AI and Autonomy in War: Understanding and Mitigating Risks

Lawrence Lewis

August 2018

CNA
ANALYSIS & SOLUTIONS

DOP-2018-U-018296-Final

CNA's Occasional Paper series is published by CNA, but the opinions expressed are those of the author(s) and do not necessarily reflect the views of CNA or the Department of the Navy.

Photography Credit: Department of the Army

Approved by: **August 2018**

Mark Geis
Executive Vice President
Center for Naval Analyses

This work was performed under Federal Government Contract No. N00014-16-D-5003.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
08-2018	Final	

4. TITLE AND SUBTITLE
(U) AI and Autonomy in War: Understanding and Mitigating Risks

5a. CONTRACT NUMBER
N00014-16-D-5003

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0605154N

6. AUTHOR(S)
Lawrence Lewis

5d. PROJECT NUMBER
R0148

5e. TASK NUMBER
D180.00

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Center for Naval Analyses
3003 Washington Blvd
Arlington, VA 22201

8. PERFORMING ORGANIZATION REPORT NUMBER
DOP-2018-U-018296-Final

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of the Chief of Naval Operations
(OPNAV N81)
Navy Department Pentagon
Washington, DC 20350-2000

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report examines commonly held concerns about AI and autonomy in war. We find that the overall premises for these concerns are either out of step with the current state of the technology, or they do not consider the way military systems are actually used. These concerns are not spurious—they can lead to much-needed debates and discussions regarding ethical issues of this emerging technology. However, the real risk in a military context (expressed in operational outcomes such as civilian casualties and fratricide) is low from these common concerns. We then examine factors associated with the current and near-future state of the technology that could introduce operational risk if not mitigated, and we identify ways to mitigate them. Finally we note that AI and autonomy provide opportunities, not just risks. States should look for opportunities to reduce risk and improve the conduct of war.

15. SUBJECT TERMS
AI, Autonomy, Technology, Safety, Ethics, Law, LOAC, Acquisition, Training, Operations

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Knowledge Center/Robert Richards
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	SAR	52	703-824-2123

Executive Summary

Significant advances in artificial intelligence (AI) over the past decade have changed our way of life, and the impacts of AI are only expected to accelerate. At the same time, the idea of adapting AI, and the related attribute of autonomy, to military applications has created considerable controversy. There are strong concerns about these technologies, even speculation that they could lead to the end of the world. Important questions to consider are: how do the actual risks of weaponizing this technology compare to those commonly discussed? And are states and the international community effectively managing these risks?

This report examines commonly held concerns about AI and autonomy in war, as reported in the media or voiced in international venues. We find that the overall premises for these concerns are either out of step with the current state of the technology, or they do not consider the way military systems are actually used (which is as part of a larger process for delivering the use of force). These concerns are not spurious—they can lead to much-needed debates and discussions regarding ethical issues of this emerging technology. However, the real risk in a military context (expressed in operational outcomes such as civilian casualties and fratricide) is low from these common concerns.

We also examine factors related to the operational use of AI and autonomy. We identify factors associated with the current and near-future state of the technology that could introduce operational risk if not mitigated, and we identify ways to mitigate them. These factors can be blind spots for militaries, which may tend to focus on developing a capability without considering the enablers necessary for the safe and effective use of AI and autonomy. We also present a framework for international and domestic discussions about the primary applicable risks of AI and autonomy in war. Finally we note that AI and autonomy provide opportunities, not just risks. States should look for opportunities to reduce risk and improve the conduct of war.

Recommendations

We offer a number of recommendations for mitigating risks from the technologies of AI and autonomy being used in war. The first set is for nations considering use of the

technology, to enable them to better address clear and present risks. Recognizing the need for additional and productive dialogue regarding AI and autonomy in war, the second set addresses needed dialogues to discuss the risks of that technology and how to mitigate them.

Recommendations for countries considering the use of AI and autonomy in war:

- Militaries interested in leveraging AI and autonomy should address risk factors impacting operational safety, including operational considerations, institutional development, and law and policy. These risk factors should be addressed to both improve effectiveness and promote safety.
- National policies for AI and autonomy should consider and address the risk of AI increasing the opacity of targeting decisions, akin to the practice of signature strikes.
- In addition to mitigating risk factors, states should also be looking for opportunities for using AI and autonomy to improve the conduct of war

Recommendations for needed dialogues to discuss the risks of the use of AI and autonomy in war:

- Separate out the two cases of general and narrow AI, since the two are distinct, carrying very different sets of risks and having different timelines for development.
- Hold deliberate, inclusive debates concerning AI and autonomy in war, requiring arguments to be supported with reason and evidence, and allowing different views to be fairly exchanged.
- Discuss the risk of AI increasing the opacity of targeting decisions and steps that can be taken to avoid this.
- International venues should consider risk factors identified in this report as a way to frame discussions on how to pursue safety of AI and autonomy in war. Those discussions should include operational considerations, institutional development, and law and policy.
- Consider potential opportunities for using AI and autonomy to improve the conduct of war.

Contents

Introduction: Understanding the Risks of AI in War 1

Artificial Intelligence: Definitions and Functions 4

Specific functions obtainable through AI 6

Commonly Voiced Concerns about AI and Autonomy 7

Concern: AI will destroy the world 7
Concern: AI and lethal autonomy are unlawful (per the Martens Clause) 9
Public polls 11
Expert opinion 11
Deliberative debate 12
Authoritative sources 12
Summary 12
Concern: Lack of accountability 13
Concern: Lack of discrimination 14
Is discrimination possible? 14
How good is good enough: what is an ethical standard for discrimination? 14
What about slaughterbots? 16
Final thoughts on general AI 17

Examining the Risks of AI and Autonomy in War 18

Military operations 19
Institutional development 21
Materiel development 21
Developing capabilities 22
Test and evaluation considerations 23
Non-materiel development 24
Law and policy 26
Assessments 28
Summary of AI and autonomy risks 29

AI for Good in War 31

Summary **34**
Recommendations 35

References **37**

List of Figures

Figure 1. Accelerating pace of computing power 5
Figure 2. Framework for comprehensive human control over the use of force 19

This page intentionally left blank.

List of Tables

Table 1. Common AI functions and descriptions 6
Table 2. Mitigating the risks of AI and autonomy in components of human control..... 30

This page intentionally left blank.

Introduction: Understanding the Risks of AI in War

The past few years have seen an exponential increase in artificial intelligence (AI) technology, defined as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings." The technology has been described as the engine of the Fourth Industrial Revolution.[1] The past two years have seen dramatic advances in the ability of machines to complete complex tasks and match or exceed human performance. One goal of AI and related information technology developments is the attribute of autonomy, in which systems can make decisions and actions with less or no need for human intervention.

The opportunities of this technology have not gone unnoticed by nations seeking a military edge. The US Department of Defense (DOD) seeks to capitalize on the rapid development of AI technology in a "Third Offset" strategy, which focuses its asymmetric advantages in AI for military applications.[2] Vladimir Putin declared that whatever country dominates AI will rule the world. China recently released a whole-of-nation strategy for AI that resembles the race to the moon in the 1960s. This strategy aims to create a $150 billion AI industry that will be the most advanced in the world, and it plans to leverage that industry for a military advantage in an unprecedented example of civil-military cooperation.

The idea of adapting AI to military applications has created considerable controversy. Google recently withdrew its support of the US DOD's Project Maven, which uses AI to scan video from drones and make suggestions for classifying objects as people, buildings, or vehicles. Similarly, hundreds of scientific organizations and thousands of individuals gathered in July 2018 at the annual

[1] "The first, from 1760 to 1840, brought the steam engine, railroads, and machine manufacturing. The second, from about 1870 to 1914, gave us electricity and mass production. The third, often called the digital revolution, encompassed the last decades of the 20th century and produced semiconductors, computers, and the Internet." David Barno and Nora Bensahel, "War in the Fourth Industrial Revolution," War on the Rocks, July 3, 2018.

[2] Larry Lewis, *Insights for the Third Offset: Addressing Challenges of Autonomy and Artificial Intelligence in Military Operations,* CNA Research Memorandum DRM-2017-U-016281-Final. Sept. 2017.

International Joint Conference on Artificial Intelligence and signed a pledge calling for laws to pre-emptively ban lethal autonomous weapons.[3] The United Nations (UN) Convention on Certain Conventional Weapons (CCW) has spent four years discussing the ethical, legal, and operational considerations of lethal autonomous weapon systems (LAWS), including whether weapon systems operating autonomously (without a human operator) should be allowed to use lethal force.[4]

In international discussions, states have consistently expressed support for approaches that mitigate potential risks of LAWS. These approaches have included requiring compliance with International Humanitarian Law (IHL) and setting requirements that mandate human control over lethal functions, which could be articulated in a political declaration or as part of IHL.

In managing risk, it is important to note that IHL, such as the Geneva Conventions and the CCW Protocols, is intended to "reflect a reasonable and pragmatic balance between the demands of military necessity and those of humanity."[5] This practical approach acknowledges that tragedies can still happen on the battlefield, but it calls for military forces to take steps to mitigate those effects by choosing actions required to achieve a legitimate purpose, conducting them in discriminating and proportional ways, and avoiding actions that are explicitly prohibited. Thus, IHL requires military forces to balance risk in every action, considering both risk to the mission and risk to noncombatants. Mitigating risk from lethal autonomous weapon systems used in an armed conflict should be understood in this context. This is different from the use of such systems in a law enforcement context, which is governed by International Human Rights Law in concert with domestic law, a subject not treated here but worthy of further exploration.

One time-tested principle of risk management is as follows: to effectively mitigate risk, first identify what the chief risks are. Otherwise, the efforts made to mitigate risk may not match the actual sources of risk. This principle is illustrated in the following example:

> *When Tony Hayward became CEO of BP in 2007, he vowed to make safety his top priority. Among the new rules he instituted were the requirements that all*

[3] Cameron Jenkins, "AI Innovators Take Pledge Against Autonomous Killer Weapons," NPR, July 18, 2018, https://www.npr.org/2018/07/18/630146884/ai-innovators-take-pledge-against-autonomous-killer-weapons.

[4] The CCW is properly referred to as 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.

[5] ICRC, "International humanitarian law and the challenges of contemporary armed conflicts," *International Review of the Red Cross* 89, no. 867, (September 2007).

> *employees use lids on coffee cups while walking and refrain from texting while driving. Three years later, on Hayward's watch, the Deepwater Horizon oil rig exploded in the Gulf of Mexico, causing one of the worst man-made disasters in history. A US investigation commission attributed the disaster to management failures that crippled "the ability of individuals involved to identify the risks they faced and to properly evaluate, communicate, and address them.*[6]

With regard to concerns about AI and autonomy in warfighting, is the international community focused on putting lids on coffee cups when a disaster is brewing unseen? How do the actual risks of weaponizing this technology compare to those commonly discussed? After providing some definitions and examples of AI, this report looks at commonly cited risks of AI and autonomy in war and discusses how applicable those risks really are. It then examines other risks that, like the oil rig example, are not commonly considered but can have a significant deleterious effect on the performance and safety of systems using AI and autonomy. The report concludes with recommendations for reframing international and domestic discussions around the primary applicable risks of AI and autonomy in war.

[6] Robert S. Kaplan and Anette Mikes, "Managing Risks: A New Framework," *Harvard Business Review*, June 2012, https://hbr.org/2012/06/managing-risks-a-new-framework.

Artificial Intelligence: Definitions and Functions

Artificial intelligence has been defined as the ability of a system "devoted to making machines intelligent," in which intelligence is that "quality that enables an entity to function appropriately and with foresight in its environment."[7] One subcomponent of AI is machine learning (ML), which refers to a set of techniques "designed to detect patterns in, and learn and make predictions from data."[8] These techniques allow machines to learn from examples and conduct tasks without explicit programming. The recent success of machine learning techniques is due largely to dramatic increases in computing power and the availability of large datasets to serve as training data for machine learning algorithms. The effectiveness of the machine learning approach depends on not just an effective algorithm design but also the quality and robustness of its training data. The power of AI, and ML in particular, is seen in everyday applications, such as the following:

- Transportation: ML powers navigation apps such as Google Maps and Waze as well as ridesharing software including Uber and Lyft.
- Banking and fraud detection: Banks can identify potentially fraudulent patterns and raise alerts concerning questionable transactions. ML is also used to interpret handwriting in mobile check deposits.
- Making recommendations: Shopping sites (e.g., Amazon), social media sites (e.g., Facebook), and entertainment sites (e.g., Netflix) analyze user preferences and suggest other content based on observed patterns.
- Virtual personal assistants: Alexa, Siri, and other applications feature voice recognition and the ability to provide requested content in conversation with users.
- Improved medical diagnoses: ML can improve the accuracy and timeliness of diagnoses from medical scans.

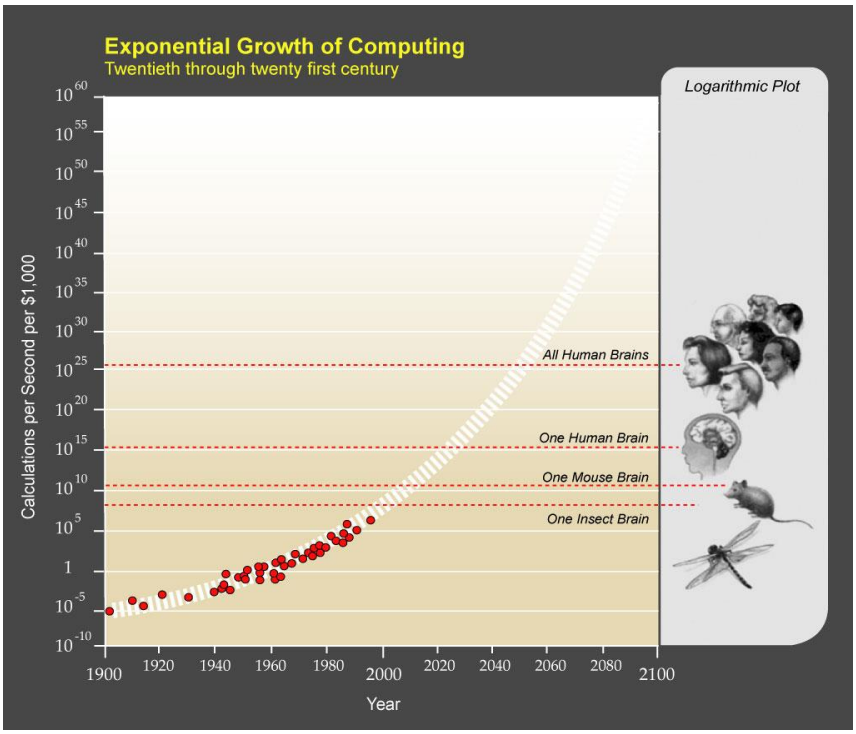
[7] Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (New York: Cambridge University Press, 2009).

[8] Andrew Ilachinski, *AI, Robots, and Swarms: Issues, Questions, and Recommended Studies,* CNA Research Memorandum DRM-2017-U-014796-Final. Jan. 2017.

- Tagging content in media: Facebook and other social media platforms identify and tag media based on recognized content, including facial recognition.[9]

The rapid adoption of AI, including ML, in the commercial sector leverages the convergence of the acceleration of computing power and the emerging availability of large datasets. For example, Figure 1 below shows the trend in computing power advances over time. In addition, over 90 percent of all data in the world was created in the last two years.[10] These two accelerating trends contribute to increasing opportunities to leverage AI for many applications.

Figure 1. Accelerating pace of computing power



Source: https://upload.wikimedia.org/wikipedia/commons/d/df/PPTExponentialGrowth of_Computing.jpg.

[9] Sabine Hauert, "Eight ways intelligent machines are already in your life," *BBC News*, April 25, 2017, https://www.bbc.com/news/uk-39657382; Gautam Narula, "Everyday examples of artificial intelligence and machine learning," Techemergence, June 28, 2018, https://www.techemergence.com/everyday-examples-of-ai/.

[10] Bernard Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," *Forbes*, May 21, 2018, https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#45fa594b60ba.

Specific functions obtainable through AI

The many commercial applications of AI typically derive from a discrete set of AI functions that find patterns, outliers, or optimal solutions in large datasets. Some of these commonly employed functions of AI are described in Table 1 below.

Table 1. Common AI functions and descriptions

AI function	Description
Automate tasks	Automate routine functions and employ man-machine teaming to reduce the time and burden on personnel for performing tasks (e.g., taking over administrative functions, improved medical diagnoses)
Process complex or large datasets	Enable analysis of more data, additional data sources (e.g., voice and image recognition)
Predict behavior	Learn from past data to anticipate possible future behavior (e.g., Google Maps predicting traffic times)
Flag anomalies or events of interest	Identify indicators of potential problems or events of interest and create alerts (e.g., banks predicting potentially fraudulent transactions)
Data tagging and error correction	Recognize content and create tags so that data can be more effectively or efficiently exploited; also improve data quality (e.g., Facebook automated image tagging)

These AI functions have advantages over human effort, including the following:

- Speed: Faster action and exploitation of data compared to human responses
- Volume: Enables analysis of more data and additional data sources
- Persistence: Can operate continuously 24/7
- Accuracy: Optimizes outcomes and improves human decision-making
- Cost-effectiveness: Can automate routine functions and employ man-machine teaming to reduce the burden on personnel for performing tasks

In the following section, we examine commonly expressed concerns about AI and autonomy and discuss whether these concerns are warranted considering the state of technology and the context of war.

Commonly Voiced Concerns about AI and Autonomy

In *2001: A Space Odyssey*, astronauts work in concert with an AI-driven computer, the HAL 9000, during a space mission. The mission ends in tragedy as the computer kills most of its crew. David Bowman, the only surviving human, survives by disabling the super-intelligent machine.[11] This picture of AI, also seen in movies such as the *Terminator* series, *RoboCop*, and *Ex Machina*, portrays a life-and-death conflict between machines and humans. As seen in articles and polling data, this conflict exists not only in the pages of sci-fi stories but also in the common concerns of the public. In an armed conflict context, the use of machines in war can be particularly worrying, evoking a host of concerns often cited by proponents of a ban on such technology or voiced by the public in opinion polls. This section examines some of the most commonly cited and consequential concerns, starting with the most consequential concern of all: the end of humanity.

Concern: AI will destroy the world

Both the media and international discussions have pointed to the potential of AI to lead to apocalyptic outcomes. Elon Musk has been a vocal leader in this regard, stating concerns about "a fleet of artificial intelligence-enhanced robots capable of destroying mankind."[12] He has also stated: "I think we should be really concerned about AI….I think we need to be proactive in regulation instead of reactive. Because I think by the time we are reactive in AI regulation, it's too late."[13] Similarly, Stephen Hawking warned that AI "could spell the end of the human race."[14]

---

[11] Arthur C Clarke, *2001: A Space Odyssey*, New York: Roc, 1968.

[12] Maureen Dowd, "Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse," *Vanity Fair*, March 26, 2017, https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x.

[13] Aatif Sulleyman, "Elon Musk: AI is a 'fundamental existential risk for human civilisation' and creators must slow down," *The Independent*, July 15, 2017, https://www.independent.co.uk/life-style/gadgets-and-tech/news/elon-musk-ai-human-civilisation-existential-risk-artificial-intelligence-creator-slow-down-tesla-a7845491.html.

[14] Ana Santos Rutschman, "Stephen Hawking warned about the perils of artificial intelligence—yet AI gave him a voice," The Conversation, March 15, 2018, http://theconversation.com/stephen-hawking-warned-about-the-perils-of-artificial-intelligence-yet-ai-gave-him-a-voice-93416.

These are dire concerns from people who have major technical achievements. How should the world treat these warnings? To interpret these concerns, one must first know what they mean when they say "AI." There are two types of AI: narrow and general. In narrow AI, machines perform specific tasks for a specific purpose; in general AI, machines solve problems like a human brain. Current AI applications are all examples of a different type of AI, narrow AI. For example, Google's development of the computer programs AlphaGo and AlphaGoZero are remarkable achievements, beating human champions at the game Go, which is so complex there are more possible moves than atoms in the universe. At the same time, these AI applications are brittle: if you change the rules or the dimensions of the board, the algorithm must start from scratch.

In general AI, a system can perform a range of cognitive functions and respond to a wide variety of input data. This has also been described as exercising different kinds of reasoning: "While computers surpass humans in their ability to reason deductively, they are currently far behind in their ability to perform either inductive or abductive reasoning."[15] As a result, although general AI has been a goal since the 1960s, it is still a distant goal, if it is achievable at all. This was also the conclusion of Stanford's One Hundred Year Study on Artificial Intelligence. In their 2016 report, they conclude: "Contrary to the more fantastic predictions for AI in the popular press, the Study Panel found no cause for concern that AI is an imminent threat to humankind. No machines with self-sustaining long-term goals and intent have been developed, nor are they likely to be developed in the near future."[16]

Apocalyptic concerns about AI center on general AI, which carries the possibility of super-intelligent machines that can act counter to the interests of the human race. That is seen, for example, in Elon Musk's comments on AI: "If one company or small group of people manages to develop god-like superintelligence, they could take over the world." He continued: "At least when there's an evil dictator, that human is going to die. But for an AI, there will be no death—it would live forever. And then you would have an immortal dictator from which we could never escape."[17]

Although this is potentially a true statement, it is predicated on the existence of general AI, which for now and in the near future can only be found in science fiction.

---

[15] Andrew Ilachinski, *AI, Robots, and Swarms: Issues, Questions, and Recommended Studies,* CNA Research Memorandum DRM-2017-U-014796-Final. Jan. 2017.

[16] Stanford University, *100 Year Study on Artificial Intelligence*, 2016 Report Executive Summary, 2016, https://ai100.stanford.edu/2016-report/executive-summary.

[17] Cadie Thompson, "Elon Musk Just Issued a Nightmarish Warning About What Will Really Happen if AI Takes Over," Science Alert, April 6, 2018, https://www.sciencealert.com/elon-musk-warns-that-creation-of-god-like-ai-could-doom-us-all-to-an-eternity-of-robot-dictatorship.

Rather, real-world applications of AI in war will be applications of narrow AI for solving specific problems. These applications of narrow AI do carry concerns and risks, but they are fundamentally different from those associated with general AI. We will discuss some risks of narrow AI in military operations in the next chapter.

Concern: AI and lethal autonomy are unlawful (per the Martens Clause)

Another concern about using AI and autonomy in war is that it is unlawful. One element of this argument often comes from consideration of the Martens Clause in IHL, first voiced in the Hague Convention of 1899 and echoed in the Geneva Conventions and in Additional Protocol I of 1977. The original clause reads:

> Until a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection and empire of the principles of international law, as they result from the usages established between civilized nations, from the laws of humanity and the requirements of the public conscience.[18]

Originally included to solve a particular dispute regarding the status of civilians taking up arms against an occupying military force, it has been repeated in other treaties involving armed conflict. This clause has been interpreted several ways:

- A vestigial element of IHL: a now redundant element of IHL given that the laws of war are more fully developed;
- A tool for interpreting IHL: a way of articulating that means and methods of warfare not expressly prohibited by IHL are not necessarily allowed, providing some means of interpreting existing law; and
- A source of IHL: an independent source of law, based on considerations of customary law, humanity, and public conscience.

Arguments from the Martens Clause that autonomous weapons, and the use of AI in warfare, are unlawful come from interpreting this phrase to be a source of law in itself. Although this interpretation is by no means universally held (the US, the United Kingdom, Russia, and other countries prescribe to either of the first two

[18] Laws and Customs of War on Land, Hague Convention II, signed July 29, 1899, https://www.loc.gov/law/help/us-treaties/bevans/m-ust000001-0247.pdf.

interpretations), it is instructive to consider the third, most expansive interpretation in the context of AI.[19]

The three sources stated in the clause are customary law ("usages established between civilized nations"), laws of humanity, and requirements of public conscience. For autonomy and AI, there are currently no sources of customary law to argue against their ethical use, which is not surprising because these are new technologies. The term "laws of humanity" is often interpreted as "prohibiting means and methods of war which are not necessary for the attainment of a definite military advantage."[20] This means, for example, that it is better to wound than to kill, it is better to create wounds that are less injurious, and that detainees will be treated as well as possible. Therefore, the means and methods of warfare using AI and autonomy should seek to cause only the necessary level of harm.

The phrase "dictates of public conscience" is more difficult to interpret. How can one discern this, and what does it entail? Human Rights Watch offers one view for determining public conscience. For example, they write:

> "States should take evolving public perspectives into account when determining whether fully autonomous weapons meet the dictates of public conscience. For many people the prospect of fully autonomous weapons is disturbing. In discussions with government and military officials, scientists, and the general public, for example, Human Rights Watch has encountered tremendous discomfort with the idea of allowing military robots to determine on their own if and when to use lethal force against a human being. A June 2013 national representative survey of 1,000 Americans found that, of those with a view, two-thirds came out against fully autonomous weapons: 68 percent opposed the move toward these weapons (48 percent strongly), while 32 percent favored their development. Interestingly, active duty military personnel were among the strongest objectors—73 percent expressed opposition to fully autonomous weapons. These kinds of reactions to fully autonomous weapons raise serious concerns under the Martens Clause."[21]

[19] Rupert Ticehurst, "The Martens Clause and the Laws of Armed Conflict," *International Review of the Red Cross*, No. 317, April 30, 1997.

[20] Rupert Ticehurst, "The Martens Clause and the Laws of Armed Conflict," *International Review of the Red Cross*, No. 317, April 30, 1997.

[21] Q&A on Fully Autonomous Weapons, Human Rights Watch. <https://www.hrw.org/news/2013/10/21/qa-fully-autonomous-weapons>.

Is public concern about autonomous weapons cause to consider these weapons unlawful? Robert Sparrow discusses several potential ways to characterize and determine the dictates of public conscience. They include public polls, the opinion of experts, and deliberate debate.[22] ICRC adds a fourth category, authoritative sources.[23]

Public polls

One way to determine the state of public conscience is to use opinion polls. For example, the example from Human Rights Watch uses opinion polls to assess the public's view of autonomous weapons. Many polls have shown very negative views overall, though other studies have shown that the level of negativity can depend strongly on how the questions are asked. Public opinion can also change dramatically over time, as seen with other previously new technologies such as the computer, the VCR, and the telephone.[24] Sparrow points out that this malleability of public opinion (which is commonly seen with emerging technologies) can make it an unreliable source of international law.[25]

Expert opinion

Instead of relying upon a random sample of the public, opinions could also be gathered from experts who are knowledgeable about the ethics of new technology for war. Although this approach would be less susceptible to irrational fears (such as the common fear that telephones were dangerous in the early days of their adoption in homes), it still involves opinions that can change over time.[26] Sparrow also mentions that this approach can prejudice the outcome because experts may have biases and

[22] Rob Sparrow, Ethics as a source of law: the Martens Clause and autonomous weapons, ICRC Blog, November 14 2017.

[23] Rupert Ticehurst, The Martens Clause and the Laws of Armed Conflict, International Review of the Red Cross, No. 317, April 30 1997.

[24] When computers became available in the 1980s, they sparked a new term: computerphobia. And in the 1990s, polls showed that about half of those polled feared technology. Adrienne LaFrance, "When People Feared Computers," *The Atlantic*, March 30, 2015, https://www.theatlantic.com/technology/archive/2015/03/when-people-feared-computers/388919/; "Fear of technology: It may be the phobia of the '90s," *Baltimore Sun*, May 9, 1994, http://articles.baltimoresun.com/1994-05-09/news/1994129089_1_fear-of-technology-alarm-clocks-electronic-services.

[25] Rob Sparrow, "Ethics as a source of law: the Martens Clause and autonomous weapons," ICRC Blog, November 14, 2017.

[26] Adrienne LaFrance, "When the Telephone was Dangerous," *The Atlantic*, September 6, 2015, https://www.theatlantic.com/notes/2015/09/when-the-telephone-was-dangerous/403609/.

loyalties to the most technologically advanced nations, so they may act in their own interests rather than the greater good.

Deliberative debate

Sparrow also discusses a more deliberate approach, in which public debate seeks to educate and to explain perspectives in an open way, seeking to move from narrow interests to a common good. Sparrow stresses that such debates must be explicit about the reasons for various positions so that the outcomes can be the result of reason and evidence. For this approach to be successful, it requires an inclusive effort with multiple perspectives represented.

Authoritative sources

One possibility not mentioned by Sparrow but noted by ICRC is that the dictates of public conscience could be conferred to decisions reached by bodies of special authority. For example, United Nations General Assembly resolutions could be seen as representing the public conscience. That case is strengthened if those resolutions are unanimous. [27]

Summary

As stated earlier, there are several possible interpretations of the Martens Clause. Given these considerations of sources, even if a wide interpretation of the Martens Clause is used, the bar for setting the dictates of public conscience is best seen as being higher than simply an opinion poll, in contrast to the view expressed by Human Rights Watch and others. The limitation of public opinion changing over time, making it an unreliable source of international law, is exacerbated by the fact that opinions about new technology can change dramatically over time. Unfavorable polling data regarding AI and autonomy is not remarkable considering similar trends for the telephone and the computer, and it should not be seen as grounds in itself for invoking the Martens Clause. At the same time, we agree with Sparrow that there is much value in a deliberative debate regarding autonomy in warfare that backs up arguments with reason and evidence, and allows different views to be openly exchanged. This would also be valuable for wider application of AI in war.

[27] Rupert Ticehurst, "The Martens Clause and the Laws of Armed Conflict," *International Review of the Red Cross*, No. 317, April 30, 1997.

Concern: Lack of accountability

Another expressed concern about AI and autonomy in warfare is a potential lack of accountability. If a human soldier violates IHL and commits a war crime, that soldier will be charged with that crime and prosecuted. If a machine decides to do that same action, who can be held accountable? Should the programmer? The civilian authorities who decided to field the system? The commander who elected to rely on the machine in that particular operation? Some, such as Matthias and Sparrow, describe this difficulty as a "gap" in responsibility or accountability, and they assert that this gap makes it unethical to use such a system in warfare.

However, this situation only applies if a general AI is truly making decisions with broad autonomy on the battlefield. Absent this possibility—which, as described earlier, is far in the future if possible at all—there is no responsibility gap with a system exercising narrow autonomy or AI. Servicemembers such as soldiers, sailors, and airmen operate within a larger context in which the use of force is governed by law, policy, doctrine, training, and other institutional processes. It is also governed by operational considerations, such as the commander's guidance, Rules of Engagement (ROE), and theater-specific processes and tactics. These are all elements of command and control—ways to ensure accountability, constrain, and positively influence the conduct and the outcome of operations. Autonomous systems and systems using AI will also fall under this overall framework, providing a means of assigning responsibility and avoiding the accountability gap that could make such weapons unethical.

Unlike a soldier on the battlefield who could make an independent and willful decision to commit a war crime, an autonomous system is unable to make a willful decision to commit a war crime because there is no such thing as general AI in the near future. Instead, if there is a tragedy on the battlefield because of an autonomous system or a narrow AI, the process will be the same as for a human soldier: an investigation will explore culpability, including the responsibility of the operator, all the way up the chain of command. The way an autonomous or AI-empowered system is treated in this process should be no different from what is done today with any other weapon, such as a cruise missile, a guided bomb, a rocket, or another munition or capability. Any malfunction is recognized; if there was a fault of the weapon, the decision-maker is held responsible if the fault could have been anticipated. The operational chain of command is responsible for decisions and for outcomes. Absent a general AI in the process, this process is not changed, and there is no responsibility gap. Though this is standard practice for operational use of technology, military policies for AI and autonomy should still pay attention to this question of responsibility to ensure clarity and transparency.

This discussion does not dismiss the need for efforts to promote safety in any use of AI and autonomy in war. We believe that extra measures can, and should, be taken to promote safety considering unique characteristics of this technology, including attention to risk factors we spell out in the following section. Here we simply state that there is no responsibility gap for a system using narrow AI or autonomy.

Concern: Lack of discrimination

One concern about AI and autonomy in warfare is the inability of these systems to adequately discriminate between valid military targets and entities protected from attack (a fundamental requirement of IHL). This concern includes distinguishing between combatants and noncombatant civilians, but it also includes situations in which a combatant surrenders and is no longer subject to attack. Could a machine respond to this situation appropriately? There are actually two parts to this question: can a machine perform this discrimination function at all, and if so, how good is good enough?

Is discrimination possible?

For the first question, context is very important. A machine may easily discriminate between a valid military target in one setting—for example, a hypersonic missile versus a civilian airliner—but have difficulty discriminating in another—such as an insurgent combatant without a uniform versus a local holding a gun for self-defense. If a system can distinguish air defense threats successfully but has trouble discriminating in the latter case, it does not mean the system is unable to make engagement decisions. Instead, it means the system should be certified and used only in the context for which it is able to exercise discrimination effectively. If a system cannot make this discrimination decision in any context, then it cannot be used in a lawful way—a situation that should surface in any weapon's legal review.

Therefore, autonomous or AI-driven systems can be used legally under IHL if they can be designed to discriminate in their specific operational contexts. The technology for this capability exists today, for example, in air defense against certain types of targets that can be distinguished through kinematic attributes.

How good is good enough: what is an ethical standard for discrimination?

But there is a larger question: can it be used ethically? That gets to the question of how good is good enough for this discrimination function. Sparrow argues that perfection is the appropriate standard, because one can expect that human soldiers

would not target civilians in war.[28] Unfortunately, this expectation of perfection is not met in reality. In operations in Afghanistan, Iraq, Syria, and Yemen, human soldiers made decisions that inadvertently and tragically led to civilian casualties.[29] Because humans are not perfect in practice, how should we think of the ethics of risk? Another possible ethical standard is posited by Simpson and Mueller. This approach is based on the fact that society tolerates necessary risk in a wide variety of settings. For example, prescription medicines can have side effects, but their risks are tolerated at certain levels because of their net positive contributions to society. Likewise, engineering projects such as bridges are designed to specific safety standards, but they can still fail if they encounter situations that exceed those standards. This is a calculated risk: if a bridge fails in situations within anticipated tolerance levels, then the designer or engineering company could be held responsible, but if conditions were beyond the tolerance levels, then no one is considered responsible.[30] That is an accepted end state of managing risk.

So, what is the acceptable tolerance level for an autonomous or AI-enabled system? The benefits of autonomy to militaries include lower operating costs and force protection benefits. However, ethicists assert that the benefits to a military force (and its country overall) should not impose greater risks to the population where military force is being used. Such a redistribution of risk, Simpson and Mueller argue, would be unethical. Thus, an ethical standard for autonomous and AI-driven systems would be as follows: can a machine pose less risk to the civilian population as a human soldier?[31] In some operational contexts, this may be an easy standard to meet, and in others it may be quite difficult. This means the ethical use of autonomous and AI is potentially possible, but it will be dependent on both the technology and the operational context. The technical and operational performance of systems employing autonomy and AI could be captured and compared with existing legacy systems using human decision-making to ensure this ethical criterion is met.[32]

Of course, that does not mean militaries should rest once they can show they meet this discrimination standard. It is both laudable and strategically wise to pursue

[28] Rob Sparrow, "Robots and respect: Assessing the case against Autonomous Weapon Systems." *Ethics and International Affairs* 30(1): 93-116. October 2017.

[29] Larry Lewis, *Redefining Human Control: Lessons from the Battlefield for Autonomous Weapons*, CNA Occasional Paper, DOP-2018-U-017258-Final. Mar. 2018.

[30] Thomas Simpson and Vincent Müeller, "Just War and Robots' Killings," *The Philosophical Quarterly* 66, no. 263, 2016.

[31] Thomas Simpson and Vincent Müeller, "Just War and Robots' Killings," *The Philosophical Quarterly* 66, no. 263, 2016.

[32] Larry Lewis, *Redefining Human Control: Lessons from the Battlefield for Autonomous Weapons*, CNA Occasional Paper, DOP-2018-U-017258-Final. Mar. 2018.

higher standards for the use of force, including setting policy constraints that can be more demanding than the requirements of international law. This was seen in US counterterrorism policy, the 2013 Presidential Policy Guidance, and is also called for in the UN Secretary General's 2018 Annual Report on the Protection of Civilians.

What about slaughterbots?

The film *Slaughterbots*, showing quadricopters coupled with sensors and weapons, demonstrates some of the potential hazards of this technology. But some aspects of the film need to be taken into account with respect to IHL. First, the context of the film was not armed conflict, but rather a repressive regime targeting its own civilians. This situation is governed by a different set of laws, so pushing for new IHL on lethal autonomous weapon systems will not address this concern. But what if those autonomous weapons were applied in an armed conflict setting? In that case, their use would be governed by IHL, and the slaughterbots' attack on civilians would fail a fundamental requirement of IHL—the distinction between valid military targets and noncombatants. In either case, the *Slaughterbots* film does not make a case for a new IHL protocol banning autonomous weapons. Either its use of the weapons is outside of an IHL context, making it irrelevant, or it duplicates one of the basic tenants of IHL, making new protocol unnecessary.

However, the type of targeting seen in Slaughterbots also raises a potential hazard in targeting that has been seen previously: using unclear standards for the discrimination process in targeting decisions. For example, US drone strikes in Pakistan, known as signature strikes, were seen to lead to the inadvertent targeting of noncombatants, including several hostages. In signature strikes, the decision was made to use lethal force based on available information that were indicators of likely militant groups, but the specific individuals being targeted were not known.[33] This opacity of targeting could potentially increase through the use of AI, such as machine learning processing through larger data sets. Since these algorithms are typically "black box" processes, their use for targeting without independent verification seems ethically problematic and may violate IHL requirements for discrimination. While this should be a general principle of existing international law, it would also be valuable for policies for the use of AI in war to specifically exclude the use of AI in "signature strike" scenarios.

[33] Scott Shane, Drone Strikes Reveal Uncomfortable Truth: U.S. Is Often Unsure About Who Will Die, New York Times, April 24 2015. https://www.nytimes.com/2015/04/24/world/asia/drone-strikes-reveal-uncomfortable-truth-us-is-often-unsure-about-who-will-die.html.

Final thoughts on general AI

Concerns about AI and autonomy are often predicated on a belief that general AI or general autonomy will be used. That is not where we are today, and given historical difficulties in this field, it is not certain we will reach this point. For example, one AI researcher estimates general AI could be possible somewhere between 50 and 250 years.[34] However, Isaac Asimov, the creator of the Three Laws of Robotics, spoke of general AI with skepticism: "It's quite possible that we will never figure out how to make computers as good as the human brain."[35] This is an artifact of a general observation of AI development over time: incremental steps such as machine learning improvements happen faster than expected, but revolutionary developments such as general AI have not been developed despite predictions to the contrary.

This is not to say that the development of general AI should not be discussed. If general AI is created, then IHL and other legal frameworks will need to consider its implications. This would be an important development in the course of human history, and anticipating its implications early is worthwhile. At the same time, the technological development of general AI is fundamentally different from the emergence of narrow AI and narrow autonomy. Because of their inherent limitations and lack of a "will" and "intent," the latter are tools rather than independent actors on the battlefield. As such, they will still be governed by the different means of military control, which we discuss in the next section. The debate on AI and autonomy should not conflate general and narrow AI, but rather clearly differentiate between the two types of AI because of their different risks.

[34] Arend Hintze, "What an artificial intelligence researcher fears about AI," The Conversation, July 13, 2017, http://theconversation.com/what-an-artificial-intelligence-researcher-fears-about-ai-78655.

[35] James Burke, Jules Bergman, and Isaac Asimov, *The Impact of Science on Society*, NASA SP-482, 1985.

Examining the Risks of AI and Autonomy in War

The last chapter discussed various concerns about the militarization of AI and autonomy, and put them in context. For example, concerns about general AI are seen to be fundamentally different from those to be expected from the near term applications of narrow AI. But there are still serious and immediate risks associated with the military use of this technology including inadvertent engagements (engaging civilian targets, fratricide) and loss of military predictability and effectiveness.

Regarding inadvertent engagements, there has been much discussion about reducing the risk of autonomy in weapons through meaningful human control over the final decision to use force. But the capacity of a military to exercise control over operations and the use of force goes far beyond a soldier's decision to pull the trigger. For example, Ekelhof describes the importance of the wider targeting cycle as a means for providing meaningful human control over decisions to use lethal force.

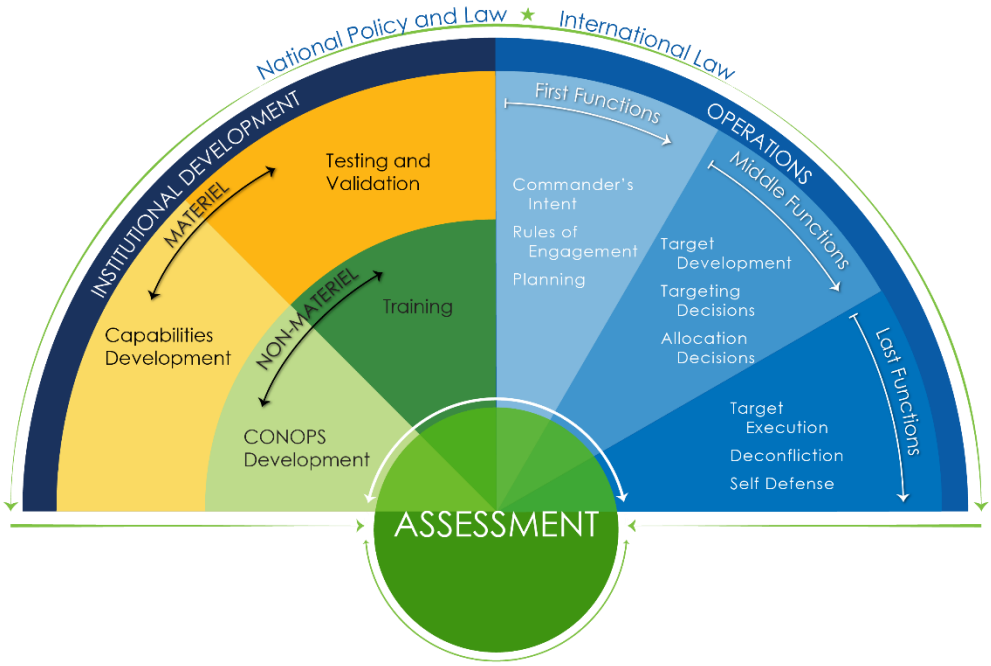
Overall, military operational outcomes are determined by three overarching components:

- operational processes, ranging from commander's intent, planning, the Rules of Engagement, and allocation decisions to the final decision to engage a target,
- institutional development, such as developing military capabilities, establishing doctrine, and training personnel, and
- law and policy that undergird both the conduct of operations and the institutional force.

These areas can be seen as methods of human control, including ways to control risk. These three areas, including major components, are shown below in Figure 2.

We present this as a framework for comprehensive human control over the use of force in operations. To manage the overall risks associated with the militarization of AI and autonomy, measures need to be taken in each of the components of the framework for human control shown in the figure. We discuss specific measures that can be taken for each of these components in the sections below. Attention to risk factors in military operations, across the military institution, and in the foundation of policy and law can strengthen human control over the use of force and help reduce risks associated with these new technologies.

Figure 2. Framework for comprehensive human control over the use of force



Military operations

In military operations involving the use of force, individuals make specific decisions to use force. In some cases, an individual may simply see an object or individual, recognize them as a valid military target, recognize there are no noncombatants in the area, and fire upon that target on their own. This can happen especially in self-defense and in dynamic targeting. However, there are other processes and policies in place that govern the use of force, even in those seemingly independent cases. These processes are represented in the right hand side of Figure 2.

For example, the Rules of Engagement control and limit the use of lethal force, in compliance with IHL and in pursuit of larger operational and policy goals. Consequently, ROE may be more restrictive than required by law. The use of force is also shaped by commander's intent, which expresses the larger operational goals and the means by which that intent will be sought. Commander's intent informs deliberate planning efforts, which shapes the means of the use of force, the desired

target sets, and the intended operational environment, all of which affect the risk of force and the risk of unintended engagements.

These factors influence the targeting process (which includes target prioritization and development, the approval process, and decisions to allocate resources) regarding intelligence and the delivery of lethal force. In this way, processes, policies, and command guidance all influence the decisions to use force. Consequently, there is no such thing as an autonomous soldier or weapon; all decisions to use force are made in this larger context. This is consistent with military doctrine regarding *control*, which includes functions such as "planning, direction, prioritization, synchronization, integration, and deconfliction."[36]

What implications are there for using AI as part of this operational process? How can risks introduced by AI or autonomous functions be addressed operationally? Insights can be gained from considering recent international operations in Afghanistan and adaptations that were made to improve civilian protection and reduce civilian casualties. As the International Security Assistance Force (ISAF) Commander, General Stanley McChrystal recognized that the continuing civilian casualties were undermining the overall mission. He emphasized the protection of the Afghan population. Under his leadership, ISAF modified its policies and procedures to help reduce the risk to civilians from international forces. This approach involved a series of adjustments across the spectrum of operational elements. For example, as ISAF Commander, he issued commander's intent that prioritized civilian protection and limited certain kinds of military actions. The planning of operations began to better consider risk factors for civilian casualties and to develop tactical alternatives, including using alternative placements of forces and placing snipers in key positions beforehand. This planning and operational shaping reduced the need for heavy weapons, such as artillery or air-delivered bombs. Pattern-of-life determination also became more of a priority. Intelligence and reconnaissance assets were increasingly allocated to establish a baseline of what was "normal," improving situational awareness and reducing the chances of mistaking regular activity as nefarious.[37]

Collectively, these efforts bore fruit. Because of improved guidance and training, ISAF forces adapted the way they conducted operations in light of civilian casualty concerns, and ISAF-caused civilian casualties decreased over time. A key reason for this progress was that the number of decisions that operators had to make in the heat of the moment to support engagement decisions was reduced. These

[36] *Operations and Organization*, Air Force Doctrinal Document 2, United States Air Force, April 3, 2007.

[37] Larry Lewis, *Redefining Human Control: Lessons from the Battlefield for Autonomous Weapons*, CNA Occasional Paper, DOP-2018-U-017258-Final. Mar. 2018.

operational adjustments moved some critical tasks to earlier in the targeting process. Though mistakes were not eliminated, these adjustments helped reduce the opportunity for mistakes in human judgment that could lead to inadvertent engagements.

This approach can also be used to make operational adjustments in response to risks regarding AI and autonomy. If an autonomous or AI-driven function carries a particular risk, modifications to the rest of the targeting cycle elements, planning, and commander's intent as well as other guidance can help reduce such risk.[38] Therefore, militaries must be aware of the potential need to modify operational processes to this end, being cognizant of potential risks and taking active steps to mitigate them. The above example of operational adjustments in Afghanistan illustrates how this can be done.

At the same time, it is not necessarily true that using AI and autonomy is riskier than using human judgment in all contexts. In an earlier report, we documented that human decision-making in operations is not perfect. The next section examines the possibility that AI and autonomy could reduce overall operational risks, something that militaries should (and per IHL, it could be argued that they must) pursue. Militaries should leverage the strengths of AI-driven and autonomous systems to reduce risk.

Institutional development

In addition to the conduct of operations, the larger military institution also has elements that collectively reduce the risk of negative operational outcomes. The development of the military as an institution includes many elements: developing and testing equipment, developing doctrine and concept of operations (CONOPS), training and education, and infrastructure such as maintaining facilities. These can be divided into two overall categories: materiel (equipment and systems) and non-materiel (training, doctrine, education). These collectively represent the left hand side of Figure 2.

Materiel development

The development of a military force includes developing and fielding advanced military systems. This represents the outer components of the military institutional

[38] Larry Lewis, *Reducing and Mitigating Civilian Casualties: Enduring Lessons, Joint and Coalition Operational Analysis,* April 12, 2013.

development in Figure 2. For militaries like the US, this is a slow and deliberate process. For example, the average length of an acquisition program is 91 months (7.6 years) from the initial analysis of the requirements to the initial availability of a system.[39] This slow pace is often intentional, because militaries are faced with missions involving the delivery of lethal force in the most challenging of environments. In light of this reality, the structure of the acquisition process ensures high quality to meet demanding internal requirements for major equipment often intended to last many decades. The process also promotes fiscal accountability through requirements that support budgetary and oversight functions.[40]

Aside from the deliberate acquisition process, militaries also have quality assurance processes to certify that the developed systems are effective and safe. These include the establishment of military standards, policies, and test & evaluation processes. Systems are only certified and approved for use when they have been validated through testing. Collectively, the acquisition process and the accompanying test & evaluation process reduces operational risk on the battlefield by screening for problems and demanding compliance with established standards.

Developing capabilities

As militaries consider the incorporation of AI, they should ask the question: what are the benefits and risks from incorporating this technology? What are the reasons to pursue AI or autonomy compared to using a human operator or an automated (but not AI-driven) system? Especially in the early stages of this technology, it makes sense to reduce risk and focus scarce computing and programming resources on high-priority applications that benefit the most from the technology.

The effective use of AI relies on sufficient and unbiased training data. Therefore, militaries need to collect the right data in quantities that are sufficient for training algorithms for AI applications. Often, militaries generate but do not record data, or the data is not archived for long-term use. Changing data use practices requires a new policy that prioritizes data collection and storage to support the effective use of AI. The new policy should be accompanied by efforts to make data more exploitable. For example, data can be in many formats, some proprietary and some not easily exploitable (e.g., PowerPoint slides). This will require the storage of a vast number of datasets, so reasonable data storage options will also need to be explored. Depending

[39] Ilachinski, *Robots, AI, and Swarms.*

[40] The standard acquisition process is described in more detail in: Julianne Nelson, Charles Porter, and Kory Fierstine, *RPED: A New Rapid Prototyping Strategy in the Department of the Navy,* CNA Research Memorandum DRM-2017-U-014757-Final. Mar. 2017.

on the approach, data may also need to be tagged to be fully exploitable, requiring additional effort that can include both human input and automated approaches. We also discuss legal and policy issues in the law and policy section below.

Test and evaluation considerations

Although the acquisition and testing processes are designed to mitigate risk, these processes will need to detect several possible risk factors associated with AI-driven and autonomous systems, including:

- **Potentially fewer communication opportunities because of autonomous operation.** In practice, many operational problems are caused by communication breakdowns. These breakdowns often do not have operational effects because they tend to be corrected over time as differences are arbitrated and resolved. But this process of correction requires continuing communication over time. This self-correcting effect may not occur in the limited time window for communication potentially associated with autonomous operation in a communication-denied or covert mode. This situation also increases the chances of breakdowns in command and control functions, increasing risk of inadvertent actions leading to escalation between States.

- **Lack of a human operator to override potential problems.** Autonomous systems will consider a range of information in making decisions. However, sometimes conflicting information will complicate a decision. In some situations, human operators must be alerted to these situations to resolve them. In addition, sometimes the data will not conflict but instead will appear to suggest what is actually not the case. Addressing this situation is clearly problematic for an autonomous system, but it can be addressed through processes and protocols using data available to multiple systems. Interface standards and systems will need to be revised to address this situation.

- **Testing for non-deterministic systems.** Standard test and evaluation processes tend to be designed for deterministic, rule-based systems. Systems that employ AI can be black-box systems, defying complete predictability. Depending on the design, they can also potentially evolve, meaning the system in operation may not be identical to the system that was tested. Different test and evaluation approaches must be developed to address these potential risks.

Non-materiel development

In addition to materiel development, the military institution also promotes operational effectiveness and avoids negative outcomes through developing doctrine and tactics, techniques, and procedures (TTP), as well as accompanying training and education to equip military personnel in their expected operational roles. This can also include the development of concepts of operations (CONOPS), which integrates doctrine, TTP, and command and control considerations to form playbooks for specific contexts. These non-materiel developments collectively form operational procedures that can reduce the risks of mistakes. This represents the inner components of the military institutional development in Figure 2.

One essential element of using AI-driven and autonomous capabilities will be educating operators regarding the proper operation of these capabilities. As systems grow in complexity, it is challenging to make sure that military personnel can operate these systems properly, and these new capabilities will be no exception.

Another important factor is the element of trust. A basic definition of *trust* is "assured reliance on the character, ability, strength, or truth of someone or something."[41] For military use of technology, the key terms in this definition are *reliance*, the willingness to use the technology, and *assured*, a reasonable confidence that the technology can be relied on. The goal is not blind trust, but rather appropriate trust, in which the operator and commander trusts the capability appropriately for its capabilities and limitations in the particular operational context. This goal has three components: building trust avoiding overconfidence, and ensuring appropriate use of systems.

- **Unwilling to employ systems.** Commanders and operators responsible for an operation are unlikely to authorize the use of a system if they do not fully understand its effects. The 2016 Defense Science Board study makes this point: "The individual making the decision to deploy a system on a given mission must trust the system."[42] A lack of consistent trust in capabilities was seen in Iraq and Afghanistan operations. In some cases, systems were fielded urgently—such as counter-IED systems or surveillance systems providing critical intelligence—and they were eagerly received and used extensively. In other cases, tactical forces generally chose less-capable weapon systems and ISR platforms that were familiar to them, or they simply

[41] "Trust." Merriam-Webster, n.d. Accessed August 16, 2018, Merriam-Webster.com.

[42] Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy*, Washington, DC: Office of the Secretary of Defense, June 2016, https://www.hsdl.org/?view&did=794641.

chose to do without. Even though systems may be provided to operating forces in the field, they are not guaranteed to be used. In those operations, there were ways to promote trust. For example, the Army began proactive training on available systems and their capabilities (and limitations) prior to deployment. Some of these systems were also made available to formal training events, such as unit training at the National Training Center. Those forces were more aware of different options and could make educated decisions regarding which one they employed. This training helped the users and commanders develop trust that these less familiar but more capable systems would advance their larger mission.

- **Overconfidence in systems.** Another challenge is avoiding overconfidence in military systems. For example, overconfidence was a problem in the PATRIOT shootdown of a Navy F/A-18C in Operation Iraqi freedom in April 2003. In that case, the system misidentified an aircraft as a ballistic missile and recommended that the operator approve it as a target. Despite information available to the operator that the entity was not a missile, the operator approved the engagement, and the aircraft was shot down, killing the pilot. This overconfidence in military systems was also seen elsewhere in that operation. For example, operators at radar surveillance systems sometimes believed that the quality of their information was better than it was. This overconfidence can be remedied through exercises combining hands-on experience with assessments that look at real system performance. Of course, those assessments should also be used to improve system performance. We discuss assessments more below.

- **Inappropriate use of systems.** This kind of training also safeguarded those forces from another problem seen in Iraq and Afghanistan: not understanding a new and unfamiliar system led to using it in a suboptimal way that was different from its intended application. At best, this led to the unproductive use of valuable technology. The lack of understanding could also lead to an increased risk of poor decisions, missed opportunities, and inadvertent engagements.

For autonomous systems, this kind of training and familiarization process at the operator and operational commander level would also promote trust and a more informed use of these kinds of capabilities. At the same time, the training requirements are likely to be more robust, because systems employing autonomy (and AI in particular) can adapt and learn depending on the operational environment, the nature of the threat, and the mission. Thus it likely will be necessary for operators and operational commanders to work with these systems more extensively and over a wider range of scenarios for such systems to become relatively predictable and acquire an appropriate degree of trust. This will also be necessary to

ensure operators use these systems properly, instead of in a suboptimal way that is different from the intended application.

It has also been noted that humans tend to be more forgiving of other humans committing breeches of trust than they are of machines doing the same.[43] Therefore, it may not be sufficient for military systems employing autonomy and AI to reach equivalence with human performance; rather, they must exceed that performance before they are accepted and trusted in practice.

Law and policy

Law and policy undergird and shape the entirety of the military enterprise, from developing and maintaining the institution of the military to shaping the conduct of operations. This situation is illustrated in Figure 2, with law and policy encompassing both institutional development and operations. Legal considerations include domestic and international law. These considerations also promote responsibility; foster transparency; avoid corruption; and protect against possible abuses including human rights violations, war crimes, and violations of privacy. Senior military and government leaders also influence the institution and the nature of military operations, including the use of specific technologies in war, through policy decisions. These policies help ensure that military activities are consistent with national principles, values, and interests. Generally, law says what a military can do, and policy says what a military will do, including values and ethics in addition to law.

The conduct of operations are governed by all applicable laws, with IHL being the primary source of international law for armed conflict. In order to comply with IHL, legal considerations are ideally built into the various components of the military institution. In this approach, weapon legal reviews (sometimes called Article 36 reviews, referring to the applicable section of Additional Protocol I) review weapons and ensure they can comply with the requirements of international and domestic law in the specific contexts they are intended for. IHL requirements are also incorporated into doctrine, training, and education to prepare forces to operate in full compliance with the law. In operations, IHL is incorporated into the Rules of Engagement and planning efforts, which reinforces the institutionalization of IHL within the military force. Investigations and assessments can provide accountability and prompt learning in cases in which IHL is not followed, improving compliance and operational outcomes.

[43] Ilachinski, *AI, Robots, and Swarms.*

Policies are also important for both the institutional force and the conduct of operations. These policies can influence levels of oversight and responsibility (e.g., DODD 3000.09 designating responsible offices and requiring a senior-level review of some types of autonomous systems), specify kinds of allowable technology (e.g., cluster munitions directive setting requirements on characteristics needed for such weapons to be used), and spell out processes and risk tolerances in certain types of operations (for example, the US counterterrorism policy outlining an approval and oversight process for specified operations).

For autonomous systems, the US military operates under DOD Directive 3000.09, which governs the approval process for the development and fielding of systems using lethal autonomy.[44] This Directive is not a ban on such systems; rather, it sets conditions for their development and approval. At the time of this report, no system has met the conditions to require such a review. As such systems are developed and fielded, additional policy could be developed to govern their operational use. For example, policy could address questions such as whether unmanned autonomous systems would be permitted to use lethal force against personnel or manned platforms in self-defense. These policy-level determinations tend to reflect the level of trust held for the reliability of these systems. Ideally this trust will be based on testing and performance.

Although there is attention to lethal autonomy internationally, there is less attention to the use of AI in the use of force. Ironically, the latter is more mature than the former, with AI being used operationally through the US Project Maven initiative. More policy attention is needed on this issue, which is both more immediate and, because of the general nature of AI functions, more all-encompassing across the institutional and operational force.

A significant legal and policy issue regarding AI is training data. Training policies and practices will need to consider how to provide an appropriate training data set for the effective employment of AI and autonomy, and account for both potential biases and larger ethical questions of using AI. Considering and addressing potential sources of bias is critical because training data can introduce biases if they are not carefully controlled for. The need for comprehensive, unbiased datasets for training purposes may also introduce new intelligence collection requirements in order to develop them, and the task of screening datasets to ensure unbiased sampling can be a major challenge. Ethical issues include privacy concerns and other legal or normative conventions.

[44] DOD Directive 3000.09, Autonomy in Weapon Systems, November 21, 2012, Incorporating Change 1, May 8, 2017.

Assessments

At the core of Figure 2 is assessments. Militaries rely on assessments at all stages of military institution building and operations. Specifically on the topic of operational risk, assessments can be a powerful way to reduce negative operational outcomes. For example, several assessments for ISAF and U.S. forces in Afghanistan aimed to help those forces reduce civilian casualties. These studies ascertained which kinds of operations were contributing the most to civilian casualties, and what practical measures could be taken to reduce them. After analyzing several hundred separate incidents, the study identified primary causal factors for different types of operations—including airstrikes, check point operations, artillery fire, and vehicle movements—and made tailored recommendations for changes in guidance and tactics to address those causal factors. ISAF made a number of changes to the conduct of operations in response, including modifying tactics, procedures, and the theater's Tactical Directive, correcting a shortcoming that persisted through four previous versions of the guidance.[45]

In addition to operational adjustments, the recommendations were shared with training centers back in the United States to be included in pre-deployment training and the fielding of certain weapons was accelerated to enhance the protection of civilians. Insights from these studies were then compiled as the main body of a handbook for soldiers addressing how to reduce civilian casualties during operations. This handbook was shared with forces in theater as well as those preparing for future deployments, and contained tailored guidance and tactics based on specific lessons from actual civilian casualty incidents. The overarching principles for these assessments were also used as the foundation for US national policy regarding civilian casualties.[46]

These assessment efforts were done periodically and sequentially, re-examining Afghanistan operations over time to observe how existing measures were doing in reducing civilian casualties, and whether there were new issues that also needed to be addressed. At the same time as enemy tactics and the environment changed, it was also necessary to conduct follow-on studies to revisit guidance and tactics in light of subsequent incidents of interest and provide fine-tuning to address new factors as they emerged.

---

[45] Larry Lewis, Reducing and Mitigating Civilian Casualties: Enduring Lessons, Joint and Coalition
Operational Analysis, April 12, 2013.

[46] Executive Order 13732, United States Policy on Pre- and Post-Strike Measures to Address Civilian Casualties in U.S. Operations Involving the Use of Force, July 1 2016.

This assessment process can also be used to identify and mitigate operational risks introduced by the use of AI and autonomy, including factors that may not have been previously anticipated.[47] For example, the process described above for Iraq and Afghanistan was found to surface insights about risk factors that had not been recognized previously. This included a new understanding of how often misidentifications led to civilian casualties and the realization that drones were more prone to cause civilian casualties than manned aircraft in the operational context of Afghanistan. Similarly, it is likely that challenges will be encountered with the use of AI that have not been anticipated, so this assessment process can aid in identifying and prompting remedies for those challenges.

Summary of AI and autonomy risks

In the previous section, we saw that some commonly perceived AI and autonomy risks are not as urgent as they may seem. In this section, we discussed other risks regarding AI and autonomy in military operations organized around our framework for human control in Figure 2 shown earlier: risks in operations, in the larger military institution, and in the foundation of law and policy. We discussed how these risks can tend to be neglected both in public discourse and in military planning. Table 2 below summarizes risks that need to be addressed. They are organized into components of our framework for human control. We believe this table is useful both for militaries looking to implement AI and autonomy as well as for national and international venues looking to understand and mitigate risk of these technologies.

[47] Larry Lewis and Diane Vavrichek, Rethinking the Drone War, Marine Corps University Press, 2017.

Table 2. Mitigating the risks of AI and autonomy in components of human control

Military element	Ways to mitigate the risks of AI and autonomy
Operations	• Military force should make operational adjustments to mitigate risks and leverage specific strengths of AI-driven and autonomous systems to improve operational outcomes
Institutional development: Capability development	• Build in protections to mitigate potentially fewer communication opportunities for autonomous systems operating in communications-denied and covert modes • Develop processes and protocols using data available to multiple systems to override and preempt potential problems associated with the lack of a human operator • Address and mitigate potential biases in training data for AI • Update intelligence and intelligence requirements to support the development of training data for planned AI applications
Institutional development: Test and evaluation	• Develop test and evaluation processes appropriate for non-deterministic and adaptive systems
Institutional development: CONOPS development	• Ensure that planned use of AI-driven and autonomous systems are consistent with their capabilities and limitations
Institutional development: training	• Train operators regarding the correct and appropriate operation of systems employing AI and autonomy • Cultivate appropriate trust, based on knowledge of system capabilities and limitations, specific to the operating environment and intended purpose
Law and policy	• Conduct legal weapon reviews (e.g., Article 36 reviews) to help ensure developed systems comply with IHL in their intended applications • Review CONOPS, doctrine, and training for use with AI and autonomy with respect to IHL • Develop and maintain policy for autonomy in weapon systems, including safeguards and limits • Develop policy for AI and its potential role in operations, including safety measures and ways to leverage its strengths • Ensure ethical and legal issues regarding the collection of training data are sufficiently addressed

AI for Good in War

Science and technology can have dramatic and positive effects on the way we live. Several examples include:

- The development of medicines, increasing the ability of mankind to resist and recover from diseases and maladies;
- The development of electricity, a diverse enabler of technology including refrigeration, electric light, and many other advancements;
- Advancements in communication methods, moving from the telegraph to the telephone to optical fiber and satellite communications;
- Advancements in ground transportation, from horseback to steam engines to gasoline-driven and then electric vehicles; and
- The development of computers, which (like electricity) have enabled changes in virtually all facets of modern life.

AI technology has similar potential to change the way we live. For example, the United Nations campaign #AI4good highlights positive ways AI can be used for the good of humanity. This campaign has emphasized areas where AI has positive applications, including medicine, education, economics, and law enforcement. How could AI relate to these areas? Here are some examples:

- Medicine: AI can design more effective medicines and obtain more accurate diagnoses of medical scans. For example, machine learning was used to improve the quality and processing speed of MRI scans;[48]
- Education: AI can enable more effective, adaptive curricula and allow broader access to educational resources;

[48] Geri Piazza, "Artificial intelligence enhances MRI scans," *NIH Research Matters*, April 10, 2018. https://www.nih.gov/news-events/nih-research-matters/artificial-intelligence-enhances-mri-scans.

- Economics: machine learning can provide insights into the root causes of complex occupational trends, identify biases (e.g., gender and age), and suggest opportunities for those out of work; and

- Law enforcement: AI can help identify victims of human trafficking and crack cold cases to enable justice.

These examples illustrate that artificial intelligence is a powerful technology that can be used for good. But this general realization has not included an important human endeavor: the waging of war. The UN reports that 2 billion people live in countries affected by conflict, violence, and fragility.[49] In this context, many civilians today face war's humanitarian tolls. But there is no conversation on how to apply artificial intelligence to ease the tragedies of war. The Geneva AI4Good conference, addressing so many areas of life, was silent on this topic.

Although AI technology can carry risks, it also offers opportunities. As seen in Table 1 earlier in this report, AI can offer a number of valuable capabilities, including processing complex and large datasets to find optimal solutions, predict behavior, flag anomalies and events of interest, and correct errors. These are all functions that could reduce the humanitarian tolls of warfare. For example, AI could be used in the following ways in war to reduce civilian casualties:

- AI technologies could reduce the number of civilians mistakenly misidentified as combatants, which is a significant cause of civilian casualties;

- AI systems could monitor targeted areas and detect when collateral damage estimates may be too low or have changed, avoiding civilian casualties;

- AI could reduce the risk to civilian infrastructure in conflict areas. This would avoid longer-term negative effects—such as the loss of power, water, and food supplies—impacting local populations; and

- AI could improve military training for civilian harm mitigation measures as a learning objective.

These are just a few ways AI could be used for good in the waging of war. Overall, AI holds promise for saving lives in war, just as in medicine. This promise could be realized if states choose to have their militaries pursue humanitarian gains from

[49] World Bank, "Pathways for Peace: Inclusive Approaches to Preventing Violent Conflict." World Bank, 2018, http://www.worldbank.org/en/topic/fragilityconflictviolence/publication/pathways-for-peace-inclusive-approaches-to-preventing-violent-conflict.

prudent use of AI, if international forums make such positive outcomes a collective goal, or if open society advocates for such goals. As states concerns themselves with the risks associated with AI, they should also look for opportunities to reduce risk and improve operational outcomes with that technology.

There is also a legal dimension to the use of AI. If states can show that AI-driven and autonomous systems do indeed reduce the risk to civilians in certain contexts, it could be argued from IHL that states have an obligation to use those technologies versus human combatants, per API Article 57 discussing feasible precautions. Such a determination is not necessarily straightforward; it requires understanding of the relative risk to civilians from autonomous systems and humans in armed conflict. But this understanding is possible—the risk of civilian casualties from specific weapon platforms can be (and in Afghanistan sometimes was) determined through analysis of operational data. This type of analysis could be made a standard practice in the future to monitor and mitigate risk from AI and autonomy in operations.[50]

[50] Larry Lewis, *Redefining Human Control: Lessons from the Battlefield for Autonomous Weapons*, CNA Occasional Paper, DOP-2018-U-017258-Final. Mar. 2018. https://www.cna.org/CNA_files/PDF/DOP-2018-U-017258-Final.pdf.

Summary

Given the rapid and significant advances in AI, the strong interest in leveraging this technology from advanced militaries, and the urgent concerns voiced in the media, we examined commonly held concerns of AI and autonomy in war. We found that these concerns, on further examination, were not quite what they seemed on first blush. Some concerns were inconsistent with the current state of the technology, such as assuming that general AI is feasible when most estimates place this development many decades away (if ever). Others do not adequately consider the way military systems are actually structured and conducted, in which AI-enabled and autonomous systems would operate as part of a larger process for delivering the use of force. This larger context helps address concerns about accountability and discrimination.

Note that we did not argue that these concerns are spurious—they have value because they can lead to much needed debates and discussions regarding ethical issues of this emerging technology. However, we emphasized that the real risk in a military context (expressed in operational outcomes such as civilian casualties and fratricide) is low from these commonly held concerns. This is important from a risk management perspective because a mismatch between efforts to mitigate risk and the actual sources of risk could lead to the pursuit of ineffective solutions.

We then examined other factors related to the operational use of AI and autonomy, based on a framework for the preparation and conduct of military operations. We identified factors associated with the current and near-future state of the technology that could introduce operational risk if not mitigated. We then identified ways to mitigate each of these factors. These risk factors and mitigation approaches could be helpful for militaries to promote safety and effectiveness when using AI and autonomy in military operations. This framework and the specific risk factors could also serve to help frame international and domestic discussions concerned with addressing the primary applicable risks of AI and autonomy in war. Finally, we showed that the use of AI and autonomy can be employed by militaries for positive outcomes such as humanitarian purposes, and we provided specific examples showing how AI can reduce civilian casualties.

Recommendations

We offer a number of recommendations for mitigating risks from the technologies of AI and autonomy being used in war. The first set is for nations considering use of the technology, to enable them to better address clear and present risks, e.g., avoiding a focus on coffee cup lids when the actual risk is quite different. Recognizing the need for additional and productive dialogue regarding AI and autonomy in war, the second set addresses needed dialogues to discuss the risks of that technology and how to mitigate them.

Recommendations for countries considering the use of AI and autonomy in war:

- Militaries interested in leveraging AI and autonomy should address risk factors impacting operational safety, including operational considerations, institutional development, and law and policy. These risk factors should be addressed to both improve effectiveness and promote safety.
- National policies for AI and autonomy should consider and address the risk of AI increasing the opacity of targeting decisions, akin to the practice of signature strikes.
- In addition to mitigating risk factors, states should also be looking for opportunities for using AI and autonomy to improve the conduct of war

Recommendations for needed dialogues to discuss the risks of the use of AI and autonomy in war:

- Separate out the two cases of general and narrow AI, since the two are distinct, carrying very different sets of risks and having different timelines for development.
- Hold deliberate, inclusive debates concerning AI and autonomy in war, requiring arguments to be supported with reason and evidence, and allowing different views to be fairly exchanged.
- Discuss the risk of AI increasing the opacity of targeting decisions and steps that can be taken to avoid this.
- International venues should consider risk factors identified in this report as a way to frame discussions on how to pursue safety of AI and autonomy in war. Those discussions should include operational considerations, institutional development, and law and policy.
- Consider potential opportunities for using AI and autonomy to improve the conduct of war.

This page intentionally left blank.

References

Barno, David, and Nora Bensahel, "War in the Fourth Industrial Revolution," War on the Rocks, July 3, 2018.

Burke, James, Jules Bergman, and Isaac Asimov, *The Impact of Science on Society*, NASA SP-482, 1985.

Clarke, Arthur C., *2001: A Space Odyssey*, New York: Roc, 1968.

Defense Science Board, *Report of the Defense Science Board Summer Study on Autonomy*, Washington, DC: Office of the Secretary of Defense, June 2016, https://www.hsdl.org/?view&did=794641.

DOD Directive 3000.09, "Autonomy in Weapon Systems," November 21, 2012, Incorporating Change 1, May 8, 2017.

Dowd, Maureen, "Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse," *Vanity Fair*, March 26, 2017. https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x.

"Fear of technology: It may be the phobia of the '90s," Baltimore Sun, May 9, 1994. http://articles.baltimoresun.com/1994-05-09/news/1994129089_1_fear-of-technology-alarm-clocks-electronic-services.

Hauert, Sabine, "Eight ways intelligent machines are already in your life," *BBC News*, April 25, 2017, https://www.bbc.com/news/uk-39657382.

Hintze, Arend, "What an artificial intelligence researcher fears about AI," The Conversation, July 13, 2017. http://theconversation.com/what-an-artificial-intelligence-researcher-fears-about-ai-78655.

ICRC, "International humanitarian law and the challenges of contemporary armed conflicts," *International Review of the Red Cross* 89, no. 867, September 2007.

Ilachinski, Andrew, *AI, Robots, and Swarms: Issues, Questions, and Recommended Studies,* CNA Research Memorandum DRM-2017-U-014796-Final. Jan. 2017.

Jenkins, Cameron, "AI Innovators Take Pledge Against Autonomous Killer Weapons," NPR, July 18, 2018, https://www.npr.org/2018/07/18/630146884/ai-innovators-take-pledge-against-autonomous-killer-weapons.

Kaplan, Robert S., and Anette Mikes, "Managing Risks: A New Framework," *Harvard Business Review*, June 2012, https://hbr.org/2012/06/managing-risks-a-new-framework.

LaFrance, Adrienne, "When People Feared Computers," *The Atlantic*, March 30 2015, https://www.theatlantic.com/technology/archive/2015/03/when-people-feared-computers/388919/.

LaFrance, Adrienne, "When the Telephone was Dangerous," The Atlantic, September 6, 2015, https://www.theatlantic.com/notes/2015/09/when-the-telephone-was-dangerous/403609/.

"Laws and Customs of War on Land," Hague Convention II, signed July 29, 1899 https://www.loc.gov/law/help/us-treaties/bevans/m-ust000001-0247.pdf.

Lewis, Larry, "Reducing and Mitigating Civilian Casualties: Enduring Lessons," Joint and Coalition Operational Analysis, April 12, 2013.

Lewis, Larry, *Insights for the Third Offset: Addressing Challenges of Autonomy and Artificial Intelligence in Military Operations,* CNA Research Memorandum DRM-2017-U-016281-Final. Sept. 2017.

Lewis, Larry, *Redefining Human Control: Lessons from the Battlefield for Autonomous Weapons*, CNA Occasional Paper, DOP-2018-U-017258-Final. Mar. 2018, https://www.cna.org/CNA_files/PDF/DOP-2018-U-017258-Final.pdf.

Lewis, Larry, Diane Vavrichek, Rethinking the Drone War, Marine Corps University Press, 2017.

Marr, Bernard, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," *Forbes*, May 21, 2018, https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#45fa594b60ba.

Narula, Gautam, "Everyday examples of artificial intelligence and machine learning," Techemergence, June 28, 2018, https://www.techemergence.com/everyday-examples-of-ai/.

Nelson, Julianne, Charles Porter, and Kory Fierstine, *RPED: A New Rapid Prototyping Strategy in the Department of the Navy*, CNA Research Memorandum DRM-2017-U-014757-Final. Mar. 2017.

Nilsson, Nils J., *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (New York: Cambridge University Press, 2009).

"Operations and Organization," Air Force Doctrinal Document 2, United States Air Force, April 3, 2007.

Piazza, Geri, "Artificial intelligence enhances MRI scans," *NIH Research Matters*, April 10, 2018, https://www.nih.gov/news-events/nih-research-matters/artificial-intelligence-enhances-mri-scans.

Rutschman, Ana Santos, "Stephen Hawking warned about the perils of artificial intelligence—yet AI gave him a voice," The Conversation, March 15, 2018. http://theconversation.com/stephen-hawking-warned-about-the-perils-of-artificial-intelligence-yet-ai-gave-him-a-voice-93416.

Simpson, Thomas, and Vincent Müeller, "Just War and Robots' Killings," *The Philosophical Quarterly* 66, no. 263, 2016.

Sparrow, Rob, "Ethics as a source of law: the Martens Clause and autonomous weapons," ICRC Blog, November 14, 2017.

Sparrow, Rob, "Robots and respect: Assessing the case against Autonomous Weapon Systems," *Ethics and International Affairs* 30(1): 93–116. October 2017.

Stanford University, *100 Year Study on Artificial Intelligence*, 2016 Report Executive Summary, Stanford University, 2016, https://ai100.stanford.edu/2016-report/executive-summary.

Sulleyman, Aatif, "Elon Musk: AI is a 'fundamental existential risk for human civilisation and creators must slow down," *The Independent*, July 15, 2017, https://www.independent.co.uk/life-style/gadgets-and-tech/news/elon-musk-ai-human-civilisation-existential-risk-artificial-intelligence-creator-slow-down-tesla-a7845491.html.

Thompson, Cadie, "Elon Musk Just Issued a Nightmarish Warning About What Will Really Happen if AI Takes Over," Science Alert, April 6, 2018, https://www.sciencealert.com/elon-musk-warns-that-creation-of-god-like-ai-could-doom-us-all-to-an-eternity-of-robot-dictatorship.

Ticehurst, Rupert, "The Martens Clause and the Laws of Armed Conflict," *International Review of the Red Cross*, no. 317, April 30, 1997.

"Trust." Merriam-Webster, n.d. Accessed August 16, 2018. Merriam-Webster.com.

White House, Executive Order 13732, United States Policy on Pre- and Post-Strike Measures to Address Civilian Casualties in U.S. Operations Involving the Use of Force, July 1 2016.

World Bank, "Pathways for Peace: Inclusive Approaches to Preventing Violent Conflict," World Bank, 2018, http://www.worldbank.org/en/topic/fragilityconflictviolence/publication/pathways-for-peace-inclusive-approaches-to-preventing-violent-conflict.

CNA
ANALYSIS & SOLUTIONS

CNA is a not-for-profit research organization
That serves the public interest by providing
in-depth analysis and result-oriented solutions
to help government leaders choose
the best course of action
in setting policy and managing operations.

Nobody gets closer—
to the people, to the data, to the problem.

CNA
ANALYSIS & SOLUTIONS

www.cna.org • 703-824-2000

3003 Washington Boulevard, Arlington, VA 22201