



Military robots should not look like humans

Kamil Mamak^{1,2} · Kaja Kowalczevska³

Accepted: 7 August 2023 / Published online: 17 August 2023
© The Author(s) 2023

Abstract

Using robots in the military contexts is problematic at many levels. There are social, legal, and ethical issues that should be discussed first before their wider deployment. In this paper, we focus on an additional problem: their human likeness. We claim that military robots should not look like humans. That design choice may bring additional risks that endanger human lives and by that contradicts the very justification for deploying robots at war, which is decreasing human deaths and injuries. We discuss two threats—epistemological and patient. Epistemological one is connected with the risk of mistaking robots for humans due to the limited ways of getting information about the external world, which may be amplified by the rush and need to fight with robots in distance. The patient threat is related to the developing attachment to robots, that in military contexts may cause additional deaths by the hesitance to sacrifice robots in order to save humans in peril or risking human life to save robots.

Keywords Military robots · Design · Human–robot interactions · Anthropomorphism

Introduction

In recent years we are experiencing a rise in interest in the deployment of robots and the societal, legal, and ethical aspects of their applications (cf. Bertolini & Aiello, 2018; Coeckelbergh, 2022). Specific issues concern particular domains in which robots are deployed, like there are specific issues with autonomous cars (cf. Mamak & Glanc, 2022; Nyholm, 2018) or sex and love robots (cf. Devlin, 2018; Mamak, Forthcoming; McArthur et al., 2017; Mamak, 2022), companion robots (cf. Danaher, 2019a; Nyholm & Smids, 2020), healthcare robots (cf. Coeckelbergh, 2018; Sparrow & Sparrow, 2006), police robots (cf. Asaro, 2016; Mamak, 2023). In this paper, we focus on the military context (cf. Sparrow, 2007).

In an equally obvious way, the development of technology affects modern military operations, which are virtually impossible to conduct without relying on disruptive technologies, the development of which is aimed at bridging the military capability gap (Farrant & Ford, 2017, pp. 393–99). Military robots are increasingly being used in all battlefield spaces, namely air, land, water and cyberspace. Some are modelled on animals (e.g. snakes, insects, birds, fish, marine and terrestrial mammals), with an increasing number being miniaturized and combined into swarms controlled by a single human operator. Despite the existence of different narratives for future robotic warfare, according to experts, the most likely scenario is one in which robotic systems will operate as cobots to support, rather than replace, the actions of human soldiers (Schmitt & Thurnher, 2012; Harris, 2016, p. 79; Scharre, 2016, p. 164).

The major military and technological powers (China, Israel, Russia, USA) created special units in government responsible for integrating algorithms, artificial intelligence and machine learning into military operations (Lewis et al., 2017; Sweijns & De Spiegeleire, 2017). We are therefore experiencing a new arms race (Bode et al., 2023).

There are many potential legal, societal, and ethical issues with military robots. This paper focuses on two normative frameworks that may be relevant in the context of their human likeness. The first is ethical, which is focused on the value of

✉ Kamil Mamak
kamil.mamak@uj.edu.pl

Kaja Kowalczevska
kaja.kowalczevska@uwr.edu.pl

¹ RADAR: Robophilosophy, AI Ethics and Datafication Research Group, University of Helsinki, Helsinki, Finland

² Department of Criminal Law, Faculty of Law and Administration, Jagiellonian University, Kraków, Poland

³ Digital Justice Center, University of Wrocław, Wrocław, Poland

the life of human beings. The second one is related to human likeness as a problem for the international humanitarian law (IHL). Those two frameworks are not entirely coherent with each other—during the armed conflict, the killing of combatants is legally permissible.

There are no human-like military robots yet on the battlefield. However, it is reasonable to assume that such robots may be in the future. Humanoid robots are starting to be deployed in other areas (sex robots, healthcare robots). Our environment is adapted to humans with particular heights, legs, and arms (stairs, doors, existing equipment, and so on), which may be a possible factor that may impact how robots are designed. The concept of a human-like soldier is also embedded in the culture—in movies and books. Moreover, the Atlas robot, a humanoid robot by Boston Dynamics, was developed for US military agency DARPA (Fox Van, 2017). We claim that military robots should not look like humans because they raise additional risks for human life, which, in a sense, contradicts the main justification of the deployment of robots in military contexts. Robots, by resembling humans may be easy to mistake for humans, and their human likeness may trigger psychological reactions in fellow humans that put in danger the others that are around.

This paper is structured as follows. After the introduction, we focus on the general—legal, and ethical aspects of using robots in military contexts. The following section is focussed on the specific issues of the human-likeness of robots. The paper ends with conclusions.

Before going further, we want to make some clarificatory remarks. We want to add that we are aware of the definitional differences arising from the interdisciplinary nature of research on robots, algorithms and artificial intelligence (AI), but for reasons of facilitating the narrative and conveying the complex nature of this field of research, we use these terms interchangeably in this article.

We focus here on military robots. One of the popular definitions of robot refers to the sense-think-act paradigm (cf. Gunkel, 2018b; Jordan, 2016; Thrun, 2010). In short, "sense" refer to the possibilities of gaining information about the external world, the "act" means the ability to impact that world, and the "think" component refers to the possibilities of analyzing information and transforming them into actions. The last element is linked to autonomy. We treat robots here broadly. It includes (potential) fully autonomous entities (probably AI-based) and human-controlled units with little or no autonomy at all (drones). By "military" we mean robots that are used in a military context, especially in the conduct of hostilities.

Use of robots in a military context

In this section, we present the context for the deployment of robots and AI on army equipment, their applications, associated risks, benefits and legal challenges. We here refer also to the literature focused on AI, and later, we focus on embodied robots. These aspects are related. Robots' development on the battlefield depends on the development first the AI. Some definitions of robots refer directly to them as embodied AI (see, e.g. Winfield, 2012, p. 8). Problematic issues associated with AI in the battlefield context may be amplified by topics related to embodied agent design (cf. Sparrow, 2021).

The debate on the use of robots for military purposes has gained momentum with the widespread use of drones (unmanned, mainly flying vehicles). The large-scale use of drones, especially in non-international armed conflicts, has contributed to the debate on the legal and ethical aspects of using new technologies on the modern battlefield. In particular, attention was drawn to the reinforcement of the asymmetric nature of warfare in the case of conflicts between a technologically advanced actor and developing states or non-state actors.

From a legal point of view, drone use has influenced the understanding of concepts such as the use of force and the scope of the right to self-defence (Heyns et al., 2016; McNab & Matthews, 2010), the supervision of the targeting process by human operators (human in the loop) or the mere temporal and geographical scope of the application of international law (Crawford, 2020). At the same time, considerable space in the doctrinal debate has been devoted to the ethical and psychological aspects of remote warfare. The progressive dehumanisation of the battlefield, which results from the increasing use of algorithmic processes and the removal of human soldiers from the battlefield, has been a source of questions about the ethos of modern warriors (Sajduk, 2015) or the moral permissibility of targeting the enemy remotely (Bober, 2015; O'Connell, 2009).

Nevertheless, for more than a decade, the subject of disruptive military technology has been dominated by the development of AI. Unlike drones, the decision-making process in AI-equipped military robots would remain outside human oversight (human out of the loop). This raises a number of legal and ethical challenges that need to be examined in the context of AI's promise of effectiveness and utility.

Military applications of robots

Although the legal debate on military robots has been expanded most around killer robots (which are discussed

below), these are not the only applications of AI for the military. In fact, the largest area of AI support for military capabilities is decision-support rather than decision-making (Cai et al., 2012; Schubert et al., 2018). Gathering and analysis of big data retrieved from internal sources and complex operation environment of the battlefield may bring clarity as to the categorization of certain objects and persons, identification of anomalies and prediction of possible scenarios. Deeks considers the use of AI-fuelled decision support systems in armed conflict settings for tasks such as: the detention review and release decisions, threat-recognition or proportionality assessment (Deeks, 2022, pp. 45–46).

The final output of the AI-fuelled systems should inform legal decision instead of replacing them. Such applications of AI are intended to help address human limitations in being able to analyse large volumes of data quickly and, by design, to harmonise and standardise decision-making interpretations. Although human control and decision-making are retained in such systems, this does not mean that they remain unproblematic. The first challenge is the overall encodability of IHL regulating the conduct of hostilities, which largely consists of highly context-dependent, open-textured and therefore highly indeterminate norms (Deeks, 2022, p. 53). The second most important issue is the non-transparency of the AI process (Kwik & Van Engers, 2021). With that said, it is important to note that it is still unclear how human actors move from qualitative to quantitative judgments (e.g. the determination of punishment in view of the proven circumstances of a crime, the military commander's recognition that a planned attack is in line with the principle of proportionality). In this context, it can be considered that human judgment is also not transparent, although we still accept its role more than decisions made by AI.

The most controversial application of robots on the battlefield is lethal autonomous weapons systems (LAWS). Since 2013, they have been discussed by experts and states at the United Nations Convention on Certain Conventional Weapons concluded at Geneva on October 10, 1980 (CCW), which restricts and prohibits the use of certain weapons. Despite the lack of formal negotiation of treaty solutions (Kayser, 2023), the CCW forum serves as a global venue for discussion of the transfer of human life and death decisions, therefore targeting process, to AI (Kowalczevska, 2021). The biggest achievement of this process was the adoption of 11 non-legally binding Guiding principles on the development and use of LAWS in 2019 (CCW/GGE.1/2019/3, 2019). Outside the area of interest were other applications such as the aforementioned decision-support systems or the use of military robots in rescue operations, logistics and transportation, bomb disposal or combat simulation and soldier training. It has been assumed that LAWS are understood to be those weapon systems that, once activated, can

identify, select, and engage targets with lethal force without further intervention by an operator, although the individual positions of states in this regard may differ slightly (CCW/GGE.1/2023/CRP.1, 2023).

Legal challenges

Given the vastness of AI applications in the military, the discussions at the CCW forum represent only a slice of the issues raised. At the same time, they encompass that element which is crucial to humanity's entire approach to AI. The targeting process can result in the deprivation of life and is therefore of the greatest concern from a legal and ethical perspective. This is why the CCW discussions take into account operational, legal and ethical issues arising from IHL and human rights law.

As far as operational issues are concerned, these primarily stem from the usefulness of AI on the battlefield. Robot systems are often shown as alleged force multipliers: while the range and time duration of military operations can be increased, the need to send large numbers of soldiers to the front is decreased, which reduces costs, but also lowers the risk of losses and inflicted suffering (Lewis, 2018; Marchant et al., 2011). Shaw argues even that the conduct of hostilities would be cleaner (Shaw, 2005). It is not uncommon to come across such slogans as that robots do not rape (Heyns, 2010) and that they are perfectly suited for the 4D missions consisting of tasks that are too monotonous (dull), performed in contaminated conditions (dirty), difficult and dangerous for humans [“Robotics (Drones) Do Dull, Dirty, Dangerous & Now Difficult”, 2018]. Nevertheless, even proponents of LAWS development understand that autonomy implies certain trade-offs, particularly regarding human control and accountability regimes, the regulation of which requires prudence and consideration of the following IHL principles.

The principle of distinction requires that attacks be directed only at military objectives (human and non-human), thus classifying persons and objects as protected from attack or not (Grzebyk, 2022). The principle of proportionality requires the determination of the direct military advantage gained from an attack and the foreseeable damage as the basis for deciding whether or not to launch an attack (Zajac, 2023). The precautionary principle imposes an obligation on belligerents to exercise constant care and take all feasible precautions to minimise civilian losses (Thurnher, 2018). In addition, a number of principles oblige belligerents to provide assistance to the wounded, sick and survivors and to treat prisoners of war appropriately or to protect objects of special status such as cultural property, medical facilities or places of worship (Sassoli, 2014; Davison, 2018). The above principles are intended to contribute to IHL's primary objective of reducing the losses and suffering caused by war. These norms are the source of principles that, unlike rules,

do not operate on a zero-sum basis and are therefore open-textured and require human judgement and interpretation. This action is a very demanding process for human belligerents and therefore, in the current state of AI development, even more so for the technology in question (cf. Arkin et al., 2012; Zurek et al., 2023).

In the legal context, beyond the technical feasibility of compliance with IHL principles, the most important issue is the attribution of individual activity for LAWS actions. The so-called „accountability gap” (Docherty, 2015) stems from the problems of ensuring the explainability of the processes occurring in LAWS, the specific and distributed process of creating algorithms and neural networks, as well as the issue of demonstrating mens rea, i.e. a mental state indicating intent (of the robot or its creator). The end state is for black box processes to become white boxes, so that it is possible to understand at what stage the "mistake" occurred that caused the breach and which human being can bear the appropriate responsibility for it (Vries, 2023) The responsibility of the state, which uses LAWS, is not problematic in this respect as it is based on the principle of objectivity (Boutin, 2023).

General ethical challenges

The issue of accountability is part of the problematic of the dehumanisation of war and is closely linked to a concept that has so far remained outside the focus of IHL. It concerns human control over decision-making processes and, more specifically, the concept of meaningful human control (MHC). MHC may in the future become a legal norm, but it finds its axiological grounding in ethics, or more precisely in the dictates of public conscience (Kowalczevska, 2019). Among other things, the report *Losing Humanity* highlighted the moral problematic of transferring life-and-death decision-making from humans to non-humans (Docherty, 2012). It became a trigger for an analysis of how, in previous methods and means of warfare, humans exercised control over this process and what this should look like with the advent of AI (Christen et al., 2023). This is the most discussed ethical issue in the context of LAWS, although not the only one.

Recently, the concept of Responsible AI (RAI) being developed in the context of military applications by countries such as the USA (U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Strategy, 2022), UK (Ambitious, Safe, Responsible: Our Approach to the Delivery of AI-Enabled Capability in Defence, 2022) or France (Report of the AI Task Force September, 2019), has also received particular attention. The RAI is based on the principles according to which: AI should be developed in accordance with national and international law (lawfulness), human responsibility should be clearly assigned and the use of AI should be done with consideration and care

(responsibility and accountability), AI applications should be subject to transparent and understandable procedures, reviews and methodologies (explainability and traceability), AI use cases should be well defined and security and robustness should be ensured throughout the life-cycle of these capabilities (reliability), adequate human-machine interaction should be ensured and safety measures such as disengagement or deactivation in case of unintended behaviour should be applied (governability) and proactive measures should be taken to reduce bias (NATO, n.d.; REAIM 2023, 2023). In general, the above ethical principles can be considered common to both military and civilian applications of AI, as apart from the issue of lethal applications, the challenges are very similar (Recommendation on the Ethics of Artificial Intelligence—UNESCO, 2022; Ethics Guidelines for Trustworthy AI | European Comission, 2019).

Ethical issues are also linked to psychological aspects, including how soldiers will interact with robots (Galliot & Wyatt, 2020). And while IHL is extremely sparse when it comes to psychological harm caused by war (with the exception of the use of terror as a weapon against civilian population), psychological issues are highly relevant to unit cohesion, morale of soldiers and operational capabilities. As a result, they can be of momentous importance for the conduct of hostilities. Surprisingly, these aspects were not addressed at all at the CCW. The debate revolved around issues related to the guts of the robot, i.e. AI, and the environment in which it operates, i.e. modern battlefield. We believe that the debate lacks an analysis of what the robot itself is supposed to look like.

At the initial stage of the discussions, while the image of Robocop or Atlas was one of the fastest brought to mind when trying to visualise LAWS, there was only a cursory mention of the Android fallacy and the risk of the anthropomorphization of the robots that would replace the soldiers. Due to the lack of specific LAWS models to be analysed (there are still no clear positions as to whether such robots already exist), discussions by constraint took place at a theoretical and general level. Hence, it was often emphasised that humanising verbs such as "decide", "think", "see" or "feel" should not be misused when describing the operation of LAWS. And while some consensus has emerged in the linguistic layer and LAWS are explicitly portrayed as means of warfare, combat systems, pieces of equipment, this does not change the fact that on the actual battlefield these robots can be perceived as humans. This risk and the subsequent threats may materialise in a scenario where military robots take the shape and behaviour of humans.

Legal challenges to human-like robots

From a legal perspective, human-like robots, should be classified unequivocally as military equipment and therefore

military objectives by nature (Grzebyk, 2022, p. 124). There should be no doubt that such robots do not have combatant status and consequent prisoners of war status—they should be treated as objects in any case. Their introduction into army equipment is difficult to justify from an operational and legal point of view. Given that the human-like robot is a part of military equipment it should be appropriately marked with the badges and symbols of the belligerent. It certainly cannot resemble civilians, wounded, prisoners of war, religious or medical personnel, as this would constitute an act of perfidy and therefore a war crime. In theory, it is not illegal to use robots to usurp combatants (human soldiers) as a ruse of war. However, the indirect consequences of such an action may have a negative impact on the adherence of the parties to the conflict to the principle of distinction, proportionality and precautions in attack.

Human-like robots can add further confusion to the modern battlefield, which is complicated and demanding enough for human soldiers even without them. The development and use of such a means of warfare should be preceded by a legal review that considers the legal, ethical, political and medical implications of using such robots (McFarland & Assaad, 2023). This should be combined with a risk assessment and the introduction of mitigation measures. However, given IHL's goal of minimising incidental loss of life, injuries to civilians and damage to civilian objects, it is impossible to defend such a robotic design from legal perspective.

Ethical issues with the human-likeness of military robots

In this section, we focus on the appearance of military robots as an ethical issue and the issue for IHL. The main claim of this paper is that military robots should not look like humans. We believe that the human shape is in contradiction with one of the main reasons for the use of robots in military settings, which is to decrease the number of human victims of war. If the reasons for deploying military robots concern human life, then the robots should not look like humans. This claim does not mean that we affirm the use of robots in the first place, but if there is a willingness to use robots in a military setting, then we should pay attention to the consequences of their human likeness.

Before going further, we want to explain what we mean by "looking like humans". We understand it broadly; it includes as well the situations in which the robots are in form looks like humans and, at first glance, are indistinguishable from them, as well as the situation in which robots from a distance may resemble humans, which means they are in the height of a human, are going on two legs, have hands and so one. The first example is not here yet outside

of popular culture (books, movies), but is doable at least potentially.

Below we dive into arguments in two groups. We draw here on Mamak's chapter entitled "Challenges of the legal protection of human lives in times of anthropomorphic robots" (Mamak forthcoming). He identifies two threats that are connected with the rise of anthropomorphic robots, the "epistemological threat" and the "patient threat". Epistemological one is connected with the limited ways in which humans are collecting information about the world. While the patient threat is related to the tendency of humans to sympathize with robots. Both threats are important in the military context.

Epistemological threat

Now we go deeper into those two threats starting with the epistemological one. As mentioned, humans have a limited apparatus from which we get information about the external world. We cannot, for example, be sure about the internal states of other people and we have no direct access to them. In philosophy, there is a popular thought experiment regarding zombies and the different issues connected with them (cf. Kirk, 2021; Vélez, 2021). One of them is how we should treat entities that look and behave like humans but do not have internal states of humans. Danaher refers to this example in the context of robots and is wondering how we should treat robots that look like entities that possess moral statuses, like humans and animals (Danaher, 2019b). He concluded that due to our epistemological limitations, it is reasonable to treat them as entities that possess such status. He calls his position ethical behaviorism because he is focused on the observable features (look and behavior) of robots.

In practice, if there is a robot that looks and behaves like a human, then it would be hard to distinguish it easily from humans. In military contexts, it may constitute a threat to human life. Now we will explain in which ways. The first problem is that if there are human-like robots adopted, then all entities that are present on the battlefield are potentially military robots. Even if the attacker wants to destroy robots and not humans, it may be an issue to distinguish between those two categories. In the military setting, there is another issue that acts to the disadvantage of humans—compared to, for example, robots in events—which is time. This issue is connected with the ethical framework focused on the value of human life.

The confrontation with robots may be deadly to human soldier, so it may be crucial to decide on its liquidation as soon as possible. The threat to the life of the attacker—the less time for making informative decisions, the bigger chance of accidentally harming humans. Even if the differences are trackable after evaluating the nature of the entity, then in a military context, time works against human safety.

Soldiers may be more willing to destroy equipment than kill a human being (even if both actions are legally permissible), but making informative decisions may be hindered by the threat related to the possibility of being endangered in close confrontation with the robot. This is why it is also problematic to create a robot that looks like humans only in a superficial way that is about the human size and is going on two legs. Such robots, from the distance, could look like humans, and again, if the direct confrontation with the robots is threatening to humans, then it may be reasonable to liquidate it from the distance, and this also increases the chances of mistakes with humans.

The existence of human-like robots in the military zones also creates a risk of providing a way of escaping from responsibility in case of the killing of a human being. It is related to the issues of differentiating between legitimate and non-legitimate targets. In short, a person targeting at a robot who happens to be a human (civilian) may not bear responsibility for a crime against a human being that is a non-legitimate target. It is connected with the institution of mistake of the fact, which could justify the perpetrator (cf. Garvey, 2009; Woodruff, 1958). It is applied in the situation if, for example, there is hunting and the person is shooting at an entity that is in the bush, is on for leg, size of the boar, and is giving a sound of boar. The shooting person has reasonable grounds to believe that it is a boar but a human instead. The person would not bear responsibility for that act, even if the person died due to the shoot. This seems justified if the person really thinks it is attacking a robot. But there is also a problem with using this justification in cases where the person deliberately shoots a protected person. The person who is about to be under investigation may use such an excuse to try to escape responsibility for causing the death human who is a civilian. A person may claim that intends to shoot the robot or combatant, not a protected person. Therefore such a mistake of fact could negate the mental element required by the crime (according to the Article 32(1) of the Rome Statute of the International Criminal Court). The more human-like robots are, the more plausible is escaping from responsibility in that way. The line of argumentation may be that the person who was shot from a distance was a civilian, but the attacker made that decision because, from a distance thought that it was a human military objective (or a military robots), and because the robot may be more dangerous while being closer, the decision was made in the state of uncertainty or mistake that was justified—in the eyes of the decision maker—by the threat for their life. We are not claiming that it may happen often, but we point out the possibility of additional arguments that may appear while deploying robots that resemble humans.

To summarize the epistemological threat, if military robots look like humans, it increases the risk of being mistaken for humans. Both humans and human-like robots

may look like military robots which put additional risks on humans. The threat not only concerns the robots that are hard to distinguish from humans but all robots that are in more or less the shape of humans, because the decisions on the attack on them may be made from the distance from the object and in a hurry, both aspects increase the chances of mistake. Human-likeness creates also a possibility of escaping from responsibility by calming that the intended aim was a robot and not a human.

Patient threat

A patient threat is not as straightforward as an epistemological threat that is based simply on the appearance of a robot. The patient threat concerns the possible attachment to human-like robots and is connected with anthropomorphization, a human tendency to see human-like qualities in non-human entities and events (cf. Guthrie, 1997).

There is a growing body of literature on human–robot interactions showing that humans treat robots not as objects but as something more. For example, Salvini et al. l. shows that people are treating the attack on robots not as vandalism but rather as bullying (Salvini et al., 2010). People do empathize with robots "suffering", they feel empathy toward them if they are under attack (cf. Rosenthal-von der Pütten et al., 2013; Rosenthal-von der Pütten et al., 2014; Suzuki et al., 2015; Malinowska, 2021).

In one study Nijssen et al. aim to examine the impact of anthropomorphism on human behavior in situations of peril and show that some people hesitate to sacrifice robots to save a human being. The experiments were based on the following idea:

"a group of people is in danger of dying or getting seriously injured, but they can be saved if the participant decides to perform an action that would mean sacrificing an individual agent (human, human-like robot, or machine-like robot) who would otherwise remain unharmed." (Nijssen et al., 2019, pp. 45–46).

In some countries, there is a duty to rescue, sometimes titled as samaritan law (cf. Feldbrugge, 1965; Heyman, 1994; McINTYRE, 1994; Pardun, 1997). If the robot is not destroyed for the purpose of saving humans in peril, then in could constitute a crime (Mamak, 2021).

In the military context, the over-attachment to robots may also be problematic; the people should have priority in being saved, and the feeling toward robots may be a burden that stops humans from acting appropriately. It is related to the ethical framework concerned with human life's value. Here is also the problem of time that was mentioned before, the decision might be made quickly, and the human-likeness of robots is an additional problem that may slow the decision. What needs to be added is the attachment to robots is not only possible if robots resemble

humans; it is also possible in the case of other robots. Even in the military context, there are known stories of treating robots as members of the team, there are for example stories of funerals made for robots by fellow soldiers (cf. Garber, 2013). Darling points out that the crucial aspect that may arise in such responses to robots is movement, if the robot is moving, then it could be interpreted by as a living object that may trigger additional responses (Darling, 2021). But it may be said that the more human-like robot is the more feeling and human-like qualities we could attribute to robots.

In this case, the problem is not the fact that we may mistake robots for humans. We know that we are dealing with robots but their features trigger some responses that are dangerous to other human beings.

This threat is also different in the case of groups of potential victims. The human-likeness threatens civilians and co-combatants. Soldiers may have a problem leaving the robot behind in dangerous situations due to their attachment to them. The hesitation in leaving behind or sacrificing robots in order to save others may cost the real lives of real humans. This is a threat that also the side that is using robots should take into account. Their own soldiers may be endangered by the overattachment of fellow robot soldiers, which may hinder the public support of a given nation for the support of deploying robots.

Way of mitigation of the threats

In response to described threats, Mamak proposed numerous measures that may decrease the negative effect on human safety, such as the call for making robots easily distinguishable from humans (Mamak, 2021; forthcoming). However, in the military context, the measures are doubtful, and it is more justified to expect to abandon the human shape from military robots. Such a proposition is made by Bryson, who is concerned about the human (emotional) responses to robots and proposes to make them in a form that does not trigger unjustified, by the nature of entities, responses (Bryson, 2018). Her proposition seems too broad to apply to all robots (like a companion or sex robots) (cf. Danaher, 2020; Gunkel, 2018a), but in military robots is justified. Taking into consideration what is at stake—which is human life, and the aspect of the military setting which is the pressure of time, that does not allow to spend too much time of think it is better to avoid human-likeness in the design of the military robots.

Resignation from the human-likeness of robots may resolve almost entirely epistemological threat and limit the patient threat. Limit, and not resolve, because the soldiers may develop attachment also to non-humans like robots.

Conclusions

There is an ongoing discussion about using robots in the military context. Many crucial decisions need to be made before deploying them on the battlefield. In this paper, we focus on the specific issue of their design. We claim that the design choices that will make military robots look like a human may bring risks to human lives and therefore undermine the objectives of IHL. Those risks would not exist or be significantly lower if the robots would not look like humans. We point to the problem of the epistemological limitations of humans, who may mistake humans for robots. The other threat that we talk about is the patienty threat which focuses on the possibility of treating robots in a way not justified by their ontological features. While outside of the military context, it is not something obviously bad, in the military context, it brings additional risks to humans, who may not be rescued or who may lose their life to saving robots. We recommend not building robots that look like humans.

The argument presented in this paper—avoidance of human-like design—could be relevant for different fields of application for robots, but not entirely. Other applications could have their own specificity that needs to be taken into account. For example, there is a discussion on the possible negative impact of sex robots (cf. Devlin, 2018; Richardson, 2015, 2016). Those worries are related to the fact that sex robots represent human beings, but it seems that the solution to those worries cannot be just the ban on creating sex robots that resemble humans (cf. Danaher et al., 2017). It would contradict the whole idea of sex robots. Specific issues of human likeness may appear in specific contexts, for example, in traffic, where the human-like robots may be "confusing" for traffic participants (humans and autonomous cars). As mentioned before, Mamak claims that robots in those situations should be easily distinguishable from humans to set priorities based on the nature of the objects and not their appearance (Mamak, 2021).

Funding Open Access funding provided by University of Helsinki including Helsinki University Central Hospital. (1) Academy of Finland, decision number 333873; (2) It came about within the framework of Academic Excellence Hub—Digital Justice Center carried out under Initiative of Excellence—Research University at the University of Wrocław.

Declarations

Conflict of interest Authors declare no conflict of interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Ambitious, Safe, Responsible: Our Approach to the Delivery of AI-Enabled Capability in Defence. (2022). GOV.UK. 2022. <https://www.gov.uk/government/publications/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence>.

Arkin, R. C., Ulam, P., & Wagner, A. R. (2012). Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception. *Proceedings of the IEEE*, 100(3), 571–589. <https://doi.org/10.1109/JPROC.2011.2173265>

Asaro, P. (2016). 'Hands up, Don't Shoot!': HRI and the automation of police use of force. *Journal of Human-Robot Interaction*, 5(3), 55–69. <https://doi.org/10.5898/JHRI.5.3.Asaro>

Bertolini, A., & Aiello, G. (2018). Robot companions: A legal and ethical analysis. *The Information Society*, 34(3), 130–140. <https://doi.org/10.1080/01972243.2018.1444249>

Bober, W. J. (2015). Czy Korzystanie z Bojowych Bezzalogowych Pojazdów Latających jest Moralnie Problematyczne? In *Systemy Dronów Bojowych. Analiza Problemów i Odpowiedź Społeczeństwa Obywatelskiego*, edited by Kaja Kowalczevska and Jacek Kowalewski. Wydawnictwo Naukowe "Scholar."

Bode, I., Huelss, H., Nadibaidze, A., Qiao-Franco, G., & Watts, T. F. A. (2023). Prospects for the global governance of autonomous weapons: Comparing Chinese, Russian, and US Practices. *Ethics and Information Technology*, 25(1), 5. <https://doi.org/10.1007/s10676-023-09678-x>

Boutin, B. (2023). State responsibility in relation to military applications of artificial intelligence. *Leiden Journal of International Law*, 36(1), 133–150. <https://doi.org/10.1017/S0922156522000607>

Bryson, J. J. (2018). Patience is not a virtue: The design of intelligent systems and systems of ethics. *Ethics and Information Technology*, 20(1), 15–26. <https://doi.org/10.1007/s10676-018-9448-6>

Cai, G., Li, G., & Dong, Y. (2012). Decision support systems and its application in military command. In *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)* (pp. 1744–1447). <https://doi.org/10.1109/CECNet.2012.6201730>.

CCW/GGE.1/2019/3. (2019). https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf.

CCW/GGE.1/2023/CRP.1. (2023). [https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_\(2023\)/CCW_GGE1_2023_CRP.1.pdf](https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.1.pdf).

Christen, M., Burri, T., Kandul, S., & Vörös, P. (2023). Who is controlling whom? Reframing 'Meaningful Human Control' of AI systems in security. *Ethics and Information Technology*, 25(1), 10. <https://doi.org/10.1007/s10676-023-09686-x>

Coeckelbergh, M. (2018). Why care about robots? Empathy, moral standing, and the language of suffering. *Kairos. Journal of Philosophy & Science*, 20(1), 141–158. <https://doi.org/10.2478/kjps-2018-0007>

Coeckelbergh, M. (2022). *Robot ethics*. The MIT Press.

Crawford, E. (2020). The temporal and geographic reach of International Humanitarian Law. In B. Saul & D. Akande (Eds.), *The Oxford guide to International Humanitarian Law*. OUP.

Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, 3(1), 5–24.

Danaher, J. (2019b). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics, June*. <https://doi.org/10.1007/s11948-019-00119-x>

Danaher, J. (2020). Welcoming robots into the moral circle: A defence of ethical behaviourism. *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-019-00119-x>

Danaher, J., Earp, B., & Sandberg, A. (2017). Should we campaign against sex robots? In J. Danaher & N. McArthur (Eds.), *Robot sex: Social and ethical implications*. New: The MIT Press.

Darling, K. (2021). *The new breed: What our history with animals reveals about our future with robots*. Henry Holt and Co.

Davison, N. (2018). A legal perspective: Autonomous weapon systems under International Humanitarian Law. In *UNODA Occasional Papers No. 30, November 2017: Perspectives on Lethal Autonomous Weapon Systems*. United Nations. <https://doi.org/10.18356/6fce2bae-en>.

Deeks, A. (2022). Coding the law of armed conflict: First steps. In M. C. Waxman & T. W. Oakley (Eds.), *The future law of armed conflict*. Oxford University Press.

Devlin, K. (2018). *Turned on: Science, sex and robots* (Illustrated). Bloomsbury Sigma.

Docherty, B. (2012). Losing humanity. *Human Rights Watch*. <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-st-killer-robots>.

Docherty, B. (2015). Mind the gap. *Human Rights Watch*. <https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots>.

Ethics Guidelines for Trustworthy AI | European Comission. (2019). Retrieved April 8, 2019, from <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

Farrant, J., & Ford, C. M. (2017). Autonomous weapons and weapon reviews: The UK Second International Weapon Review Forum. *International Law Studies*, 93(1), 13.

Feldbrugge, F. J. M. (1965). Good and bad samaritans. A comparative survey of criminal law provisions concerning failure to rescue. *The American Journal of Comparative Law*, 14(4), 630–657. <https://doi.org/10.2307/838914>

Fox Van, A. (2017). The deadly, incredible and absurd robots of the US Military. CNET. <https://www.cnet.com/pictures/deadly-incredible-absurd-robots-the-us-military/>.

Galliot, J., & Wyatt, A. (2020). Risks and benefits of autonomous weapon systems: Perceptions among future Australian Defence Force officers. *US Air Force Journal of Indo-Pacific Affairs*, 3(4), 17–34.

Garber, M. (2013). Funerals for fallen robots. The Atlantic. Retrieved September 20, 2013, from <https://www.theatlantic.com/technology/archive/2013/09/funerals-for-fallen-robots/279861/>.

Garvey, S. P. (2009). When should a mistake of fact excuse in general, should excuses be broadly or narrowly construed. *Texas Technical Law Review*, 42(2), 359–382.

Grzebyk, P. (Ed.) (2022). Human and non-human targets in armed conflicts. In *Human and non-human targets in armed conflicts* (pp. i–ii). Cambridge University Press. <https://www.cambridge.org/core/books/human-and-nonhuman-targets-in-armed-conflicts/human-and-nonhuman-targets-in-armed-conflicts/7D639078A88D93C8FAD5666DA90A490B>.

Gunkel, D. J. (2018a). The other question: Can and should robots have rights? *Ethics and Information Technology*, 20(2), 87–99. <https://doi.org/10.1007/s10676-017-9442-4>

Gunkel, D. J. (2018b). *Robot rights*. The MIT Press.

Guthrie, S. E. (1997). Anthropomorphism: A definition and a theory. In R. W. Mitchell, N. S. Thompson, & H. Lyn Miles (Eds.), *Anthropomorphism, anecdotes, and animals* (pp. 50–58). SUNY Series in Philosophy and Biology, State University of New York Press.

Harris, S. (2016). Autonomous weapons and international humanitarian law or killer robots are here: Get used to it autonomous legal reasoning: Legal and ethical issues in the technologies in conflict. *Temple International & Comparative Law Journal*, 30(1), 77–84.

Heyman, S. J. (1994). Foundations of the duty to rescue. *Vanderbilt Law Review*, 47(3), 673–756.

Heyns, C. (2010). Special rapporteur on extrajudicial, summary or arbitrary executions. https://idsn.org/wp-content/uploads/2015/02/SR_executions1.pdf.

Heyns, C., Akande, D., Hill-Cawthorne, L., & Chengeta, T. (2016). The international law framework regulating the use of armed drones. *The International and Comparative Law Quarterly*, 65(4), 791–827.

Jordan, J. M. (2016). *Robots* (1st ed.). The MIT Press.

Kayser, D. (2023). Why a treaty on autonomous weapons is necessary and feasible. *Ethics and Information Technology*, 25(2), 25. <https://doi.org/10.1007/s10676-023-09685-y>

Kirk, R. (2021). Zombies. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Spring 2021. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2021/entries/zombies/>.

Kowalczycka, K. (2019). The role of the ethical underpinnings of international humanitarian law in the age of lethal autonomous weapons systems. *Polish Political Science Yearbook*, 48(3), 464–475.

Kowalczycka, K. (2021). *Sztuczna Inteligencja Na Wojnie: Perspektywa Miedzynarodowego Prawa Humanitarnego Konfliktow Zbrojnych: Przypadek Autonomicznych Systemow Smiercionosnej Broni*. Wydanie pierwsze.

Kwik, J., & Van Engers, T. (2021). Algorithmic fog of war: When lack of transparency violates the law of armed conflict. *Journal of Future Robot Life*, 2(1–2), 43–66. <https://doi.org/10.3233/FRL-200019>

Lewis, D., Modirzadeh, N., & Blum, G. (2017). The Pentagon's new algorithmic-warfare team. Lawfare. Retrieved June 26, 2017, from <https://www.lawfareblog.com/pentagons-new-algorithmic-warfare-team>.

Lewis, L. (2018). Redefining human control: Lessons from the battlefield for autonomous weapons. <https://policycommons.net/artifacts/1542703/u-redefining-human-control/2232513/>.

Malinowska, J. K. (2021). Can i feel your pain? The biological and socio-cognitive factors shaping people's empathy with social robots. *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-021-00787-5>

Mamak, K. (2021). Whether to save a robot or a human: On the ethical and legal limits of protections for robots. *Frontiers in Robotics and AI*. <https://doi.org/10.3389/frobt.2021.712427>

Mamak, K. (2022). Should criminal law protect love relation with robots? *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01439-6>

Mamak, K. (2023). How should the law treat attacks on police robots? *Social Robots in Social Institutions*. <https://doi.org/10.3233/FAIA220662>

Mamak, Kamil. forthcoming. "Challenges of the Legal Protection of Human Lives in the Time of Anthropomorphic Robots." In *Cambridge Handbook on Law, Policy, and Regulations for Human-Robot Interaction*, edited by Woodrow Barfield, Yueh-Hsuan Weng, and Ugo Pagallo. Cambridge University Press.

Mamak, K. (Forthcoming). *Robotics, AI and criminal law: Crimes against robots*. Routledge.

Mamak, K., & Glanc, J. (2022). Problems with the prospective connected autonomous vehicles regulation: Finding A FAIR BALANCE VERSUS THE INSTINCT FOR SELF-Preservation. *Technology in Society*. <https://doi.org/10.1016/j.techsoc.2022.102127>

Marchant, G. E., Allenby, B., Arkin, R., & Barrett, E. T. (2011). International governance of autonomous military robots. *Columbia Science and Technology Law Review*, 12, 272–316.

McArthur, N., Danaher, J., Migotti, M., Wyatt, N., McArthur, N., Earp, B., Sandberg, A., et al. (2017). *Robot sex: Social and ethical implications*. Cambridge University Press.

McFarland, T., & Assaad, Z. (2023). Legal reviews of in situ learning in autonomous weapons. *Ethics and Information Technology*, 25(1), 9. <https://doi.org/10.1007/s10676-023-09688-9>

McINTYRE, A. (1994). Guilty bystanders? On the legitimacy of duty to rescue statutes. *Philosophy & Public Affairs*, 23(2), 157–191. <https://doi.org/10.1111/j.1088-4963.1994.tb00009.x>

McNab, M., & Matthews, M. (2010). Clarifying the law relating to unmanned drones and the use of force: The relationships between human rights, self-defense, armed conflict, and International Humanitarian Law Sutton Colloquium Articles. *Denver Journal of International Law and Policy*, 39(4), 661–694.

NATO. (n.d.) Summary of the NATO Artificial Intelligence Strategy. NATO. Retrieved April 17, 2023, from https://www.nato.int/cps/en/natoq/official_texts_187617.htm.

Nijssen, S. R. R., Müller, B. C. N., van Baaren, R. B., & Paulus, M. (2019). Saving the robot or the human? Robots who feel deserve moral care. *Social Cognition*, 37(1), 41–S2. <https://doi.org/10.1521/soco.2019.37.1.41>

Nyholm, S. (2018). The ethics of crashes with self-driving cars: A roadmap. I. *Philosophy Compass*, 13(7), e12507. <https://doi.org/10.1111/phc3.12507>

Nyholm, S., & Smids, J. (2020). Can a robot be a good colleague? *Science and Engineering Ethics*, 26(4), 2169–88. <https://doi.org/10.1007/s11948-019-00172-6>

O'Connell, M. E. (2009). Unlawful killing with combat drones: A case study of Pakistan, 2004–2009. SSRN Scholarly Paper. Rochester, NY. <https://papers.ssrn.com/abstract=1501144>.

Pardun, J. T. (1997). Good Samaritan Laws: A global perspective comment. *Loyola of Los Angeles International and Comparative Law Journal*, 20(3), 591–614.

REAIM 2023. (2023). Publicatie. Ministerie van Algemene Zaken. Retrieved February 16, 2023, from <https://www.government.nl/documents/publications/2023/02/16/reaim-2023-call-to-action>.

Recommendation on the Ethics of Artificial Intelligence - UNESCO. (2022). UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

Report of the AI Task Force September. (2019). Ministry of Defense. <https://www.defense.gouv.fr/sites/default/files/aid/Report%20of%20the%20AI%20Task%20Force%20September%202019.pdf>.

Richardson, K. (2015). The asymmetrical 'Relationship.' *AcM Sigcas Computers and Society*, 45(3), 290–93. <https://doi.org/10.1145/2874239.2874281>

Richardson, K. (2016). Are sex robots as bad as killing robots? *What Social Robots Can and Should Do*. <https://doi.org/10.3233/978-1-61499-708-5-27>

Robotics (Drones) Do Dull, Dirty, Dangerous & Now Difficult. (2018). Hangartech (blog). Retrieved February 28, 2018, from <https://medium.com/hangartech/robotics-drones-do-dull-dirty-dangerous-now-difficult-a860c9c182a4>.

der Pütten, R.-V., Astrid, M., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., & Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fMRI. *Computers in Human Behavior*, 33, 201–12. <https://doi.org/10.1016/j.chb.2014.01.004>

der Pütten, R.-V., Astrid, M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional

reactions towards a robot. *International Journal of Social Robotics*, 5(1), 17–34. <https://doi.org/10.1007/s12369-012-0173-8>

Sajduk, Błażej. 2015. "Problem Walki Na Odległość w Perspektywie Historycznej, Społecznej Etycznej." In *Systemy Dronów Bojowych. Analiza Problemów i Odpowiedź Społeczeństwa Obywatelskiego*, edited by Kaja Kowalczevska and Jacek Kowalewski. Wydawnictwo Naukowe "Scholar."

Salvini, P., Ciaravella, G., Yu, W., Ferri, G., Manzi, A., Mazzolai, B., Laschi, C., Oh, S. R., & Dario, P. (2010). How safe are service robots in urban environments? Bullying a Robot. In *19th International Symposium in Robot and Human Interactive Communication* (pp. 1–7). <https://doi.org/10.1109/ROMAN.2010.5654677>.

Sassoli, M. (2014). Autonomous weapons and international humanitarian law: Advantages, open technical questions and legal issues to be clarified. *International Law Studies*, 90(1), 1.

Scharre, P. (2016). Centaur warfighting: The false choice of humans vs. automation autonomous legal reasoning: Legal and ethical issues in the technologies in conflict. *Temple International & Comparative Law Journal*, 30(1), 151–66.

Schmitt, M. N., & Thurnher, J. S. (2012). Out of the loop: Autonomous weapon systems and the law of armed conflict. *Harvard National Security Journal*, 4(2), 231–81.

Schubert, J., Brynielsson, J., Nilsson, M., & Svenmarck, P. (2018). Artificial intelligence for decision support in command and control systems. In *Proceedings of the 23rd International Command and Control Research & Technology Symposium «Multi-Domain C2* (pp. 18–33).

Shaw, Martin. 2005. *The New Western Way of War: Risk-Transfer War and Its Crisis in Iraq*. Polity.

Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy*, 24(1), 62–77. <https://doi.org/10.1111/j.1468-5930.2007.00346.x>

Sparrow, R. (2021). How robots have politics. In C. Vélez (Ed.), *The Oxford handbook of digital ethics*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198857815.013.16>

Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16(2), 141–61. <https://doi.org/10.1007/s11023-006-9030-6>

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., & Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*, 5(1), 15924. <https://doi.org/10.1038/srep15924>

Sweijs, T., & De Spiegeleire, S. (2017). Artificial intelligence and the future of defense. *Artificial Intelligence and the Future of Defense*. <https://hess.nl/report/artificial-intelligence-and-the-future-of-defense/>.

Thrun, S. (2010). Toward robotic cars. *Communications of the ACM*, 53(4), 99–106. <https://doi.org/10.1145/1721654.1721679>

Thurnher, J. S. (2018). Feasible precautions in attack and autonomous weapons. In W. H. von Heinegg, R. Frau, & T. Singer (Eds.), *Dehumanization of warfare: Legal implications of new weapon technologies* (pp. 99–117). Springer.

U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Strategy. (2022). U.S. Department of Defense. <https://media.defense.gov/2022/Jun/22/2003022604-1/1-0/Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.PDF>.

Vélez, C. (2021). Moral Zombies: Why algorithms are not moral agents. *AI & SOCIETY*, April. <https://doi.org/10.1007/s00146-021-01189-x>

Vries, Barry de. 2023. *Individual Criminal Responsibility for Autonomous Weapons Systems in International Criminal Law*. Brill Nijhoff. <https://brill.com/display/title/63377>.

Winfield, A. (2012). *Robotics: A very short introduction*. Oxford: Oxford University Press.

Woodruff, O. E., Jr. (1958). Mistake of fact as a defense. *Dickinson Law Review*, 63(4), 319–34.

Zajac, M. (2023). AWS compliance with the ethical principle of proportionality: Three possible solutions. *Ethics and Information Technology*, 25(1), 13. <https://doi.org/10.1007/s10676-023-09689-8>

Zurek, T., Kwik, J., & van Engers, T. (2023). Model of a military autonomous device following International Humanitarian Law. *Ethics and Information Technology*, 25(1), 15. <https://doi.org/10.1007/s10676-023-09682-1>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.