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

Wolfgang Koch, *Fellow, IEEE*

I. Technical Assistance for Minds and Wills

'Intelligence' and 'autonomy' are omnipresent phenomena in the *biosphere*. Before any scientific reflection or technical realization, all living creatures fuse sensory perceptions with information they have learned themselves and received from other creatures. This gives them a model of their environment, the basis to act appropriately for reaching their goals and avoiding harm. In the *technosphere*, Artificial Intelligence (AI) and Information Fusion (IF) combined with comprehensive automation provides technical tools that enhance the perceptive mind and active will of persons who alone are capable to perceive consciously and to act responsibly.

Fig. 1: Cognitive and volitive assistance for the intelligent mind and autonomous will of responsible human decision-makers. © Fraunhofer FKIE

As illustrated in Fig. 1, algorithms, realized by the craft and art of programming and enabled by qualitatively and quantitatively appropriate testing and training data, drive a data processing cycle that starts from elementary real-time sensor signals and observer reports collected from multiple and heterogeneous sources. Information Fusion combines these streams of mass data with context knowledge and provides pieces of mission-relevant information at several levels that are integrated into comprehensive and near real-time situation pictures. On their basis, decision makers become aware of the current situation and decide to act according to the ends of their mission in a challenging environment. Algorithms transform their acts of will into partially or fully automated command sequences for controlling networking platforms, sensors, and effectors. Among the AI algorithms are Neural Networks and Machine Learning. The "world of algorithms", however, comprises much more than this particular type of information processing schemes. Algorithms, based on applied mathematics and running on powerful computing devices,[1] are thus the scientific core for designing cognitive and volitive tools that assist intelligent minds and autonomous wills in the "world of human beings". The concepts of mind and will, and therefore of consciousness and responsibility bring natural beings into view that are 'somebody' and not 'something', i.e. persons, and open up ethical dimensions.

General John E. Hyten (b. 1959) has characterized this situation by re-phrasing a passage from John F. Kennedy's (1917-1963) famous Moon Speech: "We set sail on this new sea because there is new knowledge to be gained, and new rights to be won, and they must be won and used for the progress of all people. For [artificial intelligence], like nuclear science and technology, has no conscience of its own. Whether it will become a force for good or ill depends on man, […] whether this new ocean will be a sea of peace or a new terrifying theater of war."[2]

II. Philosophical Preliminaries and Scope

Whoever wants to explore, develop, and leverage cognitive and volitive assistance systems, 'artificial things' that powerfully enhance the capabilities of natural minds and wills, must understand what is meant by 'artificial' and 'natural'. For this pair of notions stands as a fundamental dichotomy at the very beginning of Western thinking.[3] When speaking of 'technology', we are actually using the ancient Greek word for 'artificially created" – *technē*. "Artificial things [here: tools enabled by AI and IF], according to Aristotle, are indeed characterized by the fact that they themselves consist of a 'what' and a 'what of'," explains the philosopher Robert Spaemann (1927-2018) and continues: "Their 'how' and 'why' is not in them, but in the person who made them or use them. A natural thing, on the other hand, is characterized by the fact that their 'what' and 'to-what-end' in itself fall into one. Its end is the form of the thing itself, hence the notion of *entelecheia*: I carry the end within me."[4]

In contrast to the natural mind and will of persons, cognitive and volitive assistance systems are technical, i.e. artificially created things, and *not* natural entelechies according to this strand of thought. This contradicts emerging ideologies that consider 'conscious perception' and 'responsible will' as artifacts of

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

suboptimal information processors called 'humans' to be replaced by allegedly 'objective' and 'unbiased' AIs.

Any attempt to clarify the role of ethics in the military use of AI and IF must therefore keep two branches of mutually complementary knowledge in a proper balance.

1. *For knowledge itself is power.* Francis Bacon's (1561-1626) famous statement on achieving power as the meaning of all knowledge marks the very beginning of the modern project.[5] At the latest since the advent of AI in Defense, however, powerful technology may turn against humanity by using it irresponsibly or for unethical purposes. This admittedly age-old problem is acute with instrumental knowledge in its specific manifestation as AI and makes the modern crisis as visible as in spotlight. In addition, the Baconian paradigm has a tendency to instrumentalize even human beings as "stimulus-response machines according to a biocybernetic Image of Man"[6] with severe consequences for their status as legal subjects. "Nothing is as decisive for the style of a legal age as the Image of Man to which it is oriented," observed Gustav Radbruch (1878-1949), one of the most influential legal philosophers of the 20th century on the eve of a tyranny. According to this line of thought, a society, which sociologically standardizes man and reduces him to what is empirically observable will sooner or later manifest this Image of Man in the law established by the sovereign. For law is nothing but 'coagulated politics'.[7] Thoughtfully, the political philosopher Jürgen Habermas (b. 1929) has warned of a possibly "derailing modernization", which could very well erode the kind of solidarity on which the democratic state and the international community of states, without being able to legally enforce it, is dependent.[8]

2. Since there is an *Ecology of Man* to be taken into account, ethically guiding knowledge of the nature of persons, and their proper ends must complement Bacon's knowledge. "Man too has a nature that he must respect and that he cannot manipulate at will," reminded Benedict XVI. (b. 1927) the parliamentarians in the German *Bundestag*. "Man is not merely self-creating freedom. Man does not create himself. He is intellect and will, but he is also nature, and his will is rightly ordered if he respects his nature, listens to it and accepts himself for who he is, as one who did not create himself. In this way, and in no other, is true human freedom fulfilled."[9] In view of these insights, AI in Defense must serve the Common Good of a nation or coalitions of nations as it is discovered in the natural order as its "moral ecosystem". If a "nation's will"[10] aligns itself with partisan interests, which deny universal inalienable rights in conformity with the nature of persons, we might still speak of "government by the people", but it cannot be said to be "of the people" or "for the people",[11] paraphrasing Abraham Lincoln's (1809-1865) phrase sketching the essence of modern democracy.[12] The leadership principle *Innere Führung* of the German Armed Forces to be discussed below and the notion of a *Staatsbürger in Uniform* [citizen in uniform] underlines explicitly the importance of the 'human factor'.

We do not expect that the technological development will "naturally" lead to responsibly usable AI in Defense. Even the development of irresponsible AI-based weapon technology is possible and may be pursued by adversaries. Ethical considerations in the exploration and development of AI in Defense should therefore actively be encouraged and its responsible use technically be facilitated. This includes the conception of well-designed Rules of Engagement that take into account the risks of AI in Defense and must permeate the technical system design. 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 by international law, such considerations are also encouraged from a legal perspective.[13]

After discussing some basic preliminaries related to the Image of Man and the Common Good from the author's perspective, without which ethical problems cannot be properly addressed, we consider in this article the German Armed Forces, the *Bundeswehr*, in the sense of a case study. It serves as an example of an army acting under a parliamentary mandate that has learned lessons from tyranny and total war. Conclusions from ongoing research lead us to the notions of technical controllability and personal responsibility that appear as technical design principles for implementing AI in Defense. Although being relevant in other defense technologies as well, these issues are particularly pressing, since AI-driven automation has the inherent tendency to marginalize human involvement in decision processes. A discussion of challenges and research questions to be answered for enabling their technical implementation seems to indicate that AI in Defense may lead to a paradigm shift in systems engineering. In particular, the difference between the errors of unaided human decisions in comparison to AI-assisted and automatically executed decisions requires careful considerations on how personal accountability can technically be supported. On the other hand, Artificial Intelligence shows characteristic defects with 'blind spots' and unreal artifacts to be compensate by naturally intelligent and well-trained users. These thoughts result in the need to incorporate ethical thinking into systems engineering and ergonomics for AI in Defense staring form the highest level and the very beginning in the stages of strategic planning, defense R&D, procurement, and leverage.

If we were able to solve the ethical problems of AI in Defense, new paths are likely opened up for responsible use of AI in other sectors, such as healthcare or autonomous driving, where the impact of poor decisions on people can be significant as well. For an overview of the ongoing debates see [14].

III. AI IN DEFENSE AND THE COMMON GOOD

Since the ethical dimension of AI in Defense transcends the scope of a purely scientific discussion, any introduction to so controversial a topic must openly admit the fundamental positions of the author who is trying to provide a contribution to it. Only in this way will a transparent dialogue become possible.

Instrumental knowledge providing "power" is never an end in itself, but always meant "for" something. This is particularly true when it is used for defense, i.e. for protecting the Common Good of nations or coalitions of nations against external physical threats, i.e. against military threats emanating from other na-

tions and non-state actors. There are non-physical, e.g. economic or ideological,[15] and internal threats as well, e.g. those related to public security,[16] that we exclude here.

In this sense, we consider cognitive and volitive assistance systems under research, being developed, and to be leveraged for protecting the Common Good, to which the individual good is essentially related and which it serves. "The political community […] exists for the Common Good: this is its full justification and meaning and the source of its specific and basic right to exist," summarizes a classical text a long tradition of political thought: "The Common Good embraces the sum total of all those conditions of social life which enable individuals, families, and organizations to achieve complete and efficacious fulfilment."[17] This description of the Common Good rooted in the nature of persons and their ends seems to be quite in line with *The Declaration of Independence* of the United States of America, which describes the objective fulfilment, which the Common Good safeguards and promotes, as "Life, Liberty, and the Pursuit of Happiness."[18]

The very notion of the Common Good as understood here embraces natural law and inalienable human rights based upon it, e.g. not to be induced to act immorally. Although being fundamental to the Western, i.e. European and American, civilization[19] and having evolved in a process, it seems to have universal validity. "In the first half of the second century B.C., the social natural law developed by the Stoic philosophers came into contact with leading teachers of Roman Law. Through this encounter, the juridical culture of the West was born, which was and is of key significance for the juridical culture of mankind. This pre-Christian marriage between law and philosophy opened up the path that led via the Christian Middle Ages and the juridical developments of the Age of Enlightenment all the way to the *Declaration of Human Rights*."[20] The *Grundgesetz* of 1949, which commits modern Germany to inviolable and inalienable human rights as the foundation of every human community, and of peace and justice in the world, is in line with this development. According to this tradition, readiness for defense 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 states.[21]

Also rooted in ancient Roman and medieval juridical thinking, but from a different philosophical perspective, political thinkers such as Thomas Hobbes (1588–1679), Samuel von Pufendorf (1632–1694), and John Locke (1632–1704) have made the notion of inalienable rights a key element in *The Constitution of the United States* as well. Especially von Pufendorf's political concepts have become part of the cultural background of the American Revolution. In view of these considerations, European and U.S.-American legal traditions at least in certain, but essential, points coincide.

According to these considerations, the precepts of the natural law are to the practical reason, i.e. to the vast corpus of positive law, e.g. to the detailed and complex Rules of Engagement, what the first principles of demonstration are to the speculative reason; because both are self-evident principles.[22] In other words, knowledge of the nature of persons and their proper ends seem to be the precondition of the possibility for digital ethics for AI used in the Defense Domain.

Both over-expectation and over-fear apparently characterize the public apprehension of risks and opportunities of AI. A Google search for images illustrating 'Artificial Intelligence' reveals a psychogram of modern man with its pseudo-religious hopes and gloomy forebodings. As Fig. 2 shows, intelligently looking alien beings in black, turquoise, blue, or white are mysteriously rising from circuits and data, cool and superior, quite different from us, yet somehow "in the Image of Man". Artists' views inspired by Michelangelo's "Creation" symbolize emerging intelligence; robots pose as Rodin's "Thinker". At the same time that the Image of Man is increasingly shaped according to the model of machines, human, even superhuman qualities are somehow associated with Artificial Intelligence. Often, the collective unconscious influences researchers and developers, procurement and decision-making in an unreflected way. Google's psychogram also reveals an underlying fear: "Mark my words – A.I. is far more dangerous than nukes." [23]

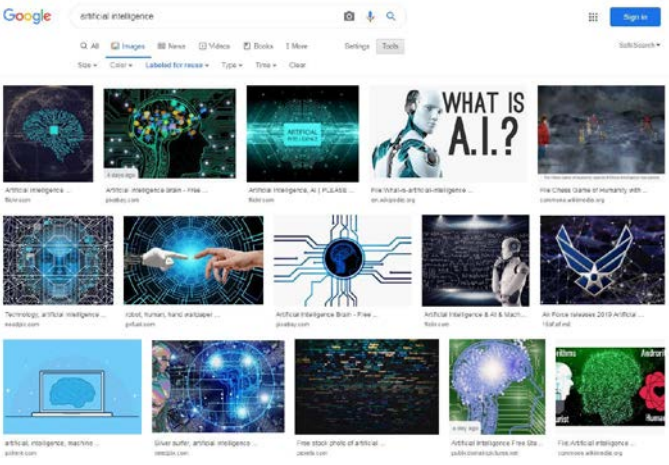

Fig. 2: Result of a Google image search with the keyword *Artificial Intelligence*. A psychogram of networked modern man? (© GIDS).

"There is a delicate balance that must be achieved between espousing the need and purpose for innovative technology, and managing expectations when it comes to execution," observes the current US Air Attaché to Germany. "Striving for perfection should always be part of the [military] profession; claiming achievement of such perfection is folly."[24] Artificial Intelligence may indeed become dangerous because of 'natural stupidity', the refusal or mental inertia to be a conscious and responsible person in the full sense that may prevent us from facing the real risks of misused AI. Do we need a new enlightenment for dealing with AI in a mature and ethical way, "man's emergence from his self-imposed nonage"?[25] *Sapere aude* – Have the courage to use your own intellect! "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 (b. 1967), a political thinker in Germany.[26]

IV. Soldiers' Minds and Wills in Hyperwar

In the West German Adenauer Era, the conceptual architects of the *Bundeswehr* wanted to establish structures that prevent any

return of what has happened in Germany from 1933-1945 and to Europe in WW II. Their minds and wills were shaped by the collapse of the Weimar Democracy that an evil man prevailing the support of a majority turned into a tyranny. Without recognition of the Common Good and inalienable rights, an unrestricted war began with all destruction, which is the fruit of unbridled nationalism and then ruthless use of technology.

At the same time when the technical term AI was coined, the architects of the *Bundeswehr* have anticipated the notion of 'hyperwar'. "The scientificization and mechanization of the military craft will lead to the dissolution of spatial boundaries and acceleration of military action", predicted Colonel Wolf Graf von Baudissin (1907-1993) the military development in 1954, one year before the Bundeswehr was founded and two years before the famous Dartmouth Summer Research Project on Artificial Intelligence took place. "The most highly mechanized combat requires that responsibility is seen and borne at many lower levels," he continues (Fig. 3). "Therefore, everything must be done to teach people the skills that make them aware of their responsibility and help them experience the consequences of their actions and omissions."[27]

How to meet these demands today? How to design cognitive and volitive assistance systems that enable conscious situational awareness, on the one hand, and encourage the responsible use of partly or fully automated systems in defense on the other? 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."[28]

Fig. 3: "The most highly mechanized combat requires […] making soldiers aware of their responsibility and making them experience the consequences of their actions and omissions." Wolf von Baudissin (1954). © Bundeswehr

Systems empowered by AI and IF are already in use in all operational domains Land, Sea, Space, Cyber, i.e. in the increasingly complex technosphere of military operations.[29] Cognitive and volitive assistance in the defense domain is thus inevitable to "information, command & control, and engagement superiority" as well as to "improving their ability to assert and respond".[30] Since potential adversaries also use or will use such digital technologies, dissolution of spatial boundaries and acceleration will characterize digitized combat with increasingly automated cause-and-effect chains, i.e. 'hyperwar'. A European example is Future Combat Air System (FCAS), a system-of-systems consisting of manned and unmanned flying platforms to protect the European airspace.[31]

Because we explicitly accept that modern armed forces must be capable "to fight at machine speed", digitization in the defense domain cannot be limited to Intelligence, Surveillance, and Reconnaissance (ISR), but must equally enable, encourage, and facilitate responsible weapon engagement. "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 von Baudissin as a highly respected Lieutenant General in 1967, in the middle of the Cold War (Fig. 4). "Without the commitment to the moral realms, the soldier threatens to become a mere functionary of violence and a manager."[32] This is all the more valid in today's sensor-to-shooter loops and Joint All Domain Command & Control.[33] The hope for deceleration and limitation by arms control policy to counter military destabilization such as flash crashes on financial markets is honorable,[34] but seems unrealistic in view of the almost unrestricted proliferation of digital technologies and their dual use.

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

For the sake of clarity, the *Bundeswehr* explicitly avoids the term 'autonomy', which obscures gradual differences. The term 'automation' seems to be more appropriate as it also includes 'full automation' after stages of partial automation. Nevertheless, there an ongoing debate on 'autonomy vs automation' at the level of international diplomacy in order to characterize so-called Lethal Autonomous Weapon Systems (LAWS) according to the regulations of the International Humanitarian Law.[35] Absurd is the interpretation of 'autonomy' as of a technical 'decision of will'. In view of these considerations, also 'Artificial Intelligence' is a problematic term, too well established, however, to be avoided.

To ensure that the *Bundeswehr* fulfils their current obligations, digitization is expanding its capability profile at 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."[36]
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."[37]

In this context, defense scientists have to clarify the ethical problems of AI and automation, while avoiding 'moralizing'. The following distinction proves to be helpful.

1. *Digital ethics* denotes thinking about the right decisions in using digital defense technologies. Required is an Image of Man that makes '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. If this is only seen from a functional and legal point of view, armed forces become a danger", observes von Bausissin.[38] "In the jungle situation of borderless wars, 'soldiers only' are no longer fit for war in the long run."[39] This aspect is intimately related with the concepts of 'Citizen in Uniform' and *Innere Führung*.
3. *Digital morality*, finally, comprises concrete guidelines for dealing with AI and automation, not only in the battlefield, but also in research, development, and procurement.

V. Responsible Design: Conclusions from Examples

As an early example of algorithmically provided assistance for situational awareness, Fig. 5 illustrates the efficiency of so-called Multiple Hypothesis Tracking (MHT).[40] The left side shows data provided by a long-range radar, accumulated over one hour. Without algorithmic assistance, decision makers were unable to extract any useful information from the data stream about two highly maneuvering pairs of fighters training air combat. MHT, an example of an AI algorithms, however, produces precise target trajectories from radar plots that are repeatedly missing and contaminated with numerous false, unwanted, or unresolved echoes (on the right).

Systems such as AWACS (Airborne Early Warning and Control System) or AGS (Alliance Ground Surveillance) massively generate information of this type on single or collective target tracks. On this basis, the algorithms of Big Track Data Analysis produces 'recognized' situation pictures for threat evaluation, guidance, and weapon assignment. Contextual information on topographical conditions or constraints, such as a plan to be followed or certain rules to be obeyed, are essential in this process, especially according to the conceptual framework *Führen mit Auftrag*, leading by mission.[41] Fusion of sensor data and non-sensor information thus provides mission-relevant insights that simply were unavailable without AI in Defense. Apparently, comprehensive Information Fusion (IF) is the key technology for integrating also formalized ethical or legal rules seamlessly into reconnaissance or combat missions.

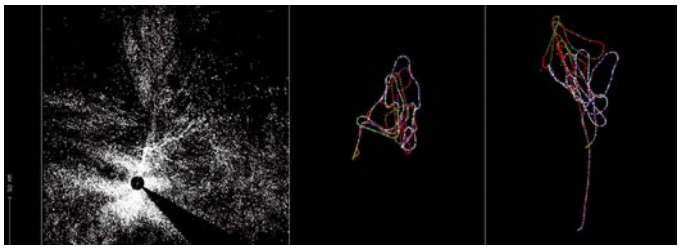

Fig. 5: Trajectories of two pairs of dogfighting fighter aircraft extracted from poor radar data by Multiple Hypotheses Tracking. © Fraunhofer FKIE

Ongoing R&D projects on using armed drones for closing capability gaps of the *Bundeswehr* have already identified research questions for responsible systems design.[42] As an example, Fig. 6 shows a convoy under attack and that was stopped in urban environment by an Improvised Explosive Device (IED). For the Forward Air Controller (FAC), coordinately operating reconnaissance fixed- and rotary-wing drones provide a comprehensive situation picture, including the expected collateral damage, thus enabling self-defense by commanding a combat drone. Here, AI in Defense provides technical prerequisites for responsible engagement with minimized risk for non-combatants. We claim that armed drones enable reliable or, compared to other weaponry, significantly more reliable target reconnaissance and control up to the final engagement decision.

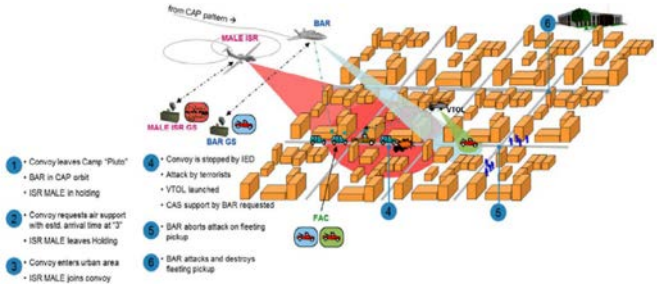

Fig. 6. Coordinated reconnaissance and armed drones enable the FOC to defend against threats in compliance with the RoE. © Fraunhofer FKIE

Rules of Engagement (RoE), which do not make any tactical specifications, but define a legally binding and mission-specific framework, are important for pre-engagement situation analysis. In compliance with legal, political, strategic, and operational requirements related to the Common Good, they concretize the *ius in bello*, i.e. the 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. Evidently, RoE are to be taken into account in designing AI-based systems (RoE Compliance by Design).

In conclusion, AI and IF combined with automation enable cognitive and volitive assistance systems. Through them, military actors acquire knowledge about threats, combatants, uninvolved parties, and options for action in the various operational theaters. At the same time, risks for friendly forces, the civil population, the infrastructure, the environment etc. can be minimized. Such assistants may thus help to master complex tasks

more adequately, to balance human subjectivities, and to protect non-combatants. Moreover, the physical presence of humans is thus increasingly dispensable in dangerous situations. In summary, cognitive and volitive assistance systems are 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 enable ethical, legal, and societal compliance.

Since decentralization and automation due to digitalization in defense implies vulnerability, own systems need to be protected against attacks from the electromagnetic spectrum and the cyber space, on the one hand, while strategies to attack enemy systems in this way are to be developed on the other.

On February 6, 2019, Germany and France launched the most ambitious European armament project of the coming decades – Future Combat Air System (FCAS), where manned fighter jets and unmanned 'remote carriers' form of a comprehensively networked system of systems. As 'loyal wingman', cooperating multiple-sensor, multiple-effector drones protect manned platforms and execute reconnaissance or combat missions, while satellites, Airborne Early Warning, in-flight refueling, or air transporting will be integrated. The core is an 'Air Combat Cloud', which makes all relevant information available to all actors on a mission in real time. In the digital era, decision superiority decides between success or failure of a mission.

Fig.7: AI und IF driven "system of systems" with Manned-Unmanned Teaming (MuM-T) for combat and reconnaissance missions. © Fraunhofer FKIE

Since AI and IF combined with automation are enabling technologies for FCAS, the ethical and legal challenges previously discussed are evident. For this reason and for the first time in Germany, a working group accompanies a major armament project from the outset with the explicit mission to operationalize ethical and legal principles through appropriate design on the technological level. While compliance with the national and international legal framework is only the "ethical minimum" of a more comprehensive approach, transparent proof that FCAS is indispensable for protecting the Common Good of Europe and its partners, is the basis for the political implementation and societal acceptance of this large project.

In accordance with the German Military Aviation Strategy 2016,[43] the responsibility of human decision makers is pivotal for all conceivable mission scenarios. The working group mentioned,[44] benefits from the complementarity between industry and research. The envisaged overall architecture is expected to contribute to responsible use by reducing the complexity of future missions and to facilitate responsible human decision-making. Technological complexity reduction is thus a key development goal. Realistic simulations accompanying the technological development from the very beginning shall ensure that ethical, legal and even societal compliance is not at the expense of effectiveness in combat. Furthermore, participation of relevant stakeholders, including the society as a whole and represented by a cross-section of political foundations, universities, and political think tanks, is decisive for the success of this initiative.[45]

VI. Controllability: Basis of Military System Design

AI in Defense enables military decision-makers to consciously perceive and responsibly decide even in the highly complex and accelerated technosphere of modern conflicts. However, are we really facing fundamentally new challenges? Certainly not. New technologies have repeatedly increased human perception and range of action. The difference between the digital revolution and earlier innovations is rather quantitative than qualitative in nature.

The most recent update of the foundational document of the *Bundeswehr* opens up a discussion on a more philosophical level when it explicitly re-affirms the principle of *Innere Führung* as "the underlying philosophy of leadership valid for the German soldiers."[46] The term *Innere Führung*, 'leadership from within', itself is not easily translated. According to von Baudissin's idea, *Innere Führung* comprises all aspects of military leadership with special consideration of the individual and social aspects of the person as such. Its overall goal is to reconcile the functional conditions of operational armed forces with the principles of a democratic constitutional state.[47] Obviously, this leadership philosophy closely links the personality development of German soldiers with the Common Good. Rooted in the tradition of Western moral thinking, *Innere Führung* therefore seems to be quite in line with a reasoning that "argues on the basis of reason and natural law, namely on the basis of what is in accord with the nature of every human being."[48] A closer analysis of the impediments of the conscious perception of the actually existing situation and responsible action according to proper ends in it, which are "not reduced to purpose and consequence", is also an indication of what is necessary in the personality development program being part of *Innere Führung*.[49]

In view of these considerations, AI in Defense poses a timeless question – How to guarantee ethical, legal, and social compliance; how to decide 'well' according to what is recognized as 'true'? In engineering, this leads to two other questions:

1. How to design *cognitive* tools that we can not only mentally but also emotionally master in each situation?
2. Which technical design principles facilitate the responsible use of artificially intelligent *volitive* tools?

As indicated in Fig. 8, AI generates situation pictures that assist "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.[50] 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 and intents. Important building blocks are furthermore inferred object interrelations. Finally, situation pictures indicate decision relevance, such as threat levels and the status of own resources and countermeasures. The situation picture as well as statements about its limitations and gaps, on the one hand, and the actual situation on the other must mutually correspond. This implies a concept of truth and its elementary foundation: "Truth consists in the equivalence between the situation picture and the situation."[51] 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.[52]

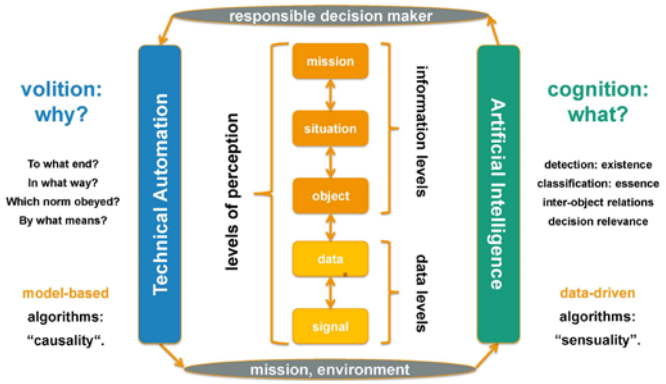

Fig. 8: Levels of AI-assisted perception and action for Command & Control, ISR, engagement, and impact assessment. © Fraunhofer FKIE

Automation translates the intentions of the decision maker into complex cause-effect chains to manage assets such as multifunctional sensors, mobile platforms, communication links, effectors etc. The question 'why' to achieve an effect is crucial for algorithm design. Aristotle's model of causality is helpful here, since it distinguishes four ways of answering to why-questions. The intentions correspond to the *final cause*, usually specified by performance parameters characterizing the intended effect. The *causa efficiens*, the effective cause, indicates by which concrete algorithms the effect is to be achieved. The *formal cause* answers the question, according to which rules this should happen. Finally, the *material cause* indicates necessary resources 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 ends 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 basic 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 to what end?*

In the algorithmically driven information processing cycle as sketched in Fig. 8, we distinguish five levels of perception. The first two levels, determined by received signals or human observer messages from reality and the corresponding signal or language processing, are *data levels* and usually hidden from military decision makers. For them, the three *information levels* are more relevant. They refer firstly to the individual objects, secondly to the situation with information about the interrelation between the objects, and thirdly to the mission itself, i.e. the underlying situation including the decision maker who wants to act in it according to his or her resources.

VII. On Responsibly Guaranteed Controllability

A challenge for consistent and responsible controllability of AI in Defense is firstly the ever-decreasing time available for human involvement in the decision making process. A further problem is the limited explainability and deceivability of algorithmically generated information and the automated execution of complex command chains.

As an example, let us consider neural networks that, seen from an abstract point of view, assign an input to an output. The output describes what the input should 'mean' for the user. Characteristic of such functions is their extremely large number of internal degrees of freedom, tunable numerical values. In a so-called 'training phase' they are adjusted by 'telling' the neural network what a particular input actually 'means', e.g. by 'understood' images. The 'labeling' of training data requires human understanding. If training has been 'long enough', the network is offered an arbitrary input and the output is considered the recognized 'what', i.e. the 'meaning' of the input. Neural networks are thus essentially function approximators. Whoever calls massive offering of interpolation points 'learning' may awaken erroneous associations in non-specialists.

As it turns out, however, only a few pixels in an input image, for example, need to be changed in a specific way to completely mislead even well-trained networks. A neural network, deceived by such "poisonous noise", may 'misrecognize' a panda bear, for example, which appears unchanged to humans, as a gibbon monkey and 'feels' certain in its judgment.[53] 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 where electronic measures call for counter measures, these for counter-counter measures and so on.

In addition, appropriately representative training data for data-driven algorithms, such as neural networks, are for most militarily relevant applications unavailable in sufficient amount. Moreover, such algorithms are often 'black boxes'. Nobody knows how these achieve their results. Furthermore, context knowledge – fundamental to every military mission – can only be learned indirectly from the training data themselves. In short, data driven algorithms are "greedy, brittle, and opaque"[54] and always provide the 'second-best solution only'.[55] At least for critical functions in the targeting cycle, 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, deliver probabilistic assessments of the estimates provided, can be developed systematically, and explicitly allow the integration of context and expert

knowledge. In certain militarily relevant cases, however, the required models are not available or too complex to be dealt with efficiently. A still unsolved problem of current research is the combination of data-driven and model-based algorithms resulting in Explainable Artificial Intelligence (XAI).

According to these considerations, the following issues are pressing in the military domain, but relevant in other domains as well and need to be addressed by systems engineering:

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 military 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 thus be reconsidered.[56] More fundamental is "accountable responsibility" to be discussed below. The use of fully automated effectors on unmanned platforms may well be justifiable, even necessary, in certain situations if appropriately designed.
3. Certification and qualification of AI-based automation are key issues. Robust military systems will comprise both data-driven and model-based algorithms, where data-driven 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 are essential prerequisites for critical components. 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 be formalized, or partly contradictory. Statistical models exploited by powerful algorithms, however, enable responsible action even on an imperfect data basis. In many cases, reliable situation pictures can be inferred from them in a much more precise, complete and faster way than humans could ever have hoped to obtain. Nevertheless, also these methods have their limitations, which decision-makers must not only be made aware of, but also be interpreted to them. Under certain circumstances, proper models are unavailable and have to be learnt provided appropriate data exist.
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 or unwanted action. Mature AI comprises detection of such deficits, which is the basis for making own assistance systems resilient towards 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. On the other hand, without AI in Defense even the cleverest and well-trained military genius will not have the informational basis and efficient reactivity that is needed in modern conflicts.

"All thinking is art," observes Carl von Clausewitz (1780-1831), 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."[57] Thus, digitization requires the ethos of digitally educated decision-makers who do not need to know how to design and program AI algorithms themselves, but are able to assess their strengths and weaknesses, risks and opportunities. 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."[58] The associated 'digital morality' is fundamentally teachable. Obviously, this addresses a key question of the soldierly ethos, which is aggravated by the digitization, but is not fundamentally new and should lead to a 'digital update' of *Innere Führung.*

VIII. Responsibility in Military Systems Engineering

The responsible use of AI in Defense requires systems that connects conscious cognition and responsible volition in the real world with data collection and automatically executed command chains for controlling sensors, platforms, and effectors.

Literally, the 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 concept has far-reaching implications: What actions or omissions are 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'? Although there exists a vast literature, covering controversial points of view and influenced by most diverse cultural backgrounds, a broader consensus seems to converge on the following aspects.

1. To speak of responsibility is only reasonable if it is assumed voluntarily. Responsibility thus presupposes the notion of 'freedom' and an Image 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 the 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".[59]

In the current debate on hyperwar, responsibility is more fundamental a notion than meaningful human control. Even the use of fully automated systems, which, after the decision to use them, achieve their effect without human intervention, can be justifiable under certain clearly defined conditions. Examples are defense against armed drones or 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, since, in a certain sense, humans are still involved even in these examples, namely with the decision to switch on such system and to select the parameter framework for enabling automated defense, so that meaningful human control is guaranteed in a broader sense.

Fig. 9 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 establishes 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 is determined by prospective action and reaction. Responsibility can only be assumed by persons for persons. Firstly, everyone is responsible for himself or herself. "Whoever assigns freedom of will to man, wants to say that man is 'himself' the cause of his acting in this very way and not in another," explains Robert Spaemann. "And this occurs in such a way that man himself is also responsible for his own existence, insofar as this existence is shaped by his own actions and certain decisions to act are at the same time decisions about what kind of person someone wants to be."[60] Secondly, responsibility is owed to own forces and the Common Good to be protected. Finally, there is responsibility to civilians and combatants. One could even speak of responsibility towards the natural habitats in the theaters of operations.
3. *Towards whom is responsibility assumed?* Responsibility implies the notion of an authority, which is exercised by judgement. The responsible person recognizes it by his or her justification. The relationship between the person and the authority is retrospective in nature. Authorities are God, the personal conscience of the responsible person, the superiors, and jurisdiction exercised by persons in the framework of courts or justice systems.

Fig. 9: Core elements of the concept of responsibility and the resulting mutual relationships. © Fraunhofer FKIE

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

According to these considerations, only natural persons and not machines, can act responsibly or irresponsibly. For technical 'things' always remain a-personnel, even if they seem to '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 situational awareness and action are acting in a 'good' or 'evil' way by responding to moral challenges in one way or another. 'Good' technical systems encourage the morally acceptable and efficient leverage of them to achieve military objectives. 'Evil' systems, which might be designed by an adversary, 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 are fruitful, this is inappropriate in a military context. For certain actions are also here *per se* – i.e. "deontologically" – qualified as immoral. On the other hand, a "consequentialist" weighing of consequences can be quite moral under certain conditions. For responsibility is graded and corresponds to the concretely existing relationship of concern and trust. Against an opponent of war, "lies", measures of electronic or cyber warfare, for example, are permitted, since he is "not at all in that moral relationship of trust that makes truthful speech necessary".[61] This does not mean, however, that consequentialism is acceptable as a moral doctrine with its tendency to deny inalienable rights and an objective moral order.[62]

IX. Selected Aspects of Moral Assistance Systems

In view of the previous discussion, systems engineering for designing responsible cognitive and volitive assistance tools, which technically supports ethically and legally compliant behavior, has to fulfill three major requirements:

1. situational awareness to enable responsible action,
2. cognitive assistance to identify responsible options to act,
3. comprehensible plausibility of the proposed options.

These are basic for ensuring responsible decisions before, during, and after the mission in order to achieve clearly defined

ends and intermediate purposes in a given operating theatre that take into account to what extent collateral effects can be tolerated. Fig. 10 illustrates how these requirements could be met in the development and deployment of assistance systems for responsible action.

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."[63] 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 on all levels of responsibility, not only in combat. These considerations correspond to some extent to the IEEE P7000 *Model Process for Addressing Ethical Concerns During System Design*,[64] by which engineers and technologists can address ethical consideration throughout the various stages of system initiation, analysis and design.[1] The VDE, the German equivalent of the IEEE, proposes comparable recommendations.[65]
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*, on the one hand, replaces outdated technology while letting procedures and processes largely unchanged, while *disruptive innovation*, on the other, 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 licensing and qualification bodies.
3. Mission-relevant decisions can be evaluated and correspond to the mission-specific Rules of Engagement (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 the spatially delimited and accelerated hyperwar, ethically relevant knowledge itself must be made electronically accessible.

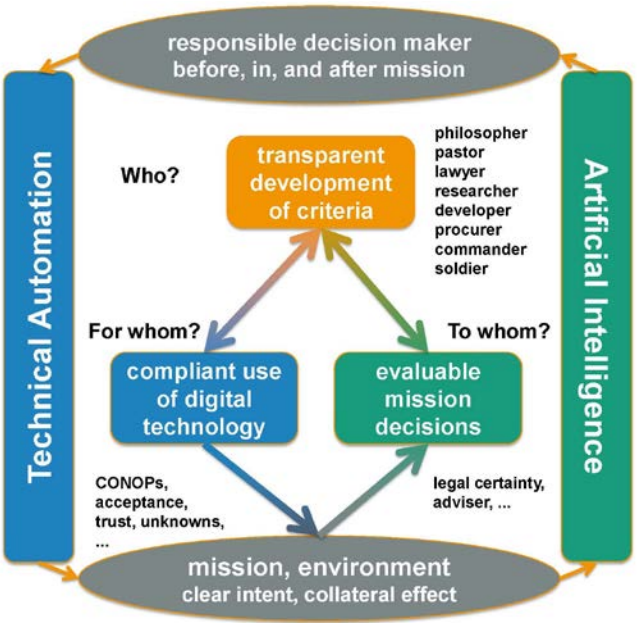

Fig. 10: Transparent criteria development for technology application and application decisions. © Fraunhofer FKIE

A concrete example is pre-engagement assessment of 'collateral damage'[66] and 'risk estimate distance', i.e. an assessment of the danger to friendly forces, that are made possible by the AI in Defense in a way that has never been achieved before. This is important for both effective and responsible weapon control, be it through precision targeting or scalable detonation effect of modern warheads. In order to avoid confusion on the term 'precision targeting', it should be recalled that it is defined as "directed at a specific target". From the aviator's perspective, for example, "the term 'precision' does not imply, as one might assume, accuracy. Instead, the word precision exclusively pertains to a discriminate targeting process. By using a word that has such specific meaning in the mind of most civilians, it is easy to see how a gap in understanding and expectations has been fostered."[67] 'Precision' is not 'perfection'.

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 the 'unknowns'. Without reliable knowledge about the limits of the available knowledge, no one can act responsibly. In addition to engineering issues,

[1] The IEEE P7000 Working Group places particular emphasis on the notion of 'values' that are to be systematically elicited, conceptualized, prioritized and finally respected via appropriate systems design [S. Spiekermann (2018), "Carousel Kittens: The Case for a Value-Based IoT," IEEE Pervasive Computing, vol. 17, no. 2, pp. 62-65, Apr.-Jun. 2018.]. The philosophical background is Material Value Ethics, first established by May Scheler (1874-1928) and Nikolai Hartmann (1882-1950). A core trait is its focus on virtue ethics, i.e. emphasis on culturally or socially desirable character traits. This design approach aims to maximize positive value potential and minimize value harms for people in IT-rich environments. We do honor the intention of this approach. Communities must share common values. This is particularly true in democracies that are only stable if the majority values rights and duties. They are based on law, however, not on moral obligation. Communities that are committed to individual freedom require observance of its laws, not the conformity with values that underlie its legal system. It may even be dangerous to speak of 'community values' because there is a tendency to undermine the legal principle in favor of a 'dictatorship of beliefs'. There have been and there are 'communities of values', where values have taken or take precedence over the law. We primarily have to talk about norms, not values only

ergonomics and cognitive sciences apparently play an important role in the digital transformation of the armed forces.

According to these considerations, ergonomic representation of the situation pictures obtained from a wide range of imperfect data sources and the limits of their informative value is a key requirement. Apparently, the research questions arising from this demand have an ethical dimension: the situation picture, despite all its abstraction, must convey to the decision maker psychologically realized awareness of the reality of the situation shown, i.e. its truth, help him to take responsibility, "and allow him to experience the consequences of actions and omissions."[68] 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 in a given situation 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".[69] 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 most cultures, and be implemented in such assistance systems in order to make them user-compliant in the sense of responsible judgement. The so-called four "cardinal virtues" of Western ethics[70] are examples with a potential of wider consent.

Only if based on a clearly defined and realistic Image of Man, his nature and his ends, that is compatible with the notion of responsible use of technology, digital assistance can support morally acceptable decisions. Awareness raising of such an Image of Man and *Innere Führung* as a guiding leadership and personality developing 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."[71]

X. 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".[72] 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*."[73]

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

1. Digital ethics and a corresponding ethos and morality are part of the skills to be built up systematically for responsibly using AI in Defense without serious harm for humanity. In particular, such skills enable military decision makers "to assess the potential and impact of digital technologies and to manage and to lead in a digitized environment."[74] In particular, consideration should be given to leadership philosophies and personality development instruments such as *Innere Führung* as a guiding principle for the development of ethical competence and to encourage its systematic development with regard to AI in Defense.
2. In addition to the operational benefit of defense digitization in closing capability gaps, expanding the range of capabilities, and developing corresponding concepts, operational procedures, and organizational measures, ethical competence in dealing with digital technologies and their ethical acceptance need to be achieved. Only then, AI in Defense will become acceptable before the conscience of the individual soldiers, but also in the broader view of the Common Good of the society as such. Success in both aspects will indicate a real innovation.
3. Digitization projects should be accompanied by ongoing analyses of technical controllability and personal accountability in a publicly visible, transparent, and verifiable manner. Otherwise, the paradigm shifts and material efforts associated with artificial intelligence and automation would hardly be politically, societally, and financially enforceable. Of course, there will be more problematic and less problematic projects, implying that an exemplary approach would be appropriate.
4. Although the International Law requires human responsibility, it defines an "ethical minimum" only. To achieve progress in the legal development, it seem promising to consider Corporate Social Responsibility. The principles anchored in this way aim to create binding standards for the responsible design and leverage of cognitive and volitive systems, including the entire supply chain. Misconduct is to be tangibly sanctioned – soft law and hard sanctions, ranging from extraordinary contract termination to contractual penalties. Although soft law is not enacted by the legislature, it may become as a source of legal knowledge.

*

"We set sail on this new sea because there is new knowledge to be gained, and new rights to be won, and they must be won and used for the progress of all people. For [artificial intelligence], like nuclear science and technology, has no conscience of its own. Whether it will become a force for good or ill depends on man, […] whether this new ocean will be a sea of peace or a new terrifying theater of war."[75]

ACKNOWLEDGEMENTS

The author wishes to thank two anonymous reviewers for their insightful suggestions, which have substantially improved this paper. The idea of making these considerations accessible to an international public emerged during a dinner conversation on "Cyber Challenges and Digital Ethics" to which Jill A. Long, Colonel, USAF, and US Air Attaché to Germany, had invited him at the US Embassy in Germany. The author has fond memories of an inspiring evening overlooking the *Brandenburg* Gate and the *Bundestag*, while he is gratefully indebted to the subsequent discussions with her. The considerations also owe a great deal of thanks to Dr. Asgar Rieks, Lieutenant General and Deputy Chief of Staff of the German *Luftwaffe*, due to his numerous hints, exchange of ideas, and personal encouragement. The author is also grateful for valuable exchange with Dr. Sibylle Bauer, Stockholm International Peace Research Institute SIPRI, Dr. Frank Sauer, Universität der Bundeswehr München UniBw M, and, especially, with his friend Prof. Dr. Roy Streit, Metron Inc. / University of Massachusetts-Dartmouth.

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 on Sensor Data Fusion, AI, and Resources Management. For many years, he is working for the German Ministry of Defense and the German Defense and Aerospace Industry. At Fraunhofer FKIE, he heads the Dept. "Sensor Data and Information Fusion". In his position as Chief Scientist FKIE, he also co-ordinates on a broader scale R&D activities related to digitization in 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. Of particular interest for him are ethical and legal aspects of AI and automation. He is one of the initiators and co-chair of the working group "Responsible Technology for a Future Combat Air System". In the international arena, Wolfgang Koch is active in the IEEE Aerospace and Electronic Systems Society AESS, the NATO Science and Technology Organization STO, and the International Information Fusion Society ISIF. He has organized numerous conferences and is member of editorial boards and committees.

---

[1] For a German perspective on Quantum Computing see: Christian Bauckhage et al. (2020), *QUANTUM MACHINE LEARNING. An analysis of competence, research and application*. Fraunhofer Big Data and Artificial Intelligence Alliance, September 2020, https://www.bigdata.fraunhofer.de/de/big-data/kuenstliche-intelligenz-und-maschinelles-lernen/quantum-ml.html.

[2] John E. Hyten (2020), "Remarks to the Joint Artificial Intelligence Symposium," September 9, 2020, https://www.defense.gov/Newsroom/Transcripts/Transcript/Article/2344135/remarks-by-general-john-e-hyten-to-the-joint-artificial-intelligence-symposium/. See also: John F. Kennedy (1962), "Address at Rice University, Houston, Texas, 12 September 1962," John F. Kennedy Presidential Library and Museum, https://www.jfklibrary.org/asset-viewer/archives/JFKPOF/040/JFKPOF-040-001.

[3] "Of things that exist, some exist by nature, some from other causes. [...] What is not natural is created and maintained by man through art and has no beginning in itself." In: Aristotle, *Physics*, II.1. http://classics.mit.edu/Aristotle/physics.2.ii.html.

[4] Robert Spaemann et al. (2005), *Natürliche Ziele. Geschichte und Wiederentdeckung teleologischen Denkens* [Natural Ends. History and Rediscovery of Teleological Thinking], Stuttgart: Klett-Cotta 2005, pp. 51-51.

[5] The phrase *ipsa scientia potestas est* occurs in Sir Francis Bacon's *Meditationes Sacrae* (*1597*) and was made popular in the work *Leviathan* by Thomas Hobbes (1588-1679), who was a secretary to Bacon as a young man.

[6] Yvonne Hofstetter (2018), *Theodor-Heuss-Preis 2018: Programmierte Freiheit – Spielräume für Verantwortung* [Programmed Freedom – Scope for Responsibility], 16.06.2018, p. 2, https://www.theodor-heuss-stiftung.de/wp-content/uploads/pv-2018-yvonne-hofstetter-ths-dank-redeformat-1.pdf.

[7] Friedrich von Westphalen and Yvonne Hofstetter (2017), "Der drohende Verlust der Privatautonomie des Verbrauchers [The Imminent Loss of the Consumer's Private Autonomy]," Anwaltsblatt 12/2017, S. 1174–1185.

[8] Jürgen Habermas and Josef Ratzinger (2004), *Dialektik der Säkularisierung. Über Vernunft und Religion* [Dialectic of Secularization. On Reason and Religion], ed. by Florian Schuller, Freiburg im Breisgau: Herder 2005, p. 26.

[9] Benedict XVI. (2011), *Address to the German Parliament*, Reichstag Building, Berlin, 22.09.2011, https://www.bundestag.de/parlament/geschichte/gastredner/benedict/speech, last access: August 2, 2020.

[10] Jill Long (2012), "What is War? A New Point of View," Small Wars Journal 12.05.2012. https://smallwarsjournal.com/jrnl/art/what-is-war-a-new-point-of-view.

[11] Raymond L. Burke (2007), "The Natural Moral Law: Foundation of Legal Realism," in: *Die fragile Demokratie – The Fragility of Democracy*, ed. by Anton Rauscher, Berlin: Duncker & Humblot 2007, p. 30.

[12] "That government of the people, by the people, for the people, shall not perish from the earth." Abraham Lincoln (1863), "Address at Gettysburg. Pennsylvania, 19 November 1863," in: *Abraham Lincoln: Speeches and Writings 1859-1865*, New York: The Library of America 1989, p. 536.

[13] Vincent Boulanin (2015), "Implementing Article 36 Weapon Reviews in the Light of Increasing Autonomy in Weapon Systems," SIPRI Insights on Peace and Security, No. 2015/1, November 2015, https://www.sipri.org/sites/default/files/files/insight/SIPRIInsight1501.pdf.

[14] See for example: Samule Lo Piano (2020), "Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward, " Humanit Soc Sci Commun 7, 9 (2020).

[15] An early vision of what is now called 'Hybrid Warfare', seen as "a continuum of engagement in order to limit the dissonance between a nation's will and that of other state and non-state actors", is sketched by Jill Long (2012): "War is the coherent execution of all means to bring about sufficient adherence to a nation's will in the international (global) arena; resulting in armed conflict only when all other means fail."

[16] Wolfgang Koch (2014), "Towards Cognitive Tools: Systems Engineering Aspects for Public Safety and Security," IEEE Aerospace and Electronic Systems Magazine, vol. 29, nr. 9, Sept. 2014, pp. 14-26.

[17] Ecclesia Romana (1965), *Pastoral Constitution* Gaudium et Spes, 07.12.1965. English translation from: *Vatican Council II: The Conciliar and Post Conciliar Documents*, ed. by Austin Flannery, O.P., Collegeville, Minnesota: Liturgical Press 1975, p. 981.

[18] Raymond L. Burke (2007), p. 31.

[19] "Unlike other great religions, Christianity has never proposed a revealed law to the State and to society, that is to say a juridical order derived from revelation. Instead, it has pointed to nature and reason as the true sources of law – and to the harmony of objective and subjective reason, which naturally presupposes that both spheres are rooted in the creative reason of God." In: *Address to the German Parliament*.

[20] *Address to the German Parliament*.

[21] "Conscious of their responsibility before God and man, inspired by the determination to promote world peace as an equal partner in a united Europe, the German people, in the exercise of their constituent power, have adopted this Basic Law." *Basic Law for the Federal Republic of Germany*, Preamble, https://www.gesetze-im-internet.de/englisch_gg/.

[22] Thomas Aquinas, STh I-IIae, q.94, a. 2, cited in Burke (2007), p. 40.

[23] Elon Musk (2018), *Keynote Speech*, SXSW 2018, 11.03.2018, https://www.youtube.com/watch?time_continue=7&v=kzlUyrccbos.

[24] Jill A. Long (2013), "The Problem with Precision: Managing Expectations for Air Power," Master of Strategic Studies, US Army War College, Carlisle Barracks, PA, p. 28, https://apps.dtic.mil/dtic/tr/fulltext/u2/a589415.pdf.

[25] Immanuel Kant (1784), *An answer to the question: What is enlightenment?* Transl. by Mary C. Smith, http://www.columbia.edu/acis/ets/CCREAD/etscc/kant.html.

[26] Ellen Ueberschär (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], opening speech of the national working group "Responsible Technology for Future Combat Air System (FCAS)", Bad Aying, 27.09.2019. www.fcas-forum.eu.

[27] Wolf von Baudissin (1951-1969), *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr* [Soldier for Peace. Drafts for a Contemporary Bundeswehr], München: Pieper 1969, p. 234.
[28] Ellen Ueberschär (2019).
[29] Yvonne Hofstetter (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: Droemer 2019.
[30] 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.
[31] Wolfgang Koch (2019), "FCAS – Herausforderungen für Sensordatenfusion und Ressourcenmanagement [FCAS – Challenges for Sensor Data Fusion and Resource Management]," cpm-Sonderheft *National FCAS SUMMIT 2019*, pp. 8-11.
[32] Wolf von Baudissin (1951-1969), p. 206 (1954).
[33] John R. Hoehn (2020), „Joint All Domain Command and Control (JADC2)," Congressional Research Service, August 25, 2020, https://fas.org/sgp/crs/natsec/IF11493.pdf.
[34] Frank Sauer (2017), "Kennzeichen des Krieges im 21. Jahrhundert. Entgrenzung und Beschleunigung [Characteristics of war in the 21st century. Spatial Dissolution and acceleration], Außerschulische Bildung 48 (1), pp. 4-10.
[35] See for example: International Committee of the Red Cross (2019), *Autonomy, artificial intelligence and robotics: Technical aspects of human control*, August 2019, https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control.
[36] 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.
[37] Konzeption der Bundeswehr 2018, 83f.
[38] Wolf von Baudissin (1951-1969), p. 84 (1967).
[39] Wolf von Baudissin (1951-1969), p. 242 (1953).
[40] Wolfgang Koch (2014), *Tracking and Sensor Data Fusion – Methodological Framework and Selected Applications*, Springer Mathematical Engineering Series, Heidelberg et al.: Springer, pp. 78ff.
[41] Mission-type Tactics. Wikipedia. https://en.wikipedia.org/wiki/Mission-type_tactics.
[42] Agile UAV in Networked Environment – Urban CAS, https://www.fkie.fraunhofer.de/de/forschungsabteilungen/sdf/AgileUAV. Open publication covering certain aspects: Christian Steffes et al. (2020), "Array-based Emitter Localization Using a VTOL UAV Carried Sensor," Proc. of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Karlsruhe.
[43] "The use of weapons is exclusively under human control." 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.
[44] The responsible use of new technologies in a Future Combat Air System (FCAS), www.fcas-forum.eu.
[45] Florian Keisinger and Wolfgang Koch (2020), "Verteidigung und Verantwortung – Nutzung neuer Technologien in einem 'Future Combat Air System' [Defence and Responsibility – Use of New Technologies in a 'Future Combat Air System']," Behördenspiegel, 16.04.2020, https://www.behoerden-spiegel.de/2020/04/16/nutzung-neuer-technologien-in-einem-future-combat-air-system/.
[46] Konzeption der Bundeswehr 2018, p. 50.
[47] See: *Global Warriors? German Soldiers and the Value of* Innere Führung, ed. by Heinrich Dierkes, Ethics and Armed Forces, special issue, 2016/1, p. 46, http://www.ethikundmilitaer.de/en/full-issues/2016-innere-fuehrung/.
[48] Benedict XVI. (2006), *Deus caritas est*, cited in Burke (2007), p. 43.
[49] Such as "impulsiveness, moral obtuseness, immaturity, and vice", in: Burke (2007), pp. 40-42.
[50] Konzeption der Bundeswehr 2018, p. 84.
[51] S.Th. I,q.21 a.2.
[52] "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.
[53] Ian Goodfellow et al. (2015), "Explaining and Harnessing Adversarial Examples," 3rd Int. Conf. on Learning Representations, ICLR 2015, San Diego, CA, USA.
[54] Gary Marcus (2018), "Deep Learning: A Critical Appraisal," https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf, or – more pointedly – Jason Pontin (2019), "'Brittle, Greedy, Opaque, and Shallow': The Promise and Limits of Today's Artificial Intelligence," Flagship Pioneering 26.09.2019, https://www.flagshippioneering.com/stories/brittle-greedy-opaque-and-shallow-the-promise-and-limits-of-todays-artificial-intelligence.
[55] A formulation by Felix Govaers, Fraunhofer FKIE.
[56] Militärische Luftfahrtstrategie 2016, p. 23.
[57] Carl von Clausewitz (1832), *Vom Kriege* [On War]. Hamburg [11]2018, II.3, p. 135. Translations provided by the author.
[58] Carl von Clausewitz (1832), I.6, p. 96.
[59] Carl von Clausewitz (1832), I.3, p. 71.
[60] Robert Spaemann (1996), *Personen. Versuche über den Unterschied von ‚etwas' und ‚jemand'* [Persons. Attempts on the difference between ‚something' and ‚someone'], Stuttgart: Klett-Cotta [3]2006, p. 209.
[61] Robert Spaemann (1982), "Wer hat wofür Verantwortung? Kritische Überlegungen zur Unterscheidung von Gesinnungsethik und Verantwortungsethik [Who is Responsible for what? Critical Reflections on the Distinction between Ethics of Conviction and Ethics of Responsibility]," in: *Grenzen. Zur ethischen Dimension des Handelns* [Boundaries. The Ethical Dimension of Action], Stuttgart: Klett-Cotta 2001, p. 232.
[62] Raymond L. Burke (2007), pp. 33-35.
[63] Yvonne Hofstetter, Wolfgang Koch, and Friedrich von Westphalen (2019), "Autonome Waffen. Das fünfte Gebot im KI-Krieg [Autonomous weapons. The fifth commandment in the AI war]," Spektrum.de , 05.07.2019, https://www.spektrum.de/kolumne/der-krieg-der-zukunft-wird-dank-robotern-und-kuenstlicher-intelligenz-ein-problem/1655406.
[64] *P7000 – Model Process for Addressing Ethical Concerns During System Design*. https://standards.ieee.org/project/7000.html. This working group places particular emphasis on the notion of 'values' that are to be systematically elicited, conceptualized, prioritized and finally respected via appropriate systems design [S. Spiekermann (2018), "Carousel Kittens: The Case for a Value-Based IoT," IEEE Pervasive Computing, vol. 17, no. 2, pp. 62-65, Apr.-Jun. 2018.]. The philosophical background is Material Value Ethics, first established by May Scheler (1874-1928) and Nikolai Hartmann (1882-1950). A core trait is its focus on virtue ethics, i.e. emphasis on culturally or socially desirable character traits. This design approach aims to maximize positive value potential and minimize value harms for people in IT-rich environments. We do honor the intention of this approach. Communities must share common values. This is particularly true in democracies that are only stable if the majority values rights and duties. They are based on law, however, not on moral obligation. Communities that are committed to individual freedom require observance of its laws, not the conformity with values that underlie its legal system. It may even be dangerous to speak of 'community values' because there is a tendency to undermine the legal principle in favor of a 'dictatorship of beliefs'. There have been and there are 'communities of values', where values have taken or take precedence over the law. We primarily have to talk about norms, not values only.
[65] *From Principles to Practice An interdisciplinary framework to operationalise AI ethics*, https://www.ai-ethics-impact.org/en.
[66] Collateral damage: "unintentional or incidental injury or damage to persons or objects that would not be lawful military targets in the circumstances ruling at the time. Such damage is not unlawful so long as it is not excessive in light of the overall military advantage anticipated from the attack." In: Hugh Smith (2005), "What Costs Will Democracies Bear? A Review of Popular Theories of Casualty Aversion," Armed Forces & Society, 31(2005), pp. 487-512.
[67] Jill A. Long (2013), pp. 11-12.
[68] Baudissin (1951-1969), p. 234 (1954).
[69] Hartwig von Schubert (2013). *Die Ethik rechtserhaltender Gewalt* [The ethics of law-preserving violence], WIFIS-aktuell 48, Opladen Berlin Toronto: Wissenschaftliches Forum für Internationale Sicherheit 2013, p. 49.
[70] Josef Pieper (1955). *Werkausgabe letzter Hand* [Last hand edition], vol. VII, Hamburg: Felix Meiner 2000.
[71] Wolf von Baudissin (1951-1969), p. 181 (1967).
[72] Erster Bericht zur Digitalen Transformation 2019, p. 27.
[73] German MoD (2019). *Umsetzungsstrategie Digitale Bundeswehr* [Implementation Strategy Digital Bundeswehr]. 14.06.2019, Nr. 102, p. 4. https://www.bmvg.de/de/themen/ruestung/digitalisierung/umsetzungsstrategie-digitale-bundeswehr.
[74] Umsetzungsstrategie Digitale Bundeswehr, Nr. 209, p. 8.
[75] John F. Kennedy (1962).