PLOUGHSHARES
SPOTLIGHT

THE NEED FOR CANADIAN LEADERSHIP IN THE DRIVE TO **REGULATE HIGH-RISK AI**

BY BRANKA MARIJAN AND GRACE WRIGHT

MAY 2021

On March 30, the Government of Canada's Advisory Council on Artificial Intelligence Public Awareness Working Group, with partners CIFAR and Algora Lab (Université de Montréal), began a series of workshops under the banner *Open Dialogue: Artificial Intelligence in Canada*. Designed to gain a better understanding of public perceptions of artificial intelligence (AI), the workshops were open to all interested Canadians, with specific sessions for youth and Indigenous communities. The series concluded on April 30. The working group is proceeding with a report of its findings, which it will submit to François-Philippe Champagne, Minister of Innovation, Science and Industry.

Open Dialogue is the latest effort by the Canadian government to develop AI policy that is informed by public and key stakeholder input and ensures human-centric AI development. In 2017, Canada was the first country to publish a national strategy on AI. The Canadian Treasury Board's Directive on Automated Decision-Making came into effect on April 1, 2019, with compliance required by April 2020. So far, the Algorithmic Impact Assessment Tool, associated with the Directive and described by the government as "a questionnaire designed to help you assess and mitigate the impacts associated with deploying an automated decision system," has only been used once in government, by the Treasury Board itself.

The Ontario government is also consulting with the public on how to develop a provincial framework. Its survey is available online until June 4.

But, as various ministries of the federal government, as well as relevant ministries at the provincial level, seek to develop policy and procedures on the use of AI, they will need clear guidance on the risks associated with different AI applications and how they should be regulated. So far, no Canadian agency has taken the lead in providing the guidance needed to plan for high-risk AI use, particularly in security and defence applications. Indeed, the separation of security and defence from other AI concerns seems to deny the multi-use nature of the technology and the reality that certain risks are associated with the use of AI in multiple realms.

IDENTIFYING AND REGULATING RISK

With this multi-use nature in mind, Canada needs a national body that will classify the risks posed by different AI systems and indicate which systems require further assessment and review. Such a classification could help to guide the design, development, and deployment of AI tools. Certain applications—for national security and the military, for example—could be identified as particularly risky and be subjected to special regulation and monitoring after deployment.

Beyond regulating AI for domestic use, there is a need to consider stronger regulations for the exporting of certain AI software, especially those with multiple—including military and security—uses.

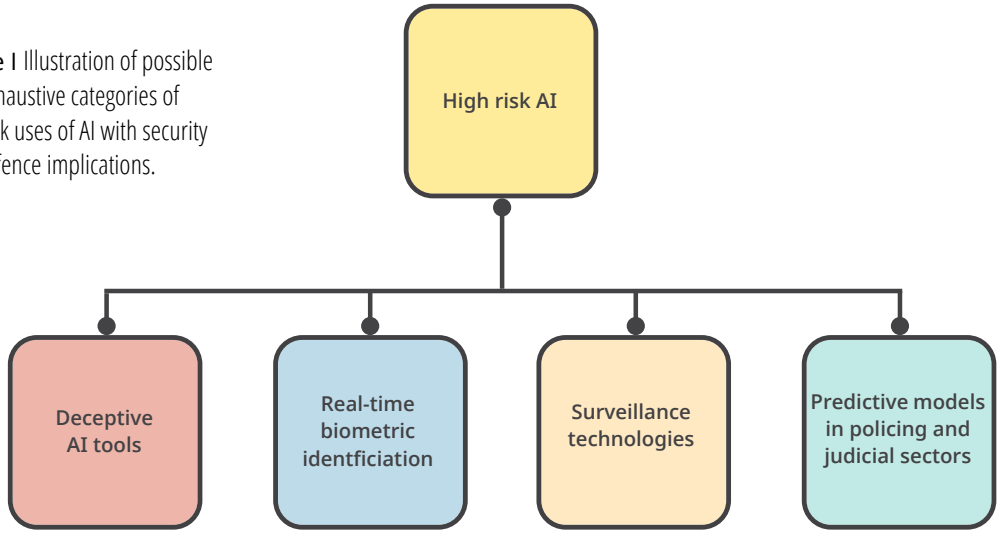
A regulation "laying down harmonised rules on artificial intelligence" that was recently proposed by the European Commission (EC) is instructive, if limited. It categorizes the risks of different AI applications, includes a road map to ensure the transparency and accountability of such applications, and offers concrete examples of an AI system that is deemed to be "an unacceptable risk," such as AI-based social scoring and real-time biometric identification in public spaces by law enforcement. While state use of surveillance tools is allowed under certain circumstances, such high-risk systems would need to be registered prior to deployment. Of particular importance is the proposed prohibition of systems that deceive. Also noteworthy is the scope of the regulation, which extends beyond EU borders to include all data on EU citizens, wherever used.

Interestingly, the United States Federal Trade Commission (FTC) has also made public its planned response to companies that deceive consumers. It cites the example of an app developer that misled individuals about how their photos were being used to train facial recognition algorithms and falsely informed them that their information would be deleted. The FTC proposes an order that would require such developers not only to delete the information, but also the models or algorithms developed from user-uploaded data.

Neither of these promising models, however, regulates AI systems and applications used by the military. National security and defence exemptions are the standard reality across different national AI strategies.

Here Canada has an opportunity to develop a regulatory process that could be a model for other countries. A number of factors and approaches must be considered.

Figure I Illustration of possible non-exhaustive categories of high-risk uses of AI with security and defence implications.



EXPORT CONTROLS

Beyond regulating AI for domestic use, there is a need to consider stronger regulations for the exporting of certain AI software, especially those with multiple—including military and security—uses. At the moment, it appears that software exports are not receiving a great deal of attention. However, technologies that could be used to violate the human rights of populations in other countries should be controlled, perhaps by being placed on export control lists.

But there is a debate about using export controls as tools to regulate AI. Some concern has been expressed that cumbersome export controls would negatively impact AI companies and technological development. On the other hand, risk assessments and particular scrutiny of certain exports can provide clarity and help in the development of a more tailored approach.

Export controls can at times be a dragnet approach that captures a wider number of technologies than intended, as illustrated in the following example. In January 2020, the United States added "Software Specially Designed To Automate the Analysis of Geospatial Imagery" to its export control list, renewing this control for an additional year in January 2021. While the United States is likely most interested in restricting applications that are used in military intelligence gathering, one possible result of such a control is that some software critical for autonomous vehicles could be restricted. Nonetheless, export controls remain an important tool that should be considered, particularly when exported goods are being sent to countries that are experiencing conflict or when there is a reasonable risk that the technology will be misused and will cause harm.

AI BIAS AGAINST MARGINALIZED COMMUNITIES

The risk that marginalized communities, at home and abroad, will be harmed by AI technology is significant. AI systems are shaped—and slanted—by their creators and by the data that these humans supply to train AI systems. When the data contains historical prejudices or software teams have a narrow set of perspectives and experiences, AI systems will likely perpetuate historical and contemporary inequalities. Both predictive algorithms and facial recognition technology, with civil as well as security and defence functions, contain biases that can cause disruption and harm to individuals from these communities.

Bias can be particularly dangerous in policing, security, and defence applications. The effects of AI bias on civilians who interact with police are particularly troubling.

For example, in the United States, the bias present in AI-based predictive policing programs has led to the disproportionate targeting of racialized groups, with individuals from marginalized communities, particularly individuals of colour, wrongly apprehended and denied parole. Also in the United States, "at least one in four law enforcement agencies are able to use facial recognition technology." This technology, which is typically less accurate in identifying members of minority groups and racialized communities, has been used by police to surveil peaceful protests.

Recognizing the risks, several U.S. cities and states have banned the use of facial recognition technologies by law enforcement. A United Kingdom court also ruled that police use of facial recognition technology is unlawful. Predictive policing has also come under scrutiny particularly as minority communities continued to be over policed and added to watch lists often seemingly due to their racial background. Police departments across Canada have invested in predictive tools and as push-back from civil liberties groups mounts these tools should also be examined. The City of Santa Cruz, California is the first city in the US to ban predictive policing. It seems clear that facial recognition technology and predictive models for policing should be seen as high-risk AI applications and be subjected to regulation, possibly including prohibition of use in certain cases.

THE EXPLOITATION OF PERSONAL DATA

Around the world, the privacy of individuals is already under fire. In this increasingly digital age, some companies are harvesting personal data without permission. Canada is not immune.

New York City-based company Clearview AI, for example, has scraped social media websites and collected images of Canadians, without their knowledge or consent, and then used the data to develop facial recognition technology that has been used by, among others, the Royal Canadian Mounted Police and the Ontario Provincial Police. Not only is the facial recognition technology flawed and biased, but it is based on illegally obtained information.

Clearview AI is not alone. State and non-state actors around the world are using invasive data collection methods, including data brokers that collect information from various phone applications and sell databases of that information to eager customers, including government agencies.

Major powers such as the European Union are beginning to take action. Here the scope of the proposed EC regulation, which covers any use of data of EU citizens anywhere in the world, is relevant. Canada should ensure a similar level of data protection for Canadians, particularly when it comes to data sharing within security and defence partnerships and networks.

SECURITY- AND DEFENCE-SPECIFIC CONSIDERATIONS

Governing the sharing of data with security and defence partners: Data will power the AI defence innovations of tomorrow—from drones to decision-assistance programs to complete weapons systems. Allies are even now cooperating in research and development of AI defence applications by sharing data. However, the mechanisms and boundaries for data sharing and use remain unclear.

In the current Canadian security and defence context, there is no unified AI governance framework that would ensure the transparency and guidance required to maintain accountability in AI research, development, and use. Canadian allies, including the European Union and the United Kingdom, have acknowledged the importance of strong data governance, but are still at the proposal stage.

Ensuring Canada supports international efforts to ban autonomous weapons: In 2019, when he was Minister of Foreign Affairs, François-Philippe Champagne was given the mandate to support international efforts to ban lethal autonomous weapons (LAWs). However, Canada has not yet publicly engaged with this issue, although other countries have taken up the cause.

Canada's defence policy *Strong, Secure, Engaged* states, "The Canadian Armed Forces is committed to maintaining appropriate human involvement in the use of military capabilities that can exert lethal force." Canada still needs to define what it considers appropriate human involvement, which directly links with its commitment to support a ban on fully autonomous systems.

ACHIEVING CLARITY

As Professors Fenwick McKelvey and Jonathan Roberge noted recently in *The Globe and Mail*, "Canada's approach [to AI regulation] is ad hoc, with ample room for interpretation and gaming the scattershot rules." Such an approach should worry everyone who uses security and defence applications of AI or who could be affected by their use.

Canada must address a glaring gap in regulation and develop a robust data protection framework that values individual privacy over third-party interests. Without such protection, ordinary Canadians are vulnerable to actors who will use or sell their data for unintended, harmful, and nefarious purposes. In some cases, applications of predictive technologies should be prohibited to ensure that a human decision-maker is not overly relying on biased technologies.

The use of unpredictable and inaccurate systems in matters of life and death could produce truly horrific results. The time for serious talk and action is past due. Canada can't afford to drag its heels any longer. Canada, as a country committed to human-centric AI, should assume a leadership role in developing national and international regulations that will plot a helpful path for the future development of AI. A national body that classifies AI risk and has the regulatory means and capacity to address systems from the design phase to post-deployment is crucial to capturing the benefits of AI technologies for all Canadians.

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.