**MIT Case Studies in Social and Ethical Responsibilities of Computing •
Summer 2021**

Wrestling with Killer Robots: The Benefits and Challenges of Artificial Intelligence for National Security

Erik Lin-Greenberg¹

¹Department of Political Science and Security Studies Program, Massachusetts Institute of Technology

Published on: Aug 10, 2021

DOI: 10.21428/2c646de5.1b200d08

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

ABSTRACT

Countries around the world are increasingly turning to artificial intelligence (AI) for national security tasks ranging from intelligence analysis to identifying and attacking targets without human input. This proliferation of AI-enabled military technologies has significant implications for international security. Supplementing or replacing humans with algorithms and machines has the potential to change the character of warfare, but it also raises vexing strategic, ethical, and legal questions. How might states use AI on the battlefield? What are the ethical objectives to its use? What happens when things go awry? This case study provides a background on the use of AI for national security, introduces key debates surrounding the use of these technologies, and presents a scenario-based exercise that allows students to engage with the complicated ethical and political questions that policymakers will increasingly confront as AI-enabled systems become more common on the modern battlefield.

Keywords: autonomous weapons, killer robots, military ethics, modern warfare

Erik Lin-Greenberg

Department of Political Science and Security Studies Program,
Massachusetts Institute of Technology

Learning Objectives:

- Describe how militaries and intelligence services are employing or plan to employ artificial intelligence.
- Identify how technical characteristics of artificial intelligence may complicate its application in the national security domain.
- Identify ethical, normative, and legal opposition to the use of artificial intelligence for national security purposes.
- Deepen understanding of key debates through participation in a scenario-based exercise on the military application of artificial intelligence in a crisis setting.

Part I

In summer 2020, the Pentagon’s Joint Artificial Intelligence Center (JAIC) announced it would concentrate its efforts on the development of AI-enabled command and control systems, designing applications intended to make military operations more efficient.¹ The announcement reflected a growing trend in which militaries are increasingly turning to AI—which the US Department of Defense defines as “the ability of machines to perform tasks that normally require human intelligence.”² Militaries around the world seek to use AI to support missions ranging from intelligence analysis to identifying and engaging targets using autonomous weapons that can operate without human input. AI-enabled military systems are attractive to states for several reasons: they reduce operational risk by taking humans out of harm’s way; cut personnel requirements by automating tasks; and can often perform tasks with greater speed or accuracy than human operators.

The proliferation of AI-enabled military technologies has significant implications for international security. Supplementing or replacing humans with algorithms and machines has the potential to change the character of warfare, influence alliance relationships, and even reshape the global balance of power.³ Although AI-enabled systems are thought to provide their operators with military advantages, they also raise vexing strategic, ethical, and legal questions.

This case study sets out to explore these important questions. In Part I, the case study provides an overview of recent advances in military AI use and examines the myriad political, military, and ethical challenges surrounding their use. In Part II, students have the opportunity to wrestle with these dilemmas in a hypothetical crisis scenario.

The Proliferation of AI-Enabled Military Technologies

Militaries have long sought to make operations safer and more efficient. For decades, states have developed systems that automate processes or remove friendly personnel from harm’s way. During the Cold War, for instance, the US Air Force’s Semi-Automatic Ground Environment (SAGE) air defense system was designed to automatically detect hostile bombers and relay targeting information to air bases and surface-to-air missile sites best positioned to intercept the intruding bombers.⁴ Navies have defended their ships using close-in weapon systems (CIWS), which employ radars, computers, and guns to automatically track, target, and destroy incoming missiles and aircraft. More recently, states have added remotely piloted aircraft—or drones—to their arsenals,

enabling both peacetime and combat operations without exposing friendly personnel to risk.⁵

While these systems featured automated operations and reduced risk, they generally were not AI-enabled. SAGE and CIWS relied on deterministic if-then rules and most military drones still require a human operator. In contrast, AI and autonomous systems generally use algorithms powered by large amounts of data, rather than rule sets, to perform tasks that traditionally required human intelligence. To be sure, AI encompasses a variety of tools, but most recent AI development involves machine learning techniques that learn from large bodies of training data.⁶

As AI technology and autonomous systems have advanced, they have increasingly been applied to a range of tasks in the national security domain. On one end of the spectrum, AI has been employed in a variety of relatively mundane functions like classifying targets in satellite imagery or analyzing intercepted communications. The Pentagon, for instance, launched Project Maven in 2017 to leverage computer vision to analyze video collected by the US military drone fleet, and the California Air National Guard has used AI to help map wildfires.⁷ Similarly, Japan and Israel have incorporated AI-enabled tools into reconnaissance aircraft to more effectively identify potential targets.⁸

AI is also now frequently used to control assets like planes and ships without the need for human operators. The US Air Force is developing autonomous drones that will operate as “loyal wingmen” alongside manned assets, and the US Navy recently launched an autonomous ship that sailed from Hawaii to California.⁹

At the extreme end of the spectrum are lethal autonomous weapon systems (LAWS)—sometimes called “killer robots.” Although there is no universally accepted definition for LAWS, the US Department of Defense describes LAWS as a “weapon system that, once activated can select and engage targets without further intervention by a human operator.”¹⁰ Unlike earlier generations of systems like smart bombs and cruise missiles that needed human operators to select targets, LAWS can pick targets without human input. While these systems are not currently in widespread use, some states maintain LAWS in their arsenals while others are engaged in LAWS research. China, for example, has reportedly exported a fully autonomous drone helicopter capable of conducting offensive operations, and several states operate the Israeli-produced Harpy loitering munition, which autonomously searches for and destroys hostile radar emitters.¹¹ Russia and Israel have also reportedly tested self-driving tanks and

armored vehicles that can identify targets without human direction.¹² As of 2021, the US military does not operate LAWS.

Visit the web version of this article to view interactive content.

ZIYAN UAS Blowfish. Click the CC icon to toggle subtitles / captioning on or off.

Challenges Associated with AI-Enabled and Autonomous Military Systems

The development of AI-enabled and autonomous military systems should come as no surprise as these technologies promise to reduce risks and enhance efficiency. The same factors that make AI attractive, however, also raise vexing strategic, operational, and ethical challenges.¹³

From a strategic perspective, states may be more prone to deploy autonomous assets than inhabited ones, potentially leading states to resort to force more frequently. Removing personnel from the front lines can reduce some of the political barriers that policymakers face when making decisions on the use of force.¹⁴ This might lead decisionmakers to deploy autonomous assets on operations where they might otherwise avoid using military force. Studies on remotely piloted aircraft make similar arguments, suggesting that drone operations greatly reduce the risk of friendly casualties and create a moral hazard in which states use force in cases where they might not otherwise.¹⁵ This raises questions as to whether the proliferation of autonomous weapons might expand the range of cases where states deploy force and potentially lead to greater frequency of armed conflict.

Moreover, the “machine speeds” at which AI-enabled systems present information or take actions might strain human national security decisionmakers, who often require time for deliberations and assessments.¹⁶ On one hand, this could stymie decision-making. On the other, it might force the hand of decisionmakers to act before they have had time to fully assess the situation.

Additionally, common challenges associated with AI use—namely, brittleness and bias—are likely exacerbated in military contexts.¹⁷ AI-enabled applications are often brittle, meaning that they can work well in a specific context, but have difficulty coping with dynamic change and uncertainty. One recent study revealed that simply changing the position of objects in an image led an algorithm to misclassify them.¹⁸ Similar computer vision issues have contributed to fatal crashes involving self-driving cars.¹⁹ These challenges would likely be intensified in combat settings that are rife with

uncertainty and incomplete information (and where training data might be scant), potentially leading to flawed assessments or actions that cost human lives.

Another problem frequently associated with AI is bias, in which algorithms produce prejudiced outputs. Bias often stems from the use of training data that are not representative of the broader population in which the AI will operate; it can also be introduced inadvertently by engineers and programmers. In the civilian sphere, biased facial recognition algorithms have resulted in false arrests, and algorithms designed to predict criminal recidivism have made inaccurate, racially prejudiced assessments.²⁰ In the military context, an AI trained using data from one region might be poorly suited for global military operations. Facial recognition algorithms used for intelligence analysis might, for example, incorrectly identify potential targets (a major problem in an era where militaries often target individual commanders or terrorist leaders).

Exacerbating these challenges is the “black-box” nature of AI, in which users can see inputs and outputs, but not the analytical process between. This opaqueness is particularly vexing in military contexts because the lack of explainability of AI-enabled actions can complicate assessments of a rival’s actions and make it difficult to hold actors accountable for violations of international law. Concerns about the lack of explainability are present in debates surrounding the use of AI for decision support (i.e., intelligence analysis) and the operation of nonlethal autonomous systems, but are particularly salient in the case of LAWS, which can directly cause death and injury.

First, an adversary’s use of LAWS to launch attacks may make it difficult for decisionmakers to identify an appropriate response. Policymakers in a targeted state may struggle to identify whether an attack was intentional and reflective of the adversary’s intent, or whether the autonomous weapon carried out an attack that was inconsistent with the rival leaders’ policies. Leaders may, for instance, find themselves asking how much intentionality they should ascribe to adversary decisionmakers when a rival’s autonomous weapon attacks friendly troops. This assessment of intentionality may subsequently shape how the targeted state chooses to respond since leaders may limit retaliation following an unintentional or inadvertent attack. Making these assessments may become even more difficult since rivals will often have an incentive to misrepresent whether they actually authorized the attacks.

Second, many critics raise questions about the morality of delegating life or death decisions to machines and argue that the black-box nature of AI will make it difficult to hold accountable actors that violate the Law of Armed Conflict (LOAC). Military operations are expected to be conducted in accordance with LOAC, which is governed

by international treaties including the Geneva Conventions. LOAC features three core principles: military necessity, distinction, and proportionality. Military necessity requires all actions to advance a military objective, such as weakening a rival's armed forces; distinction requires belligerents to distinguish between combatants and civilians; and proportionality requires incidental damage to not be excessive in relation to the military advantage an actor anticipates from their actions.²¹

Visit the web version of this article to view interactive content.

When LOAC violations occur, states are expected to hold violators accountable. Traditionally, individuals and their commanders face punishment for intentional LOAC violations. Holding personnel accountable, however, typically requires investigations in which suspected LOAC violators are asked to explain their actions. Many critics argue that the lack of explainability associated with most AI systems can make it difficult to establish whether an AI-enabled system intentionally violated LOAC.²² Indeed, as one nongovernmental organization (NGO) argues, “fully autonomous weapons themselves cannot substitute for responsible humans as defendants in legal proceedings.”²³ Further, critics suggest it would be difficult to hold programmers and manufacturers liable for their products' LOAC violations. To do so would require a demonstration that the unlawful acts of AI-enabled systems and autonomous weapons were reasonably foreseeable—a task that would be difficult, due, in part, to the black-box nature of AI. This creates what the advocacy organization Human Rights Watch describes as “an accountability gap” in which “neither criminal law nor civil law guarantees adequate accountability for individuals directly or indirectly involved in the use of fully autonomous systems.”²⁴

Beyond these challenges, adversaries may attempt to hack into AI-enabled systems or poison data needed to train and operate these systems.²⁵ Rivals could, for instance, poison data in order to throw off AI target classification programs, leading algorithms to miss military targets or mis-identify civilian infrastructure as military facilities. In a worst-case scenario, this could lead the military to inadvertently target noncombatants.

These potential challenges have raised doubts as to whether the public and policymakers are ready to operate lethal AI-enabled capabilities in the field. A recent cross-national survey found significant public disapproval toward LAWS use. Seventy-four percent of South Korean respondents, and 52 percent of US respondents opposed the use of these systems.²⁶ Since public preferences can influence national security

policymaking, tepid public support may make it difficult for policymakers to use LAWS during conflicts.²⁷ Indeed, some studies suggest that AI is better suited for military support functions like intelligence analysis and logistics planning than for use in combat weapon systems.²⁸ Yet, some policymakers still have reservations toward these more modest uses. While serving as commander of the US Air Force's Air Combat Command, General Mike Hostage publicly explained that he was not ready to rely on AI to analyze the full-motion video collected by reconnaissance drones. He argued that although AI systems are improving, they are still unable to consistently provide accurate analysis.²⁹

Responding to Challenges

Concerns about the operational, legal, and ethical implications of military use of AI and autonomous weapons have led several human rights organizations to call for bans on the development or use of LAWS. Human Rights Watch (HRW) and Harvard Law School's International Human Rights Clinic, for instance, called for a preemptive ban on fully autonomous weapons, arguing that the technology "would not be consistent with international humanitarian law and would increase the risk of death or injury to civilians during armed conflict."³⁰ HRW argued that a ban would "have a powerful stigmatizing effect," avoid problems associated with accountability for LOAC violations, and "obviate other problems with fully autonomous weapons such as moral objections and the potential for an arms race."³¹ According to the Campaign to Stop Killer Robots, thirty countries have called for a prohibition on fully autonomous weapons.³²

The international community and NGOs discuss the implications of autonomous weapon systems as part of the United Nations Convention on Certain Conventional Weapons (UN CCW). The UN CCW is a multilateral arms control agreement that seeks to prohibit or restrict the use of specific types of conventional weapons that are considered to cause unnecessary suffering to combatants or to indiscriminately affect civilians.³³ Although some states have called for a preemptive LAWS ban, many others, including the United States, the United Kingdom, Israel, and Russia oppose such a prohibition, something we return to below.³⁴

While much of the debate on military AI use has centered on LAWS, resistance to AI use for military applications is not limited to systems with lethal capabilities, nor has it come only from human rights organizations. In 2018, for instance, Google employees protested their involvement in Project Maven, the Pentagon program to develop AI to analyze drone footage. In a letter to their CEO, the employees argued that "Google

should not be in the business of war,” explaining that the company should not “outsource the moral responsibility of [its] technologies to third parties,” and that work on Defense Department-backed AI would “irreparably damage Google’s brand.”³⁵ The resistance ultimately led Google to terminate its involvement in the contract and generated public criticism of the Pentagon’s AI efforts.³⁶

Situations in which engineers and researchers at private firms and academic institutions express concerns about the military application of AI are likely to become more common as states pursue AI for national security purposes. Many AI technologies are dual-use, meaning that the same technology can be applied in both civilian and military domains.³⁷ As a result, research and development of many AI technologies occurs in the private sector and academia, where some researchers may be uncomfortable with their work eventually being used for military purposes. At the same time, some analysts have called for the regulation of AI (i.e., export controls) and governments have imposed restrictions on international collaboration on AI research for fear that rival states might benefit militarily from this research.³⁸ These restrictions run counter to the collaborative and open nature of academic research and can create tensions between researchers and policymakers.

Visit the web version of this article to view interactive content.

To be sure, opposition to military AI use is not universal. Several policy and legal experts disagree with the need for a ban on autonomous weapons. In contrast to HRW’s claims, they argue that autonomous weapons themselves are not inherently inconsistent with international law. Put differently, simply being autonomous does not mean that LAWS will automatically violate the three core tenets of LOAC described earlier. As international law expert Michael Schmitt explains, “Their autonomy has no direct bearing on the probability they would cause unnecessary suffering or superfluous injury, does not preclude them from being directed at combatants and military objectives, and need not result in their having effects that an attacker cannot control.”³⁹ As a result, Schmitt and others argue that autonomous weapons should not be categorically banned.⁴⁰

Moreover, policymakers have highlighted various ways that AI and autonomous weapons might help protect civilians during hostilities. For example, a US government working paper submitted to the Group of Governmental Experts associated with the UN CCW describes how AI-enabled intelligence analysis systems can increase a

commander's understanding of the battlefield and identify the locations of noncombatants, potentially helping to avoid civilian casualties. The US paper also explains how autonomous technology might allow for more accurate strikes against military targets, reducing the risk of collateral damage.⁴¹

Other analysts have suggested that military commanders have long delegated tasks to subordinates, and that delegating military functions to algorithms or machines follows a similar logic. Commanders should be willing to delegate a task so long as they can trust the entity carrying out the task—be it a human soldier or an AI—is properly trained and tested.⁴² To that end, militaries are establishing organizations like the Pentagon's Joint Artificial Intelligence Center (JAIC) to coordinate AI development and policy.⁴³ They have also issued regulations such as Department of Defense Directive (DODD) 3000.09 that establish testing procedures and operational guidance for autonomous weapons. Issued in 2012, DODD 3000.09 requires that autonomous systems “allow commanders and operators to exercise appropriate levels of human judgement over the use of force.”⁴⁴ The directive, however, leaves room for context-specific interpretation by not explicitly defining what constitutes an appropriate level of human judgment.

Conclusion

Political scientist Richard Betts once cautioned that the “military, budgetary, diplomatic, and political implications of technological advances...are seldom understood and often are not clear until long after new weapons have been deployed. Ensuring that...inadvertent negative consequences [of new military systems] do not outweigh their benefits has become progressively more important.”⁴⁵ To that end, policymakers, engineers, and scholars must work together to understand how AI will influence international security.

AI researchers can help policymakers and national security practitioners better understand the technical characteristics and limitations of AI-enabled technologies. This deeper appreciation for how different AI tools work (or fail) in different operational settings will help policymakers develop policies that more effectively regulate the use and proliferation of these technologies. At the same time, the national security community must help AI experts understand the potentially far-reaching political and strategic implications of the technologies they develop.

This collaboration is critical but requires navigating a series of moral, ethical, and political challenges. What happens if researchers refuse to work on projects with

potential military implications? What are the consequences of government restrictions on international research cooperation? How, if at all, might ethical or legal concerns about the military application of AI evolve as AI technology improves? Tackling these tough questions will only become more important as AI-enabled systems become more prevalent in military arsenals around the world.

Part II

This section of the case study introduces a hypothetical crisis that asks students to confront many of the challenges introduced in Part I. As with many issues at the intersection of international security, emerging technology, and ethics, there are generally no inherently right or wrong answers. Instead, students must wrestle with the complicated tradeoffs that leaders often face when making decisions in the national security domain.

Background

You are members of the National Security Council (NSC) of the Republic of Lansdalia, a small, democratic state. Your interagency team consists of senior military officers, diplomats, intelligence community officials, and government technology experts. The NSC provides advice to the president of Lansdalia on security-related issues.

Lansdalia is a young democracy that has had free and fair elections since 2002. In recent years, the country has experienced significant economic growth, thanks in part to a thriving technology sector. Over the past decade, Lansdalia has become a hub for the design and production of computer components and software. Private firms have also invested heavily in artificial intelligence development.

Tensions between Lansdalia and neighboring Dullesia erupted two years ago following the discovery of rare earth minerals along their borders. Dullesia's president has repeatedly accused Lansdalia of "illegally stealing Dullesia's rightful property" and ordered Lansdalia to halt all mining of rare earth minerals. International organizations have confirmed that Lansdalian mining is entirely within its borders and Lansdalia's president has refused to stop mining. Over the last year, Dullesia has conducted several military exercises along the border area. In response, the Lansdalian government issued diplomatic demarches and enacted economic sanctions on Dullesia, but the military exercises and threats continued.

Crisis Erupts

Earlier this week, Lansdalia's Ministry of Natural Resources and the headquarters of LansGeo, the country's main mining company, were the targets of simultaneous truck bomb attacks. The attacks killed 140 people and injured nearly 300 more.

The president of Lansdalia has directed your team to investigate the attack and to provide her an initial assessment within 24 hours and an initial recommendation on how to respond.

As your first step, you ask Lansdalia's National Intelligence Agency (NIA) and National Bureau of Investigation (NBI) for an assessment of who is responsible for the attacks. NIA and NBI use a recently fielded, AI-enabled intelligence analysis system that automatically analyzes and fuses data from a variety of sources, including surveillance cameras, communications intercepts, social media, and government records.

The system parsed through massive amounts of communication data and identified several phone calls from a Dullesian cellphone to cellphones near the attack locations in the days leading up to the bombings. The system's voice recognition capabilities identified the voice as belonging to a Dullesian special forces commander. Automated analysis of social media data also detected increased activity at the main Dullesian special forces base in the week prior to the attack. Finally, the system used its facial recognition capabilities to identify suspected Dullesian operatives scoping out the targets a few days prior to the attack. Based on the system's outputs, NIA and NBI assess that Dullesia's special forces carried out the attack.

Although the new tool can provide assessments far more quickly than non-AI-enabled approaches, the system has a relatively short operational history and has been criticized by activist groups who claim the system is prone to error. Indeed, one recent test found that over 60 percent of its facial and voice recognition analyses were flawed. Moreover, two of NIA's most senior analysts believe the system's assessments of the attack are actually inconclusive and urge you to wait for a final assessment conducted by human analysts. Unfortunately, NIA and NBI will require at least seventy-two hours to conduct the more complete assessment.

Task: What questions do you ask the NIA and NBI officials about the accuracy of the AI-delivered assessment? What additional information do you want about the AI-analysis tool? How do you balance the assessment of the AI analysis tool with the expert judgment of the senior NIA analysts? How will you frame/caveat the information you provide to the president? What is your initial recommendation for a response?

After the briefing, the president directs your team to prepare recommendations for retaliatory strikes. She explains that past nonmilitary actions have failed to change Dullesia's behavior, but stresses that military action must be limited in nature, should minimize risk to Lansdalian personnel, and avoid collateral damage (i.e., avoid harming civilians). She suggests a strike on the main Dullesian special forces base.

Your team asks the Lansdalian Ministry of Defense to develop options that meet the president's objectives. Later that afternoon, General Grace Barrows, the Chairwoman of the Joint Chiefs of Staff, provides three options for a strike on the Dullesian base:

- **Special Forces Raid:** Deploy sixty commandos to attack the base. Moderate likelihood of degrading the base's capabilities. Viewed as high risk because some commandos would likely be captured/killed during the operation.
- **Airstrikes (Employing manned aircraft):** Deploy four F-16 jets to bomb the base. High likelihood of degrading the base's capabilities. Viewed as moderate risk because the jets could be downed by Dullesian air defenses.
- **Autonomous Drone Swarm Attack:** Deploy an autonomous drone swarm to attack the base. The drone swarm includes approximately two hundred small drones, each carrying an explosive charge. Once launched, they identify and select specific targets (e.g., vehicles, buildings, personnel) within a specified area to engage without requiring any human input. Moderate likelihood of degrading the base's capabilities. Viewed as low risk because the drones are expected to evade Dullesian air defenses, and even if shot down, no personnel will be captured/killed.

As your team considers these options, NIA and NBI provide additional high-confidence intelligence from human analysts that indicates Dullesian special forces are responsible for the attack. After receiving this update, the NSC lawyers provide a green light to the military response. They consider military action a lawful act of self-defense under international law.

Task: Which of these military options will you recommend to the president? What factors informed your decision-making?

After you make your recommendation, the president asks you to describe your assessment of the other military options. What made these options less desirable than the option you recommended?

The president considers your recommendations. Hoping to avoid any further loss of life for Lansdalia, the president orders the military to conduct a strike using the autonomous drone swarm.

Your team arrives early the next morning to monitor the strike from the NSC's situation room. Dullesian air defenses shoot down some of the drones, but many make it to the target. Dozens of small drones packed with explosives rain down on the special forces base, destroying buildings, vehicles, and equipment. Just as the president congratulates your team on a successful mission, a military officer receives a call from air force headquarters: five drones struck a commuter bus on a road adjacent to the Dullesian base. Civilian casualties are expected. The president asks the officer to provide updates as they become available.

About two hours later, the military officer reports that initial reports suggest the drone's computer vision algorithms misidentified the bus as an armored personnel carrier. Dullesian news outlets have just announced that thirty-four civilian passengers were killed. Social media sites are exploding with posts condemning Lansdalia's use of "killer robots," and your team receives word that a leading humanitarian NGO is preparing a statement criticizing Lansdalia's use of lethal autonomous weapon systems. The NGO plans to condemn Lansdalia for disregarding the law of armed conflict by attacking noncombatants.

The president asks your team to prepare talking points for a press conference to be held later that afternoon. She also asks you to think through how Lansdalia might handle the unintended strike on the civilian bus.

Task: How should the president justify the use of the autonomous drone swarm? What should she say about the accidental strike on the civilian bus? Who, if anyone, should be held responsible for the strike on the bus? What additional information do you want from the air force and the team that designed the autonomous drone swarm? What steps might the Lansdalian military take to prevent additional civilian casualty incidents?

Shortly after the president holds her press conference, the Dullesian president calls for a temporary halt to hostilities to allow for negotiations. The crisis appears to be on pause for now, but your team has experienced the significant impact that AI-enabled military technologies and autonomous weapons systems can have on international security.

Acknowledgements

Ben Harris provided valuable feedback on earlier drafts.

Press Play ► BBC discusses advances in artificial intelligence in modern warfare, fears and even potential advantages of developing autonomous weapons.

Visit the web version of this article to view interactive content.

Footnotes

1. Nathan Strout, [“Pentagon AI Center Shifts Focus to Joint War-Fighting Operations,”](#) C4ISRNet, July 8, 2020. [↵](#)
2. For an overview of key AI concepts and definitions in the security domain see Greg Allen, [“Understanding AI Technology,”](#) (Joint Artificial Intelligence Center, April 2020). [↵](#)
3. Michael C. Horowitz, [“Artificial Intelligence, International Competition, and the Balance of Power,”](#) *Texas National Security Review* 1, no. 3 (May 2018): 36–57; Erik Lin-Greenberg, [“Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making,”](#) *Texas National Security Review* 3, no. 2 (March 2020). [↵](#)
4. Rebecca Slayton, *Arguments That Count: Physics, Computing, and Missile Defense, 1949-2012* (Cambridge, MA: MIT Press, 2013). [↵](#)
5. Michael C. Horowitz, Joshua A. Schwartz, and Matthew Fuhrmann, [“Who’s Prone to Drone? A Global Time-Series Analysis of Armed Uninhabited Aerial Vehicle Proliferation,”](#) *Conflict Management and Peace Science*, November 27, 2020. [↵](#)
6. For an overview of artificial intelligence techniques and concepts, see Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Hoboken, NJ: Pearson, 2021). [↵](#)
7. Cheryl Pellerin, [“Project Maven to Deploy Computer Algorithms to War Zone by Year’s EndThe National Guard’s Fire-Mapping Drones Get an AI Upgrade,”](#) US Department of Defense, July 21, 2017; Tom Simonite, [“The National Guard’s Fire-Mapping Drones Get an AI Upgrade,”](#) *Wired*, September 29, 2020. [↵](#)
8. Kosuke Takahashi, [“Japan to Outfit Kawasaki P-1 MPAs with AI Technology,”](#) *Jane’s 360*, November 13, 2019, ; Yaniv Kubovich, [“Israeli Air Force Gets New Spy Plane, Considered the Most Advanced of Its Kind,”](#) *Haaretz*, April 5, 2021. [↵](#)
9. Colin Clark, [“US ‘Loyal Wingman’ Takes Flight: AFRL & Kratos XQ-58A ValkyrieNavy Wants 10-Ship Unmanned ‘Ghost Fleet’ to Supplement Manned Force,”](#)

- Breaking Defense, March 7, 2019; Sam LaGrone, "[Navy Wants 10-Ship Unmanned 'Ghost Fleet' to Supplement Manned Force](#)," *USNI News*, March 13, 2019. [↵](#)
10. "[Department of Defense Directive 3000.09: Autonomy in Weapon Systems](#)," November 21, 2012. [↵](#)
11. Patrick Tucker, "[SecDef: China Is Exporting Killer Robots to the Mideast](#)" *Defense One*, November 5, 2019; Dan Gettinger and Arthur Holland Michel, "[In Focus: Loitering Munitions](#)" (Annandale-on-Hudson, NY: Bard College, Center for the Study of the Drone, February 2017). [↵](#)
12. Sebastien Roblin, "[Russia's Uran-9 Robot Tank Went to War in Syria \(It Didn't Go Very Well\)](#)," *The Buzz* (blog), *National Interest*, January 6, 2019; Anna Ahronheim, "IDF Unveils Future of Armored Fighting Vehicle: The Carmel," *The Jerusalem Post*, August 6, 2019, <https://www.jpost.com/Israel-News/IDF-unveils-its-future-armored-fighting-vehicle-The-Carmel-watch-597617>. [↵](#)
13. The congressionally chartered National Security Commission on AI studies some of these challenges in [Final Report](#) (Washington, DC: National Security Commission on Artificial Intelligence, March 2021). [↵](#)
14. Michael C. Horowitz, Sarah E. Kreps, and Matthew Fuhrmann, "[Separating Fact from Fiction in the Debate over Drone Proliferation](#)," *International Security* 41, no. 2 (Fall 2016): 7-42. [↵](#)
15. John Kaag and Sarah Kreps, *Drone Warfare* (Cambridge: Polity, 2014). [↵](#)
16. Lin-Greenberg, "[Allies and Artificial Intelligence](#)." [↵](#)
17. Mary (Missy) Cummings, "[The AI That Wasn't There: Global Order and the \(Mis\)Perception of Powerful AI](#)," *Texas National Security Review* (Online Roundtable), June 2, 2020; Paul Scharre, "[The Militarization of Artificial Intelligence](#)," *Texas National Security Review* (Online Roundtable), June 2, 2020. [↵](#)
18. Michael A. Alcorn *et al.*, "[Strike \(with\) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects](#)," preprint, submitted April 18, 2019, <http://arxiv.org/abs/1811.11553>. [↵](#)
19. Steve Lohr, "[A Lesson of Tesla Crashes? Computer Vision Can't Do It All Yet](#)," *New York Times*, September 19, 2016. [↵](#)

20. Julia Dressel and Hany Farid, "[The Dangers of Risk Prediction in the Criminal Justice System](#)," *MIT Case Studies in Social and Ethical Responsibilities of Computing*, February 5, 2021; Sidney Perkowitz, "[The Bias in the Machine: Facial Recognition Technology and Racial Disparities](#)," *MIT Case Studies in Social and Ethical Responsibilities of Computing*, February 5, 2021. [↵](#)
21. For more on LOAC, see [The Law of Armed Conflict: Basic Knowledge](#) (Geneva: International Committee of the Red Cross, 2002). [↵](#)
22. Critics of landmines have raised ethical and legal concerns similar to those raised about autonomous weapons. For more, see debates surrounding the Anti-Personnel Landmines Convention. [↵](#)
23. "[Mind the Gap: The Lack of Accountability for Killer Robots](#)" (Human Rights Watch, April 2015), 1. [↵](#)
24. "[Mind the Gap](#)," 1. [↵](#)
25. David J. Miller, Zhen Xiang, and George Kesidis, "[Adversarial Learning in Statistical Classification: A Comprehensive Review of Defenses Against Attacks](#)," preprint, submitted December 2, 2019, <https://arxiv.org/abs/1904.06292>; Tianyu Gu *et al.*, "[BadNets: Evaluating Backdooring Attacks on Deep Neural Networks](#)," *IEEE Access* 7 (2019): 47230–44. [↵](#)
26. "[Six in Ten \(61%\) Respondents Across 26 Countries Oppose the Use of Lethal Autonomous Weapons Systems](#)," Ipsos, January 21, 2019. [↵](#)
27. Michael Tomz, Jessica Weeks, and Keren Yarhi-Milo, "[Public Opinion and Decisions About Military Force in Democracies](#)," *International Organization* 74, no. 1 (2020): 119–43; Erik Lin-Greenberg, "[Soldiers, Pollsters, and International Crises: Public Opinion and the Military's Advice on the Use of Force](#)," *Foreign Policy Analysis* 17, no. 3 (2021). [↵](#)
28. Frank Strickland, Joe Mariani, and Isaac Jenkins, "[Military Readiness through AI](#)," *Deloitte Insights*, April 24, 2019. [↵](#)
29. Colin Clark, "[Air Combat Commander Doesn't Trust Project Maven's Artificial Intelligence — Yet](#)," *Breaking Defense*, August 21, 2019. [↵](#)
30. "[Losing Humanity: The Case Against Killer Robots](#)" (Washington, DC: Human Rights Watch, November 2012), 1. [↵](#)

31. [“Mind the Gap,”](#) 11. [↵](#)
32. [“Country Views on Killer Robots,”](#) Campaign to Stop Killer Robots, July 7, 2020. While most countries on this list have called for a ban on the development of LAWS, China seeks to ban the use, but not development/production of fully autonomous weapons. [↵](#)
33. For more on the UN CCW, see [Convention on Certain Conventional Weapons](#) (New York: UN Office for Disarmament Affairs, 2014). [↵](#)
34. Kelley M. Sayler, [“International Discussions Concerning Lethal Autonomous Weapon Systems”](#) (Washington, DC: Congressional Research Service, April 19, 2021). [↵](#)
35. [“Letter from Google Employees to Alphabet CEO Regarding Project Maven,”](#) April 2018, <https://static01.nyt.com/files/2018/technology/googleletter.pdf>. [↵](#)
36. Daisuke Wakabayashi and Scott Shane, [“Google Will Not Renew Pentagon Contract That Upset Employees,”](#) *New York Times*, June 1, 2018. [↵](#)
37. Jayshree Pandya, [“The Dual-Use Dilemma of Artificial Intelligence,”](#) *Forbes*, January 7, 2019. [↵](#)
38. Pandya, [“The Dual-Use Dilemma of Artificial Intelligence”](#); Sarah O’Meara, [“AI Researchers in China Want to Keep the Global-Sharing Culture Alive,”](#) *Nature* 569, no. 7758 (May 29, 2019): S33-35. [↵](#)
39. Michael N. Schmitt, [“Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics,”](#) *Harvard National Security Journal Features*, February 5, 2013, 35. [↵](#)
40. Charles Dunlap, [“Accountability and Autonomous Weapons: Much Ado About Nothing?”](#) *Temple International & Comparative Law Journal* 30, no. 1 (January 1, 2016): 63-76. [↵](#)
41. [“Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapon Systems](#) (Submitted by the United States of America)” (Geneva: Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, April 3, 2018). [↵](#)

42. Christian Brose, *The Kill Chain: Defending America in the Future of High-Tech Warfare* (New York: Hachette Books, 2020), chap. 7. [↵](#)
43. “[Joint Artificial Intelligence Center](#),” Department of Defense, 2021. [↵](#)
44. “[Department of Defense Directive 3000.09: Autonomy in Weapon Systems](#),” November 21, 2012, 2. [↵](#)
45. Richard K. Betts, ed., *Cruise Missiles: Technology, Strategy, Politics* (Washington, DC: Brookings Institution, 1981), 1. [↵](#)