



Debate on the Ethics of Developing AI for Lethal Autonomous Weapons

Jai Galliot,¹ John Forge²

¹ Values in Defence & Security Technology Group, University of New South Wales, Australia

² School of History and Philosophy of Science, Sydney University, Australia

Abstract

In this philosophical debate on the ethics of developing AI for Lethal Autonomous Weapons, Jai Galliot argues that a “blanket prohibition on ‘AI in weapons,’ or participation in the design and engineering of artificially intelligent weapons, would have unintended consequences due to its lack of nuance.” In contrast to Galliot, John Forge contends that “the only course of action for a moral person is not to engage in weapons research.”

Keywords

Artificial Intelligence; Ethics; Lethal Autonomous Weapons; War; Weapon Design.

DOI: 10.22618/TP.PJCIV.20215.1.139009

The PJCIV Journal is published by Trivent Publishing



Debate on the Ethics of Developing AI for Lethal Autonomous Weapons

Jai Galliot,¹ John Forge²

¹ Values in Defence & Security Technology Group, University of New South Wales, Australia

² School of History and Philosophy of Science, Sydney University, Australia

Abstract

In this philosophical debate on the ethics of developing AI for Lethal Autonomous Weapons, Jai Galliot argues that a “blanket prohibition on ‘AI in weapons,’ or participation in the design and engineering of artificially intelligent weapons, would have unintended consequences due to its lack of nuance.” In contrast to Galliot, John Forge contends that “the only course of action for a moral person is not to engage in weapons research.”

Keywords

Artificial Intelligence; Ethics; Lethal Autonomous Weapons; War; Weapon Design.

Some Thoughts Concerning the Ethics of Developing AI for LAWS

In April 2018, hundreds of United Nations member nations agreed to continue diplomatic talks that started in 2014 to consider concerns raised over Lethal Autonomous Weapons (LAWS), defined by the United States as systems that, once activated, can select and engage targets without further intervention by a human operator, and which are known in hyperbolic terms as ‘killer robots.’ Every UN meeting since that commencing the international discussion on these weapons has ended in disappointment for all parties, especially advocates hoping that the world would make progress on regulating or banning ‘killer robot’ technologies, because the UN group of governmental experts has barely even scratched the surface of the potential options for regulating lethal autonomous weapons.

The dialogue on autonomous weapon systems initially proceeded quite cautiously on the part of the states with responsibility for steering the discussion, on the basis that few understood what it was some were seeking to outlaw with a pre-emptive ban. Seemingly wise at the time, it is my view that this uncertainty allowed a small number of advocate groups to sway the debate in the vacuum of informed opinion, giving rise to a debate that has ever since been very heavily one-sided. Some contend, on legal and moral grounds, that military operations should be immune from the progress of automation and artificial intelligence evident in other areas of society. There are those who deplore all weapons research and believe that the expertise and resources devoted to the development of new ways to kill would be much better spent in devising new ways to help rather than harm people and are thus opposed to LAWS simply because they are new means of inflicting harm. But there have also been appeals to ban the research and development of LAWS on more nuanced grounds.

The Campaign to Stop Killer Robots, operated by a consortium of non-government interest groups, deploys the full range of possible arguments against LAWs, with over 1,000 in artificial intelligence, as well as science and technology luminaries such as Stephen Hawking, Elon Musk, Steve Wozniak, Noam Chomsky, Skype co-founder Jaan Tallinn and Google DeepMind co-founder Demis Hassabis, expressing several problems on their website:

Allowing life or death decisions to be made by machines crosses a fundamental moral line. Autonomous robots would lack human judgment and the ability to understand context. These qualities are necessary to make complex ethical choices on a dynamic battlefield, to distinguish adequately between soldiers and civilians, and to evaluate the proportionality of an attack. As a result, fully autonomous weapons would not meet the requirements of the laws of war. Replacing human troops with machines could make the decision to go to war easier, which would shift the burden of armed conflict further onto civilians. The use of fully autonomous weapons would create an accountability gap as there is no clarity on who would be legally responsible for a robot's actions: the commander, programmer, manufacturer, or robot itself? Without accountability, these parties would have less incentive to ensure robots did not endanger civilians and victims would be left unsatisfied that someone was punished for the harm they experienced.¹

It is the latter issue of the responsibility of LAWs designers and engineers that I want to engage for it is one to which John Forge has dedicated much effort and I have personally grappled in my work for military forces, foreign and domestic. There has been significant debate in the literature about responsibility gaps, concerning whether these would or could arise if LAWS were deployed and the impact this should have on their development. We already have machines in development, and a limited number already in use, which are task-autonomous and can decide on a course of action in some limited scenarios without any human input. Going forward, many suggest that all indications point to there being lethal machines with rules for action that are not fixed by their manufacturers during the design and/or production process and which are open to be changed by the machine itself during its operation. That is, these machines will be capable of learning from their surroundings and experiences. Conventionally, there are several loci of responsibility for the actions of machines, but both Andreas Matthias² and Robert Sparrow³ argue that these more advanced robots will bring about a class of actions for which nobody is responsible, because no individual or group has sufficient control of these systems, including the designer.

I have elsewhere argued, in discussing the nature of this alleged responsibility gap and showing its inadequacy as a justification for a ban on autonomous systems, that the scope of the conditions for imposing responsibility, and hence, accountability, are frequently overstretched or considered in too wide a frame.⁴ Interestingly, whilst usually a point of

¹ Campaign to Stop Killer Robots, "The Problem," available at www.stopkillerrobots.org/learn/, Accessed April 18, 2021.

² Andreas Matthias, "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata," *Ethics and Information Technology* 6 (2004): 177.

³ Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24/ 1 (2007): 62-77.

⁴ See Jai Galliot, *Force Short of War in Modern Conflict: Jus ad Vim* (Edinburgh: Edinburgh University Press, 2019) ; Jai Galliot, *Military Robots: Mapping the Moral Landscape* (New York: Routledge, 2016); J. Galliot, J. Scholz, "Artificial Intelligence and Space Robotics: Questions of Responsibility," in *Commercial Space Exploration: Ethics, Policy & Governance*, ed. Jai Galliot (New York: Routledge, 2016), 211-227; J. Galliot, "Cyber warfare, asymmetry, and responsibility: Considerations for defence theorem," in *Handbook of Research on Civil Society and National Security in the Era of Cyber Warfare*, ed. Metodi

contention for those with opposing viewpoints in the LAWs debate, Forge and I share the view that designers can and should be held responsible and to account for certain uses of artefacts of war, generally those that are either intended and could reasonably be foreseen, although we draw different conclusions about the implications of the attribution of said responsibility.

To develop guidance on which uses a designer is responsible for, Forge has developed a threefold taxonomy of the uses of an artefact that can be applied to LAWs⁵ and provides the following summary:

Thus, the *primary* purpose of an artefact is the purpose for which it is intended by the designer: it is what the designer designs it to do. The primary purpose of an assault rifle, for example, is to kill or otherwise harm others, by firing single shots or in semi-automatic mode, and whether a particular use of an assault rifle is justified depends on the circumstances. If the threatened use of the weapons is sufficient to stop some incident, then I call this a *derivative* purpose because it is contingent on the primary purpose (and not vice versa). Assault rifles are not intended to hold flowers, but a long-stemmed flower could be lodged down the barrel as a way to cheer up a barrack room. If so, this would be a *secondary* purpose of the weapon, and as such is not contingent on its primary purpose: it is quirky and fortuitous.⁶

The primary purpose of a lethal autonomous weapon, Forge would then suggest, is to select and engage its targets (something intended by the designer), and a derivative purpose may be consequential suppression of insurgent activity while such groups are under surveillance or within reach of LAWs (this is something contingent on the primary purpose of the weapon). By this account, designers, engineers and the like are accountable when someone is engaged and killed by an autonomous weapon, but they would not necessarily be responsible for secondary uses of the same weapon, such as uses with means or in contexts that could not be anticipated. But should non-combatants or other innocents be targeted and killed, I would share Forge's view that the designer retains a degree of responsibility (perhaps with the individuals who put it into use) by virtue of the weapon having been designed with the intent to kill and maim, knowing that the programming may be imperfect or that it could be misused or incorrectly deployed in the future, just as Kalashnikov should have foreseen the potential broader consequences/applications of his AK-47's intended purpose: killing with bullets.⁷ Such cases cannot be dismissed as being 'secondary' in Forge's taxonomy. Nor can we say that a designer is absolved of responsibility for a weapon that somehow operates out of control, because one designs weapons knowing that whilst unlikely, such erratic technical events are, at least, theoretically possible.

Where I depart from Forge is in his insisting that designers abstain from participation in design of autonomous weapons in the absence of a guarantee of their justified use. In fact, I would argue that the opposite is true. James Garvey has argued that in much the same way that 'ought implies can,' 'can implies ought' in a range of other circumstances where the

Hadji-Janev & Mitko Bogdanoski (Hershey: IGI, 2016), 1-21 ; J. Galliot, "Responsibility for War Machines," in *Rethinking Machine Ethics in the Age of Ubiquitous Technology*, ed. Jeffrey White & Rick Searle (Hershey: IGI, 2015), 152-165.

⁵ John Forge, *The Responsible Scientist* (Pittsburgh: Pittsburgh University Press, 2008), 156-159.

⁶ John Forge, "Closing the Gaps: Lethal Autonomous Weapons and Designer Responsibility," *Morality Matters* (2018), available at: <https://www.moralitymatters.net/on-weapons-research/closing-the-gaps-lethal-autonomous-weapons-and-designer-responsibility/>, Accessed January 29, 2019.

⁷ John Forge, "No Consolation for Kalashnikov," *Philosophy Now* 59 (2007): 6-8.

financial or political means behind the ‘can’ have contributed to the problem that ought to be corrected or mitigated. This also seems to hold true in the LAW’s debate with which we are engaged. Generally speaking, the more power an agent has and the greater the resources at their disposal – whether intellectual, economic or otherwise – the more obliged s/he is to take *reasonable* action when problems arise. Withdrawing from the design process does not satisfy the ‘reasonableness’ requirement, from my realist social contract view,⁸ given that we occupy a world in which the most complex of international problems are resolved through using force, either in full scale or ‘short of war’ modes.⁹ As I have argued more fully elsewhere, “the companies of the military-industrial complex are in a unique position, and have it well within their power, to anticipate risks of harm and injury and theorise about the possible consequences of developing learning systems.”¹⁰ The costs of doing so, after the fact, or once a less reputable manufacturer has taken up the task, are great and many. Moreover, responsible manufacturers are best positioned to create opportunities for designers and engineers to do what is right without fear of reprimand, whether that would be going ahead as planned, designing in certain system limitations or simply refusing to undertake certain projects, especially now that autonomous weapons manufacturers are able ‘ethics’ into their design guidance and contractual requirements.¹¹ It therefore seems reasonable to impose forward-looking responsibility upon manufacturers and their designers. However, we know that profit can sometimes trump morality for these collective agents of the military-industrial complex, so concerned parties should also seek to share forward-looking responsibility and ascribe some degree of responsibility to the governments which oversee these manufacturers and set the regulatory framework for the development and deployment of their product, perhaps through enhance Article 36 provisions, should manufacturers fail to self-regulate and establish appropriate ethical design standards.

A blanket prohibition on ‘AI in weapons,’ or participation in the design and engineering of artificially intelligent weapons, would have unintended consequences due to its lack of nuance. There is an important distinction to be made between those kinds of AI that would have humanitarian benefits and those have no such promise. This lack of nuance is also evident in the case against chemical weapons. For example, pepper spray or tear gas is a chemical agent banned in warfare under the Chemical Weapons Convention of 1993, making it illegal for use by militaries except in law enforcement. The denial of tear gas to military forces removes a less-than-lethal option from the inventory, which could lead to the unnecessary use of lethal force. Another consequence of a ban would be to deny the use of autonomous weapons as a countermeasure against other autonomous weapons. Meanwhile, with fears about non-existent sentient robots stalling debate and halting technological progress, one can see in the news that the world faces real ethical and humanitarian problems in the use of existing weapons. A gun stolen from a police officer and used to kill, guns used for mass shootings, vehicles used to mow down pedestrians, a bombing of a religious site, a guided-bomb strike on a train bridge as an unexpected passenger train passes over it, a missile strike on a Red Cross facility, and so on – many of which may be preventable by using AI in weapons and in autonomous systems more generally.

Confusion about the means to achieve desired nonviolence is, of course, not new. A general disdain for simple technological solutions aimed at a better state of peace was prevalent in the anti-nuclear campaign spanning the confrontation period with the Soviet Union, recently renewed with the invention of miniaturized warheads, and during the

⁸ Jai Galliot, *Military Robots: Mapping the Moral Landscape*, chapter 3.

⁹ Jai Galliot, *Force Short of War in Modern Conflict: Jus ad Vim*.

¹⁰ Jai Galliot, *Military Robots: Mapping the Moral Landscape*, 226.

¹¹ Jai Galliot, J. Scholz, “Artificial Intelligence and Space Robotics: Questions of Responsibility.”

campaign to ban land mines in the late nineties. Yet, it does not seem unreasonable to ask why weapons with advanced autonomous seekers could not embed AI to identify a symbol of the Red Cross and abort an ordered strike. Or why the location of protected sites of religious significance, schools or hospitals might be programmed into LAWS to constrain their actions, just as we could prevent guns from being firing by an unauthorized user pointing it at humans. And why initiatives cannot begin to test these innovations and how they might be ensconced in International weapons review standards? I assert that while autonomous systems are likely to be incapable of action leading to the attribution of moral responsibility in the near term,¹² they might today autonomously execute value-laden decisions embedded in their design and in code, so they can perform actions to meet enhanced ethical and legal standards, even holding their designers more accountable through digital responsibility-tracking means.¹³

Jai Galliot

Reply to Jai Galliot

I'm very pleased to have the opportunity to reply to Jai Galliot and I'm grateful to the editor of this special issue of the *PJCV* for inviting me to do so, especially because the topic, weapons research, design and responsibility, is one that I have been engaged with for some years. It is an important topic but one that has been neglected in the past, although now there is more interest, with Galliot (and the editor of this issue) making notable contributions. This has come about, in part, because of problems in regard to a particular kind of weapons system, namely Lethal Autonomous Weapons Systems (or LAWS), which are the main focus of Galliot's contribution. My own interest in weapons research had a different origin: the development of the atomic bomb. It is well-known that the 'context' in which the atomic bomb was conceived and developed was different from the one in which it was used, to the dismay of many of those who had done the original research. The lesson to be learnt from this episode applies to all weapons research: it is that weapons intended for one particular purpose may come to be used in quite unanticipated and unacceptable ways in the future. The suggestion that some weapons research, into nuclear or biological weapons for instance, should be banned is not too contentious. I maintain that *all* weapons research should be banned, and that puts me firmly at one end of the spectrum of possible positions.¹⁴ Galliot notes while that some (like me) call for a ban on LAWS because they deplore all weapons research, he advocates a more 'nuanced' approach. What I hope to do here is to convince him to move along the spectrum, in my direction.

Galliot mentions my taxonomy of the purposes of an artefact, a weapon for example, and says that I hold the designer — here I take “designer” and “researcher” to be synonymous — responsible for the primary purpose. This is entirely correct. What is not clear from the passage quoted in Galliot's paper is that I take the designer to be responsible for providing a *means* to kill, which is the primary purpose of a weapon, but I do not think that the designer is necessarily also responsible for each and every individual occasion in which ‘her’ weapons kills someone. If that were true, then Mikhail Kalashnikov would be responsible for millions of deaths since 1947, and Leo Szilard, Enrico Fermi, Otto Frisch and others would be responsible for Hiroshima and Nagasaki. And to give one more example, in the first two years

¹² Ibid.

¹³ Ibid.

¹⁴ For the complete argument, see John Forge, *The Morality of Weapons Research* (Dordrecht: Springer, 2019) ; *Designed to Kill: The Case against Weapons Research* (Dordrecht: Springer, 2013).

of World War Two, the standard infantry weapons of the Wehrmacht was the Mauser Carbine, designed in 1898. It was also the weapon used by the *Einsatzgruppen* that carried out the ‘holocaust by bullets’ in the Soviet Union after the German invasion, when at least a million innocent people were killed. But one cannot surely hold Paul Mauser responsible for these atrocities. I believe that the designer of a weapon is responsible for individual killings if she had reason to believe her weapon would be used in the conflict were the killings took place. However, I argue that it is *always* morally wrong to design weapons because weapons are the means to harm and because it is morally wrong to provide the means to harm. One could say here, though this is not normally how I put the matter, that weapons research is morally wrong because it risks harm.

At the beginning of the paragraph after the one in which he cites the quotation about purposes, Galliot writes “Where I depart from Forge is in [his] insisting that designers refrain from participation in autonomous weapons design in the absence of a guarantee of their justified use.” I do indeed insist on this, not just for LAWS but for all weapons. This is because the particular system of morality which underpins my account is based on the principle that it is morally wrong to harm without justification, and that the only justification for harming is in preventing at least as much harm as is caused. So, one agent is only permitted to harm another if she has good reason to believe that this will prevent at least as much harm, otherwise the judgement that what she does is wrong stands. I argue that the same goes for providing the means to harm: the only justification for providing new means to harm is if there is good reason to believe that this will prevent at least as much harm. So, for example, using weapons to defend against harm or using them to deter others from harming could stand as possible justifications. The Soviet tank factories in World War Two provided T-34 and KV tanks to fight the Germans and it seems clear that this was altogether justified. But providing weapons and providing *designs* for weapons are entirely different; the latter can be reproduced many times over, at different times and places. Numerically different but qualitatively the same T-34 tanks were used to suppress the East German uprising in 1953 and the Hungarian Revolution of 1956, which was not justified. And Kalashnikovs been made in at least ten countries and copied in many others, with tens of millions being made. It is therefore impossible for a weapons designer to have good reason to believe that her creation will only have justified uses because she cannot know, or guess, all of the ways it will come to be used.

Galliot has a different view about responsibility, which is evidently based on a different conception of morality — his ‘realist social contract view.’ I take this to imply a *positive duty* to engage in weapons research in certain circumstances, for example to design LAWS that have the right sorts of safeguards built in. Translating this into the language I have been using, we can say, I think, that Galliot adopts a moral system that mandates the prevention of harm. This seems plausible: if harming is morally wrong, then surely preventing harm is morally praiseworthy. Applying this intuition to weapons research, one might then argue that it can be justified if the aim is to prevent harm, either by defence or deterrence — I take it that this is what Galliot is saying in his second last paragraph. He says at the beginning of that paragraph that a blanket prohibition on ‘AI weapons’ would have unintended consequences. My response to that is that if everyone agreed to the ban, we would have no such weapons, and I think that is a good thing. But the problem with weapons research, as I have pointed out, is that there often are unintended consequences, in the form of the weapon in question been used ‘out of context,’ in new and unacceptable ways. The weapons designer may be held to account for these, but on my view of morality, she is not accountable for the consequences of her failing to engage in weapons research. Failure to prevent harm is not, according to my system, morally blameworthy.

I will be interested to read Jai Galliot's reply to these comments. To conclude my remarks, I will come back to LAWS and responsibility gaps. The so-called problem of responsibility gaps comes about according to a scenario in which a LAW is able to select its own targets: it does so, but picks out innocent civilians and kills them. Who is responsible and who is to blame? There is no shortage of candidates, including the designer, the person who set up the LAW for the mission in question and the local commander. The designer may be responsible for these particular killings, but my account can add nothing to what has already been said about this question, by Rob Sparrow and others for example. However, if there were to be operational LAWS in existence, then according to my account, moral wrongdoing has *already* occurred, before the weapon is even deployed. A new way of harming has come into the world, and that for me is wrong. What the scenario just outlined does, or would do were it to become reality, is both illustrate my claim that this instance of weapons research is wrong because of unforeseeable wrongful uses and confirm my overall account that all weapons research is wrong, for the same reason. It would be more evidence to add to the long list of examples of the unjustified use of weapons.

John Forge

Response to John Forge

John Forge's response, while humbling, has not yet convinced me to move along the spectrum in his direction. But let us briefly rehash his illustrative argument before I outline my reasons: Forge accepts that where we depart is at his insistence that "designers refrain from participation in autonomous weapons design in the absence of a guarantee of their justified use" and goes on to argue that developing the Kalashnikov for its originally intended purpose or building T-34 and KV tanks to fight Nazi Germans in World War Two may have been altogether justified, but that the case of providing "weapons and *designs* for weapons" is entirely different, noting that such tanks went on to be used against East Germany in the post-war period and the AK-47 would go on to kill millions in other conflicts. Forge believes this illustration is demonstrative of the fact that it is impossible for a weapon designer to have reason to believe their creation will have only justified uses or knowledge of all possible uses.

While it is somewhat unfair to critique Forge's argument without full reference to his voluminous work in the area, which is impossible in this short response, it strikes me from his writings that there exist a couple of problems with his argument and I therefore make just a few brief points. The first is to say that Forge's argument is, in my view, too demanding of us in that it admits no shades of grey. Rather than admit that there are degrees of responsibility and corresponding levels of accountability, which is likely more in line with the collective nature of all modern weapon design, Forge attempts to force us into accepting some form of black-and-white argument: we either accept that Kalashnikov would be responsible for all deaths stemming from the invention of his rifle; or that a designer is causally responsible for the weapon they design and should also foresee that they are designing something that can only kill and maim and therefore renounce their profession; or bite the bullet and accept the existence of some kind of 'responsibility gap' (pertaining to actions for which nobody is responsible).

To briefly address this latter point and pre-empt Forge's likely comments concerning responsibility gaps, I am often amazed to find that thought experiments, usually relied upon to conveniently explain away problems, are used in the lethal autonomous weapons space to manufacture problems that simply do not exist. In the case of Sparrow,¹⁵ which represents

¹⁵ Robert Sparrow, "Killer Robots," *Journal of Applied Philosophy* 24/ 1 (2007): 62-77.

the most prominent example, we are presented with the idea of a robot-child soldier analogy. The idea he advances is that there is a conceptual space in which child soldiers and military robots are sufficiently autonomous to make the full attribution of responsibility to an adult or conventional moral agent problematic, but not autonomous enough to be held fully responsible themselves. Sparrow argues¹⁶ that his opponents try to close this space by stipulating that the relevant entities hold more or less responsibility than they should and thus fit within one of the polar boundaries, but it is my view that he is incorrectly trying to widen the responsibility gap by fundamentally failing to understand the limitations of artificial intelligence and machine learning, and overlooking the huge human element in even the most autonomous systems available today.¹⁷

But in returning to Forge's argument and applying it to technology writ large, I think we can see how easy it would be to extrapolate from this a *reductio ad absurdum*-type counter argument based on the account's ability to lead us toward potentially society-destroying anarcho-primitivism. Cars, for instance, are not designed to kill people, but they do, and because of it we strictly regulate their design and drivers. We do not prohibit the design of new cars. Under my account, the fact that one of the intended uses of weapons is killing is of little moral relevance in the grand picture. Forge seems to be invoking — when placed in just war-theoretical terms — the concept of proportionality in saying that it is “weapons research [which] is morally wrong because it risks harm,” but with little to no regard for the positive uses of technology. I have already pointed to the moral benefits of embedding artificial intelligence in lethal autonomous weapons and of further relevance here is that states often fail to engage in humanitarian crises because of an aversion to casualties, with it thought that autonomous technology may assist in overcoming this reluctance to intervene what might be regarded as supererogatory cases, thus enhancing the proportionality calculus.¹⁸ Of course, under Forge's account, failure to prevent harm is not morally blameworthy. This is, I suppose, a key difference between his account and my social-contract view.

The absence of a positive duty to prevent harm to others is interesting because, at least in some respects, Forge has a rather idealistic view of personal morality. On the one hand, he thinks it too onerous to hold a designer responsible for unintended uses of one's design and yet thinks one should abstain from participation in the design of weapons in an absolute fashion. Central to the whole argument is that war is bad, its conduct leads to harm and war is something that we can somehow avoid. It is of course true that without weapons there can be no *armed* conflict of the kind we typically envision in the modern day, but that is not to say there will be no conflict or war of other kinds. What Forge may think is a framework for design liability in modern society, I would suggest, is a framework for liability in something much nearer to a fictional utopia. Under my account, as in reality, there is an understanding that we need people to be tools of the state. Given this reality then, to later say that there is no positive duty to develop weapons that we can responsibly deploy, or to say that there is a responsibility gap for something that the state and its people have asked of them, is ridiculous. If anything, one could argue that it is morally wrong for a designer in some circumstances to

¹⁶ Ibid., 72.

¹⁷ Jai Galliot, “The Unabomber on Robots: The Need for a Philosophy of Technology Geared Toward Human Ends,” in *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, ed. Patrick Lin, Ryan Jenkins & Keith Abney (New York: Oxford University Press, 2016), 355-369 ; Jai Galliot, “Artificial Intelligence in Weapons: The Moral Imperative for Minimally-Just Autonomy,” *US Airforce Journal of Indo-Pacific Affairs* 1/2 (2018): 57-67.

¹⁸ Z. Beauchamp, J. Savulescu, “Robot Guardians: Teleoperated Combat Vehicles in Humanitarian Military Intervention,” in *Killing by Remote Control: The Ethics of an Unmanned Military*, ed. Bradley Strawser (New York: Oxford University Press, 2013), 106-125.

decide of their own accord not to conduct a general action that everyone else has sanctioned and requires him to do. If the autonomous technology or war is not something the public endorse, of course, that is another matter entirely and not the responsibility of the designer, but rather the politicians who have established such circumstance.

Jai Galliot

Rejoinder to Jai Galliot

I thank Jai Galliot for his response. I have only a space for two quick comments by way of a rejoinder. One of Galliot's criticisms is the same as one by David Resnick¹⁹ when he reviewed my book *Designed to Kill* and I will start with this.

In his fourth paragraph, Galliot suggests that my argument should — for me to be consistent? — apply to all technological design, and he thinks this leads to a *reductio*. I argue that weapons designers cannot know how, where or when the outcomes of their work will be used. For instance, Kalashnikov could not have known that 'his' assault rifle would be used to suppress dissent in East Germany in 1953, when he started his research in 1941. Nevertheless, I claim that, among others, he is responsible. Galliot mentions cars and says that they can kill people, but that we do not therefore look askance at people who design cars, who could not know that this could happen. So, suppose a Mack truck is used by a terrorist to kill people, do we then hold the designer responsible? Of course not! And I am not committed to saying that we do, because I maintain that designers are only always responsible for, and committed to, the *primary purpose* of the artefacts that they design, and killing people is not the primary purpose of a truck. I first introduced a taxonomy of purposes of artefacts in my book *The Responsible Scientist*,²⁰ along with my claim about designer responsibility, and have elaborated it in 2013 and again in 2019.²¹ Very quickly, an artefact can have three purposes or functions: the primary purpose, which is what it is designed to do, a derivative purpose which supervenes on the primary purpose, and a secondary purpose, which is fortuitous. If the primary purpose of truck is to haul heavy goods, then I think using it as a weapon to kill people is secondary. However, I claim that weapons are the only artefacts designed to kill, that is, they are the only artefacts whose primary purpose is to kill. I believe therefore that Galliot, and Resnick, are mistaken in thinking that my argument leads to a *reductio*.

Elaborating a little further, as Galliot remarks I do not think responsibility comes in degrees, but I think blame does. Moral persons do not want to act in ways that are blameworthy — by definition — and acts that harm are *prima facie* blameworthy. Such acts can be justified, for instance, if they prevent more harm than they cause. If weapons researchers intentionally provide the means to harm by designing weapons and then the weapons are used to harm, they are responsible, but not of course solely responsible, for that harm. Just who, if anyone, is to be *blamed* depends on the circumstances. It is no *excuse* for the weapons designer to maintain that she only wanted her work to have justifiable uses, to prevent harm, because it is impossible to design a weapon that can only have uses that are good and just, as Mikhail Kalashnikov conceded in his old age. Thus, the only course of action for a moral person is not to engage in weapons research.

John Forge

¹⁹ David Resnick, "Is Weapons Research Immoral?" *Metascience* 23 (2014): 105-107.

²⁰ John Forge, *The Responsible Scientist*.

²¹ John Forge, *Designed to Kill: The Case against Weapons Research*; John Forge, *The Morality of Weapons Research*.

References

- Beauchamp, Z, Savulescu, J. "Robot Guardians: Teleoperated Combat Vehicles in Humanitarian Military Intervention." In *Killing by Remote Control: The Ethics of an Unmanned Military*, ed. Bradley Strawser, 106-125. New York: Oxford University Press, 2013.
- Campaign to Stop Killer Robots. "The Problem." Available at www.stopkillerrobots.org/learn/. Accessed April 18, 2021.
- Forge, John. "Closing the Gaps: Lethal Autonomous Weapons and Designer Responsibility." *Morality Matters* (2018). Available at: <https://www.moralitymatters.net/on-weapons-research/closing-the-gaps-lethal-autonomous-weapons-and-designer-responsibility/>. Accessed January 29, 2019.
- Forge, John. *The Morality of Weapons Research*. Dordrecht: Springer, 2019.
- . *Designed to Kill: The Case against Weapons Research*. Dordrecht: Springer, 2013.
- . *The Responsible Scientist*. Pittsburgh: Pittsburgh University Press, 2008.
- . "No Consolation for Kalashnikov." *Philosophy Now* 59 (2007): 6-8.
- Galliot, Jai. *Forge Short of War in Modern Conflict: Jus ad Vim*. Edinburgh: Edinburgh University Press, 2019.
- . "Artificial Intelligence in Weapons: The Moral Imperative for Minimally-Just Autonomy." *US Airforce Journal of Indo-Pacific Affairs* 1/2 (2018): 57-67.
- . *Military Robots: Mapping the Moral Landscape*. New York: Routledge, 2016.
- Galliot, J., Scholz, J. "Artificial Intelligence and Space Robotics: Questions of Responsibility." In *Commercial Space Exploration: Ethics, Policy & Governance*, ed. Jai Galliot, 211-227. New York: Routledge, 2016.
- Galliot, J. "Cyber warfare, asymmetry, and responsibility: Considerations for defence theorem." In *Handbook of Research on Civil Society and National Security in the Era of Cyber Warfare*, ed. Metodi Hadji-Janev & Mitko Bogdanoski, 1-21. Hershey: IGI, 2016.
- . "The Unabomber on Robots: The Need for a Philosophy of Technology Geared Toward Human Ends." In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, ed. Patrick Lin, Ryan Jenkins & Keith Abney, 355-369. New York: Oxford University Press, 2016.
- . "Responsibility for War Machines." In *Rethinking Machine Ethics in the Age of Ubiquitous Technology*, ed. Jeffrey White & Rick Searle, 152-165. Hershey: IGI, 2015.
- Matthias, Andreas. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata." *Ethics and Information Technology* 6 (2004): 175-83.
- Resnick, David. "Is Weapons Research Immoral?" *Metascience* 23 (2014): 105-107.
- Sparrow, Robert. "Killer Robots." *Journal of Applied Philosophy* 24/ 1 (2007): 62-77.