On Ethically Aligned Information Fusion for Defence and Security Systems

Wolfgang Koch *Fellow, IEEE*
Fraunhofer FKIE, University of Bonn
Wachtberg, Germany
w.koch@ieee.org

***Abstract*—Research on Information Fusion does not consider whether and with what capabilities defense and security systems should be developed. Nor does it have to assess the ethical dimension of their use. Nevertheless, its tasks include systematically investigating technical options and limitations of such systems, researching methods for their use and control, advising users, and finally developing the required technology. At the same time, it must assess development risks associated with this technology at the technical level and estimate future development trends. Fusion research for defense and security therefore provides a factual basis which must be taken into account in any ethical evaluation. In this context, it is also necessary to consider what is substantially new about innovative technology compared to systems already in use. As concrete examples we consider armed drones and hazardous material localization in public spaces.**

***Index Terms*—digital ethics, information fusion, artificial intelligence, technical autonomy, decision support, responsibility.**

I. Information Fusion for AI and Automation

Information Fusion is an omnipresent phenomenon. Every living creature fuses impressions of different, complementary sensory organs with previously learned knowledge and messages from other creatures. From this, it forms a mental model of its environment, the basis for situationally appropriate action. Fusion algorithms provide tools that powerfully enhance the perceptive mind and active will of users who consciously perceive and responsibly act. They are the backbones of Multiple Source Fusion Engines, see [1], for example, that transform data streams from a variety of sources along with context knowledge into situation pictures, the basis for decision making in an ever increasing range of applications. Examples are manned-unmanned teaming and platform management, use cases in manufacturing, process control, or supply chain management, in health or elderly care, as well as in defense and public security.

Situational awareness is not only basic to reaching goals efficiently, but to reaching them also in an ethically and societally acceptable and responsible way [2], [3]. For a comprehensive overview of ongoing ethical debates see [4] and the IEEE P7000 Model Process for Addressing Ethical Concerns During System Design [5].

Fusion engines form the core of artificially intelligent and technically automated systems that assist actors to acquire knowledge about options for action in various operational theaters. We therefore use the term Artificial Intelligence in a broader sense, which includes Machine Learning and Neural Networks as special examples of the more comprehensive family of data fusion algorithms. They help to master complex tasks more adequately, to balance human subjectivities, and to protect uninvolved persons. This requires 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, and
- to guarantee ethical, legal, and societal compliance.

A. On the Problems of Armed Drones and Dirty Bombs

Any ethical evaluation of armed drones, for example, leads to the more profound question of whether and to what extent stand-off weapons can responsibly be used, and this always means whether their use can be justified by *persons*. A precursor to the current discussion may be seen in the outlawing of the crossbow, which, however, had hardly any practical consequences. The ethical quality of the use of armed drones would therefore, despite significant differences in detail, be analogous to the use of weapons such as slingshot, crossbow, rifle, artillery, machine gun, bomb, rocket, cruise missile etc.

In this view, public debates on armed drones, for example in Germany, rather fundamentally problematize the use of military force than to critically examine a novel technology of long-range weapons from an ethical point of view. In this sense, a former German Minister of Defense spoke of a catch-up debate: “The ethical core question does not seem which weapon is used, but the question of the legitimization of the military use of weapons as such [6].” Research on Information Fusion can contribute to objectifying this discussion by answering a question that it is actually entitled to be answered clearly: What are the technical fundamentals of armed drones, on the basis of which the conditions for its ethically responsible use can be discussed? Apparently, the quality and trustworthiness of instrumental knowledge, as obtained by fusion engines, has a key position to answer the question of responsible action.

The Nuclear Security Summit 2016 named radiological terrorism as one of the greatest challenges to international security, which is constantly growing [7]. In this example, a conventional explosion, a ‘dirty’ bomb, releases radioactive isotopes to contaminate hot spots of public life. Experts speak

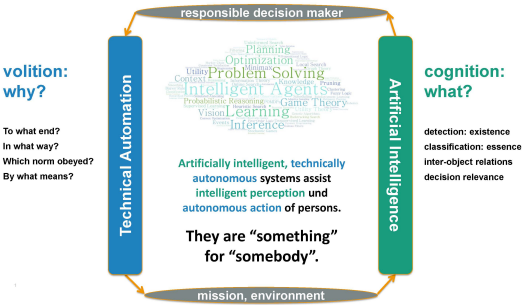

Fig. 1. Cognitive and volitive assistance for responsible decision-makers.

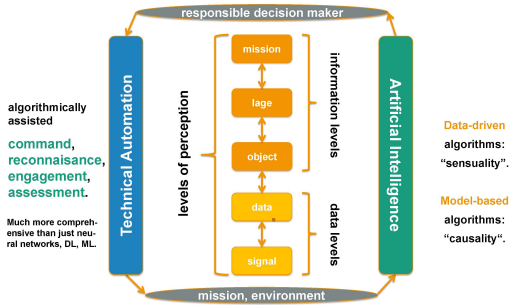

Fig. 2. Levels of perception for algorithmically assisted decision making.

of Improvised Radiological Dispersion Devices (IRDD). Such material is used in hospitals, research centers and industrial plants in almost every country. According to the Nuclear Threat Initiative, many installations of this type are considered poorly secured and prone to theft. In addition to the actual damage and injuries caused by radiological bombs, the health, psychosocial and economic consequences are long and extensive (Weapons of Mass Disruption).

How to identify in a flow of people a nuclear terrorist without violating informational self-determination of the vast majority? Biometric recording should only be carried out after there is sufficient suspicion. To this end, assistance systems are needed that direct the attention of security personnel to potential radiological hazards. Since many substances suitable for 'dirty' bombs emit γ-radiation, miniaturized spectrometer provide data on the existence of a 'dirty' bomb, the type of material, and its intensity. Assigning this signature to a specific person, however, is only possible via a multiple sensor fusion. Distributed γ-sensors are networked with cameras, thus enabling a spatio-temporal fusion of the data stream generated by a flow of people. How to design such fusion engines in a societally acceptable way in terms of privacy?

B. Cognitive and Volutive Assistance

The previous examples pose a timeless question: How to comprehensively guarantee ethical, legal, and societal compliance; how to decide 'well' according to what is recognized as 'true'? This breaks downs to two engineering questions:

1) How to design *cognitive* tools that we mentally and emotionally are always mastering?
2) Which design principles facilitate the responsible use of artificially intelligent *volitive* tools?

As illustrated by Fig. 1, AI-based automated systems cognitively assist the perceiving minds of personally responsible decision makers in understanding complex situations and volitively support the enforcement of their will in terms of appropriate and responsible action.

Decisive for cognitive assistance is the question of 'what' is to be recognized and represented by situation pictures. 'Detection' informs about the existence of relevant objects and phenomena, 'classification' about their properties, i.e. their essence. Important are inferred object interrelations. Finally, situation pictures indicate relevance, such as threat levels and the state of own resources. The situation picture as well as statements about its limitations and gaps must correspond to the actual situation. This implies the concept of 'truth' according to its classical definition: "Truth consists in the equivalence between the situation picture and the situation." We may distinguish between the 'logical truth' of the situation picture and its 'ergonomic truth', in that it corresponds to the tasks, roles, and abilities of decision makers [8].

Automation translates the intentions of decision makers into complex cause-effect chains to manage the available resources. The question of 'why' to achieve an effect is crucial for algorithm design. We may distinguish four ways of answering to why-questions. The goals correspond to the *final cause*, usually specified by performance parameters. The *efficient cause* indicates which concrete algorithms are used to achieve them. The *formal cause* answers the question, according to which rules this happens. Finally, the *material cause* indicates which means are to be used with their respective properties.

In general, we distinguish data-driven from model-based algorithms. The first family, e.g. Deep Learning, corresponds to intuitive perception – What do I see? The second family, in the sense of Bayesian reasoning, enables causal reasoning for rational action – What shall I do? In the information processing circle in Fig. 2, we distinguish five levels of perception. The first two of them, determined by received signals and signal processing, are summarized as *data levels*. The three *information levels* refer to the individual objects, to the situation with information about the interaction of objects, and to the mission, which represents both, a situation vignette and the decision maker who wants to act in it.

C. Contribution and Structure

In this paper, we present ethically relevant aspects of Information Fusion, Artificial Intelligence, and Technical Automation, i.e. resources management, from a systems engineering perspective for applications in defense and public security. If we were able to solve the ethical problems here, new pathways will open up for a wider use of digital technologies. After discussing examples of ethically aligned engineering for armed drones and radiological material detection, we sketch core elements that characterize the notion of responsibility. On this basis, we derive technical prerequisites of responsible systems design, discuss selected aspects of moral assistance,

and comment on norm-based versus value-driven engineering. This discussion leads to the recommendation to explicitly implementing ethical aspects in the various stages of strategic planning. In the appendix, we discuss the difference between 'natural beings' and 'artificial things' from a philosophical perspective. A clear understanding of this fundamental dichotomy seems essential for any ethically aligned engineering.

II. Examples of Ethically Aligned Engineering

Essentially, there are two aspects of drone technology that are innovative and by no means of purely military interest.

1) It uses a mature technology for the management and end-to-end control of mobile platforms and the communication, sensor, and effector systems on board.
2) It uses high performance technology to fuse vast streams of heterogeneous sensor data with each other and with all available context information.

We distinguish fixed-wing from rotary-wing drones. Drones of the first category generally have longer standing times and carry larger payloads than those of the second category, which can be used much more flexibly, e.g., by vertically take-off and landing from a military vehicle and hovering at a specific location for more precise situation assessments.

For military drone operations, a distinction is made between sensors for the reliable operation of the drones and sensors for situation assessment. The first category includes navigation sensors for platform control and avoiding collisions in airspace, as well as sensors for observing the internal status of the subsystems. The second category includes signal detection and imaging sensors, e.g. for detecting, classifying, localizing and tracking of communications, radar, and acoustic signals, such as shots. Laser warners indicate enemy target acquisition systems. In order to meet the resulting requirements, AI-based automation as previously sketched makes drones technically controllable, the basis for their responsible use.

A. Drone Support for Forward Air Controllers

In Fig. 3, a convoy is stopped in urban environment by an Improvised Explosive Device (IED) and attacked. Coordinately operating drones make it possible to estimate the expected collateral damage and to provide the Forward Air Controller (FAC) with a comprehensive, situation picture. The example illustrates, in which way drone technology can in principle provide technical prerequisites for responsible use of long-range weapons. The thesis claimed here is: Armed drones in principle enable reliable or, compared to other weapon systems, significantly more reliable target reconnaissance and weapon control up to the final engagement decision. This is the *technical* prerequisite for their responsible use with minimized risk for uninvolved people. So-called 'fire-and-forget weapons' with multiple sensor seekers have long been available. It would be a perfectly legitimate question whether these weapons should not be replaced those that enable control until the weapon effect is achieved.

An important source of information for pre-engagement situation analysis are so-called Rules of Engagement (RoE).

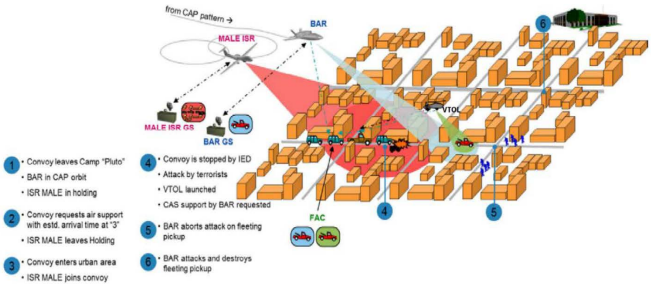

Fig. 3. Coordinated drones assisting FAC decisions aligned with RoE

They are to be kept up-to-date and to be taken into account right down to the design of information fusion algorithms. RoE do not make any tactical specifications, but rather define a legally binding and mission-specific framework of action. In accordance with legal, political, military strategic, and operational requirements, they concretize the *ius in bello*.

We do not expect that technological development will "naturally" lead to responsible stand-off weapon systems. Moreover the informational basis for engagement decisions can always be enhanced. Even the development of irresponsible drone technology is quite possible and may even be pursued by enemies. This does imply that the ethical development of this technology should actively be encouraged and technically supported. This includes the conception of well-thought-out Rules of Engagement that take into account the risks of this technology, which will permeate all technical system components. In short, 'responsible design' should mean developing armed drones in such a way that it closes the existing protection gaps in the sense described.

B. Multisensor Detection of Radiological Threats

Multiple sensor drones are also suitable means to complement military situation pictures by identifying and localizing radiological threats in an operational theater. Since security of public life is a basic human desire and a fundamental prerequisite of liberal societies, however, such threats have to be countered also here. In other words, we are considering the dual problem, i.e. a network of stationary sensors and a flow of people in which a carrier of radiological material may be moving [11].

This rises a crucial question: How can public security be improved by ethically and legally aligned as well as societally acceptable surveillance systems in public spaces? Let us consider the following goals:

1) People tracking provides a temporal basis for exploiting spatially distributed γ-sensors, i.e. enhanced resolution.
2) Before a clear hazard-to-suspect association is made, no biometric parameters for identification are collected.
3) Since this space-time approach is basically 'blind' to uninvolved people, 'normal' public life may be regained.

Perhaps rather unexpectedly, *indistinguishable* target tracking [12] seems to play a key role seen from a systems engineering

perspective, where the problem of reconciling the values of greater security with the values of personal dignity and privacy that an individual foregoes is to be solved. The idea is based on non-classifying sensors, where the set Z_k of measurements at each time t_k does not reveal the identity of n persons characterized by their kinematic state vectors $\mathbf{x}_k^{1:n}$. Mathematically speaking, the likelihood function of non-classifying sensors obey a symmetry property with respect to permutations $\sigma \in S_n$ of the indices of the kinematic states:

$$\forall \sigma \in S_n : \quad \ell(\mathbf{x}_k^{1:n}; Z_k) = \ell(\mathbf{x}_k^{\sigma(1:n)}; Z_k). \tag{1}$$

Under rather general assumptions, this symmetry property is transferred to the conditional probability densities given the accumulated sensor data $Z_{k:1}$ that describe the available knowledge on the kinematic people states:

$$\forall \sigma \in S_n : \quad p(\mathbf{x}_k^{1:n}|Z_{k:1}) = p(\mathbf{x}_l^{\sigma(1:n)}|Z_{k:1}). \tag{2}$$

As a more detailed analysis shows, we may even distinguish between "fermionic" and "bosonic" tracking in analogy to multiple particle quantum physics [12].

In the context of public surveillance, this tracking approach guarantees "Indistinguishability of the Uninvolved", a principle which seems to play the role of a quite fundamental, and even *certifyable* systems design principle. By considering persons to be tracked as indistinguishable targets, such security systems will be able to preserve the anonymity of the vast majority of persons until a certain level of suspicion is reached which may finally justify the identification of an individual.

This research was part of a project, which investigated the vulnerability of the transnational high-speed train systems [13]. While maintaining an open transport concept as far as possible, an analysis of the infrastructure usually available in and around railway stations shows that there are always areas suitable for continuous radiological monitoring. For further details see [14]. Fig. 4 shows an experimental system realized in this project, where persons are walking around γ-sensors. In practical realizations these sensors may well be hidden in the walls or in the floor. The association of positive γ-signatures to an individual and its tracking over time is produced by Track-while-Classify (TwC) as described in [15]. Indistinguishable target tracking is essential in TwC that here treats persons as 'fermionic' targets. The benefit of indistinguishable target tracking in public security lies less in the fact that "better" tracks are produced. In a crowd of people, however, 'fermionic' trackers may also provide a certain gain in terms of track continuity, since the well known phenomenon of track coalescence is mitigated.

III. Responsibility in Digital Systems Engineering

The notion of responsible use of digital technologies realizes a connection between human cognition and volition that leads to action in the real world on the one hand and automatically running processes in the world of algorithms on the other. It therefore seems to 'fuse' two realms of knowledge that are in need to be properly balanced, i.e. instrumental knowledge according to the famous statement of Francis Bacon "for knowledge itself is power" [9] and the "Ecology of Man", i.e. knowledge of the human nature [10].

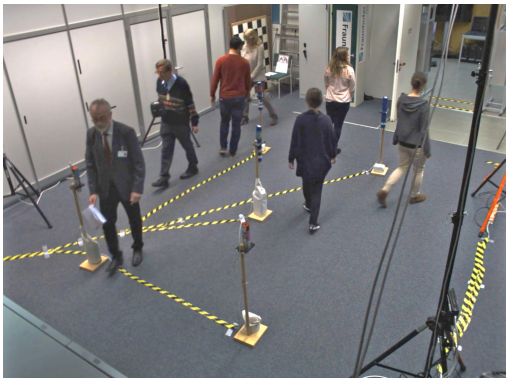

Fig. 4. IRDD localization in person streams with γ-sensors and cameras.

The word 'responsibility" is rooted in the language at Courts of Justice. A responsible person may be called upon to 'respond' to questions about his or her actions by a judge. 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 is 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 and philosophical backgrounds. Nevertheless, there seems to be a broader consensus on the following aspects.

1) To speak of responsibility is only reasonable if it is assumed voluntarily. Its very notion thus presupposes '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 International Law.
3) Responsibility also implies the 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 'understanding'. The 'true' and the 'good' thus form the intellectual basis of responsible action.

Fig. 5 illustrates core elements of responsibility, insofar as it is relevant to the technical design of cognitive and volitive assistance systems for defense and security. It implies three persons or groups of persons and characteristic relationships between them.

1) *Who bears responsibility?* Any capability development for defense and security takes place at various levels and requires responsible action in research, development, certification, and qualification of command & control, surveillance and weapon systems as well as in the preparation and execution of 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

Fig. 5. Elements of responsibility and the resulting mutual relationships.

determined by prospective action and reaction. Everyone is primarily responsible himself. Secondly, responsibility is owed to own forces, combatants or civilians. Improperly, 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.

In the end, only voluntarily assumed responsibility, which shows itself in care and trust and is ready to justify itself, keeps human societies and relationships stable, even on the battlefield. Purely legal constructs, such as liability for one's actions, are not sufficient, especially in military operations. Only persons, who use cognitive and volitive assistance responsibly or irresponsibly act 'good' or 'evil' by responding to moral challenges in the one way or the another. 'Good' technical systems encourage the morally acceptable and efficient use of them to achieve defense or security objectives. 'Evil' systems facilitate or even encourage their irresponsible use.

A. Prerequisites of Responsible Systems Design

A serious challenge for comprehensive and consistent controllability in digital defense and security systems is certainly the ever-decreasing time available for human-involved decision making. A further problem is the limited explainability and deceivability of algorithmically generated information.

From an abstract point of view, neural networks assign an input to an output that states what the input should 'mean' for the user. Characteristic of these 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'. 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 tiny details in the input need to be changed in a specific way to completely mislead even a well-trained network. Deceived by "poisonous noise", it may 'recognize' a panda bear, which appears unchanged to humans, as a gibbon monkey and 'feels' certain about it [16]. The military relevance of this discovery is obvious. Attack systems against AI systems are already under development, own AI systems are to be hardened against such "adversarial attacks." In addition, for appropriate training of data-driven algorithms no sufficient amount of representative training data are available in many military applications. Moreover, neural networks are 'black boxes' - they provide correlation only and no "tell me why". Furthermore, context knowledge – fundamental to every military mission – can only be learned indirectly, i.e. from the data. In short: Neural Networks are greedy, brittle, and opaque, always the 'second-best' solution. For critical functions, meaningful human control is required.

Model-based algorithms, on the other hand, allow logical reasoning according to the Bayesian paradigm also in case of uncertainty. They uncover probable cause-effect chains, can be developed systematically, and enable the explicit integration of context and expert knowledge. 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 considerations, the following aspects need to be addressed by ethically aligned systems design:

1) Any responsible use of technology requires comprehensive and consistent controllability. In some applications, occasional malfunction of AI and automation may have no consequences. In defense and security, however, rigorous safety requirements must be guaranteed with all legal consequences. The use of technically incontrollable technology is immoral *per se*.
2) The notion of *meaningful human control*, on the other hand, needs to be interpreted more broadly than the related concept of *human-in/on-the-loop* suggests. *Accountable responsibility* seems to be the more fundamental concept, since the use of fully automated effectors may well be justifiable, even necessary, in certain well-defined situations.
3) Certification and qualification are key issues. Robust AI-driven automated systems will comprise both, data-driven and model-based algorithms, where former are 'enclosed' by the latter. 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 is to be guaranteed *by design*.
4) Sensor and context data never meet ideal expectations. They are always imperfect, inaccurate, ambiguous, unresolved, false, corrupted or deceptive, difficult to formalize, or partly even 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 a fundamental requirement to any use of AI-based systems: Are valid and representative sensor and context data available at all? Are they produced reliably and do the unavoidable deficits correspond to the underlying statistical models? In naive systems, violated data integrity easily turns data fusion into *confusion*, resources management into *mismanagement*.
6) Finally, artificially intelligent algorithms of Information Fusion almost always generate artifacts that do not exist in reality, or have 'blind spots', i.e. do not show what is actually there. 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 if necessary.

Artificially intelligent 'self-criticism' of technical systems requires naturally intelligent critical capabilities of decision makers towards AI. Otherwise, there is a danger of voluntary subordination and uncritical acceptance of machine offers, of 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 questions rise from these considerations. "All thinking is art," observes Carl von Clausewitz, the 19th century 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." [17] Digitization in defense and security thus 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 teachable [18].

B. Selected Aspects of Moral Assistance Systems

For ethically aligned cognitive and volitive assistance systems, which technically support responsible behavior, three major requirements result from the previous considerations:

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

These requirements are basic 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 be tolerated or not. Fig.6 illustrates their impact on the development of assistance systems for responsible action.

1) Transparent development of criteria must accompany any capability development for defense and security from the outset. Philosophers, pastors, and lawyers bring in basic insights. Legal standards that apply to defense research, development, and procurement are indispensable. Finally yet importantly, the experience of soldiers and security personnel must be taken into account. Analogous to industrial quality assurance, these considerations support responsible action not only in battle, but at all levels of responsibility. 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 [5].

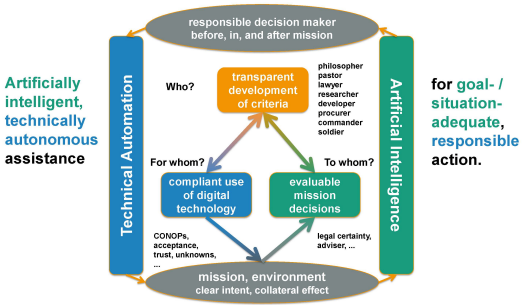

Fig. 6. Transparent criteria development for technology and applications.

2) Any technology that complies with these criteria must be integrated into 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 digitization in defense and security are only realizable if it takes into account how operational forces think about and handle technology or how certification and qualification bodies work.
3) Decisions can be evaluated and correspond to the mission-specific Rules of Engagement (RoE), which have an impact deep into the information technology design. 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" [19] are indispensable for identifying RoE-compliant options for action. In the spatially delimited and accelerated hyperwar, ethical and legal knowledge itself must be made accessible through digitization.

In order to realize the potential of Information Fusion 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 [20]. Without reliable knowledge about the

limits of the available knowledge, no one can act responsibly. A central aspect is therefore ergonomic representation of the situation pictures, which have an ethical dimension in that it must convey to the decision maker also psychological awareness of the reality of the situation represented, encourage him to take responsibility, and allow him to experience the consequences of his actions and omissions.

A more recent study emphasizes the rationality of ethical judgement, for which algorithmic support is 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" [21]. A digital assistant for "moral decision support" would have to implement such reasoning. It should also be considered how technical design principles could be derived from classical virtues, which appear under different names in many cultures, in order to make them user-compliant in the sense of responsible judgement. The so-called four "cardinal virtues" of Western ethics [22] 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. A biologistic Image of Man as a cybernetic stimulus-response machine or poststructuralist thinking that postulates the "death of the subject" [23], for example, are ruled out.

C. Comments on Norm-based vs. Value-driven Engineering

The IEEE P7000 Working Group aims to establish a process model for addressing ethical concerns during system design [5]. Particular emphasis is placed on notion of "values" that are to be systematically elicited, conceptualized, prioritized and finally respected via appropriate systems design [24]. 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.

"The sharpest weapon of democracy is legislation. Therefore, civil society cannot help but call on its governments to establish globally binding standards for cognitive systems and to make the corresponding agreements under international law [25]." International law requires that human responsibility be demanded. However, since law circumscribes an ethical minimum only, the obligation of IT companies remains. It therefore seems reasonable to use the proven concept of Corporate Social Responsibility (CSR). The principles to be anchored in this way should aim to create binding standards of conduct for all actors in the responsible use of cognitive and volitive systems. The aim is to establish a system of standards, including the entire supply chain. Tangible sanctions should be introduced in the event of misconduct – *soft law and hard sanctions*. The instruments range from extraordinary contract termination to damages and contractual penalties. Although soft law is not enacted by the legislature, it may be "hardened", e.g., if the courts are using it as a source of legal knowledge. We primarily have to talk about norms, not values only.

IV. Responsible Systems Design: Recommendations

The future of digitization in defense and security does not choose between man and AI, but lies in a scalable combination of man and AI to ensure the best possible performance of tasks. This includes an ethical dimension in digital systems engineering. Since we feel that there might be a broader consent within the Information Fusion community to these considerations, we are closing with some recommendations.

1) Ethically aligned systems design is a fundamental capability that we need to develop systematically in order to be able to use digital technologies in such a way that harm for humanity is prevented. In particular, consideration should be given to systematically develop ethical competence along with technological progress in defense and security at all stages.
2) In addition to their operational benefit in closing capability gaps, expanding the range of capabilities, and developing corresponding concepts, procedures, and organizational measures, ethical competence of military and police forces in dealing with digital technologies as well as personal and societal acceptance needs to be achieved. This would enable successful innovation in defense and security.
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 Technical Automation based on Information Fusion would hardly be politically and socially enforceable.

Appendix: What means 'Natural' and 'Artificial'?

Anyone involved in digitization is talking about technology. The Association of German Engineers VDI defines technology as "a set of utility-oriented, artificial, objective systems." [26] Since utility is the main focus, technology implies persons who use it. According to this definition, technical systems are "objective systems", i.e. objects, not subjects. The attribute 'artificial' is decisive. Whoever wants to know what is meant by 'artificial' must understand what 'natural' is. For the

pair of terms *natural vs. artificial* stands as a fundamental dichotomy at the beginning of Western thinking [27]. The author of this paper believes that the fundamental dichotomy between 'natural' and 'artificial' has to be clear for researchers, developers, procurers, and users of cognitive and volitive assistance systems, which are 'artificial things'.

We are using the language of ancient Greek philosophy when we speak of 'technology'. For the Greek word for 'artificially created" is *technē*. Perhaps the thinking of German philosopher Robert Spaemann (1927-2018) offers a key: "*Artificial* things [especially digital technologies], according to Aristotle, are indeed characterized by the fact that they themselves consist of a 'what' and a 'what of'. 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 'wherefore' in itself fall into one. The purpose is the form of the thing itself, hence the notion of *entelecheia*: I carry the purpose within me." [28] What are the consequences if a decision-maker in defence and security is seen as an entelechy in this sense? Is there a mental bridge between the thought "I carry the goal within me" and the principle "leading by mission"? For the philosopher Josef Pieper (1904-1997), *entelechia* is "an original idea of Western metaphysics in general: the idea of *entelechia* as the inner *telos*. This inner purpose, the immanent goal of every thing is its own essence, its inner form. To be 'complete' means: to be completely this essence. To 'become' means: realization of this form, the realization of the sense and inner purpose. Therefore, renunciation of this inner purpose is the opposite of becoming, i.e. 'corruption'; therefore, return to it is 'recovery'." [22, vol. EB, p. 7]

Technical things, created by the 'art' of Information Fusion, Artificial Intelligence, and Technical Automation, are according to this strand of thought by definition and in contrast to human beings NOT *entelechies*. This point of view contradicts to transhumanistic ideologies that consider 'conscious perception', 'free will' and 'responsibility' as artifacts of suboptimal information processors called 'humans' to be replaced by presumably 'objective' and 'unbiased' AIs.

References

[1] W. Koch, *Tracking and Sensor Data Fusion. Methodological Framework and Selected Applications*. Springer: Mathematical Engineering Series, 2014.

[2] W. Koch, "Zur Ethik der wehrtechnischen Digitalisierung. Informations- und ingenieurwissenschaftliche Aspekte" [On the Ethics of Military Digitization. Computer Science and Engineering Aspects], in: M. Rogg, S. Scheidt, H. von Schubert (Eds.), *Ethische Herausforderungen digitalen Wandels in bewaffneten Konflikten*, Hamburg 2020, 17-54.

[3] W. Koch, "Towards cognitive tools: Systems engineering aspects for public safety and security," in: IEEE AESS Magazine, vol. 29, no. 9, Sept. 2014, 14–26.

[4] A. Jobin, M. Ienca, E. Vayena, *The global landscape of AI ethics guidelines*. Nat Mach Intell 1, 389–399 (2019).

[5] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition. IEEE, 2019.

[6] Ch. Strack (2013). Bundesverteidigungsminister wirbt für Drohnen-Einsatz. Deutsche Welle, 24. April 2013.

[7] *Nuclear Security Summit*, Washington 2016.

[8] "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." In: Thomas Aquinas, S.Th. I, q.21 a.2.

[9] F. Bacon. *Scientia potentia est*. https://en.wikipedia.org/wiki/Scientia_potentia_est, last access March 12, 2020.

[10] "There is also an ecology of man. Man too has a nature that he must respect and that he cannot manipulate at will." In: Benedict XVI., Speech in the German Parliament on September 22, 2011. http://w2.vatican.va/content/benedict-xvi/en/speeches/2011/september/documents/hf_ben-xvi_spe_20110922_reichstag-berlin.html, last access March 12, 2020.

[11] W. Koch. "On Detecting Radiological Bombs with Potential Applications to Field Camp and Soldier Protection," in: Proc. 4th Int. Symp. Development of CBRN Protection Capabilities, Sept. 2017, Berlin, Germany.

[12] W. Koch, "On Indistinguishability and Antisymmetry Properties in Multiple Target Tracking," in: J. Adv. Inf. Fusion JAIF, vol. 4, nr. 2, Dec. 2019, 199-212.

[13] REsilience of the Franco-German High Speed TRAIN Network RE(H)STRAIN. http://rehstrain.w3.rz.unibw-muenchen.de, last access March 12, 2020.

[14] F. Govaers, T. Fiolka, J. Heinskill, J. Biermann, W. Koch, "A Detection System for Dirty Bombs in Open Environments," in: Proc. of the Transport Research Arena (TRA), Vienna, 2018.

[15] F. Govaers, "A classify-while-track approach using dynamical tensors," in: Proc. of the 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Curacao, 2017.

[16] I. Goodfellow et al., "Explaining and Harnessing Adversarial Examples," in: Proc. 3rd Int. Conf. on Learning Representations, ICLR 2015, San Diego, CA, USA.

[17] C. von Clausewitz, *Vom Kriege* [On War]. Hamburg 2018, II.3, p. 135.

[18] W. Koch et al., *Artificial Intelligence for Military ISR Decision Makers*, NATO STO Lecture Series SET-290, Hamburg, Rome, Budapest, Fall 2020.

[19] M. Wunder et al., *Synthetic Legal Adviser – AI-based Decision Making in Hyperwar*, NATO STO Research Task Group IST-HFM-182.

[20] "We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – there are things we do not know we don't know." In: *DoD News Briefing - Secretary Rumsfeld and Gen. Myers*, February 12, 2002. https://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636, last access March 12, 2020.

[21] H. von Schubert, *Die Ethik rechtserhaltender Gewalt* [The ethics of law-preserving violence], WIFIS-aktuell 48. Opladen et al. 2013, 49.

[22] J. Pieper, *Werkausgabe letzter Hand*, Hamburg, vol. VII.

[23] "There is no need to get particularly upset about the end of man; it is only a special case or, if you like, one of the visible forms of a much more general dying. By this I do not mean the death of God, but the death of the subject, the subject as the origin and basis of knowledge, freedom, language and history." In: Michel Foucault, *Dits et Ecrits*, Frankfurt am Main 2003, vol. I, 1002.

[24] S. Spiekermann, "Carousel Kittens: The Case for a Value-Based IoT," in: IEEE Pervasive Computing, vol. 17, no. 2, pp. 62-65, Apr.-Jun. 2018.

[25] Y. Hofstetter, W. Koch, F. von Westphalen, Friedrich, "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, last access on March 12, 2020.

[26] *Technology Assessment – Concepts and foundations*, VDI Richtlinie 3780, Sept. 2000. https://www.vdi.de/richtlinien/details/vdi-3780-technikbewertung-begriffe-und-grundlage2, last access March 12, 2020.

[27] "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, last access March 12, 2020

[28] R. Spaemann et al., *Natürliche Ziele. Geschichte und Wiederentdeckung teleologischen Denkens* [Natural purposes. History and rediscovery of teleological thinking], Stuttgart 2005, 51ff.