

The image artistically portrays the self-searching and identity evolution of humans as they sort through how to incorporate more and more sophisticated technology. The author of the article examines how jus in bello responsibility will have to shift radically as autonomous weapons systems grow in sophistication. (Photo courtesy of Pixabay)

Warbots and Due Care The Cognitive Limitations of Autonomous and Human Combatants

Maj. Jules Hurst, U.S. Army Reserve

uring World War II, pilots relied on analog calculators and heuristics to place their bombs on target. By the 1980s, targeting computers notified pilots precisely when to release their munitions. The failure of the pilot to release munitions at the correct moment could result in munitions missing their target by large distances,

all dependent on the aircraft's altitude, speed, and orientation. Today, the large launch acceptability regions of precision-guided munitions (PGMs) require pilots to exercise significantly less skill to place bombs on target. If a PGM misses a positively identified target, fault likely lies with the warhead, not the pilot's release technique. The proliferation of semiautonomous systems on the battlefield will further this trend. As computers take on increased warfighting responsibilities, equipment malfunctions will make up an increased percentage of weapon employment errors. Soldiers can easily digest the error rates and circular-error probables of precision weapons, but the errors a semiautonomous warbot could make might be more troublesome to divine. Accurate analysis of these failures may require both knowledge of the algorithms that drove the robot's decisions and the code that executed them. An F-35 Joint Strike Fighter already requires eight million lines of code to power its flight, but future semiautonomous systems will need more.¹

The complexity of future warbot decision-making will potentially make combatant attempts to exercise jus in bello responsibilities—just conduct in war—increasingly difficult. In parallel with their capability, warbots will exercise greater latitude on the battlefield, and their decision-making parameters will be largely set before they leave the factory, potentially unmodifiable by the soldiers who employ them. Society imposes ethical responsibility proportionate to an individual or an organization's ability to control the actions in question. When lethal autonomous systems proliferate on the battlefield, combatants may not be the dominant force in controlling their activities. Instead, the acquisition officials and engineers who acquire and construct these systems will likely have the most influence on warbot battlefield behavior. Accordingly, jus in bello concepts must be updated to account for the greater responsibility that noncombatants in the weapon-system acquisition chain will hold in affecting ethical behavior on the battlefield.

Which Autonomous Systems?

Not all lethal autonomous weapon systems (LAWS) will complicate combatant attempts to exercise ethical responsibilities in war. If advances in neural networks and computer science provide LAWS with decision-making capabilities and sensory perception analogous to human beings; logically, jus in bello worries will be no worse than those for human combatants. Regardless, there will undoubtedly be a transition period between the achievement of human-level artificial intelligence and the appearance of warbots on the battlefield.

Expert estimates of when human-like artificial intelligence (AI) will appear vary wildly. In 2012, analysts from the Machine Intelligence Research Institute examined 257 expert and nonexpert literary predictions of when machines would achieve human-comparable cognitive performance. These predictions stretched from 1980 to beyond 2100. The majority clustered between 2020 and 2060.² No expert consensus on the arrival date of human-comparable AI exists, nor is there a precise definition of the term. Besides, the creation of machines with human-like intelligence will not necessarily coincide with machines achieving a human-comparable ability to perceive their environments.3 Intelligence and sensory perception are independent abilities; ask someone handicapped by hearing loss or blindness at your peril. Human-like AI could remain a long way away, and even upon its arrival, human-level AI may only be achievable through the use of a room-sized computer or quantum processors in subzero temperatures.⁴ Moreover, it will likely take many years to cost-effectively miniaturize human-level AI for deployment in tactical weapon systems, just like it took decades for personal computers to become financially and technologically viable.5

The Transition Period between Human-level AI and Combat-Capable Warbots

Regardless of when robots achieve intellectual and sensory parity to human beings, there will almost certainly be a transition period prior to human-equivalent AI where nations place LAWS on the battlefield because of machine advantages in performing narrowly defined tasks.⁶

With few exceptions, machines can outperform human beings in any narrowly defined role. The utilization of automated weapon systems in well-defined, routine tasks already offers the United States and other sophisticated forces tremendous tactical advantages. These advantages will grow in

Maj. Jules "Jay" Hurst, U.S. Army Reserve,

has previously served as the operations officer for U.S. Special Operations Command's Analytic Innovation and Technology Cell and as the civilian senior intelligence analyst for 1st Battalion, 75th Ranger Regiment. parallel with technological gains and military cultural comfort with their employment. The capability to employ machines with subhuman levels of artificial intelligence in structured tasks (narrow AI) already factors heavily into a nation's military prowess. Over the long term, it will become even more important.⁷ The temptation for nations to grant robots lethal autonomy before they reach human levels of cognition and perception will be enormous.

Current U.S. policy restricts development of robotic vehicles to those that "allow commanders and operators to exercise appropriate levels of human judgment over the use of force," but the U.S. military might abandon this directive during conflict for practical reasons.⁸ Unless the international community successfully imposes an arms control regime that bans killer robots or constrains their use through international law or established norms, belligerent nations could easily justify the use of LAWS as a necessity in a protracted or even limited conflict.⁹ Even if the international community establishes an arms control regime or legislation governing the use of autonomous systems, history shows that states quickly violate these agreements when necessity or desire arises. In the 1930s, the Japanese empire and Third Reich both withdrew from or violated arms-related provisions of the Washington Naval Treaty and the Treaty of Versailles.¹⁰ More liberal regimes also tend to disregard international provisions or agreements as it suits them. Just fifteen years ago, President George W. Bush withdrew the United States from the Anti-Ballistic Missile Defense Treaty signed with the Soviet Union in 1972 to begin construction of a national missile defense system.¹¹

When a significant party to an arms control treaty defects for military advantage, potential adversaries to that party and signatories to the treaty lose incentives to comply with its provisions; one bad apple ruins the whole bunch. If the military advantage of fielding lethal systems with narrow AI seems large enough, countries could rapidly modify internal restrictions or disregard treaties and norms. These policy changes could even arise out of well-intentioned military uses. The Western world continues to deploy tens of thousands of soldiers to combat terrorist threats in Africa, Asia, and the Middle East. Deployment of LAWS in place of human soldiers may be appealing to democratically elected regimes who need to balance electorate demands for national security with public unwillingness to endure wartime casualties.¹²

If the international community does manage to effectively enforce bans on killer robots, autonomous systems may still have opportunities to make lethal battlefield decisions. Assuming that nations field offensive robotic systems, even with a human in-the-loop, warbots suffering electronic- or cyberattack could find themselves cut off from human controllers by enemy (or friendly) electronic attacks.¹³ Militaries worldwide have embraced the expansion of electronic warfare capabilities to all echelons, largely spurred by modern reliance on wireless communication and threats from improvised explosive devices and drones. China and Russia, in particular, aspire to block the tactical communications of an opposing force, a potential difficulty for U.S. forces accustomed to unimpeded command and control.¹⁴ Warbots roaming the battlefield could easily be cut off from human control by an electronic attack or modified through cyber means.

Alternatively, damage to warbots could render them incapable of receiving or processing human control inputs. Malfunctioning warbots or autonomous systems separated from controllers could act like twenty-first century war elephants, weapon systems capable of inflicting tremendous damage to friend or foe under tenuous control. Whether damaged or under electronic attack, fail-safes could be installed that force autonomous systems to cease lethal activities if they lose contact with human controllers, but other nations may not follow the same rules of engagement. If warbots become essential to military success, denying robots a form of autonomous lethal authority could be a war loser.

Placing humans in-the-loop or on-the-loop may not drastically reduce the risk of jus in bello violations by warbots because of human tendencies to defer to machine judgment when information is limited or in stressful situations. Human warbot controllers, like the drone operators of today, might make lethal decisions based on information supplied to them through remote sensors or even nonvisual readouts in a bandwidth-constrained area. Future warbots may even possess an ability to tersely explain their process for reaching a decision to an operator, but this will not necessarily remove the risk of human deference to machines if operators lack combat or system experience.¹⁵ As Dr. John Hawley notes in his



A standard missile 3 (SM-3) is launched from the Aegis Combat System-equipped Arleigh Burke class destroyer USS *Decatur* 22 June 2007 during a Missile Defense Agency ballistic missile flight test. Minutes later the SM-3 intercepted a ballistic-missile threat target launched from the Pacific Missile Range Facility, Barking Sands, Kauai, Hawaii. It was the first time such a test was conducted from a ballistic missile defense equipped-U.S. Navy destroyer. The Aegis Combat System has an automatic mode that allows fire-control computers to identify and engage targets with no humans in the loop. (Photo courtesy of U.S. Navy)

paper "Patriot Wars: Automation and the Patriot Air and Missile Defense System," "an automated system in the hands of an inadequately trained crew is a de facto fully automated system."¹⁶ Other researchers have noted the same kind of trust in machines during emergencies. In a 2015 study at Georgia Tech, participants followed a robot guide to a conference room and observed the robot make numerous navigational mistakes on the trip there. After arriving at the conference room, the researchers set off the building's fire alarm. Despite observing the robot make navigational mistakes just minutes prior, every participant in the study chose to follow the robot guide out of the potentially burning building.¹⁷

Despite U.S. reservations about the employment of future lethal autonomous machines, air defense systems have already entered a transitory period where "smart" weapons possess the capability and authority to make lethal decisions. The Patriot Air Defense System, Aegis Combat System, and Close-In Weapon System all feature automatic modes that place target classification and engagement systems in the hands of fire control computers.¹⁸ All three have seen combat, and each system has misclassified and engaged a friendly or neutral target resulting in injuries to allies or noncombatants.¹⁹ In many of these instances, fratricide or collateral damage resulted from human misunderstanding of air defense system algorithms or operators placing too much trust in fire control system target identification. These errors are not limited to U.S. systems. The 2015 shoot down of Malaysian Air Flight 17 by a Ukrainian separatist-controlled SA-11 likely typifies the same kind of human-machine teaming error.²⁰

Future lethal autonomous systems will encounter significantly more complex situations than the automated air defense systems of today because they will be self-mobile and deployed across a greater range of environments (air, land, sea, space). Lethal autonomous weapon systems will also interact with a greater diversity of threats and more importantly, interact with human beings in a greater number of scenarios.²¹ Additionally, LAWS may operate at considerable distance from their human controllers and supervisors. This mobility will give warbots the ability to place themselves in unforeseen circumstances, magnifying the likelihood that robots will encounter situations outside of their logical depth, beyond the scope or capacity of their internal algorithms to solve.

Warbot Difficulty in Following Just War Principles

Recognizing and employing lethal force against potential targets indiscriminately on the battlefield takes significantly less cognitive calculation than employing force justly. It does not matter if the combatant is human or robotic. Human soldiers concerned with following the jus in bello principles of discrimination, proportionality, and military necessity draw on vast stores of cultural and experiential knowledge to avoid inflicting unnecessary suffering on noncombatants. They also unconsciously develop heuristics and quickly integrate seemingly unrelated pieces of information into their decision-making. Machines generally struggle to create their own rules of thumb or relate disparate data as well as human beings and may not attain human-like capacity to do so.²² The sections below explain three major jus in bello principles and why warbots may struggle to adhere to them.

Discrimination. Once war begins, able-bodied soldiers and combatants are subject to attack at any time.²³ When an individual takes up arms, he or she loses the civilian immunity that prevents being targeted by lethal force. As these combatants lose their civilian rights, they gain others in turn. They are now free to attack other combatants. Combatants only regain their immunity from attack by physically losing their ability to harm others (by becoming severely wounded) or permanently resigning their role; if a soldier's term of enlistment ends and he or she returns home, that person is no longer a legitimate military target.

If tasked with target identification, LAWS will likely struggle to discriminate between combatants

and noncombatants outside of conventional war, just as human soldiers do. In the multitude of conflicts raging across the earth (e.g., Syria, Ukraine, Iraq, and Afghanistan), large portions of the combatant population exist outside of conventional military structures. Homogenous uniforms or identical weapon systems do not identify them as warfighters. Today, object recognition software allows machines to detect and classify objects by comparing them to pictures held in internal databases—warbots can classify an AK-47 or a T-72 tank and conclude that the individuals associated with it are a threat.

Software, however, cannot make these same conclusions if the combatant is not holding a traditional weapon system or an object with dual uses—an insurgent using a cellphone to detonate an improvised explosive device closely resembles a bystander sending a text. It takes the processing power of a human mind to do that. To make this kind of judgment, a young infantryman has to evaluate intentions from the insurgent's visual cues and conduct pattern-of-life analysis of civilians and insurgent activity in the area, rapidly pulling information from past experiences and training. All of this occurs in a matter of seconds. Lethal autonomous weapon systems will not perform this task or others that require the complicated heuristic models of biological minds well. Google's self-driving car sometimes struggles to get through a four-way stop because it is unable to read the intentions of the human drivers around it.24

These microanalyses are important. If a LAWS comes across a wounded combatant, how will it determine if he or she is incapacitated? How will a machine accept a surrender from weapon-carrying soldiers so that prisoners of war are treated fairly? Algorithms can be written that supply the machines with some capacity to do this, but they will likely fall far short of human levels for the foreseeable future. Warbots may identify blatant combatants with ease but struggle to separate noncombatants and irregular soldiers.²⁵

Proportionality. In addition to discrimination in the use of force, just war theory demands that "the destruction needed to fulfill a military goal is proportional to the good of achieving it."²⁶ In short, to use force appropriate to the target—no need to drop a five hundred-pound bomb on an AK-47 wielding insurgent when a 5.56 mm round will do. Failure to abide by this principle increases the likelihood of unnecessary collateral damage and excessive loss of combatant lives.

Warbots will struggle to make proportionality decisions without human input.²⁷ Software developers can engineer a LAWS to follow preprogrammed rules of engagement: restrictions on the caliber of weapons employed in urban areas, no fire zones, etc., and commanders can apply all kinds of geographic boundaries and fire support control measures to retard excessive use of force and fratricide by robots. But, even with these restrictions, LAWS will inevitably fall short if human controllers fail to predict what kinds of control measures are needed. In mercurial combat, this will severely test human foresight.

Estimating collateral damage, an essential element of proportionality, requires a multitude of predictions and assumptions. Human minds just make it appear



A materials researcher examines experimental data on the Autonomous Research System (ARES) artificial intelligence planner. The system, developed by the Air Force Research Laboratory, uses artificial intelligence to design, execute, and analyze experiments at a faster pace than traditional scientific research methods. (Photo courtesy of the Department of Defense)

easy. If an infantryman takes fire from a building in an urban environment, he or she automatically considers the function of that structure and the potential presence of noncombatants before retaliating. Humans unconsciously examine the structural material, design, and building signage among other inputs before comparing them to an archive of architectural information developed over a lifetime. Human minds have massive storage capacity and incredible recall capabilities that allow them to quickly retrieve seemingly unrelated pieces of information and apply them to new problems.²⁸ Current computer systems do not.

Proportionality judgments require more than structural identification. Without strain, human beings analyze the effects of time of day, cultural settings, and days of the week on civilian patterns of life—you are unlikely to kill many civilians by bombing a church at midnight on Wednesday during most of the year but could kill scores if it was Christmas Eve. The ability of LAWS to make these kinds of conclusions under combat timelines remains suspect at best. Programmers can write algorithms to approximate these human decision cycles, but accounting for all possible variations will be nearly impossible. Differences between robotic and

> human sensory capabilities will pose difficulties.²⁹ Furthermore, tactical-level warbots may not possess sufficient memory to accommodate the needed databases or computing power to cross-reference them without becoming cost-ineffective or tying into a nearby cloud.

> **Military necessity.** Targets of violence in war must be legitimate, attacked to accomplish an objective that aids in the defeat of an enemy force. Even the murder of combatants for purposes separate from a military objective can be unnecessary.³⁰ This requirement poses difficulty for warfighters at both the strategic and tactical levels. The course of war is unpredictable. It is often hard to evaluate if an attack is necessary or not.

Lethal autonomous weapon systems will have trouble making

this determination as well. Imagine that a LAWS locates a terrorist plotting to drive a vehicle-borne improvised explosive device (VBIED) into a U.S. embassy sometime in the next few days. The terrorist has his eight-year-old child next to him. Is it a military necessity for the LAWS to attack him now? Or should it wait until the terrorist is alone even if it might allow the VBIED attack to occur? Decisions such as these are challenging even if the greater context that necessitates action is understood. A human decision-maker could make probabilistic calculations regarding how likely the terrorist would be to successfully execute the VBIED attack or how successful ground forces would be in capturing the terrorist if the warbot did not strike. A tactical-level robot will probably not have access to the external information or the broader situational awareness to make an estimate.

Human soldiers struggle to fulfill their jus in bello responsibilities to perfection now. Warbots with subhuman intelligence will likely struggle even more. If human controllers and supervisors hope to prevent LAWS from violating the rules of discrimination, proportionality, and military necessity, they will need to be thoroughly versed in the parameters that guide their warbots' selection and engagement of targets. Soldiers will be asked to assume responsibility for the decisions automated weapon systems make on the battlefield. Air defense systems and fire-and-forget munitions carry similar burdens now, but the greater autonomy of self-maneuvering warbots will make these burdens greater.

Though imperfect, military working dogs offer the closest analogy. When a handler releases a military working dog on the battlefield, he does not fire a weapon. He deploys a weapon system that makes its own engagement decisions. To ensure the effectiveness of the military working dog in combat and prevent ethical violations and fratricide, the military working dog goes through months of intense training alongside its handler. The handler relies on this training to guide the animal through simple tasks and learns to understand the strengths and weaknesses of his or her military working dog.³¹ Warbot handlers might undergo a similar train-up with LAWS in virtual and physical environments but may have greater difficulty in using the experience to gain insight into warbot behaviors. Despite differences in scale, shared senses (e.g., sight, hearing, smell), thought patterns, and emotions (e.g., fear, excitement) provide human and canine with an evolutionary baseline to understand one another. A human being can watch a dog interact with his environment and at least partially understand its intent, motivations, and thoughts. Stark differences between human and machine perception and cognition will frustrate human efforts to develop this same level of understanding with warbots.

Understanding Warbot Decision-Making

Algorithms—processes defined by coders for computers to solve problems-are the building blocks of software. Software developers may combine hundreds or thousands of algorithms to create a software program that makes decisions without any awareness from the user. Users manipulate a graphic-user interface that executes scripts from a higher-level programming language, which is then translated into binary inputs for the computer's central processing unit. These levels of abstraction hide the machine's actual decision-making process from the user and the numerous heuristics, assumptions, and flaws that coders intentionally or unwittingly include within programs. As AI experts increasingly use machine learning and deep learning techniques to allow machines to craft their algorithms, these biases may become more opaque. Machine learning techniques allow algorithms to create generalizations and analyze patterns based on the evaluation of training data; the algorithms effectively teach themselves through trial and error. Machine learning reliance on training data allows computer scientists to unconsciously insert significant biases into algorithms through their selection of data inputs.³² For instance, a facial recognition algorithm that used high school yearbooks from the American Midwest as training data might struggle to identify ethnic minorities when employed because of a dearth of examples in its training data. In a commercial context, this kind of bias resulting from human selection of training data is embarrassing. With LAWS, it is deadly.

Imagine designing an algorithm to identify an armed military-age male. Programmers could accomplish this in a variety of ways. The software could use sensors to measure the height of a potential target to confirm adulthood, look for facial hair or measure shoulder-to-waist ratios to assess gender, and estimate the potential target's body mass to compare to averages for adults and children. Sensor accuracy permitting, the program could even measure objects held by the potential target to determine if they match specifications for weapons stored in the warbot's memory.



Airman 1st Class Colten Carey, 23d Maintenance Squadron (MXS) precision-guided munitions technician, examines Joint Direct Attack Munitions (JDAM) 11 January 2018 at Moody Air Force Base, Georgia. The JDAM guidance kit converts standard bombs into "smart" munitions that are capable of autonomously engaging designated targets. (Photo by Airman Eugene Oliver, U.S. Air Force)

Each of these determinations would require the execution of algorithms that compare and apply data captured by the warbot's sensors to internally held databases and predeveloped algorithms. This warbot algorithm could fail in its task to identify armed military-age males in various ways: (1) The robot could lack the precision to make the determination; its visual sensors might be unable to identify objects or measure heights beyond a certain distance, and the algorithm might force it to make a decision anyway. (2) The warbot's databases could contain insufficient information; the target, for instance, might be holding a model of firearm that was not in the algorithm's training data or it could have modifications that make it unrecognizable to the machine. (3) The programmer's heuristics or assumptions for making the decision could possess flaws or make false assumptions. If the algorithm attempted to identify objects as female or male based on height and body mass, it could run into issues if it failed to factor in the potential target's nationality and/ or ethnicity—the average Scandinavian woman is taller than the average Chinese male and likely weighs an equal amount.³³ Alternatively, a weapon-identifying algorithm might classify a hunting rifle as a tool and an AK-47 as a weapon based on the judgments of a software engineer, even though both are lethal.

Almost every algorithm contains the inherent biases and value judgments of the people that create them and the training data they select.³⁴ Soldiers tasked with monitoring and controlling lethal autonomous systems will need to be familiar with the biases and value judgment imparted into weapon system software to avoid accidents and law of armed conflict violations.

Understanding these value judgments and recognizing when they will affect operational performance may be easier than it sounds. The algorithms that power LAWS will be incredibly dense. Warbots may rely on millions of lines of code to operate, and each of them will have embedded biases and assumptions from human software engineers working on them. The volume of code will make complete understanding of the intricacies of every algorithmic decision and its effect on a warbot in diverse wartime situations very hard. And, the use of machine learning techniques will potentially deepen this obscuration.

Military members may experience even greater difficulty in understanding these algorithms because they lack a technical background as rigorous as the software engineers who will create the LAWS they operate. Warbot controllers may not even have access to the algorithms because of security concerns; governments could guard them as national or proprietary secrets.³⁵ This obscuration could be troublesome. Current concepts of jus in bello place responsibility for ethical employment of lethal autonomous systems on combatants and commanders, even though the decisions made by acquisition officials and software engineers will have equal or greater impact on warbot behavior.

Extending Jus in Bello Responsibilities

As the majority of combatants providing guidance to LAWS will be incapable of fully understanding how their warbots actually make lethal decisions, the software engineers and acquisition officials responsible for placing this equipment in their hands must bear partial ethical responsibility for the jus in bello violations of warbots on the battlefield. Western legal systems do not hold defendants criminally responsible if they are mentally incapable of understanding their crime, so society should not hold combatants singularly responsible for decisions LAWS make on the battlefield that fall beyond their ability to comprehend.³⁶ Instead, society holds individuals responsible for events in proportion to their ability to shape them. While combatants will always be responsible for jus in bello violations that result from operator error, system designers and acquisition officials should bear blame proportionate to their responsibility for the violation.³⁷

Acquisition officials have ethical responsibilities to author requirements for lethal autonomous systems that minimize the likelihood that they will violate the law of armed conflict. Specifically, government representatives will need to establish rigorous standards for training algorithms and the testing of robotic systems in simulated and real-world environments to preclude likely errors. Defense contractors, in turn, have a duty to provide warbots that meet those requirements to the highest technological level possible and to inform governments of known vulnerabilities in algorithmic decision-making and sensor perception. Companies that manufacture self-driving cars hold responsibility for accidents caused by design flaws.³⁸ Therefore, companies that produce war machines with negligent flaws should rightfully face civil and potentially criminal penalties. Finally, both civilian and military leaders hold responsibilities to decree policies and rules of engagement that minimize combatant opportunities to place lethal autonomous systems in situations outside of their analytic and logical depth.

Regardless of the difficulty, military leaders must make extensive efforts to prepare soldiers to understand the risks of employing lethal autonomous systems in predictable combat situations.³⁹ Future militaries fielding LAWS may find that many of the cost advantages of automation dissipate in light of increased operator training costs and the testing required to fine-tune AI algorithms. The automation of battlefield tasks performed by humans will probably forge militaries with smaller numbers of combatants who possess greater technical aptitude.

Militaries should consider creating special classes of soldiers similar to joint tactical air controllers who receive enhanced training on the complexities of autonomous systems. Commanders could potentially restrict control of warbots to these soldiers outside of emergency situations until artificial intelligence matures further.⁴⁰ A host of methodologies and specialized personnel exist to guide commanders in the employment of airstrikes and long-range fires to include targeteers, joint tactical air controllers, and forward observers. Militaries deploying LAWSs should consider creating analogous positions. Robotic experts could accompany field commanders and help them make difficult decisions regarding the employment of warbots on the battlefield just as targeteers help military leaders determine the risk of collateral damage during air strikes. Militaries may also need to form algorithm test and evaluation cells that cultivate training data and create simulations that allow LAWS to be calibrated for specific theaters of combat and rules of engagement.

Methodologies for estimating the risk of employing LAWS could also aid commanders in their decisions to deploy warbots and establish norms for their use.

During the next few decades, combatants will enter an age where their weapon systems will bear increasing portions of responsibility for their ability to successfully exercise force against legitimate military targets. Ethical concepts and policies need to advance in lockstep to ensure that technological changes do not result in ethical lapses by distributing jus in bello responsibilities proportionately on the human actors involved.

Notes

1. "A Digital Jet for the Modern Battlespace," Lockheed Martin (website), accessed 31 January 2018, <u>https://www.f35.com/</u> about/life-cycle/software.

2. Stuart Armstrong and Kaj Sotala, "How We're Predicting Al—or Failing To," in *Beyond Al: Artificial Dreams*, eds. Jan Romportl et al. (Pilsen, Czech Republic: University of West Bohemia, 2012), 52–75.

3. Noel E. Sharkey, "The Evitability of Autonomous Robot Warfare," *International Review of the Red Cross* 94, no. 886 (2012): 788–89.

4. One example of a current room-sized computer is the National Supercomputing Center in Wuxi, China, for more, see their website, <u>http://www.nsccwx.cn/wxcyw/</u>. IBM developed a quantum computer that researchers can use to further develop the potential of such vast computing power. Cade Metz, "IBM Is Now Letting Anyone Play with its Quantum Computer," *Wired* (website), 4 May 2016, accessed 31 January 2018, <u>https://www.wired.</u> com/2016/05/ibm-letting-anyone-play-quantum-computer/.

5. History.com Staff, "Invention of the PC," History.com, 2011, accessed 31 January 2018, <u>http://www.history.com/topics/</u>inventions/invention-of-the-pc.

6. Patrick Lin, George Bekey, and Keith Abney, "Autonomous Military Robotics: Risk, Ethics, and Design," Prepared for the Office of Naval Research, 20 December 2008, accessed 12 February 2018, <u>http://ethics.calpoly.edu/onr_report.pdf</u>.

7. Greg Allen and Taniel Chan, "Artificial Intelligence and National Security" (Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School, July 2017), 17–18, accessed 31 January 2018, <u>https://www.belfercenter.org/</u> publication/artificial-intelligence-and-national-security.

8. Department of Defense Directive 3009.09, Autonomy in Weapon Systems (Washington, DC: U.S. Government Publishing Office [GPO], 21 November 2012); Joseph Breecher, Heath Niemi, and Andrew Hill, "My Droneski Just Ate Your Ethics," War on the Rocks (website), 10 August 2016, accessed 31 January 2018, <u>https://warontherocks.com/2016/08/</u> my-droneski-just-ate-your-ethics/.

9. Groups such as the Campaign to Stop Killer Robots have already begun lobbying for a worldwide ban on lethal weapon systems. For more on this group, see https://www.stopkillerrobots.org/.

10. C. Peter Chen, "Japan's Refusal of Washington Treaty, 19 Dec 1934," WWII Database (website), accessed 31 January 2018, https://ww2db.com/battle_spec.php?battle_id=45.

11. Wade Boese, "U.S. Withdraws from ABM Treaty; Global Response Muted," *Arms Control Today* (website), July/August 2002, accessed 31 January 2018, <u>https://www.armscontrol.org/</u> act/2002_07-08/abmjul_aug02. 12. Jules Hurst, "Intervention and the Looming Choices of Autonomous Warfighting," War on the Rocks (website), 25 August 2016, accessed 31 January 2018, <u>https://warontherocks.com/2016/08/intervention-and-the-looming-choices-of-autono-mous-warfighting/</u>.

13. "In the loop" references situations where human beings must give warbots an order to exercise lethal force. Human controllers are in the kill-chain. "On the loop" references situations where human beings supervise robots as they exercise lethal force.

14. Thomas Gibbons-Neff, "'We Don't Have the Gear': How the Pentagon is Struggling with Electronic Warfare," *Washington Post* (website), 9 February 2016, accessed 31 January 2018, <u>https://www.washingtonpost.com/news/checkpoint/</u> wp/2016/02/09/we-dont-have-the-gear-how-the-pentagon-isstruggling-with-electronic-warfare/.

15. Dave Gershgorn, "We Don't Understand How AI Make Most Decisions, So Now Algorithms are Explaining Themselves," Quartz (website), 20 December 2016, accessed 12 February 2018, <u>https://qz.com/865357/we-dont-understand-how-ai-make-</u>most-decisions-so-now-algorithms-are-explaining-themselves/.

16. John K. Hawley, *Patriot Wars: Automation and the Patriot Air and Missile Defense System*, Voices from the Field series (Washington, DC: Center for a New American Security, January 2017), accessed 31 January 2018, <u>https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Report-EthicalAutonomy5-PatriotWars-FINAL.pdf.</u>

17. Paul Robinette et al., "Overtrust of Robots in Emergency Evacuation Scenarios" (lecture, The 11th ACM/IEEE International Conference on Human Robot Interaction, Christchurch, New Zealand, 7 March 2016), 101, accessed 12 February 2018, <u>http://</u> ieeexplore.ieee.org/document/7451740/.

18. John Canning, "A Concept of Operations for Armed Autonomous Systems" (PowerPoint presentation, 3rd Annual Disruptive Technology Conference, Washington, DC, 6–7 September 2010), accessed 12 February 2018, <u>http://www.dtic.mil/</u> ndia/2006/disruptive_tech/canning.pdf.

19. Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]), "Report of the Defense Science Board Task Force on Patriot System Performance Report Summary" (report, Washington, DC: Department of Defense, January 2005); George C. Wilson, "Navy Missile Downs Iranian Jetliner," *Washington Post* (website), 4 July 1988, accessed 12 February 2018, <u>http://www. washingtonpost.com/wp-srv/inatl/longterm/flight801/stories/</u> july88crash.htm; A. J. Plunkett, "Iwo Jima Officer Killed In Firing Exercise," *Daily Press* (website), 12 October 1989, accessed 12 February 2018, <u>http://articles.dailypress.com/1989-10-12/</u> news/8910120238_1_iwo-jima-ship-close-in-weapons-system.

20. Simon Tomlinson, "Russian Missile Killed Pilots and Cut Jet in Half but Passengers Could Have Been Conscious for up to a Minute as Plane Plunged, Reveals Official Report into MH17 Downed over Ukraine," *Daily Mail* (website) 13 October 2015, accessed 12 February 2018, <u>http://www.dailymail.co.uk/news/</u> article-3270355/Doomed-flight-MH17-shot-Russian-BUK-missile-fired-rebel-held-territory-eastern-Ukraine-Dutch-investigators-set-rule.html.

21. Andrew Ilachinski, "AI, Robots, and Swarms Issues, Questions, and Recommended Studies" (Arlington, VA: CNA, January 2017), vi–vii, accessed 12 February 2018, <u>https://www.cna.org/</u> CNA_files/PDF/DRM-2017-U-014796-Final.pdf.

22. Sharkey, "The Evitability of Autonomous Robot Warfare," 789.

23. Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 4th ed. (1977; repr., New York: Basic Books, 2006), 138.

24. Matt Richtel and Conor Dougherty, "Google's Driverless Cars Run into Problem: Cars with Drivers," *New York Times* (website), 1 September 2015, 12 February 2018, <u>https://www.nytimes.</u> <u>com/2015/09/02/technology/personaltech/google-says-its-not-</u> the-driverless-cars-fault-its-other-drivers.html.

25. Sharkey, "The Evitability of Autonomous Robot Warfare," 788–89.

26. Brian Orend, *The Morality of War* (Peterborough, Ontario: Broadview Press, 2006), 119.

27. Sharkey, "The Evitability of Autonomous Robot Warfare," 789. 28. Jeneen Interlandi, "New Estimate Boosts the Human Brain's Memory Capacity 10-Fold," *Scientific American* (website), 5 February 2016, accessed 12 February 2018, <u>https://www.scientificamerican.com/article/new-estimate-boosts-the-human-brain-s-memorycapacity-10-fold/.</u>

29. Ruth A. David and Paul Nielsen, "Report of the Defense Science Board Summer Study on Autonomy" (Washington, DC: OUSD[AT&L], June 2016), 14, accessed 12 February 2018, <u>http://</u>www.dtic.mil/dtic/tr/fulltext/u2/1017790.pdf.

30. Orend, The Morality of War, 119.

31. Department of the Army Pamphlet 190-12, *Military Working Dog Program* (Washington, DC: U.S. GPO, 30 September 1993), 24.

32. Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," *Science* 356, no. 6334 (2017): 183-86.

33. Ian Langtree, "Height Chart of Men and Women in Different Countries," Disabled World (website), last updated 19 December 2017, accessed 12 February 2018, <u>https://www.</u> <u>disabled-world.com/artman/publish/height-chart.shtml</u>.

34. Felicitas Kramer, Kees van Overveld, and Martin Peterson, "Is There an Ethics of Algorithms?," *Ethics and Information Technology* 13, no. 3 (2010): 251.

35. Tom Simonite, "For Superpowers, Artificial Intelligence Fuels New Global Arms Race," *Wired* (website), 8 September 2017, accessed 12 February 2018, <u>https://www.wired.com/story/</u> for-superpowers-artificial-intelligence-fuels-new-global-arms-race/.

36. "The Infancy Defense: Criminal Law Basics," The 'Lectric Law Library, accessed 12 February 2018, <u>http://www.lectlaw.com/mjl/</u>cl032.htm.

37. Aaron M. Johnson and Sidney Axinn, "The Morality of Autonomous Robots," *Journal of Military Ethics* 12, no. 2 (2013): 131.

38. Anjali Singhvi and Karl Russell, "Inside the Self-Driving Tesla Fatal Accident," *New York Times* (website), 12 July 2016, accessed 12 February 2018, <u>https://www.nytimes.com/interac-</u> tive/2016/07/01/business/inside-tesla-accident.html.

39. This is easier said than done. Commanders would likely only be found liable for failing to prepare troops for a combat scenario if it existed as part of a standardized training curriculum.

40. David and Nielsen, "Report of the Defense Science Board," 38.