Peace Review

A Journal of Social Justice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cper20

Algorithms, AI, and Ethics of War

John R. Emery

To cite this article: John R. Emery (2021) Algorithms, AI, and Ethics of War, Peace Review, 33:2, 205-212, DOI: 10.1080/10402659.2021.1998749

To link to this article: https://doi.org/10.1080/10402659.2021.1998749

Published online: 03 Jan 2022.

Peace Review: A Journal of Social Justice, 33:205–212, 2021
Copyright © Taylor & Francis Group, LLC
ISSN: 1040-2659 print / 1469-9982 online
DOI: 10.1080/10402659.2021.1998749

Algorithms, AI, and Ethics of War

John R. Emery

In the episode, "A Taste of Armageddon," of the science fiction television show *Star Trek: The Original Series,* the crew of the Enterprise visit a pair of planets that have been engaging in computer-simulated war for over 500 years. To prevent the destruction of their societies, the two planets signed a treaty whereby the wars would be fought fictitiously with a computer-generated outcome, yet the casualties would be real, with the tabulated victims voluntarily reporting to be killed. Captain Kirk destroys the war simulation computers and is condemned because without the computers to fight the war, real war would be inevitable. Yet, the reason the war had gone on for so long was precisely because the simulation had insulated both societies from the horrors of war, thus, they had little reason to end it.

While based on science fiction, the threat of an AI-enabled battlefield of the future raises ethical and practical concerns about the horrors of war. The logic driving countries to adopt lethal autonomous weapons systems (LAWS) is indeed seductive. Humans are fallible, emotional, and irrational; we can protect both our soldiers and their civilians via LAWS. Thus, this line of reasoning constructs LAWS as inherently rational, predictable, and even ethical. Killer robots, despite their name, will actually save lives.

Such logic, however, is folly. There are a number of potential pitfalls of an AI-enabled warfare that focuses on perfecting the means of war, while neglecting the ends of war. Like in *Star Trek*, the allure of riskless war is compelling, and yet it has real consequences for those who inevitably end up killed, maimed, and left displaced. In what follows, I argue that there are serious ethical issues with the prospect of LAWS that cannot be solved by advanced technology. Ethics cannot be pre-programmed to apply across context or conflict, and meaningful human control neglects the ways in which the automation bias shapes human-machine interaction in decision-making.

The concept of meaningful human control, especially in lethal decision-making has been proposed by both military entities and NGOs alike as a necessity. The concept of a human-in-the-loop, or human-on-the-loop (either to positively push the button to use lethal force or to actively stop an AI-generated weapons deployment, respectively) however, neglects the psychology of human-machine interactions and the speed at which wars of the future may take place. The concept of meaningful human control is easy because it holds a human accountable for inevitable killing of innocents with laws, it smooths over the difficult questions of LAWS in favor of a legal/ethical paradigm that we are familiar with: rights, duties, and accountability. And yet, this "solution" to the problem of LAWS neglects the ways in which human programmers shape the decisions that AI makes, and how humans interact with AI-generated systems. This is where the issue of trust in AI and the automation bias come to the forefront. What happens when military commanders have too much trust in AI systems? Outside of the military there is overwhelming empirical evidence on how humans interact with advanced technology in what we call an automation bias. This automation bias is a strong tendency of humans to defer to automated technology; such that humans assume positive design intent even when the tech is malfunctioning, and this holds true even when presented evidence of a system's failure. More troubling, automation bias occurs in both naive and expert participants, it cannot be prevented by training or instructions, and can affect decision making in individuals as well as in teams.

The assumption that LAWS would be at least as good or better than humans at life and death decisions relies on a disconnect between how these algorithms function and how they are utilized. Human judgment is never eliminated from the equation, but it transfers judgment from democratically accountable military practitioners to computer programmers. As Frank Pasquale notes in his book *The Black Box Society*, in the financial sector, the attraction of the black box algorithm is that it actually promotes an automation bias. There is an assumption that a machine-driven, software-enabled system is going to offer better results than human judgment. And when the stakes are high enough, automation bias can degenerate into wishful thinking or worse: opportunistic misuse of models to validate existing practices. This was especially true with predictive policing systems across the U.S. that was plagued with both horribly biased data, and the tendency of police to disregard or not search for contradictory information in light of a computer-generated solution that they desired to accept as correct. While trust may currently be hard to come by with military/AI interactions, there is also a danger in that an automation bias may take over to where AI is used to justify existing practices rather than modify behavior, giving an aura of objective legitimacy over legally or morally questionable targeting practices.

This was at least partially an issue when it came to collateral damage estimation algorithms, which never took into account empirical data of actual numbers of civilians killed to retrain algorithms, and were discursively used to tick the box of ethics of due care in war throughout the War on Terror. In my previous work "Probabilities Toward Death," I examined the over two and half decades of human-computer interaction in the U.S. Air Force with these algorithms and found that there was a tendency to utilize them to defer accountability for the killing of innocents. My fear with LAWS is that we are assuming that they will be better than human judgment, neglecting the ways in which AI is both shaped by, and shapes, decision-making in war.

For example, in the Iraq War, the USAF had a casualty cutoff value of over 30 civilian casualties, when the collateral damage estimation algorithm predicted over 30 the commander would need to seek higher up approval for the strike. It was not that the strike would not occur, only that they needed either the highest commander, Secretary of Defense or President's office to approve. Thus, some field commanders would adjust the fuse delay on a bomb to get the software to predict 29 casualties thereby avoiding higher-up scrutiny. What is interesting here is that this arbitrary ceiling shaped the way in which this algorithm was selectively utilized, and it was divorced from the actual empirical outcomes. It predicted 0 civilian casualties and yet we had 50, we ran the algorithm and therefore we exercised due care. Thus, the concept of meaningful human control quickly fell out even when rudimentary algorithms were introduced as perhaps a pragmatic faith was placed in them if only to defer accountability for the killing of innocents. Rather than their proposed and purpose of protecting civilian life.

The second issue with meaningful human control is that the speed at which a LAWS-enabled battlefield can operated can become the next useless measure against which policy-makers judge military success or preparedness. AI can simultaneously make decisions on a time scale incomprehensible to humans and will enable rapid decisions across multiple domains and multiple levels of war that humans cannot outpace. Thus, if humans cannot by definition keep up with the pace of battle at which AI can make decisions, how then could they be expected to exercise meaningful human control over such processes?

This is especially worrisome, because while speed can be beneficial in certain circumstances, it can also be a detriment in others. Especially in crisis situations where the concern is inadvertent escalation where we tend to rely on fast—System 1—thinking that relies heavily on bias rather than slow—System 2—thinking that relies on deliberate consciousness that can override the knee-jerk, time-crunched decisions of System 1. This is drawn from Danial Kahneman in, *Thinking Fast and Slow*, and applied to issues such as nuclear crisis escalation, where essentially the goal is to slow down the decision-making process to encourage System 2 thinking rather than seeking fast cognitive closure. Speed in and of itself is not a valuable goal to strive toward, as it can be as much of a liability as asset depending on context. Thus, the prospect of a hyper-speed conflict with swarm drones or other LAWS systems, could quickly and inadvertently escalate a situation beyond the desired outcome of decision-makers.

Moreover, a focus on efficiency alone misses the necessity of meaningful inefficiencies to be built into the system to allow for things such as ethical deliberation, democratic debate, and re-thinking priorities or strategies. The emphasis on the speed of the future of war often seems to treat military endeavors as if they take place outside of civilian control over the military, not within a military that always acts in the name of the U.S. public. Within battle as well, a focus on accelerated efficient ordering of warfare neglects the fact that military necessity and proportionality are in a constant state of flux depending on the political objectives that we are trying to achieve at that particular time/context. The laws of war cannot be pre-programmed to apply across conflict/context as a universal/timeless rule to follow. It is a deliberative process of weighing the potential outcomes and projected gains within an uncertain environment of war that does not lend itself readily to quantification.

Ethics of war rests on meaningful inefficiencies because such probability-based computation cannot dictate values. What is at stake in these techno-practices of war is nothing less than the erosion of effective constraints on the use of lethal force because the techno-rationalization of risk assessment has supplanted genuine ethical deliberation about which strikes constitute necessary and proportionate responses. Thus, the ethics of war are about deliberation and weighing the pros and cons of political and military objectives against probable (yet unknown outcomes) and feeling the weight of those decisions. When we integrate LAWS and AI into that decision-making process, it becomes a self-justificatory system that assumes out these meaningful inefficiencies, the essence of ethical and democratic deliberation. Ultimately, this stems from the assumption that human fallibility can be eliminated, and a science of war can be created with AI-systems, that render the horrors of war rational, controllable, and more predictable than previous conflicts.

Perhaps one of the most fundamental arguments against LAWS is the faith that those employing LAWS have in the technology that it somehow describes the world rather than constructs it based on pre-programmed assumptions. The issue here is that there is an assumption of some objective criterion of superior or the “best decision.” There are better and worse decisions/outcomes, but that is always linked to the political objectives one is trying to achieve in war. Many argue by abstracted example that humans should follow the recommendations of the weapon-target model if it produces the “best” decision 95 percent of the time if a human only finds the “best” decision 80 percent of the time. AI constructs the world as much as it does explain it. That is dependent on the assumptions of what is written into the algorithms for what counts as success or failure. What counts as the “best” decision will vary based upon the deep context of immensely complex situations that often do not map nearly onto other scenarios.

Discussion of these types of quantifiable confidence intervals necessitate a certain epistemological leap of faith that ultimately gives human interpreters of AI a false sense of certainty in the external validity of these numbers as better than human judgment. As if it were somehow determined outside of human judgment at the programming stage. The desire for a number, no matter how unscientifically achieved has the danger of steering military practitioners and policy-makers away from viable alternatives outside of the conceived outputs of the AI model.

Programmers themselves do not know why AI makes the decisions that it does, because of the nature of AI. Google DeepMind’s AlphaGo that defeated the world champion in what is arguably the most difficult game in the world, Go, best demonstrates how AI works. First the AI had the rules of the game programmed, and played a number of Go players to where it became a decent Go player. What happened next was that AlphaGo played against itself in millions of games, until it learned the best of all possible strategies. AlphaGo made a move that was completely unexpected, however, as no human had never been made in a Go game; it was entirely unpredictable, yet later deemed as brilliant. What is now deemed “Move 37” (as this was the 37th move of the second game against world champion Lee Sedol) is a major problem for LAWS, because AI can make wild and unpredictable moves in crisis situations.

Thus, when it played the world champion, it was making moves no human had ever made in the game, and DeepMind could not explain why it would do that because no human could track all the millions of iterations it played; hence, the essence of AI is it is always already beyond meaningful human control in the first instance. What is important for AI in warfare, is that there are not strict “rules of the game” like in Go or chess, when the enemy doesn’t play by the rules, you can’t hit the reset button and try again.

The world is complex; war is an experiment in catastrophe where the complexities of the social world are amplified exponentially. My hesitation about AI rests in the fact that once Pandora’s box is opened, we cannot know why it makes the decisions it does, which has enormous consequences when we give AI the power to take human life. I will end with why LAWS challenge these ethical frameworks that help to reduce the horrors and uncertainty of war.

LAWS are most dangerous, not because they kill from a distance or desensitize us to the horrors of war, but they give decision-makers a delusion that war can be controlled and is rendered scientific and predictable. In fact, AI-enabled warfare may do the opposite. As retired Colonel Andrew Bacevich aptly notes, “War remains today as it has always been–elusive, untamed, costly, difficult to control, fraught with surprise, and sure to give rise to unexpected consequences.” Indeed, the enduring nature of war is that it is an experiment in catastrophe, yet we strive to construct a science of warfare and program away algorithmically the ethical-political dilemmas of killing in war. Hence, the aura of objectivity and neutrality that techno-ethics purports to offer decision-makers not only allows them to bury the ethical dilemmas of practical judgment into the algorithmic code, it simultaneously removes them one causal step from the act of killing. In the end, there is always uncertainty in warfare.

The risk I see for LAWS based on the empirical evidence from collateral damage estimation algorithms, is that technology offers military decision-makers an alluring appeal to technological fixes to ethical-political dilemmas of killing in war. The assumption that LAWS would be at least as good or better than humans at life and death decisions relies on a disconnect between how these algorithms function and how they are utilized. Human judgment is never eliminated from the equation, but it transfers judgment from military practitioners to computer programmers. LAWS opens up what Elke Schwarz has defined as moral vacuums as it reshapes our capacity to think ethically:

> A moral vacuum opens when certain parameters of harm are no one’s responsibility; when the decision that harm is permissible has been determined through technological means. This moment is, paradoxically, also the very moment of moral responsibility. In other words, the moral vacuum exists exactly in the moment when neither law nor existing moral guides have adequate reach. It is in this moment where responsibility resides.

The appeals of LAWS and a more technological battlefield are not only a new blind faith, but may actually enable what it seeks to constrain: that is, making war more horrible and unpredictable than previously imagined. Technology does not inherently make war a more ethical space. Instead, LAWS function to replace difficult ethical-political decision-making in war with a fantasy of control over the uncertainties of conflict, while simultaneously absolving decision-makers of responsibility for killing by removing them one causal step further from the act of killing. Despite the rhetoric of "just war" that often accompanies praise of technological advances in targeting and killing, virtue ethics and practical judgment has been abandoned and replaced by a predetermined utilitarian calculation conceived as objective and neutral techno-innovation in the eyes of practitioners. Such a (r)evolution in framing war as a technical problem to be resolved speaks to a wider drive of quantifying the uncertainties of war into a numerically calculable risk assessments that our capacity to make ethical decision in war.

What is at stake in these techno-practices of war is the erosion of effective constraints on the use of lethal force because this techno-rationalization of risk assessment has supplanted genuine ethical deliberation about the consequences of contemporary conflict. Moral vacuums created by LAWS ultimately eliminate meaningful inefficiencies of ethical and political deliberation in favor of speed and riskless warfare. Yet, the horrors of war have not been eliminated nor has it become more scientific; instead we are developing methods of killing and maiming that further remove us from the act of killing. Like Captain Kirk, we must disable the war-machine that is driving the debate on LAWS and recognize its inherent limitations before it is too late.

RECOMMENDED READINGS

Amoore, Louise. 2014. "Security and the Incalculable." *Security Dialogue* 45(5): 423–439. doi:10.1177/0967010614539719.

Crawford, Neta C. 2013. *Accountability for Killing: Moral Responsibility for Collateral Damage in America's Post-9/11 Wars*. Oxford: Oxford University Press.

Elish, Madeline Clare. 2019. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction." *Engaging Science, Technology, and Society* 5: 40–60.

Emery, John R. 2020. "Probabilities toward Death: Bugsplat, Algorithmic Assassinations, and Ethics of Due Care." *Critical Military Studies* doi:10.1080/23337486.2020.1809251.

Emery, John R. 2020. "Historical and Contemporary Reflections on Lethal Autonomous Weapons Systems" *E-International Relations*. April 15, 2020. https://www.e-ir.info/2020/04/15/historical-and-contemporary-reflections-on-lethal-autonomous-weapons-systems/.

Kahneman, Daniel. 2013. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux

Pasquale, Frank. 2016. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Renic, Neil C. 2018. "Justified Killing in an Age of Radically Asymmetric Warfare." *European Journal of International Relations* 25(2): 408–430. doi:10.1177/1354066118786776.

Roff, Heather M. 2014. "The Strategic Robot Problem: Lethal Autonomous Weapons in War." *Journal of Military Ethics* 13(3): 211–227.

Schwarz, Elke. 2018. "Technology and Moral Vacuums in Just War Theorising." *Journal of International Political Theory* 14(3): 280–298.

Schwarz, Elke. 2018. *Death Machines: The Ethics of Violent Technologies*. Manchester: Manchester University Press.

Zehfuss, Maja. 2011. "Targeting: Precision and the Production of Ethics." *European Journal of International Relations* 17(3): 543–566. doi:10.1177/1354066110373559.

Zehfuss, Maja. 2018. *War and the Politics of Ethics*. Oxford: Oxford University Press.

John R. Emery is an assistant professor of international security at the university of Oklahoma in the department of international and area Studies. His research focuses on issues of security, technology, and ethics of war in international relations. Email: john.emery@ou.edu