CHARTING THE WAY FORWARD

EXAMINING THE PRINCIPLES OF RESPONSIBLE USE OF MILITARY AI BY CANADA'S ALLIES

by Branka Marijan, PhD

SEPTEMBER 2021

PROJECT PLOUGHSHARES

Charting the way forward

Examining the principles of responsible use of military AI by Canada’s allies

Author: Branka Marijan, PhD
Research Assistants: Grace Wright and Emily Standfield
Editor: Wendy Stocker
Designer: Tasneem Jamal

First published: September 2021

With funding from the Mobilizing Insights in Defence and Security (MINDS) program
of the Canadian Department of National Defence

Project Ploughshares
140 Westmount Road North
Waterloo, Ontario N2L 3G6
Canada
Telephone: 519-888-6541
Email: plough@ploughshares.ca

TABLE OF CONTENTS

SUMMARY

Many democratic countries are eager to collaborate with like-minded states in developing artificial intelligence (AI) defence capabilities, but are also starting to recognize the need for the international community to establish basic global principles or standards that would together constitute a regulatory regime. Indeed, a multilateral regulatory effort and greater international cooperation are necessary to address the growing competitive use of AI for defence purposes and potential misperceptions about intent and capability.

This report concentrates on the example of the United States-led AI Partnership for Defense (AIPfD), which commonly cites as reasons for developing AI applications for defence the geopolitical threats from actors such as China and Russia, and the need to automate processes that are dangerous, repetitive, or cannot be adequately supported by staff. The AIPfD currently includes: Canada, Australia, Denmark, Estonia, Finland, France, Israel, Japan, Norway, the Republic of Korea, Sweden, the United Kingdom, and the United States.

At the same time, the AIPfD acknowledges the need for normative leadership, aiming to set global standards for defence applications of AI. It expresses a commitment to the development of trustworthy, human-centric technologies and responsible uses of AI-supported systems. As well, the group aims to promote greater transparency and accountability for the effects of such systems. While the AIPfD is concentrated on like-minded states and Western allies, the norms that it promotes are likely to shape the wider standards adopted by the global community on AI in military applications.

There is a growing global consensus that all AI technology should exhibit the characteristics of transparency, justice and fairness, non-maleficence, and privacy. While a specific blueprint of responsible AI in defence applications has not yet emerged, shared commitments to reliable technologies that operate with an appropriate role for human judgement and experience are increasingly accepted. However, as the European Union and the United States appear to have different approaches to developing commercial AI, it seems likely that each will also approach AI in defence applications from a different perspective. There is already an economic competition among allies, which will also have an impact on developments in military AI.

Government experts from different countries express a range of views, with some pushing for faster and widespread adoption of AI in defence, while others approach AI for defence more cautiously. Critically, balancing normative commitments with security interests—and, in particular, disclosing capabilities and functioning of systems—will need to be thought-

fully addressed. Finally, ethical concerns about the use of AI feature prominently in discussions among democratic countries and in legislation for various domestic applications. These norms and standards will also have some impact on various state positions on defence uses of AI. As standards and commitments might differ across countries and regions, international norms and standards will be necessary to build confidence and avoid deployment of systems that are not reliable.

Ultimately, the real test of expressed commitments will be in the behaviour that follows and in the engagement of countries that are considered adversaries. The AIPfD should consider the limits to allies-only engagement, and the security implications of potential misperceptions among adversaries. A more globally oriented process would allow for ownership and a stake in norm development by a greater number of countries.

INTRODUCTION

Responsible uses of artificial intelligence (AI) feature prominently in national discussions and increasingly in multilateral forums. However, the definition of “responsible use of AI” varies from state to state. Often the terms "ethical" and "responsible" are used interchangeably. While there is considerable overlap, we believe that there are also differences. However, in this report, we will not be focused on the differences but on the commonalities.

AI is a multi-purpose technology that has many uses, some of which need to be closely regulated. With this in mind, many countries have started to draft policy directives that outline the principles of responsible use of AI. But they face real challenges.

One of the key challenges is that most responsible AI discussions do not specifically address military applications. As some experts have explained, AI use in military technology is in the early stages and there is a great deal of uncertainty about just how it will be developed. What is clear is that militaries around the world see AI as critical to future warfare, with AI tools allowing for more efficient and speedier responses. What is not clear is who will develop global standards for AI defence applications. Until a regulatory regime is established and accepted by the majority of states, the widespread incorporation of AI technology in military settings remains a subject of concern and even fear.

In October 2020, the United States launched the AI Partnership for Defense (AIPfD), a multinational effort of 13 like-minded democratic countries, including Canada. The partnership is meant to create standards of responsible uses of AI, which would lead to better integration and interoperability among military partners. The partnership is also widely seen as a counterbalance to AI efforts by China and Russia. But the work of the AIPfD could have broader effects, establishing global normative positions on acceptable and unacceptable behaviours and uses of AI.

This paper provides an overview of the positions of AFPfD members, plus Germany, Spain, and the European Union, on responsible uses of AI. It also includes specific commitments that are emerging from the Partnership as a whole. Information was garnered by examining various states' policies and literature from think tanks and academia; expert analysis was acquired through a survey and webinar discussions.

AI governance frameworks are still in process and so this paper should be viewed as only an initial overview of key developments. Still, it is important for countries like Canada to understand this nascent regulatory landscape in preparation for the accelerating adoption of AI technologies.

Crucially, while no single blueprint on responsible AI development in defence has emerged, there is a convergence of views in national efforts by members of the AIPfD on the development of trustworthy, human-centric technologies and responsible uses of systems. At the same time, norms regarding transparency, for example, will be challenged in practice by the efforts of various countries to safeguard information about capabilities and uses. Still, clearer policies and the release of national strategies on defence applications of AI would be useful to start crafting broader standards. Balancing the perceived and potential national security advantages of AI with commitments to responsible use and development are at the crux of these governance efforts.

CONTEXT

The national AI strategies of most AIPfD members and other like-minded states are still under development. The United States and the European Union aim to lead in shaping the norms and policies that relate to responsible use of AI, but each has so far taken a different approach to regulating commercial, security, and defence uses of AI.

The United States seems to take a more industry-focused, voluntary approach in the commercial sector. This approach seems to carry over to defence applications, with the prioritization of domestic industry and competitiveness. The US is undoubtedly the leader in shaping the regulatory landscape in terms of AI defence applications. Still, the various agencies and experts involved also view defence applications of AI in a variety of ways. Among the key proponents of AI for defence is the National Security Commission on Artificial Intelligence (NSCAI), created by the US Congress in 2018. Its March 2021 Final Report highlighted the view that AI is central to US geopolitical interests and pushed for greater adoption of AI to maintain a US military advantage.

The EU has a more regulation-oriented approach that focuses on ethics and rights. However, it should be noted that some EU members, such as France, appear eager to place AI technology on the battlefield. Other member states, such as Germany, are more cautious in adopting new technologies for their military. These different approaches seem likely to produce a fragmented EU response to AI applications for defence.

To date, national AI strategies seem to contain general references to defence uses of AI rather than well-developed policies. Mostly, they focus on responsible AI use in healthcare, governance, and business, with defence usually excluded. Some broader efforts, such as the Organisation for Economic Co-operation and Development (OECD) 2019 AI principles, have shown a wider commitment to "robust, safe, fair and trustworthy" AI. Although the OECD did not focus on defence applications, they participate in the wider governance response to the development and use of AI. Different initiatives, documents, and stated positions all provide insight into how countries are considering responsible defence applications of AI.

Noteworthy is the ongoing United Nations dialogue on autonomous weapons, which began in 2014 at the Convention on Certain Conventional Weapons (CCW). Here, discussion on AI relates to the implications of greater autonomy in the selection and engagement of targets. So far, most of the members of the AIPfD are committed to ensuring that lethal force is directed by a human operator. What human control means, however, varies, with some states supporting significant levels of human input while others believe less human involvement is appropriate. The positions that the members of the AI Partnership for Defense take in CCW discussions—and all international discussions—will likely be shaped by their engagement with other AIPfD members. Some countries, like Canada, could come under more pressure to conform to partners' positions at the CCW.

The AIPfD will also likely attempt to ensure the interoperability of the AI defence tech of all members. Whether the partners will develop a specific shared position on AI in weapons remains to be seen.

This public commitment to responsible AI by the United States and its allies will be important in controlling the global competition in military AI. At times described as an arms race between the United States and its allies with opponents China and Russia, such a competition is of concern. The oft-cited remark by Russian President Vladimir Putin that the country that leads in AI will rule the world seems to signal a prominent role for AI in geopolitics.

Along with others, Paul Scharre and Heather Roff have noted that, because AI is a general-purpose technology, framing the global competition for AI as an arms race is inaccurate and potentially dangerous. As Scharre notes, “Perceptions of a ‘race’ to field AI systems before competitors do could cause nations to cut corners on testing, leading to the deployment of unsafe AI systems that are at risk of accidents that could cause unintended escalation or destruction.” Thus, efforts to agree upon shared norms and commitments in the use of AI for defence applications are critical.

As the AI Partnership for Defense is for like-minded democracies, it has so far not made efforts to engage with countries outside of the more traditional allies. Nor does it focus on the challenges of creating shared norms on the use of AI for defence between Western-style countries/democracies and China and Russia. In fact, as the March 2021 NSCAI report indicates, the US sees its work with allies as a direct defence of democratic norms and values. Nevertheless, AIPfD efforts to achieve norms within the partnership offer possible ways to develop knowledge building and advance dialogue on a more global level.

It is not clear that Russia or its allies will agree with any of the AIPfD norms. Samuel Bendett notes that there are some indications that Russia is considering particular controls on AI applications, including those used for defence. However, to date, Russia, along with other states, has been uncooperative at international discussions such as the CCW talks on autonomous weapons. Moreover, as Oleg Khramov, Deputy Secretary of the Russian Security Council has pointed out, Russia sees developed countries such as the United States promoting approaches and norms to AI regulation that directly benefit themselves. Khramov indicated that this creates a concern for Russia’s security.

The United States considers China, which views a military-civil fusion of technologies as key to its national security, as the chief competitor in the race to achieve dominance in the field of AI. While this Chinese fusion is, as yet, largely aspirational, Elsa B. Kania and Lorand Laskai have found that misperceptions about the level of a threat can result in policies with negative impacts on technological transfer and scientific engagement. Therefore, a clearer sense of the specific security challenges that China poses are needed to avoid an unnecessarily heavy-handed response by the United States and its allies.

In any event, China must be included in international and regional dialogues on defence applications of AI. China’s position that humans need to be in control of autonomous weapons systems seems to provide some room for engagement. The Beijing Academy of Artificial Intelligence has even released a set of “Beijing AI Principles,” which have, however, been met with skepticism. While the principles are remarkably similar to those proposed by the US and some of its allies, it has been noted that some commitments do not seem to reflect how AI has been used by the Chinese state against minorities. To be fair, though, China is not alone in its misuse of new technologies and there is a concern about the commitments being translated into practice for all countries.

It seems clear that engaging China, Russia, and their allies in the development of normative frameworks will be a challenge. However, as has been the case on other issues in arms control and disarmament, states need to feel ownership and have a stake in the norm development. Otherwise, the norms being promoted will not be adopted by a wide enough group of states. Therefore, AIPfD and other states need to consider the ways in which diplomatic overtures can be made and what specific confidence-building measures will be necessary to contribute to global norms that support international stability.

Interestingly, as Ulrike Franke notes, there is also the concern about "AI nationalism" or protectionist policies that could emerge among allies and are already present in the broader US-China competition. Franke points to the case of the Dutch semi-conductor company ASML that came under US pressure not to export semiconductor chips to China. Such pressures are likely to continue and should also be the subject of dialogue. When economic interests are significant, alliances are going to be tested, as was seen in the case of the response to the announcement of a new security pact by the UK, US, and Australia. In this case, France lost a contract to build submarines for Australia, which had decided to purchase nuclear-powered submarines from the US. France responded by recalling its ambassadors to Australia and the US. This case showed that when economic interests are at stake there will also be competition between allies that could impact their relationship and support for common principles.

Against this background, it may appear that efforts to develop shared norms will be futile. However, it is precisely because of this context that countries need to start broader discussions on global standards on responsible use of AI by militaries. By exploring the ways in which members of the AI Partnership for Defense approach commitments to responsible AI use, we can gain insights into the types of norms that could be palatable to adversaries as well as allies.

NATIONAL AND REGIONAL POSITIONS

The countries and regions included here approach the responsible development of AI applications for defence in a variety of ways. The main focus is on the AI Partnership for Defense, which currently includes the following: Canada, Australia, Denmark, Estonia, Finland, France, Israel, Japan, Norway, the Republic of Korea, Sweden, the United Kingdom, and the United States. Germany and Spain, not currently partners, are included as like-minded states. Finally, the EU, an active leader in AI, is the one regional actor discussed in this paper. Note that Denmark, Estonia, Finland, France, Germany, Spain, and Sweden are all EU member states.

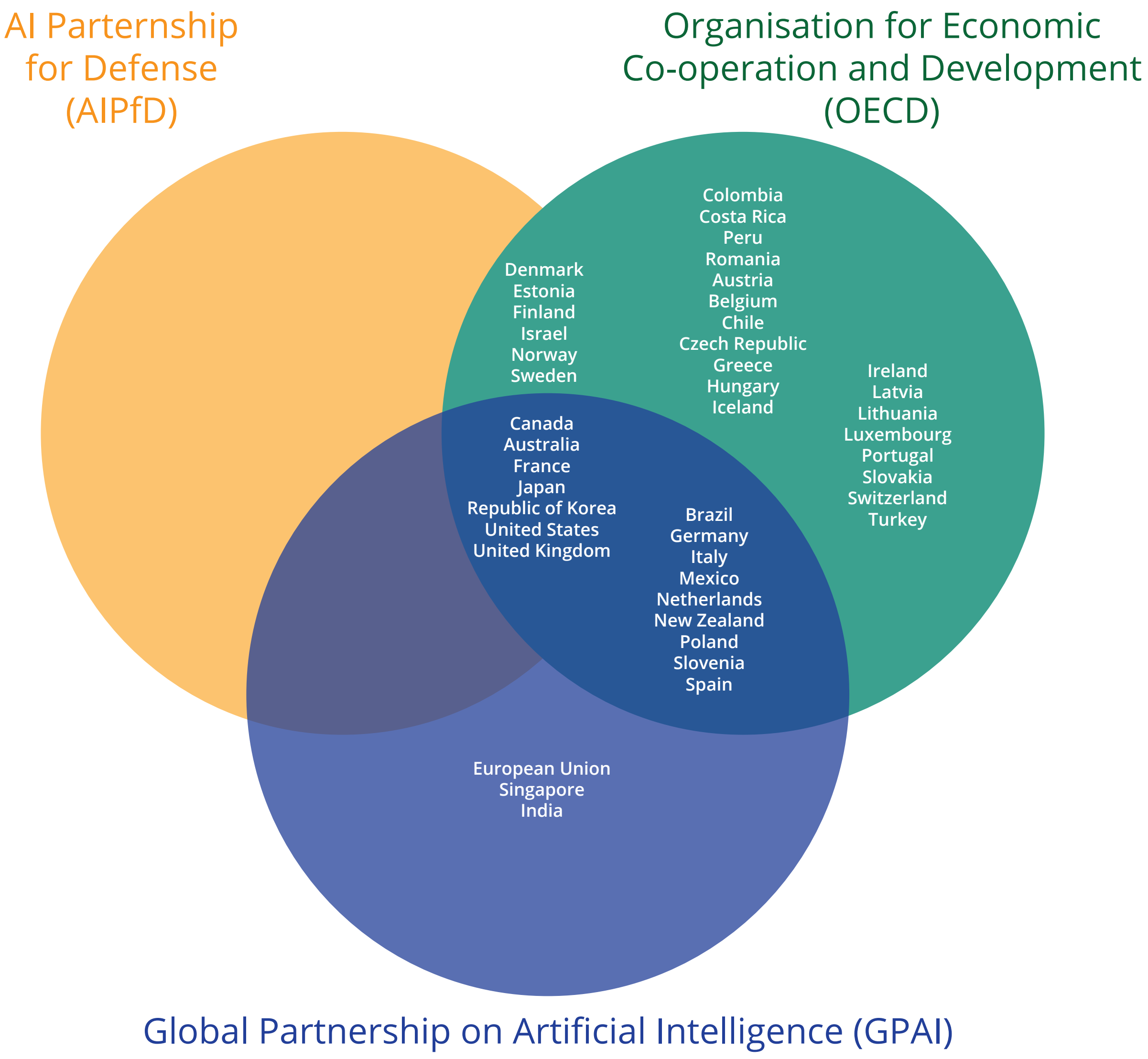

Figure 1. Key regimes on responsible AI

Members of the AI Partnership for Defense

The positions of countries that are members of the AI Partnership for Defense are outlined below. These brief overviews highlight key approaches to AI for defence purposes and note any other relevant positions on broader responsible use of AI.

UNITED STATES

The United States is the foremost Western state involved in discussions on responsible use of AI and was one of the first to consider AI principles for defence applications. It is keen to use AI for such defence capabilities as intelligence collection and analysis, cyber operations, logistics, information operation, semi-autonomous and autonomous vehicles, and command control. It has already operationalized AI in the military and is engaged in multiple AI-based projects, some of which leverage big data, machine learning, and robotics in various domains of warfare.

Other US initiatives focus on nano-drone technology, augmented reality training systems, autonomous emergency medical treatment, data analytics software improvement, and strategic decision-making. The Defense Advanced Research Projects Agency (DARPA) is engaged in numerous AI initiatives, including its Explainable AI program and the AI Next campaign, which is aimed at improving processes and systems across the US Department of Defense (DoD).

The DoD has developed a set of core ethical principles on the use of AI that focus on responsibility, equitability, traceability, reliability, and governability. These principles are to be operationalized across DoD's personnel, processes, partnerships, and policy. The DoD Joint Artificial Intelligence Center (JAIC) is spearheading, among other initiatives, the enhancing of AI literacy of DoD personnel and advising leadership on AI ethical issues.

DoD-shaped principles will likely inform much of the conversation on AI for defence by allied and like-minded countries. However, US leadership could be challenged by the EU, which is developing a reputation as the hub of responsible AI technologies.

AI Strategy/Framework: The United States has developed a National Artificial Intelligence Research and Development Strategic Plan, the National Artificial Intelligence Initiative, a National Strategy for Critical and Emerging Technologies, as well as a specific DoD AI strategy.

The United States is also involved in bilateral and multilateral AI initiatives. The Global Partnership on AI seems to lack a specific focus on security and defence, but some of the standards and norms it sets on AI use could have an impact on the regulation of defence and security applications. As a member of the G7, the United States is committed to the 2018 G7 statement, the Charlevoix Common Vision for the Future of Artificial Intelligence, which aims to promote human-centric AI, respect for privacy and personal data protection, and the involvement of women and underrepresented groups in the development and implementation of AI. The United States also contributes to the D10 initiative of leading democracies. The Five Eyes alliance, to which the United States belongs, has issued a Contested Urban Environment Strategic Challenge that relates to the use of emerging technologies by the military in urban settings.

AUSTRALIA

Australia's AI Ethics Framework does not include military applications. The Department of Defence has developed an AI checklist, an Ethical Risk Matrix, and a Data Item Descriptor that could be used by contractors to develop a Legal, Ethical, and Assurance Program Plan. However, these tools have not been formally endorsed by the Australian Government.

Australia has already begun to deploy AI in military operations. For example, the Australian Navy is employing AI-automated and cognitive assistants as well as AI advisors.

AI Strategy/Framework: In 2019, Australia produced an AI Roadmap that focuses on health, infrastructure, and natural resources. Australia is a founding member of the Global Partnership on AI and participates in the Five Eyes strategic challenge. It supports the D10 initiative on AI.

CANADA

Canada is home to some of the early leaders of AI and Deep Learning, such as Yoshua Bengio and Geoffrey Hinton. The government has invested billions of dollars in AI research and development, fostering AI "clusters" in Edmonton, Montreal, and metropolitan Toronto. As a rich middle power and a member of strategic alliances, Canada is well positioned to influence the conversation on responsible AI development.

Canada's national defence policy *Strong, Secure, Engaged* identifies plans to invest in a range of technologies, including armed aerial systems. It also indicates that "Canada is committed to employing new technological capabilities in a manner that rigorously respects all applicable domestic and international law, is subject to proven checks and balances, and ensures full oversight and accountability" (p. 55).

In 2018, the Department of National Defence launched the Defence Excellence and Security Program, which focuses on AI, cyber, remotely piloted systems, data analytics, human performance, space, and surveillance. The Canadian armed forces already integrate AI into some operations, such as voice assistants on naval warships.

AI Strategy/Framework: In 2017, Canada became the first country to launch a national AI strategy—the Pan-Canadian AI Strategy. One of its four main objectives is to understand the societal implications of AI. Canada is a founding member of the Global Partnership on Artificial Intelligence and is involved in other global AI initiatives, such as the D10 initiative, the National Technology and Industrial Base, and the Five Eyes/TTCP AI strategic challenge. As a member of the G7, Canada supports the Charlevoix Common Vision for the Future of Artificial Intelligence.

DENMARK

Denmark's National Strategy for Artificial Intelligence ranks responsibility, privacy, and ethics most highly when developing AI. While it does not focus on a framework for military applications of AI, it does cite the need to invest in cyber and information security. Denmark has also established an inter-ministerial group to determine how its legislative framework must be altered to incorporate the regulation of AI.

AI Strategy/Framework: Denmark's National Strategy for Artificial Intelligence has four major objectives, including the creation of a common basis for ethical and human-centric AI. Denmark has signed on to the declaration *AI in the Nordic-Baltic Region*, which commits it to collaborate with other northern European states on human-centric AI guidelines, standards, principles, and values. Denmark also belongs to the Council of Europe's Ad hoc Committee on Artificial Intelligence.

ESTONIA

Estonia has focused on the development and use of AI in non-lethal uncrewed vehicles. Since a cyberattack on government and other websites in 2007, Estonia has invested heavily in robust cybersecurity. It now has cybersecurity agreements with several countries, including the United States. Estonia's defence department collaborates with the private sector; a partnership with Wise Guys Cyber offers an accelerator program to develop AI and cybersecurity for defence.

AI Strategy/Framework: Estonia does not appear to have a military-specific AI strategy. It has signed on to the declaration *AI in the Nordic-Baltic Region*. It does not support a categorical ban on autonomous weapons; a "food for thought" paper co-written by Estonia, Finland, France, Germany, and the Netherlands indicates that such a ban is not conducive to advancing ethical conduct in the military. Estonia is a member of the Council of Europe's Ad hoc Committee on Artificial Intelligence.

FINLAND

Finland recognizes the utility of AI in security applications and the need for common ethical principles on the use of AI across industries. It has challenged the private sector “to create ethical principles for AI,” but the implications for defence are unclear. The European Centre of Excellence for Countering Hybrid Threats, based in Finland, already acts as a hub for such AI work.

AI Strategy/Framework: Finland’s national strategy aims to raise its profile in the defence sector. Finland has signed on to the declaration *AI in the Nordic-Baltic Region* and, as a co-author of the “food for thought” paper, does not support a categorical ban on autonomous weapons. It is a member of the Council of Europe’s Ad hoc Committee on Artificial Intelligence.

FRANCE

France, with aspirations of global AI leadership, has one of the most rigorous approaches to AI, actively collaborating with other states and leveraging the expertise of the French private sector. France has identified defence as one of its top AI priorities, and its Ministry of Defence is investing 100 million euros in AI research. France is also working with the World Economic Forum to develop a governance framework for facial recognition technology.

Cédric Villani, leader of a task force appointed by the French Prime Minister to develop an AI strategy for France and Europe, writes about lethal autonomous weapons in *For a Meaningful Artificial Intelligence: Towards a French and European Strategy* (2018): “From a French point of view it is, however, possible to be a driving force behind proposed regulations or the development of good practices without having to forego advanced capabilities *ex ante* or fall behind other States in this important strategic domain” (p. 125).

The French military is investing in six major AI capabilities: decision-making, intelligence, collaborative combat, robotics, cyberspace, and logistics & maintenance. It has already begun testing AI in some of its systems, including those that can detect, recognize, and identify vehicles using infrared imaging. With Germany, France has formed the FCAS (Future Combat Air System) Expert Commission on Responsible Use of Technologies, which aims to determine ethical guidelines based on international law for this developing system.

AI Strategy/Framework: France has adopted a national AI strategy, as outlined in the Villani Report, and also has a national digital security strategy and cybersecurity strategy. As a member of the G7, it is committed to the Charlevoix Common Vision for the Future of Artificial Intelligence and also participates in the D10 initiative. With Canada, France began the Global Partnership on AI. France is a partner in the Trilateral French-Japanese-German Research Projects on Artificial Intelligence and a member of the Council of Europe's Ad hoc Committee on Artificial Intelligence. A co-author of the "food for thought" paper, France opposes a categorical ban on autonomous weapons.

ISRAEL

Israel has a robust technology sector and demonstrates significant collaboration between private industry, the defence department, and the military. The Israeli military views AI as critical to Israel's survival in today's world. However, Israel does not have a clear AI strategy, only some policies. The Israel Innovation Authority is urging the development of an AI strategy to "maintain its leading position."

Israel has already integrated AI into military applications, including the SPICE 250 missile system that uses deep learning, a networked sensor-to-shooter system that leverages computer vision and AI to assist in targeting, and automated robots. The Sigma branch of the Israel Defense Forces is responsible for developing, researching, and implementing the latest AI and advanced software.

There is active cooperation in AI projects between the US and Israeli private sectors. A key interest is in cybersecurity.

AI Strategy/Framework: At the request of the government, a committee prepared a draft report on a national AI plan in 2019. The strategy was launched at the end of 2020, although political turmoil threatened delays.

JAPAN

Japan is eager to integrate AI and other emerging technologies into Japanese society to achieve what it calls "Society 5.0"—a human-centric, sustainable society that leverages technology to blend cyberspace with physical space.

Japan's strategic planning for defence includes using AI and other "potentially game changing" technologies. The Ministry of Defense first identified uncrewed vehicles, cyberspace, and support for decision-making and data processing as immediate priorities for AI applications. Then, in late 2019, "new threats" prompted a new R&D focus on cyber, underwater technologies, the electromagnetic spectrum, hypersonics, wide-area intelligence, surveillance and reconnaissance, and network operations. Japan has already made significant investments in uncrewed drones and submarines and AI-based maritime surveillance platforms.

AI Strategy/Framework: Japan has a national AI strategy and an Integrated Innovation Strategy. As a member of the G7, it is committed to the Charlevoix Common Vision for the Future of Artificial Intelligence and also participates in the D10 initiative. Japan belongs to the Global Partnership on AI and is a partner in the Trilateral French-Japanese-German Research Projects on Artificial Intelligence.

NORWAY

Norway's defence priorities for AI applications focus on defending the "High North," particularly from Russia and with an emphasis on cyber threats. It is interested in developing uncrewed vehicles for security.

AI Strategy/Framework: In 2019, Norway launched a cybersecurity strategy involving government agencies and departments, including Justice and Public Security, Defence, Local Government and Modernisation, and Foreign Affairs. In early 2020, it released its National Strategy for Artificial Intelligence. As it states, this strategy "does not cover the defence sector." However, it does expect to take "a leading position" in applying AI in some areas related to defence, including the maritime and marine industries. Norway has signed on to the declaration *AI in the Nordic-Baltic Region* and is a member of the Council of Europe's Ad hoc Committee on Artificial Intelligence.

REPUBLIC OF KOREA

With its population declining, South Korea is under pressure to maintain a strong and effective military with fewer personnel. It is looking for AI and automation to move into the military realm, which currently relies on mandatory conscription. The world leader in autonomous sentry weapons, it plans to focus AI development primarily on surveillance and reconnaissance, specifically of its border with North Korea. South Korea has already developed a semi-autonomous weapon system to protect the demilitarized zone from North Korean attacks.

While it does not have an explicit code for military uses of AI, it does have guidelines for an "intelligent information society" that is based on the principles of "publicness, accountability, controllability, and transparency."

Like many other nations, South Korea considers civil-military cooperation on AI essential to national defence. Such cooperation, which has caused controversy in the past, exists between leading defence firm Hanwha Systems and the state-run Korea Advanced Institute of Science and Technology in the development of AI for military weapons.

The Ministry of National Defense is focusing on the development of AI applications for surveillance and reconnaissance, training, equipment and asset management and inspection, and medical data. AI is already being used to process battlefield data from a growing number of surveillance systems and make subsequent decisions.

AI Strategy/Framework: South Korea has a national AI strategy. In 2018, it released its plan to restructure and modernize its military forces and defence systems. Defense Reform 2.0 aims to take advantage of new technology to promote "smart defense" that encompasses, inter alia, safety, human resources, housing for troops, and disaster management. South Korea is also a member of the Global Partnership on AI and participates in the D10 initiative.

SWEDEN

Sweden has identified a need to develop national and international rules, standards, norms, and ethical principles to guide AI development and use. Its *National approach to artificial intelligence* recognizes the connections between civil research (especially in cyber) and defence and promotes the exploitation of these synergies.

Sweden has established a Committee for Technological Innovation and Ethics (KOMET) "to identify policy challenges, contribute to reducing uncertainty surrounding existing regulations, and accelerate policy development linked to fourth industrial revolution technologies." Other initiatives on AI ethics include its Data Factory & Arena, a venue that brings together actors from the private and public sectors to work on AI research and innovation, including ethics.

AI Strategy/Framework: Sweden's national AI strategy says little about defence applications, other than to acknowledge the value of synergies between civil and defence research and promote enhanced cyber expertise. Sweden has signed on to the declaration *AI in the Nordic-Baltic Region* and is a member of the Council of Europe's Ad hoc Committee on Artificial Intelligence.

UNITED KINGDOM

The UK is committed to fostering intergovernmental collaboration and cooperation on the research and development of AI. It does not have an overarching AI strategy, but a variety of agreements and guidelines that direct AI research and innovation. The UK government has established a Centre for Data Ethics and Innovation that is charged with ensuring safe and ethical innovation in AI and other data-driven technologies. The Centre has partnered with the Alan Turing Institute, which has a Public Policy Programme that advises the public sector on AI, data science, and ethics. Digital Catapult, a "leading advanced digital technology innovation centre," is involved in "accelerating the adoption of new and emerging technologies" and is also involved in the ethical and responsible adoption of AI in UK industries, including the defence and aerospace sectors.

The Ministry of Defence has stated that it does not possess and does not plan on developing fully autonomous weapon systems; however, it is noteworthy that the UK's definition of lethal autonomous weapon systems (LAWS) does not align with those more commonly used by other NATO members. The UK does not support a LAWS ban and believes that existing international human rights law and the UN Convention on Certain Conventional Weapons provide sufficient regulation. However, the UK has also stated that lethal force must be directed by a human and that humans must be held accountable for its use.

The Ministry of Defence is set to launch a range of innovation programs that make use of AI, machine learning, and autonomous systems to enhance decision-making, situational awareness, surveillance and reconnaissance, data analysis, and information distribution. It has also established a Defence and Security Accelerator that brings together various stakeholders to develop innovative solutions to national security challenges. More recently, GCHQ, a "world-leading intelligence, cyber and security agency," has also created an ethical framework for AI development and use that focuses on integrating principles of fairness, transparency and accountability, empowerment, and privacy.

AI Strategy/Framework: The primary UK AI strategy is its AI Sector Deal, which addresses safety and ethics in part and establishes an AI Council and an interim Centre for Data Ethics and Innovation. As a member of the G7, the UK is committed to the Charlevoix Common Vision for the Future of Artificial Intelligence and contributes to the D10 initiative. The UK also participates in the Five Eyes/TTCP AI strategic challenge, the Global Partnership on AI, and the Council of Europe's Ad hoc Committee on Artificial Intelligence.

A Like-minded Region and Two States

EUROPEAN UNION

The EU aims to be a leader in AI and actively encourages member states to collaborate on AI initiatives, including those for defence. The EU is adamant that ethical principles, particularly those that are human-centric, guide the development and use of AI. In other words, AI must be aligned with values such as transparency, non-discrimination and fairness, accountability and oversight, safety, privacy and sound data governance, and the advancement of societal and environmental well-being. The EU's General Data Protection Regulation provides a standard for data collection, privacy, and consent that is the most rigorous in the world.

While the EU is enthusiastic about using AI in defence applications, it insists that these applications be evidence-based, necessary, proportionate, and show respect for basic human rights. In January 2021, the EU adopted a report calling for a legal framework that covers definitions and ethical principles for the use of AI, including by the military. It also calls for an EU strategy to prohibit lethal autonomous weapons.

The European Union has funded multiple AI-related defence projects related to cyber security, uncrewed aerial vehicles, and hybrid uncrewed systems. These highly collaborative projects have brought together various EU countries and subject matter experts from academia, research centres, and industry. The EU has also published a directive to enhance cybersecurity within its borders by, among other means, the exchange of strategic information.

AI Strategy/Framework: The European AI Alliance has developed a broad approach to AI and an agreement to encourage cooperation among European countries. The EU contributes to the D10 initiative and is a member of the Global Partnership on AI.

GERMANY

Germany's national AI strategy, which is focused on social and economic good, says little about military applications, but does indicate that the federal government will promote research on cybersecurity. In 2018, Germany adopted key points for a national strategy on AI, which indicate that AI is to be developed responsibly and for the good of society. In collaboration with Ghana, Rwanda, South Africa, Uganda, and India, Germany participates in the German Development Cooperation initiative FAIR Forward to promote more open, inclusive, and sustainable approaches to AI on the global level.

Germany is involved in numerous EU initiatives to leverage AI in defence, particularly in uncrewed systems. With France, Germany has formed the Future Combat Air System Expert Commission on Responsible Use of Technologies, which aims to determine ethical guidelines based on international law for this developing system.

Germany favours rules for autonomous weapons systems. As its foreign minister said in 2019, "Killer robots that lord over life and death on the basis of anonymous datasets and entirely beyond human control are already a frighteningly real prospect today. This constitutes nothing less than an attack on humanity itself." However, as one of the authors of the "food for thought" paper on autonomous weapons, Germany does not support a categorical ban.

AI Strategy/Framework: A new cybersecurity agency was formed in 2020. Germany belongs to a number of international AI initiatives, including the Council of Europe's Ad hoc Committee on Artificial Intelligence, the D10 initiative, and the Global Partnership on AI. Germany is a partner in the Trilateral French-Japanese-German Research Projects on Artificial Intelligence. As a member of the G7, it is committed to the Charlevoix Common Vision for the Future of Artificial Intelligence.

SPAIN

Spain's national AI strategy is human-centric; Spain wants to ensure that AI is used to promote inclusivity and sustainability. The strategy is based on six pillars, the last of which is "the establishment of an ethical and regulatory framework that guarantees the protection of individual and collective rights, with social welfare and sustainability as structuring cornerstones." However, it is not clear that this framework extends to the defence sector. Spain has established an AI Advisory Council to conduct a broad assessment of the implications of AI, but, again, it is not clear that its mandate includes defence.

Spain is already a key member in some EU AI defence projects. As well, its Ministry of Defense is planning to integrate AI into its army as part of Fuerza 35 – the modernization of its armed forces to make them ready to engage in multi-domain operations. This plan includes nanotechnology, robotics, and neural networks, which all have practical relationships with AI.

Primary applications of AI relate to command, intelligence, logistical support, protection, and manoeuvring for combat functions. The Ministry has indicated that mandatory tests will be conducted to gauge the use of these technologies in both peace and conflict settings.

AI Strategy/Framework: Spain has a national AI strategy, but it is not clear how it relates to AI defence applications. In December 2020, Spain joined the Global Partnership on AI and is also a member of the Council of Europe's Ad hoc Committee on Artificial Intelligence.

KEY POINTS OF CONVERGENCE

The national and regional approaches outlined above are a good start. As one expert notes, focusing on joint interests, such as "safety, reliability, verification and validation is a great first step to build some mutual trust." Even national policies that do not explicitly mention

AI for defence purposes – including those of Australia, Denmark, Estonia, Finland, and Spain – might have an indirect effect on military uses of AI.

However, analysis indicates that further confidence-building measures and more specific commitments are needed to achieve a robust international regulatory framework. The private sector has also been active in promoting norms, producing a variety of commitments to the ethical and responsible use of AI. A scan of 84 documents highlights principles that recur and which could become the basis for much needed global cooperation on applications of AI, particularly for national defence.

Specific Defence Principles

RELIABILITY

Users of military technology require reliability. Commanders will not adopt unpredictable technologies that could harm the operator or have unexpected or extreme consequences.

The US Department of Defense lists five DoD AI Ethical Principles. According to the DoD, “reliable” means the following: “The Department’s AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across AI capabilities’ entire life-cycle.”

ACCOUNTABILITY

CCW discussions on autonomous weapons aim to establish accountability for actions or results. While AI allows weapons systems to operate with greater autonomy, humans will still need to be held accountable for the results of the use of those systems. While countries broadly agree that humans need to be in charge and therefore responsible for AI systems, more specific norms on the level and type of human engagement will be necessary. Some countries have claimed that existing international humanitarian law (IHL) applies and is sufficient. However, others disagree, and so there is a need to address concerns about who would be held accountable for systems that function with a great deal of autonomy.

RESPECTING RIGHTS

Various statements on the responsible use of AI for defence require AI systems to comply with IHL (the law of war) and international human rights law (IHRL). At a minimum, such AI systems must be unbiased and not cause superfluous harm.

Broader Commitments to Responsible AI

Many states have committed to the responsible employment of AI for security and defence. Ideally, the developers and users of these AI applications would all buy into the same values and norms on what constitute responsible actions.

The principles discussed below have already been identified by Anna Jobin et al. (2019) as the points of convergence in the guidelines on ethical AI more broadly. These key principles would also be relevant to responsible uses of AI for defence. In many cases, these principles are widely shared.

TRANSPARENCY

Appropriate definitions of transparency capture the need for explainability and openness about the ways in which AI systems function and how they are making decisions.

Applying transparency means developing and deploying AI technology in such a way that relevant personnel have an understanding of the ways in which the technology operates. In a military context, this is essential. Transparency allows actors to properly communicate and explain outcomes of an AI system and, importantly, to pinpoint where things have gone wrong if the outcome is unexpected.

Transparency measures help to overcome the ‘black box’ nature of some systems, in which the decisions made by the systems are not understood by humans. They also make the use of AI in national defence and security more predictable, reliable, and secure. More broadly, a lack of transparency around a state’s intentions with AI could lead to an arms race, with actors incentivized to arm to protect themselves from the unknown capabilities of their opponents. Transparency would also be critical in the case of accidents that result from the use of AI systems; if countries or other actors can demonstrate that an action was not intentional, they can perhaps prevent escalatory responses.

JUSTICE AND FAIRNESS

AI developers who deliberately and consciously design security systems with a full awareness of how design can inadvertently exclude or harm groups are better able to achieve inclusivity in their products. Users are likelier to trust information that comes from systems that are seen to be unbiased and objective. The need to address bias in AI systems—from data collection to data analysis to the deployment of particular technologies—is being seen as critical as we learn more about how gender and racial bias are embedded in AI systems. In the case of military applications of AI, there is a need for developers to adhere to norms of justice and fairness that ensure that principles of IHL are upheld.

NON-MALEFICENCE

The concept of non-maleficence refers to a commitment to do no harm or, if harm must be done, then no more than is necessary to achieve a beneficial outcome. In the case of defence and military applications, AI must be deployed so that violence is minimized and innocent civilians are not harmed. In some applications, such as those found in autonomous weapon systems, this could involve a prohibition on targeting people.

PRIVACY

AI has the capacity to invade an individual's privacy, exploit their personal data, and reinforce discriminatory practices. Intelligence and security services that use AI-enhanced capabilities must make a serious commitment to the preservation of the privacy of innocent civilians.

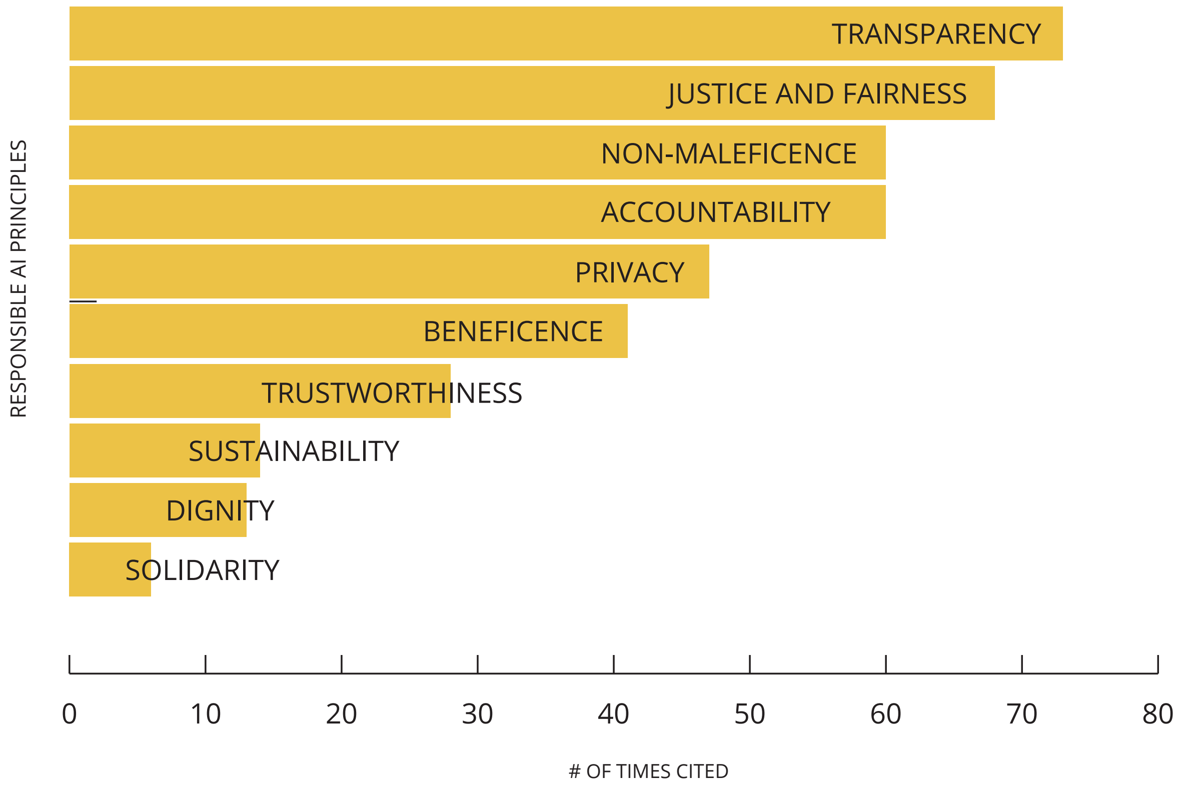

Figure 2. Most frequently cited responsible AI principles

Defence Principles

The principles identified below are drawn from state documents that explicitly discuss principles/values for AI/emerging technologies in defence.

PRINCIPLE	TERMS INCLUDED	STATE
Trustworthiness	Trustworthy	United States, France
Equity	Equitable, fair	United States, United Kingdom
Reliability	Evidence-based, resilient, secure, efficient, effective	United States, European Union, France
Subject to governance	Governable, controllable, sovereignty, interoperability	United States, France
Promotion/maintenance of freedom	Freedom	France
Accountability	Accountability, checks and balances, oversight, explainability, traceable	Canada, United Kingdom, United States
Empowerment	Empowering	United Kingdom
Privacy	Privacy	United Kingdom, United States
Respect for rights	Human rights, international law (necessity, proportionality), civil rights, civil liberties	European Union, Canada, United States,United Kingdom, France

Non-Defence Principles (State/Non-State Actors)

The principles below are drawn from some 84 AI ethics documents from around the world from private companies, the public sector, and research institutions.

PRINCIPLE	NUMBER OF CITATIONS	INCLUDED TERMS
Transparency	73/84	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing
Justice & Fairness	68/84	Justice, fairness, consistency, inclusion, equality, equity, (non-)bias, (non-)discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Non-maleficence	60/84	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion
Accountability	60/84	Accountability, liability, acting with integrity
Privacy	47/84	Privacy, personal or private information
Beneficence	41/84	Benefits, beneficence, well-being, peace, social good, common good
Trustworthiness	28/84	Trust
Sustainability	14/84	Sustainability, environment (nature), energy, resources (energy)
Dignity	13/84	Dignity
Solidarity	6/84	Solidarity, social security, cohesion

GETTING TO GLOBAL NORMS

International stability in the AI era requires clear commitments by all significant actors to a robust regulatory framework that encompasses all uses of AI, not only those related to defence and the military. However, there is a clear lag in terms of the policy developments on military applications of AI at the national and international levels. One step toward that framework is the promotion of norms.

But which states will take leading roles in developing norms and then possibly binding international standards? The AIPfD is at the forefront of the discussions on such standards and will likely shape the type of norms that emerge at a broader level. However, experts suggest that the actors that develop normative standards on how AI is used for defence and national security could be those with the political will rather than those with the most capacity or capability. In this event, smaller states could play an important role.

Experts express a measure of confidence in the ability of some smaller countries to adapt to emerging tech challenges; indeed, they emphasize the necessity of not overlooking the capabilities of smaller states when it comes to technologies such as AI, which appear to be more accessible to a wider number of actors.

Ultimately, the norms developed by a few states will need to be adopted by the majority of states. The challenge for the AIPfD will be to garner support beyond the traditional allies engaged in these networks. A necessary first step will be for each member of the partnership to develop clear policies for military uses of AI. At the moment, such policies are not readily available; middle powers and smaller states can play an important role in shaping the normative frameworks that emerge.

While defence AI policies have not been a key focus, private sector efforts at developing particular understandings of what responsible AI means have been particularly active. The role of the private sector in the AI space is an important one to watch from the military AI perspective. Tech companies have come under closer scrutiny for their work with the military and there have been very public efforts to pay greater attention to tech giants, such as Google, which are developing tech tools in the United States. It was Google employees who took a stand and advocated for their company not to continue a contract with the US military that would see them developing AI for use in weapons. Many companies have continued to engage with the military in other ways, through subcontracts, for example, and some have reverted to business as usual once the attention faded away. Still, the role of industry and the broader tech community in creating standards and norms should not be overlooked.

As noted, much of the discussion on responsible AI uses the terms “responsible” and “ethical” interchangeably. While there is significant overlap, there are also differences. What constitutes ethical and responsible use in the context of defence is likely to be debated. But this is the subject of another paper.

It seems clear that there is much food for thought for Canada in our findings. To be sure, further research and ongoing attention are needed to track normative developments. Canada should be prepared to stake a claim in leading the world to a regulated use of AI in all spheres of activity, including national defence and security. To do so, Canada too needs to develop its own national policy on military applications of AI. Such a policy could also help to guide its normative leadership.

ABOUT THE AUTHOR

Branka Marijan is a Senior Researcher with Project Ploughshares.
Her work examines the ethical, legal, and social implications of the development of autonomous weapons systems and the impact of artificial intelligence and robotics on security policies and trends in warfare.

Dr. Marijan holds a PhD from the Balsillie School of International Affairs, with a specialization in conflict and security.

PROJECT PLOUGHSHARES

140 Westmount Road North
Waterloo ON N2L 3G6
Canada
www.ploughshares.ca
519-888-6541
plough@ploughshares.ca

Project Ploughshares is a Canadian peace research institute with a focus on disarmament efforts and international security, specifically related to the arms trade, emerging military and security technologies, nuclear weapons, and outer space.