TITLE Artificial Intelligence, Ethics, and Future Conflicts
Date: Dec. 8, 2018

Disclaimer: this briefing note contains the encapsulation of views presented throughout the evening and does not exclusively represent the views of the speaker or the Canadian Association for Security and Intelligence Studies.

KEY EVENTS

The following key events will be considered in this briefing note:

Key Event 1: The Canadian Intelligence Community met in October 2018 to discuss the following themes:

1. Artificial Intelligence -Operational Context,
2. How is AI revolutionizing the Intelligence and Security Enterprise,
3. How Artificial intelligence will reshape the Canadian Intelligence Enterprise from a Cyber Security and Technological Perspective,
4. Technology Supporting the Intelligence Enterprise,
5. Innovation and Ethics, and
6. Artificial Intelligence and how it will transform the Intelligence Enterprise.

Key Event 2: The continued interest in AI and the role of Ethics to inform AI development and AI as an ethical machine.

Key Event 3: Operationalization of AI within current national and international conflicts (war, terrorism, and crime) with specific reference to 5th Gen warfare[1] and conflicts which involve CAF personnel during deployment with NATO or other allied partners.

NATURE OF DISCUSSION

The nature of this discussion will involve the role of ethical standards in the development of an AI system to be used by the CAF. In particular, it will focus on what ethical standards could be utilized and how such standards could influence AI development. In addition, this note will examine the potential ethical problems pertaining to the use of an AI system against 5th generation threats.

BACKGROUND

[1] The term '5th Gen warfare' is defined as "Identity based group-on-group conflict for the purpose of enacting a change in society." (Canadian Association for Security and Intelligence Studies, 2018). CASIS Vancouver. (2018). Retrieved from https://casisvancouver.ca/1_19_fifth-generation-warfare.html/publications/lexicon/

This note is organized as follows: Section 1 introduces AI; Section 2 explores the potential bias which may occur during programming and inputting data; Section 3 explores the intersection of Ethics and AI, and how these will influence / support the decision making processes for the CAF Intelligence Enterprise.

Artificial intelligence (AI) refers to one or more of four categories "Systems that think like humans", "Systems that act like humans", "Systems that think rationally", and "Systems that act rationally.[2] AI is capable of managing the data of an autonomous system and from this, it "reasons probabilistically given a set of inputs" and "guesses about best possible courses of action given sensor data input"[3]. In other words, AI and autonomous systems in the battlefield can work together.

Section 1 - AI now and in the future

The Canadian government is currently exploring the role of AI within various departments and programs. The DND/CAF adoption of AI may be informed by a larger federal government mandates proposed by Ann Cavoukian, and Pan-Canadian AI strategy.

Cavoukian, the former Information and Privacy Commissioner of Ontario, has developed seven principles of AI. These principles are:[4]

- Transparency and accountability of algorithms essential
- Ethical principles applied to the treatment of personal data
- Algorithmic oversight and responsibility must be assured
- Respect for privacy as a fundamental right
- Data protection/personal control via privacy as the default
- Proactively identify the security risks, thereby minimizing the harms
- Strong documentation to facilitate ethical design and data symmetry

These principles according to Cavoukian could help policy makers both in the government and the private sectors to develop ethical AI standards.[5]

Moreover, Simpson (2018) has reported that "The Canadian government itself is currently working on its Pan-Canadian Artificial Intelligence Strategy, dividing $125 million in funding between three prominent Canadian AI institutes, with one of its goals to develop ideas around the economic, ethical, policy and legal implications of advancements in AI"[6]. Furthermore, according to Simpson (2018) the Chief Information Officers Strategy Council says "it is ready to create standardized policies to help Canadian

---

[2] Russell, S. & Norvig, P. (1995). Artificial Intelligence: A Modern Approach Pg. 5. Retrieved from https://www.cin.ufpe.br/~tfl2/artificial-intelligence-modern-approach.9780131038059.25368.pdf

[3] Cummings, M.L. (2017). Artificial Intelligence and the Future of Warfare. International Security Department and US and the Americas Programme.

[4] Jackson, B. (2018). Holding AI accountable: public leaders organize to ensure algorithms influencing government are ethical (May 17, 2018). Pg. 4. Retrieved from https://www.itworldcanada.com/article/holding-ai-accountable-public-leaders-organize-to-ensure-algorithms-influencing-government-are-ethical/404793

[5] Ibid

[6] Simpson, M. (2018). Ethical AI standards and the Canadian group making it happen. ITWC.ca (Aug 23, 2018). Retrieved from https://www.itworldcanada.com/article/ethical-ai-standards-and-the-canadian-group-making-it-happen/408281

companies and governments". The CIO Strategy suggests there are two topics that the standards may consider: "data access and privacy" and "ethical design and use of automated decision systems". The CIO Strategy Council reportedly "aims to be a leader in developing the first-ever globally recognized standards for the ethical use of AI and big data".[7]

Both models, Cavoukian and the CIO Strategic Council, may help CAF/DND inform their insights into AI and Ethics. The challenge will be to ensure the CAF/DND core CAF values, such as Respecting the "dignity of all persons", integrity and public accountability[8] continue to be respected. Doing so, will ensure CAF members are able to bear the closest public scrutiny in addition to obeying the letter of the law.[9] For example, Cavoukian's principle which calls for a need to have transparency and accountability in the AI systems would provide one reason for CAF/DND to use a transparent AI model (See BN Types of AI and CAF/DND needs) which can be audited. Hence both the CAF codes of conducts and Cavoukian principles on AI ethical standards have the values of integrity, trust, and accountability as their common denominator and therefore should be used to design a CAF/DND AI ethical framework.

Section 2 Effects of Bias on AI Development and Deployment.

AI has the potential to be fast and efficient, but it is still susceptible to the bias' of its coder. The foundational level of an AI may be built in the image of its creator and includes their thoughts and personality. The data input given to an AI for learning is also susceptible to bias as a human decides what and how data is original collected in a computer / information legacy system. It is this legacy data which will be feed into the AI and the AI learns from.[10, 11] The AI takes on the cultures and ethical codes of conducts that are aligned with the locale and ideals of the AI's place of origin. For example an AI created by the Canadian Armed Forces may well be different from one created by ISIS or another terror organization.

A summary of bias is provided in IBM's "Ethics of Artificial Intelligence". The summary includes, for example confirmation bias, blind spots, and stereotyping. These examples are used to demonstrate how intelligence analysis would be biased when encountering a 5th Generation Warfare group Night Wolves [12], Accuracy of AI to detect based on contextual nature of operational environment? Examples of Bias towards 5G groups (ex Night Wolves) according to IBM's "Ethics for Artificial Intelligence"[13]

---

[7] Ibid

[8] Department of National Defence. (2012). Code of Values and Ethics. Ottawa, ON.

[9] Ibid

[10] Osobo, O. (2018). Keeping artificial intelligence accountable to humans. TechCrunch (Aug 20, 2018). Retrieved from https://techcrunch.com/2018/08/20/keeping-artificial-intelligence-accountable-to-humans/

[11] Jones, D. (2017). AI Uninterrupted: Culture, Representation and Ethics of AI. iPass (October 13th, 2018). Retrieved from https://www.ipass.com/blog/ai-uninterrupted-culture-representation-and-the-ethics-of-ai/

[12] Night Wolves is a motorcycle club which is a identity based Russian ethnic background. The group is identity based and will use violence as necessary to assert their Russian heritage. This group is currently operating in Russia, annexed regions by Russia, and has been involved with Russian expansion. (Higgins, A. Mar 31, 2018). Russia's feared 'Night Wolves' bike gang came to Bosnia. Bosnia giggled.' New York Times online, https://www.nytimes.com/2018/03/31/world/europe/balkans-russia-night-wolves-republika-srpska-bosnia.html

[13] Cutler, A., Pribic, M., & Humphrey, L. (2018). Everyday ethics for artificial intelligence. Retrieved from https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf Pg. 14

There are several potential biases that may influence ethical decision-making in an AI. For example, *confirmation bias* refers to the "tendency to search for, interpret, or focus on information in a way that confirms one's preconceptions." The AI may only seek information that aligns with its preconceived beliefs, consequently biasing the information that is collected and analyzed. With a confirmation bias in its decision-making, the AI may only search for certain features or personality types when examining groups like the Night Wolves or its individual members.

Another example of bias; *blind spot bias*, "The tendency to see oneself as less biased than others, or to be able to identify more cognitive biases in others than in oneself", is also problematic. Intelligence Analysts may view an AI's findings as more objective than a human's due to the sheer fact that the analyst may not be familiar with AI constructs. On the contrary, due to the lack of knowledge, analysts may be skeptical of the findings in AI generated intelligence products and therefore ignore AI reports.

Given that the algorithm is designed by humans, the issue of *stereotyping*, "Expecting a member of a group to have certain characteristics without having actual information about that individual" also arises. As the programmer's biases pertaining to certain groups may not be addressed, their bias may be unintentionally programmed in the AI, resulting in an algorithm that is biased in data processing and decision-making.

Given the previous examples of bias being introduced into the AI model, there is a need to create an CAF/ DND ethics framework with oversight to examine AI development, operationalization, and auditing of outputs. This oversight would / may have detected the flaw which occurred with Amazon where Amazon's experimental hiring tool with an artificial intelligence framework displayed a favourable bias towards male candidates. The data set used resulted in an inherent favorability of male candidates, as such the system sought out male candidates (confirmation bias) and penalized resumes that indicated that the applicant is a woman[14].

Arguably, when an AI system is employed in a military context, the problem of bias/ flawed algorithms creating inaccurate assessments will still be present. In particular, these programmed biases may reduce the operational effectiveness of AI in the battlefield and undermine the confidence of the decision maker having to rely upon AI outputs. This in turn leads to a potential problem where a commander acts on a flawed assessment.

Section 3 - Ethics and AI to support decision making

The deployment of artificial intelligence in a CAF/DND environment will have to consider two types of conflict; the 5th generation warfare aspect and the traditional warfare model. 5th generation warfare (identity warfare) is characterized by fluid evolving groups who are in conflict without state support. Whereas, traditional warfare is characterized as fully supported armed forces with a specific tasking. The key concept to consider is the ethical postmodernist consideration; First and foremost the claim - No Authority beyond the Self, which means that ethics is the logical outgrowth of a prior commitment to

[14] Dastin, J. (Oct, 9 2018) Amazon scraps secret AI recruiting tool that showed bias against women. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

a particular theology.[15] A similar thought is noted by Dickson (2006) the "postmodern war has no pre-set rules or recognized code of conduct"[16].

Secondly, decision making occurs within a specific context. Consistent with the above notion of an evolution in warfare from traditional armies to engaging fluid identity forming groups, CAF / DND decision making will need to evolve. AI can help that evolution by capturing the local community's moral and philosophical truths which inform decision making. As the CAF / DND AI encounters the 5th generation group conflict data set, new outputs may challenge traditional command and control decision making.

KEY POINTS OF DISCUSSION AND WEST COAST PERSPECTIVES

1. The problem of bias is linked to humans, previous data sets, and policy. How do you train human intelligence analysts to critique or test AI outputs?

2. AI and Ethics is a continually evolving field. The use of AI, although appealing, must be approached with caution. Is it possible to develop testing platforms or environments to run various war gaming simulations to test the reliability and trustworthiness of the AI engine?

3. Machine learning is not AI. This must be very clearly differentiated in all debates. How does CAF / DND ensure this differentiation is maintained? How does CAF / DND dictate this criteria to bidders, developers, and contractors?

4. National security threats will evolve whereby internal and external threats of the 5th generation warfare groups will be encountered by local and national authorities. How will the Government of Canada ensure there is a coherent AI intelligence strategy? .

[15] Richard Rorty, *Achieving Our Country: Leftist Thought in Twentieth-Century America* (Cambridge, MA: Harvard University Press, 1998), 16.
[16] Dickson, D. K. (2006) War in (Another) New Context: Postmodernism. The Journal of Conflict Studies.