Reframing Autonomous Weapons Systems

Autonomous systems designed to cause physical harm have additional ethical dimensions as compared to both traditional weapons and autonomous systems not designed to cause harm. Multi-year discussions on international legal agreements around autonomous systems in the context of armed conflict are occurring at the United Nations (UN), but professional ethics about such systems can and should have ethical standards covering a broad array of issues arising from the automated targeting and firing of weapons.

Broadly, we recommend that technical organizations promote a number of measures to help ensure that there is meaningful human control of weapons systems:

- That automated weapons have audit trails to help guarantee accountability and control.
- That adaptive and learning systems can explain their reasoning and decisions to human operators in transparent and understandable ways.
- That there be responsible human operators of autonomous systems who are clearly identifiable.
- That the behavior of autonomous functions should be predictable to their operators.
- That those creating these technologies understand the implications of their work.
- That professional ethical codes are developed to appropriately address the development of autonomous systems and autonomous systems intended to cause harm.

Specifically, we would like to ensure that stakeholders are working with sensible and comprehensive shared definitions, particularly for key concepts relevant to autonomous weapons systems (AWS). Designers should always ensure their designs meet the standards of international humanitarian law, international human rights law, and any treaties or domestic law of their particular countries, as well as any applicable engineering standards,

Reframing Autonomous Weapons Systems

military requirements, and governmental regulations. We recommend designers not only take stands to ensure meaningful human control, but be proactive about providing quality situational awareness to operators and commanders using those systems. Professional ethical codes should be informed by not only the law, but an understanding of both local- and global-level ramifications of the products and solutions developed. This should include thinking through the intended use or likely abuse that can be expected by users of AWS.

While the primary focus of this document is with kinetic AWS that cause physical harm, it is recognized that many of these concerns and principles may also apply to cyber-weapons. This is, of course, also pertinent to cyber-weapons that have kinetic effects, such as those that destroy civilian infrastructures or turn civilian objects, vehicles, or infrastructure into kinetic weapons.

Additionally, society must be aware of the variety of political and security threats posed by AWS. Miniaturized AWS will pose additional threats because they are small, insidious, or obfuscated, and may therefore be non-attributable to the deploying entity. Depending upon payload or weapons (such as chemical, biological, or nuclear weapons), these may autonomously deploy weapons of mass destruction (WMD), or themselves constitute a new form of WMD. Additional ethical recommendations are needed to prevent the development of systems having these dangerous properties.

- Issues 1–3 raise general high-level questions regarding the definition of AWS and their relation to existing law and ethics.
- Issues 4–10 raise socio-political concerns over the likely uses and effects of AWS development and use.
- Issue 11 raises engineering concerns over the specific challenges posed by autonomous systems capable of targeting and deploying weapons.

Disclaimer: *While we have provided recommendations in this document, it should be understood these do not represent a position or the views of IEEE but the informed opinions of Committee members providing insights designed to provide expert directional guidance regarding A/IS. In no event shall IEEE or IEEE-SA Industry Connections Activity Members be liable for any errors or omissions, direct or otherwise, however caused, arising in any way out of the use of this work, regardless of whether such damage was foreseeable.*

Reframing Autonomous Weapons Systems

Issue 1:

Confusions about definitions regarding important concepts in artificial intelligence (AI), autonomous systems (AS), and autonomous weapons systems (AWS) stymie more substantive discussions about crucial issues.

Background

The potential for confusion about AWS definitions is not just an academic concern. The lack of clear definitions regarding what constitutes AWS is often cited as a reason for not proceeding toward any kind of international governance over autonomous weapons. As this is both a humanitarian issue and an issue of geopolitical stability, the focus in this area needs to be on how the weapons are controlled by humans rather than about the weapons' technology *per se*.

The term *autonomy* is important for understanding debates about AWS; yet there may be disputes — about what the term means and whether what the definition identifies is technically possible today. This prevents progress in developing appropriate policies to regulate AWS design, manufacture, and deployment. Consistent and standardized definitions are needed to enable effective discussions of AWS, but they should be general enough to enable flexibility to ensure that those definitions do not become quickly technologically outdated.

Moreover, the phrases "human in the loop" and "human on the loop" also lack clarity and only contribute further confusion. Depending upon what one means, "in the loop" or "on the loop" means different things to different people. It could be used to describe the command chain that authorizes weapon release, where the commands flow down to a human and a weapon system to take specific actions. Yet, there are micro-level decisions where a human operator may have an opportunity to question the command. What often matters is the time delay between the fielding of an autonomous system, the decision to engage a weapon against a target, and the impact time.

Contrarily, "in the loop" obscures another temporal question: that whether in these scenarios clearance to fire at a target entails an authorization to prosecute that target indefinitely, or whether there are necessarily predetermined limits on the amount of time or ordinance each clearance provides. Central to this issue is how long a target that has been designated and verified by an authorized human in a given situational context remains a legitimate target.

This notion of autonomy can be applied separately to each of the many functions of a weapons system; thus, an automatic weapons system could be autonomous in searching for targets, but not in choosing which ones to attack, or vice versa. It may or may not be given autonomy to fire in self-defense when the program determines that the platform is under attack, and so on. Within each of these categories, there are also many intermediate gradations in the way that human and machine decision-making may be coupled.

Reframing Autonomous Weapons Systems

Candidate Recommendations

The term *autonomy* in the context of AWS should be understood and used in the restricted sense of the delegation of decision-making capabilities to a machine. Since different functions within AWS may be delegated to varying extents, and the consequences of such delegation depend on the ability of human operators to forestall negative consequences via the decisions over which they retain effective control, it is important to be precise about the control of specific functions delegated to a given system, as well as the ways in which control over those functions are shared between human operators and AWS.

We support the working definition of AWS offered by the International Committee of the Red Cross (ICRC) and propose that it be adopted as the working definition of AWS for the further development and discussion of ethical standards and guidelines for engineers. The ICRC defines an AWS as: "any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention."

Further Resources

- Dworkin, G. *The Theory and Practice of Autonomy.* Cambridge, U.K.: Cambridge University Press, 1988.
- Frankfurt, H. G. "Freedom of the Will and the Concept of a Person," in *The Importance of What We Care About*, Cambridge, U.K.: Cambridge University Press, 1987.
- DoD Defense Science Board, The Role of Autonomy in DoD Systems, Task Force Report. July 2012, 48.
- DoD Defense Science Board, Summer Study on Autonomy. June 2016.
- Young, R. *Autonomy: Beyond Negative and Positive Liberty.* New York: St. Martin's Press, 1986.
- Society of Automotive Engineers. J3016, Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. SAE International, 2014.
- Roff, H. M. "An Ontology of Autonomy: Autonomy in Weapons Systems," in *The Ethics of Autonomous Weapons*, edited by C. Finkelstein, D. MacIntosh, and J. D. Ohlin. Cambridge, U.K.: Oxford University Press, forthcoming.
- Sharkey, N. "Towards a Principle for the Human Supervisory Control of Robot Weapons." *Politica and Società* 2 (2014): 305–324.
- U.K. Ministry of Defence. UK Joint Doctrine Note (JDN) 3/10, "Unmanned Aircraft Systems: Terminology, Definitions and Classification." May 2010.
- U.K. Ministry of Defence. UK Joint Doctrine Note (JDN) 2/11, "The UK Approach to Unmanned Aircraft Systems." March 2011.
- United Nations Institute for Disarmament Research (UNIDIR). "Framing Discussions on the Weaponization of Increasingly Autonomous Technologies." 2014
- International Committee of the Red Cross (ICRC). "Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons." September 1, 2016.

Reframing Autonomous Weapons Systems

Issue 2:

The addition of automated targeting and firing functions to an existing weapon system, or the integration of components with such functionality, or system upgrades that impact targeting and automated weapon release should be considered for review under Article 36 of Additional Protocol I of the Geneva Conventions.

Background

According to Article 36 of Additional Protocol I to the Geneva Conventions (1977), "In the study, development, acquisition or adoption of a new weapon, means or methods of warfare," weapon systems must be internally reviewed for compliance with international humanitarian law (IHL). Alterations to the critical functions or targeting and weapons release of an already-reviewed weapons systems should be considered for review, and any system automating those functions should be reviewed to ensure meaningful human control.

International human rights law (IHRL) also guarantees, by way of international and bilateral treaties, rights to life, human dignity, fair trial, and further positive and negative human rights. Society and engineers must consider the ways in which these rights may be threatened by the deployment and/or use of AWS, during armed conflict, policing, or other security operations.

There are situational and operational limitations of all engineered systems, and complete knowledge is not something that can be expected or required. However, there must be a multi-level effort to:

- Evaluate the conformity of a system to the law
- Evaluate its reliability and applicability for a given mission
- Evaluate its ability to conform to rules of engagement

Further, key decision makers need to understand the engineering constraints and limitations of weapons systems with high degrees of autonomy.

Candidate Recommendations

- All engineering work should conform to the requirements of international law, including both IHL and IHRL, as well as national and local laws. While this is not the primary responsibility of an individual engineer, there ought to be opportunities for engineers to learn about their obligations, their responsibilities with respect to AWS, as well as keeping their employing agencies accountable.
- Meaningful human control over the critical functions in weapons systems can help ensure that weapons can be used in conformity with the law in each instance. It is

Reframing Autonomous Weapons Systems

also necessary for all stakeholders to consider design and implement accountability measures to help ensure all weapons are used in conformity with the law.

- Engineering constraints should be clearly identified, defined, and communicated to Article 36 weapons reviewers, to operators in their training for a system, and to military commanders and their legal counsel charged with specifying the rules of engagement.
- All those with responsibilities for weapon systems should ensure that Article 36 reviews will be held and provide all evidence needed at them. This should include any data which will lead to restrictions on their use, which will also be needed for Article 36 reviews and for military staff to set rules of engagement for the weapon system's use.
- There should be greater engineering input into the weapons reviews, and greater communication between engineers and lawyers in the weapons review process to ensure meaningful human control over weapons.

Further Resources

- International Committee of the Red Cross (ICRC). "Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons." September 1, 2016.

Issue 3:

Engineering work should conform to individual and professional organization codes of ethics and conduct. However, existing codes of ethics may fail to properly address ethical responsibility for autonomous systems, or clarify ethical obligations of engineers with respect to AWS. Professional organizations should undertake reviews and possible revisions or extensions of their codes of ethics with respect to AWS.

Background

- The ethical requirements for engineering have an independent basis from the law, although they are hopefully aligned with written laws and written codes of professional ethics. Where agreed upon, ethical principles are not reflected in written laws and ethical codes, individuals and organizations should strive to correct those gaps.
- Ethical requirements upon engineers designing autonomous weapon systems may go beyond the requirements of meeting local, national, and international laws.

Reframing Autonomous Weapons Systems

Many professional organizations have codes of conduct intended to align individuals' behaviors toward particular values. However, they seldom sufficiently address members' behaviors in contributing toward particular artifacts, such as creating technological innovations deemed threatening to humanity, especially when those innovations have significant probabilities of costly outcomes to people and society. Foremost among these in our view are technologies related to the design, development, and engineering of AWS.

Organizations such as the IEEE, the Association for Computing Machinery (ACM), the Association for the Advancement of Artificial Intelligence (AAAI), the UK Royal Academy of Engineering, the Engineering Council, Engineers Canada, and the Japanese Society for Artificial Intelligence (JSAI) have developed codes of ethics. Some of these groups are currently reviewing those codes in light of current and future developments in autonomous systems and AI.

While national laws may differ on what constitutes responsibility or liability for the design of a weapon system, given the level of complicity or the causal contribution to the development of a technology, ethics looks for lines of moral responsibility. Determining whether an individual is morally responsible requires understanding the organizations in which they work and to establish relevant facts in relation to the individual's acts and intentions.

Candidate Recommendations

Codes of conduct should be extended to govern a member's choice to create or contribute to the creation of technological innovations that are deemed threatening to humanity. Such technologies carry with them a significant probability of costly outcomes to people and society. When codes of conduct are directed toward ensuring positive benefits or outcomes for humanity, organizations should ensure that members do not create technologies that undermine or negate such benefits. In cases where created technologies or artifacts fail to embody or conflict with the values espoused in a code of conduct, it is imperative that professional organizations extend their codes of conduct to govern these instances so members have established recourse to address their individual concerns. Codes of conduct should also more broadly ensure that the artifacts and agents offered into the world by members actively reflect the professional organization's standards of professional ethics.

Professional organizations need to have resources for their members to make inquiries concerning whether a member's work may contravene (IHL) or (IHRL).

How one determines the line between ethical and unethical work on AWS requires that one address whether the development, design, production, and use of the system under consideration is itself ethical. It is incumbent upon a member to engage in reflective judgment to consider whether or not his or her contribution will enable or give rise to AWS and their use cases. Members must be aware

Reframing Autonomous Weapons Systems

of the rapid, dynamic, and often escalatory natures of interactions between near-peer geopolitical adversaries or rivals. It is also incumbent upon members of a relevant technical organization to take all reasonable measures to inform themselves of the funding streams, the intended use or purpose of a technology, and the foreseeable misuse of their technology when their contribution is toward AWS in whole or in part. If their contribution to a system is foreseeably and knowingly to aid in human-aided decisions — that is, as part of a weapon system that is under meaningful human control — this may act as a justification for their research.

Further Resources

- Kvalnes, Ø. "Loophole Ethics," in *Moral Reasoning at Work: Rethinking Ethics in Organizations*, 55–61. Palgrave Macmillan U.K., 2015.
- Noorman, M. "Computing and Moral Responsibility," *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta , Summer 2014 Edition.
- Hennessey, M. "Clearpath Robotics Takes Stance Against 'Killer Robots'." Clearpath Robotics, 2014.
- "Autonomous Weapons: An Open Letter from AI & Robotics Researchers." Future of Life Institute, 2015.
- Noorman, M. "Computing and Moral Responsibility," in *The Stanford Encyclopedia of Philosophy* (Summer 2014 Edition), edited by Edward N. Zalta.
- "Engineers Canada Code of Ethics," 2017.
- The Japanese Society for Artificial Intelligence Ethical Guidelines, 2017
- Engineering Council and Royal Academy of Engineering, Statement of Ethical Principles for the Engineering Profession.

Issue 4:

The development of AWS by states is likely to cause geopolitical instability and could lead to arms races.

Background

The widespread adoption of AWS by nation states could present a unique risk to the stability of international security. Because of the advantages of either countering an adversary through concomitant adoption of arms or being the first or prime mover is an offset advantage, the pursuit of AWS is likely to spur an international arms race. Evidence of states seeking greater adoption of artificial intelligence and quantum computing for security purposes already exists. The deployment of machine learning and other artificial intelligence applications on weapons systems is not only occurring, but will continue to advance. Thus it is important to look to previous scholarship on arms race dynamics to be informed about the first- and second-order effects of these races, such as the escalatory effects, arms development, decreasing international stability, and arms proliferation.

Reframing Autonomous Weapons Systems

Candidate Recommendations

Autonomous weapons designers should support the considerations of the United Nations to adopt a protocol to ensure meaningful human control over AWS under the Convention on Certain Conventional Weapons (CCW) treaty, or other similar effort by other international bodies seeking a binding international treaty.

It is unethical to design, develop, or engineer AWS without ensuring that they remain reliably subject to meaningful human control. Systems created to act outside of the boundaries of "appropriate human judgment," "effective human control," or "meaningful human control," violate fundamental human rights and undermine legal accountability for weapons use. Various scenarios for maintaining meaningful human control over weapons with autonomous functions should be further investigated for best practices by a joint workshop of stakeholders and concerned parties (including, but not limited to, engineers, international humanitarian organizations, and militaries), and that those best practices be promoted by professional organizations as well as by international law.

Further Resources

- Scharre, P., and K. Sayler. "Autonomous Weapons and Human Control" (poster). Center for a New American Security, April 2016.
- International Committee for Robot Arms Control. "LAWS: Ten Problems for Global Security" (leaflet). April 10, 2015.
- Roff, H. M., and R. Moyes. "Meaningful Human Control, Artifical Intelligence and Autonomous Weapons." Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons, April 2016.
- United Nations Institute for Disarmament Research (UNIDIR). "The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward." 2014.

Issue 5:

The automated reactions of an AWS could result in the initiation or escalation of conflicts outside of decisions by political and military leadership. AWS that engage with other AWS could escalate a conflict rapidly, before humans are able to intervene.

Background

One of the main advantages cited regarding autonomous weapons is that they can make decisions faster than humans, enabling rapid defensive and offensive actions. When opposing autonomous weapons interact with each other, conflict might escalate without explicit human military or political decisions, and escalate more quickly than humans on either side will be able to understand or act.

Reframing Autonomous Weapons Systems

Candidate Recommendations

- Consider ways of limiting potential harm from automated weapons. For example: limited magazines, munitions, or maximum numbers of platforms in collaborative teams.
- Explore other technological means for limiting escalation, for example, "circuit breakers," as well as features that can support confidence-building measures between adversaries. All such solution options ought to precede the design, development, deployment, and use of weapons systems with automated targeting and firing functions.
- Perform further research on how to temper such dynamics when designing these systems.

Further Resources

- Scharre, P. "Autonomous Weapons and Operational Risk." Washington, DC: Center for New American Security, February, 2016.

Issue 6:

There are multiple ways in which accountability for the actions of AWS can be compromised.

Background

Weapons may not have transparency, auditability, verification, or validation in their design or use. Various loci of accountability include those for commanders (e.g., what are the reasonable standards for commanders to maintain meaningful human control?), and operators (e.g., what are the levels of understanding required by operators to have knowledge of the system state, operational context, and situational awareness?).

Ideally all procurers, suppliers, and users of weapons systems components have accountability for their part of every weapons system, potential incorporation in future systems, and expected and potential users.

Candidate Recommendations

- Designers should follow best practices in terms of design process, which entails clearly defined responsibilities for organizations, companies, and individuals within the process.
- Systems and components should be designed to deter the easy modification of the overall weapon after the fact to operate in fully autonomous mode.
- Further exploration of black box recording of data logs, as well as cryptographic, block-chain, and other technical methods for tracing access and authorization of weapons targeting and release is needed.

Reframing Autonomous Weapons Systems

- System engineers must work to the same high standards and regulations of security for AWS design from a cybersecurity perspective than they would for any other work. Weapons systems ought to be designed with cybersecurity in mind such that preventing tampering, or at least undetected tampering, is a highly weighted design constraint.
- Procurement authority: only contract with contractors who have proper legal and security processes; carry out Article 36 reviews at all major steps in the procurement process; maintain database of design, tests, and review evidence.
- Contractors: ensure design meets relevant engineering and defense standards for military products; deliver evidence for Article 36 reviews using, but not restricted to, design reviews and simulation models; provide evidence requested by user for setting ROE; ensure design has clear criteria for decisions made by their product.
- Acceptance body: have validation and test plans for behavior of actual system produced; test weapons systems in a number of representative scenarios; have plans to ensure upgrades are reviewed against IHL criteria such as Article 36.
- User/military commanders: only operate weapons systems with meaningful human control and in accordance with delegated authority.
- Weapons systems must have default modes of operation agreed with campaign planners before operation commences.
- Ensure as many aspects of weapons systems as possible are designed with fail-safe behaviors.
- Ensure clear embedded lines of accountability in the design, deployment, and operation of weapons.
- Trusted user authentication logs and audit trail logs are necessary, in conjunction with meaningful human control. Thorough human-factors-driven design of user interface and human–computer/robot interaction design is necessary for situational awareness, knowability, understandability, and interrogation of system goals, reasons, and constraints, such that the user could be held culpable.
- Tamper-proof the equipment used to store authorization signals and base this on open, auditable designs, as suggested by Gubrud and Altmann (2013). Further, the hardware that implements the human-in-the-loop requirement should not be physically distinct from operational hardware.

There will need to be checks that all these bodies and organizations have discharged their responsibilities according to IHL and their domestic laws. Even if this is the case, weapons system operations may be compromised by, for example, equipment failure, actions by

Reframing Autonomous Weapons Systems

opponents such as cyber-attacks, or deception so that the automated functions act according to design but against an incorrect target.

There are currently weapons systems in use that, once activated, automatically intercept high-speed inanimate objects such as incoming missiles, artillery shells, and mortar grenades. Examples include SEA-RAM, C-RAM, Phalanx, NBS Mantis, and Iron Dome. These systems complete their detection, evaluation, and response process within a matter of seconds and thus render it extremely difficult for human operators to exercise meaningful supervisory control once they have been activated, other than deciding when to switch them off. This is called *supervised autonomy* by the U.S. Department of Defense (DoD) because the weapons require constant and vigilant human evaluation and monitoring for rapid shutdown in cases of targeting errors, change of situation, or change in status of targets. However, most of these systems are only utilized in a defensive posture for close-in weapons systems support against incoming lethal threats.

Further Resources

- Gubrud, M., and J. Altmann. "Compliance Measures for an Autonomous Weapons Convention." International Committee for Robot Arms Control, 2013.
- U.K. Ministry of Defence. "The UK Approach to Unmanned Aircraft Systems (UAS)," Joint Doctrine Note 2/11, March 2011.
- Sharkey, N. "Towards a Principle for the Human Supervisory Control of Robot Weapons." *Politica and Società* 2 (2014): 305–324.
- Owens, D. "Figuring Forseeability." *Wake Forest Law Review* 44 (2009): 1277, 1281–1290.
- Roff, H. M., and R. Moyes. "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons Systems." Briefing Paper for the Delegates at the Convention on Certain Conventional Weapons Meeting of Experts on Lethal Autonomous Weapons Systems, Geneva, April 2016.
- Roff, H. M. "Meaningful Human Control or Appropriate Human Judgment." Briefing Paper for the Delegates at the 5th Review Conference at the Convention on Certain Conventional Weapons, Geneva, December 2016.
- Scherer, M. "Who's to Blame (Part 4): Who's to Blame if an Autonomous Weapon Breaks the Law?" *Law and AI*, February 24, 2016.
- Rebecca C, "War Torts: Accountability for Autonomous Weapons." *University of Pennsylvania Law Review* 164, no. 6 (2016): 1347–1402.
- Gillespie, T., and R. West. "Requirements for Autonomous Unmanned Air Systems Set by Legal Issues." *International C2 Journal* 4, no. 2 (2010): 1–32.
- Defense Science Board. "Summer Study on Autonomy." Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, June 2016.
- Rickli, J.-M. "Artificial Intelligence and the Future of Warfare" (Box 3.2.1). *2017 Global Risk Report,* Geneva: World Economic Forum, 2017.

Reframing Autonomous Weapons Systems

Issue 7:

AWS offer the potential for severe human rights abuses. Exclusion of human oversight from the battlespace can too easily lead to inadvertent violation of human rights. AWS could be used for deliberate violations of human rights.

Background

The ethical disintermediation afforded by AWS encourages the bypassing of ethical constraints on people's actions that should require the consent of multiple people, organizations, or chains of commands. This exclusion concentrates ethical decision-making into fewer hands.

The potential lack of clear lines of accountability for the consequences of AWS might encourage malicious use of AWS by those seeking to avoid responsibility for malicious or illegal acts.

Candidate Recommendations

Acknowledge that the design, development, or engineering of AWS for anti-personnel or anti-civilian purposes are unethical. An organization's values on respect and the avoidance of harm to persons precludes the creation of AWS that target human beings. If a system is designed for use against humans, such systems must be designed to be semi-autonomous, where the control over the critical functions remains with a human operator, (such as through a human-in-the-loop hardware interlock). Design for operator intervention must be sensitive to human factors and intended to increase, rather than decrease, situational awareness.

Under no circumstances is it morally permissible to use AWS without meaningful human control, and this should be prohibited. Ultimately, weapons systems must be under meaningful human control. As such, design decisions regarding human control must be made so that a commander has meaningful human control over direct attacks during the conduct of hostilities. In short, this requires that a human commander be present and situationally aware of the circumstances on the ground as they unfold to deploy either semi-autonomous or defensive anti-materiel AWS. Organizational members must ensure that the technologies they create enhance meaningful human control over increasingly sophisticated systems and do not undermine or eliminate the values of respect, humanity, fairness, and dignity.

Further Resources

- Heller, K. J. "Why Preventive Self-Defense Violates the UN Charter." *Opinio Juris*, March 7, 2012.
- Scherer, M. "Who's to Blame (Part 5): A Deeper Look at Predicting the Actions of Autonomous Weapons." *Law and AI*, February 29, 2016.

Reframing Autonomous Weapons Systems

- Roff, H. M. "Killer Robots on the Battlefield: The Danger of Using a War of Attrition Strategy with Autonomous Weapons." *Slate,* 2016.
- Roff, H. "Autonomous Weapons and Incentives for Oppression." *Duck of Minerva,* March 13, 2016.

Issue 8:

AWS could be used for covert, obfuscated, and non-attributable attacks.

Background

The lack of a clear owner of a given AWS incentivizes scalable covert or non-attributable uses of force by state and non-state actors. Such dynamics can easily lead to unaccountable violence and societal havoc.

Features of AWS that may contribute to their making covert and non-attributable attacks easier include: small size; the ability to swarm; and ability to act at great distance and time from the deployment of a weapon from responsible operators; layers of weapons systems within other systems.

States have a legal obligations to make attacks practically attributable. There are additional legal obligations not to booby trap autonomous systems. Self-destructive functions, such as those aimed at preventing access to sensitive technologies or data, should be designed to not cause incidental or intentional harm.

There are significant concerns about the use of AWS by non-state actors, or individuals, and the potential for use in terror attacks against civilians, and non-attributable attacks against states. Designers should be concerned about the potential of systems to be used by malicious actors.

Candidate Recommendation

Because AWS are delegated authority to use force in a particular situation, they are required to be attributable to the entity and human that deployed them. Designers should ensure that there is a clear and auditable authorization of actions taken by the AWS when in operation.

Further Resources

- Bahr, E. "Attribution of Biological Weapons Use," in *Encyclopedia of Bioterrorism Defense.* Hoboken, NJ: John Wiley & Sons, 2005.
- Mistral Solutions. "Close-In Covert Autonomous Disposable Aircraft (CICADA) for Homeland Security," 2014.
- Piore, A. "Rise of the Insect Drones." *Popular Science*. January 29, 2014.
- Gillespie, T., and R. West. "Requirements for Autonomous Unmanned Air Systems Set by Legal Issues." *International C2 Journal* 4, no. 2 (2010): 1–32.

Reframing Autonomous Weapons Systems

Issue 9:

The development of AWS will lead to a complex and troubling landscape of proliferation and abuse.

Background

Use of AWS by a myriad of actors of different kinds, including states (of different types of regime) and non-state actors (militia, rebel groups, individuals, companies, including private military contractors), would lead to such systems becoming commonplace anywhere anyone favors violence due to the disintermediation and scalability afforded by their availability.

There will be incentives for misuse depending upon state of conflict and type of actor. For example, such misuse may include, but is not limited to, political oppression, crimes against humanity, intimidation, assassination, and terrorism. This can lead to, for example, a single warlord targeting an opposing tribe based on their respective interests as declared on Facebook, their DNA, their mobile phones, or their appearance.

Candidate Recommendations

- One must design weapons with high degrees of automation in such a way that avoids tampering for unintended use. Further work on technical means for nonproliferation should be explored, for example, cryptographic chain authorization.
- There is an obligation to consider the foreseeable use of the system, and whether there is a high risk for misuse.
- There is an obligation to consider, reflect on, or discuss possible ethical consequences of one's research and/or the publication of that research.

Issue 10:

AWS could be deployed by domestic police forces and threaten lives and safety. AWS could also be deployed for private security. Such AWS may have very different design and safety requirements than military AWS.

Background

Outside of military uses of AWS, other likely applications include use by domestic police forces, as well as coast guards, border patrols, and other domestic security applications. Police in Dallas, Texas used a bomb disposal robot to deliver a bomb to kill a suspect in the summer of 2016. While that was a remotely operated weapon delivered by a remote operated platform, the path to more autonomous forms of police robots using weapons seems highly likely.

Beyond use by governments, AWS could potentially also be deployed for other private

Reframing Autonomous Weapons Systems

security applications, such as guarding property, patrolling areas, and personal protection.

Tyrants and despots might utilize AWS to gain or retain control over a population which would not otherwise support them. AWS might be turned against peaceful demonstrators when human law enforcement might not do the same.

Candidate Recommendations

- Police and private security systems should not be permitted to deploy weapons without meaningful human control.
- Police and security systems should deploy non-lethal means to disrupt and avert security threats and threats to the physical safety of humans.

Further Resources

- Asaro, P. "Will #BlackLivesMatter to RoboCop?" WeRobot 2016, University of Miami School of Law, Miami, FL, April 1–2, 2016.
- Asaro, P. "'Hands Up, Don't Shoot!' HRI and the Automation of Police Use of Force," Special Issue on Robotics Law and Policy, *Journal of Human-Robot Interaction* 5, no. 3 (2016): 55–69.

Issue 11:

An automated weapons system might not be predictable (depending upon its design and operational use). Learning systems compound the problem of predictable use.

Background

Autonomous systems that react and adapt to environmental and sensor inputs results in systems that may be predictable in their general behavior, but may manifest individual or specific actions that cannot be predicted in advance.

As autonomous systems become more complex in their processing of data, the ability of designers to anticipate and predict their behavior becomes increasingly difficult.

As adaptive systems modify their functional operations through learning algorithms and other means, their behavior becomes more dependent upon the content of training data and other factors which cannot be anticipated by designers or operators.

Even when a single system is predictable, or even deterministic, when such systems interact with other systems, or in large masses or swarms, their collective behavior can become intrinsically unpredictable. This includes unpredictable interactions between known systems and adversarial systems whose operational behavior may be unknown.

Reframing Autonomous Weapons Systems

Modeling and simulation of AWS, particularly learning systems, may not capture all possible circumstances of use or situational interaction. They are underconstrained cyberphysical systems. Intrinsic unpredictability of adaptive systems is also an issue: one cannot accurately model the systems of one's adversary and how an adversary will adapt to your system resulting in an inherently unpredictable act.

Candidate Recommendations

- Systems that exhibit intrinsically unpredictable behavior should be considered illegal and not deployed.
- Similarly, deploying systems with otherwise predictable behavior in situations or contexts in which the collective behavior of systems cannot be predicted should be avoided. In particular, deploying AWS swarms in which the emergent dynamics of the swarm have a significant impact on the actions of an individual AWS must be avoided.
- The predictability of weapons systems should be assessed with confidence levels with respect to specified contexts and circumstances of use. Systems should not be used outside of the contexts of use for which their operational behavior is understood and predictable. Engineers should explicitly examine their systems and inform their customers of their qualitative and quantitative confidence in the predictability of the actions of the autonomous functions of weapons systems in response to representative scenarios, specific contexts of use, and scope of operations.
- Commanders and operators should be trained to understand and assess confidence in the behavior of a system under specific contexts and scope of operations. They should maintain situational awareness of those contexts where weapons systems are deployed, and prevent those systems from being used outside the scope of operations for which their behavior is predictable.
- To ensure meaningful human control, operators should be able to query a system in real-time. Such a query should offer the evidence, explanation, and justification for critical determinations made by the systems, i.e., identification of a target, or key recommendations.
- Weapons systems with advance automation should also keep records and traces of critical functional and operational decisions that are made automatically. Such traces and records should be reviewable in instances where the behavior of the system was not as predicted.
- To the extent that systems contain adaptive or learning algorithms, any critical decision made by systems based upon those algorithms should be transparent and explainable by the designing engineers. Any data used for training and adaptation should be reviewed as to its integrity so as to ensure that learned functions can behave in reliably predictable ways.

Reframing Autonomous Weapons Systems

Further Resources

- International Committee for Robot Arms Control. "LAWS: Ten Problems for Global Security" (leaflet). April 10, 2015.
- Owens, D. "Figuring Forseeability." *Wake Forest Law Review* 44 (2009): 1277, 1281–1290.
- Scherer, M. "Who's to Blame (Part 5): A Deeper Look at Predicting the Actions of Autonomous Weapons." *Law and AI*, February 29, 2016.
- Arquilla, J., and D. Ronfeldt. *Swarming and the Future of Conflict*, Santa Monica, CA: RAND Corporation, 1997.
- Edwards, S. J. A. *Swarming and the Future of Warfare*, Santa Monica, CA: RAND Corporation, 2004.
- Rickli, J.-M. "Some Consideration of the Impact of Laws on International Security: Strategic Stability, Non-State Actors and Future Prospects." Meeting of Experts on Lethal Autonomous Weapons Systems Convention on Certain Conventional Weapons (CCW) United Nations Office Geneva, April 16, 2015.
- Scharre, P. *Robotics on the Battlefield Part II: The Coming Swarm*, Washington, DC: Center for a New American Security, 2014.