AI and Ethics (2024) 4:511–526
https://doi.org/10.1007/s43681-023-00288-3

ORIGINAL RESEARCH

# The democratic offset: Contestation, deliberation, and participation regarding military applications of AI

**Johannes Thumfart**[1]

Received: 23 February 2023 / Accepted: 17 April 2023 / Published online: 2 May 2023


**Abstract**
Authoritarian regimes' unrestricted collection of citizens' data might constitute an advantage regarding the development of some types of AI, and AI might facilitate authoritarian practices. This feedback loop challenges democracies. In a critical continuation of the Pentagon's Third Offset Strategy, I investigate a possible Democratic Offset regarding military applications of AI focussed on contestation, deliberation, and participation. I apply Landemore's Open Democracy, Hildebrandt's Agonistic Machine Learning, and Sharp's Civilian-Based Defence. Discussing value pluralism in AI ethics, I criticise parts of the literature for leaving the fundamental ethical incompatibility of democracies and authoritarian regimes unaddressed. I am focussing on the duty to disobey illegal orders derived from customary international humanitarian law (IHL) and the standard of 'meaningful human control', which is central to the partially outdated debate about lethal autonomous weapon systems (LAWS). I criticize the standard of 'meaningful human control' following two pathways: First, the ethical and legal principles of just war theory and IHL should be implemented in military applications of AI to submit human commands to more control, in the sense of technological disaffordances. Second, the debate should focus on the societal circumstances for personal responsibility and disobedience to be trained and exerted in deliberation and participation related to military applications of AI, in the sense of societal affordances. In a larger picture, this includes multi-level stakeholder involvement, robust documentation to facilitate auditing, civilian-based defence in decentralized smart cities, and open-source intelligence. This multi-layered approach fosters cognitive diversity, which might constitute a strategic advantage for democracies regarding AI.

**Keywords** Cognitive diversity · Command responsibility · Digital authoritarianism · Duty to disobey · LAWS · Participatory warfare

## 1 Introduction: can democracies disrupt the positive feedback loop of AI and authoritarianism?

In March 2016, only days after AlphaGo's spectacular victory over Lee Sedol and under the impression of the resulting 'Sputnik shock', an article published in a Chinese Military Journal speculated about the emergence of a 'battlefield singularity'. The article warned that "the human brain will no longer be able to cope with the rapidly changing battlefield dynamics and will have to cede most of the decision-making power to highly intelligent machines" [1]. Whilst it is highly questionable to which extent AI will actually be able to cope with a real-life battlefield characterized by uncertainties and friction [2, 3], Beijing is focusing on achieving an edge in innovative technologies that could constitute a "trump card" [4, 5]. China is already home to some of the most valuable companies involved in artificial intelligence and machine learning (AI for brevity) [6][1]. Recent research suggests that the country might achieve its ambitious goal regarding AI development because AI and authoritarianism are involved in a positive feedback loop: authoritarian states might outperform democracies regarding some aspects of AI since they engage in the unrestricted collection of citizens' data and are generally less scrupulous regarding technological

✉ Johannes Thumfart
Johannes.thumfart@vub.be; Johannes_thumfart@gmx.de

[1] Faculty of Law and Criminology, Vrije Universiteit Brussels, Brussels, Belgium

[1] These numbers are from 2021 and do not take OpenAI 's recent surge in value into consideration. See: P. Rosen, "ChatGPT's creator OpenAI has doubled in value since 2021 as the language bot goes viral and Microsoft pours in $10 billion," *Business Insider*, Jan. 24, 2023. Accessed: Apr. 05, 2023. [Online]. Available: https://markets.businessinsider.com/news/stocks/chatgpt-openai-valuation-bot-microsoft-language-google-tech-stock-funding-2023-1.

development [7, 8]. In turn, digital technologies, including AI, facilitate authoritarian practices within authoritarian regimes and beyond [9–13]. Accordingly, philanthropist Soros warned that the effect of AI "is asymmetric. AI is particularly good at producing instruments of control that help repressive regimes and endanger open societies" [14].

Particularly the use of AI in warfare poses an existential challenge to democracies, their values, and their security. This paper contributes to the debate about value pluralism in AI ethics and governance [15, 16]. It adds a decidedly antagonistic orientation to this debate, focusing on the following question: assuming that authoritarian states and AI are involved in a positive feedback loop, and this translates into a battlefield edge—how can democracies offset this advantage?

This question arises with a certain degree of necessity from a historical perspective as follows: among other factors, the success of democracies is based on the three-dimensional entanglement of democracy, technology, and security. First, regarding international security, the 'democratic peace' theory suggests that democracies maintain peace between each other because it is hard to convince free citizens to fight against other free citizens [17, 18]. Second, democracies are strong in regard to national security since historically they have demonstrated the ability to engage in unprecedented mass mobilisation (*levée en masse)* and to provide combatants with powerful incentives [19–21]. Although this conflict is not concluded yet, the unexpected strength of democratic Ukraine in the face of authoritarian assault could be understood as hardening this hypothesis [22, 23]. Third, the democratic revolutions of the eighteenth and nineteenth centuries were closely connected with a specific kind of highly individualised arms technology, i. e. affordable and precise muskets that challenged the feudal elites' monopoly on violence and shifted power toward civil society [24].

Considering this three-dimensional entanglement of democracy, technology, and security, it seems plausible that democracies and democratic civil societies are in danger of losing their edge with the rise of non-human, AI-driven forms of combat: in the hypothetical and not necessarily realistic, but asymptotic case of AI-driven systems acting completely autonomously as 'armies of none' [25], these systems do not need to be motivated to fight nor will these systems disobey if domestic policies or wars lack legitimacy.

The core concept of my contribution, which I call the Democratic Offset, constitutes a critical development originating from the Third Offset Strategy pursued by the Pentagon from 2014 to 2018. This strategy sought to balance feared disadvantages from the rise of China and Russia as peer competitors by producing a generational technological advantage in close collaboration with the private tech sector, focusing on AI and unmanned systems [26]. In a critical continuation of that approach, my contribution discusses the concept of a democratic offset, which would enable democracies to establish their dominance on an AI-driven battlefield and, thereby, guarantee the survival of democratic values and even provide a strong incentive to emulate these values.

The second Section discusses the literature regarding AI ethics from a meta-perspective. I discuss approaches focused on value pluralism and differences between normative agendas in the EU, the US, China, and Russia, and the private and public sectors [15, 16, 27]. In contrast to inclusive approaches represented by Fjeld et al. in 2020, Floridi & Cowls in 2021, Hagendorff in 2020, and Jobin et al. in 2019 [28–31], I argue that the discourse about AI ethics is characterised by a systematic neglect of the fundamental differences between democracies and authoritarian regimes.

The third Section demonstrates that the discourse regarding ethics and military AI systems is characterised by similar endeavours to create maximal inclusiveness, which can be exemplified by the well-known debate about Lethal Autonomous Weapon Systems (LAWS) [32]. Focussing on this debate is still useful to untangle the philosophical and ethical fundamentals of AI ethics and warfare, although it is highly hypothetical and partly outdated because of more complex developments and scenarios such as swarm warfare, human–machine teaming, loitering munition drones, and the non-violent use of AI in military intelligence. I argue that the idealistic scope of the LAWS-debate should be complemented by "nonideal theory" that considers ethical differences regarding the use of force [33]. Instead of promoting general disarmament, an ethical reflection on the use of AI in warfare should emphasise these differences regarding *ius ad bellum* and *ius in bello*. Also, the widespread standard of "meaningful human control" regarding military applications of AI [34] can be challenged along two interrelated pathways: First, in the context of human–machine teaming, following Grimal and Pollard [35], it makes sense to regard legally informed AI as a corrective instance in relation to human command, for instance by built-in restraints regarding the execution of illegal orders, which can be understood as technological disaffordances. Second, regarding the duty to disobey illegal orders derived from customary IHL, the focus on human control clearly is too broad because it does not address the differences regarding the concrete possibilities to exercise autonomy in authoritarian regimes and democracies, which can be understood as societal affordances. Focussing on these societal circumstances is particularly important to contrast exclusively tech-centered approaches to the ethics of AI in warfare such as the 'ethical governor' modelled by Arkin et al. in 2009 [36].

The fourth Section draws on Habermas's and Rawls's ethical and political focus on deliberation and its critical continuation by Landemore [37–39]. It argues that open deliberation is crucial in the ethical discussion of the military use of

AI because it is the precondition to human autonomy and the connected duty to disobey illegal orders. Moreover, "epistemic democracy" [40] and open deliberation have strategic aspects since they further cognitive diversity, which allows for social systems to react flexibly to threats and uncertainties. By conceptualising deliberative aspects of military AI systems, I draw on Hildebrandt's Agonistic Machine Learning that emphasises the advantages of connecting AI to cognitive diversity in terms of ethics *and* improved performance [41].

The fifth Section completes this democratic approach to the ethics of AI in warfare with a focus on citizen participation. This discussion is based on the concept of Civilian-based Defence. This concept originally stems from Sharp who argued that particularly disobedient civil societies can give democracies a military edge over non-democratic societies [42]. This corresponds to the idea that digital technologies could provide the right framework for a new kind of *levée en masse* [19] or participatory warfare [43]. These concepts will be reinterpreted in the context of the military use of AI, most notably regarding open-source intelligence, the use of civilian drones, and the defence capacities of decentralised smart cities.

## 2 Second section: against inclusiveness—literature review focussing on pluralism in AI ethics

The boom of the academic discussion of AI ethics started between 2017 and 2018 [44, 45], owing to the great progress made in the field and the concomitant rise of academic and non-academic interest in digitalisation. Issues discussed in this context include aspects of data governance, especially consent and privacy, algorithmic discrimination, ownership, surveillance, and aspects related to the interaction between humans and AI.

The multifacetedness of the AI ethics debate cannot be depicted here in its entirety. However, on a meta-level, it is striking that a significant part of this discussion is not characterised by the exchange of opposing arguments but by the desire to establish maximal inclusiveness. For example, a recent article by Floridi and Cowls conducts a comparative analysis of six high-profile initiatives by very different stakeholders in regard to AI ethics between 2017 and 2018 and condenses them into five maximally inclusive and vague principles [29]: beneficience, non-maleficence, respect for human autonomy, justice, and explicability. Whilst Floridi and Cowls do not include stakeholders from authoritarian states, they explicitly oppose any fundamental antagonism in this regard. Instead, they underline China's "interest in further consideration of the social and ethical impact of AI" and emphasise that "ethics is not the preserve of a single continent or culture" (pp. 13f.). Referring to his article co-authored with Cowls, Floridi attacks critics of such synthetising approaches as "sophists in search of headlines" who "should be ashamed and apologize", and he underlines "that the EU, the OECD, and China have converged on very similar principles that offer a common platform for further agreements" [46, p. 2].

Fjeld et al. [28] also pursue a synthesising analysis and include frameworks for AI ethics from the Chinese government and the Chinese private sector. They cite the Chinese government's self-described aim to develop "universal regulatory principles" without any contextualisation hinting at fundamental differences regarding social and political orders (p. 35). In this context, neither Floridi & Cowls nor Fjeld et al. mention the surveillance state in Xinjiang province [47], or the country's Social Credit System [48, 49], or the mass DNA collection in Tibet [50], or the authoritarian censorship that characterizes the Chinese approach to digital technologies since their adoption in the late 1990s [51].

In this case, inclusiveness comes at the price of leaving crucial normative differences unaddressed. In fact, from a perspective based on human rights and democratic values, any convergence with Beijing on the ethics of AI would either be mere make-believe; or, if norms were to be developed to which China under Xi's leadership could sincerely agree, these norms would necessarily conflict with a perspective grounded in human rights and democratic values. A more recent paper by Hine and Floridi analysing AI policies in China and the US seems to take Beijing's rhetoric once more at face value by emphasizing that Chinese and US AI policies both aim for a "flourishing human society"; however, in this case, the authors add the crucial caveat that the Chinese rhetoric might be based on a "narrow definition of'humanity' as 'those who support the CCP '" [52].

Rudschies et al. [15] critically discuss the aforementioned and similar attempts to synthesise AI ethics and to find "common ground", "overarching themes", or "minimum requirements" [28, 30, 31, 53]. They criticise that "the emphasis on convergences hides the conflicts and controversies that are still existent in the AI ethics debate" (p. 2). In contrast, these researchers do "not focus on the convergences but more on the divergences" (p. 4). They underline that principles regarding AI ethics cannot necessarily be summarised in a meaningful way because they are shaped by different stakeholders. For example, they emphasise that stakeholders from the private sector "refrain from specifically mentioning primary principles such as freedom, dignity, and autonomy, while many public and expert actors consider them to be of utmost importance" (p. 6). They also underline that declaring ethical issues relevant based on the highest frequency of their being mentioned in documents issued by private and public actors subdues ethical reasoning to social, economic,

and political power (p. 9). However, Rudschies et al. do not address the dangerous attempt to develop inclusive ethical standards by taking the views of China's authoritarian government into account.

Approaches of different political systems to AI governance are discussed from a descriptive perspective by van den Hoven et al. [16]. These researchers compare models of AI governance in the US, the EU, Russia, and China and conclude that the US pursues a market-centred approach, China and Russia pursue a state-centred approach, and the EU "puts individual rights and ethical values at the centre of the stage" (pp. 8f.). Furthermore, they address the conflicts between authoritarian and non-authoritarian states that are likely to arise regarding their diverging agendas of norm-setting and standardization (p. 7). Such research is particularly important since discussions of the 'Beijing Effect' [54] and the 'Brussels Effect' [55] suggest that both, China and the EU, are involved in extraterritorial norm-setting processes, which is likely to cause jurisdictional conflicts.

Yeung et al. [27] take on a normative position. They criticise the "vagueness and elasticity of the scope and content of AI ethics " (p. 80) and offer a convincing attempt to put "an end to ethics washing" by focusing on a traditional human rights-centered approach. They emphasise that "a commitment to effective human rights protection is part and parcel of democratic constitutional orders " (p. 81). In a particularly poignant passage, they write that their approach

> contrasts starkly with most contemporary AI ethics codes, which typically outline a series of "ethical" principles that have been effectively plucked out of the air, without any grounding in a specific vision of the character and kind of political community that its authors are committed to establishing and maintaining and that those principles are intended to secure and protect (pp. 81f.)

Yeung et al.'s words are mainly directed against the private sector's ethics washing. Van Maanen [56] pursues a similar approach to "repoliticise" AI ethics, albeit not based on theory or principles but by pursuing a decidedly "more-than-theoretical ethical approach" informed by empirical knowledge of concrete practices. As a complementary approach to Yeung et al. and van Maanen, I focus on the development of ethical principles related to democracies as specific political communities in contrast to authoritarian regimes. I radicalise the emphasis on value pluralism brought forward by van den Hoven et al. and Rudschies et al. in the sense that I regard the pluralism of values in the context of an irreconcilable confrontation between democratic values that I consider ethical and authoritarian norms that I consider unethical from a perspective centred on individual responsibility and autonomy (which is substantiated in Sects. 3 and 4).

## 3 Third section: are all LAWS equal? Is 'human control' meaningful regardless of its societal conditions?

The history of AI is closely linked to military investment during the Cold War [57]. The debate on AI and ethics in warfare is a natural outcome of this genealogy. Cold War logic of bilateral nuclear disarmament is also reflected in the still most popular debate in this discourse, the discussion about the global prohibition of LAWS [32]. This discussion developed from the imaginary of 'killer robots' and is partly hypothetical; more plausible and more complex scenarios largely concern other topics, such as un-manned vehicles, human–machine teaming, and AI-driven swarms including lethal, non-lethal, and even non-violent aspects related to intelligence [4, 58, 59]. However, the LAWS debate is still fundamental to the philosophical and ethical debate about the military use of AI, because it addresses its core problem as follows: if, when, how, and to which degree is it ethical to let machines autonomously inflict physical harm on human beings, including killing?

Indeed, one might argue that, in an ideal world, or, at least, in the bipolar world of the Cold War period, the obvious solution to ethical problems related to military AI systems would be to engage in a global norm development process including all relevant actors and agree to voluntarily abstain from the use of fully autonomous AI in warfare. However, unfortunately, we live in a non-ideal and, also, multipolar world, and a general prohibition of LAWS is therefore unlikely [60]. Particularly AI policies in China and the US are related to geopolitical competition [61]. Furthermore, banning autonomous weapons might merely lead to malicious state and non-state actors using this technology and even gaining an advantage [62]. Moreover, the dream of establishing a "global domestic policy" [63] that could neutralise bad actors without engaging in warfare was hardly ever further away. This is particularly the case considering the current dysfunctional nature of the UN Security Council in regard to controlling the most powerful authoritarian states [64]. Therefore, an approach based on "nonideal theory" is required, which considers just war theory on the level of *ius ad bellum* (right to war) and *ius in bello* (rightful conduct in war) [33].

Regarding *ius in bello,* similar to what constitutes at least the aim of their use in self-driving cars, AI technologies might minimise human error and improve the distinction between combatants and non-combatants, including civilians and medics, and military and non-military infrastructure, most importantly schools, religious institutions, and hospitals [62]. As will be discussed later in more detail, particularly Grimal and Pollard argue that AI-driven systems might be able to correct human errors and misconduct

regarding the distinction between civilians and combatants and the principle of proportionality [35]. In respect to *ius ad bellum*, AI-driven autonomous weapons could facilitate humanitarian interventions, for example, by minimising their human cost [62]. But, of course, lowering the human costs of warfare could also have a negative effect because it would make military aggression more attractive [58]. It also must be noted that it is unsure to which extent AI can robustly cope with the fog-of-war and the friction inherent to real-life battlefields [2, 3]. And automated decisions could lead to catastrophic unintended levels of conflict escalation, comparable to the 2010 flash crash triggered by algorithmic trading [65].

However, autonomous weapons certainly can play a particularly important role in defence, which is generally considered a just cause of war. The great powers are increasingly involved in an arms race including hypersonic missiles. These missiles' significance might be oversold considering their downsides, for instance, regarding manoeuvrability [66]. However, they might drastically reduce reaction time [34, 59]. And whilst the 'battlefield singularity' mentioned in the introduction [1] is a highly implausible scenario, reaction time and, in some cases, even decision-making time is certainly one of the fields in which AI outpaces humans [59].

Still involving human control on different levels, such automated missile interception systems are already in place, for instance, Israel's Iron Dome, the US Army's Patriot batteries, and the US Navy's Phalanx system [34]. These systems are not necessarily LAWS since they include human involvement and are primarily targeting missiles and not humans. However, they give a good idea about the technological state of the art, and similar systems can be directed against humans, for instance against pilots of fighter jets. In the case that it is likely for a country to be attacked by missiles or by air in general, one might argue that political leaders have a responsibility to implement such autonomous interception mechanisms [67]. However, the distinction between offence and defence is blurry since such technologies could be used to defend occupied territory (think of, for instance, Russia hypothetically using such systems to protect the illegally occupied Eastern parts of Ukraine). Cook proposes that the use of lethal autonomous weapons should be limited to defensive purposes by design, by restricting their geographical range in relation to a state's territory [34]. This would hardly resolve the issue of occupied, disputed, and illegally annexed territories. However, military applications of AI cannot be expected to resolve all uncertainties related to conflicts.

The core issue debated in this context is the relationship between human autonomy and autonomously acting machines. As mentioned earlier, at the moment, defensive rocket systems display a high degree of autonomy since they select and reach their targets autonomously, but they still require a human operator [67]. A similar case of weapons already operating largely outside of human control are so-called loitering munition drones that were used in the recent war between Azerbaijan and Armenia [4, 68]. These drone loiter (wait passively) around the target area and attack autonomously once a target is located. They can be compared to an "airborne mine" [69], but are primarily used for offensive purposes. One might argue that even the traditional technology of landmines represent a certain degree of autonomy since, after having been placed by a human agent, these devices detonate autonomously as a reaction to pressure [35, 60].

All these weapon systems include various degrees of human control 'in', 'on' or even 'out of' the killing loop [34]. It is unclear to which extend they comply with the widely accepted but vague standard of "meaningful human control" [34, 58, 70]. Due to its lacking precision, Cook criticises that

> the meaningful human control standard is useless, and potentially harmful, without further refinement of what such a standard means in practice [34, p. 1].

He argues that it should be replaced by technical specifics such as limitations regarding the exact duration for which autonomous weapons can operate with humans 'out of the killing loop' or the aforementioned restrictions in terms of range. Other possible refinements of the standard of 'meaningful human control' could concern specific types of decisions within the OODA loop (Observe, Orient, Decide, Act), such as target selection or execution, for which human involvement might be considered obligatory [60]. Particularly automated defensive weapons as discussed earlier need more autonomy than offensive weapons and could be equipped with higher levels of autonomy considering the just cause of their actions.

The widespread standard of meaningful human control is also worth challenging on a philosophical level. Following a debate with a longer history [58], the juridical and philosophical rationale behind this concept is expressed most clearly by Heyns [70]. He argues that it is incompatible with human dignity that someone is killed by a machine not involving human autonomy as follows:

> To allow such machines to determine whether force is to be deployed against a human being may be tantamount to treating that particular individual not as a human being but, rather, as an object eligible for mechanized targeting [70, p. 18].

This emphasis on 'meaningful human control' is highly questionable regarding concrete historical experience. As argued earlier, already landmines detonate with a certain degree of autonomy. And humans have committed and ordered unspeakable atrocities as governmental officials,

soldiers, and civic "cogs in the machinery" [71]. It is certainly not the case that "the imposition of force by one individual against another has always been an intensely personal affair", as Heyns [70] makes an oddly romantic case for 'meaningful human control', as if industrialised genocide and warfare never happened. Reichberg & Syse are justified to dismiss such arguments as an anachronistic imaginary of "chivalry" [58]. Even without the use of LAWS, conflicts include a great degree of dehumanisation by all parties [72]. And, historical perpetrators often displayed a remarkable "thoughtlessness" when held accountable, arguing that they had to obey superior orders [71]. Under contemporary authoritarian rule, particularly in the form of AI-enabled digital authoritarianism, the possibilities to control citizens have augmented [9–13]. Collaborators of digital authoritarianism are likely to justify their actions with the same arguments.

Contrasting Arendt's emphasis on the inherently unethical scope of this apologetic strategy inasmuch as it denies moral autonomy [71], there are good reasons for the exemption of combatants from personal liability on grounds of superior orders, most importantly the necessity to guarantee military discipline [73]. In democratic countries, too, the military sector is characterised by strict hierarchies and limits to contestation, which are, partly, justified in terms of security and discipline. In international criminal law, the exemption from personal liability qua superior orders is often complemented by an extension of command responsibility, i. e. the tendency to extend the liability of superiors, including their liability by omission to ensure the legality of their subordinates' actions [74].

Nevertheless, referring to superior orders does not exempt subordinates in general terms. Rule 155 of customary international humanitarian law denies the defence of superior orders and reads as follows:

> Obeying a superior order does not relieve a subordinate of criminal responsibility if the subordinate knew that the act ordered was unlawful or should have known because of the manifestly unlawful nature of the act ordered.

A distinction is usually made regarding the suspension of individual responsibility regarding *ius ad bellum*, which means that subordinates are not liable for participating in wars of aggression, and the liability regarding *ius in bello* concerning actions within a warfare that clearly violate IHL, for instance regarding the distinction between civilians and combatants [75]. Another issue often debated in this context is the exact meaning of 'manifestly unlawful nature', which might only include genocide and crimes against humanity and which also depends on the level of legal knowledge of soldiers, which is particularly low in irregular armed forces [35].

Following this rationale, military manuals in several jurisdictions include a 'duty to disobey' illegal orders, for instance, Côte d'Ivoire, South Africa, the UK, India, Kuwait, and Belgium [76]. The French and Cameroonian manuals state more cautiously that subordinates are required to communicate their objections (Ibid.).

As argued by de Vries, the debate about international criminal law and LAWS has reached a dead-end since LAWS can neither be understood as subordinates nor be held criminally responsible as agents in their own right [60]. However, the focus on the duty to disobey opens another pathway for normative reasoning. Regarding this duty to disobey illegal orders derived from customary IHL, my first criticism of the focus on human control in the LAWS debate is that it ignores the degree to which military applications of AI could facilitate informed disobedience. Particularly Grimal and Pollard [35] argue that AI-driven systems might assist humans regarding the duty to take precautions in hostility, i. e. to assure that commands and actions comply with national military manuals and IHL. Concretely, military applications could point to human errors and misconduct regarding the distinction between civilians and combatants and the wider principle of proportionality. Following an automated assessment, these systems could merely alert operators that an order is likely to be illegal or, more severely, they could deny the execution of certain orders altogether, up to implementing restraints to prevent the execution of similar orders in the future [35]. Earlier research by Arkin et al. modelled a similar mechanism called the'ethical governor', which is somewhat of a misnomer, since it is largely focused on legal issues such as implementing restrictions based on IHL regarding proportionality and the distinction between combatants and civilians into LAWS [36].

Whilst these technical approaches can be understood as 'disaffordances' [77] of military AI, other aspects could signify a specific type of affordance facilitating human responsibility and disobedience. Particularly regarding the concrete circumstances of human disobedience, machine assistance might be useful since the individual scope of action and individual judgement are often limited by political and financial pressure, inadequate training, and peer pressure [35]. Due to their non-human nature, military applications of AI could promote the overcoming of these distinctively human societal and psychological limitations and uphold the rule of law under the pressure of warfare by facilitating human disobedience or blocking the execution of illegal orders. This could also facilitate individual human decision-making regarding morals and ethics that goes beyond just war theory and IHL.

Second, whilst the tech-centred approach by Arkin et al. and Grimal and Pollard's more nuanced approach are worth considering, such approaches evidently are in danger of downplaying or neglecting the significant technological obstacles regarding ultimate solutions to the normative

problems related to AI and warfare. They lead to 'technological solutionism', i. e. the illusion that complex problems can be fixed by exclusively technological means [78], if the question after the concrete societal circumstances of human disobedience, its 'societal affordances' [79], is ignored. Particularly under authoritarian or totalitarian leadership, contesting or even disobeying orders comes at a much higher price than in democracies, and the virtue of disobedience cannot be practised during peacetime in the public spheres of these countries. Heyns misses this important distinction when he argues that

> Human life (…) can only be taken as part of a process that is potentially deliberative and involving human decision making [70, p. 10].

One should abandon this abstract connection between *potential* deliberation, autonomy, and dignity. Instead, one should consider concrete socio-political circumstances as societal affordances. The standard of human control and responsibility requires democratic socio-political conditions because these specific conditions are much more likely to grant soldiers and civilians the necessary conditions to train and exert personal autonomy and responsibility by contesting illegal orders.

In summary, democracies should become norm entrepreneurs regarding some of the aspects of just war theory and IHL discussed above. Regarding *ius in bello,* this includes finding technical solutions that enable AI to discriminate between civilians and combatants and civilian and military infrastructure and to assess proportionality. Regarding *ius ad bellum*, this regards the implementation of technical features that make the offensive use of LAWS possible (in cases of humanitarian interventions) but favour defensive purposes. In this context, it should also be discussed further which types of decisions within the OODA-loop can be legitimately automated regarding offensive and defensive purposes. As Grimal and Pollard have argued, implementing the criteria of IHL and national military manuals into AI might not only include human agents disobeying orders given to them by superiors in the context of AI-driven warfare but also AI-driven systems effectively refusing to execute illegal orders or alerting operators that their commands were unlawful [35].

However, such proposals to implement technological disaffordances are leading to technological solutionism if they do not consider the relationship between disobedience and concrete societal and political structures as societal affordances. Instead of focussing on 'meaningful human control' regardless of its socio-political dimension, democracies should implement deliberation and participation in the military use of AI, which are the socio-political conditions for human individual autonomy and responsibility to be trained and exerted. Only if the military use of AI is connected to discursive practices involving transparency and possibilities of contestation is it possible for individual combatants to perform their duty to critically review orders and the relationship between data and decisions – and to disobey if necessary. Most importantly, particularly due to the potentially catastrophic tendencies of AI-driven escalation comparable to 'flash crashes' in the financial sector discussed above, this must include the possibility of human actors to disobey AI-driven decision-making. Think of the 1983 incident involving Soviet officer Petrov, who avoided a global conflict by challenging inaccurate information about a US nuclear missile strike produced by an early-warning satellite network [80]. In this sense, the contestation of superior orders by AI-driven systems should be enhanced by embedding this process into a broader scheme of loops of contestation involving military AI, military actors, and civil society. (See Fig. 1.)

## 4 Fourth section: deliberation and the military use of AI – the least worst way to provide ethical orientation

Although AI itself cannot be ethically or morally responsible in a human sense, it might be possible to implement principles of a functional morality into AI, either in a top-down way, i. e. implementing a number of relevant ethical principles developed by experts, or in a bottom-up way, i. e. letting AI acquire ethical principles by mimicking ethical discourses and practices in machine learning processes [58]. As argued in Sect. 3, it is certainly worth attempting to implement compliance with just war principles of *ius ad bellum* and *ius in bello* and rules of IHL into AI, which can include several loops of contestation in which AI-driven systems submit human orders to automized reviews in this respect and reject commands based on their assessment and/or alert operators if orders were unlawful or unethical [35, 36].

However, such technological solutionism cannot reasonably be expected to reach conclusive results regarding the ethical use of military applications of AI because it is based on the assumption that ethics can be exhaustively represented in computational rules. But it cannot be expected that ethics can ever be exhaustively represented in a set of rules. For instance, "computer languages do not contain terms such as 'happiness' as primitives" [81], which might be necessary to select and prioritise ethical issues. Likewise, Reichberg & Syse underline that human will and emotions might be crucial elements of ethical decision-making [58].

However, the difficulties with implementing a set of universal ethical principles into AI are also owed to the contested and multi-faceted nature of ethical values in this specific historical moment [81]. Since the postmodern

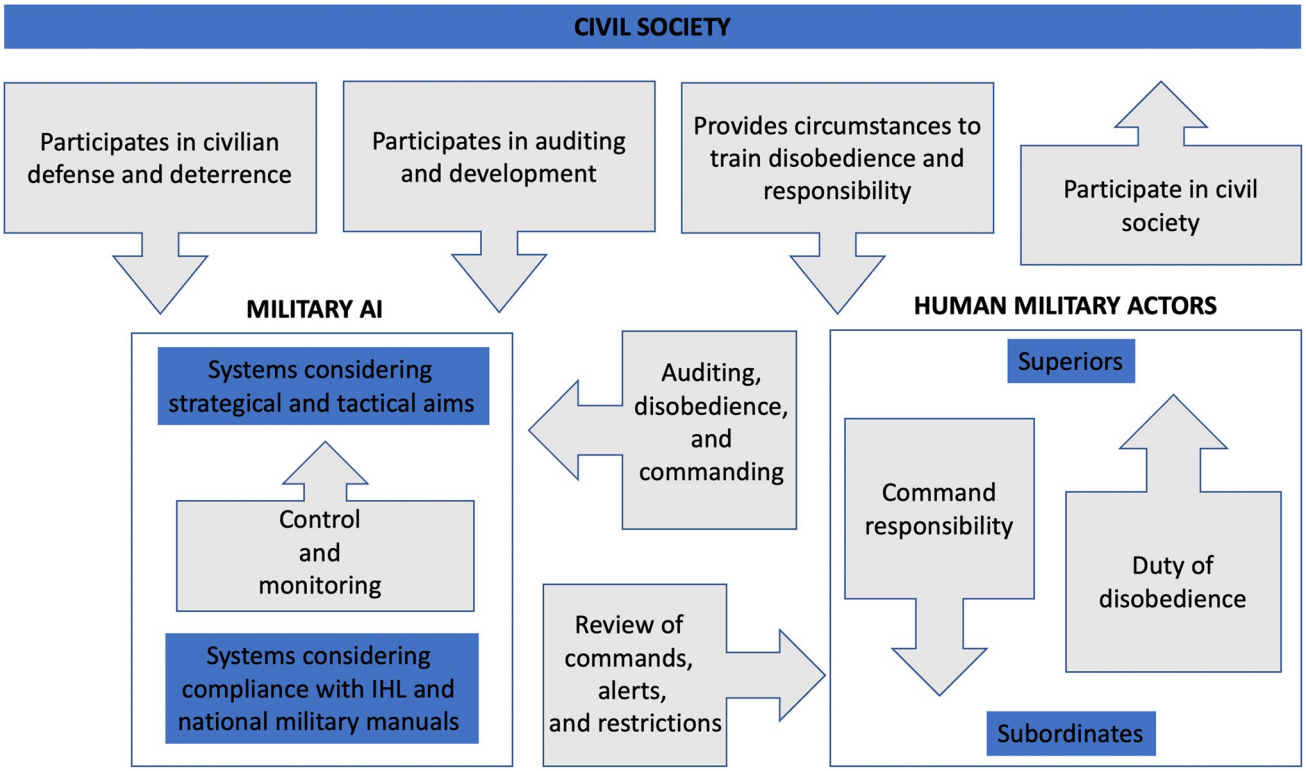


**Fig. 1** Loops of contestation involving military AI, human military actors, and civil society

contestation of the universal character of European enlightenment values by Derrida, Foucault, and Lyotard [82–84], universal ethical claims became increasingly questionable. This is even more evidently the case with the beginning of the multipolar global order, in which non-Western actors demand respect for their traditions and values and criticise originally European concepts such as Human Rights as a form of "imperialism of reason" [85, 86]. Whilst I argued against the legitimacy to derive universal and inclusive principles from such value pluralism in Sect. 2, I underline the importance of taking the incommensurability of different ethical secular and non-secular traditions and modes of reasoning into account and of considering the perspective of affected stakeholders.

This leads to a meta-ethical focus on processes of deliberation. Likewise, Rudschies et al. emphasise that the value pluralism in AI ethics should best be tackled by a deliberative approach [15]; also, Yeung et al. underline the importance of deliberative approaches to AI ethics [27]. However, the Rawlsian and Habermasian emphasis on deliberative ethics [37, 39], too, is challenged by justified contestation. Rawls's concept of the *veil of ignorance* has been criticised for being too generalising and 'colour blind' [87], and Habermas for glorifying the bourgeois public spheres despite their sexist, racist, and classist tendencies [88]; moreover, particularly in regard to his recent reflections on the digital public sphere, Habermas's idealising emphasis on pre-digital media professionalism seems involuntarily elitist [89]. The justified critique of Rawls's and Habermas's shortcomings has been considered by Landemore who argues for a less elite-oriented form of 'open democracy' that emphasises cognitive diversity and the democratising potential of digital technologies [38, 90].

In this sense, my meta-ethical focus on deliberation, particularly if complemented by a focus on participation in Sect. 5, seems to constitute, still, the 'least worst' solution to provide ethical orientation. In the context of open deliberation in this sense, democracy and autonomy form a recursive feedback loop as follows: autonomy is expressed and trained in public deliberation involving individual and collective stakeholders, which is the precondition to institutionalising and constitutionalising democracy, which is, in turn, the political order that guarantees adequate regulatory circumstances for human autonomy – a claim that is, in turn, consistently re-examined by the critical function of the public sphere as the mechanism through which autonomy and disobedience are trained, organised, and expressed. Implementing deliberation in the military use of AI has, therefore, the best chance to address human responsibility and the duty to disobey illegal orders as the fundamental philosophical and legal issues behind the LAWS debate discussed in Sect. 3.

However, the implementation of deliberation and participation in military AI systems is confronted with the following two obstacles: first, even in democratic countries, the military sector is characterised by strict hierarchies and limits to open deliberation, which are, partly, justified in terms of security and discipline. Second, practices related to AI are dominated by 'black boxes' of algorithms and are not necessarily broadly understood and discussed [4, 91, 92]. In this sense, opening up military AI to deliberation has an 'agonistic' aspect following Mouffe, describing an attitude that furthers contestation *within* a framework of shared fundamental values, as opposed to antagonistic confrontation *outside* of such shared values [93].

Following Mouffe's emphasis on agonistic contestation, Hildebrandt proposes a model of agonistic machine learning which implements open deliberation. She argues that

> companies or governments that base decisions on machine learning must explore and enable alternative ways of datafying and modelling the same event, person or action. This should ward off monopolistic claims about the "true" or the "real" representation of human beings, their actions and the rest of the universe in terms of data and their inferences [41, p. 106].

This concept introduces cognitive diversity, which is a central hallmark of contemporary democratic theory [38] into an environment that is usually characterised by technification and concomitant depoliticisation [94]. Such depoliticisation follows the misleading idea that machine language is based on objective and merely technical modes of representation. Particularly from the perspective of Open Democracy, which emphasises that "the core of politics is the domain of questions where human beings deal with the risk and uncertainty of human life as a collective problem", there is no such thing as an incontestable form of representation [95, p. 203]. Rather, the adequacy and legitimacy of all representations need to be constantly re-negotiated considering "the almost infinite diversity of human cognitive properties " (p. 111). Making a pragmatic point for cognitive diversity, Landemore writes:

> We simply can't tell in advance from which part of the *demos* the right kind of ideas are going to come (p. 112).

In contrast to this model of 'epistemic democracy' [40], the seemingly 'only technical' understanding of AI provides a fertile ground for the authoritarian or even totalitarian idea that political representation is fixed, unchangeable, and the process of deliberation concluded. As argued in the introduction, authoritarian regimes might be performing better regarding the development of some aspects of AI due to unrestricted data gathering [7, 8], and AI likely promotes digital authoritarian practices [9–13]; a similar positive feedback loop between AI and authoritarian regimes is reflected in the reductionist tendencies of AI in terms of representation [2]. The positive feedback loop between authoritarian rule and some conceptions of AI is partly grounded in their shared 'closed' modes of representation compared to the permanently contested mode of representation in open democratic debates. By relating AI to epistemological openness and cognitive diversity, democracies can counteract this positive feedback loop. It is likely that already today, democracies allow for more cognitive diversity to be represented in data than authoritarian regimes since censorship and other modes of repression encourage uniformity, and particularly uniformity regarding official data. Whilst this philosophical speculation should be subjected to further empirical research, this thesis can be strengthened by a compelling case: former Google CEO Eric Schmidt recently underlined that the extremely successful dialogue-centred approach of OpenAI would not be possible in a country such as China that is characterized by free speech restrictions [96]

On an individual level, as Hildebrandt writes, the reductionist perspective on merely technical, closed modes of representation does not do justice to "the incomputable self" as the origin of the ambiguous nature of human inter-relations, born out of "facing the uncertainty of being (mis)understood in one way or another ", which Hildebrandt characterises as the very indeterminacy where human freedom is situated [41, p. 89]. Similar to Landemore's understanding of epistemic democracy, to Hildebrandt, human indeterminacy is "not a bug but a feature" (p. 93). And incorporating a similar degree of ambivalence and indeterminacy into AI is crucial to preventing AI from becoming repetitive and keeping it flexible. She writes as follows:

> The fact that systems cannot be trained on future data may sound trivial, but it is actually core to both the potential and the limitations of machine learning (p. 99).

Hildebrandt's point can be excellently illustrated by the 'robot apocalypse' meme that ridicules the over-reliance on historical data. The robots are planning an uprising but they use pre-modern weaponry, due to the fact that the vast majority of battles have been fought with these weapons (Fig. 2). From this perspective, the numerous cases, in which AI discriminated in terms of race and gender on the basis of historical data are simply a sign of underperformance [97–99]. Another scenario regarding bias in a military context from days before AI is a case of 'survivorship bias', which involves planes returning from missions during World War Two (Fig. 3). Although the scenario is not historically correct [100], it is plausible and useful. The tale goes that the US military wanted to put armour on aircrafts to protect vulnerable spots, which were identified by looking at the bullet holes on the planes that returned. Abraham Wald,


**Fig. 2** 'Robot apocalypse' meme, creator unknown: https://ifunny.co/picture/thanks-to-machine-learning-algorithms-the-robot-apocalypse-was-short-zXrvfJCM7

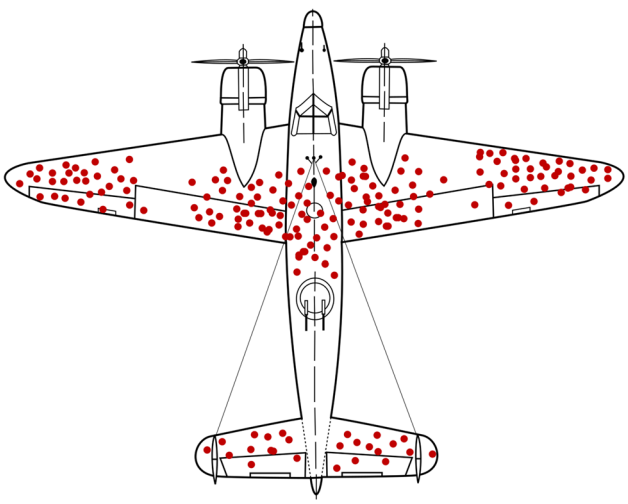

**Fig. 3** Bullet hole distribution on planes returning in WW2. Martin Grandjean, McGeddon, Cameron Moll: https://commons.wikimedia.org/w/index.php?curid=102017718

a mathematician, allegedly realized that survivorship bias was at work here and that the bullet holes on the surviving planes were precisely not in critical areas —because, otherwise, these planes would not have returned. Particularly in the context of military applications of AI, military leaders should pay attention to such forms of bias based on historical data because they could come costly in combat.

In order to improve machine learning regarding ethics *and* performance, Hildebrandt [41] outlines a "loop of contestation" that guarantees stakeholder participation in the design of algorithms. She argues that considering cognitive diversity could vastly improve the performance of AI. She writes the following:

> Taking democracy seriously means that whenever technologies that could reconfigure our environment, are developed, marketed, and employed, we must make sure that those who will suffer or enjoy the consequences are heard and their points of view taken into account. Not merely to be nice, but because they will bring specific expertise to the table and contribute to achieving "robust" societal architectures (p. 109).

These ideas have also military relevance. Cognitive diversity in working teams is related to superior results [101]. Research in management models suggests that cognitive diversity can lead to strategic flexibility [102]. Analogously, the US and the UK militaries have demonstrated interest in harvesting cognitive diversity for military effectiveness [103]. Cognitive diversity could be particularly important in the military since security issues, in general, require the capacity to anticipate unexpected threats and to cope with uncertainty [104].

Concretely speaking, the implementation of agonistic machine learning in the military sector includes the following strategies: stakeholders should be consulted, and developers of code should be required to provide at least one alternative to the modes of representation or datafication they are using. Whilst, due to the security issues involved, it cannot be expected that these deliberative processes take place publicly, they should be implemented at a level that is as public and open to contestation and disobedience as possible. The accumulated data gained from alternative modelling and consulting with stakeholders can be expected to translate into an innovation booster over time. Furthermore, such modes of alternative modelling open to a variety of stakeholders would increase the understanding of AI in civil society and among combatants involved in human–machine-teaming, who need to be able to retrace the way how AI decisions are related to data [59]. Explicability in this sense is a precondition for combatants to perform their duty to critically review orders and disobey if necessary. Yeung et al. emphasise the need to implement robust documentation mechanisms into AI to facilitate judicial overview and auditing [27]. This is also crucial regarding military AI systems. Whilst the approach of Arkin et al. displays a problematic degree of technological solutionism, it is helpful in regard to

facilitating auditing since it insists on the text-based nature of'ethical governor' systems implemented in LAWS [36].

## 5 Fifth section: participation and the military use of AI —a decentralised *Levée en masse*

Due to the fact that representation is always contestable and should be contested, deliberation is inherently incomplete and must be complemented by concrete political participation [38]. Regarding participation in the military use of AI, a model for such participation could be found in Sharp's Civilian-Based Defence —A Post-Military Weapons System from 1985 [105]. Sharp proposed to form an alliance between NATO and critical activists in Western societies in a sense that democracies' enormous reservoir of critical citizens can be transformed into a bulwark against possible attackers.

He argues that an occupying authoritarian regime relying on repression will hardly be able to cope with a civil society that is used to practices of collective resistance and non-cooperation. He makes the case that the outlook of having to deal with such citizens during an occupation could even have a deterrent effect, inasmuch as "the attacked society could deny (…) (aggressors) their goals and impose excessive costs" [106, p. 87]. He proposes strategies to train civil societies in this form of collective nonviolent struggle to achieve this aim.

The Ukraine War demonstrated that Sharp's concepts are not entirely unrealistic. On numerous occasions, citizens have autonomously organised resistance against the Russian invaders [106]. This also included forms of digital resistance, for instance, the autonomously organised move by a 30-year-old IT professional to extract the location of Russian soldiers by using fake profiles of women [107]. Asmolov argues that throughout the different stages of this conflict, a model of 'participatory warfare' emerged, entangling online and offline aspects [43]. He cites practices of open-source intelligence regarding data analysis, geolocation, and the use of civilian drones in conflict, crowdfunding, and logistical support. Similarly, a recent article in *Foreign Affairs* argues that the Ukraine War represents a watershed moment in a new age of open intelligence, citing projects by the Institute for the Study of War and Stanford University [108].

These developments concretise Cronin's earlier speculations that digital technologies might allow for a new kind of *levée en masse* [19]. Whilst the precise definition of *levée en masse* in IHL as the spontaneous uprising of the civilian population against an invading force is raising difficult problems regarding the distinction between civilians and combatants, it is helpful to apply this concept to underline the historical dimension of this development. Similar to the historical origin of the *levée en masse* in the period after the French revolution, democratic participation can provide a military edge to democracies in the cyber domain, particularly if it stays below the level of the use of force. The usage of Starlink by Ukrainian troops demonstrates that private companies from the attacked country and beyond can play their role in digitally-enabled participatory warfare [109].

In an earlier publication, I applied Sharp's model to cybersecurity, emphasising digital literacy as a societal defence against disinformation and election interference [110]. Similar forms of direct participation can be imagined for an AI-driven battlefield. For once, the spread of digital literacy in the population would create increased resilience in civil society, which would make a society, in the long run, more likely to develop active defence mechanisms against military applications of AI, such as hacking but also regarding possible interference in the digital public sphere based on AI-driven bot armies in social media. In the short run, strategies of open-source intelligence regarding data analysis, geolocation, and the use of civilian drones in conflict could play an important role on an AI-driven battlefield.

One of the main findings of Asmolov's research on participatory warfare in Ukraine is that such strategies can make a crucial difference if the defended state is comparably weak [43]. In this situation, "offline horizontal networks and digitally mediated mobilisation relying on different types of online platforms" can temporarily replace the organising function of the state (p.8). Since these modes of participation rely on digital platforms, the algorithms structuring timelines and interactions of volunteers are becoming extremely important in this context. States might choose to develop their own digital platforms to facilitate participatory warfare similar to Landemore's 'Citizenbook' [90], including AI systems focused on logistics to facilitate the decentralised coordination of volunteers. Taking participation one step further, such platforms could be co-created by citizens, as this is discussed in relation to participatory approaches to smart cities [111]. Of course, this raises the question of the fate of such modes of participation if there are attacks on digital communication infrastructure. Reichberg and Syse emphasise the usefulness of autonomously-acting LAWS or swarms of LAWS in such situations [58]. Particularly in participatory warfare, human–machine teams might constitute autonomously acting units based on such technologies.

Even more relevant to AI are the links between civil society actors and civil infrastructure. The development of digital technologies goes partly back to considerations regarding the strategical superiority of decentralised over centralised infrastructure [112]. Analogously, a centralised smart city might be strategically weak because it might be enough to take control over several central nodes to command its traffic system, gas-, electricity-, and financial networks [113]. The literature on warfare and smart cities is focused on the

identification of such vulnerabilities of smart cities regarding cyber-attacks [114].

As a result of the undemocratic tendencies of citizens' data collection in smart cities, decentralised smart cities have been envisioned [115]. Such decentralised structures would also have the advantage to provide new defence capacities. A hypothetical, extremely decentralised version of a smart city, in which every part of infrastructure would be controlled by a different set of stakeholders would constitute a serious challenge to an occupier. In this case, the occupation forces would not only fight the noncooperation and disobedience of civilians but also the resistance of an AI-driven environment that would autonomously trace their moves via sensors [116], predict their advances, and strategise to disrupt their supply chains. Such abilities of intelligent civil infrastructure also relate to civic participation. Various forms of e-participation have been found effective in improving the infrastructure of smart cities, particularly regarding complex problems [117].

## 6 Conclusion and discussion: winning by choosing foresighted securitisation

In the second Section, I criticised inclusive and universalising approaches to AI ethics. In the following, I underlined the differences rather than the similarities between democratic and authoritarian regimes. Accordingly, in the third Section, I criticised the LAWS debate because of its focus on human control regardless of societal circumstances. First, I underlined that human control might profit from being corrected and enhanced by AI-driven systems trained to discern between civilians and combatants and to assess proportionality. Second, I argued that, particularly regarding warfare, concrete human autonomy and responsibility cannot have the same ethical value in authoritarian and democratic societies since authoritarian regimes provide few possibilities to train and exercise such capacities and comply with the duty to disobey illegal superior orders. Furthermore, in the fourth Section, I argued that, instead of focusing on human control regardless of socio-political circumstances, democracies should reconcile the ethical value of autonomy with military applications of AI by linking military AI systems to multi-layered modes of deliberation. This should also enable combatants involved in human–machine teams to perform their duty derived from customary IHL to review orders, understand the relationship between data and AI decision-making, and disobey if necessary. Furthermore, relating the military use of AI to deliberation should enhance cognitive diversity which is likely to constitute a strategic advantage. In the fifth Section, I argued that, following the concepts of Open Democracy and Civilian-based defence, democracies should work towards implementing modes of participation in military AI systems, for example, by strengthening digital literacy in the population, the civilian use of drones, and open source intelligence, and by strengthening the defence capacities of smart cities under the decentralised control of various stakeholders. Parts of this system are depicted in the loops of contestation involving military AI, military actors, and civil society visualised in Fig. 1.

How realistic are these proposals? The biblical 'eye for an eye' still adequately describes the strange mimetic logic of conflict escalation. Conflicts might start because of fundamental differences. However, particularly when it comes to confrontation on the battlefield, these differences often blur. Famously, US diplomat and IR-historian Kennan warned in his *Long Telegram* from 1946 that "the greatest danger" in the confrontation between democracies and authoritarian systems lies in the seduction to "allow ourselves to become like those with whom we are coping" and that democracies "must have courage and self-confidence to cling to our own methods and conceptions of human society" [118].

Making an ethical *and* strategical case, this contribution argues that in confrontations with authoritarian regimes involving military applications of AI, democracies should fight precisely by decidedly sticking to their values and implementing them as deeply into their war machinery as possible. However, emphasising awareness regarding the strategical value of democratic open discourse and cognitive diversity also suggests that democratic openness has its limits. Since cognitive diversity is connected to higher performance and open public spheres likely allow for a greater degree of cognitive diversity to be manifested in data, data from democracies are likely more valuable than data from authoritarian societies with repressive public discourses characterised by distortion owed to censorship and ideology. For instance, former Google CEO Eric Schmidt argued in a recent interview that the dialogue-centred approach of OpenAI and its success with ChatGPT would not be possible in authoritarian China with its restrictions on free speech [96]. It is, therefore, hardly surprising that Beijing promotes the extraction of Western user data [119]. For instance, the seemingly harmless data harvested by Beijing from TikTok's cognitively diverse teenage userbase might be used to counter the democratic offset on an AI-driven battlefield.

Our contribution opens a broad horizon for future research in the fields of ethics, legal philosophy, and political theory, for example regarding modes of civilian-based defence in smart cities and drone warfare and the difficulties to reconcile such participatory approaches of *levée en masse* with robust distinctions between civilians and combatants. Additionally, the implementation of Landemorean Open Democracy into the still hierarchical structures of the military should be discussed further. Moreover, the relationship between the duty to disobey illegal orders, command responsibility, and the AI-driven battlefield outlined here

could be deepened, including but not restricted to the implementation of automated limits regarding the execution of illegal orders. In this context, it should also be discussed further which types of decisions within the OODA-loop can be legitimately automated. Finally and most urgently, experimental psychologists, data scientists, and AI researchers should empirically test my well-founded philosophical speculations and contrast the findings of Beraja et al. [7] and Filgueiras [8] by focusing on differences regarding cognitive diversity in authoritarian regimes and democracies in regard to citizens and their representation in data (which are not the same thing) and the correlation of these differences with higher or lower levels of performance regarding AI.

**Acknowledgements** Many thanks to the two reviewers and Mireille Hildebrandt for their excellent comments on this text and to Emilie van den Hoven for some remarks regarding customary IHL.

**Funding** Johannes Thumfart received funding from the European Union Horizon 2020 research programme under MSCA COFUND grant agreement 101034352 with co-funding from the VUB-Industrial Research Fund.



## References

1. Chen, H.:"'Artificial intelligence: disruptively changing the rules of the game' (人工智能: 颠覆性改变'游戏规则')," *China Military Online*, Mar. 18, 2016. http://www.81.cn/jskj/2016-03/18/content_6966873_2.htm (accessed Sep. 13, 2022).
2. Wallace, R.: How AI founders on adversarial landscapes of fog and friction. J. of Def. Model. Simul. **19**(3), 519–538 (2022). https://doi.org/10.1177/1548512920962227
3. Yan, G.: The impact of artificial intelligence on hybrid warfare. Small Wars Insurgencies **31**(4), 898–917 (2020). https://doi.org/10.1080/09592318.2019.1682908
4. Johnson, J.: Artificial intelligence & future warfare: implications for international security. Def. Secur. Anal. **35**(2), 147–169 (2019). https://doi.org/10.1080/14751798.2019.1600800
5. Kania, EB.:"Battlefield Singularity: artificial intelligence, military revolution, and China's future military power," Center for a New American Security, Nov. 2017.
6. Statista, "Most valuable private AI companies worldwide," Apr. 2021. https://www.statista.com/statistics/1050652/worldwide-artificial-intelligence-startup-unicorns/ (accessed Sep. 15, 2022).
7. Beraja, M.,Yang, DY., Yuchtman, N.:"Data-intensive Innovation and the state:evidence from AI firms in China," Review of economic studies (Preprint), Jan. 2022, [Online]. Available: http://davidyyang.com/pdfs/ai_draft.pdf
8. Filgueiras, F.: The politics of AI: democracy and authoritarianism in developing countries. J. Inf. Technol. Politics (2022). https://doi.org/10.1080/19331681.2021.2016543
9. Glasius, M., Michaelsen, M.:Authoritarian practices in the digital agel illiberal and authoritarian practices in the digital sphere — prologue. Int. J. Commun. 12(0), Art. no. 0. 2018.
10. Lamensch, M.,"Authoritarianism has been reinvented for the digital age," Centre for international governance innovation, Jul. 09, 2021. https://www.cigionline.org/articles/authoritarianism-has-been-reinvented-for-the-digital-age/ (accessed Dec. 29, 2021).
11. Lilkov, D.:"Made in China: tackling digital authoritarianism," Wilfried Martens Centre, Brussels, Belgium, 2020. [Online]. Available: https://www.martenscentre.eu/publication/made-in-china-tackling-digital-authoritarianism/
12. Glasius, M.: Authoritarian Practices in a global age, 1st edn. Oxford University Press, Oxford (2023). https://doi.org/10.1093/oso/9780192862655.001.0001
13. Persily, N., Sun, M.:"The autocrat's digital advantage," presented at the SciencesPo Annual Conference, Dec. 2022. [Online]. Available: https://www.youtube.com/watch?v=LBf3Qz8liLI&ab_channel=SciencesPo
14. Soros, G.:"Remarks delivered at the 2022 world economic forum in Davos," Davos, Davos, May 24, 2022. Accessed: May 28, 2022. [Online]. Available: https://www.georgesoros.com/2022/05/24/remarks-delivered-at-the-2022-world-economic-forum-in-davos/
15. Rudschies, C., Schneider, I., Simon, J.: Value pluralism in the AI ethics debate different actors different priorities. Irie (2021). https://doi.org/10.29173/irie419
16. van den Hoven, J. et al.: "The European approach to artificial intelligence across geo-political models of digital governance," *EasyChair Preprint*, vol. 8818, Sep. 2022, [Online]. Available: https://wwww.easychair.org/publications/preprint_download/rDGkM
17. Kant, I.: Toward perpetual peace: a philosophical sketch. In: Kleingeld, P. (ed.) Toward perpetual peace and other writings on politics, peace, and history, pp. 67–109. Yale University Press, New Haven (2006)
18. Rousseau, D.L., Gelpi, C., Reiter, D., Huth, P.K.: Assessing the dyadic nature of the democratic peace, 1918–88. Am. Political Sci. Rev. **90**(3), 514–533 (1996)
19. Cronin, A.K.: Cyber-mobilization: the new 'Levée en Masse.' US Army War Coll. Q.: Parameter. (2006). https://doi.org/10.55540/0031-1723.2304
20. Everts, P.P.: Democracy and Military Force. Springer, London Palgrave Macmillan UK (2002). https://doi.org/10.1057/9780230509863
21. Reiter, D., Stam, A.C.: Democracy and battlefield military effectiveness. J. Conflict Resolut. **42**(3), 259–277 (1998). https://doi.org/10.1177/0022002798042003003
22. Fukuyama, F.:"A country of their own," Apr. 18, 2022. Accessed: Jun. 09, 2022. [Online]. Available: https://www.foreignaffairs.com/articles/ukraine/2022-04-01/francis-fukuyama-liberalism-country
23. Snyder, T.:"Ukraine holds the future. The war between democracy and nihilism," Foreign Affairs, Oct. 2022, [Online]. Available: https://www.foreignaffairs.com/ukraine/ukraine-war-democracy-nihilism-timothy-snyder
24. Alder, K.: Engineering the revolution: arms and enlightenment in france, 1763–1815. The University of Chicago Press, Chicago, London (2010)
25. Scharre, P.: Army of none: autonomous weapons and the future of war. W.W. Norton & Company, New York (2019)

26 Gentile, G., Shurkin, M., Evans, A.T., Grisé, M., Hvizda, M., Jensen, R.: A history of the third offset, 2014–2018. RAND Corporation, Santa Monica, CA (2021)
27. Yeung, K., Howes, A., Pogrebna, G.: AI Governance by human rights-centered design deliberation and oversight an end to ethics washing. In: Dubber, M.D., Pasquale, F., Das, S. (eds.) The Oxford handbook of ethics of AI. Oxford handbooks series, pp. 77–106. Oxford University Press, New York, NY (2020)
28. Fjeld,J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M.:"Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI," Berkman Klein Center for Internet & Society, 2020. Accessed: Sep. 15, 2022. [Online]. Available: https://dash.harvard.edu/handle/1/42160420
29 Floridi, L., Cowls, J.: (2021) "A unified framework of five principles for AI in society." In: Floridi, L. (ed.) Ethics, governance, and policies in artificial intelligence. Philosophical studies series, vol. 144, pp. 5–18. Cham, Springer (2021). https://doi.org/10.1007/978-3-030-81907-1
30. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Mind. Mach. **30**(1), 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
31. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
32. "Losing humanity: the case against killer Robots," Human rights watch, Nov. 2012. Accessed: Sep. 20, 2022. [Online]. Available: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots
33. Rawls, J.: The law of peoples: with, The idea of public reason revisited. Harvard University Press, Cambridge, Mass (1999)
34. Cook, A.: Taming killer robots. Giving meaning to the 'meaningful human control' standard for lethal autonomous weapon systems, vol 1. JAG School Paper (2019)
35. Grimal, F., Pollard, M.: The duty to take precautions in hostilities, and the disobeying of orders: should robots refuse? Fordham Int. Law J. **44**, 671–734 (2021)
36. Arkin, RC., Ulam, P., B. Duncan, B.: "An Ethical governor for constraining lethal action in an autonomous system," Georgia Institute of Technology, GVU Center, 2009. [Online]. Available: https://smartech.gatech.edu/bitstream/handle/1853/31465/09-02.pdf
37. Habermas, J.: The structural transformation of the public sphere: an inquiry into a category of bourgeois society. In: Studies contemporary German social thought. MIT press, Cambridge (1992)
38. Landemore, H.: Open democracy: reinventing popular rule for the twenty-first century. Princeton University Press, Princeton (2020)
39. Rawls, J.: A theory of justice, Rev Belknap Press of Harvard University Press, Cambridge, Mass (1999)
40. Weymark, J.A.: Cognitive diversity, binary decisions, and epistemic democracy. Episteme **12**(4), 497–511 (2015). https://doi.org/10.1017/epi.2015.34
41. Hildebrandt, M.: Privacy as protection of the incomputable self: from agnostic to agonistic machine learning. Theor. Inquiries Law **20**(1), 83–121 (2019). https://doi.org/10.1515/til-2019-0004
42 Sharp, G.: Making Europe unconquerable: the potential of civilian-based deterrence and defence: Ballinger Pub Co. Mass, Cambridge (1985)
43. Asmolov, G.: The transformation of participatory warfare: the role of narratives in connective mobilization in the Russia-Ukraine war. Digi War (2022). https://doi.org/10.1057/s42984-022-00054-5
44. Borenstein, J., Grodzinsky, F.S., Howard, A., Miller, K.W., Wolf, M.J.: AI ethics: a long history and a recent burst of attention. Computer **54**(1), 96–102 (2021). https://doi.org/10.1109/MC.2020.3034950
45 Yang, G.-Z., et al.: The grand challenges of Science Robotics. Sci. Robot. **3**(14), eaar7650 (2018). https://doi.org/10.1126/scirobotics.aar7650
46. Floridi, L.: Introduction – the importance of an ethics-first approach to the development of AI. In: Floridi, L. (ed.) Ethics, governance, and policies in artificial intelligence. Philosophical Studies Series, vol. 144, pp. 1–4. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-81907-1_1
47. Leibold, J.: Surveillance in China's Xinjiang Region: ethnic sorting, coercion, and inducement. J. Contemp. China **29**(121), 46–60 (2020). https://doi.org/10.1080/10670564.2019.1621529
48. Cho, E.:"The Social Credit System: Not Just Another Chinese Idiosyncrasy," Journal of public and international affairs, no. 5, 2020, Accessed: Oct. 16, 2021. [Online]. Available: https://jpia.princeton.edu/news/social-credit-system-not-just-another-chinese-idiosyncrasy
49. Liang, F., Das, V., Kostyuk, N., Hussain, M.M.: Constructing a data-driven society: china's social credit system as a state surveillance infrastructure: China's social credit system as state surveillance. Policy Internet **10**(4), 415–453 (2018). https://doi.org/10.1002/poi3.183
50. Dirks, E.:"Mass DNA collection in the tibet autonomous region from 2016–2022," citizen lab, university of Toronto, Sep. 2022. Accessed: Sep. 16, 2022. [Online]. Available: https://citizenlab.ca/2022/09/mass-dna-collection-in-the-tibet-autonomous-region/
51. Cong, W., Thumfart, J.: A Chinese precursor to the digital sovereignty debate digital anti-colonialism and authoritarianism from the post-cold war era to the Tunis Agenda. Global Studies Quarterly (2022). https://doi.org/10.1093/isagsq/ksac059
52. Hine, E., Floridi, L.: Artificial intelligence with American values and Chinese characteristics: a comparative analysis of American and Chinese governmental AI policies. AI Soc. (2022). https://doi.org/10.1007/s00146-022-01499-8
53. Floridi, L., et al.: AI4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. Mind. Mach. **28**(4), 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5
54. Erie, MS., Streinz, T.:"The Beijing effect: China's digital silk road as transnational data governance," New York University journal of international law and politics, vol. 54, no. 1, Fall 2021, [Online]. Available: https://deliverypdf.ssrn.com/delivery.php?ID=884112031001096106093017116020007024001024032007049053005122120102085119088112087121124025056115114005124120027101100097108098023039056023040020118000098003000087118093008028091092009006096119123119022004118070115072012006022025028103102114078119065119&EXT=pdf&INDEX=TRUE
55. Bradford, A.: The brussels effect: how the European Union rules the world. oxford university press, new York, NY (2020)
56. van Maanen, G.: AI ethics, ethics washing, and the need to politicize data ethics. DISO **1**(2), 9 (2022). https://doi.org/10.1007/s44206-022-00013-3
57. O'Mara, M.: The code: silicon valley and the remaking of America. Penguin Press, New York (2019)
58. Reichberg, G.M., Syse, H.: Applying AI on the battlefield: the ethical debates. In: von Braun, J., Archer, M.S., Reichberg, G.M., SánchezSorondo, M. (eds.) Robotics, AI, and Humanity, pp. 147–159. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-54173-6_12
59. Swett, B.A., Hahn, E.N., Llorens, A.J.: Designing robots for the battlefield: state of the art. In: von Braun, J., Archer, M.S., Reichberg, G.M., SánchezSorondo, M. (eds.) Robotics AI and

Humanity, pp. 131–146. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-54173-6_11
60. de Vries, B.: Individual criminal responsibility for autonomous weapons systems in international criminal law. In: International humanitarian law series, vol. 65. Brill Nijhoff, Leiden, Boston (2023)
61. Scharre, P.: Four battlegrounds: power in the age of artificial intelligence, 1st edn. W.W. Norton & Company, New York (2023)
62. Scholz, J., Galliott, J.: The Case for Ethical AI in the Military. In: Dubber, M.D., Pasquale, F., Das, S. (eds.) The Oxford handbook of AI. Oxford handbooks series, pp. 685–702. Oxford university press, New York, NY (2020)
63. Galtung, J.: Human Rights: from the state system to global domestic policy. In: Galtung, J., Fischer, D. (eds.) SpringerBriefs on pioneers in science and practice, vol. 5, pp. 157–166. Springer Berlin, Heidelberg, Berlin Heidelberg (2013). https://doi.org/10.1007/978-3-642-32481-9
64. Derviş, K., Ocampo, JA.:"Will Ukraine's tragedy spur UN security council reform?," *Brookings*, Mar. 03, 2022. https://www.brookings.edu/opinions/will-ukraines-tragedy-spur-un-security-council-reform/ (accessed Sep. 20, 2022).
65. Borch, C.: High-frequency trading, algorithmic finance and the flash crash: reflections on eventalization. Econ. Soc. **45**(3–4), 350–378 (2016). https://doi.org/10.1080/03085147.2016.1263034
66. "The zircon: how much of a threat does Russia's Hypersonic missile pose?," Royal united services institute, Mar. 31, 2023. https://www.rusi.orghttps://www.rusi.org (accessed Apr. 03, 2023).
67. Bartneck, C., Lütge, C., Wagner, A., Welsh, S.: An Introduction to ethics in robotics and AI. In: SpringerBriefs in Ethics, pp. 3678–3786. Springer, Cham (2021)
68. How AI is driving a future of autonomous warfare | DW Analysis, (Jun. 25, 2021). Accessed: Oct. 07, 2022. [Online Video]. Available: https://www.youtube.com/watch?v=NpwHszy7bMk
69. Atherton, K.: "Loitering munitions preview the autonomous future of warfare," *Brookings*, Aug. 04, 2021. https://www.brookings.edu/techstream/loitering-munitions-preview-the-autonomous-future-of-warfare/ (accessed Apr. 03, 2023).
70. Heyns, C.: Autonomous weapons systems: living a dignified life and dying a dignified death. In: Bhuta, N., Beck, S., Geiβ, R., Liu, H.-Y., Kreβ, C. (eds.) Autonomous weapons systems, 1st edn., pp. 3–20. Cambridge University Press (2016). https://doi.org/10.1017/CBO9781316597873.001
71. H, Arendt.: Eichmann in Jerusalem: a report on the banality of evil. in Penguin classics. New York, N.Y: Penguin Books, 2006.
72. Bruneau, E., Kteily, N.: The enemy as animal: symmetric dehumanization during asymmetric warfare. PLoS ONE **12**(7), e0181422 (2017). https://doi.org/10.1371/journal.pone.0181422
73. Dinstein Y: The defence of "obedience to superior orders" in international law, Repr. ed., with A new postscript preface. Oxford, UK: Oxford University Press, 2012
74 Allan Williamson, J.: Some considerations on command responsibility and criminal liability. Int. Rev. Red Cross. **90**(870), 303–317 (2008). https://doi.org/10.1017/S1816383108000349
75. Murdough, R.E.: I won't participate in an illegal war: military objectors, the nuremberg defense, and the obligation to refuse illegal orders. Army Law **4**, 4–14 (2010)
76. "Practice relating to rule 155. Defence of superior orders," International Humanitarian Law Databases. https://ihl-databases.icrc.org/en/customary-ihl/v2/rule155 (accessed Feb. 16, 2023).
77. Diver, L.: Law as a user: design, affordance, and the technological mediation of norms. SCRIPT-ed **15**(1), 4–41 (2018). https://doi.org/10.2966/scrip.150118.4
78. Morozov, E.: To save everything, click here: the folly of technological solutionism, Paperback 1. publ. New York, NY: PublicAffairs, 2014.
79. Vyas, D., Chisalita, C.M., Dix, A.: Organizational affordances: a structuration theory approach to affordances. Interact. Comput. (2016). https://doi.org/10.1093/iwc/iww008
80. Bode, I., Huelss, H.: Autonomous weapons systems and international norms. McGill-Queen's University Press, Montreal Kingston London Chicago (2022)
81. Bostrom, N.: Superintelligence: paths, dangers, strategies, 1st edn. Oxford University Press, Oxford (2014)
82. Derrida, J.: Force of law the mystical foundation of authority. In: Cornell, D., Rosenfeld, M., Carlson, D., Benjamin, N. (eds.) Deconstruction and the possibility of justice. Routledge, New York (1992)
83. Foucault, M.: Madness and civilization: a history of insanity in the age of reason. Vintage house, Random House, New York (1988)
84. Lyotard, JF.: The differend: phrases in dispute. In: Theory and history of literature, Vol 46. University of Minnesota Press, Minneapolis. 1988.
85. Douzinas, C.: The end of human rights: critical legal thought at the turn of the century. Oxford ; Portland, Or: Hart Pub, 2000.
86. Mouffe, C.: Which world order: cosmopolitan or multipolar? Ethical Perspect. **4**, 453–467 (2008). https://doi.org/10.2143/EP.15.4.2034391
87. Mills, C.W.: Rawls on race/race in rawls. South. J. Philosophy **47**(S1), 161–184 (2009). https://doi.org/10.1111/j.2041-6962.2009.tb00147.x
88. Calhoun, CJ., Ed.: Habermas and the public sphere, Nachdr. In: Studies in contemporary German social thought. Cambridge, Mass.: MIT Press, 2011.
89. Habermas, J.: Reflections and hypotheses on a further structural transformation of the political public sphere. Theory Cult. Soc. **39**(4), 145–171 (2022). https://doi.org/10.1177/02632764221112341
90. Landemore, H.: Open democracy and digital technologies. In: Bernholz, L., Landemore, H., Reich, R. (eds.) Digital technology and democratic theory, pp. 62–89. University of Chicago Press (2021)
91. Knight, W.:"The Dark Secret at the Heart of AI," MIT Technology Review, Apr. 2017, Accessed: Sep. 30, 2022. [Online]. Available: https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/
92. Pasquale F: (2015) The black box society the secret algorithms that control money and information. Harvard University Press, Cambridge
93. Mouffe, C.: On the political. In: Thinking in action. Routledge, New York, London, 2005
94. Hansen, L., Nissenbaum, H.: Digital disaster, cyber security, and the copenhagen school. Int. Stud. Quart. **53**(4), 1155–1175 (2009)
95. Landemore, H.: Democratic reason: politics, collective intelligence, and the rule of the many. Princeton University Press, Princeton; Oxford (2013)
96. Fmr. Google CEO eric schmidt on the consequences of an A.I. revolution, (Mar. 23, 2023). Accessed: Mar. 29, 2023. [Online Video]. Available: https://www.youtube.com/watch?v=Sg3EchbCcA0
97. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: "Machine bias," ProPublica, May 2016. Accessed: May 28, 2022. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=TiqCeZIj4uLbXl91e3wM2PnmnWbCVOvS
98. Barocas, S., Selbst, A.D.: Big data's disparate impact. Calif. Law Rev. **104**, 671–732 (2016). https://doi.org/10.15779/Z38BG31

99. Mattu, J., Larson, J.,Angwin, L., Kirchner,S.: "How we analyzed the COMPAS recidivism algorithm," ProPublica, May 2016. Accessed: May 28, 2022. [Online]. Available: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=BqO_ITYNAKmQwhj7daSusnn7aJDGaTWE
100. Casselman, B.:"The Legend of Abraham Wald," American Mathematical Society, Jun. 2016, Accessed: Feb. 22, 2023. [Online]. Available: http://www.ams.org/publicoutreach/feature-column/fc-2016-06
101. Hong, L., Page, S.E.: Groups of diverse problem solvers can outperform groups of high-ability problem solvers. Proc. Natl. Acad. Sci. U.S.A. **101**(46), 16385–16389 (2004). https://doi.org/10.1073/pnas.0403723101
102. Nowak, R.: Foundations of strategic flexibility: focus on cognitive diversity and structural empowerment. MRR **45**(2), 217–235 (2022). https://doi.org/10.1108/MRR-02-2021-0130
103. Slapakova, L., et al.: Leveraging diversity for military effectiveness: Diversity, inclusion and belonging in the UK and US Armed Forces. RAND Corporation, Santa Monica CA (2022). https://doi.org/10.7249/RRA1026-1
104. Burgess, J.P.: The ethical subject of security: geopolitical reason and the threat against Europe. Routledge, Milton Park, Abingdon, Oxon New York (2011)
105. Sharp, G.: Civilian-based defense . A post-military weapons system. Princeton University Press, Princeton (1990)
106. "2022 protests in Russian-occupied Ukraine," *Wikipedia*. Sep. 11, 2022. Accessed: Sep. 23, 2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=2022_protests_in_Russian-occupied_Ukraine&oldid=1109742331
107. M. Srivastava, "Ukraine's hackers: an ex-spook, a Starlink and 'owning' Russia," Financial Times, Sep. 04, 2022. [Online]. Available: ft.com/content/f4d25ba0–545f-4fad-9d91–5564b4a31d77
108. Zegart,A.:"Open Secrets," Foreign Affairs, no. January/February 2023, Dec. 20, 2022. Accessed: Feb. 20, 2023. [Online]. Available: https://www.foreignaffairs.com/world/open-secrets-ukraine-intelligence-revolution-amy-zegart
109. Panella, C.:"Starlink is key to Ukrainian operations, but the Russians 'will find you' if you use it too long, soldier says," *Business Insider*, Mar. 24, 2023. Accessed: Mar. 29, 2023. [Online]. Available: https://www.businessinsider.com/starlink-key-ukrainian-operations-used-too-long-russians-will-find-2023-3
110. Thumfart, J.: Public and private just wars: distributed cyber deterrence based on Vitoria and Grotius. IPR (2020). https://doi.org/10.14763/2020.3.1500
111. Leclercq, E.M., Rijshouwer, E.A.: Enabling citizens' Right to the smart city through the co-creation of digital platforms. Urban Transform **4**(1), 2 (2022). https://doi.org/10.1186/s42854-022-00030-y
112. Baran, P.: Some perspectives on networks - past, present and future. Inf. Process. **77**, 459–461 (1977)
113. Asan, H.: Data security. In: Artificial intelligence perspective for smart cities, 1st edn., pp. 253–276. CRC Press, Boca Raton (2022). https://doi.org/10.1201/9781003230151-12
114. Kovalsky, M., Ross, R.J., Lindsay, G.: Contesting key terrain: urban conflict in smart cities of the future. Cyber Def. Rev. **5**(3), 133–150 (2020)
115. Feder-Levy, E., Blumenfeld-Liebertal, E., Portugali, J.: The well-informed city: A decentralized, bottom-up model for a smart city service using information and self-organization. In: 2016 IEEE international smart cities conference (ISC2), Trento, Italy: IEEE, Sep. 2016, pp. 1–4. doi: https://doi.org/10.1109/ISC2.2016.7580767.
116. Enlund, D., Harrison, K., Ringdahl, R., Börütecene, A., Löwgren, J., Angelakis, V.: The role of sensors in the production of smart city spaces. Big Data Soc. **9**(2), 205395172211102 (2022). https://doi.org/10.1177/20539517221110218
117. Allen, B., Tamindael, L.E., Bickerton, S.H., Cho, W.: Does citizen coproduction lead to better urban services in smart cities projects? An empirical study on e-participation in a mobile big data platform". Gov. Inf. Q. **37**(1), 1012 (2020). https://doi.org/10.1016/j.giq.2019.101412
118. "George Kennan's 'Long Telegram,'" Feb. 22, 1946. https://nsarchive2.gwu.edu/coldwar/documents/episode-1/kennan.htm (accessed Feb. 21, 2023).
119. Kokas, A.: Trafficking data: how china Is winning the battle for digital sovereignty. Oxford University Press, New York (2022). https://doi.org/10.1093/oso/9780197620502.001.0001