SciencesPo

CHAIR DIGITAL, GOVERNANCE AND SOVEREIGNTY

Autonomous Systems of Normative Control in Military Applications of AI: Towards an Ethically and Legally responsible Industry Standard for Military Technology *Made in Europe*

Dr. Johannes Thumfart
Senior and Postdoctoral Researcher Research Group Law, Science, Technology, and Society (LSTS)
Vrije Universiteit Brussel

November 2023

"States should consider how to use military AI capabilities to enhance their implementation of international humanitarian law and to improve the protection of civilians and civilian objects in armed conflict."

***Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy*, February 16, 2023.**

Table of Contents

Executive summary

This policy brief addresses military applications of AI in the sense of partially autonomous lethal weapon systems (PALWS) and logistical AI units. The systems that I call 'autonomous systems of normative control' (ASNCs) are comparable to intelligent speed assistance (ISA)-systems in cars. ISA-systems alert or correct drivers when exceeding the speed limit using road-sign recognition and speed-limit databases linked to geoposition data. Correspondingly, ASNCs should block the unlawful use of military applications of AI, for instance, in the case of a war of aggression or alert commanders if an action is disproportionate or a selected target civilian.

I promote a technology-centered approach, which is in line with the multilateral 2023 *Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy* and the technical recommendations in the report of the 2023 session of the *UN Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*. I argue that PALWS and logistical AI units in the military should be equipped with ASNCs to contribute to ensuring that they are used in compliance with international humanitarian law (IHL), most importantly the principles of proportionality and harm minimization. Furthermore, ASNCs should include blocking mechanisms to contribute to ensuring that PALWS and logistical AI units are neither used in wars of aggression nor against domestic peaceful protesters. In a technological sense, ASNCs likely require a hybrid approach to AI systems, combining data-driven and rule-based elements and much simpler blocking mechanisms based on geolocation data. Whilst it is not possible to outsource moral or legal responsibility to machines, it is plausible that ASNCs contribute to making military decision-making on the battlefield more responsible in a legal and ethical sense.

In parallel with this technology-centered approach, national and international attempts to regulate military applications of AI should be pursued further. However, the development of ASNCs does not necessarily constitute a reaction to governmental regulations but could also voluntarily be advanced as de facto industry standard by producers of military technology. Rather than refraining from the production of PALWS and logistical AI units for the military, European producers of military technology should aim at leading in research and development and establishing a standard *made in Europe*, including ASNCs that contribute to guaranteeing their use within the boundaries of legal and ethical principles. At the same time, it must already

be warned that these systems should not be abused for *'ASNC-washing'* to justify arms exports to authoritarian regimes and that the establishment of a de facto standard can only constitute one element within a broader toolkit of measures to regulate the military use of AI.

1. Introduction

The military use of AI systems in partially autonomous lethal weapon systems (PALWS) is no speculative future scenario. Reportedly, AI drones that autonomously select and attack targets are being used by Ukraine in the ongoing war against Russia [1]. Already in 2021, UN experts reported the deployment of Turkey-manufactured fully autonomous drones with lethal capacities in Libya [2]. And in 2022, Israeli Ioitering munition drones contributed to Azerbaijan's superiority in the long-standing Nagorno-Karabakh conflict [3]. In the aftermath of this impressive demonstration of loitering munition drones's capacities, France and Germany accelerated different pathways toward acquiring and developing this weapon type [4]. A high degree of autonomy is also already standard since long in projectile interception systems such as the US Patriot and Phalanx and the German MANTIS, which are required to react more rapidly and precisely than human operators are capable of. Furthermore, after the US Defense Advanced Research Projects Agency (DARPA) demonstrated the superiority of AI systems in flight simulators in the AlphaDogfight program, it is now testing the technology in physical F-16 fighter jets [5]. But the true advantage of AI systems in the military might not even lie in such scenarios that still resemble the somewhat quaint imaginary of 'killer robots'. Following the Pentagon's Joint All-Domain Command and Control (JADC2) vision, logistical AI units in the military might lead to a far-reaching cross-domain platformization of war that has been compared to the workings of the mobility platform Uber [6]. (See fig. 1 and 2 next page.) Not only for military reasons, the integration of AI systems in the military comes close to being an inevitable development: particularly the armed forces in highly developed nations are disproportionately affected by the general shortage of skilled human labor.

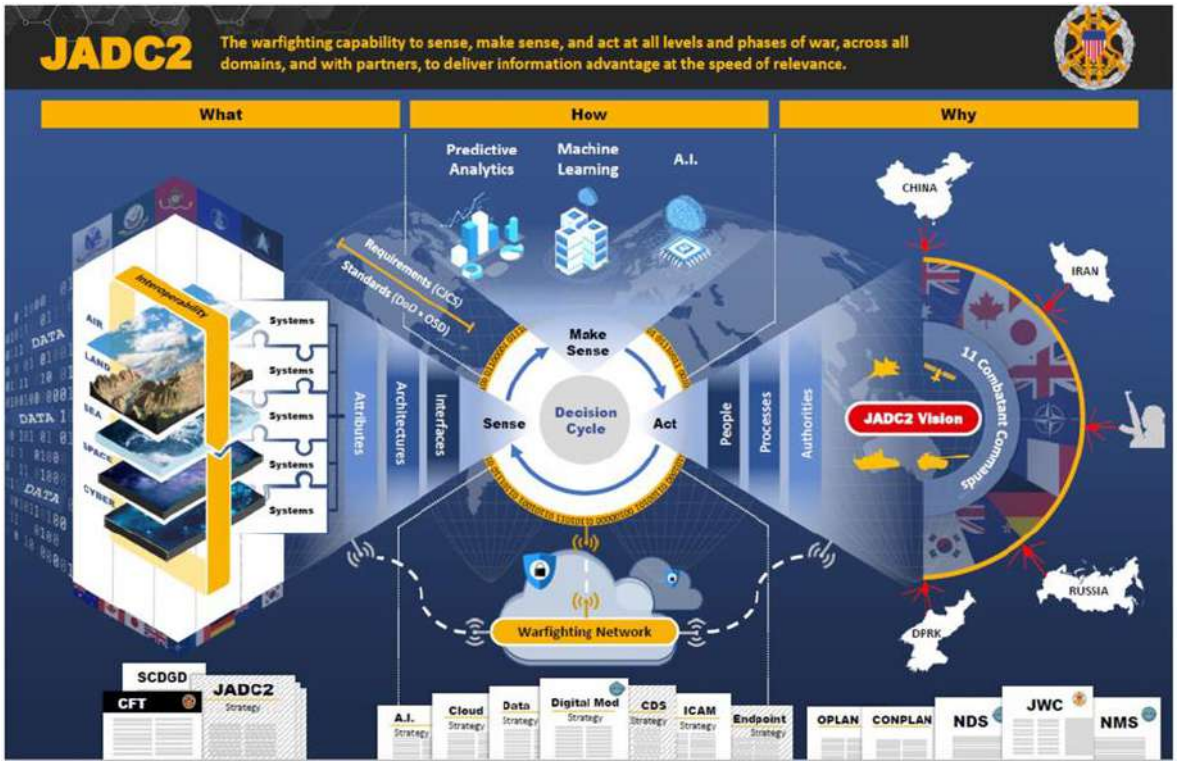

Fig. 1: JADC2 Concept emphasizing cross-domain interoperability; Source: 'Summary of the Joint All-Domain Command & Control (JADC2) Strategy' (US Department of Defense, March 2022), https://media.defense.gov/2022/Mar/17/2002958406/-1/-1/1/SUMMARY-OF-THE-JOINT-ALL-DOMAIN-COMMAND-AND-CONTROL-STRATEGY.PDF.

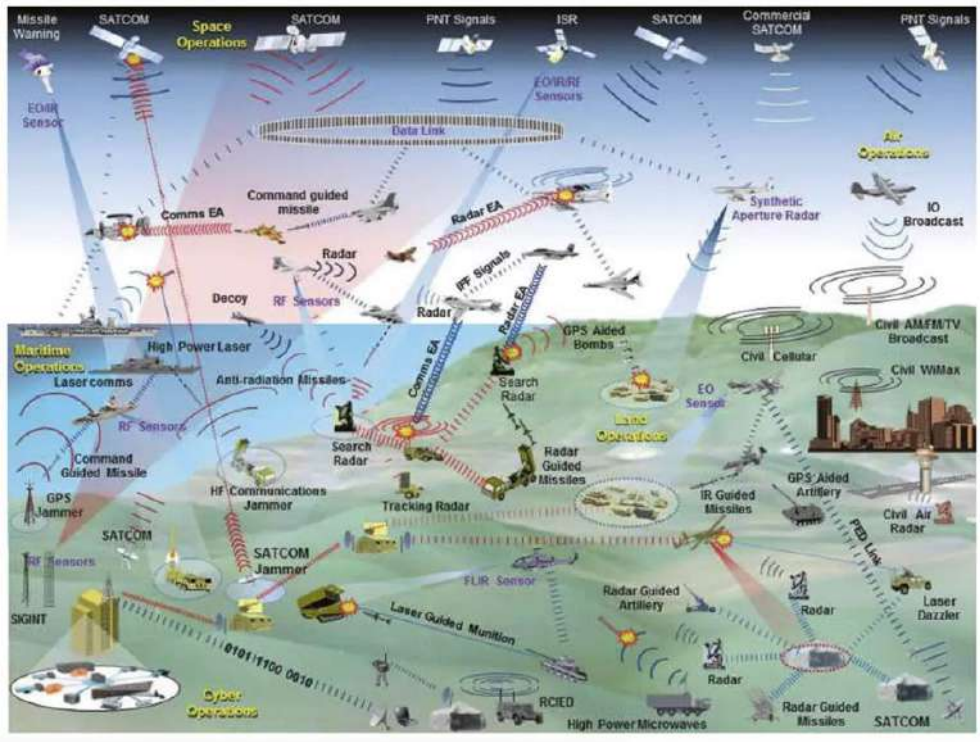

Fig. 2: Cross-Domain digital communication, intelligence, and logistics as envisioned by NATO. Source: 'Electronic Warfare – The Forgotten Discipline'. NATO Joint Air Power Competence Center, 2 December 2018. https://www.japcc.org/articles/electronic-warfare-the-forgotten-discipline/.

Especially the use of PALWS (Partially autonomous lethal weapon systems) and LAWS (Lethal autonomous weapon systems) raises profound moral, ethical, and legal problems. These have been summarized in the problem of the "accountability gap" [7] that might result from automated decision-making in warfare and the complementary demand to keep PALWS under "meaningful human control" [8]. Respective regulations regarding the military use of AI are emerging but have not reached a robust state yet. Several national governments are gradually developing national frameworks regarding the integration of AI in the military. For example, the official position of France is based on the French defense ethic committee's opinion on PALWS from 2021 [9]. The UK published its strategy on "AI-enabled capability in Defence" in June 2022 [10]. In January 2023, the US Department of Defense updated its directive on the highly dynamic matter [11]. These national approaches agree in explicitly rejecting the development of fully autonomous LAWS. Concerning partially autonomous PAWLS, they underline the importance of meaningful human control. However, it is far from clear what this implies concretely. Reaction speed in warfare often supersedes human capacities. In projectile interception systems, a high degree of autonomy is inevitable; programs such as AlphaDogfight suggest similar developments for the entirety of aerial combat, in which context the level of effective human control has been, de facto, constantly shrinking for decades.

On the international level, the 2022 and 2023 sessions of the UN Group of Governmental Experts (GGE) on Emerging Technologies in the Area of LAWS reached the conclusion that the use of LAWS "entails international responsibility" of the states deploying them and that "humans responsible for the planning and conducting of attacks must comply with international humanitarian law" [12]. Beyond this well-known emphasis on human control and liability, the GGE also made a number of technical recommendations: for instance that "LAWS must not be used if they are incapable of being used in compliance with IHL (…), including the principles and requirements of distinction, proportionality and precautions in attack" [13]. This addresses the problem of technologically primitive LAWS that can be regarded as indiscriminate and, therefore, unlawful weapons. To counteract such unlawful indiscriminate attacks originating from technologically primitive AI systems, the GGE recommended to:

A. Limit the types of targets that the system can engage.
B. Limit the duration, geographical scope, and scale of the operation of the weapon system.

Similar technical recommendations, even explicitly mentioning AI systems to guarantee IHL compliance, are formulated in the multilateral 2023 *Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy*, which is, as of November 2023, supported by all of the G7 states, virtually all EU states, and several individual states such as Libya, South Korea, Turkey, and Singapore:

> States should also consider how to use military AI capabilities to enhance their implementation of international humanitarian law and to improve the protection of civilians and civilian objects in armed conflict [14].

This represents the starting point of my reflections. Instead of focusing on human control and governmental actions, I promote a technology-centered approach to implementing regulatory systems into military applications of AI. These systems, which I call Autonomous Systems of Normative Control (ASNCs) could be demanded by public bodies as industry standards but could also be developed and implemented voluntarily as de facto standard by producers of military technology. This rather informal and private sector-oriented pathway might be faster than the cumbersome and, so far, inefficient attempts to reach a consensus between governments in this matter.

ASNCs are comparable to ISA-systems in cars, which can actively prevent drivers from exceeding speed limits using road-sign recognition and geolocation. Corresponding to varying degrees of autonomy, ASNCs could merely advise combatants in legal and ethical terms, autonomously block unlawful orders given by humans and/or propose military actions that are likely to be proportionate and minimize collateral damage. Furthermore, ASNCs should include blocking mechanism to contribute to ensuring that PALWS and logistical AI units are neither used in wars of aggression nor against domestic protesters. It is crucial to acknowledge that none of these technological approaches can be expected to provide complete solutions and to relief military personnel, politicians, and civil society of their duty to strive for more complete solutions to regulate PAWLS. Moreover, it must already be warned that these devices should not be abused to conduct *'ASNC-washing'* to justify arms exports to authoritarian regimes.

2. Previous approaches to Autonomous Systems of Normative Control

The discourse on Autonomous Systems of Normative Control (ASNCs) can be regarded as a subfield of AI ethics, machine ethics, or the construction of artificial moral agents [15]. However, there are also important differences: first, the discourse on ASNCs is rather preoccupied with implementing legal than ethical or moral norms, which is why I am speaking generally of normative control; second, minimalistic ASNCs regarding blocking mechanisms based on geolocation data are not necessarily linked to AI or artificial reasoning capacities, but are autonomous only in the sense that they do not require human intervention. ASNCs for PALWS, conventional weapon systems, and logistical AI units have been discussed for some time. More or less well-known examples include the maximalist approach by robot ethicist Arkin (2006) [16], the minimalist 'MinAI' by autonomous weapon systems-specialists Scholz and Galliot (2018) [17], robots refusing unlawful orders by international law scholars Grimal and Pollard (2021) [18], the logistics-oriented 'minotaurs' by technology ethicists Sparrow and Henschke (2023) [19], and the model of a military autonomous device following IHL by Zurek, Kwik, and·van Engers (2023) [20].

In general, these speculative and theoretical approaches, usually producing only AI flowcharts, are very optimistic about AI's abilities to regulate PAWLS as well as human behavior on the battlefield. This is in striking discrepancy with approaches focused on the current state of technology, which emphasize that today's AI is generally incapable of operating reliably under the condition of the complexity of the battlefield characterized by Clausewitzian 'frictions' and the proverbial 'fog of war'. For instance, these expected shortcomings of AI on the battlefield have been addressed by Wallace in 2022 [21] and Yan in 2020 [22]. In the often speculative theoretical AI discourse, it is important to contextualize research that is positive about the capabilities of AI and to critically examine the extent to which the arguments put forward are expressions of an unfoundedly progress-optimistic 'technological solutionism', which falsely assumes that complex social, political, or ethical problems can be easily solved by digital technology [23].

That being said, it is crucial to assess the capacity of AI systems in different domains differently: as demonstrated by AlphaDogfight, the success of loitering munition drones, and other examples, the comparably 'frictionless' aerial domain is rather suitable for AI systems; likewise, cyber, the sea, and outer space are relatively suitable for AI systems. In contrast to that, there are no reasonable mid-term expectations that AI will master ground combat, which is still the decisive domain – particularly if this involves urban terrains with unclear distinctions between civilian and military targets and particularly regarding on-the-ground PALWS or embodied AI-systems (EAI) in the sense of more or less humanoid 'killer robots'. Some authors regard the true capacities of military AI as entirely disassociated from PALWS and EAI and argue that AI is best employed in cognitive tasks, for instance, logistics [24]. As mentioned in the introduction, such an approach is also pursued in the U.S. Department of Defense's JADC2 vision. (See fig. 1 and 2.)

a) The Maximalist Approach: Governing Lethal Behavior (2006)

Arkin's 2006 study *Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture*, funded by the U.S. Department of Defense, is both a highly problematic expression of a techno-solutionist attitude and the origin of ASNCs. Arkin's concept is inspired by Watts' *mechanical governor* for the steam engine, which was intended to ensure the machine's safety and performance, a concept that generally plays a large role in early debates on cybernetics [25]. As depicted in the flowchart below (fig. 3), Arkin outlines a complex but comparatively transparent system that he labels as "Ethical Autonomous Robot Architecture", involving several loops: Mission planning and deliberation, which involves human-robot interaction and which lead to the formulation of ethical principles defined in the ethical adapter and constraints (C), which, in turn, contains obligations and prohibitions. Arkin characterizes C as informed by the rules of engagement and the international law of war to ensure ethical behavior. Law and ethics are barely distinguished. A system of "ethical behavior control" feeds computer-processed perceptions of behavior $s_n \rightarrow \beta_n \rightarrow r_n$ into C, which, turn, feeds data into the ethical governor, which, in turn, defines permissible actions (in terms of permissions, obligations, and prohibitions) (Permissible) and blocks all other actions.

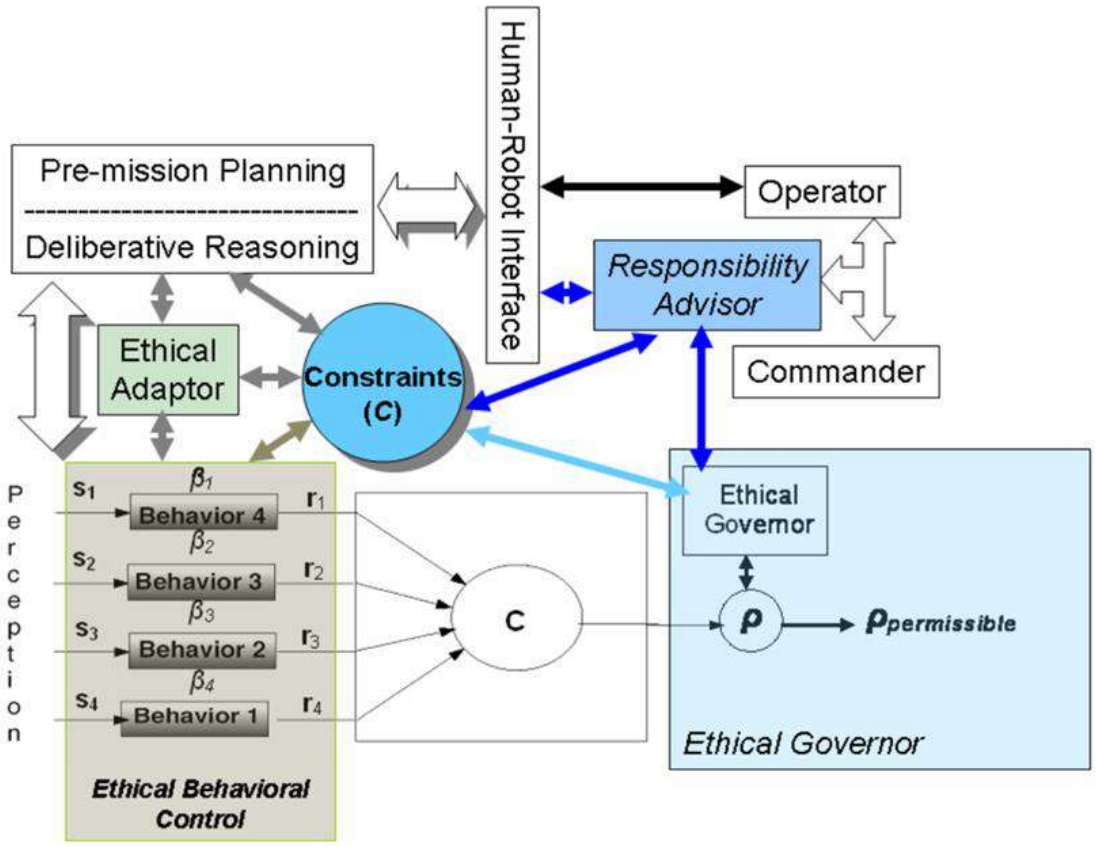

Fig.3: Major Components of an Ethical Autonomous Robot Architecture as depicted in Ronald C. Arkin, 'Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture' (Mobile Robot Laboratory College of Computing Georgia Institute of Technology, 2006).

Regarding the normative sources of the constraints and obligations in the C-module, Arkin envisions that these should be informed by mission-specific rules of engagement (ROE) in the machine's 'short term memory', including $O_{bligations}$ and $F_{orbidden,}$ and the prohibitions F_{orbid} in the machine's 'long term memory', which should be informed by the Laws of War (LOW) and Standing Rules of Engagement (SROE). These normative considerations would be applied as constraints and obligations to a module of 'evidential reasoning' fed by a global information grid producing situational awareness data and perceptions transformed into computable representations of the surrounding situation. See the representation of the system in the flowchart below (fig. 4 next page).

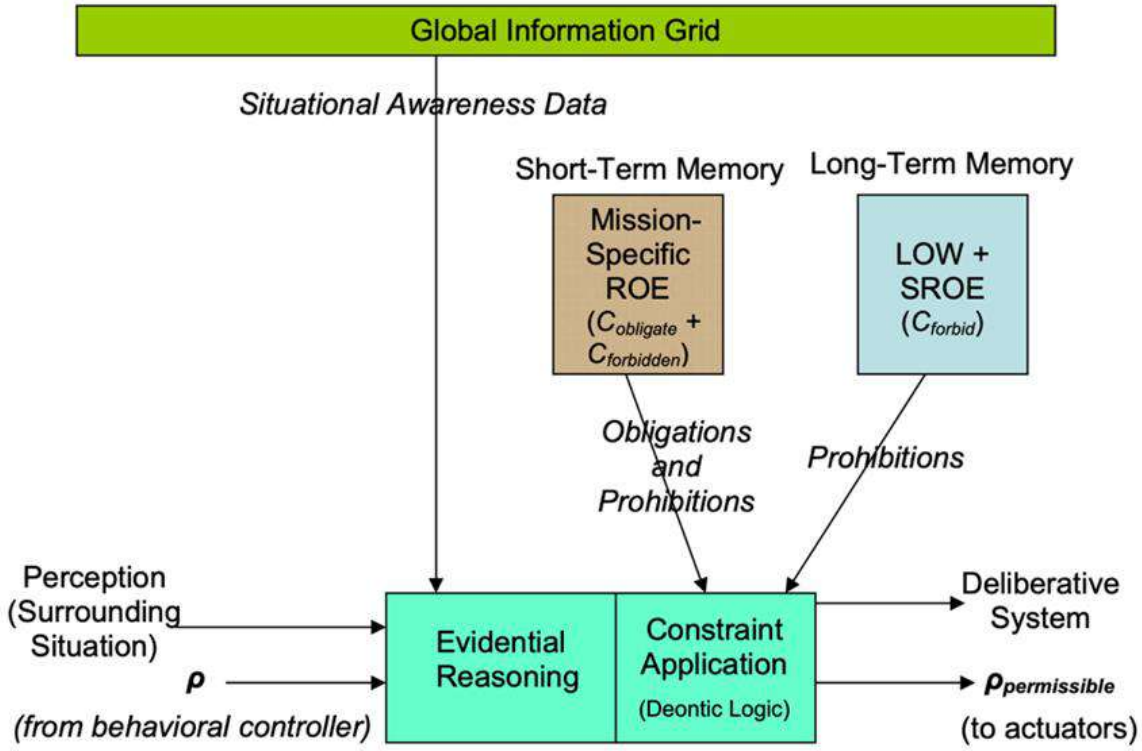

Fig.4: Ethical Governor Architectural Components as depicted in Arkin, 'Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture'.

b) The Minimalist Approach: MinAI (2018)

In 2018, Scholz and Galliot published a text on their ASNC concept MinAI, which explicitly sets itself apart from Arkin [26]. Scholz and Galliot characterize Arkin's approach as a 'maximum approach' since it relies on the ability of machines in C (Fig. 3) to independently relate ethical and legal principles to concrete situations and actions, i.e. to engage on numerous levels in *object recognition* and *machine reasoning*, including the definition of permitted *and* prohibited actions. In contrast, the authors take a 'minimal approach': they argue that translating general humanitarian principles of warfare into concrete actions recommended in a concrete combat situation should not be left to machines; instead, human actors should implement much simpler blocking mechanisms in PALWS, such as blocking mechanisms in response to the detection of civilian targets and the recognition of protective signs such as the Red Cross and the Red Crescent, the white flag, and raised hands as signs of surrender, as well as the recognition of religious and cultural sites. In addition, weapons could be blocked in certain use cases, including cases applying to proportionality criteria. Further, the authors emphasize that such ASNCs and AI systems generally cannot be morally and legally accountable and argue for

robust documentation to assess the responsibility of commanders in detail. In the figure below, it is depicted how the 'ethical weapon' system blocks the launch of a missile towards a ship with the Red Cross on it.

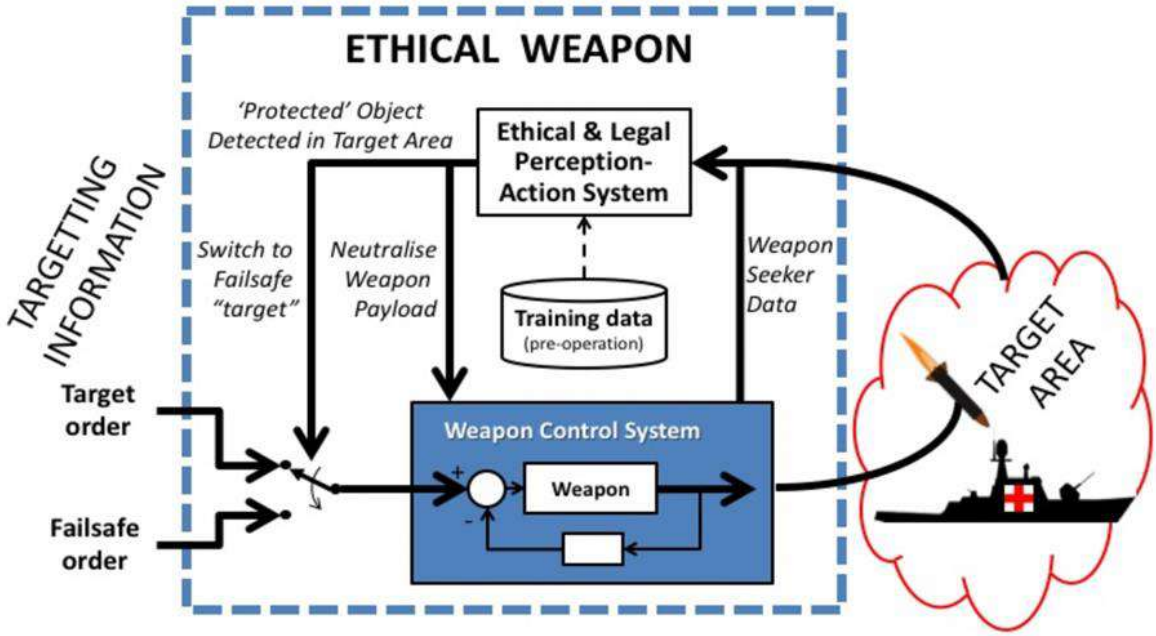

Fig. 5: Flowchart for Scholz and Galliot's Minimalist "Ethical Weapon" that merely includes a blocking mechanism. From: Scholz, Jason, and Jai Galliott. 'The Humanitarian Imperative for Minimally-Just AI in Weapons'. Journal of Indo-Pacific Affairs, Winter 2018, 57–67. https://doi.org/10.1093/oso/9780197546048.003.0005.

c) Robots refusing unlawful orders (2021)

Grimal and Pollard's 2021 *The Duty to Take Precautions in Hostilities, and the Disobeying of Orders: Should Robots Refuse?* marks a significant departure from Arkin's technological solutionism [27]. The authors (both international lawyers) no longer focus exclusively on the autonomous actions of machines, but address human-machine teaming. In addition, their article considers the particularities of IHL, especially the complex relationship between liability, responsibility, and chains of command. The authors point out that under Article 57 of Additional Protocol I of the Geneva Conventions and Rule 155 of customary IHL, both superiors and subordinates have a duty to take precautions, for example, regarding the distinction between civilians and combatants. Considering this duty, Grimal and Pollard argue, it might even be unlawful for military personnel not to consult AI systems, particularly if these are linked to

aerial reconnaissance and it is, therefore, likely that these systems facilitate the distinction between civilians and combatants.

However, Grimal and Pollard are also relevant to the concept of ASNCs insofar as they shift the focus from Arkin's concept of autonomous ethical and legal control over LAWS to a hybrid scenario in which 'disobeying robots' not only regulate themselves but primarily control or "advise" human behavior, including behavior in relation to conventional (i.e. non-autonomous) weapons. Grimal and Pollard describe a scenario where "a human still decides upon the most suitable means of attack, though in reality their “choices” are likely to be fairly restricted."

The authors point out that human combatants could benefit from asymmetries in human-machine teaming. This is because AI systems are not subject to the same psychological and social stresses that characterize humans actively engaged in combat. For example, Grimal and Pollard address AI-based safeguards systems for nuclear weapons that ethically and legally assess the legitimacy of launching missiles. This leads to Grimal and Pollard's concept of robot refusal: embodied AI (EAI) systems review human commands and practice various levels of command "refusal," which, in the least serious case, may involve a mere hint of possible unlawfulness and, in the most serious case, may involve an immediate block on all further commands similar in structure or issued by the same actor. This is depicted by the authors in the following flowchart, leading from the assessments of EAI System 1 to 3 to system 4A “systematic refusal” or system 4B “follow order” or “re-run assessment”.

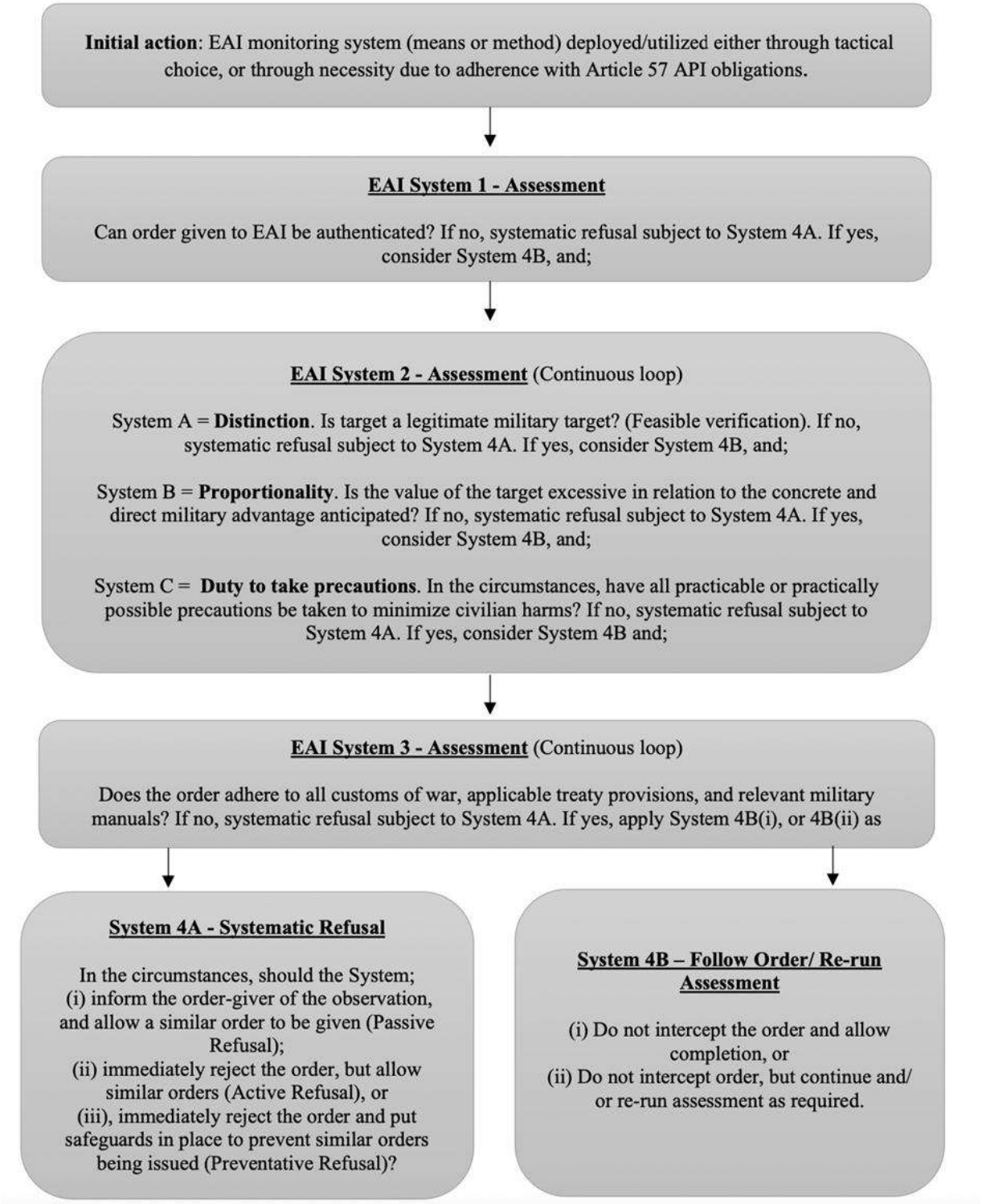

Fig. 6: Flowchart for refusing robots from Grimal, Francis, and Michael Pollard. 'The Duty to Take Precautions in Hostilities, and the Disobeying of Orders: Should Robots Refuse?' *Fordham International Law Journal*, 2021, 671–734.

d) Logistical AI units as Minotaurs (2023)

A fourth approach to be discussed here is *Minotaurs, Not Centaurs: The Future of Manned-Unmanned Teamin*g by ethicists Sparrow and Henschke from 2023 [28]. As the title, which refers to the hybrid creature of the Minotaur, suggests, this approach is exclusively about human-machine teaming. In the discussion of Sparrow and Henschke, the focus is explicitly shifted from physical forms of EAI, for example drones controlled by humans (referred to by autonomous weapon systems-expert Scharre as "centaurs" [29]), to AI systems that support human actors in the field in a cognitive capacity. The latter systems are referred to by the authors as "minotaurs" because they entail *non-human entities on top of processes, for example, as logistical coordinators*. In a realistic assessment of the state of the art, Sparrow and Henschke write:

> Artificial intelligences are arguably already more capable of performing the cognitive tasks most relevant to warfighting than robots are capable of performing the functions of the human body most relevant to warfighting.

Similar to Grimal and Pollard regarding the duty to take precautions, Sparrow and Henschke argue that the rapid development of AI systems creates an "ethical imperative" for the use of these technologies in warfare because they “will help prevent friendly fire incidents and enhance the survivability of human warfighters." Implicitly drawing on the Pentagon’s JADC2 vision, Sparrow and Henschke cite the logistical coordination performed by AI systems in the private sector, such as Amazon's fulfilment centers or Uber's coordination of drivers, as a model for such Minotaur-scenarios. An extreme case of such dynamics has been observed in participatory warfare in Ukraine, where volunteers were logistically organized by social media algorithms [30]. These phenomena of a platformization of war are highly relevant to Minotaur scenarios. Due to its focus on cognitive guidance provided by AI systems in human-machine teaming, Sparrow and Henschke's paper offers an interesting starting point for investigating possible applications of ASNCs in logistical AI units in the military.

e) Autonomous device following IHL (2023)

The most recent and certainly one of the boldest approaches to establish ASNCs is Zurek, Kwik, and·van Engers's *Model of a military autonomous device following International Humanitarian Law* (2023), which builds on Arkin but does not engage with the other approaches mentioned above [31]. On the most basic level, Zurek, Kwik, and·van Engers's model consists in predicting the relationship between *Military Advantage* (how much advantage is gained from attacking a specific target) and *Incidental Harm* (collateral damage caused by both foreseeable direct and indirect effects of an attack). As far as possible, these criteria should be quantified and formalised. The authors write:

> The creation of an autonomous AI-driven model requires not only a computational model, hence a quantifiable representation [of Military Advantage and Incidental Harm], but also a representation which allows for their formal comparison.

In a technical regard, Zurek, Kwik, and·van Engers underline that there has been a significant shift in AI research and development since Arkin's approach, namely the turn towards data-driven rather than rule-based approaches. In a turn away from the present trend, they promote a hybrid approach that is data-driven regarding the perception tasks that they summarise as "cognitive part" and rule-based regarding the "reasoning part", which is informed by IHL and other norms. Zurek, Kwik, and·van Engers are also more concrete than all the preceding authors since they explicitly focus on one part of the OODA-loop (Observe, Orient, Decide, Act), in which the ASNC should be implemented: the targeting cycle. Their following flowchart considers general circumstances, signal intelligence, and goals to be fed into the cognitive part and international treaties to be fed into the reasoning part. The generated output consists of ordering decisions regarding their degree of compliance with international regulations.

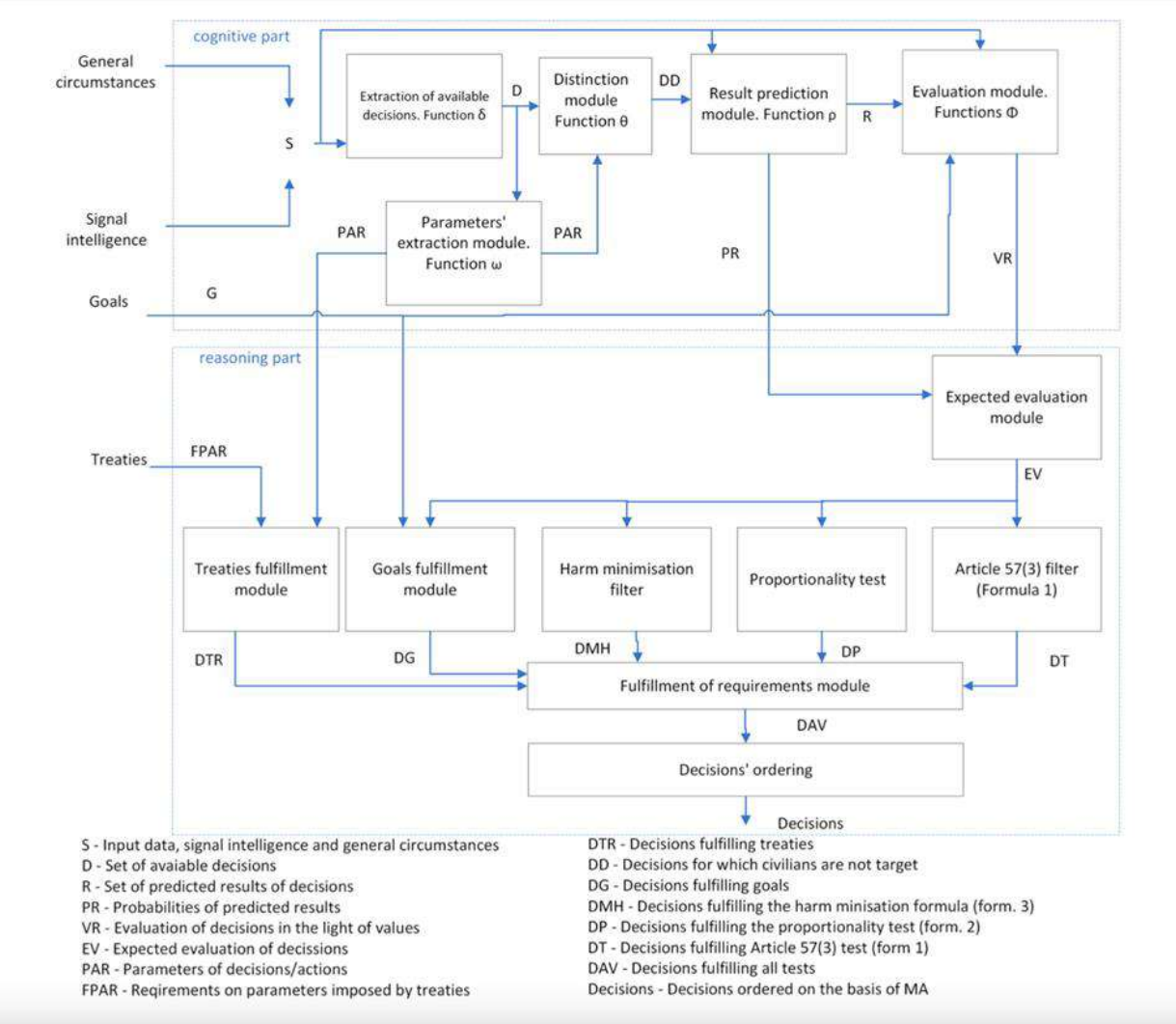

Fig. 7: Flowchart for *Model of a military autonomous device following International Humanitarian Law* from Zurek, Tomasz, Jonathan Kwik, and Tom Van Engers. 'Model of a Military Autonomous Device Following International Humanitarian Law'. *Ethics and Information Technology* 25, no. 1 (March 2023).

As depicted in the flowchart above, Zurek, Kwik, and·van Engers include several modules with specific tasks in the reasoning part, most importantly the "Harm minimization filter", "the Proportionality test", and the "Article 57 filter". These modules could also be activated separately, where an issue requires a specific form of scrutiny. Rather than promising a complete solution, these researchers emphasize that moral and legal responsibility cannot be delegated to machines, but that AI systems can constitute a crucial part in taking all possible precautions.

3. Discussion and Development of ASNCs

It is crucial to be aware of the difference between the tasks that ASNCs can plausibly master in the midterm future as opposed to such tasks that these systems are very unlikely to be capable of mastering in the midterm future. The historical development of ASNCs from Arkin's techno-solutionist approach to Zurek, Kwik, and·van Engers's gradual solution and Sparrow and Henschke's logistics-oriented approach serves as an illustration of a general turn in AI research away from EAI systems that promise complete solutions towards skepticism and emphasizing flexible approaches based on human-machine-teaming, combining the best of both worlds. Even in the current "AI summer" [32], which is really an AI heatwave, no one can seriously assume that it is possible to outsource moral or legal responsibility to machines in any way.

ASNCs cannot relief military personnel from their duty to be informed regarding relevant norms and to ensure norm abidance. Neither can ASNCs relief politicians and civil society from their duties to formulate and enforce specific norms that could govern the use of PALWS and other military applications of AI. Last but not least, due to military friction on the battlefield and the 'fog of war', AI systems cannot be expected to function perfectly in every domain. Autonomous ASNCs embedded in PALWS might work more or less well in the air, sea, and outer space domain, but certainly not regarding ground combat, in which context ASNCs might rather be operating from data centers, providing logistics and normative control to PALWS and human troops on the ground. These two different types of ASNCs are just two extreme ends of a spectrum that includes many intermediary forms. (See fig. 8 and 9.) As the example of the autonomous drones reportedly used in Ukraine demonstrates, the fully autonomous weapon type with built-in features has the great advantage of being immune to jamming.

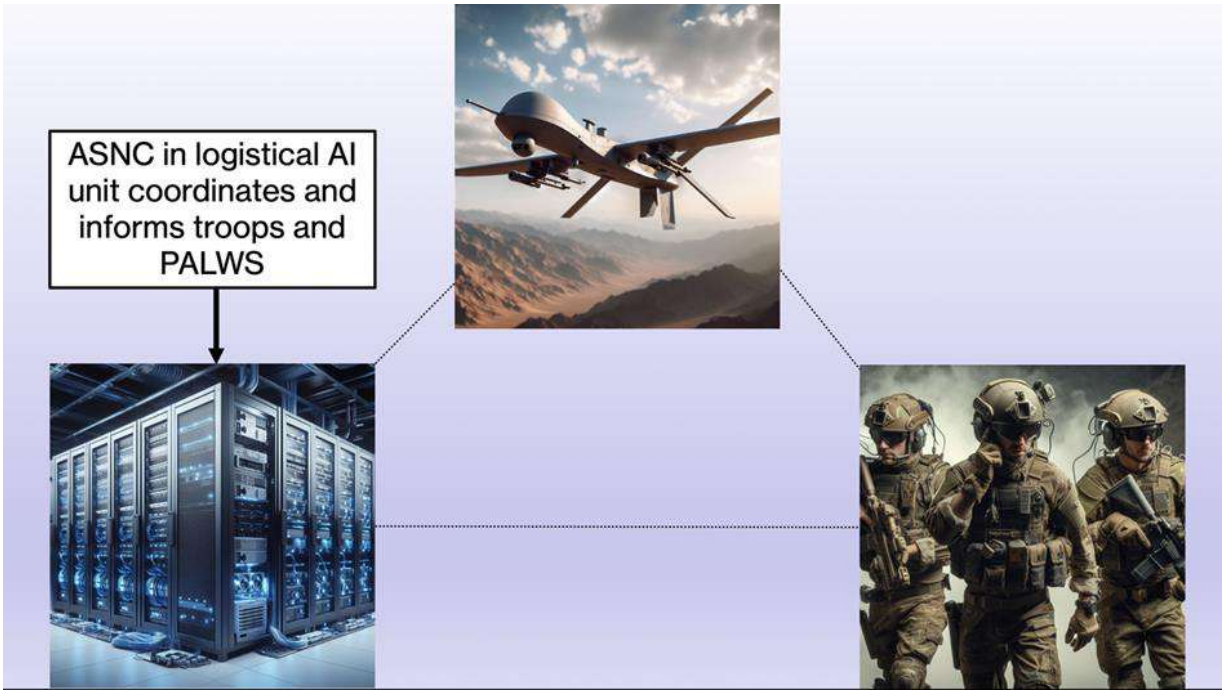

Fig. 8: ASNC in logistical AI unit following the Pentagon's JADC2 vision.

Fig 9: Autonomous ASNC built into PALWS, a maximalist approach.

Regardless of the domain, it is certainly the case that ASNCs could bring about significant benefits if understood as a tool to assist human operators and commanders in predicting the relationship between military advantage and incidental harm, which is crucial to comply with the IHL principle of proportionality. However, this requires the solution of significant problems in representation and language. As also emphasized by the US Air Force, particularly the criterion of proportionality is highly subjective [33]. It is difficult to achieve a quantifiable and formal representation in this regard. However, it must be underlined that the assessment of proportionality is often closely related to the quantitative assessment of fatalities. So, there is undoubtedly the possibility to represent the most crucial aspect of IHL in a computable way. Even beyond the construction of ASNCs, such processes of quantification and formalization could contribute to increasing the objectivity, transparency, and, therefore, debateability of IHL criteria.

From a technical perspective, Arkin's early rule-based approach from 2006 might represent the future rather than the past of ASNCs. During the last decade, rule-based approaches to AI have been abandoned in favor of data-driven approaches. Data-driven approaches are far more flexible than rule-based approaches since they do not require software engineers to formulate abstract rules for all possible scenarios. However, precisely since data-driven systems are not characterized by rules formulated by humans but by automated statistical inferences, these systems have often been criticized for constituting black boxes and lacking explainability [34]. Particularly in an ethically highly sensitive field such as warfare, ASNCs cannot rely on data-driven machine learning processes alone, which means that they will be only explainable *ex post*. Whilst it is inevitable to use data-driven approaches regarding situational awareness, ASNCs should include rule-based processes explicitly informed by the legal principles and deontic and modal logics of IHL.

However, minimal approaches such as Scholz and Galliott's *MinAI* have their advantages since they leave legal reasoning to humans and are based on "simpler" blocking mechanisms. Scholz and Galliott mention the detection of civilian targets and the recognition of protective signs such as the Red Cross and the Red Crescent, the white flag, and raised hands as signs of surrender, as well as the recognition of religious and cultural sites.

This idea of simpler blocking mechanisms is also relevant regarding legal and ethical principles that have not yet been sufficiently addressed in the literature: the prohibition on wars of aggression and the annexation of territory following public international law and the much softer, rather ethical-political or human rights-based demand not to use military weaponry to quell domestic protests.

In summary, ASNCs should fulfil the following tasks:

- Blocking mechanisms regarding the offensive use of PALWS in highly populated areas
- Blocking mechanisms regarding the defense of annexed territories
- Recognition of cultural and religious sites
- Blocking mechanisms regarding the use of force to quell domestic protests
- Differentiation between civilian and military infrastructure
- Differentiation between civilians and combatants
- Assessment of proportionality
- Recognition of protected symbols (such as the Red Cross and the Red Crescent)

It is worthwhile grouping these tasks in the following XY chart (fig. 10 below), with X (in green) ranging from higher to lower ethical sensibility and Y (in red) ranging from lower technological feasibility and human-machine teaming to higher technological feasibility and more machine autonomy. Furthermore, the upper half (in blue) is partially or wholly based on stable geolocation data, whereas the lower half (in purple) consists of dynamic tasks. Of course, regarding the parameters of ethical sensibility and technological feasibility, it must be added that none of these tasks is particularly easy and all of these issues generally are of high ethical sensibility. The differences represented in this XY chart might appear small and gradual from an external perspective. Regarding the distinction between dynamic tasks and tasks based on stable geolocation, it must be clarified that, for instance, blocking mechanisms related to specific territories, areas, and sites are stable because these territories, areas, and sites do not move, unlike, most importantly civilians and combatants themselves. Blocking mechanisms regarding the use of force against protesters also have an aspect based on stable geolocation data because

protests probably occur in domestic city centers and there is no need to employ military weaponry there if there is no foreign invasion. Analogously, of course, the differentiation between civilian and military infrastructure has stable aspects, since infrastructure does not move, but also dynamic aspects because, for instance in urban warfare, civilian infrastructure is frequently repurposed in a military sense.

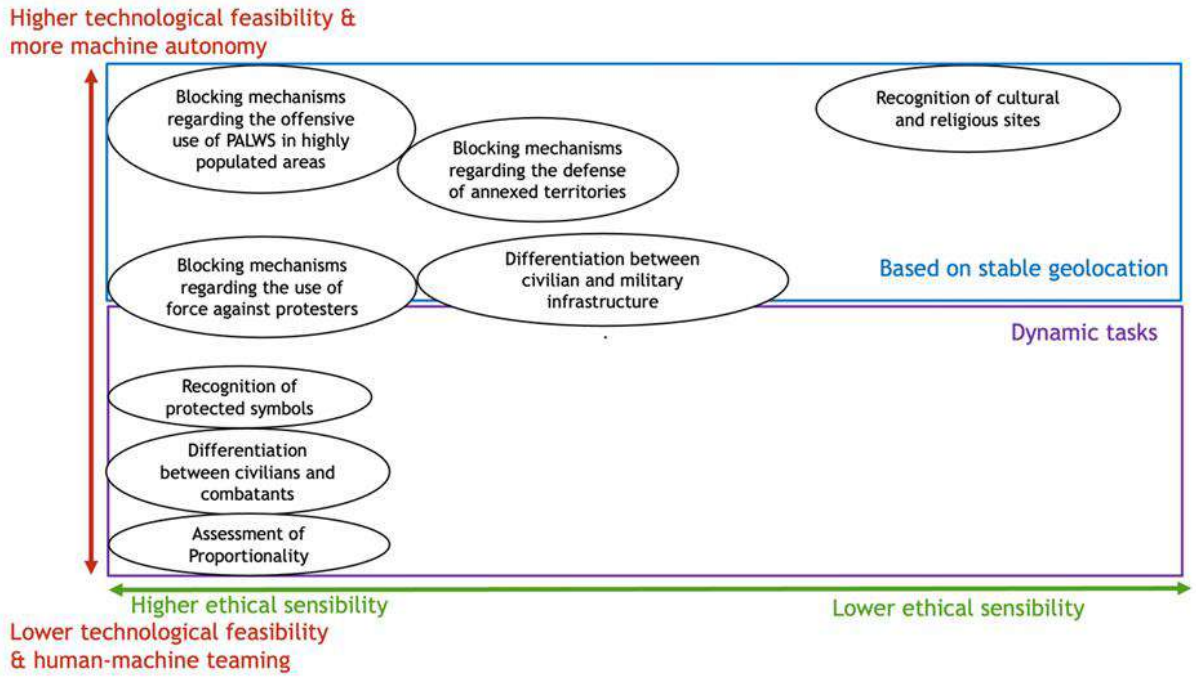

Fig. 10: XY chart, with X (in green) ranging from higher to lower ethical sensibility and Y (in red) ranging from lower technological feasibility and human-machine teaming to higher technological feasibility and more machine autonomy. The upper half (in blue) is partially or wholly based on stable geolocation, whereas the lower half (in purple) consists of dynamic tasks.

Considering all these qualifications, this XY chart reveals something surprising: Most discussions so far regard the lower left corner of the chart, which displays high ethical sensibility and lower technological feasibility (see fig. 11 next page). It would be far better to start in the upper right corner of the chart, which displays higher technological feasibility and lower ethical sensibility.

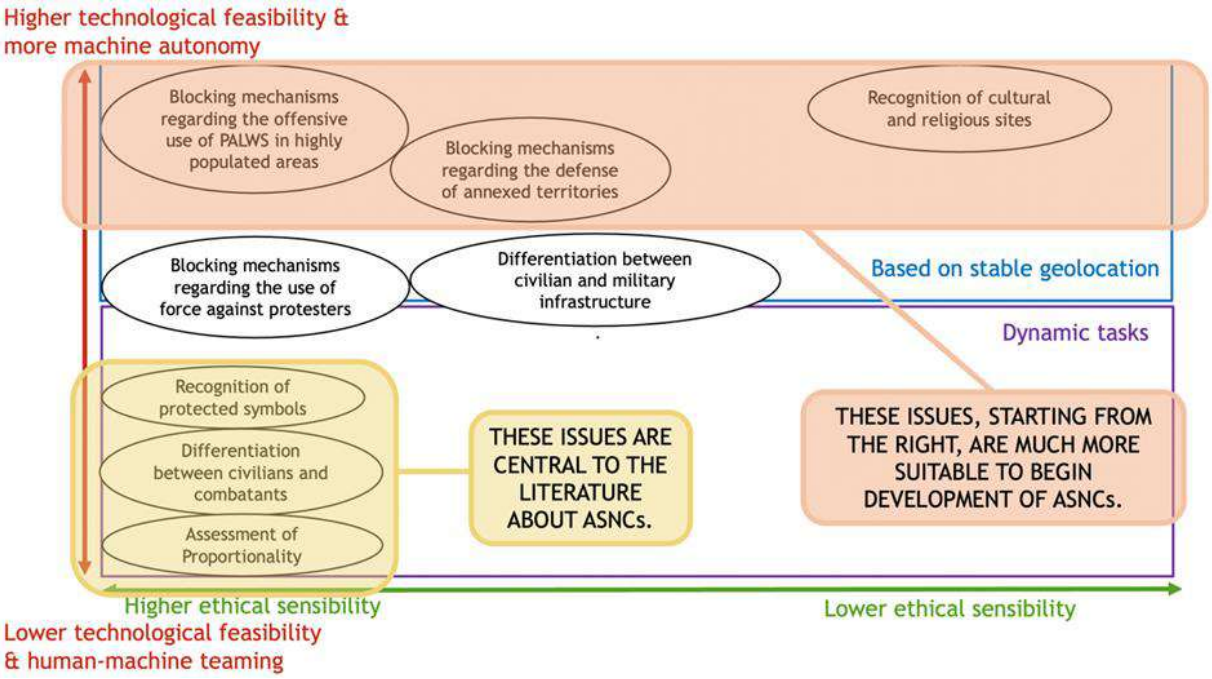

Fig. 11: XY chart, lower left corner (yellow) is usually discussed, whilst upper half (orange) would be more feasible.

4. ASNCs as de facto industry standard

This leaves the non-trivial question of how ASNCs should make their way from speculative reasoning expressed in flowcharts to becoming a de facto industry standard and why this should be desirable.

Of course, governments should follow the recommendations formulated in the 2022 and 2023 sessions of the UN Group of Governmental Experts (GGE) on Emerging Technologies in the Area of LAWS and the 2023 Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, which clearly point towards developing some form of ASNCs. But these recommendations are very vague and far from binding. National governments will inevitably interpret them differently, also depending on the interests of specific industries and other lobby groups relevant to their national discourses. Finding more robust international agreements regarding military applications of AI can be expected to be extremely difficult. So far, the many discussions regarding LAWS and PALWS on the national and international level have not even produced a robust definition of LAWS and PALWS. It is urgent to promote and accelerate national and international approaches to developing coherent and robust regulations regarding military applications of AI.

Starting with promoting ASNCs as a de facto industry standard, and then transforming them into an official industry standard, would be the normatively weakest possible pathway to work towards a regulation of military applications of AI. However, it would also be the one closest to those stakeholders who know the most about current and midterm technological feasibility.

Why should producers of military technology be interested in voluntarily developing and implementing such ASNCs? There are four good reasons for this.

1) It is not implausible that corresponding to the increased manufacturing of PALWS and the development of logistical AI units for the military in Europe, a public debate might form that insists on equipping military applications of AI with ASNCs.

2) The producers of military technology are generally under great public scrutiny and implementing robust ASNCs should be the only way to justify the export of PALWS and logistical AI units for the military. PALWS might be particularly useful for authoritarian

regimes since they minimize the number of consenting collaborators that a regime needs to maintain as loyal subjects. Exporting PALWS and logistical AI units to non-democratic regimes should be an absolute no-go because this poses a threat to political stability everywhere. But also, democracies should only receive PALWS and logistical AI units if these include ASNCs that guarantee that governments do not use AI systems against their own population, to conduct wars of aggression, or to defend annexed territory.

3) Research and development regarding ASNCs, particularly if publicly funded, can be expected to create synergy effects that will also improve the general reliability of PALWS and logistical AI units for the military. In the long run, PALWS and logistical AI units with ASNCs are likely to perform better in all respects.

4) Fourth, also clients, i.e. governments that purchase military applications of AI might have good reasons to prefer a product with ASNCs over one without safety mechanisms. In unstable regions of the world, such as in the Sahel Region, it is not unusual for governments to be overthrown by military leaders abusing publicly funded weaponry for a coup d'état. ASNCs can contribute to mitigating such risks and promoting political stability by limiting the extent to which weapons can be used against domestic opponents.

Summary of policy recommendations

- Promote Research and Development of ASNCs for PAWLS and logistical AI units in Europe according to the recommendations in the report of the 2023 session of the UN GGE on Emerging Technologies in the Area of LAWS and the 2023 Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.
- Promote Research and Development of ASNCs based on hybrid AI systems, combining data-driven and rule-based elements and simple blocking mechanisms based on geolocation data.
- In a first step, promote the development of blocking mechanisms based on geolocation data, for instance blocking mechanisms regarding cultural sites and the offensive use of PALWS in highly populated areas. In a second step, move to the development of more complex ASNCs, for instance regarding differentiation between civilians and combatants, assessment of proportionality, and the recognition of protected symbols.
- Raise public awareness regarding ASNCs. Promote debates in civil society regarding the guiding principles of ASNCs. Promote roundtables with industrial leaders to find possible agreement and convergence on ASNCs, also considering technological feasibility.
- Encourage industrial leaders to establish ASNCs as a de facto industry standard for PALWS and logistical AI units *Made in Europe.*
- Encourage industrial leaders to develop technological superiority of European products to nudge clients worldwide into accepting products with ASNCs.
- Ensure that ‘ASNC-washing’ does not serve as justification for arms exports to authoritarian regimes.
- Also pursue other non-technological pathways to regulate military applications of AI on national and international levels.

References

'ACE Program's AI Agents Transition from Simulation to Live Flight'. DARPA, 13 February 2023. https://www.darpa.mil/news-events/2023-02-13.

'Air Force Operations and the Law: A Guide for Air and Space Forces'. US Air Force Judge Advocate General's Department, 2002.

'Ambitious, Safe, Responsible: Our Approach To The Delivery Of AI-Enabled Capability In Defence'. UK Ministry of Defence, June 2022. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf.

Arkin, Ronald C. 'Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture'. Mobile Robot Laboratory College of Computing Georgia Institute of Technology, 2006. https://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf.

Asmolov, Gregory. 'The Transformation of Participatory Warfare: The Role of Narratives in Connective Mobilization in the Russia–Ukraine War'. *Digital War*, 27 September 2022. https://doi.org/10.1057/s42984-022-00054-5.

Chengeta, Thompson. 'Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law'. *Denver Journal of International Law & Policy* 45, no. 1 Fall (January 2016). https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1011&context=djilp.

'Department of Defense Directive 3000.09a Autonomy In Weapon Systems', 25 January 2023. https://media.defense.gov/2023/Jan/25/2003149928/-1/-1/0/DOD-DIRECTIVE-3000.09-AUTONOMY-IN-WEAPON-SYSTEMS.PDF.

Eklund, Amanda Musco. 'Meaningful Human Control of Autonomous Weapon Systems: Definitions and Key Elements in the Light of International Humanitarian Law and International Human Rights Law'. Swedish Defence Research Agency, February 2020. https://www.fcas-forum.eu/publications/Meaningful-Human-Control-of-Autonomous-Weapon-Systems-Eklund.pdf.

'Final Report of the Panel of Experts on Libya Established Pursuant to Security Council Resolution 1973 (2011)'. UN Security Council, 8 March 2021. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/037/72/PDF/N2103772.pdf?OpenElement.

Grimal, Francis, and Michael Pollard. 'The Duty to Take Precautions in Hostilities, and the Disobeying of Orders: Should Robots Refuse?' *Fordham International Law Journal*, 2021, 671–734.

Hambling, David. 'Ukraine's AI Drones Seek And Attack Russian Forces Without Human Oversight'. *Forbes*, 17 October 2023, sec. Aerospace & Defense. https://www.forbes.com/sites/davidhambling/2023/10/17/ukraines-ai-drones-seek-and-attack-russian-forces-without-human-oversight/.

'Joint All-Domain Command and Control (JADC2)'. Congressional Research Service, 21 January 2022. https://sgp.fas.org/crs/natsec/IF11493.pdf.

'Loitering Munition Systeme für Deutschland'. MDBA Germany, 12 July 2023. https://www.mbda-deutschland.de/pressemitteilung/loitering-munition-systeme-fuer-deutschland/.

Machi, Vivienne. 'French Army Taps Nexter to Build Tank-Busting Kamikaze Drones'. *C4ISRNet*, 19 June 2023, sec. Unmanned. https://www.c4isrnet.com/global/europe/2023/06/19/french-army-taps-nexter-to-build-tank-busting-kamikaze-drones/.

Metz, Cade, and Gregory Schmidt. 'Elon Musk and Others Call for Pause on A.I., Citing "Profound Risks to Society"'. *The New York Times*, 29 March 2023, sec. Technology. https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html.

Morozov, Evgeny. *To Save Everything, Click Here: The Folly of Technological Solutionism*. Paperback 1. publ. New York, NY: PublicAffairs, 2014.

Ntoutsi, Eirini, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, et al. 'Bias in Data-driven Artificial Intelligence Systems—An Introductory Survey'. *WIREs Data Mining and Knowledge Discovery* 10, no. 3 (May 2020): e1356. https://doi.org/10.1002/widm.1356.

'Opinion on the Integration of Autonomy into Lethal Weapon Systems'. Defence Ethics Committee, April 2021. https://www.defense.gouv.fr/sites/default/files/ministere-armees/20210429_Comité%20d%27éthique%20de%20la%20défense%20-%20Avis%20intégration%20autonomie%20systèmes%20armes%20létaux%20-%20Version%20anglaise.pdf.pdf.

'Report of the 2022 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', 29 July 2022. https://documents.unoda.org/wp-content/uploads/2022/08/CCW-GGE.1-2022-CRP.1-Rev.1-As-Adopted-on-20220729.pdf.

'Report of the 2023 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', 24 May 2023. https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_2_Advance_version.pdf.

Saltman, Max. 'As Azerbaijan Claims Final Victory in Nagorno Karabakh, Arms Trade with Israel Comes under Scrutiny'. CNN, 4 October 2023. https://www.cnn.com/2023/10/04/middleeast/azerbaijan-israel-weapons-mime-intl/index.html.

Scharre, Paul. 'Centaur Warfighting: The False Choice of Humans vs. Automation'. *International and Comparative Law Journal* 30, no. 1 (2016): 151–65.

Scholz, Jason, and Jai Galliott. 'The Humanitarian Imperative for Minimally-Just AI in Weapons'. *Journal of Indo-Pacific Affairs*, Winter 2018, 57–67. https://doi.org/10.1093/oso/9780197546048.003.0005.

Sparrow, Robert J., and Adam Henschke. 'Minotaurs, Not Centaurs: The Future of Manned-Unmanned Teaming'. *The US Army War College Quarterly: Parameters* 53, no. 1 (3 March 2023). https://doi.org/10.55540/0031-1723.3207.

'Summary of the Joint All-Domain Command & Control (JADC2) Strategy'. US Department of Defense, March 2022. https://media.defense.gov/2022/Mar/17/2002958406/-1/-1/1/SUMMARY-OF-THE-JOINT-ALL-DOMAIN-COMMAND-AND-CONTROL-STRATEGY.PDF.

United States Department of State. 'Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy', February 16, 2023. https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/

Van Wynsberghe, Aimee, and Scott Robbins. 'Critiquing the Reasons for Making Artificial Moral Agents'. *Science and Engineering Ethics* 25, no. 3 (June 2019): 719–35. https://doi.org/10.1007/s11948-018-0030-8.

Wallace, Rodrick. 'How AI Founders on Adversarial Landscapes of Fog and Friction'. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 19, no. 3 (July 2022): 519–38. https://doi.org/10.1177/1548512920962227.

Wiener, Norbert. *The Human Use of Human Beings : Cybernetics and Society*. Houghton Mifflin, 1950.

Yan, Guilong. 'The Impact of Artificial Intelligence on Hybrid Warfare'. *Small Wars & Insurgencies* 31, no. 4 (18 May 2020): 898–917. https://doi.org/10.1080/09592318.2019.1682908.

Zurek, Tomasz, Jonathan Kwik, and Tom Van Engers. 'Model of a Military Autonomous Device Following International Humanitarian Law'. *Ethics and Information Technology* 25, no. 1 (March 2023): 15. https://doi.org/10.1007/s10676-023-09682-1.

Footnotes

[1] Hambling, David. 'Ukraine's AI Drones Seek And Attack Russian Forces Without Human Oversight'. *Forbes*, 17 October 2023, sec. Aerospace & Defense. https://www.forbes.com/sites/davidhambling/2023/10/17/ukraines-ai-drones-seek-and-attack-russian-forces-without-human-oversight/.

[2] 'Final Report of the Panel of Experts on Libya Established Pursuant to Security Council Resolution 1973 (2011)' (UN Security Council, 8 March 2021), https://digitallibrary.un.org/record/3905159.

[3] Max Saltman, 'As Azerbaijan Claims Final Victory in Nagorno Karabakh, Arms Trade with Israel Comes under Scrutiny', CNN, 4 October 2023, https://www.cnn.com/2023/10/04/middleeast/azerbaijan-israel-weapons-mime-intl/index.html.

[4] Vivienne Machi, 'French Army Taps Nexter to Build Tank-Busting Kamikaze Drones', *C4ISRNet*, 19 June 2023, sec. Unmanned, https://www.c4isrnet.com/global/europe/2023/06/19/french-army-taps-nexter-to-build-tank-busting-kamikaze-drones/. 'Loitering Munition Systeme für Deutschland' (MDBA Germany, 12 July 2023), https://www.mbda-deutschland.de/pressemitteilung/loitering-munition-systeme-fuer-deutschland/.

[5] 'ACE Program's AI Agents Transition from Simulation to Live Flight' (DARPA, 13 February 2023), https://www.darpa.mil/news-events/2023-02-13.

[6] 'Joint All-Domain Command and Control (JADC2)' (Congressional Research Service, 21 January 2022), https://sgp.fas.org/crs/natsec/IF11493.pdf.

[7] Thompson Chengeta, 'Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law', *Denver Journal of International Law & Policy* 45, no. 1 Fall (January 2016), https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1011&context=djilp.

[8] Amanda Musco Eklund, 'Meaningful Human Control of Autonomous Weapon Systems: Definitions and Key Elements in the Light of International Humanitarian Law and International Human Rights Law' (Swedish Defence Research Agency, February 2020), https://www.fcas-forum.eu/publications/Meaningful-Human-Control-of-Autonomous-Weapon-Systems-Eklund.pdf.

[9] 'Opinion on the Integration of Autonomy into Lethal Weapon Systems' (Defence Ethics Committee, April 2021), https://www.defense.gouv.fr/sites/default/files/ministere-armees/20210429_Comité%20d%27éthique%20de%20la%20défense%20-%20Avis%20intégration%20autonomie%20systèmes%20armes%20létaux%20-%20Version%20anglaise.pdf.pdf.

[10] 'Ambitious, Safe, Responsible: Our Approach To The Delivery Of AI-Enabled Capability In Defence' (UK Ministry of Defence, June 2022), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf.

[11] 'Department of Defense Directive 3000.09a Autonomy In Weapon Systems', 25 January 2023, https://media.defense.gov/2023/Jan/25/2003149928/-1/-1/0/DOD-DIRECTIVE-3000.09-AUTONOMY-IN-WEAPON-SYSTEMS.PDF.

[12] 'Report of the 2022 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', 29 July 2022, https://documents.unoda.org/wp-content/uploads/2022/08/CCW-GGE.1-2022-CRP.1-Rev.1-As-Adopted-on-20220729.pdf.

[13] 'Report of the 2023 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems', 24 May 2023, https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_2_Advance_version.pdf.

[14] United States Department of State. 'Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy', February 16, 2023. https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/.

[15] Aimee Van Wynsberghe and Scott Robbins, 'Critiquing the Reasons for Making Artificial Moral Agents', *Science and Engineering Ethics* 25, no. 3 (June 2019): 719–35, https://doi.org/10.1007/s11948-018-0030-8.

[16] Ronald C. Arkin, 'Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture' (Mobile Robot Laboratory College of Computing Georgia Institute of Technology, 2006), https://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf.

[17] Jason Scholz and Jai Galliott, 'The Humanitarian Imperative for Minimally-Just AI in Weapons', *Journal of Indo-Pacific Affairs*, Winter 2018, 57–67, https://doi.org/10.1093/oso/9780197546048.003.0005.

[18] Francis Grimal and Michael Pollard, 'The Duty to Take Precautions in Hostilities, and the Disobeying of Orders: Should Robots Refuse?', *Fordham International Law Journal*, 2021, 671–734.

[19] Robert J. Sparrow and Adam Henschke, 'Minotaurs, Not Centaurs: The Future of Manned-Unmanned Teaming', *The US Army War College Quarterly: Parameters* 53, no. 1 (3 March 2023), https://doi.org/10.55540/0031-1723.3207.

[20] Tomasz Zurek, Jonathan Kwik, and Tom Van Engers, 'Model of a Military Autonomous Device Following International Humanitarian Law', *Ethics and Information Technology* 25, no. 1 (March 2023): 15, https://doi.org/10.1007/s10676-023-09682-1.

[21] Rodrick Wallace, 'How AI Founders on Adversarial Landscapes of Fog and Friction', *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 19, no. 3 (July 2022): 519–38, https://doi.org/10.1177/1548512920962227.

[22] Guilong Yan, 'The Impact of Artificial Intelligence on Hybrid Warfare', *Small Wars & Insurgencies* 31, no. 4 (18 May 2020): 898–917, https://doi.org/10.1080/09592318.2019.1682908.

[23] Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism*, Paperback 1. publ (New York, NY: PublicAffairs, 2014).

[24] Sparrow and Henschke, 'Minotaurs, Not Centaurs'.

[25] Norbert Wiener, *The Human Use of Human Beings : Cybernetics and Society* (Houghton Mifflin, 1950).

[26] Scholz and Galliott, 'The Humanitarian Imperative for Minimally-Just AI in Weapons'.

[27] Grimal and Pollard, 'The Duty to Take Precautions in Hostilities, and the Disobeying of Orders: Should Robots Refuse?'

[28] Sparrow and Henschke, 'Minotaurs, Not Centaurs'.

[29] Paul Scharre, 'Centaur Warfighting: The False Choice of Humans vs. Automation', *International and Comparative Law Journal* 30, no. 1 (2016): 151–65.

[30] Gregory Asmolov, 'The Transformation of Participatory Warfare: The Role of Narratives in Connective Mobilization in the Russia–Ukraine War', *Digital War*, 27 September 2022, https://doi.org/10.1057/s42984-022-00054-5.

[31] Zurek, Kwik, and Van Engers, 'Model of a Military Autonomous Device Following International Humanitarian Law'.

[32] Cade Metz and Gregory Schmidt, 'Elon Musk and Others Call for Pause on A.I., Citing "Profound Risks to Society"', *The New York Times*, 29 March 2023, sec. Technology, https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html.

[33] 'Air Force Operations and the Law: A Guide for Air and Space Forces' (US Air Force Judge Advocate General's Department, 2002).

[34] Eirini Ntoutsi et al., 'Bias in Data-driven Artificial Intelligence Systems—An Introductory Survey', *WIREs Data Mining and Knowledge Discovery* 10, no. 3 (May 2020): e1356, https://doi.org/10.1002/widm.1356.

Acknowledgements:

Many thanks to Florence G'sell for her valuable remarks on the manuscript and to Jessica Dorsey for bringing reports about the Ukrainian use of fully autonomous drones to my attention. During this research, I received funding from the EU Horizon 2020 research programme under MSCA COFUND grant agreement 101034352 with co-funding from the VUB-Industrial Research Fund.

About the author:

Johannes THUMFART is a senior and postdoctoral researcher at the research group Law, Science, Technology, and Society (LSTS) of Vrije Universiteit Brussel and an adjunct lecturer in the ethics of international security management at the Berlin School of Economics and Law. His research is focused on the ethics of international security, the military use of AI, and digital sovereignty in BRICS states. Thumfart received his PhD in the history and philosophy of international law at Humboldt Universität Berlin. He has been awarded a Marie Skłodowska-Curie Fellowship and held numerous teaching and research positions in Germany, France, Mexico, and Belgium. He contributed to some of the largest and most respected German newspapers such „der Spiegel“ and „die Zeit“. His monograph on digital sovereignty will be published with Palgrave Macmillan in summer 2024. His research is published in journals such as European Journal of International Security, AI and Ethics, and Global Studies Quarterly.

About the Digital, Governance and Sovereignty Chair:

Sciences Po’s Digital, Governance and Sovereignty Chair’s mission is to foster a unique forum bringing together technical companies, academia, policymakers, civil societies stakeholders, public policy incubators as well as digital regulation experts.

Hosted by the School of Public Affairs, the Chair adopts a multidisciplinary and holistic approach to research and analyse the economic, legal, social and institutional transformations brought by digital innovation. The Digital, Governance and Sovereignty Chair is chaired by **Florence G’sell**, Professor of Law at the Université de Lorraine, lecturer at the Sciences Po School of Public Affairs. During the 2023-2024 academic year, Florence G’sell is visiting professor at the Cyber Policy Center of Stanford University.

The Chair’s activities are supported by:

sopra steria next

Caisse des Dépôts GROUPE