

# Bulletin of the Atomic Scientists

<http://bos.sagepub.com/>

---

## Stopping killer robots

Mark Gubrud

*Bulletin of the Atomic Scientists* 2014 70: 32

DOI: 10.1177/0096340213516745

The online version of this article can be found at:

<http://bos.sagepub.com/content/70/1/32>

---

Published by:



<http://www.sagepublications.com>

On behalf of:

Bulletin of the Atomic Scientists

Additional services and information for *Bulletin of the Atomic Scientists* can be found at:

**Email Alerts:** <http://bos.sagepub.com/cgi/alerts>

**Subscriptions:** <http://bos.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Jan 6, 2014

[What is This?](#)



Feature

Bulletin of the Atomic Scientists

2014, Vol. 70(1) 32–42

© The Author(s) 2014

Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/0096340213516745

<http://thebulletin.sagepub.com>



## Stopping killer robots

Mark Gubrud

### Abstract

Autonomous weapons are robotic systems that, once activated, can select and engage targets without further intervention by a human operator. Advances in computer technology, artificial intelligence, and robotics may lead to a vast expansion in the development and use of such weapons in the near future. Public opinion runs strongly against killer robots. But many of the same claims that propelled the Cold War are being recycled to justify the pursuit of a nascent robotic arms race. Autonomous weapons could be militarily potent and therefore pose a great threat. For this reason, substantial pressure from civil society will be needed before major powers will seriously consider their prohibition. However, demands for human control and responsibility and the protection of human dignity and sovereignty fit naturally into the traditional law of war and imply strict limits on autonomy in weapon systems. Opponents of autonomous weapons should point out the terrible threat they pose to global peace and security, as well as their offensiveness to principles of humanity and to public conscience.

### Keywords

autonomous weapons, Convention on Certain Conventional Weapons, international humanitarian law, killer robots, unmanned weapons

Since the first lethal drone strike in 2001, the US use of remotely operated robotic weapons has dramatically expanded. Along with the broader use of robots for surveillance, ordnance disposal, logistics, and other military tasks, robotic weapons have spread rapidly to many nations, captured public attention, and sparked protest and debate. Meanwhile, every dimension of the technology is being vigorously explored. From stealthy, unmanned jets like the X-47B and its Chinese and European counterparts, to intelligent missiles, sub-hunting robot ships, and machine gun-wielding micro-tanks, robotics is now

the most dynamic and destabilizing component of the global arms race.

Drones and robots are enabled by embedded autonomous subsystems that keep engines in tune and antennas pointed at satellites, and some can navigate, walk, and maneuver in complex environments autonomously. But with few exceptions, the targeting and firing decisions of armed robotic systems remain tightly under the control of human operators. This may soon change.

Autonomous weapons are robotic systems that, once activated, can select and engage targets without further intervention by a human operator (Defense



The X-47B unmanned combat aircraft during a 2011 test.  
*Photo credit: US Air Force*

Department, 2012). Examples include drones or missiles that hunt for their targets, using their onboard sensors and computers. Based on a computer's decision that an appropriate target has been located, that target will then be engaged. Sentry systems may have the capability to detect intruders, order them to halt, and fire if the order is not followed. Future robot soldiers may patrol occupied cities. Swarms of autonomous weapons may enable a preemptive attack on an adversary's strategic forces. Autonomous weapons may fight each other.

Just as the emergence of low-cost, high-performance information technology has been the most important driver of technological advance over the past half-century—including the revolution in military affairs already seen in the 1980s and displayed to the world during the 1991 Gulf War—so the emergence of artificial intelligence and autonomous robotics will likely be the most important development in both civilian and military technology to unfold over the next few decades.

Proponents of autonomous weapons argue that technology will gradually take over combat decision making: "Detecting, analyzing and firing on targets will become increasingly automated,

and the contexts of when such force is used will expand. As the machines become increasingly adept, the role of humans will gradually shift from full command, to partial command, to oversight and so on" (Anderson and Waxman, 2013). Automated systems are already used to plan campaigns and logistics, and to assemble intelligence and disseminate lethal commands; in some cases, humans march to orders generated by machines. If, in the future, machines are to act with superhuman speed and perhaps even superhuman intelligence, how can humans remain in control? As former Army Lt. Colonel T. K. Adams observed more than a decade ago (2001), "Humans may retain symbolic authority, but automated systems move too fast, and the factors involved are too complex for real human comprehension."

Almost nobody favors a future in which humans have lost control over war machines. But proponents of autonomous weapons argue that effective arms control would be unattainable. Many of the same claims that propelled the Cold War are being recycled to argue that autonomous weapons are inevitable, that international law will remain weak, and that there is no point in seeking restraint since adversaries will not agree—or would cheat on agreements. This is the ideology of any arms race.

### **Is autonomous warfare inevitable?**

Challenging the assumption of the inevitability of autonomous weapons and building on the work of earlier activists, the Campaign to Stop Killer Robots, a coalition of nongovernmental organizations, was launched in April 2013. This effort has made remarkable progress in

its first year. In May, the United Nations Special Rapporteur on extrajudicial killings, Christof Heyns, recommended that nations immediately declare moratoriums on their own development of lethal autonomous robotics (Heyns, 2013). Heyns also called for a high-level study of the issue, a recommendation seconded in July by the UN Advisory Board on Disarmament Matters. At the UN General Assembly's First Committee meeting in October, a flood of countries began to express interest or concern, including China, Russia, Japan, the United Kingdom, and the United States. France called for a mandate to discuss the issue under the Convention on Certain Conventional Weapons, a global treaty that restricts excessively injurious or indiscriminate weapons. Meeting in Geneva in November, the state parties to the Convention agreed to formal discussions on autonomous weapons, with a first round in May 2014. The issue has been placed firmly on the global public and diplomatic agenda.

Despite this impressive record of progress on an issue that was until recently virtually unknown—or scorned as a mixture of science fiction and paranoia—there seems little chance that a strong arms control regime banning autonomous weapons will soon emerge from Geneva. Unlike glass shrapnel, blinding lasers, or even landmines and cluster munitions, autonomous weapon systems are not niche armaments of negligible strategic importance and unarguable cost to humanity. Instead of the haunting eyes of children with missing limbs, autonomous weapons present an abstract, unrealized horror, one that some might hope will simply go away.

Unless there is a strong push from civil society and from governments that

have decided against pursuing autonomous weapons, those that have decided in favor of them—including the United States (Gubrud, 2013)—will seek to manage the issue as a public relations problem. They will likely offer assurances that humans will remain in control, while continually updating what they mean by control as technology advances. Proponents already argue that humans are never really out of the loop because humans will have programmed a robot and set the parameters of its mission (Schmitt and Thurnher, 2013). But autonomy removes humans from decision making, and even the assumption that autonomous weapons will be programmed by humans is ultimately in doubt.

Diplomats and public spokesmen may speak in one voice; warriors, engineers, and their creations will speak in another. The development and acquisition of autonomous weapons will push ahead if there is no well-defined, immovable, no-go red line. The clearest and most natural place to draw that line is at the point when a machine pulls the trigger, making the decision on whether, when, and against whom or what to use violent force. Invoking a well-established tenet of international humanitarian law, opponents can argue that this is already contrary to principles of humanity, and thus *inherently* unlawful. Equally important, opponents must point out the threat to peace and security posed by the prospect of a global arms race toward robotic arsenals that are increasingly out of human control.

### **Humanitarian law vs. killer robots**

The public discussion launched by the Campaign to Stop Killer Robots has mostly centered on questions of legality

under international humanitarian law, also called the law of war. “Losing Humanity,” a report released by Human Rights Watch in November 2012—coincidentally just days before the Pentagon made public the world’s first open policy directive for developing, acquiring, and using autonomous weapons—laid out arguments that fully autonomous weapons could not satisfy basic requirements of the law, largely on the basis of assumed limitations of artificial intelligence (Human Rights Watch and International Human Rights Clinic at Harvard Law School, 2012).

The principle of distinction, as enshrined in Additional Protocol I of the Geneva Conventions—and viewed as customary international law, thus binding even on states that have not ratified the treaty—demands that parties to a conflict distinguish between civilians and combatants, and between civilian objects and military objectives. Attacks must be directed against combatants and military objectives only; weapons not capable of being so directed are considered to be indiscriminate and therefore prohibited. Furthermore, those who make attack decisions must not allow attacks that may be expected to cause excessive harm to civilians, in comparison with the military gains expected from the attack. This is known as the principle of proportionality.

“Losing Humanity” argues that technical limitations mean robots could not reliably distinguish civilians from combatants, particularly in irregular warfare, and could not fulfill the requirement to judge proportionality.<sup>1</sup> Distinction is clearly a challenge for current technology; face-recognition technology can rapidly identify individuals from a limited list of potential targets, but more

general classification of persons as combatants or noncombatants based on observation is well beyond the state of the art. How long this may remain so is less clear. The capabilities to be expected of artificial intelligence systems 10, 20, or 40 years from now are unknown and highly controversial within both expert and lay communities.

While it may not satisfy the reified principle of distinction, proponents of autonomous weapons argue that some capability for discrimination is better than none at all. This assumes that an indiscriminate weapon would be used if a less indiscriminate one were not available; for example, it is often argued that drone strikes are better than carpet bombing. Yet at some point autonomous discrimination capabilities may be good enough to persuade many people that their use in weapons is a net benefit.

Judgment of proportionality seems at first an even greater challenge, and some argue that it is beyond technology in principle (Asaro, 2012). However, the military already uses an algorithmic “collateral damage estimation methodology” (Defense Department, 2009) to estimate incidental harm to civilians that may be expected from missile and drone strikes. A similar scheme could be developed to formalize the value of military gains expected from attacks, allowing two numbers to be compared. Human commanders applying such protocols could defend their decisions, if later questioned, by citing such calculations. But the cost of this would be to degrade human judgment almost to the level of machines.

On the other hand, IBM’s Watson computer (Ferruci et al., 2010) has demonstrated the ability to sift through millions of pages of natural language and

weigh hundreds of hypotheses to answer ambiguous questions. While some of Watson's responses suggest it is not yet a trustworthy model, it seems likely that similar systems, given semantic information about combat situations, including uncertainties, might be capable of making military decisions that most people would judge as reasonable, most of the time.

"Losing Humanity" also argues that robots, necessarily lacking emotion,<sup>2</sup> would be unable to empathize and thus unable to accurately interpret human behavior or be affected by compassion. An important case of the latter is when soldiers refuse orders to put down rebellions. Robots would be ideal tools of repression and dictatorship.

If robot soldiers become available on the world market, it is likely that repressive regimes will acquire them, either by purchase or indigenous production. While it is theoretically possible for such systems to be safeguarded with tamper-proof programming against human rights abuses, in the event that the world fails to prohibit robot soldiers, unsafeguarded or poorly safeguarded versions will likely be available. A strong prohibition has the best chance of keeping killer robots out of the hands of dictators, both by restricting their availability and stigmatizing their use.

Accountability is another much-discussed issue. Clearly, a robot cannot be held responsible for its actions, but human commanders and operators—or even manufacturers, programmers, and engineers—might be held responsible for negligence or malfeasance. In practice, however, the robot is likely to be a convenient scapegoat in case of an unintended atrocity—a technical failure occurred, it was unintended and unforeseen,

so nobody is to blame. Going further, David Akerson (2013) argues that since a robot cannot be punished, it cannot be a legal combatant.

These are some of the issues most likely to be discussed within the Convention on Certain Conventional Weapons. However, US Defense Department policy (2012) preemptively addresses many of these issues by directing that "[a]utonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force."

Under the US policy, commanders and operators are responsible for using autonomous weapons in accordance with the laws of war and relevant treaties, safety rules, and rules of engagement. For example, an autonomous weapon may be sent on a hunt-and-kill mission if tactics, techniques, and procedures ensure that the area in which it is directed to search contains no objects, other than the intended targets, that the weapon might decide to attack. In this case, the policy regards the targets as having been selected by humans and the weapon as merely semi-autonomous, even if the weapon is operating fully autonomously when it decides that a given radar return or warm object is its intended target. The policy preapproves the immediate development, acquisition, and use of such weapons.

Although the policy does not define "appropriate levels," it applies this rubric even in the case of fully autonomous lethal weapons targeting human beings without immediate human supervision. This makes it clear that appropriate levels, as understood within the policy, do not necessarily require direct human involvement in the decision to kill

a human being (Gubrud, 2013). It seems likely that the United States will press other states to accept this paradigm as the basis for international regulation of autonomous weapons, leaving it to individual states to determine what levels of human judgment are appropriate.

### **Demanding human control and responsibility**

As diplomatic discussions about killer robot regulation get under way, a good deal of time is apt to be lost in confusion about terms, definitions, and scope. “Losing Humanity” seeks to ban “fully autonomous weapons,” and Heyns’s report used the term “lethal autonomous robotics.” The US policy directive speaks of “autonomous and semi-autonomous weapon systems,” and the distinction between these is ambiguous (Gubrud, 2013). The Geneva mandate is to discuss “lethal autonomous weapon systems.”

Substantive questions include whether non-lethal weapons and those that target only matériel are within the scope of discussion. Legacy weapons such as simple mines may be regarded as autonomous, or distinguished as merely automatic, on grounds that their behavior is fully predictable by designers.<sup>3</sup> Human-supervised autonomous and semi-autonomous weapon systems, as defined by the United States, raise issues that, like fractal shapes, appear more complex the more closely they are examined.

Instead of arguing about how to define what weapons should be banned, it may be better to agree on basic principles. One is that any use of violent force, lethal or non-lethal, must be by human decision and must at all times be under human control. Implementing this principle as strictly as possible implies that the

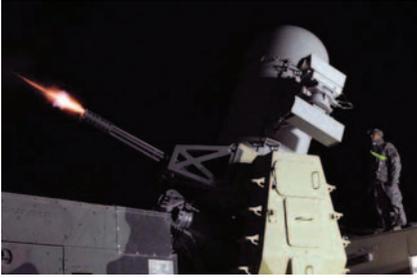
command to engage an individual target (person or object) must be given by a human being, and only *after* the target is being reliably tracked by a targeting system and a human has determined that it is an appropriate and legal target.

A second principle is that a human commander must be responsible and accountable for the decision, and if the commander acts through another person who operates a weapon system, that person must be responsible and accountable for maintaining control of the system. “Responsible” refers here to a moral and legal obligation, and “accountable” refers to a formal system for accounting of actions. Both elements are essential to the approach.

Responsibility implies that commanders and operators may not blame inadequacies of technological systems for any failure to exercise judgment and control over the use of violent force. A commander must ensure compliance with the law and rules of engagement independently of any machine decision, either as to the identity of a target or the appropriateness of an attack, or else must not authorize the attack. Similarly, if a system does not give an operator sufficient control over the weapon to prevent unintended engagements, the operator must refuse to operate the system.

Accountability can be demonstrated by states that comply with this principle. They need only maintain records showing that each engagement was properly authorized and executed. If a violation is alleged, selected records can be unsealed in a closed inquiry conducted by an international body (Gubrud and Altmann, 2013).<sup>4</sup>

This framing, which focuses on human control and responsibility for the decision to use violent force, is both



The Counter-Rocket, Artillery, and Mortar (C-RAM) gun can fire thousands of rounds per minute at incoming projectiles.

*Photo credit: US Air Force*

conceptually simple and morally compelling. What remains then is to set standards for adequate information to be presented to commanders, and to require positive action by operators of a weapon system. Those standards should also address any circumstances under which other parties—designers and manufacturers, for instance—might be held responsible for an unintended engagement.

There is at least one exceptional circumstance in which human control may be applied less strictly. Fully autonomous systems are already used to engage incoming missiles and artillery rounds; examples include the Israeli Iron Dome and the US Patriot and Aegis missile defense systems, as well as the Counter Rocket, Artillery, and Mortar system. The timeline for response in such systems is often so short that the requirement for positive human decision might impose an unacceptable risk of failure. Another principle—the protection of life from immediate threats—comes into play here. An allowance seems reasonable, if it is strictly limited. In particular, autonomous return fire should not be permitted, but only engagement of unmanned munitions directed against

human-occupied territory or vehicles. Each such system should have an accountable human operator, and autonomous response should be delayed as long as possible to allow time for an override decision.

### **The strategic need for robot arms control**

Principles of humanity may be the strongest foundation for an effective ban of autonomous weapons, but they are not necessarily the most compelling reason why a ban must be sought. The perceived military advantages of autonomy are so great that major powers are likely to strongly resist prohibition, but by the same token, autonomous weapons pose a severe threat to global peace and security.

Although humans have (for now) superior capabilities for perception in complex environments and for interpretation of ambiguous information, machines have the edge in speed and precision. If allowed to return fire or initiate it, they would undoubtedly prevail over humans in many combat situations. Humans have a limited tolerance of the physical extremes of acceleration, temperature, and radiation, are vulnerable to biological and chemical weapons, and require rest, food, breathable air, and drinkable water. Machines are expendable; their loss does not cause emotional pain or political backlash. Humans are expensive, and their replacement by robots is expected to yield cost savings.

While today's relatively sparse use of drones, in undefended airspace, to target irregular forces can be carried out by remote control, large-scale use of robotic weapons to attack modern military forces would require greater

autonomy, due to the burdens and vulnerabilities of communications links, the need for stealth, and the sheer numbers of robots likely to be involved. The US Navy is particularly interested in autonomy for undersea systems, where communications are especially problematic. Civilians are sparse on the high seas and absent on submarines, casting doubt on the relevance of humanitarian law. As the Navy contemplates future conflict with a peer competitor, it projects drone-versus-drone warfare in the skies above and waters below, and the use of sea-based drones to attack targets inland as well.

In a cold war, small robots could be used for covert infiltration, surveillance, sabotage, or assassination. In an open attack, they could find ways of getting into underground bunkers or attacking bases and ships in swarms. Because robots can be sent on one-way missions, they are potential enablers of aggression or preemption. Because they can be more precise and less destructive than nuclear weapons, they may be more likely to be used. In fact, the US Air Force's Long Range Strike Bomber is planned to be both nuclear-capable and potentially unmanned, which would almost certainly mean autonomous.

There can be no real game-changers in the nuclear stalemate. Yet the new wave of robotics and artificial intelligence-enabled systems threatens to drive a new strategic competition between the United States and other major powers—and lesser powers, too. Unlike the specialized technologies of high-performance military systems at the end of the Cold War, robotics, information technology, and even advanced sensors are today globally available, driven as much by civilian as military uses. An autonomous

weapons arms race would be global in scope, as the drone race already is.

Since robots are regarded as expendable, they may be risked in provocative adventures. Recently, China has warned that if Japan makes good on threats to shoot down Chinese drones that approach disputed islands, it could be regarded as an act of war. Similarly, forward-basing of missile interceptors (Lewis and Postol, 2010) or other strategic weapons on unmanned platforms would risk misinterpretation as a signal of imminent attack, and could invite preemption.

Engineering the stability of a robot confrontation would be a wickedly hard problem even for a single team working together in trust and cooperation, let alone hostile teams of competing and imperfectly coordinated sub-teams. Complex, interacting systems-of-systems are prone to sudden unexpected behavior and breakdowns, such as the May 6, 2010 stock market crash caused by interacting exchanges with slightly different rules (Nanex, 2010). Even assuming that limiting escalation would be a design objective, avoiding defeat by preemption would be an imperative, and this implies a constant tuning to the edge of instability. The history of the Cold War contains many well-known examples in which military response was interrupted by the judgment of human beings. But when tactical decisions are made with inhuman speed, the potential for events to spiral out of control is obvious.

## The way out

Given the military significance of autonomous weapons, substantial pressure from civil society will be needed before the major powers will seriously consider accepting hard limits, let alone

prohibition. The goal is as radical as, and no less necessary than, the control and abolition of nuclear weapons.

The principle of humanity is an old concept in the law of war. It is often cited as forbidding the infliction of needless suffering, but at its deepest level it is a demand that even in conflict, people should not lose sight of their shared humanity. There is something inhumane about allowing technology to decide the fate of human lives, whether through individual targeting decisions or through a conflagration initiated by the unexpected interactions of machines. The recognition of this is already deeply rooted. A scientific poll (Carpenter, 2013) found that Americans opposed to autonomous weapons outnumbered supporters two to one, in contrast to an equally strong consensus in the United States supporting the use of drones. The rest of the world leans heavily against the drone strikes (Pew Research Center, 2012), making it seem likely that global public opinion will be strongly against autonomous weapons, both on humanitarian grounds and out of concern for the dangers of a new arms race.

In the diplomatic discussions now under way, opponents of autonomous weapons should emphasize a well-established principle of international humanitarian law. Seeking to resolve a diplomatic impasse at the Hague Conference in 1899, Russian diplomat Friedrich Martens proposed that for issues not yet formally resolved, conduct in war was still subject to “principles of international law derived from established custom, from the principles of humanity, and from the dictates of public conscience.” Known as the Martens Clause, it reappeared in the second Hague Convention (1907), the Tehran Conference on

Human Rights (1968), and the Geneva Convention additional protocols (1977). It has been invoked as the source of authority for retroactive liability in war crimes and for preemptive bans on inhumane weapons, implying that a strong public consensus has legal force in anticipation of an explicit law (Meron, 2000).

Autonomous weapons are a threat to global peace and therefore a matter of concern under the UN Charter. They are contrary to established custom, principles of humanity, and dictates of public conscience, and so should be considered as preemptively banned by the Martens Clause. These considerations establish the legal basis for formal international action to prohibit machine decision in the use of force. But for such action to occur, global civil society will need to present major-power governments with an irresistible demand: Stop killer robots.

### **Funding**

This research was supported in part by a grant from the John D. and Catherine T. MacArthur Foundation.

### **Notes**

1. Philosopher Peter Asaro (2012) takes this a step further, arguing that those who plan or decide military attacks are assumed to be human, and that proportionality, in particular, is inherently subjective and represents not just an algorithmic criterion but a moral burden upon commanders to give due consideration to the human costs in judging whether a lethal action is justified.
2. However, emotion is an active area of research in robotics and artificial intelligence. Research goals include the recognition of human emotion, the simulation of emotional responses, social interaction, and internal states as moderators of robot behavior. Moreover, it is not clear that emotion is the best way to govern behavior in robots.

Hard rules such as “Don’t open fire on civilians” might be preferable (Arkin, 2009).

3. Mines have already been addressed by the Convention on Certain Conventional Weapons and by the landmines and cluster munitions treaties. Their inclusion in an autonomous weapons ban would strengthen it from the point of view of conceptual and moral clarity, but might needlessly complicate the negotiations.
4. Such records would logically include the data on which the engagement decision was based, as well as the identities of commander and operator, and some evidence of their human action. Identities can be kept secret, but the commander’s account of the justification for an attack may be crucial. The records can be time-stamped, encrypted, and archived, and a digital signature (hash) of each record published or held by an international body, in order to prove that records were not later altered. Some data, such as the time-stamped engagement command, could be recorded in real time by tamper-proofed “glass box” verification devices (Gubrud and Altmann, 2013).

## References

- Adams TK (2001) Future warfare and the decline of human decisionmaking. *Parameters* 31(4): 57–71.
- Akerson D (2013) The illegality of offensive lethal autonomy. In: Saxon D (ed.) *International Humanitarian Law and the Changing Technology of War*. Leiden: Martinus Nijhoff, 65–98.
- Anderson K and Waxman M (2013) Killer robots and the laws of war. *Wall Street Journal*, November 4, A19. Available at: <http://online.wsj.com/news/articles/SB1000142405270230465104579163361884479576>.
- Arkin RC (2009) *Governing Lethal Behavior in Autonomous Robots*. Boca Raton, FL: CRC.
- Asaro P (2012) On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross* 94(886): 687–709.
- Carpenter C (2013) US public opinion on autonomous weapons. Available at: [www.whiteoliphant.com/duckofminerva/wp-content/uploads/2013/06/UMass-Survey\\_Public-Opinion-on-Autonomous-Weapons\\_May2013.pdf](http://www.whiteoliphant.com/duckofminerva/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons_May2013.pdf).
- Defense Department (2009) Chairman of the Joint Chiefs of Staff instruction 3160.01: No-strike and the collateral damage estimation methodology. February 13. Available at: [www.aclu.org/files/dronefoia/dod/drone\\_dod\\_3160\\_01.pdf](http://www.aclu.org/files/dronefoia/dod/drone_dod_3160_01.pdf).
- Defense Department (2012) Directive 3000.09: Autonomy in weapon systems. November 21. Available at: [www.dtic.mil/whs/directives/corres/pdf/300009p.pdf](http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf).
- Ferruci D, Brown E, Chu-Carroll, et al. (2010) Building Watson: An overview of the DeepQA Project. *AI Magazine* 31(3): 59–79. Available at: [www.aaai.org/ojs/index.php/aimagazine/article/view/2303](http://www.aaai.org/ojs/index.php/aimagazine/article/view/2303).
- Gubrud M (2013) US killer robot policy: Full speed ahead. *Bulletin of the Atomic Scientists*, web edition, September 20. Available at: <http://thebulletin.org/us-killer-robot-policy-full-speed-ahead>.
- Gubrud M and Altmann J (2013) Compliance measures for an autonomous weapons convention. International Committee for Robot Arms Control Working Paper no. 2. Available at: [http://icrac.net/wp-content/uploads/2013/05/Gubrud-Altman\\_Compliance-Measures-AWC\\_ICRAC-WP2.pdf](http://icrac.net/wp-content/uploads/2013/05/Gubrud-Altman_Compliance-Measures-AWC_ICRAC-WP2.pdf).
- Heyns C (2013) Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions. United Nations Document A/HRC/23/47, April 9. Available at: [www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47\\_en.pdf](http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf).
- Human Rights Watch and International Human Rights Clinic at Harvard Law School (2012) Losing humanity: The case against killer robots. November 19. Available at: [www.hrw.org/reports/2012/11/19/losing-humanity-o](http://www.hrw.org/reports/2012/11/19/losing-humanity-o).
- Lewis GN and Postol TA (2010) How US strategic antimissile defense could be made to work. *Bulletin of the Atomic Scientists* 66(6): 8–24.
- Meron T (2000) The Martens Clause, principles of humanity, and dictates of public conscience. *American Journal of International Law* 94(1): 78–89.
- Nanex (2010) Analysis of the ‘Flash Crash’. June 18. Available at [www.nanex.net/20100506/FlashCrashAnalysis\\_Intro.html](http://www.nanex.net/20100506/FlashCrashAnalysis_Intro.html).
- Pew Research Center (2012) Global opinion of Obama slips, international policies faulted. June 13. Available at: [www.pewglobal.org/2012/06/13/global-opinion-of-obama-slips-international-policies-faulted/](http://www.pewglobal.org/2012/06/13/global-opinion-of-obama-slips-international-policies-faulted/).
- Schmitt MN and Thurnher JS (2013) ‘Out of the loop’: Autonomous weapon systems and the law of armed conflict. *Harvard National Security Journal* 4(2): 231–281.

**Author biography**

**Mark Gubrud** is a member of the International Committee for Robot Arms Control. Research for this article was done while he was a postdoctoral research associate in the Program on

Science and Global Security at Princeton University. He did his doctoral work in low-temperature and nanoscale experimental physics at the University of Maryland and has taught physics at the University of North Carolina.