MEMORANDUM FOR ACTION

TO:	**The Minister of Foreign Affairs**
CC:	**The Digital Inclusion Lab, Office of Human Rights Freedoms and Inclusion**
SUBJECT:	The International Role of Government in Countering Online Extremism

SUMARY:

This memorandum proposes an international course of action to promote the ethical and responsible use of automated decision-making systems (commonly referred to as Artificial Intelligence - AI) in extremist digital content takedown. The increasing use of the internet generally, and social media platforms more specifically, by terrorist organizations warrants an internationally coordinated effort amongst Western liberal democracies in countering online extremism using emerging technologies such as AI.

Strategies devised by states to deal with the issue of online extremism have been nation-centric. The international nature of the internet requires an international approach. Global Affairs Canada is uniquely positioned to play a leading role in the fight against online extremism.

RECOMMENDATION(S):

- **Promote the adoption of an internationally accepted definition of "extremist content."** Canada should use its current G7 Presidency to raise this issue on the international policy agenda. The definition of "extremist content" must reflect Canada's commitment to human rights, democracy, and an open society.

- **Devise an international agreement for the ethical takedown of online extremist content using AI.** It is important to ensure that only content which incites violence is subject to AI-systems' automated flagging and removal. Removing content which is not harmful is an impediment to free speech and must be avoided.

[MOHAMMAD ROSTAMI]
[STUDENT, CONCORDIA - MIGS]

☐ I wish to discuss
☐ I concur   ☐ I do not concur

_____

Minister

BACKGROUND:

1.   Terrorist organizations such as ISIS are using social media platforms (SMPs) and so-called "dark social" channels (i.e. private messaging on SMPs, messenger apps) for propaganda, communication, and recruitment. The 2017 Public Report on the Terrorist Threat to Canada by Public Safety Canada recognizes SMPs as a significant tool for terrorist organizations to reach their strategic objectives.

2.   Beyond its military campaigns in Iraq and Syria, the Global Coalition against Daesh is committed to countering the group's propaganda, communication, and recruitment efforts. This includes preventing the dissemination of extremist content online, the focus of this briefing note.

3.   SMPs (especially Facebook) are facing mounting pressures from government, civil society groups, NGOs, academics, and the public to address the issue of extremist content dispersed through their platforms. There is an increasing use of Artificial Intelligence-powered algorithms that automate the process of extremist content detection.

4.   Mark Zuckerberg's testimony before members of Congress on 10 April 2018 confirmed that AI-powered systems are currently used only to identify unwanted content; human analysts are responsible for reviewing flagged content and deciding on its removal.

CONSIDERATIONS:

5.   The UN Counter-Terrorism Committee Executive Directorate (UN CTED) provides a space for tech companies and states to exchange ideas to facilitate counter-terrorism efforts. Working closely with UN CTED can help develop international approaches and methodologies for the Canadian government to counter violent extremism online. As Canada seeks to get a seat on the Security Council, it should engage and support CTED in its efforts to fight online terrorist activity. An international approach to the problem of online extremism is crucial.

6.   Prime Minister Theresa May has urged G7 members to unite to regulate what tech companies like Google, Facebook and Twitter allow to be posted on their platforms. In late September 2017, French President Emmanuel Macron and Italian Prime Minister Paolo Gentiloni urged SMPs to remove more swiftly terrorist content from the internet. Canada is to hold the G7 Presidency for 2018. This allows Canada to promote a coordinated, responsible, and ethical development and use of AI in countering online extremism. Western states must establish a standard of development and use of AI to discourage potential misuse of these systems by non-democracies such as Russia, Iran, and China.

7.   On 1 January 2018, Germany enacted the Network Enforcement Act, commonly referred to as the "Facebook Law". It requires social media platforms with more than 2 million users in the country to erase posts that violate German hate speech laws. Failure to do so within 24 hours can result in a fine of up to 50 million euros. Germany's approach is undoubtedly informed by its experience with the extreme-right; this may not be suitable for countries which don't share this unique history. Nation-specific policies must be avoided, as the internet is transnational in nature.

8.   The United Kingdom funded ASI Data Science in developing an AI-powered tool that can detect 94% of ISIS propaganda with 99.995% accuracy. These measurements of performance are self-reported. It is claimed that the tool can be used on any platform and can remove content within 2 hours. It is unclear if this is possible; this tool may not be usable on all platforms. UK is considering forcing SMPs in using this system. It is problematic if every state

develops its own tools and forces SMPs to use it. The international nature of the internet requires a coordinated effort amongst states.

9.   There are different definitions of what is considered "extremist content." On 25 October 2017, eleven activists went on trial for terror offenses in Turkey, including two local Amnesty International chiefs. Turkey has detained tens of thousands of activists, journalists, and opposition political figures. Removing content which is not considered harmful, offensive, extremist, or illegal, even if distasteful, is an impediment to free speech and must be avoided. Without a common definition of "extremist content," AI-systems may impede basic human rights. Non-democracies may use such systems to silence their opposition. There is the need for an international definition of "extremist content."

10.   On 10 and 11 April 2018, Mark Zuckerberg stated that Facebook estimates 5 to 10 years before effective AI tools are available to automatically detect hate speech. In 2016, Zuckerberg said AI could be outperforming humans in the next 5 to 10 years. Solutionism, the belief that every problem has a solution based in technology, often has too optimistic of an outlook. Given the well-documented issues of AI bias and decision-making problems, it may not be an effective solution in eliminating extremist content; it may instead introduce a whole new set of unexpected problems. AI systems may remove legitimate content and limit free speech. During the hearings, members of Congress discussed how Facebook is removing the lawful content of their constituents. The definition of "extremist content" must reflect democratic values and norms.

11.   YouTube's AI-powered algorithms have removed thousands of videos which documented the atrocities of the Syrian civil war. This evidence could have potentially been used to prosecute Syrian authorities for their crimes. A coordinated international effort to define "extremist content" will partly address this issue.

12.   The solution is for Western liberal democracies to coordinate their strategy in regard to extremist content takedown by first, defining "extremist content" and second, devising an international agreement for its ethical takedown using AI. Automated systems must respect human rights and target only content which incites violence. Western SMPs are an important tool in promoting human rights, democracy, and an open society. Overregulation may favor Chinese (Weibo, WeChat) and Russian (VKontakte) in increasing their reach and user base.

COMMUNICATIONS IMPLICATIONS/ACTIONS:

13.   The proposed course of action may generate some media coverage, especially since Facebook has come under severe scrutiny lately.

14.   The media and civil society groups may criticize the proposed course of action on the basis of its focus on terrorism. It is important to state that while the extreme-right/left poses important security concerns, one of the biggest challenges to international security remains terrorist organizations.

15.   It may also be criticized on the basis of regulating a space (the internet generally, and SMPs specifically) which has traditionally been free of government interference in Western states. It is important to state that any government interference in online content is solely aimed at promoting human rights, democracy, and an open society.