Artificial Intelligence, Robotics, Ethics, and the Military: A Canadian Perspective

Sherry Wasilow, Joelle B. Thorpe

■ *Defense and security organizations depend upon science and technology to meet operational needs, predict and counter threats, and meet increasingly complex demands of modern warfare. Artificial intelligence and robotics could provide solutions to a wide range of military gaps and deficiencies. At the same time, the unique and rapidly evolving nature of AI and robotics challenges existing polices, regulations, and values, and introduces complex ethical issues that might impede their development, evaluation, and use by the Canadian Armed Forces (CAF). Early consideration of potential ethical issues raised by military use of emerging AI and robotics technologies in development is critical to their effective implementation. This article presents an ethics assessment framework for emerging AI and robotics technologies. It is designed to help technology developers, policymakers, decision makers, and other stakeholders identify and broadly consider potential ethical issues that might arise with the military use and integration of emerging AI and robotics technologies of interest. We also provide a contextual environment for our framework, as well as an example of how our framework can be applied to a specific technology. Finally, we briefly identify and address several pervasive issues that arose during our research.*

Artificial intelligence is based on the assumption that aspects of human thought can be mechanized. Although AI has existed for decades in philosophical debate, mathematical models, and computer-science labs (IEEE 2017; JASON 2017), it is only in the last five or six years that a massive increase in computer-processing power, access to a profusion of data, and advances in algorithmic techniques have together propelled AI to the forefront of public, media, government, and military attention (DASA R&T 2017). We anticipate that the relationship between AI and robotics[1] will become interdependent with time, as robots become the hardware that use, for example, machine learning algorithms to perform a manual or cognitive task (UK Atomic Energy Authority 2016).[2]

AI can be both transformative and disruptive, due largely to its dual-use properties (Allan and Chan 2017) and capabilities.[3] The benefits can be numerous. Driverless cars are anticipated to save hundreds of thousands of lives (Osoba and Wesler 2017). AI already assists clinicians with medical diagnoses (Amato et al. 2013). Neural networks can scrutinize surveillance video and alert soldiers to specific frames that contain objects of interest such as vehicles, weapons, or persons. Facial-recognition software could alert soldiers when an individual of interest is observed in video surveillance or in real time. AI might also help military personnel amalgamate and fuse large amounts of data from numerous sensors in a battlespace, and find relationships within the data, to help make more informed and more rapid decisions than if the data were processed manually. Furthermore, AI-enhanced robotic systems can be given dull, dirty, and dangerous jobs, reducing physical risk to soldiers and enabling them to concentrate their efforts elsewhere.

Yet AI applications also raise a number of red flags. Facial-recognition capabilities and databases for surveillance and protection purposes can prompt individual privacy concerns. Surveillance tools targeting criminals can also be used to collect personal information on ordinary citizens or even to commit intelligence espionage. For example, Project Arachnid is an automated web crawler used by the Winnipeg-based Canadian Centre for Child Protection that detects online child sexual abuse images and videos (Beeby 2018) and then sends a notice to the host service provider to have it removed — resulting in nearly 700 removal notices every day. Conversely, Edward Snowden used web-crawler software to collect roughly 200,000 top secret documents from the US National Security Agency servers (Sanger and Schmitt 2014). In addition, the use of cyberspace for sharing news and opinions can be manipulated by "social bots" for the purposes of disinformation and political agitation (Lazer et al. 2018). As AI moves along the spectrum of technical sophistication in conjunction with an anticipated increase of autonomy, public concerns can increase. For example, fear that AI technology is rapidly evolving toward autonomy in weapons has led to opposition to the development of so-called lethal autonomous weapons systems (LAWS)[4] and to spirited public debate both within Canada and at meetings of the Convention on Certain Conventional Weapons (CCW) in Geneva.

Although civilian acceptance of AI in daily life has noticeably increased, its adoption in the military realm is much more complicated given the high stakes involved. The difference in pace between the scientific development of these technologies and the creation of policy to regulate their use can lead to gaps when it comes to understanding the legal, ethical, and social implications of adopting these technologies for military purposes. While AI can provide a number of benefits in the areas of military surveillance/intelligence, detection/protection, decision-making, and weapons, it is important to consider the ethical implications of these technologies in advance of their use in order to mitigate potential issues on the battlefield before they occur. In 2016, the Office of the Chief Scientist, Defence Research and Development Canada, initiated work on the ethical implications of AI, which led to the creation of an ethics assessment framework for emerging AI and robotics technologies in future military systems.[5]

Why an Ethics Framework?

Concurrent with rapid developments in AI technologies, academic interest in the ethics of AI has grown exponentially in the last several years — in conferences,[6] initiatives (MIT Media Lab 2017), longitudinal studies (Stone et al. 2016), and principles and policy positions (Future of Life Institute 2017; IEEE 2016, 2017; Montreal Declaration[7]). Government attention has increased in the US (Executive Office of the President 2016a, 2016b), the European Union (European Parliament 2016), and France (Villani 2018). Several private companies have established their own ethics codes on AI (for example, DeepMind[8]), and a number of industry players have created the Partnership on AI to Benefit People and Society to formulate best practices for the use of AI technologies.[9] It is not clear, however, how much cross-fertilization of ideas is taking place across academia, layers of governance, and public and private sectors in Canada and elsewhere (House of Commons 2016). It is not clear if, and how much, consensus exists regarding the ethics of AI.

Frameworks are guidance tools. In this case, a framework on ethics can help invested parties identify ethical issues that might be raised by the use of a technology of interest. While several ethics frameworks for emerging technologies currently exist (for example, Wright [2011]), and some reports provide in-depth examination of the potential ethical impact of AI-enabled robotics use by the military (Lin, Bekey, and Abney 2008), to our knowledge there is no existing framework designed to be used as a tool to guide scientists and policymakers in their ethics assessments for emerging AI and robotics technologies of interest to the military.

The framework we present consists of 12 broad categories with guiding questions to help technology developers, policymakers, decision makers, and other stakeholders identify and broadly consider potential military ethical issues that might arise with the use and integration of specific emerging technologies of interest in the fields of AI and robotics. We believe that when ethics are considered early in the development process, potential ethical issues can be mitigated by changes either to fundamental algorithmic design or in the creation of policies regulating technology use within a military or society. It is important to note that while this framework is designed to help individuals identify potential military ethical issues, it is not designed to provide immediate solutions to these issues, advocate for or against the use of any particular technology, make specific policy recommendations, or rank the importance of ethical issues.

In the remainder of this article, we present our framework (and sample guiding questions), demonstrate the framework's utility in identifying potential ethical issues raised by an example technology area of interest (swarming), and discuss several overarching ethical issues raised by AI and robotics technologies.

Ethics Assessment Framework: Emerging AI and Robotics Technologies

The first three categories of our framework address Canadian and international codes and norms. The Defence Ethics Programme (DEP 2015) — a comprehensive values-based ethics program put in place to meet the needs of Department of National Defence and the Canadian Armed Forces (CAF), also known as the Canadian Forces (CF), at both the individual and the organizational levels — is foundational. A key component of the DEP is the DND and CF Code of Values and Ethics (2014), which defines the values and behaviors to which Canadian military members must adhere. We have also included the international rules that must be followed before and during times of conflict. The remaining nine categories encompass ethical concerns that were identified by research as important to consider, but they do not necessarily fall under national or international laws and norms.[10] The majority of sample questions raised in each of the categories have been derived from the existing literature.

1. Compliance with the DND and CAF Code of Values and Ethics

> *Definition:* Common values and expected behaviors that guide CAF members and DND employees.

The code is made up of three principles and five values. The principles are respect the dignity of all persons; serve Canada before self; and obey and support lawful authority. The values are integrity, loyalty, courage, stewardship, and excellence. A sampling of questions related to this category: Could robotic coworkers undermine group loyalty, cohesion, and group effectiveness? Could the use of AI-enhanced technologies that enable soldiers to remain further removed from danger serve to devalue the military value of courage? Or could such use increase risk taking?

2. Compliance with *Jus Ad Bellum* Principles

> *Definition:* Criteria to be met before entering a conflict so that all conflicts are justified.

Just war theory (Wertheimer 2010) is a philosophy of military ethics that aims to ensure that war is permissible and fair. Generally speaking, *jus ad bellum* is the part of just war theory that includes principles designed to ensure that all conflicts entered into are justified. These are principles such as that the aim of a conflict must be for self-defense, and must not serve the narrow self-interests of the state but serve to reestablish peace; that conflict must be waged only by a legitimate authority; that there must be a reasonable expectation the conflict will achieve its desired outcome; that all nonviolent options must be tried before entering into a conflict; that a state's response must be proportional to the threat received; and that the intent of the conflict must be legitimate. A sampling of questions related to this category: Could the use of AI-enhanced surveillance or weapons technologies that reduce physical risk to soldiers lead to lowered barriers to entering conflict, and could this violate the principle of last resort? Could a vast increase in technological asymmetry against our adversaries gained through use of AI and robotics technologies be considered unethical and violate the principle of proportionality because we could engage in conflict in a much more risk-free way?

3. Compliance with Law of Armed Conflict (LOAC) and International Humanitarian Law

> *Definition:* International laws that must be followed during times of conflict.

LOAC is an international law that exists to protect those affected by conflict and to regulate the means of warfare that are used (Solis 2016). LOAC includes *jus in bello* principles, which ensure that the means of warfare are permissible and just. Several major principles are that a soldier must distinguish between combatants and noncombatants; that damage and loss of life in pursuit of a military objective must not be excessive compared to the direct military advantage gained by the action; that prisoners of war (POWs) must be treated humanely, and adversaries who are injured or who surrender must not be targeted; that no means of war that are evil in themselves — such as ethnic cleansing or rape — nor excessive force, nor weapons banned by international law may be used; and that there must be no discrimination of individuals based on gender, race, religion, or any other aspect of humanity. A sampling of questions related to this category: Could AI-enhanced autonomous systems[11] effectively distinguish between combatants and noncombatants? Could AI-enhanced surveillance and detention capabilities such as robot guards be ethically used with POWs? If AI-enhanced weapons were able to target with far greater accuracy and precision than a human soldier, leading to less collateral damage and fewer casualties, would it be ethical to avoid using these weapons if they were developed?

4. Health and Safety Considerations

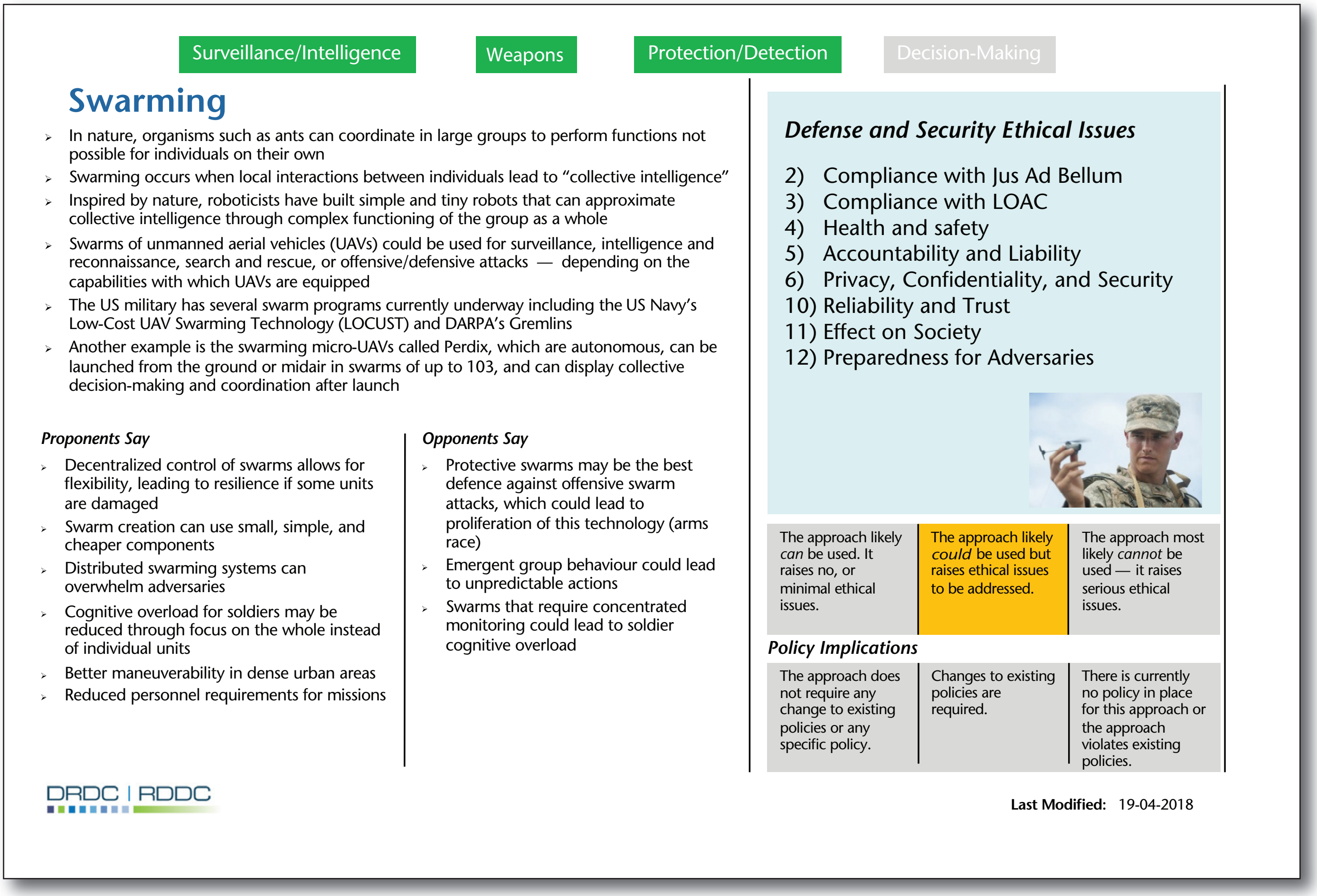
> *Definition:* Questions about the direct and indirect impact of AI or robotic technologies on soldiers' and civilians' physical and psychological well-being.

A sampling of questions related to this category: Could ground robots lessen physical and psychological injury to noncombatants? Further, could they be safer because they will lack an immediate emotional response to the death of a comrade that could lead to acts of revenge? Would unmanned aerial vehicle (UAV) pilots operating thousands of miles away from their targets be classified as combatants? Could the use of UAVs expand the theater of war and put more civilians' safety at risk? Could the use and supervision of, or responsibility for, multiple AI-enhanced systems lead to cognitive overload on soldiers and place their safety and that of others at risk?

5. Accountability and Liability Considerations

> *Definition:* Questions about risk and responsibility for AI- and robotic-technology failures, as well as unanticipated and/or undesired effects.

A sampling of questions related to this category: Who would be accountable for the decisions and actions of semi- or fully autonomous systems as well



Surveillance/Intelligence | Weapons | Protection/Detection | Decision-Making

Swarming

- In nature, organisms such as ants can coordinate in large groups to perform functions not possible for individuals on their own
- Swarming occurs when local interactions between individuals lead to "collective intelligence"
- Inspired by nature, roboticists have built simple and tiny robots that can approximate collective intelligence through complex functioning of the group as a whole
- Swarms of unmanned aerial vehicles (UAVs) could be used for surveillance, intelligence and reconnaissance, search and rescue, or offensive/defensive attacks — depending on the capabilities with which UAVs are equipped
- The US military has several swarm programs currently underway including the US Navy's Low-Cost UAV Swarming Technology (LOCUST) and DARPA's Gremlins
- Another example is the swarming micro-UAVs called Perdix, which are autonomous, can be launched from the ground or midair in swarms of up to 103, and can display collective decision-making and coordination after launch

Proponents Say

- Decentralized control of swarms allows for flexibility, leading to resilience if some units are damaged
- Swarm creation can use small, simple, and cheaper components
- Distributed swarming systems can overwhelm adversaries
- Cognitive overload for soldiers may be reduced through focus on the whole instead of individual units
- Better maneuverability in dense urban areas
- Reduced personnel requirements for missions

Opponents Say

- Protective swarms may be the best defence against offensive swarm attacks, which could lead to proliferation of this technology (arms race)
- Emergent group behaviour could lead to unpredictable actions
- Swarms that require concentrated monitoring could lead to soldier cognitive overload

Defense and Security Ethical Issues

2) Compliance with Jus Ad Bellum
3) Compliance with LOAC
4) Health and safety
5) Accountability and Liability
6) Privacy, Confidentiality, and Security
10) Reliability and Trust
11) Effect on Society
12) Preparedness for Adversaries

The approach likely *can* be used. It raises no, or minimal ethical issues.	The approach likely *could* be used but raises ethical issues to be addressed.	The approach most likely *cannot* be used — it raises serious ethical issues.

Policy Implications

The approach does not require any change to existing policies or any specific policy.	Changes to existing policies are required.	There is currently no policy in place for this approach or the approach violates existing policies.

DRDC | RDDC

Last Modified: 19-04-2018

Figure 1. Sample Chart on Swarming Technology Assessment.

as decision-support systems? The programmer? Manufacturer? Soldier? Commander? Government officials? Given that robots could have better situational awareness by being able to see through walls, see in the dark, or network with other computers — if soldiers chose not to use these systems, leading to a civilian casualty, would the soldiers become liable due to their choice? What would happen if a soldier disagrees with a decision rendered by AI technologies?

6. Privacy, Confidentiality, and Security Considerations

Definition: Questions about sharing, storing, protecting, and using information obtained by AI technologies.

A sampling of questions related to this category: What expectation of privacy would soldiers have in scenarios involving surveillance or data-collection technologies? How would the acquired information be used, stored, and protected? Could robots with biometric capabilities, such as the detection of faces from a distance or weapons under clothing or inside a house, blur the line between surveillance and search (which requires a warrant)? How would we ensure that training data — both private and open sourced — for machine learning systems are safe from exploratory hacking that could discover and exploit program weaknesses for later use?

7. Equality Considerations

Definition: Questions about the influence of AI and robotics on fairness and functionality within the CAF and society.

A sampling of questions related to this category: Could the use and distribution of AI capabilities lead to changes in unit cohesion? (For example, if soldiers were asked to work alongside AI-enhanced or autonomous robots with cameras that recorded soldier actions.) How could the military guard against using AI containing algorithmic bias or stereotyping? Would human and machine interactions be equal? Who or what would be in charge?

8. Consent Considerations

Definition: Questions about consent to, or approval of, AI technologies.

A sampling of questions related to this category: Do soldiers need to provide consent to observation by AI-enabled surveillance technologies or to working with robots? Is it possible to obtain truly informed consent from soldiers (or civilians) if AI can infer private or undisclosed information, such as gender, from publicly available data such as online behavior?

9. Humanity Considerations

Definition: Questions about the impact of AI and robotic technologies on morality, personal responsibility, and human dignity.

A sampling of questions related to this category: Could operating remotely piloted vehicles or supervising remote autonomous systems have the feel of a video game war, removing the emotional link to the consequences of engaging in conflict? Could such emotional distance encourage unethical behavior? Could it be considered inhumane? Alternately, could physical distance from a battlefield give soldiers more time and distance to make more calculated, deliberated, and humane decisions since they are not at risk of injury or death? Could a robot have the right to act in self-defense, for example, to protect classified information? What responses would become permissible for a robot acting in self-defense?

10. Reliability and Trust Considerations

Definition: Questions about trust in AI-enhanced technologies, and human and machine interactions.

A sampling of questions related to this category: Could distrust in AI decision aids lead to battlefield soldiers disregarding recommendations made by these systems? Conversely, if a robot and a soldier were to “disagree” on a course of action, could an overabundance of trust in the system lead the soldier to disregard his/her training and instincts? Could a soldier mistakenly trust an AI-enhanced or robotic technology that has been hacked and is no longer trustworthy?

11. Effect on Society Considerations

Definition: Questions about how the use of AI-enhanced and robotics technologies could affect civilians, and how civilians could respond to these technologies.

A sampling of questions related to this category: Should AI and robotic technologies be regulated to the same degree in the civilian world as in the military world? Could using robotics in conflicts abroad hinder the ability of soldiers to connect with, and win the hearts and minds of, civilians on the ground? If robotic technologies were equipped with self-destruct capabilities that are triggered when captured, and injured civilians, how could this influence civilian attitudes?

12. Considerations Regarding
Preparedness for Adversaries

Definition: Questions about the use of AI-enhanced technologies and robotics by our adversaries, and how our adversaries might view our use of these tools.

A sampling of questions related to this category: Could our AI technologies be hacked or spoofed by adversaries? (For example, could our robotics systems be captured and reprogrammed to act against us, such as hacking an unmanned ground vehicle and driving it into a crowd?) Could new AI technology that enables realistic audio/visual impersonation be used by our adversaries as propaganda or to spread false information about our military actions? Could a nation’s development and use of AI-enhanced and robotics technologies contribute to an international AI arms race?

Application of Ethics Assessment Framework: Swarming Technologies Case Study[12]

Here we provide a case study wherein we use our framework to identify potential ethical issues raised by possible future military use of swarming technologies.[13]

In nature, organisms such as ants can coordinate in groups of large numbers to perform functions not possible for individuals on their own (Mlot, Tovey, and Hu 2011). This swarming behavior results from local interactions between individual entities that lead to collective intelligence and emergent group behavior (Couzin and Krause 2003). Inspired by nature, roboticists have built simple robots that can exhibit this swarming behavior (Rubenstein, Cornejo, and Nagpal 2014). Swarming capabilities could be useful for military purposes if applied to UAVs. For example, swarms of UAVs could be used for intelligence, reconnaissance, as well as — depending on capabilities — defensive and offensive purposes (Hurst 2017; Scharre 2014). Swarming capabilities have been developed and tested by the US military, including Perdix (US Department of Defense 2017) and low-cost UAV swarming technology called LOCUST (Smalley 2015). If adopted, swarming could offer several military advantages, including greater resilience due to decentralized control (Scharre 2014); the ability to overwhelm adversaries because of a swarm’s distributed nature (Scharre 2014); and superior maneuverability in dense urban areas or other locations too dangerous for humans (Higgins, Tomlinson, and Martin 2009). Despite these and other advantages, there are ethical questions raised by swarming technologies, identified with the help of our framework in the following categories (referenced by category number):

Compliance with Jus Ad Bellum *Principles* (category 2). Some questions related to *jus ad bellum* principles: Could UAV swarming technology, which enables soldiers to remain farther from danger, lead to lowered barriers to entering conflict. Could this violate the principle of last resort? Could the use of swarming technology, with its inherent advantages (that is, greatly reduced risk to soldiers), against a less technologically advanced adversary lead to a violation of the principle of proportionality?

Compliance with Law of Armed Conflict and International Humanitarian Law (category 3). Some questions related to this category: Can swarms of UAVs used for persistent surveillance distinguish between combatants and non-combatants? Or recognize an adversary that has surrendered or is injured?

Health and Safety Considerations (category 4). Some questions related to this category: Could swarms of UAVs lead to psychological injury to civilians on the ground who might feel "spied on"? Could soldiers tasked with supervising swarms become overwhelmed and experience cognitive overload? Could this lead to mistakes that place soldiers or civilians on the ground at risk?

Accountability and Liability Considerations (category 5). Some questions related to this category: If swarms display emergent, unanticipated behavior (that is, they "decide" to carry out orders without human input), who would be held accountable for potentially negative consequences? How would the use of swarm technologies be regulated if and when used by one country that is part of an alliance or coalition?

Privacy, Confidentiality, and Security Considerations (category 6). Some questions related to this category: Could a swarm of UAVs be hacked by adversaries to prevent it from acting (for example, jamming communications capabilities), cause it to act against us, or obtain surveillance data? Could the pervasive use of swarms for persistent surveillance negatively impact the privacy of civilians?

Reliability and Trust Considerations (category 10). Some questions related to treliability and trust: Can we trust swarms of UAVs that display emergent behavior that is not programmed? Could emergent behavior lead to unpredictable actions of the swarm or unanticipated by-products of the behavior?

Effect on Society Considerations (category 11). Some questions related to the effect of society: Could our use of UAV swarms be viewed negatively by civilians in an area of conflict, and could this impact our ability to win their hearts and minds? How much input should the public and interest groups have regarding the use of swarming technology, particularly if used for offensive purposes?

Considerations Regarding Preparedness for Adversaries (category 12). A question related to this category: Can we defend against swarms of UAVs as this technology proliferates?

We believe that early attention to ethical considerations related to technologies of interest — with the assistance of a framework such as the one presented — can enable militaries to take advantage of the benefits offered by technologies such as swarming while avoiding potential negative consequences.

Key Ethical Issues and Patterns: Emerging AI and Robotics Technologies in the Military Sphere

While emerging AI and robotics technologies might have associated ethical questions that our framework can help identify, our research also revealed convergence on several issues and themes across different emerging technologies that make use of AI or robotics, discussed below.

Privacy

A number of privacy issues can be raised for both soldiers and civilians through use of surveillance/intelligence and detection/protection technologies. The collection, analysis, use, and sharing of personal data have become increasingly attractive features of AI systems, particularly for marketing and political purposes. Simply hiding or even deleting sensitive variables in the data-collection process often fails to solve the problem, as machine learning methods are capable of probabilistically inferring hidden variables (Campolo et al. 2017). In short, traditional expectations of data privacy and anonymity might no longer be realistic because modern machine learning algorithms are capable of reidentifying data easily and robustly (Osoba and Welser 2017). How might this impact the use of soldiers' personal information, for example, if AI were used to identify patterns and make predictions about their mental-health status, in the course of care during service? What happens to this personal information when a soldier leaves the force? Could the technology be hacked, giving adversaries unauthorized access to sensitive information that could then be manipulated? These issues and others led to development of the European Union's new data privacy regulation, General Data Protection Regulation (GDPR), enacted on May 25, 2018, in order to protect EU citizens from privacy and data breaches. GDPR compliance is becoming the de facto expectation worldwide.

Although news concerning Cambridge Analytica's targeting of 50 million Facebook users for political purposes during the 2016 US election has garnered recent media attention (Rosenberg, Confessore, and Cadwalladr 2018), a more concerning and underlying issue is that "AI challenges current understandings of privacy and strains the laws and regulations we have in place to protect personal information" (Compolo et al. 2017, 28). While members of the military community might rely on Facebook, Twitter, Instagram, and other social networks to stay connected with families, friends, and current events, research shows that extremists, conspiracy theorists, and foreign actors use social media to spread subversive disinformation to influence opinions and discussion within the US military community (Gallacher et al. 2017).

Bias

Data are used to train AI software. Research has shown that the amount of data used to train machine learning algorithms has a greater effect on prediction accuracy than the type of machine learning method used (Banko and Brill 2001). The central role that data plays is one of the reasons that successful companies such as IBM and Google are eager to acquire massive amounts of it. Google's Chief Scientist, Peter Norvig, has been quoted as saying: "We don't have better algorithms than anyone else; we just have more data" (Buchanan and Miller 2017, 13). Military systems likewise have access to massive amounts of data collected over decades.

However, AI software is only as smart as the data used to train it. Human-generated data labeling and algorithms can contain biases — and if the data sample and labeling are biased, then so too will the outputs be tainted. For example, a 2016 ProPublica investigation (Angwin et al.) revealed that the COMPAS program — an algorithm-based risk-assessment tool used to assess criminal risk in the US — was inherently biased against African Americans. Another 2016 study determined that facial-recognition technology used for law-enforcement purposes in the US disproportionately implicated African Americans because they are disproportionately represented in mug-shot databases (Garvie, Bedoya, and Frankle 2016). A more recent analysis of three commercial technologies that identify people of different races and gender — owned by Microsoft, IBM, and Megvii of China — found that when the person in the photo was a white man, the software was correct 99 percent of the time; however, the darker the skin, the more errors arose, especially for darker-skinned women, who were scarcely represented in the system (Buolamwini and Gebru 2018).

Given that algorithmic bias has been found in private industry, it might also exist within military databases. How then can that data confidently be used for AI training purposes? For decision support in foreign and/or unfamiliar regions of the world that in no way are represented by the data being used to generate options? As has been noted by Immigration, Refugees and Citizenship Canada: "Data is not neutral, nor can it be neutralized. Data will always bear the marks of its history. In using data to train a system to make recommendations or decisions, we must be fully aware of the workings of this history" (IRCC 2018, 33). The aura of objectivity and infallibility that our culture ascribes to algorithms (Bogost 2015) is sadly misplaced and, in the case of military use, could have serious and long-term implications.

Safety and Security

The use of AI and robotic technologies raises questions about soldiers' and civilians' physical and psychological well-being, both domestically and internationally. For example, the March 2018 fatality involving a pedestrian and an autonomous Uber car in Tempe, Arizona, has led to intensified scrutiny of autonomous vehicles on public roads (Coppola, Beene, and Hull 2018). What about the safety of soldiers and civilians in battlefield scenarios where remotely piloted air and ground vehicles — and possibly autonomous vehicles — are used?

The Uber accident also raises questions about the statistics and foresight that have propelled autonomous technology forward. For example, experience garnered from commercial aviation developments has shown there is often an increase in the rate of adverse events when new automated systems are introduced (Airbus 2017). This statistic raises questions about soldier safety and testing: How will outcomes for AI and robotics technologies that are generated in a laboratory or safe and controlled sandbox areas be transferred to real-world scenarios where rough terrain, obstacles, combatants, and debris might complicate testing and place soldiers at risk during trials (Anderson and Matsumura 2015)?

As machine learning systems become more powerful and central to society, so too might the potential harm from hacking become greater. If machine learning algorithms are driving cars, guiding robots on patrol, and piloting systems, then not only are the safety and security stakes higher, but the response speed of individual decisions will need to be faster as well. Hackers who compromise systems will have a much greater capacity to do enormous damage more quickly, while defenders might find it harder to identify the threat or intervene in time (Buchanan and Miller 2017). Finally, as military strategy evolves toward greater human/machine teaming, the ramifications of as-yet-unknown incompatibilities, pressures, and rights and responsibilities might arise. For example, placing greater value on the use of AI-powered swarming technologies in field operations could risk the mental health of remote pilots who might become overloaded (Chung 2018), which could then jeopardize the physical health and safety of civilians on the ground.

Accountability and Responsibility

Emerging AI and robotics technologies are complex. When a complex or autonomous system fails or causes unanticipated and/or undesired effects, it can be very difficult to determine the cause or ascribe responsibility for the failure. While AI is a tool that can offload certain tasks from humans, it does not possess the agency to ultimately take responsibility for recommendations, decision-making, or even its impact on decision-making processes.

Much of the current conversation concerning accountability and AI-enabled systems has taken place at the far end of the machine-autonomy spectrum, where the LAWS debate resides, and has revolved around definitions such as "appropriate human involvement" or "meaningful human control."[14] The US Department of Defense already recognizes AI, both commercially derived and military-specific, as a key enabling technology that will become integral to most future systems and platforms as part of a "Third Offset Strategy" that seeks a unique, asymmetric advantage over near-peer adversaries (JASON 2017). Furthermore, the US Center for

Strategic and International Studies recommends that, instead of "LAWS 'never,' our policy should be 'not until they can outperform human/machine intelligence collaboration,' including making ethically acceptable choices about when to 'pull the trigger'" (Carter, Kinnucan, and Elliot 2018, 23).

However, there are day-to-day accountability issues related to AI and robotics that should be addressed long before dystopian scenarios, issues such as malware and the destruction it can cause, technology failure and/or unintended activity, and the use of AI for law-enforcement purposes or social monitoring. For example, robotic police officers debuted in Dubai in 2017 (Cellan-Jones 2017). If these robots were to carry weapons, new questions would arise about how to determine when the use of force is appropriate (Campolo et al. 2017). China is creating a pervasive algorithmic surveillance system designed to produce a "citizen score" (Mitchell and Diamond 2018). How will democratic societies that are building smart cities that incorporate similar surveillance technologies use, analyze, and store collected citizen data? Furthermore, if a soldier has been teamed with an AI-enhanced robot that fails or is hacked, will they be accountable for the actions of the robot? Will they be expected to intervene as best they can and, if they don't, will they be held liable for the consequences? And if so, why would soldiers agree to partnering with these systems if they could be blamed for actions they cannot necessarily predict?

Reliability

It is not clear that reliability — defined as achieving the same performance under diverse conditions, whether in the lab or during field operations — currently exists for a number of existing AI paradigms. Any aura of scientific reliability might in fact be based on algorithmic flaws. Many current AI systems are frequently "brittle" — meaning their narrow applications can generate "dumb results" when activated or projected outside of initial constraints.

AI researchers are grappling with a *replication crisis,* a term coined close to two decades ago when researchers were facing a similar challenge in the fields of chemistry, social psychology, medicine, and others (Baker 2016). According to Nicolas Rougier, a computational neuroscientist at France's National Institute for Research in Computer Science and Automation in Bordeaux, reproducibility is not guaranteed just because AI applications are built by code (Hutson 2018). In addition, researchers often do not share their source code. While emerging movements have encouraged publishing algorithms, or making them open source, this approach has come under fire (see Brundage et al. [2018]) due to concerns that code might be used by parties with nefarious intentions.

The need for software engineering validation and verification is particularly acute for law enforcement and military applications with respect to accountability and liability issues. Employment of AI within future battlespaces could create new and unexpected operational risks, such as potential malfunctions, adversarial interference and/or counterattacks, or unexpected emergent behaviors (Scharre 2016). For example, the recently developed algorithmic capacity to create indistinguishable counterfeits of audio and video demonstrates how quickly new and unexpected threats can arise.[15] Elsa B. Kania, who has written extensively on China's aggressive use of AI in the development of its future military might (2017), has noted the country's focus on reliability considerations, quoting a Chinese Academy of Sciences researcher: "What the military cares most about is not fancy features. What they care most is the thing does not screw up amid the heat of a battle" (2018).

Trust

Trust has historically been a social contract, based on our understanding of how people around us think and our experiences of their behaviors toward us and others. AI-enhanced technologies and human-machine interactions can challenge that convention. In the civilian sector, trust seemingly exists everywhere. Consumers invite virtual personal assistants such as Amazon's Alexa and the Internet of Things (IoT) into their personal living spaces; travelers assume that it is safe to journey on airplanes equipped with autopilot; and patients trust in certain types of data-driven medical diagnoses and treatment options. Civilians can even place too much trust in AI and robotics to the point of risking their security and physical safety (Booth et al. 2017).

In the military sector, however, human operators need to understand and trust AI enough to leverage it effectively in a combat role. Too much trust could mean that soldiers do not sufficiently question AI assistance. For example, during the 2003 invasion of Iraq, the downing of a British Tornado aircraft on March 23 (Loeb 2003) was found to be due to "automation bias," an unwarranted and uncritical trust in automation that led to control responsibility being ceded to a machine (Hawley 2011). A subsequent internal army investigation criticized the Patriot community culture for "reacting quickly, engaging early, and trusting the system without question" (Hawley 2007, 4).

Conversely, too little trust — often due to a lack of explainability or transparency — can likewise have tragic results. For example, the crash of Air France Flight 447 on June 1, 2009, killing all 228 people on board, was likely caused by pilot misunderstanding of AI-generated data — a problem of transparency that likely would not have existed in a similar situation on a simpler aircraft (Scharre 2016). Aware that future fighters need to understand, appropriately trust, and effectively manage an emerging generation of AI-machine partners, in 2017 the Defense Advanced Research Projects Agency (DARPA) initiated the Explainable AI (XAI) program to help humans understand how AI works and why it reached the decision(s) it did (Gunning 2018).

Summary and Future Considerations

Technology is developing at a rapid pace, and ethical, social, and legal gaps are widening because of the slower process of policymaking. Certainly, the civilian world has ethical quandaries to face: an often-cited dilemma facing self-driving cars is related to the philosophical "trolley problem" — referring to an autonomous vehicle's choice between killing five people on the tracks versus one person off to the side (Lin 2016). However, the risk to human lives associated with the military use of AI, robotics, and machine and deep learning for defense and security purposes raises ethical concerns to a much higher level.

We propose our framework of ethical considerations and questions as a means to initiate an early and meaningful discussion. We further suggest that part of that baseline discussion will need to address the current lack of clear definitions and common language (National Academy of Sciences 2018). This might be a challenging first step, given that AI as a field of research has numerous subcomponents that coexist with a wide range of interested parties possessing varied and sometimes opposing perspectives and terminologies.

The subsequent development of a professional military code of ethics and policy for AI will need to be a thoughtful and inclusionary process that proceeds carefully toward policy development, regulation, and oversight. Fundamental questions will need to guide the process: Which values should inform the design of ethical statements and standards for AI? How and by whom should these statements and standards be implemented and enforced? How do we measure or assess AI performance? Who will be legally responsible? How should design values be weighed or adjusted in the face of conflict, and by whom? How will behavior by different cultures and political systems — specifically, Russia and China[16] — inform and influence values? What impact might less-ethical use of AI technologies have on our own ethical resolve?

Much of the contemporary energy and discussion surrounding AI has revolved around dire apocalyptic warnings associated with a perceived inevitable march toward LAWS (Kerr 2017). It would be more productive if supporters and opponents of autonomous systems could come to an agreement on what capabilities actually exist, and what they will reasonably be in the future. Otherwise, devoting disproportionate attention and resources to an unlikely AI apocalypse could distract policymakers from addressing AI's more immediate challenges cited earlier in this article and, furthermore, discourage research on AI's numerous social and legal impacts. Also pressing are threats posed by the adversarial use of AI technologies by nonstate actors such as hackers, terrorists, black marketeers, and drug cartels as well as by competitive nation states, likely necessitating proactive policies.

There are many reasons why militaries should invest in AI and robotics technologies — greater speed, accuracy (and therefore civilian and soldier safety), efficiency, extended reach, multilevel coordination — but the human consequences of military actions necessitate early consideration of the ethical implications of choices that will be made. When ethical and policy analysts consider the repercussions of machine learning advances, they are in essence trying to peer into the future: they are trying to plan for the world of tomorrow by anticipating issues and acting today. We hope to contribute some clarity to that unknown world of tomorrow by offering this framework of ethical considerations to technology developers, policymakers, decision makers, and other stakeholders so they can identify and broadly consider potential military ethical issues. It will take time to address the technical, institutional, legal, and regulatory elements of a national or international AI code of ethics, but acting early on ethical issues and gathering the endorsement of key players is imperative in order to develop a cohesive and forward-thinking strategy.

Acknowledgements

The authors wish to acknowledge Paul Comeau, Chief Scientist, for his guidance of, and support for, this project and thank Anton Minkov, strategic analyst, Office of the Chief Scientist, for his helpful comments. We also wish to thank the Mitacs Canadian Science Policy Fellowship Program. The opinions expressed in this article are those of the authors and do not represent the official position of the Department of National Defence, the Government of Canada, or any of its other departments and agencies.

Notes

1. Currently, *robotics* refers to machines that are capable of carrying out a series of actions on behalf of humans, typically operating without possession of any AI.

2. This connection between AI and robotics is called the *embodiment problem,* and many researchers in AI agree that intelligence and embodiment are tightly coupled issues (Baillie 2016).

3. "Dual-use technologies like artificial intelligence or synthetic biology … have the potential to be used in both good and evil ways. While the technologies themselves are not the subject of treaties and conventions, we are now faced with controlling the proliferation of weapons employing these technologies" (Latiff 2016, 87).

4. Artificial general intelligence is a proportionately small and much more challenging research area within AI that seeks to build machines with general cognitive abilities that can go far beyond performing specific tasks. It is AGI rather than AI that has garnered high public visibility, uncertainty, and fear disproportionate to its size or success (JASON 2017). While the Campaign to Stop Killer Robots is fueled largely by concerns about future LAWS possessing AGI-like cognitive abilities that allow independence from human control, AGI has not been developed and is considered unlikely in the near future (Stone et al. 2016).

5. Ethics refer to the principles that govern a person's behavior or their oversight of an activity — that is, questions about what we should or ought to do — as well as general concerns related to social, political, legal, and cultural impacts and risks (Lin, Bekey, and Abney 2008).

6. Such as the 2018 Artificial Intelligence, Ethics, and Society conference hosted by the Association for the Advancement of Artificial Intelligence and cohosted by the

Association for Computing Machinery, www.aies-conference.com.

7. See the Montreal Declaration on the Responsible Development of Artificial Intelligence, produced at the 2017 Forum on the Socially Responsible Development of Artificial Intelligence, nouvelles.umontreal.ca/en/article/2017/11/03/montreal-declaration-for-a-responsible-development-of-artificial-intelligence/.

8. See deepmind.com/applied/deepmind-ethics-society

9. See www.partnershiponai.org.

10. We have chosen to use a broad definition of ethics because evidence-informed policymaking also has a broad base — considering elements of societal and political pressures, resources, safety and security, and cultural norms.

11. Machine autonomy exists on a spectrum. Our definitions specific to autonomy adopt the following approach. *Semiautonomous* or *human in the loop* indicates that a weapons system waits for human command and permission before taking action. *Supervised autonomy* or *human on the loop* refers to systems that may track, target, and act defensively, but that are supervised by humans who can monitor and, if necessary, intervene in the weapon's operation, as with, for example, the Phalanx Close-In Weapons System, which is used to defend ships against incoming enemy missiles. *Full autonomy* or *human out of the loo*p refers to when human input activates a weapon that then selects and engages targets without further operator intervention, for example, the Harpy drone. Full autonomy that is based on AGI refers to LAWS. While it is accurate to say there are a number of weapon systems in existence today that can perform independent actions, these systems act in accordance with a defined rule set based on complex sensor(s) input, and thus, would be better described as *automated*.

12. Other technologies we have addressed in brief overviews include object recognition, facial recognition, and gait recognition; using AI to monitor mental health; sentiment analysis; AI for dis/misinformation; robotic casualty evacuation; robotic telesurgery; robotics and sensors for IED, explosive, and chemical detection; and smart cities.

13. Please note that our research supports future policy development, which is why we included a "policy implications" category.

14. Canada officially supports the term "appropriate human involvement" (Canada's National Statement 2016), introduced in 2016 as a bridge between the terms "meaningful human control" and "appropriate human judgment."

15 In 2017, Adobe demonstrated a new product that, with just 10 minutes of audio, can exactly replicate a person's voice in limitless artificial audio (Carter, Kinnucan, and Elliot 2018).

16. Notably, China has adjusted its strategic focus from yesterday's informatized ways of warfare to tomorrow's intelligentized warfare, for which AI will be critical (Kania 2017). Russia has already demonstrated its willingness to engage in information warfare (Floridi and Taddeo 2014) during the 2016 US presidential election and its ability to target more than 10,000 Twitter users in the US Defense Department (Calabresi 2017).

References

Allen, G. and Chan, T. 2017. Artificial Intelligence and National Security. A US Intelligence Advanced Research Projects Activity Study. Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School.

Airbus. 2017. *A Statistical Analysis of Commercial Aviation Accidents 1958-2016.* Annual Investigative Report. Blagnac Cedex, France: AIRBUS S.A.S. flightsafety.org/wp-content/uploads/2017/07/Airbus-Commercial-Aviation-Accidents-1958-2016-14Jun17-1.pdf.

Amato, F.; López, A.; Peña-Méndez, E. M.; Vanhara, P.; Hampl, A.; and Havel, J. 2013. Artificial Neural Networks in Medical Diagnosis. *Journal of Applied Biomedicine* 11(2): 47–58. doi.org/10.2478/v10136-012-0031-x.

Anderson, J. M., and Matsumura, J. M. 2015. Civilian Developments in Autonomous Vehicle Technology and Their Civilian and Military Policy Implications. In *Autonomous Systems: Issues for Defence Policymakers,* edited by Andrew P. Williams and Paul D. Scharre, 127–48. Technical Report AD10110077. The Hague, Netherlands: NATO Communications and Information Agency.

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2016. Machine Bias. *ProPublica*, May 23, 2016. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Baillie, J. C. 2016. Why AlphaGo Is Not AI. *IEEE Spectrum*, March 17, 2016. spectrum.ieee.org/automaton/robotics/artificial-intelligence/why-alphago-is-not-ai.

Baker, M. 2016. 1,500 Scientists Lift the Lid on Reproducibility. *Nature* 533(7604): 452–54. doi.org/10.1038/ 533452a.

Banko, M., and Brill E. 2001. Scaling to Very Very Large Corpora for Natural Language Disambiguation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, 26–33. San Francisco: Morgan Kaufmann.

Beeby, D. 2018. Liberal Government Looks to Update Fight Against Online Child Porn. *CBC News*, January 10, 2018. www.cbc.ca/news/politics/child-pornography-online-sexploitation-rcmp-goodale-public-safety-spencer-1.4477563.

Bogost, I. 2015. The Cathedral of Computation. *The Atlantic*, January 15, 2015. www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300.

Booth, S.; Tompkin, J.; Pfister, H.; Waldo, J.; Gajos, K.; and Nagpal, R. 2017. Piggybacking Robots: Human-Robot Overtrust in University Dormitory Security. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 426–34. New York: Association for Computing Machinery. doi.org/10.1145/2909824.3020211.

Brundage, M.; Avin, S.; Clark, J.; Toner, H.; Eckersley, P.; Garfinkel, B.; Dafoe, A.; Scharre, P.; Zeitzoff, T.; Filar, B.; et al. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.* Workshop Report. arXiv preprint arXiv: 1802.07228 [cs.AI]. Oxford, UK: Future of Humanity Institute, Centre for the Study of Existential Risk, Centre for the Future of Intelligence.

Buchanan, B., and Miller, T. 2017. *Machine Learning for Policymakers: What It Is and Why It Matters.* The Cyber Security Project. Cambridge, MA: Harvard Kennedy School, Belfer Center for Science and International Affairs.

Buolamwini, J., and Gebru T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research* 81: 1–15.

Calabresi, M. 2017. Inside Russia's Social Media War on America. *Time*, May 18, 2017. time.com/4783932/inside-russia-social-media-war-america.

Campolo, A.; Sanfilippo, M.; Whittaker, M.; and Crawford, K. 2017. *AI Now 2017 Report.* Edited by A. Selbst and S. Barocas. New York: New York University, AI Now Institute. assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf.

Canada's National Statement. 2016. Presented at the Experts Meeting on Lethal Autonomous Weapons Systems Convention on Certain Conventional Weapons (CCW). Geneva, Switzerland, April 11–15. www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2016 /meeting-experts-laws/statements/11April_Canada.pdf.

Carter, W. A.; Kinnucan, E.; and Elliot, J. 2018. *A National Machine Intelligence Strategy for the United States: A Report of the CSIS Technology Policy Program.* Washington, DC: Center for Strategic and International Studies.

Cellan-Jones, R. 2017. Robot Police Officer Goes on Duty in Dubai. *BBC News Technology*, May 24, 2017. www.bbc.com/news/technology-40026940.

Chung, T. 2018. *Offensive Swarm-Enabled Tactics (OFFSET)*. Defense Advanced Research Projects Agency, Program Information. Washington, DC: US Department of Defense. www.darpa.mil/program/offensive-swarm-enabled-tactics.

Coppola, G.; Beene, R.; and Hull, D. 2018. Arizona Became Self-Driving Proving Ground Before Uber's Deadly Crash. *Bloomberg Technology*, March 20, 2018. www.bloomberg. com/news/articles/2018-03-20/arizona-became-self-driving-proving-ground-before-uber-s-deadly-crash.

Couzin, I. D., and Krause, J. 2003. Self-Organization and Collective Behavior in Vertebrates. In *Advances in the Study of Animal Behavior,* edited by P. Slater, J. Rosenblatt, C. Snowdon, and T. Roper, 1–75. Boston, MA: Academic Press.

Defence Ethics Programme (DEP). 2015 About the Defence Ethics Programme. Department of National Defence. www. forces.gc.ca/en/about-policies-standards-defence-admin-orders-directives-7000/ 7023-1.page#int.

Deputy Assistant Secretary of the Army for Research and Technology (DASA R&T). 2017. *Emerging Science and Technology Trends: 2017-2047. A Synthesis of Leading Forecasts*. Unclassified Report. Providence, RI: FutureScout.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2016. *Ethically Aligned Design: A Vision for Prioritizing Wellbeing with Artificial Intelligence and Autonomous Systems*, Version 1. Piscataway, NJ: Institute for Electrical and Electronics Engineers. standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2017. *A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*, Version 2. Piscataway, NJ: Institute for Electrical and Electronics Engineers. standards. ieee.org/develop/indconn/ec/ead_ brochure _v2.pdf.

European Parliament. 2016. *European Civil Law Rules in Robotics*. Study prepared for the European Parliament's Committee on Legal Affairs. Brussels, Belgium: Policy Department for Citizens' Rights and Constitutional Affairs. www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU%282016%29571379_EN.pdf.

Executive Office of the President. 2016a. *The National Artificial Intelligence Research and Development Strategic Plan*. Washington, DC: National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee. www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf.

Executive Office of the President. 2016b. *Preparing for the Future of Artificial Intelligence*. Washington, DC: National Science and Technology Council, Committee on Technology. obamawhitehouse.archives. gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

Floridi, L., and Taddeo, M., eds. 2014. *The Ethics of Information Warfare*. Law, Governance and Technology 14. New York: Springer. doi.org/10.1007/978-3-319-04135-3.

Future of Life Institute. 2017. Asilomar AI Principles. Winchester, MA: Future of Life Institute. futureoflife.org/ai-principles.

Gallacher, J. D.; Barash, V.; Howard, P. N.; and Kelly, J. 2017. *Junk News on Military Affairs and National Security: Social Media Disinformation Campaigns Against US Military Personnel and Veterans*. Data Memo 2017.9. Oxford, UK: University of Oxford, Project on Computational Propaganda. comprop.oii.ox.ac.uk/research/working-papers/vetops/.

Garvie, C.; Bedoya, A.; and Frankle J. 2016. The Perpetual Line-Up: Unregulated Police Face Recognition in America. Washington, DC: Georgetown Law School, Center on Privacy and Technology. www.law.georgetown.edu/privacy-technology-center/publications/the-perpetual-line-up/.

General Data Protection Regulation (GDPR). 2018. 2018 Reform of EU Data Protection Rules. Brussels, Belgium: European Union, European Commission. ec.europa. eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en.

Gunning, D. 2018. Explainable Artificial Intelligence (XAI). Program Information. Washington, DC: Defense Advanced Research Projects Agency. www.darpa.mil/program/explainable-artificial-intelligence.

Hawley, J. K. 2007. *Looking Back at 20 Years of MANPRINT on Patriot: Observations and Lessons*. Technical Report ARL-SR-0158. Adelphi, MD: Army Research Laboratory. www.arl.army.mil/arlreports/2007/ARL-SR-0158.pdf.

Hawley, J. K. 2011. Not by Widgets Alone: The Human Challenge of Technology-intensive Military Systems. *Army Forces Journal*, February 1, 2011. www.armedforcesjournal.com/not-by-widgets-al.

Higgins, F.; Tomlinson, A.; and Martin, K. M. 2009. Threats to the Swarm: Security Considerations for Swarm Robotics. *International Journal of Advances in Security* 2(2–3): 288–97.

House of Commons. 2016. *Robotics and Artificial Intelligence: Fifth Report of Session 2016–17*. Together with formal minutes relating to the report. London: UK Parliament, House of Commons, Science and Technology Committee. publications.parliament. uk/pa/cm201617/cmselect/cmsctech/145/145.pdf.

Hurst, J. 2017. Robotic Swarms in Offensive Maneuver. *Joint Force Quarterly* 87(4):105-11.

Hutson, M. 2018. Artificial Intelligence Faces Reproducibility Crisis. *Science* 359(6377): 725–26. doi.org/10.1126/science.359.6377.725.

Immigration, Refugees and Citizenship Canada (IRCC). 2018. *Digital Transformation at IRCC: Benefits, Risks and Guidelines for the Responsible Use of Emergent Technologies*. White Paper. Ottawa, Canada: Government of Canada, Strategic Policy and Planning.

JASON. 2017. *Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD*. Technical Report JSR-16-Task-003. McLean, Virginia: The MITRE Corporation.

Kania, E. B. 2017. *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power*. CCW Report, Ethical Autonomy Project. Washington, DC: Center for a New American Security. www.cnas.org/publications/ reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power

Kania, E. B. 2018. Chinese Sub Commanders May Get AI Help for Decision Making. *Defence One*, February 12, 2018. www. defenseone.com/ideas/2018/02/chinese-sub-commanders-may-get-ai-help-decision-making/145906.

Kerr, I. 2017. Weaponized AI Would Have Deadly, Catastrophic Consequences. Where Will Canada Side? *The Globe and Mail,* November 6, 2017. www.theglobeandmail. com/opinion/weaponized-ai-would-have-deadly-catastrophic-consequences-where-will-canada-side/article36841036.

Latiff, R. H. 2016. *Future War: Preparing for the New Global Battlefield*. New York: Alfred A. Knopf.

Lazer, D. M. J.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M. J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; Schudson, M.; Sloman, S. A.; Sunstein, C. R.; Thorson, E. A.; Watts, D. J.; and Zittrain, J. L. 2018. The Science of Fake News. *Science* 359(6380): 1094–6.doi.org/10.1126/science.aao2998.

Lin, P. 2016. Why Ethics Matters for Autonomous Cars. In *Autonomous Driving: Technical, Legal and Social Aspects,* edited by M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, 69-85. New York: Springer Nature. doi.org/10.1007/978-3-662-48847-8_4.

Lin, P.; Bekey, G.; and Abney K. 2008. *Auton-*

omous Military Robotics: Risk, Ethics, and Design. Investigative Report. Washington, DC: US Department of the Navy, Office of Naval Research. digitalcommons.calpoly.edu/cgi/viewcontent. cgi?article=1001&context=phil_fac.

Loeb, V. 2003. Patriot Downs RAF Fighter. *The Washington Post*, March 24, 2003. www.washingtonpost.com/archive/ politics/2003/03/24/patriot-downs-raf-fighter/d231ba70-080a-450b-a12c-ba6e4ee2e10f/?utm_term=.a676939a4b29.

MIT Media Lab. 2017. MIT Media Lab to Participate in $27 Million Initiative on AI Ethics and Governance. *MIT News*, January 10, 2017. news.mit.edu/2017/mit-media-lab-to-participate-in-ai-ethics-and-governance-initiative-0110.

Mitchell, A., and Diamond, L. 2018. China's Surveillance State Should Scare Everyone. *The Atlantic*, February 2, 2018. www.theatlantic.com/international/archive/2018/02/china-surveillance/552203.

Mlot, N. J.; Tovey, C. A.; and Hu, D. L. 2011. Fire Ants Self-Assemble into Waterproof Rafts to Survive Floods. *Proceedings of the National Academy of Sciences of the United States* 108(19): 7669-73. doi.org/10.1073/pnas.1016658108.

National Academy of Sciences. 2018. *The Frontiers of Machine Learning: 2017 Raymond and Beverly Sackler U.S.-U.K. Scientific Forum*. Washington, DC: The National Academies Press. https://www.nap.edu/read/25021/chapter/1.

Osoba, O., and Welser, W. 2017. *An Intelligence in Our Image: The Risk of Bias and Errors in Artificial Intelligence*. Technical Report RR-1744-RC. Santa Monica, CA: RAND Corporation. doi.org/10.7249/RR1744.

Rosenberg, M.; Confessore, N.; and Cadwalladr, C. 2018. How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*, March 17, 2018. www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.

Rubenstein M.; Cornejo, A.; and Nagpal, R. 2014. Programmable Self-Assembly in a Thousand-Robot Swarm. *Science* 345(6198): 795–99. doi.org/10.1126/science.1254295.

Sanger, D. E., and Schmitt, E. 2014. Snowden Used Low-Cost Tool to Best N.S.A. *The New York Times*, February 8, 2014. www.nytimes.com/2014/02/09/us/snowden-used-low-cost-tool-to-best-nsa.html.

Scharre, P. 2014. *Robotics on the Battlefield Part II: The Coming Swarm*. CCW Report, Ethical Autonomy Project. Washington, DC: Center for a New American Security. www.cnas.org/publications/reports/robotics-on-the-battlefield-part-ii-the-coming-swarm.

Scharre, P. 2016. *Autonomous Weapons and Operational Risk*. CCW Report, Ethical Autonomy Project. Washington, DC: Center for a New American Security. www.cnas.org/publications/reports/autonomous-weapons-and-operational-risk.

Smalley, D. 2015. LOCUST: Autonomous, Swarming UAVs Fly into the Future. *Office of Naval Research News and Media Center*, April 14, 2015. www.onr.navy.mil/Media-Center/ Press-Releases/2015/LOCUST-low-cost-UAV-swarm-ONR.aspx.

Solis, G. D. 2016. *The Law of Armed Conflict: International Humanitarian Law in War*. Cambridge, UK: Cambridge University Press. doi.org/10.1017/CBO9781316471760.

Stone, P.; Brooks, R.; Brynjolfsson, E.; Calo, R.; Etzioni, O.; Hager, G.; Hirschberg, J.; Kalyanakrishnan, S.; Kamar, E.; Kraus, S.; Leyton-Brown, K.; Parkes, D.; Press, W.; Saxenian, A.; Shah, J.; Tambe, M.; and Teller, A. 2016. 2016. *Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence*. Report of the 2015–2016 Study Panel. Stanford, CA: Stanford University.

UK Atomic Energy Authority. 2016. Written Testimony Submitted by RACE, UK Atomic Energy Authority (ROB0041). In *Robotics and Artificial Intelligence: Fifth Report of Session 2016–17*. London, UK: UK Parliament, House of Commons, Science and Technology Committee. data.parliament.uk/ writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-c ommittee/robotics-and-artificial-intelligence/written/32640.htm.

US Department of Defense. 2017. Department of Defense Announces Successful Micro-Drone Demonstration. Press Release NR-008-17. January 9, 2017. Washington, DC: US Department of Defense. www.defense.gov/News/News-Releases/News-Release-View/Article/1044811/department-of-defense-announces-successful-micro-drone-demonstration.

Villani, C. 2018. For a Meaningful Artificial Intelligence. Towards a French and European Strategy. Paris: Villani Mission on Artificial Intelligence. www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf

Wertheimer, R., ed. 2010. *Empowering Our Military Conscience: Transforming Just War Theory and Military Moral Education*. Burlington, VT: Ashgate Publishing, Ltd.

Wright, D. 2011. A Framework for the Ethical Impact Assessment of Information Technology. *Ethics and Information Technology* 13(3): 199-226. doi.org/10.1145/968261.968263.

Sherry Wasilow received her PhD in communication from Carleton University (2017), where her research addressed Canadian military and media relations through an examination of embedded reporting, using Canada's involvement in the Afghanistan War as a case study. She also holds an MA in journalism from the University of Texas at Austin, a graduate diploma in journalism from Concordia University, and a BA in political science from the University of Calgary. Wasilow's professional background includes communicating addiction-science research to nonscientists, health planning and policy, and media and legislative analysis. Wasilow investigated the ethical and policy implications of emerging AI and robotics technologies for the military as a policy analyst with Defence Research and Development Canada in the Department of National Defence.

Joelle B. Thorpe received her PhD in psychology, neuroscience, and behaviour from McMaster University in Hamilton, Ontario, Canada (2013), and has a master of science degree in biology from Queen's University in Kingston, Ontario, Canada (2009). During her time as a clinical research associate from 2014 to 2016, Thorpe developed an interest in ethics and sat as a board member on the Queen's University and Affiliated Teaching Hospitals Health Sciences Research Ethics Board. In 2016 and 2017, Thorpe completed a Mitacs Canadian Science Policy Fellowship with Defence Research and Development Canada in the Department of National Defence, where she investigated the ethical and policy implications of emerging human enhancement technologies in the military. Thorpe is currently employed as a policy analyst in the Office of the Chief Scientist at Defence Research and Development Canada.