impossible to extract useful information from the data stream about two pairs of fighters training air combat. From the measurements, which are repeatedly missing, even consecutively, and contaminated with many false, unwanted, or unresolved radar echoes, the MHT produces precise target trajectories (on the right).[14]

The Airborne Early Warning and Control System (AWACS) or Alliance Ground Surveillance (AGS) are 'machines' that massively generate such information. On this basis, Big Track Data Analysis produces 'recognized' situation pictures for threat evaluation, guidance, and weapon assignment. Contextual information on topographical conditions or spatio-temporal rules, such as a plan to be followed, are essential, especially according to the concept *Führen mit Auftrag*, leading by mission.[15] If fused with sensor data, context information provides important insights. Fusion technology is thus suitable to integrate also formalized moral rules into reconnaissance or combat missions.

Studies are investigating how coordinated multiple ground robots and multiple sensor drones can support troops by providing comprehensive situation pictures locating, e.g., camouflaged snipers by fusing spatially distributed acoustic and multispectral sensor data, classifying the calibers of ammunition fired, and identifying radio communication links of enemy forces.[16]

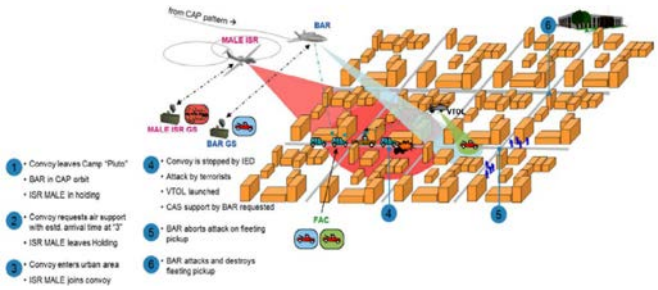

Fig. 3. Coordinated drone deployment enables the FOC to defend against threats in accordance with the RoE. © FKIE

Armed drones may close capability gaps. Let us consider a convoy in an urban environment that is stopped by an Improvised Explosive Device and attacked. Coordinate drones make it possible to estimate the expected collateral damage and to provide the Forward Air Controller (FAC) with a comprehensive, real situation picture, the basis of responsible and potentially lethal decisions. All engagement decisions follow the Rules of Engagement (RoE), which are to be taken into account right down to the information technology design.[17]

In conclusion, artificially intelligent automation aims at 'cognitive machines' that assist their users. Through them, military actors acquire knowledge about threats, uninvolved parties, and options for action in the various operational areas. At the same time, risks for own forces are minimized. Such assistants help to master complex tasks more adequately, to balance human subjectivities, and to protect non-combatants. The physical presence of humans is thus increasingly dispensable in dangerous situations. Cognitive assistance is of central importance to

- evaluate imperfect and incomplete mass data,
- to fuse context knowledge with current data streams,
- to fuse complementary and heterogeneous sources,
- to estimate the plausibility of the information content,
- to enable Manned-Unmanned Teaming and action,
- to guarantee ethical, legal, and societal compliance.

Decentralization and automation implies vulnerability, i.e. the necessity to protect own systems against cyber-attacks and attacks from the electromagnetic spectrum and to develop strategies to attack enemy systems in this way.

III. Controllability: Basis of Military System Design

Artificially intelligent automation enables military decision-makers to do what they can do only as persons, i.e. to perceive with real intelligence and to make decisions with responsible autonomy. Are we facing fundamentally new challenges? Certainly not. Technology has repeatedly increased perception and the range of action. The difference of digitization to earlier revolutions are more on a quantitative than on a qualitative level.

The core document of the *Bundeswehr* explicitly opens up to a discussion on a philosophical level when it speaks of the principle of *Innere Führung* as "the underlying philosophy of leadership valid for the German soldiers."[18] The term *Innere Führung* is not easily translated. Von Baudissin defined it: "*Innere Führung* is military leadership with special consideration of the individual and social aspects of the person. Its overall goal is to reconcile the functional conditions of operational armed forces with the principles of a democratic constitutional state."[19]

Therefore, defense digitization poses a timeless question: How to guarantee comprehensively ethical, legal, and social compliance; how to decide 'well' according to what is recognized as 'true'? In engineering, this breaks down to two questions:

1. How to design cognitive tools that we mentally and emotionally always are 'masters' of them?
2. Which design principle facilitates the responsible use of artificially intelligent automates systems?

As illustrated by Fig. 4, AI and automation assist the perceiving minds of "personally responsible" decision makers in understanding complex situations and support "the enforcement of their will in every situation" in terms of appropriate and responsive action.[20]

Fig. 4: Algorithmic assistance for the perceptive mind and active will of responsible decision-makers. © FKIE

AI generates contributions to situation pictures. Decisive is the question of 'what' is to be recognized. 'Detection' informs about the existence of relevant objects and phenomena, 'classification' about their properties, i.e. their essence. Important

building blocks are inferred object interrelations. Finally, situation pictures indicate decision relevance, such as threat levels and the state of own forces, platforms, or weapons. The situation picture as well as statements about its limitations and gaps and the actual situation must mutually correspond. This implies a concept of truth: "Truth consists in the equivalence between the situation picture and the situation."[21] We may distinguish between a logical truth of the situation picture and its 'ergonomic truth', in that it corresponds to the tasks, roles, and abilities of the decision maker.[22]

Automation translates the intentions of the decision makers into complex cause-effect chains to manage, e.g., multifunctional sensors, mobile platforms, and effectors. The question 'why'to achieve an effect is crucial for algorithm design. Aristote's model of causality is helpful, which distinguishes four ways of answering to why-questions. The goals correspond to the *final cause*, usually specified by performance parameters. The *causa efficiens* indicates which concrete algorithms are used to achieve them. The *formal cause* answers the question, according according to which rules this happens. Finally, the *material cause* indicates which means are to be used with their respective properties. The close link that Aristotle sees between the formal and the final cause corresponds to the principle that military goals are to be achieved according to the Rules of Engagement. Mission preparation corresponds to the link between material and formal cause. Finally, impact assessment determines the extent to which the final cause of the military action has actually been achieved and is the prerequisite for further action.

In general, we distinguish data-driven from model-based algorithms. The first family, for which Deep Learning is exemplary, corresponds to intuitive sensory perception – *What do I see*? The second family, in the sense of Bayesian reasoning, enables rational causal action – *What do I do?*

In the algorithmically driven information processing circuit, sketched in Fig. 5, we distinguish five levels of perception. The first two levels, determined by received signals and signal processing, are summarized as *data levels* and are usually hidden from military decision makers. For them, the three *information levels* above are more relevant. They refer to the individual objects, secondly to the situation with information about the interaction of the objects, and thirdly to the mission. It describes a situation and the decision maker who wants to act in it.

Fig. 5: Levels of perception for Command & Control, ISR, engagement, and impact assessment (© FKIE).

IV. On Responsibly Guaranteed Controllability

A challenge for the consistent and responsible controllability of defense digitization is firstly the ever-decreasing time available for human-involved decision making. A further problem is the limited explainability and deceivability of algorithmically generated information and automation.

From an abstract point of view, neural networks assign an input to an output. The output describes what the input should 'mean' for the user. The special feature of such functions is their extremely large number of degrees of freedom, tunable numerical values. In a 'training phase' they are adjusted by 'telling' the neural network what the input 'means', e.g. 'understood' photos. This 'labeling' requires human understanding. If training has been 'long enough', the network is offered an arbitrary input and the output is considered the recognized 'what'. Neural networks are thus function approximators. Whoever calls massive offering of interpolation points 'learning', awakens erroneous associations in non-specialists.

As it turns out, however, only a few pixels in the input need to be changed in a specific way to completely mislead even a well-trained network: The neural network, deceived by "poisonous noise", may 'recognize' a panda bear, which appears unchanged to humans, as a gibbon monkey and 'feels' certain.[23] The military relevance of this discovery is obvious. Attack systems against AI systems have already been developed, own AI systems are to be hardened against such "adversarial attacks". A situation occurs as in Electronic Warfare.

In addition, for appropriate training of data-driven algorithms no sufficient amount of representative training data are available in most military applications. Moreover, neural networks are 'black boxes'. Nobody knows how they achieve their results. Furthermore, context knowledge – fundamental to every military mission – can only be learned indirectly, i.e. from the data itself. In short: Neural Networks are 'greedy', 'brittle', and 'opaque', i.e. they are always the 'second-best' solution. At least for critical targeting functions, meaningful human control is required.

Model-based algorithms, on the other hand, allow logical reasoning also in case of uncertainty, uncover probable cause-effect chains, can be developed systematically, and explicitly allow the integration of context and expert knowledge. Sometimes, however, the required models are not available or are too complex to be used efficiently. An unsolved problem of current research is the combination of data-driven and model-based algorithms resulting in Explainable Artificial Intelligence (XAI).

In view of these problems, the following aspects need to be addressed by appropriate system design:

1. Responsible use of technology requires consistent controllability. In some applications, occasional malfunction of AI and automation may have no consequences. In military use, however, rigorous safety requirements must to be guaranteed with all legal consequences. The use of technically uncontrollable technology is immoral *per se*.
2. The notion of 'meaningful human control', on the other hand, 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 of the German *Luftwaffe* should be reconsidered.[24] More fundamental is "accountable responsibility" to be discussed below. The use of fully automated effectors may well be justifiable, even necessary, in certain situations.

3. Certification and qualification are key issues. Robust systems will comprise both, data-driven and model-based algorithms, where data-driven algorithms could be 'contained' 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 well as explainability for critical components are essential prerequisites. Finally yet importantly, compliance to a code of conduct must be guaranteed *by design*.
4. Sensor and context data never meet ideal expectations. They are always imperfect, inaccurate, ambiguous, unresolved, corrupted or deceptive, difficult to formalize, or partly contradictory. Statistical models, however, enable responsible action even on an imperfect data basis. In many cases, reliable situation pictures can be inferred from them in a more precise, complete and faster way than humans could ever hope to obtain. Nevertheless, these methods have limitations, which must not only be made aware of, but also be interpreted.
5. Data integrity is fundamental to any use of AI-based systems: Are valid sensor and context data available at all? Are they produced reliably and do the unavoidable deficits correspond to the statistical 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 the military context, enemies may take over sensors or subsystems, which then produce deceptive data. Mature AI comprises detection of such deficits, which is the basis for making own systems resistant to interference and deception or to deceive enemy systems.

Artificially intelligent 'self-criticism' of technical systems requires naturally intelligent critical capabilities of military decision makers towards AI. Otherwise, there is a danger of voluntary subordination and uncritical acceptance of machine offers, of the mental refusal to actually bear responsibility, of blind trust. AI-based systems must therefore train the alertness of their users and teach them how the AI offers were developed. AI must not stupify its users. Only alert natural intelligence is able to assess plausibility, to actually develop understanding, and to regain control if digitization fails. Many research question in systems design rise from these considerations.

"All thinking is art," observes Carl von Clausewitz, the Prussian general and military theorist who stressed the moral, psychological, and political aspects of war. "Where the logician draws the line, where the prefixes end, there art begins."[25] Thus, digitization requires the ethos of digitally educated decision-makers who do not need to know how to design AI algorithms, but are able to assess their strengths and weaknesses, risks and opportunities. The associated digital morality is communicable. Obviously, this addresses a more fundamental question, which is aggravated by digitization, but not fundamentally new. "Firmly confident in his better inner knowledge, the military leader must stand like the rock where the wave breaks."[26] The associated 'digital morality' is fundamentally teachable. Obviously, this addresses a fundamental question of the soldierly ethos, which is aggravated by the digitization, but is not fundamentally new.

V. Responsibility Military Systems Engineering

The notion of responsible use of digital technologies realizes a connection between conscious cognition and volition leading to action in the real world and automatically running programs and processes in the world of algorithms that are dealing with sensors, actuators, and effectors.

As a word 'responsibility' is rooted in the language at courts of justice and means being called upon to *respond* to questions about one's actions by a judge, a primal situation of human existence as a person. This concept has far-reaching implications: What action or omission is owed? Why, under which circumstances, and according to which law is there an obligation to respond? What form of accountability is expected? Who is called to accuse, who to judge? According to which standards do we speak of acquittal with 'praise' or conviction with 'punishment'? There exists a vast literature, influenced by most diverse cultural backgrounds. There seems to be a broader consensus on the following aspects.

1. To speak of responsibility is only reasonable if it is assumed voluntarily. Responsibility thus presupposes 'freedom' and the idea of Man as a free person.
2. The concept of free will as the decisive cause of actions implies the idea of accountability, which is legally relevant and an essential criterion in International Law.
3. Responsibility also implies the ability and willingness to act 'well' even in case of absent or contradicting rules. Casuistry, formalization of human action, seems impossible.
4. The will, responsible in freedom, is not absolute, but depends on the understanding mind. The 'true' and the 'good' thus form the intellectual basis of responsible action.

Von Clausewitz speaks of "the courage of responsibility, be it before the judgment seat of some external power or the inner one, namely conscience". It is a "disposition of the mind," which he equates with "courage against personal danger".[27]

In the current debate, responsibility is more fundamental than meaningful human control, because even the use of technically autonomous systems, which, after the decision to use them, achieve their effect without human intervention, can be justifiable under certain clearly defined conditions. Examples are the armed swarms of drones and highly reactive protection against approaching missiles. As a purely reactive measure, these examples do not involve a fully automated targeting cycle. In a certain sense, humans are involved in these examples as well, namely with the decision to switch on the system and select the parameter framework for enabling independent defense, so that meaningful human control is guaranteed in a broader sense.

Fig. 6 illustrates core elements of the concept of responsibility, insofar as it is relevant to the technical design of defense systems. It implies three persons or groups of persons and characteristic relationships between them.

1. *Who bears responsibility?* Military capability development takes place at various levels and requires responsible action in research, development, certification, and qualification of military C2, ISR, and weapon systems as well as in the preparation and execution of military operations.
2. *For whom is responsibility borne?* The relationship between a responsible person and those for whom he or she is responsible is characterized by 'care' and 'trust' and determined by prospective action and reaction. Responsibility can only be assumed for persons. Everyone is responsible for himself. "Whoever grants freedom of will to human beings, means that human beings are 'themselves' the reason for their actions and not for their actions towards others. In fact [...] in such a way that Man himself bears responsibility for his own existence."[28] Secondly, responsibility is owed in particular to own forces, combatants or civilians. In an improper sense, one could speak of a responsibility towards society or the natural habitats in the area of operations.
3. *Towards whom is responsibility assumed?* Responsibility implies the notion of an authority that is exercised by judgement and recognized by justification by the person responsible. The relationship between him and authority is retrospective in nature. Authorities are God, the personal conscience of the responsible person, the superiors, and jurisdiction exercised by persons.

Fig. 6: Core elements of the concept of responsibility and the resulting mutual relationships. © FKIE

Ultimately, it is voluntarily assumed responsibility, which shows itself in care and trust and is ready to justify itself, which keeps all human societies stable, not only combat units. Purely legal constructs, such as liability for damage caused by one's actions, are not sufficient, especially in military operations.

Apparently, only persons, not machines, can act responsibly or irresponsibly. For technical 'things' always remain a-personnel, even if they 'speak' and, due to anthropomorphic system design, a psychologically realized distinction between Man and machine may become difficult. Only people who use cognitive machines responsibly or irresponsibly for reconnaissance and action act 'good' or 'evil' by responding to moral challenges in one way or another. 'Good' technical systems encourage the morally acceptable and efficient use of them to achieve military objectives. 'Evil' systems, which might be used by an enemy, facilitate their irresponsible use.

Those who place the concept of responsibility at the center of military action must avoid representing a particular strand in the philosophy of ethics. Without denying that such distinctions may be fruitful, this is inappropriate in a military context. For certain actions are also here *per se* – "deontologically" – qualified as immoral. On the other hand, "consequentialist" weighing can be utterly moral, since responsibility is graded and corresponds to the concretely existing relationship of care and trust. For example, 'lies', such as measures of electronic or cyber warfare, are permitted towards an opponent of war, since he is "not at all in that moral relationship of trust that makes a truthful speech necessary."[29]

VI. Selected Aspects of Moral Assistance Systems

For Responsible Cognitive Systems Engineering, which technically supports moral behavior, three major requirements result from the previous considerations:

1. AI-based situational awareness to enable responsible action,
2. cognitive assistance, in to identify responsible options for action,
3. comprehensible plausibility of the proposed situation and options.

They are the basis for ensuring responsible decisions before, during and after the mission in order to achieve clearly defined goals in a given operating theatre taking into account the collateral effects that may to be tolerated or not. Fig. 7 illustrates how these requirements could be met in the development and deployment of artificially intelligent and automated assistance systems for responsible action that is appropriate to the objective and situation.

Fig. 7: Transparent criteria development for technology application and application decisions. © FKIE.

1. Transparent criteria development must accompany military capability development from the outset. Philosophers, pastors, and lawyers bring in basic insights. Legal standards that apply to defense research, development, and procurement are indispensable: "The sharpest weapon of democracy is legislation. For this reason, civil society cannot help but call on its governments to establish binding standards for cognitive weapons systems and to make the corresponding agreements under international law."[30] Finally yet importantly, the experience of commanders and soldiers must be taken into account. Analogous to industrial quality assurance processes, these considerations support responsible action not only in battle, but also at all levels of responsibility, not only in combat. These considerations correspond to the IEEE P7000 *Model Process for Addressing Ethical Concerns During System Design*, by which engineers and technologists can address ethical consideration throughout the various stages of system initiation, analysis and design.[31]

2. 2) Any technology that complies with these criteria must be integrated into military procedures and processes, e.g. in appropriately formulated Concepts of Operations (CONOPS). Evolutionary innovation replaces outdated technology while letting procedures and processes largely unchanged. Disruptive innovation, on the other hand, opens up fundamentally new applications, which require both conceptual and organizational changes. Ultimately, the innovation potential of defense digitization are only realizable if it takes into account how operational forces think about and handle technology or how licensing and qualification bodies work.
3. Mission decisions can be evaluated and correspond to the mission-specific Rules of Engagement (RoE), which have an impact deep into the information technology design. RoEs do not impose tactical specifications, but define the framework for action in an also legally binding manner. In accordance with legal, political, strategic, and operational requirements, they concretize principles of the International Law and Soft Law. Examples are discrimination (engagement only if targets are fully identified), proportionality (choice of threat-adequate effectors), care and imputability to a person. RoEs can be so complex, however, that computer-aided 'synthetic legal advisors' are indispensable for identifying RoE-compliant options for action. In the spatially delimited and accelerated hyperwar, moral knowledge itself must be made accessible through digitization.

A concrete example is provided by pre-engagement collateral damage assessment, which is made possible by the algorithms of artificial intelligence in a way that has never been achieved before. In military use, this is the basis for efficient control of the effectors, be it through precision targeting or scalable detonation effect of modern warheads.

In order to realize the potential of AI for responsible action in critical situations, decision makers must be made aware in an intuitively comprehensible manner of remaining inaccuracies, ambiguities, and aspects of the situation that have not yet been clarified. For situational awareness does not consist of algorithmically generated symbols on a screen, but rather arises in the minds of decision-makers. It is imperative that situational awareness includes information on unknown aspects. Without reliable knowledge about the limits of the available knowledge, no one can act responsibly. In addition to engineering issues, ergonomics and cognitive sciences apparently play an important role in the digital transformation of the armed forces.

A central aspect is therefore the ergonomic representation of the situation images obtained from a wide range of data sources and the limits of their informative value. Apparently, the research questions arising from this have an ethical dimension: the situation picture, despite all its abstraction, must convey to the decision maker psychological awareness of the reality of the situation shown, help him to take responsibility, "and allow him to experience the consequences of actions and omissions."[32] Decision making must not remain on a purely virtual level. Such considerations are entirely in line with political declarations of the Federal Government of Germany and all planning documents of the *Bundeswehr*.

A more recent study emphasizes the rationality of ethical judgement, for which algorithmic support is principally possible. According to it, the concrete case is at the center of ethical judgement. The assessment, which abstract and concrete, normative and descriptive, cognitive and emotional aspects are to be placed in relation to one another, is to be done in a "culture of reasoned consideration".[33] A digital assistant for "moral decision support" would have to implement and reproduce such structures of thought. From a system ergonomic point of view, it should also be considered how technical design principles could be derived from classical virtues, which appear under different names in many cutures, in such assistance systems in order to make them user-compliant in the sense of responsible judgement. The so-called four "cardinal virtues" of European ethics[34] are examples with a potential of wider consent.

Only if based on a clearly defined Image of Man that is capable of responsible use of technology, digital assistance can be designed to support morally acceptable decisions. The maintenance of such an Image of Man and the *Innere Führung* as a guiding principle towards this direction is especially a task of military pastoral care. Since the Hippocratic Oath is regarded as a fundamental formulation 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 Wolf von Baudissin, the architect of the *Innere Führung*, 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."[35]

We point out that Future Combat Air System is the first major armament project that is accompanied from the very beginning by a high-ranking working group on "Ethics & Technology Responsibility".[36] This is a unique initiative in the field of defense technology development. We expect it to provide insights into how ethical and moral considerations can be concretized and operationalized as technical design principles in such way that they can be transferred to other defense projects as well.

VII. Recommendations for Strategical Planning

The digital defense council to the German Minister of Defence, states that "the future of AI in the armed forces [...] does not lie in the decision between man and AI, but in an effective and scalable combination of man and AI to ensure the best possible performance of tasks".[37] This includes the ethical dimension of digital technologies: "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*."[38]

Since we feel encouraged to assume that there might be a broader consent within the international defense science and data fusion community to these considerations, we are closing with some recommendations.

1. Digital ethics and a corresponding morality are part of the skills that we need to built up systematically in order in order to be able to use digital technologies to avoid serious

harm for humanity. In particular, they enable military decision makers "to assess the potential and impact of digital technologies and to manage and to lead in a digitized environment."[39] In particular, consideration should be given to instruments such as the proven *Innere Führung* as da guiding principle for the development of ethical competence and to encourage its systematic development with regard to defense digitization.

2. In addition to the operational benefit of digitization for the *Bundeswehr* in closing capability gaps, expanding its range of capabilities, and developing corresponding concepts, procedures, and organizational measures, ethical competence in dealing with digital technologies and ethical acceptance of them before the conscience of individual soldiers, but also before civil society need to be achieved. Both are key characteristics of successful innovation.
3. Digitization projects should be accompanied by ongoing analyses of technical controllability and personal accountability in a publicly visible and verifiable manner. Otherwise, the paradigm shifts and material efforts associated with artificial intelligence and automation would hardly be politically and socially enforceable. Of course, there will be more problematic and less problematic projects, so an exemplary approach would be appropriate.

Digital ethos and the corresponding moral formation are essential elements of a "digital self-image" that will shape the military capabilities of the future for the better.

ACKNOWLEDGEMENTS

The idea of making these considerations accessible to an international public emerged during a dinner conversation on "Cyber Challenges and Digital Ethics" to which Dr. Jill A. Long, Colonel, USAF, and US Air Attaché to Germany, had invited at the US Embassy in Germany. The author is very grateful to Jill for the conversations with her and the group she collected. This evening, overlooking the Ronald-Reagan-Terrace on the roof of the Embassy, the Brandenburg Gate, and the German *Bundestag* was much inspiring, indeed. The considerations themselves owe a great deal of thanks to Dr. Asgar Rieks, Lieutenant General and Deputy Chief of Staff of the German *Luftwaffe*. The author is deeply indebted to him for numerous hints, exchange of ideas, and his personal encouragement.

Johann Wolfgang Koch (M'00–SM'09–F'11) studied Physics and Mathematics at the Aachen Technical University RWTH, where he earned a PhD degree in Theoretical Physics. At Bonn University, he holds a habilitation degree and teaches as a Professor for Computer Science. For many years, he is working for the German Ministry of Defence and the German Defence and Aerospace Industry. At the Fraunhofer Institute FKIE, he heads the Department "Sensor Data and Information Fusion". In his role as Chief Scientist of Fraunhofer FKIE, he also co-ordinates on a broader scale R&D activities related to digitization in the domains of defense, aerospace, and public security.

Within the areas of his scientific interests, he has published a well-referenced monography, 18 handbook chapters and about 300 journal and conference articles. He is one of the co-editors and the authors of the handbook "Novel Radar Techniques and Applications". Of particular interest for him are ethical and legal aspects of Artificial Intelligence and Technical Autonomy in defense and public security. He is one of the initiators and co-chair of the high-rank working group "Responsible Technology for Future Combat Air System".

He is internationally active in the IEEE Aerospace and Electronic Systems Society and the NATO Science and Technology Organization.

[1] Baudissin, Wolf Graf von (1951-1969). *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr*. München 1969, 234.
[2] Ueberschär, Ellen (2019). *Anmerkungen zur Politischen Ethik und dem Verhältnis zwischen Ethik und Autonomen Systemen* [Notes on political ethics and the relationship between ethics and autonomous systems]. Impus zur Eröffnung der Arbeitsgruppe „Technikverantwortung bei FCAS", Bad Aying, 27.09.2019. www.fcas-forum.eu.
[3] Musk, Elon (2018). *Keynote Speech*. SXSW 2018, 11.3.2018. https://www.youtube.com/watch?time_continue=7&v=kzlUyrccbos.
[4] Hofstetter, Yvonne (2019). *Der unsichtbare Krieg. Wie die Digitalisierung Sicherheit und Stabilität in der Welt bedroht* [The Invisible War. How digitization threatens security and stability in the world]. Munich.
[5] German MoD (2019). *Erster Bericht zur Digitalen Transformation* [First report on digital transformation]. October 2019, p. 3. https://www.bmvg.de/resource/blob/143248/7add8013a0617d0c6a8f4ff969dc0184/20191029-download-erster-digitalbericht-data.pdf.
[6] Koch, Wolfgang (2019). *FCAS – Herausforderungen für Sensordatenfusion und Ressourcenmanagement* [FCAS – Challenges for Sensor Data Fusion and Resource Management]. In: cpm-Sonderheft „National FCAS SUMMIT 2019", pp. 8-11.
[7] Baudissin 1954, 206.
[8] Sauer, Frank (2017). *Kennzeichen des Krieges im 21. Jahrhundert. Entgrenzung und Beschleunigung* [Characteristics of war in the 21st century. Spatial Dissolution and acceleration]. In: Außerschulische Bildung 48 (1), S. 4-10.
[9] German MoD (2018). *Konzeption der Bundeswehr* [Concept of the Bundeswehr]. 20.07.2018, p. 48. https://www.bmvg.de/resource/blob/26544/9cedddf6df2f48ca87aa0e3ce2826348d/20180731-konzeption-der-bundeswehr-data.pdf.
[10] Konzeption der Bundeswehr 2018, 83f.
[11] Baudissin 1967, 84.
[12] Baudissin 1953, 242.
[13] Ueberschär 2019, 3.
[14] Koch, Wolfgang (2014). *Tracking and Sensor Data Fusion. Methodological Framework and Selected Applications.* Springer Mathematical Engineering Series, p. 78f.
[15] *Mission-type Tactics*. Wikipedia. https://en.wikipedia.org/wiki/Mission-type_tactics.
[16] Automated Augmented Battle Field Reconnaissance. https://www.fraunhofer-innovisions.de/kuenstliche-intelligenz/automatische-augmented-gefechtsfeldaufklaerung.
[17] Agile UAV in Vernetzter Umgebung – Urban CAS. https://www.fkie.fraunhofer.de/de/forschungsabteilungen/sdf/AgileUAV.
[18] Konzeption der Bundeswehr 2018, 50.
[19] Bernzen, Enno et al. (2016). Innere Führung – *Leadership Culture in Camouflage*. In: *Global Warriors? German Soldiers and the Value of* Innere Führung. http://www.ethikundmilitaer.de/en/full-issues/2016-innere-fuehrung/.
[20] Konzeption der Bundeswehr 2018, 84.
[21] S.Th. I,q.21 a.2.
[22] "Therefore, if things are the measure and guide of the mind, truth consists in the fact that the mind conforms to the thing. But if the mind is the guide and measure of things, truth consists in the conformity of the things with the mind." Ibid.
[23] Goodfellow, Ian (2015) et al., "Explaining and Harnessing Adversarial Examples", 3rd Int. Conf. on Learning Representations, ICLR 2015, San Diego, CA, USA.
[24] German MoD (2016). *Militärische Luftfahrtstrategie 2016* [Military Aviation Strategy]. 19.01.2016, p. 23. https://www.bmvg.de/resource/blob/11504/3e76c83b114f3d151393f115e88f1ffb/c-19-01-16-download-verteidigungsministerium-veroeffentlicht-militaerische-luftfahrtstrategie-data.pdf.

[25] Clausewitz, Carl von (1832). Vom Kriege [On War]. Hamburg [11]2018, II.3, p. 135.
[26] Clausewitz 1832, I.6, 96.
[27] Clausewitz 1835, I.3, 71.
[28] Spaemann, Robert (1996*). Personen. Versuche über den Unterschied von ‚etwas' und ‚jemand'* [Persons. Attempts on the difference between ‚something' and ‚someone']. Stuttgart [3]2006, p. 209.
[29] Spaemann 1996, 232.
[30] Hofstetter, Yvonne/Koch, Wolfgang/Westphalen, Friedrich Graf von (2019). Autonome Waffen. Das fünfte Gebot im KI-Krieg [Autonomous weapons. The fifth commandment in the AI war]. In: Spektrum.de , 05.07.2019. . https://www.spektrum.de/kolumne/der-krieg-der-zukunft-wird-dank-robotern-und-kuenstlicher-intelligenz-ein-problem/1655406.
[31] P7000 – Model Process for Addressing Ethical Concerns During System Design. https://standards.ieee.org/project/7000.html.
[32] Baudissin 1954, 234.
[33] Schubert, Hartwig von (2013). *Die Ethik rechtserhaltender Gewalt* [The ethics of law-preserving violence]. In: Wissenschaftliches Forum für Internationale Sicherheit, WIFIS-aktuell 48. Opladen Berlin Toronto, p. 49.
[34] Pieper, Josef (1955). *Werkausgabe letzter Hand* [Last hand edition]. Bd. VII. Hamburg.
[35] Baudissin 1967, 181.
[36] www.fcas-forum.eu.
[37] Erster Bericht zur Digitalen Transformation 2019, 27.
[38] German MoD (2019). Umsetzungsstrategie Digitale Bundeswehr [Implementation Strategy Digital Bundeswehr]. 14.06.2019, Nr. 102, 4. https://www.bmvg.de/de/themen/ruestung/digitalisierung/umsetzungsstrategie-digitale-bundeswehr. Umsetzungsstrategie Digitale Bundeswehr 2019.
[39] Umsetzungsstrategie Digitale Bundeswehr 2019, Nr. 209, 8.