A Case for 'Killer Robots': Why in the Long Run Martial AI May Be Good for Peace

Ognjen Arandjelović

North Haugh
University of St Andrews
St Andrews KY16 9SX
Fife, Scotland
United Kingdom

Tel: +44(0)1334 46 28 24
E-mail: ognjen.arandjelovic@gmail.com
ORCID ID: 0000-0002-9314-194X

The author has no funding sources or conflicts of interest to declare.

Abstract

Purpose: The remarkable increase of sophistication of artificial intelligence in recent years has already led to its widespread use in martial applications, the potential of so-called 'killer robots' ceasing to be a subject of fiction.

Approach: Virtually without exception, this potential has generated fear, as evidenced by a mounting number of academic articles calling for the ban on the development and deployment of lethal autonomous robots (LARs). In the present paper I start with an analysis of the existing ethical objections to LARs.

Findings: My analysis shows the contemporary thought to be deficient in philosophical rigour, these deficiencies leading to an alternative thesis.

Value: I advance a thesis that LARs can in fact be a force for peace, leading to fewer and less deadly wars.

1 Introduction

Artificial intelligence and machine learning in particular, that is computer based systems capable of learning from experience or supervision, have remarkably quickly found themselves integrated into our daily lives (Elliott, 2019). One could say that this rise of artificial intelligence has perhaps taken place somewhat by stealth, in that its increased use is taking place in a manner rather different from that depicted in the popular culture. Indeed, most of the society is unaware of the role that artificial intelligence already plays in a variety of mundane activities (Anderson and Smith, 2017). In contrast to these, there are numerous application domains of artificial intelligence with much more obvious potentially serious consequences. Not the least amongst these is warfare (Cummings, 2017). The employment of artificial intelligence in martial applications can hardly come as a surprise considering that military has for a long while been an ardent adopter of new technology, that it invests heavily in research collaborations with academia (Barker, 2017) (thereby steering the direction of academic research) and with the technology industry (Yoshida, 2016), and that it has massive research programmes of its own (Kania, 2019). The words from the USA Defense Science Board itself summarize this clearly:

> "The DoD dominates the world's military organizations in being able to use basic research results to create new and enhanced military capabilities, by dint of financial resources, infrastructure,

and national culture."

Equally unsurprising is the reaction of the portion of the society aware of the increasing use of artificial intelligence in war, not seldom led by voices from academia (Goose and Wareham, 2016). A particularly controversial issue is that of lethal autonomous robots (LARs) (Burri, 2018), often emotively referred to as "killer robots" (Young and Carpenter, 2018). Indeed, at the time of writing this, Google Scholar retrieves 4,450 articles matching the search query {"killer robots" "artificial intelligence"}. All but unanimously, these articles call for the cessation of the development of LARs (Sparrow, 2007; Sharkey, 2019; Gubrud, 2014; Sauer, 2016; Gibbs, 2017). In the present work I would like to offer a radically different view, a view that diverges substantially even from the small amount of published thought on the permissibility of 'killer robots', and argue that autonomous killing machines are not only permissible but rather potentially even *desirable* if the goal is that of world peace.

2 Arguments for and against LARs

As I have already noted, quite understandably the landscape of contemporary thought in the published academic work regarding the use of lethal autonomous robots is characterized by vehement opposition to the development and deployment of the technology. The views of Gubrud (2014) summarize the overwhelming attitude of the community well:

> "Opponents of autonomous weapons should point out the terrible threat they pose to global peace and security, as well as their offensiveness to principles of humanity and to public conscience."

In the present overview of these views I would like to approach the topic though the structure set up by one of the few dissenting voices, namely Burri (2018), for this will allow me at the same time to present a balanced picture of the mainstream as well as to differentiate my argument from Burri's itself, whose rejection of the mainstream is far weaker than mine. I also note that herein I do not delve into the related legal concerns, such as those raised by Krishnan (2016) and others, which although undoubtedly important, fall out of the sphere of ethics which is where the focus of my present article lies.

Burri (2018) delineates four main groups of objections to LARs, namely based on (i) non-codifiability, (ii) rightness of reasons for actions, (iii) responsibility, and (iv) heartlessness. I examine these in order next.

2.1 Non-codifiability of morality

The first group of objections to the use of LARs discussed by Burri (2018) is founded on the argument centering on the non-codifiability of moral decision-making (n.b. Burri uses the term 'anti-codifiability thesis' which I find less clear; hence my preference for *non*-codifiability) (Hooker, 2000; Roeser, 2012). What is meant by non-codifiability is, in its stronger form, that the decision-making process *cannot be* reduced to a set of rules, or, in its weaker form, that

the formulation of such a set of rules *is excessively complex to be considered practicable* (Kadar and Palatinus, 2022; Siegel and Pappas, 2021; Wallach et al., 2020).

The rebuttal offered by Burri (2018) does not focus on the ethical fundamentals of the aforementioned objections, but instead sidesteps them by effectively proposing a more constrained use of LARs, that is their deployment within limited bounds imposed by human actors' moral reasoning:

> "...LARs don't have to be morally sophisticated deliberators to almost exclusively inflict only permissible harm. It suffices, instead, that a conscientious human commanding officer deploys them only in contexts where they are able to identify sufficient conditions for the morally permissible infliction of lethal harm. For LARs to be usefully and permissibly employable, they don't have to be able to replace human soldiers across all possible circumstances, nor do they have to be able to strategize and reason about entire missions the way higher-ranking military personnel have to."

While I do not find fault with this rebuttal in that it does show the permissibility of LARs under *some* circumstances, as described by Burri (2018), I find it unnecessarily limiting and as such left wanting in strength.

Rather, the key realization that should lead us to reject the non-codifiability based arguments concerns the reasons for this non-codifiability; understanding these reveals the double standards hidden behind the surface. The

non-codifiability emerges not from some mystical aspect of human ethical decision-making which would make it inherently inexpressible as a set of rules but rather from something much more mundane: from the inconsistencies (and hence imperfection) in how *an individual* forms moral judgements (Krebs et al., 1997; Monin and Merritt, 2012), as well as from the differences between the processes and outcomes of moral judgements made by *different individuals* (Faulhaber et al., 2019). In other words, those who reject the use of LARs on the basis of the non-codifiability thesis, demand of machines more than they demand of humans (Grover, 2005). Why would this be? The reason can be but one: veiled in the cloth of the non-codifiability thesis is the true focus of the objectors which is the lack of an obvious moral agent that would bear the responsibility, and therefore suffer the punishment, when an objectionable action is performed, as demanded by the human nature for satisfaction (Carpenter, 2007; Orth, 2003).

A similar inconsistency in how intelligent agents are treated based on the aetiology of their intelligence (artificially created vs natural) can be found elsewhere too in Burri's work; I quote:

> "...correctly applying a moral principle to a specific situation can never be done purely mechanically; it always requires interpretation."

The reader will readily note the hidden presumption in the form of implication that interpretation is not 'mechanistic'. This is a blatant example of *petitio principii*; it is precisely what is meant by non-codifiability that

Burri (2018) is attempting to support by this sentence. And yet, what else could interpretation be but mechanistic? Does our brain not obey the laws of physics, just as a boiling kettle of water or an apple falling off a tree do (Schopenhauer, 2009)?

2.2 Acting for the right reasons

Unlike the previous one, the next group of arguments against the use of LARs discussed by Burri (2018) is distinctly not consequentialist in nature. Quoting Purves et al. (2015), Burri (2018) summarizes the gist behind this group of objections as lying in the moral insufficiency of ethical decisions which are "perfect" but were made without "the right reason" and, the argument goes, robots cannot act for any reason whatsoever because "an attitude of belief or desire (or some further propositional attitude) is a conceptual prerequisite of acting for a reason" while "something which runs on algorithms cannot possess such an attitude".

Burri (2018) starts her rebuttal by quite correctly pointing out that the proponents of this group of objections never actually explain *why* the absence of "the right reason" (more on this soon) of an agent which always makes morally agreeable decisions matters, and, admirably, tries her best to reconstruct plausible explanations herself. She firstly and quite correctly rejects a possible analogy of a morally perfect LAR with that of a sociopathic soldier which always obeys orders and never receiving a morally objectionable order thus always acts in a morally right manner, by recognizing the flaw in this

comparison which stems from the the obvious imperfection of the sociopathic soldier (contrasting the *ex hypothesi* perfection of the LAR) which is merely constrained by a different, morally righteous agent (the superior officer). She next addresses the rather nebulous thesis based on "a lack of respect" (closely related to objections on the grounds of human dignity (Sharkey, 2019)) of a justifiable killing by an unreasoned robot. While agreeing with Purves et al. (2015) that this would be a valid objection for an agent capable of reasoning, Burri (2018) nevertheless rejects it as invalid in cases of agents which, again by the very premise of the objectors, do not have the capacity of reason in the first place. I protest against this rebuttal on several grounds, some of which I shall return to shortly; for now, it suffices to say that the very conception of respect in this circumstance is ill-conceived. As I have demonstrated in my previous work (Arandjelović, 2022, 2023), the entire notion of respect for life is nothing but an uncomfortable anachronistic remnant of theological morals left floating in the air without anything to support it now that its theological foundations have been stripped away. Burri (2018) next considers what she, rather strangely, describes as a "Kantian idea", that:

> "...actions that are performed for the right reasons are accorded a special moral status — unlike other actions, they have moral worth — because the will behind them is of unconditional moral value.

Her rejection of this argument is effectively identical to that of the previous one, arguing for its inapplicability to agents which are not capable of having

reasons. Yet again, I find her rebuttal wanting. Firstly, I find it rather bizarre to describe the original objection as being Kantian. Kant's moral imperative, as wonderfully lucidly and convincingly deconstructed by Schopenhauer (2009), is not only not a law and void of any particular prescription for action in the real world, but also utterly lacks the key elements which make an action morally worthy, namely compassion and love. Kant's attempt at reducing morality to mere reason, void of any sympathy which would give its impetus, any possible impetus behind it ever emanating from selfishness and egoism, is an absolute antithesis of morality, elevated by Kant's successors, and Fichte in particular, to grotesque heights. Related is the claim of the "unconditional moral value", the phrase whose meaninglessness is obscured by its superficial appeal and strength, aimed at instilling awe and fear in the reader, lest it be challenged. The claim of an unconditional moral value, or indeed unconditional value of anything at all, is nonsensical, a *contradictio in adjecto*, for the very meaning of the word "value" is comparative in nature and thus conditional. That something has value inherently implies a hypothesised fair *exchange*. When it is talked about the worth of a house, it is understood that the worth is hypostatized by the mutual willingness of its owner and its potential owner to make an exchange of the house for a certain sum of (usually) money. That 'a bird in the hand is **worth** two in the bush' means that my exchanging a bird I have in the hand for two that are in the bush leaves me no better or worse off. In short, the "Kantian objection" is vacuous, a casuistic slight of hand, not worthy of a serious consideration.

For completeness, I find it worthwhile to make two additional points as regards the correctness of reasons objection. The first of these is the implicit suggestion that human soldiers in general act for the right reasons. While this may be so if the rightness is interpreted as meaning 'conforming with the law' (the *jus in bello* rules), the righteousness of interest here is based on an appeal to emotion. Do soldiers really engage in lethal combat for the right reasons? How many soldiers truly understand the morality of the reasons for them being placed in combat situations in the first place? Few, evidence would suggest (McMahan, 2008; Finlay, 2019). Rather, I would contend that in practice a *professional* soldier seldom makes a decision to kill for a right reason, any appearance of righteousness being merely incidental. The reason is to be found in the professional soldier's surrender of autonomy over such a monumental choice as is that to engage in a war, to a structure that has repeatedly been shown to be but a poor moral actor. The only partial defence of this surrender — hence the restraint in my position and wording — can be sought in individuals' lack of knowledge and full appreciation of the said choice (Arandjelović, 2021).

Another challenge which Burri (2018) fails to make is to the claim that robots cannot act with a *reason*. With no justification at all, with little more than a wave of the hand, the proponents of the objection summarily dismiss the tenant *operari sequitur esse*. Their lack of sophistication in understanding the crucial underpinnings of modern artificial intelligence and its conception is reflected with lucidity by their choice of words "something which runs on

algorithms" (in full: "something which runs on algorithms cannot possess such an attitude"). There is indeed no basis to reject the ability of machines to act with a reason, 'reason' merely being a word that we use to denote a representation of knowledge that acts as an impetus for an acting agent. Whether that representation be in the form of synaptic connections between biological neurons or, say, weights of connections between artificial neurons, is a matter of irrelevance.

2.3 Responsibility

The third and rather eminent group of anti-LARs arguments discussed by Burri (2018) revolves around the notion of responsibility (Hellström, 2013; Lokhorst and Van Den Hoven, 2012; Nyholm, 2018; Bigman et al., 2019) and in particular:

> "[the] risk that they [LARs] will inflict wrongful harm for which no one is morally responsible."

This is a widely supported objection. For example, Sparrow (2007) writes:

> "I argue that in fact none of these [loci of responsibility] are ultimately satisfactory. Yet it is a necessary condition for fighting a just war, under the principle of *jus in bellum* (sic), that someone can be justly held responsible for deaths that occur in the course of the war. As this condition cannot be met in relation to deaths

> caused by an autonomous weapon system it would therefore be unethical to deploy such systems in warfare."

whereas Gubrud (2014) raises the concerns around responsibility alongside the already discussed issue of 'human dignity':

> "However, demands for human control and responsibility and the protection of human dignity and sovereignty fit naturally into the traditional law of war and imply strict limits on autonomy in weapon systems."

Well-advisedly, Burri (2018) approaches the challenge by considering the (α) possibility of a human agent (or agents) being held responsible for wrongful harm inflicted by a LAR, and (β) the possibility of responsibility lying with the LAR itself (in which case the pronoun "themself" would probably be more appropriate). In considering the former, Burri (2018) correctly points out that the proponents of the argument seldom elucidate with any precision as to why they reject the possibility and, generously and quite reasonably, makes the best attempt at surmising the possible thinking behind it:

> "...a human agent is not morally responsible for harm inflicted by an LAR when the harm was not, in some meaningful sense, under the human agent's control...and the machine behaved in a way that was not foreseeable."

Burri (2018) counters this with an analogy of a programmer, say, who "decides to hide the fact that the software comes with crucial unpredictabilities",

concluding that

> "...the moral responsibility for any unforeseeable wrongful harm that an LAR running on the software might cause remains with him or her. His or her actions are not only negligent but downright reckless: he or she is pretending that it is relatively safe to use an incredibly dangerous tool."

This is a rather poor challenge, bordering on sophistic. If the programmer in question is *hiding* the knowledge about a robot's unpredictability, then this violates any reasonable interpretation of the premise of the argument which is that the robot's behaviour was not foreseeable. The behaviour in this instance can only be described as being unforeseeable from the subjective viewpoint of, say, a military operative who engages the LAR and from whom vital knowledge about its behaviour was withheld (who, consequently, indeed cannot be held responsible), or by virtue of semantic dishonesty and casuistry, the *precise* sequence of actions performed by the LAR was unforeseeable. The latter is as convincing as claiming that my firing a gun into somebody's head has unpredictable consequences because one cannot be certain as to what precise areas of the brain will get damaged. The consequences are foreseeable in the contextually relevant sense.

Burri's rejection of the impossibility presumed by the proponents of the responsibility based objection to LARs, of holding the robot itself accountable is equally unconvincing. Focusing on the reasons behind the claim offered by (Sparrow, 2007), whose view is representative, and which stands on

the premise that a robot cannot be held accountable for wrongdoing because it cannot be meaningfully punished as it cannot suffer, Burri (2018) offers two counterarguments. Firstly, she dismisses the implied obviousness of the claim that robots cannot suffer:

> "For one thing, I am not convinced that the type of LAR that Sparrow envisages would necessarily be incapable of suffering. Once LARs have goals and desires of their own, why wouldn't they suffer if they had these thwarted?"

In addition to the appeal to the intuition rejecting the possibility of sentience of robots in the form in which they exist at present (Picard, 2003; Velik, 2010; Turkle, 2017; Feil-Seifer and Matarić, 2011; Arandjelović, 2021), if Burri (2018) truly believed that LARs are capable of suffering, I would find it odd, to say the least, that she would not be far more concerned about the creation of this artificial sentience, the effects of our design choices on these sentient (but non-biological) beings, etc. It is difficult to take this belief as being genuine and hence I consider it unworthy of further consideration.

Burri's second counterargument is rather different in spirit; to summarize it succinctly in her words, it rests on the observation that:

> "...our practices of holding wrongdoers accountable for their actions are not limited to making them suffer."

While this claim is true, it too is a superficial linguistic veil covering an evasion of the crux of the matter. Firstly, the alternative or additional practices

of holding wrongdoers accountable (e.g. through the use of apology and the expression of remorse (Bibas and Bierschbach, 2004)) also rest on sentience, requiring it for the hypostatization of accountability as a meaningful concept. Secondly, Burri (2018) ignores the importance and the value of, and indeed the need for retributive justice which emanates from the very nature of the human mind and which can have positive effects on victims (McClelland, 2010; Seton, 2001; Zaibert, 2006).

2.4 Heartlessness

Lastly, Burri (2018) turns her attention to the objections to the use of LARs premised on the claim of heartlessness inherent in the killing of humans by non-sentient agents. The argument is summarized well by O'Connell (2014):

> "[g]iving up the decision [to kill] entirely to a computer program will...remove, literally, the humanity that should come to bear in all cases of justifiable killing."

and, to offer an alternative phrasing, by Ekelhof and Struyk (2014):

> "War is about human suffering, the loss of human lives, and consequences for human beings. Killing with machines is the ultimate demoralization of war. Even in the hell of war we find humanity, and that must remain so."

Interestingly, Burri (2018) largely agrees with the spirit of this thinking, stating that:

> "...in cases where the risk of harm to a just combatant is very small, the morally best killing of an unjust enemy combatant takes place when a just combatant feels the weight of the decision and finally kills the enemy combatant with empathy and for the right reasons.",

rejecting merely the conclusions drawn by their proponents, by arguing that even if the risk to the killing combatants is small, it is still reasonable to eliminate this risk in its entirety if possible, as would indeed be done through the use of robots. Yet, even if the vacuous shibboleth 'humanity', a mere appeal to emotion discussed before, is put aside, there is so much to be objected to. Firstly, let us remember that *ex hypothesi*, we are comparing materially the same decisions and actions of a human agent and a non-human, automatic one. With this in mind, asking a sentient being, capable of suffering, reflection, and remorse to undertake a task which we know is traumatic and with long-lasting psychological consequences on the individual (MacNair, 2007; Maguen et al., 2017; Purcell et al., 2018; Pitts et al., 2013), is surely precisely an argument in favour of the opposite of what Burri (2018) agrees with, that is the killing by a LAR should be seen as not merely morally justifiable but rather morally *preferable* for the reasons of compassion. This is precisely why the administration of capital punishment in those Western societies in which it is still practised, is realized by means which divorce the executioner as much as possible from the executed and the proximally lethal act itself (Seal, 2016; Ebury, 2021; Osofsky et al., 2005).

2.5 Burri's argument *for* LARs

Having rejected the popular arguments against the development and the use of LARs which I contend she did with varying degrees of success as we have seen in the preceding sections, Burri (2018) finally lays out her positive challenge. In other words, she puts forward her reasons for favouring LARs in the battlefield. Her argument is fairly brief and it boils down to the following points:

> "Simply put, if we are able to develop LARs that can replace human soldiers in the theater of war, taking a wide perspective on the principle of necessity implies that we should do so as it helps us minimize the extent to which we have to put our soldiers at risk of harm when pursuing just goals."

and:

> "It follows that if LARs have the potential to help us shield our soldiers from emotional and mental harm, then this provides us with a valid reason in favor of developing autonomous weapons technology further."

While I broadly agree with both of these, though it should be noted that I have already highlighted how some of Burri's views do not cohere with the intent expressed here, I find them insufficiently strong. Hence, I put forward a stronger argument, one absent from the published academic literature, next.

2.6 My challenge

Hitherto, my focus has been on the most supported objections to the use of LARs. My rejection of these has thus far been what one may described as proximal: proximal in the sense that I have in my analysis and critique thereof, for the sake of argument and with the aim of providing as comprehensive rebuttal as possible, (temporarily) accepted a particular well-hidden premise underlying them. Yet, this premise is key to the most practically important distal realization in the context of the present discussion. I am referring to the assumption made by all of the groups of objections discussed, namely that the LARs would *actually be killing*. This may sound odd, I understand. After all, is not the very *purpose* of *killer* robots to do exactly as their name suggests, that is, to kill? Not necessarily, I say. Let me explain.

Consider a time when sufficiently sophisticated killer robots can be built. It is all but inconceivable to imagine only a single state actor having access to this capability (Mori, 2019; Lukin, 2021; Johnson, 2021). Firstly, much of the requisite technology needed in LARs is built upon openly accessible research (and there is a significant drive to maintain this research as widely accessible as possible (Vicente-Saez and Martinez-Fuentes, 2018)), whether that research be coming out of academia or industry, especially as most of it is conducted with a view of its use in much more mundane, everyday applications, its martial employment being but a consequence of translational opportunism (Edgerton, 1988). Specifically military oriented work in academia, often in collaboration with and funded by the military

and weapons manufacturers, is also abound with an ever-increasing amount of work on computer vision based military target detection (Eismann et al., 1996; Tiwari et al., 2011; Wang et al., 2018), target classification (Thiagarajan et al., 2010; Lampropoulos et al., 2008), vehicle tracking from aerial views (Ma'Sum et al., 2013; Arandjelović, 2015), and many other relevant problems (Gonzalez-Aguilera and Rodriguez-Gonzalvez, 2017; Akbari et al., 2021). Some technical information on commercial LARs is also in the public domain, such as the Boston Dynamics LS3 (Michael, 2012) or the Vision60 Q-UGV (Ghost Robotics, 2021).

Secondly, espionage between nation states, aided both by benevolent (which does not necessarily mean well-advised) and malevolent actors, is rife (Rubenstein, 2014; Lindsay, 2017; Banks, 2016), leaving few secrets between powerful parties. Hence, a major state with access to LARs can very much count on other dominant powers having a comparable LAR technology (Mori, 2019; Cheung et al., 2017; Johnson, 2021). Any military confrontation between two or more such states would therefore not involve human soldiers at all. As Burri (2018) quite correctly pointed out, while failing to take her reasoning to its logical conclusion, why would either state risk its own people when sophisticated but non-sentient machines would do? And yet, what would a confrontation like that, between two armies of LARs, achieve? Very little, if anything at all.

At the same time, it is an equally difficult possibility to imagine that *all* nation states would have LARs, at least for some time to come. At first

sight this seems like a rather perilous situation. However, it is precisely in the obvious asymmetry of strength (and the virtually *symmetric* understanding thereof) wherein the incentive against a potential war lies; the less powerful actor would be nothing short of insane in engaging in a war with the odds so obviously set against it (Renic, 2020; Grafen, 1987). While this does not mean that the result would be acceptable, in that the hypothetical powerful state would in principle be able to take over another with no hindrance of force, it is, most importantly, clear that lives which would have otherwise been lost on the battlefield would be saved. An unresisted, at least by means of arms, occupation is certainly undesirable, but were war to be waged the same end result would ensue, but with the additional cost to human life preceding it. Moreover, while this is not my main point here, it is also worth adding that the unresisted takeover scenario does not seem particularly likely as a general rule: the deterrent in the form of international reputation is not to be forgotten lightly (Tang, 2005; Guzman, 2005; Brewster, 2009; Downs and Jones, 2002).

3 Conclusions

The recent rapid advancements of artificial intelligence and its increasing use in martial applications has made the possibility of manufacture of lethal autonomous robots (LARs) a part of reality. This possibility of their use in actual warfare has largely been met with understandable fear. Indeed,

numerous academic articles and books have already been published on the topic, outlining a variety of associated concerns, and all but unanimously calling for a ban on such machines. I started this article by first discussing the most popular objections to the use of LARs, approaching the task through the lens of one of the few dissenting voices, showing deficiencies in both sides' arguments. Hence, with a view on the fundamental error shared by these, which was previously unrecognized in the published academic literature and elsewhere, namely that the potential ubiquity of LARs changes both the nature of warfare and the decisions to engage in the same, I explained why this would likely result in fewer wars and less lethal wars.

References

Akbari, Y., Almaadeed, N., Al-Maadeed, S., and Elharrouss, O. (2021). Applications, databases and open computer vision research from drone videos and images: a survey. *Artificial Intelligence Review*, 54:3887–3938.

Anderson, M. and Smith, A. (2017). Automation in everyday life. *Pew Research Center*.

Arandjelović, O. (2015). Automatic vehicle tracking and recognition from aerial image sequences. In *International Conference on Advanced Video and Signal Based Surveillance*, pages 1–6. IEEE.

Arandjelović, O. (2021). AI, democracy, and the importance of asking the right questions. *AI & Ethics Journal*.

Arandjelović, O. (2022). On the value of life. *International Journal of Applied Philosophy*, 35(2):227–241.

Arandjelović, O. (2023). On the subjective value of life. *Philosophies*, 8(2):23.

Banks, W. C. (2016). Cyber espionage and electronic surveillance: Beyond the media coverage. *Emory Law Journal*, 66:513.

Barker, K. (2017). The quiet military buyout of academia. *Preventing War and Promoting Peace: A Guide for Health Professionals*, page 141.

Bibas, S. and Bierschbach, R. A. (2004). Integrating remorse and apology into criminal procedure. *Yale Law Journal*, 114:85.

Bigman, Y. E., Waytz, A., Alterovitz, R., and Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in Cognitive Sciences*, 23(5):365–368.

Brewster, R. (2009). Unpacking the state's reputation. *Harvard International Law Journal*, 50:231.

Burri, S. (2018). What is the moral problem with killer robots. *Who Should Die: The Ethics of Killing in War*, pages 163–185.

Carpenter, J. P. (2007). The demand for punishment. *Journal of Economic Behavior & Organization*, 62(4):522–542.

Cheung, T. M., Anderson, E., and Yang, F. (2017). Chinese defense industry reforms and their implications for us-china military technological competition. *SITC Research Briefs*, (2017-4).

Cummings, M. (2017). *Artificial intelligence and the future of warfare*. Chatham House for the Royal Institute of International Affairs London.

Downs, G. W. and Jones, M. A. (2002). Reputation, compliance, and international law. *The Journal of Legal Studies*, 31(S1):S95–S114.

Ebury, K. (2021). Justice and punishment in executioners' life-writing. In *Modern Literature and the Death Penalty, 1890-1950*, pages 89–116. Springer.

Edgerton, D. (1988). The relationship between military and civil technology: a historical perspective. In *The Relations between Defence and Civil Technologies*, pages 106–114. Springer.

Eismann, M. T., Schwartz, C. R., Cederquist, J. N., Hackwell, J. A., and Huppi, R. J. (1996). Comparison of infrared imaging hyperspectral sensors for military target detection applications. In *Imaging Spectrometry II*, volume 2819, pages 91–101. SPIE.

Ekelhof, M. and Struyk, M. (2014). Deadly decisions: 8 objections to killer robots. *PAX*. `https://paxforpeace.nl/news/overview/stop-killer-robots-while-we-still-can`.

480 Elliott, A. (2019). *The culture of AI: Everyday life and the digital revolution.*
481 Routledge.

482 Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld,
483 L. R., Stephan, A., Pipa, G., and König, P. (2019). Human decisions in
484 moral dilemmas are largely described by utilitarianism: Virtual car driving
485 study provides guidelines for autonomous driving vehicles. *Science and*
486 *Engineering Ethics*, 25(2):399–418.

487 Feil-Seifer, D. and Matarić, M. J. (2011). Socially assistive robotics. *IEEE*
488 *Robotics & Automation Magazine*, 18(1):24–31.

489 Finlay, C. J. (2019). Justification and legitimacy at war: on the sources of
490 moral guidance for soldiers. *Ethics*, 129(4):576–602.

491 Ghost Robotics (2021). Vision60 Q-UGV. https://www. ghostrobotics.io.

492 Gibbs, S. (2017). Elon musk leads 116 experts calling for outright ban of
493 killer robots. *The Guardian*, 20:2017.

494 Gonzalez-Aguilera, D. and Rodriguez-Gonzalvez, P. (2017). Drones—an open
495 access journal. *Drones*, 1(1):1.

496 Goose, S. D. and Wareham, M. (2016). The growing international movement
497 against killer robots. *Harvard International Review*, 37(4):28–34.

498 Grafen, A. (1987). The logic of divisively asymmetric contests: respect for
499 ownership and the desperado effect. *Animal Behaviour*, 35(2):462–467.

Grover, S. L. (2005). The truth, the whole truth, and nothing but the truth: The causes and management of workplace lying. *Academy of Management Perspectives*, 19(2):148–157.

Gubrud, M. (2014). Stopping killer robots. *Bulletin of the Atomic Scientists*, 70(1):32–42.

Guzman, A. T. (2005). Reputation and international law. *Georgia Journal of International and Comparative Law*, 34:379.

Hellström, T. (2013). On the moral responsibility of military robots. *Ethics and Information Technology*, 15(2):99–107.

Hooker, B. (2000). Moral particularism: wrong and bad. In Hooker, B. and Little, M. O., editors, *Moral Particularism*, pages 1–22. Oxford, UK: Oxford University Press.

Johnson, J. (2021). The end of military-techno Pax Americana? Washington's strategic responses to Chinese AI-enabled military technology. *The Pacific Review*, 34(3):351–378.

Kadar, E. E. and Palatinus, Z. (2022). Reinventing Kantian autonomy for artificial agents: Implications for self-driving cars. In *Towards Trustworthy Artificial Intelligent Systems*, pages 169–177. Springer.

Kania, E. B. (2019). Chinese military innovation in the AI revolution. *The RUSI Journal*, 164(5-6):26–34.

Krebs, D. L., Denton, K., and Wark, G. (1997). The forms and functions of real-life moral decision-making. *Journal of Moral Education*, 26(2):131–145.

Krishnan, A. (2016). *Killer robots: legality and ethicality of autonomous weapons*. Routledge.

Lampropoulos, G. A., Liu, T., Qian, S.-E., and Fei, C. (2008). Hyperspectral classification fusion for classifying different military targets. In *IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages 262–265. IEEE.

Lindsay, J. R. (2017). Cyber espionage. *The Oxford Handbook of Cyber Security*.

Lokhorst, G.-J. and Van Den Hoven, J. (2012). *Responsibility for Military Robots*, pages 145–156. MIT Press Cambridge, MA.

Lukin, A. (2021). The Russia–China entente and its future. *International Politics*, 58(3):363–380.

MacNair, R. M. (2007). Killing as trauma. *Trauma Psychology: Issues in Violence, Disaster, Health, and Illness*, 1:147–162.

Maguen, S., Burkman, K., Madden, E., Dinh, J., Bosch, J., Keyser, J., Schmitz, M., and Neylan, T. C. (2017). Impact of killing in war: A randomized, controlled pilot trial. *Journal of Clinical Psychology*, 73(9):997–1012.

Ma'Sum, M. A., Arrofi, M. K., Jati, G., Arifin, F., Kurniawan, M. N., Mursanto, P., and Jatmiko, W. (2013). Simulation of intelligent unmanned aerial vehicle (uav) for military surveillance. In *International Conference on Advanced Computer Science and Information Systems*, pages 161–166. IEEE.

McClelland, R. T. (2010). The pleasures of revenge. *The Journal of Mind and Behavior*, pages 195–235.

McMahan, J. (2008). The morality of war and the law of war. In Rodin, D. and Shue, H., editors, *Just and Unjust Warriors: The Moral and Legal Status of Soldiers*, pages 19–43. Oxford University Press.

Michael, K. (2012). Meet boston dynamics' ls3-the latest robotic war machine. *Faculty of Engineering and Information Sciences — Papers: Part A*. `https://ro.uow.edu.au/eispapers/2773`.

Monin, B. and Merritt, A. (2012). *Moral hypocrisy, moral inconsistency, and the struggle for moral integrity*, pages 167–184. American Psychological Association.

Mori, S. (2019). Us technological competition with china: The military, industrial and digital network dimensions. *Asia-Pacific Review*, 26(1):77–120.

Nyholm, S. (2018). Attributing agency to automated systems: Reflections

on human–robot collaborations and responsibility-loci. *Science and Engineering Ethics*, 24(4):1201–1219.

O'Connell, M. E. (2014). Banning autonomous killing: The legal and ethical requirement that humans make near-time lethal decisions. In *The American Way of Bombing: Changing Ethical and Legal Norms from Flying Fortresses to Drones*, pages 224–235. Cornell University Press.

Orth, U. (2003). Punishment goals of crime victims. *Law and Human Behavior*, 27(2):173–186.

Osofsky, M. J., Bandura, A., and Zimbardo, P. G. (2005). The role of moral disengagement in the execution process. *Law and Human Behavior*, 29(4):371–393.

Picard, R. W. (2003). What does it mean for a computer to "have" emotions. *Emotions in Humans and Artifacts*, pages 213–235.

Pitts, B. L., Chapman, P., Safer, M. A., Unwin, B., Figley, C., and Russell, D. W. (2013). Killing versus witnessing trauma: Implications for the development of PTSD in combat medics. *Military Psychology*, 25(6):537–544.

Purcell, N., Burkman, K., Keyser, J., Fucella, P., and Maguen, S. (2018). Healing from moral injury: A qualitative evaluation of the impact of killing treatment for combat veterans. *Journal of Aggression, Maltreatment & Trauma*, 27(6):645–673.

Purves, D., Jenkins, R., and Strawser, B. J. (2015). Autonomous machines, moral judgment, and acting for the right reasons. *Ethical Theory and Moral Practice*, 18(4):851–872.

Renic, N. C. (2020). *Asymmetric killing: risk avoidance, just war, and the warrior ethos*. Oxford University Press.

Roeser, S. (2012). Emotional engineers: Toward morally responsible design. *Science and Engineering Ethics*, 18(1):103–115.

Rubenstein, D. (2014). Nation state cyber espionage and its impacts. *Deptartment of Computer Science and Engineering WUSTL, Saint Louis*.

Sauer, F. (2016). Stopping 'killer robots': Why now is the time to ban autonomous weapons systems. *Arms Control Today*, 46(8):8–13.

Schopenhauer, A. (2009). *The two fundamental problems of ethics*. Cambridge University Press.

Seal, L. (2016). Albert pierrepoint and the cultural persona of the twentieth-century hangman. *Crime, Media, Culture*, 12(1):83–100.

Seton, P. H. (2001). On the importance of getting even: A study of the origins and intention of revenge. *Smith College Studies in Social Work*, 72(1):77–97.

Sharkey, A. (2019). Autonomous weapons systems, killer robots and human dignity. *Ethics and Information Technology*, 21(2):75–87.

Siegel, J. and Pappas, G. (2021). Morals, ethics, and the technology capabilities and limitations of automated and self-driving vehicles. *AI & Society*, pages 1–14.

Sparrow, R. (2007). Killer robots. *Journal of applied philosophy*, 24(1):62–77.

Tang, S. (2005). Reputation, cult of reputation, and international conflict. *Security Studies*, 14(1):34–62.

Thiagarajan, J. J., Ramamurthy, K. N., Knee, P., Spanias, A., and Berisha, V. (2010). Sparse representations for automatic target classification in sar images. In *International Symposium on Communications, Control and Signal Processing*, pages 1–4. IEEE.

Tiwari, K. C., Arora, M. K., and Singh, D. (2011). An assessment of independent component analysis for detection of military targets from hyperspectral images. *International Journal of Applied Earth Observation and Geoinformation*, 13(5):730–740.

Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. Hachette UK.

Velik, R. (2010). Why machines cannot feel. *Minds and Machines*, 20(1):1–18.

Vicente-Saez, R. and Martinez-Fuentes, C. (2018). Open science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88:428–436.

Wallach, W., Allen, C., and Smit, I. (2020). Machine morality: bottom-up and top-down approaches for modelling human moral faculties. In *Machine Ethics and Robot Ethics*, pages 249–266. Routledge.

Wang, X., Cheng, P., Liu, X., and Uzochukwu, B. (2018). Fast and accurate, convolutional neural network based approach for object detection from UAV. In *Annual Conference of the IEEE Industrial Electronics Society*, pages 3171–3175. IEEE.

Yoshida, K. (2016). Tripartite collaboration among industry, academia and government-changing scenarios in the era of megacompetition. *Journal of International Association of P2M*, 11(1):1–10.

Young, K. L. and Carpenter, C. (2018). Does science fiction affect political fact? yes and no: A survey experiment on "killer robots". *International Studies Quarterly*, 62(3):562–576.

Zaibert, L. (2006). Punishment and revenge. *Law and Philosophy*, 25(1):81–118.