Philosophy Study, May 2018, Vol. 8, No. 5, 231-239
doi: 10.17265/2159-5313/2018.05.004

Lethal Autonomous Weapon Systems and Responsibility Gaps

Anne Gerdes
The University of Southern Denmark

This paper argues that delegation of lethal decisions to autonomous weapon systems opens an unacceptable responsibility gap, which cannot be effectively countered unless we enforce a preemptive ban on lethal autonomous weapon systems (LAWS). Initially, the promises and perils of artificial intelligence are brought forward in pointing out (1) that it remains an open question whether moral decision making, understood as situated ethical judgement, is computationally tractable, and (2) that the kind of artificial intelligence, which would be required to cause ethical reasoning, would imply a system capable of operating as an independent reasoner in novel contexts (sec. 2). In continuation thereof, issues of responsibility are discussed (sec. 3 and 3.1) and it is claimed that unacceptable responsibility gaps may occur since unpredictability would presumably follow full system autonomy. These circumstances call for a strong precautionary principle, in the form of a preemptive ban.

Keywords: LAWS, artificial intelligence (AI), responsibility

1. Introduction

Weaponized (semi-) autonomous technologies are rapidly entering the scene of warfare, and although their capabilities already by now allow for autonomous actions, we have not yet witnessed these systems initiating lethal action without humans in the decision loop. But this could very well change in a near future. If this is the case, we may envision warfare scenarios with no a clear chain of accountability and nobody to be held responsible for serious moral wrongdoing.

When dealing with the topic of weaponized technologies and the implications of their autonomous lethal capacities, discussions are often carried away by arguments, which initially come with some reservations, but nevertheless end up exaggerating the potentials of AI (artificial intelligence). In general public debates as well as in research, optimistic AI-arguments are brought forward as good reasons for permitting lethal autonomous weapon system (LAWS). Hence, Sharkey, in addressing the impact of anthropomorphism on AI, points to the "mythical narrative from science fiction and media" based on unrealistic assessments of robot capabilities, and underscores that "the real danger is in the language being used by military researchers and others to describe robots and what they can do" (Sharkey 2012, 787). In this sense, the AI-narrative may potentially mislead politicians into releasing LAWS without proper testing and scientific evidence (Sharkey 2012, 791).

Similarly, within research circles, arguments are cast around the assumption that machine-based rational engagement might outperform irrational emotion driven human actions in the fog of war and thereby prevent war crimes (Arkin 2009; Müller & Simpson 2014; 2016). Thus, proponents accentuate the benefits as if a

Anne Gerdes, Ph.D., Associate Professor, Department of Design and Communication, The University of Southern Denmark, Denmark; main research fields: Robot Ethics, Philosophy of Artificial Intelligence, and Privacy.

Acknowledgements: The author is grateful to the colleague, Stig Børsen Hansen, whose valuable insights and comments helped shape this paper.

breakthrough in AI is upcoming, assuming that we shall soon be seeing autonomous weapon systems demonstrating capabilities for situational awareness, as well as situated ethical judgment. Proponents argue that sophisticated AI-models with the right kind of programming might eventually lead to LAWS capable of acting by the laws of war. Yet, as illustrated in section 2, discussions about the feasibility of artificial intelligence are still very open, as is the issue of whether ethics is computationally tractable, in the sense of autonomous AI-agents capable of the kind of situational awareness needed to react responsibly in dynamic real-life settings.

However, on the other side, in complex technologically mediated contexts (Singer 2009; Cummings 2006), humans have started to move out of the loop a long time ago. Consequently, seeking to anchor responsibility in a notion of Kantian autonomy may seem overly idealistic since humans often work as mere interfaces between technologies, which makes it hard to delimit the level of human involvement and control in the first place. Hence, some suggest (Verbeek 2009; Floridi & Sanders 2004; Coeckelbergh 2011) that issues of robo-responsibility are best described by introducing a distributed notion of responsibility (section 3). In dealing with responsibility gaps in an armed context, Müller and Simpson (2014; 2016) maintain that the loci of responsibility can be singled out at the level of persons and handled within the legal war framework already in place, assuming that we may treat responsibility gaps in a warfare context similarly to gaps in civilian contexts. But, alas, as shall be demonstrated in section 3.1, both approaches leave behind undesirable responsibility gaps, which stress the need for a preemptive ban.

2. Artificial Intelligence and LAWS

The ethical issues related to LAWS primarily reside in the context of *Jus in Bello*, which concerns the morally acceptable conduct of war as spelled out in International Humanitarian Law (IHL). Hence, these regulations clarify the legal framework and form a springboard for establishing, which moral requirements we need to pay attention to in assessing the performance of LAWS. The hallmark criteria of IHL are the principles of proportionality and discrimination. Thus, LAWS must be capable of balancing ends and goals proportionately, implying that an autonomous weapon system must restrain from using unnecessary violence or coercion in obtaining its military goals. Presumably, LAWS cannot be programmed to encompass all eventualities. However, the fact that they will be capable of making calculations faster than humans and without emotional distractions, are, by some (see for instance Arkin 2007; 2009; Lokhorst & van den Hoven 2012; Lin et al. 2008), considered to make it plausible that LAWS shall eventually come with the capacity to outperform humans and lower the risk for war crimes. Moreover, such systems would be able to monitor and report battlefield misconduct. Likewise, in cases in which an officer commands against IHL and specific rules of engagement, such systems can reject to follow orders and thereby prevent war crimes from happening; whereas soldiers often tend to follow orders blindly (Arkin 2007, 6; Lin et al. 2008, 50). In continuation thereof, the principle of discrimination refers to the obligation to discriminate between combatants and noncombatants, and avoid intentionally causing harm to civilians. The reference to intentions discloses the acceptance of the Doctrine of the Double Effect (Quinn 1989), which provides a deontological explanation of why it may sometimes, for instance in self-defense, be acceptable to kill or harm others. Hence, there are situations in which it is permissible to cause harm to others as a side effect of doing good and acting from the right intentions. Or, as stated in its first explicit version by Aquinas: “Nothing hinders one act from having two effects, only one of which is intended, while the other is beside the intention” (Aquinas, Summa II-II, 64, 7). Consequently, during the conduct of a military mission, collateral damage to civilians can be accepted as long

as the harm is either unforeseen or is not the goal in itself or the mean to obtain the goal, but rather represents unintended, all though possible foreseeable, harm in the cause of reaching a necessary military goal. According to Arkin, the standards for LAWS might be heightened by implementing Walzer's (1977) Principle of Double Intention, which furthermore lowers the tolerance for collateral damages, inherent in the Double Effect doctrine, by emphasizing the necessity of intentionally seeking to reduce civilians' sufferings (Arkin 2007, 12).

However, *Jus in Bello* decisions regarding proportionality and discrimination are complicated and context-sensitive, resulting in possible incommensurability of the relevant factors that demand attention in a dynamic real-life setting. Consequently, it is impracticable to produce a set of strict decisions rules to capture situated moral reasoning, which relies on experience grounded in practice. This kind of moral knowledge is reflected in the Aristotelian notion of *phronesis*—i.e., the kind of situated practical wisdom needed in order to know good means to bring about good ends (Aristotle 1909)—and further refined in the presupposition of a conception of psychological concepts in virtue ethics (Anscombe 1958) emphasizing that being a moral agent requires more than being rational. Nevertheless, Abney strikes an optimistic tone and argues for a "robotic turn" assuming that outsourcing of war to artificial moral agents shall improve warfare, or even pave the road for a peaceful world:

> Autonomous lethal robots could herald, not a Terminator scenario of the "rise of the machines" and the death of humanity, but an enforceable and lasting peace, based on an international moral consensus that disagreements are always to be resolved by negotiations or robotic "proxy wars", and never by coercive violence against humans. (Abney 2013, 348)

Elaborating on this idea, Abney presents a so-called hybrid approach to the programming of lethal autonomous robots, which consists of both symbol manipulating and self-learning programming components. Presumably this would allow for the development of a virtuous robot (Abney 2013, 347). This line of argument is also reflected in his work together with Lin et al., which suggests incorporating the appropriate virtues related to roles in a military context, such as "warrior fierceness towards enemies and gentle kindness towards comrades" (Lin, Bekey, & Abney 2008, 38).

This notion of a "hybrid approach" was originally introduced by Wallach and Allen (2009) in distinguishing between three different approaches towards the programming of moral machines.[1] Here, they also emphasize that these approaches be infeasible in the near future, if ever. Moreover, Wallach assumes that "killer robots are *mala in se*" (Wallach 2015, 219), and suggests the establishment of an international humanitarian principle "that machines should not be making decisions about killing humans" (Wallach 2015, 214).

With these reservations in place, their framework offers perspectives on different engineering models. Hence, a top-down approach refers to rule-based programming, which is not promising due to the frame problem (Wallach & Allen 2009), which refers to the challenge of capturing conditions in dynamic changing environments by formal knowledge representation (Dennett 1988; Ford 1996). On the other side, a purely bottom-up approach, in the shape of connectionist networks, which implies learning by doing (performance optimization through representation of different trial-and-error methods), will not do either, since we cannot be sure of the outcome—i.e., that such machines will develop the kind of morality we wish for (Wallach & Allen 2009, 113).

Therefore, Wallach and Allen suggest a hybrid-model, which integrates top-down and bottom-up approaches by incorporating virtue ethics as a theoretical foundation for implementation of the idea of how we

develop into virtuous persons through habituation. This mirrors a model of connectionism, understood as bottom-up learning through deep learning algorithms and pattern recognition. Traditionally, a connectionist network provides a strategy for the accumulation of data as a foundation for the build-up of generalized responses, which extend beyond the original data training set. However, learned patterns emerge without explanation for why certain actions are chosen over others. Moreover, the lack of general rules makes it difficult to decide what to do when faced with novel situations. Here, a hybrid model would allow for virtues to be represented as top-down patterns, which might scaffold the systems evaluation of actions. Nevertheless, while such approaches guide considerations about moral machines, they do not stipulate that LAWS are computationally tractable; a point emphasized by Wallach in the quotation below:

> Nevertheless, the prospect is low for developing robot soldiers any time soon with the ability of making an appropriate judgement in a complex situation. For example, a robot would not be good at distinguishing a combatant from a non-combatant, a task that humans also find difficult. Humans, however, bring powers of discrimination to bear for meeting the challenge; capabilities that will be difficult, if not impossible, for robots to emulate. (Wallach 2015, 218)

Likewise, if one takes a look at the capabilities demonstrated at the recent flagship conference in artificial intelligence, namely DARPA Robotics Challenge 2015,[2] one will see robots dwelling for several minutes before carrying out simple tasks, such as turning a doorknob. Moreover, while pattern recognition, which is important to master in order to discriminate between objects, has advanced in static contexts, it is still futuristic to imagine weaponized autonomous technologies capable of discriminating between combatant and non-combatant, friend and foe, and e.g., figuring out whether a person, waging a stick, is about to surrender or attack. Despite AI-breakthroughs in machine learning (based on representation learning methods, especially deep learning methods (LeCun et al. 2015)), the implementation of situational awareness in AI has not yet succeeded in overcoming the frame problem, i.e., the challenge of formalizing the kind of common world knowledge, which is a presumption for adequate behavior in dynamic changing environments (McCarthy & Hayes 1969; Dennett 1988; Ford 1996).

Nevertheless, ongoing research efforts in the field of weaponized autonomous technologies are motivated by expectations that full-blown AI is looming in a not too distant horizon. Recently, in 2014, the Office of Naval Research granted $7.5 million to develop an artificial moral agent (Tucker 2014). Although LAWS are currently not in use, the expectation is that such autonomous weapon systems can participate in carrying out ethical assessments, deliberating about moral dilemmas, and the implementation of military strategies. The US Department of Defense presents these prospects in the Unmanned Systems Integrated Roadmap FY2011-2036[3] and the Unmanned Ground Systems Roadmap.[4]

At the same time, it is worth noting that in 2012, the US Department of Defense issued a directive, which constitutes a time-limited (five to ten years) moratorium on LAWS with the possibility for certain abdications. However, as pointed out by Sparrow, “there is an obvious tension involved in holding that there are good military reasons for developing autonomous weapon systems but then not allowing them to fully exercise their ‘autonomy’” (Sparrow 2007, 68).

To sum up, it remains an open question whether Artificial General Intelligence on pair with human intelligence is a tractable problem of AI or not. Hence, we cannot know for sure that we are not going to see machines with genuine moral capabilities in the future, which have emerged from, e.g., sophisticated evolutionary models. Such moral systems would be independent reasoners and thereby cease to count as

functional tool-like agents operating in restricted domains. The kind of all-inclusive agency, which would follow from their ability to reason and act in dynamic real-life contexts, would make it highly complicated to safeguard such systems, as noted below:

> It is relatively easy to envisage the sort of safety issues that may result from AI operating only within a specific domain. It is a qualitatively different class of problem to handle an AGI operating across many novel contexts that cannot be predicted in advance. (Bostrom & Yudkowsky 2014, 318)

On the backdrop of these observations, issues of responsibility in technology-mediated contexts are discussed (sec. 3), serving as a springboard for discussions of how and why LAWS will inevitably be accompanied by unacceptable responsibility gaps (sec. 3.1).

3. Issues of Responsibility

Despite the objections mentioned above concerning the possibility of AI, some question whether it makes sense to uphold a distinction between human and machine responsibility, i.e., the idea that machines are limited by reasoning mechanisms governed by laws of causation and therefore only causal, but never moral responsible for their actions. According to this approach, it might make better sense to introduce a notion of distributed responsibility in complex technologically mediated settings. Consequently, in what follows, proponents of a distributed view on human-machine responsibility are introduced. Hence, Verbeek (2009), in dealing with ambient intelligence and persuasive technologies, proposes a post-phenomenological perspective. Here, Verbeek points to that we are situated in a material world, and therefore we cannot escape interacting with technologies. Consequently, freedom and intentionality each become “a hybrid affair...distributed over people and artefacts” (Verbeek 2009, 238). As a consequence, technology frames our capacity for acting free. By the same token, all though technological artefacts possess no form of human-like intentions, it is still possible to assign intentionality to technology in the limited sense that technologies can be said to play a directing role in relation to our actions and experiences (Verbeek 2011, 57). Hence, the interrelatedness of human and technology makes it impossible to locate freedom solely in humans (or technologies for that sake). Still, Verbeek suggests that we are still able to discuss issues of human responsibility since he assumes that “technological mediations can create the space for moral decision making” (Verbeek 2011, 60-61).

Likewise, Floridi and Sanders (2004) emphasize that moral agency may be attributed to intelligent artificial systems capable of adaptive behaviors and autonomous responses to the environment. Consequently, moral accountability can be assigned to an artificial agent, which is causally responsible for a given action. Here, what counts is whether an agent’s actions are morally qualifiable, that is if the agent is capable of carrying out actions that may be deemed as good or evil. Using this approach, they illustrate ways in which normative actions can unfold without necessarily implying moral responsibility based on intentional states (Floridi & Sanders 2004, 371). “[...] moral accountability is a necessary but insufficient condition for moral responsibility. An agent is morally accountable for x if the agent is the source of x and x is morally qualifiable... To be also morally responsible for x, the agent needs to show the right intentional states” (Floridi & Sanders 2004, 371).

According to Floridi and Sanders, their view on agency may enhance the moral discussion, concerning accountability, in non-human contexts. Presumably, maintaining consistency between human and artificial moral agents will make it possible to establish steps for censoring of immoral artificial agents in a continuum of

censure stretching from monitoring till annihilation (death) from cyberspace, i.e., deletion with no back up (Floridi & Sanders 2004, 373).

> We can stop the regress of looking for the *responsible* individual when something evil happens, since we are now ready to acknowledge that sometimes the moral source of evil or good can be different from an individual or group of humans... The greatest advantage is a better grasp of the moral discourse in non-human contexts. (Floridi & Sanders 2004, 376)

These suggestions hold that it might be fruitful to move beyond viewing human agency as the primary locus of responsibility and accountability. Similarly, in a military setting, Coeckelbergh offers an explanation of responsibility based on a relational ontology (Coeckelbergh 2011, 5). Here, he presents the concept of a "swarm," which refers to the military use of intelligent swarm systems seeking to simulate self-organized intelligent behavior in biological organisms and animals. The ambition is to be able to develop algorithms by analogy to the kind of collective intelligent behavior displayed by bird flocks, ants or bee swarms, and other self-organized systems (such as bacteria). Using this notion as a stepping stone, Coeckelbergh suggests that when talking about human-machine interactions, we should stop separating the parties involved, and instead assume "that humans and things are *already* connected (network metaphor) and buzzing together (swarm metaphor) in common activity before one can zoom in on particular connections and movements" (Coeckelbergh 2011, 6).

Due to the lack of a clear chain of accountability, Coeckelbergh realizes that this relational turn complicates issues of moral responsibility in a military setting (Coeckelbergh 2011, 7). Nevertheless, Coeckelbergh concludes with some optimism for our future endeavors within the field of (machine) ethics. Hence, the epistemological assumptions tied to his relational turn may work as a fruitful platform for tackling the challenges of responsibility gaps:

> My suggestions concerning possible moral, social, and epistemic implications of Singer's swarm conception, for instance, may be perceived as casting a dark shadow over the future. Lack of control and lack of predictability are the horror *par excellence* to the modern mind. However, it might be a consolation to consider that ethical reflections […] may develop into swarms too: initially invisible, but potentially powerful networks of people and ideas that can help us to better understand what awaits us, to find unexpected corridors for change, and to open up different windows of possibility. (Coeckelbergh 2011, 10)

3.1. LAWS and Responsibility Gaps

Moving on to a discussion of whether we should regulate or ban LAWS, Müller and Simpson (2014; 2016) acknowledge the challenge of regulation; indeed they claim that this is the only real problem in relation to weaponized autonomous systems: to clarify the responsibility of the user (criminal liability) as well as that of the maker (product liability). The context of warfare does not imply any special requirements and the mere existence of responsibility gaps does not count as an argument *per se* for banning the use of such technologies. Instead, we have to accept the inevitability of responsibility gaps in a war context in the same way as we accept such gaps in other situations. They exemplify this by referring to bridge building, claiming that if a bridge collapses and causes a lethal accident, we might find ourselves in a situation in which nobody can be held accountable. Thus, if the building of the bridge has taken place in accordance with legislation and given rules and standards, and if the conditions, which led to its collapse, could not have been foreseen, despite careful considerations, we accept the existence of a normal gap in the form of *a certain engineering tolerance* (Müller & Simpson 2016).

In continuation thereof, Müller and Simpson (2014; 2016) argue that the deployment of autonomous weapon systems does not mean that the primary locus can escape responsibility. By analogy, they bring in the use of child-soldiers, which does not create a gap, since responsibility is assigned to the commander in charge. Here, they are in alignments with Champagne and Tonkens (2013), who present the idea of "blank check responsibility," which means that responsibility is allocated to the level of a general or a president, who decides whether or not to deploy LAWS and hence after are held accountable should something go wrong. Apparently, under these circumstances, deployment of autonomous weapon technologies does not seem to collide with the doctrines of just war (Champagne & Tonkens 2013, 44).

These arguments underestimate that one should not at the outset aim for a model, which stretches the chain of accountability far away from the involved parties. Of course, the doctrine of command responsibility for negligence may assign responsibility to a commander for not preventing or punishing an action. Yet, in situations in which a commander fails to prevent a wrongful action, he or she is held responsible on behalf of subordinates because the commander is assumed to have effective control and to be able to forestall upcoming possible wrongdoings. Consequently, the doctrine of command responsibility requires a human-based perspective, a point mentioned by Sparrow (2007) and reflected upon by Heyns in referring to lethal autonomous robot systems (LARS) and questioning "whether military commanders will be in a position to understand the complex programming of LARs sufficiently well to warrant criminal liability (Heyns 2013, 15).

Moreover, the claim of Müller and Simpson that LAWS may be regulated with reference to existing civilian legal frameworks downplays that LAWS radically alters the conception of responsibility. Already by now, we are faced by ethical challenges from autonomous systems (e.g., driverless cars) in civilian contexts. There is no reason to believe that these challenges will not be intensified in armed contexts in which LAWS are considered as autonomous agents to which we intend to delegate lethal force based on the prediction that they shall be able to outperform humans. As such, the combination of agency and the idea of an engineering tolerance is a dangerous cocktail, because autonomy will be followed by the possibility for unpredictable behavior, which ought to call for the enforcement of a strong precautionary principle rather than a preferential treatment of an engineering tolerance.

> Under the precautionary principle, it is not necessary to resolve scientific uncertainty in order for preventive measures to be warranted. At this time, there is no absolute proof as to whether one or more technological improvements could eliminate the threat posed by fully autonomous weapons. [...] Today's scientific uncertainty combined with the potential threat to the civilian population, however, suffices to open the door to immediate preventive measures. (IHCR, 2013, 12)

Even in cases concerning AI-systems working in restricted domains, it may be hard to fully predict their behavior. Moreover, from a pragmatic perspective, reliance on the notion of "meaningful human control" (Article 36, 2016), as a framework for scaffolding ongoing policy discussions and negotiations about regulations of LAWS through arms control agreements, will presumably be hard to enforce in practice since regulative initiatives may lag behind technology thereby complicating enforcement of agreements.

4. Concluding Remarks

It is unwise to operate with an engineering tolerance threshold and thereby let the notion of a responsibility gap slip into the discussion of LAWS as an inevitable precondition, which, *per se*, deserves to be incorporated and taken for granted in the regulation of LAWS. Presumably, there is no guarantee that we shall be able to come up with reasonable robust autonomy impact assessments when dealing with LAWS capable of

operating in complex domains. Given these circumstances, it makes sense to place a preemptive ban on the use of LAWS rather than setting out to regulate the development and use thereof. Moreover, LAWS will be unpredictable, but so are humans too. However, contrary to humans, LAWS cannot be held morally responsible for their actions in the strong sense thereof, since this implies morality and mortality. By the same token, international humanitarian laws emphasize a clear chain of accountability as a compelling prerequisite to ensure that someone, namely *a person*, can and will be held morally responsible for causalities and war crimes. There are convincing reasons to keep it this way.

Notes

[1]. We have discussed issues related to the programming of moral machines at length in Gerdes & Øhrstrøm, 2015 and Gerdes, 2014.

2. See for instance this YouTube video, which summarizes the finals: https://www.youtube.com/watch?v=wwWHfBS9tuw.

3. Accessed July 19, 2018: https://fas.org/irp/program/collect/usroadmap2011.pdf.

4. Accessed July 19, 2018: http://www.dtic.mil/docs/citations/ADA570570.

Works Cited

Abney, K. “Autonomous Robots and the Future of Just War Theory.” Eds. F. Allhoff, N. G. Evans, and A. Henschke. *Routledge Handbook of Ethics and War Just War Theory in the 21st Century*. 2013. eBook ISBN: 9780203107164

Anscombe, E. “Modern Moral Philosophy.” *Philosophy* 33 (1958): 1-19.

Aquinas, T. “Summa Theologica.” *The Morality of War: Classical and Contemporary Readings.* Eds. L. May, E. Rovie, and S. Viner. New Jersey: Prentice Hall, 2005. 26-33.

Aristotle. *Nicomachean Ethics*. Trans. L. H. G. Greenwood. Cambridge, UK: University Press, 1909.

Arkin, R. C. “Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture.” *Technical Report GIT-GVU-07-11*. 2007. Accessed, March 5, 2018. <http://www.cc.gatech.edu/ai/robot-lab/online-publications/formalizationv35.pdf>.

---. “Ethical Robots in Warfare.” *IEE Technology and Society Magazine* 28.1 (2009): 30-33.

“Article 36 Key Elements of Meaningful Human Control.” 2016. Accessed, March 5, 2018. <http://www.article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf>.

Bostrom, N., & Yudkowsky, E. “The Ethics of Artificial Intelligence.” Eds. W. Ramsey and K. Frankish. *The Cambridge Handbook of Artificial Intelligence*. Cambridge, UK: University Press, 2014.

Champagne, M., & Tonkens, R. “Bridging the Responsibility Gap in Automated Warfare.” *Philos. Technol.* 2013. doi:10.1007/s13347-013-0138-3

Coeckelbergh, M. “From Killer Machines to Doctrines and Swarms, or Why Ethics of Military Robotics Is Not (Necessarily) About Robots.” *Philos. Technol.* 2011. doi:10.1007/s13347-011-0019-6

Cummings, M. L. “Integrating Ethics in Design Through the Value-sensitive Design Approach.” *Science and Engineering Ethics* 12 (2006): 701-15.

Dennett, D. C. “Cognitive Wheels: The Frame Problem of AI.” Ed. Z. W. Pylyshyn. *The Robot’s Dilemma: The Frame Problem in Artificial Intelligence*. NJ: Ablex, 1988. 41-65.

Floridi, L. C., & Sanders, J. W. “On the Morality of Artificial Agents.” *Minds and Machine* 14 (2004): 349-79.

Ford, K. M. (Ed.). *The Robot’s Dilemma Revisited.* NJ: Ablex, 1996.

Gerdes, A. “Ethical Issues Concerning Lethal Autonomous Robots in Warfare.” Red. J. Seibt, R. Hakli, and M. Nørskov. *Sociable Robots and the Future of Social Relations.* IOS Press, s. 277-89. *Frontiers in Artificial Intelligence and Applications* 273 (2014). doi:10.3233/978-1-61499-480-0-277

Gerdes, A., & Øhrstrøm, P. “Issues in Robot Ethics Seen Through the Lens of a Moral Turing Test.” *Journal of Information, Communication and Ethics in Society* 13.2 (2015): 98-109. doi:10.1108/JICES-09-2014-0038

Grut, C. “The Challenge of Autonomous Lethal Robotics to International Humanitarian Law.” *Journal of Conflict & Security Law*. Oxford: Oxford University Press, 2013. doi:10.1093/jcsl/krt002

Heyns, C. “Report of the Special Rapporteur on Extrajudicial Summary or Arbitrary Executions on Lethal Autonomous Robot Systems.” *A/HCR/23/47*. 2013. Accessed, March 5, 2018. <http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf>.

IHRC. “The Need for New Law to Ban Fully Autonomous Weapons: Memorandum to Convention on Conventional Weapons Delegates. Human Right Watch and Harvard Law School’s International Human Right Clinic.” Nov. 2013. Accessed, March 5, 2018. <https://www.hrw.org/sites/default/files/supporting_resources/11.2013_memo_to_ccw_delegates_fully_autonomous_weapons.pdf>.

LeCun, Y., Bengio, Y., & Hinton, G. “Deep Learning.” *Nature* 521 (28 May, 2015): 436-44. doi:10.10381/nature14539

Lin, P., Bekey, G., & Abney, K. “Autonomous Military Robotics: Risk, Ethics, and Design.” *CALPOLY*. 2008. Accessed, March 5, 2018. <http://ethics.calpoly.edu/ONR_report.pdf>.

Lokhorst, G. J., & van den Hoven, J. “Responsibility for Military Robots.” Eds. P. Lin, K. Abney, and G. A. Bekey. *Robot Ethics—The Ethical and Social Implications of Robotics*. Cambridge, Massachusetts: MIT Press, 2012. 145-57.

Müller, V., & Simpson, T. “Killer Robots: Regulate, Don’t Ban.” *BSG Policy Memo*. Nov. 2014. Accessed, March 5, 2018. <http://www.bsg.ox.ac.uk/sites/www.bsg.ox.ac.uk/files/2014-Killer-Robots-Policy-Paper.pdf>.

---. “Autonomous Killer Robots Are Probably Good News.” Eds. E. Di Nuccy and F. Santonni de Sio. *Drones and Responsibility: Legal, Philosophical and Socio-Technical Perspectives on the Use of Remotely Controlled Weapons.* London: Ashgate, 2016.

Quinn, W. S. “Actions, Intentions, and Consequences: The Doctrine of Double Effect.” *Philosophy & Public Affairs* 18.4 (Autumn, 1989): 334-51. Princeton University Press.

Sharkey, N. E. “The Inevitability of Autonomous Robot Warfare.” *International Review of the Red Cross* 94.886 (2012): 787-99.

Singer, P. *Wired for War—The Robotics Revolution and Conflict in the 21st Century*. New York Times Bestseller, 2009.

Sparrow, R. “Killer Robots.” *Journal of Applied Philosophy* 24.1 (2007): 62-77.

Tucker, P. “Now the Military Is Going to Build Robots That Have Morals. *Defense One*. May 13, 2014. Accessed, March 5, 2018. <http://www.defenseone.com/technology/2014/05/now-the-military-going-build-robots-have-morals/84325/>.

Verbeek, P. P. “Ambient Intelligence and Persuasive Technology—The Blurring Boundaries Between Human and Technology.” *Nanoethics* 3 (2009): 231-42. doi:10.1007/s11569-009-0077-8

---. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: The University of Chicago Press, 2011.

Wallach, W. *A DangerousMaster: How to Keep Technology From Slipping Beyond Our Control*. New York: Basic Books, Jun. 2015.

Wallach, W., & Allen, C. *Moral Machines—Teaching Robots Right From Wrong*. New York: Oxford University Press, 2009.

Walzer, M. *Just and Unjust Wars: A Moral Argument With Historical Illustrations*. New York: Basic Books, 1977.