NOTE FOR NATIONAL DEFENCE:
Public and Defense Policy Challenges and Innovations on Artificial Intelligence, Autonomous Systems, and Cybersecurity, Part 7: Ethical AI in Defence

Authors: Neshat Elhami Fard [1], Rastko R. Selmic [2], Khashayar Khorasani [3]
[1] Graduate student, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada
[2] Professor, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada; rastko.selmic@concordia.ca
[3] Professor, Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada; kash@ece.concordia.ca

Summary

- Artificial intelligence (AI) is a rapidly developing field. Various studies have shown that it holds great potential for enhancing military capabilities and decreasing operational risks [1].
- For defense, a delay in adopting emerging technologies can lead to a military disadvantage, while immature adoptions without enough investigation and analysis may trigger unintended adverse effects [1].
- It appears that significant work has to be done in order to avoid the adverse outcomes and the negative consequences of the introduction of AI technology in military applications [1].

CONTEXT

Background: The Royal Australian Air Force's (RAAF's) Plan Jericho1 found out that specialists from various fields (Plan Jericho, Defense Science and Technology Group (DSTG), and the Trusted Autonomous Defense Cooperative Research Centre (TASDCRC)) have to come together to consider and solve the lack of understanding of the ethical issues related to emerging technology. Moreover, RAAF believed that the specialists have to gather and define a robust and relevant framework to guide the development and operation of AI systems in the defense sector [1], [2]. The ethical issues are defined based on the following topics:

- **War-fighting and Combat Functions:**

The war-fighting functions are force application (FA), force protection (FP), force sustainment (FS), and situational understanding (SU).

- **Non-war-fighting Functions:**

The non-war-fighting functions are including personnel (PR), enterprise logistics (EL), business process improvement (BP), and others (OR).

Policy, Rules, and Ethical Principles: Ethical principles for different aspects of AI in defense should meet and answer the following issues [1], [2]:

- **Responsibility:** Determining which person or organization is responsible for AI's performance and operations [1], [2]!
- **Education:** Just as military officers are taught various viewpoints of human behavior, cognition, and social factors during leadership and management training, to be a leader or manage different aspects of AI in the defense industry, one must know all the different perspectives of AI. Without sufficient knowledge and awareness about AI, it cannot be managed or leadership. Therefore, early AI education for military and other national security personnel is essential [1].
- **Command:** Multiple decision-makers, e.g., commanders, designers, acquisition agencies, and operators, are involved in making critical decisions in such a way that all of them have the opportunity to exert authority and make mistakes. To attain accountability for military decisions, these significant decisions with aid from or produced by AI must be taken by international and domestic law as legal reference frameworks [1].

The International Committee of the Red Cross: The AI decisions should be made in a human-centered manner to guarantee that humans are finally engaged and responsible [1].

- **Governance:** Determining how AI systems are controlled [1], [2].
- **Effectiveness:** Experimentation, simulation, limited live trials, etc., are all excellent ways to display the efficacy of AI systems before they are used. It will be necessary to conduct robust testing to evaluate the AI's decision-making. As different scenarios are presented, it will be possible to determine whether a system can cope with diverse levels of risk, dynamics, and decision requirements.
- **Integration:** A robust and diverse decision-making process would be improved through system integration.
- **Transparency:** An operator is conscious and aware of the autonomous agent's actions, decisions, intentions, and behaviors because of transparency. Research shows that trust in autonomous systems could be enhanced by transparency. In terms of situation awareness and workload reduction, it seems that a certain level of transparency enhances operator performance. However, excessive transparency can negatively affect performance.
- **Human Factors:** It is better to develop AI systems by considering cognitive psychology and neurophysiology into account. In order to prevent poor, wrong, and helpless decision-making, such as automation bias and mistrust of the system, human-machine collaboration and coordination need to be optimized.
- **Scope:** The caution is given concerning over- or under-reliance on AI. In order to build enough trust in the system and use it, it is also important for the human to construct an accurate mental model of the system's capabilities.
- **Confidence:** It is essential for confidence levels to be based on knowledge of when it is appropriate to use (or trust) the AI and when humans must intervene. Basically, humans and AI require vast amounts of information, and, finally, only real-world experience can determine the information required to enable efficient and adequate interaction.
- **Resilience:** It is essential to build a resilient system capable of anticipating, containing, and recovering from anomalous events and situations.
- **Trust:** Determining the reliability of AI and how it can be trusted [1], [2]!

  Users, operators, commanders, support staff, military, government, and civilian population of a nation must trust the human-AI systems. AI must be lawful, ethical, and robust in order to be considered trustworthy.
- **Sovereign Capability:** In March 2017, the National Security Science and Technology

Interdepartmental Committee was created in Australia as part of measures to address sovereign ability requirements in terms of AI (Defense, 2018). Cybersecurity, intelligence, border security and identity management, investigative support and forensic science, preparedness, prevention, intervention, and technology foresight are national security science and technology priorities endorsed by the committee. Using suitable AI applications can improve the mentioned issues.

- **Safety:** Safety should be a priority for AI systems. Various experiments, simulations, limited live trials, etc., can be conducted to show the safety state. It is required to test AI systems in various scenarios to evaluate their ability to perform in diverse environments with different dynamics, decision requirements, and risks. For AI to be safe, it needs to avoid adverse side effects while seeking its objects, prevent 'reward' hacking, have scalable oversight so actions can be checked and reviewed, be able to investigate the safety and be able to handle situations different from those it was trained to handle.
- **Supply Chain:** AI developed by an insecure supply chain can be unsafe and be vulnerable to hacking or comprise backdoors. Not just the end product for military applications, but all perspectives and aspects of generating AI for the military should be examined for its ethical status (from initial stages to the final product).
- **Test and Evaluation:** The experts and specialists emphasized the need to test and evaluate AI under significant adversarial situations. It may be necessary to impose strict test and evaluation criteria for AI systems. AI experts strongly recommended iterative testing during the AI development process.
- **Misuse and Risks:** According to the predicted level of autonomy and planned application contexts, AI risks and misuses may be various and more extensive than others. The development of AI should consider "cyber hygiene" processes and systems while considering potential concerns of "cyber interference".
- **Authority Pathway:** Specialized researchers examined how AI can help and support tactical decision-makers to make significant ethical decisions. AI is utilized to:
  - Assist decision-makers before and at the trigger point to make more ethical and accurate judgments and assessments;
  - Build ethical consciousness, habits, reasoning, and actions in combination with interactive interfaces;
  - Help medical decision-makers;
  - Assist a human-on-the-loop or in-the-loop in determining whether objects are combatants or non-combatants, potentially decrease the risk of civilian death by taking all available information into account and processing it efficiently.
- **Data Subjects:** In some national defense uses of AI, information must be secure and not available to individuals. However, Defense personnel and their data must be treated ethically. Consequently, AI-based systems should be reviewed concerning the impact on personnel. Additionally, scientists have expressed concern over the dangers of using data analysis AI tools, such as big data.
- **Law:** Determining how AI can be used legally and lawfully [1], [2]!
  - A human-centered approach is applied to the legal framework for Defense activities. Accordingly, conforming AI to these standards will yield more ethical outcomes. A superior humanitarian consequence would be possible if AI reinforces human decision-making.
- **Protected Symbols and Surrender:** It might be possible to identify surrender and protected symbols (e.g., red crescent sign or red cross sign) utilizing defense AI. In this way, it was hoped that AI might decrease the number of operational accidents caused by human error that have significant negative humanitarian, social, and political ramifications. However, some parties use protected symbols to profit, take advantage, and deceive AI. Therefore, AI should be defensive in such a way as to anticipate and detect deception and misinformation.
- **De-escalation:** The use of autonomous systems and AI may aid in the de-escalation of conflict. AI could provide commanders with better information, data, knowledge, and understanding of the conflict, enabling them to manage and lead escalation and de-escalation.

- **Traceability:** Determining how AI functions are recorded and their actions traceable [1], [2].
  - The manner of records is significant because it can identify the involved and decision-maker systems that include a chain of events, humans, and AI systems.
- **Explainability:** AI systems require providing explanations for their information. In terms of ethical principles for AI in defense, this means that human oversight, understanding, and explainability should be considered.
- **Accountability:** According to legislative obligations, human-AI logistics systems should provide explanations of their decisions. A straightforward decision-making process increases the ability to educate the system, produce feedback on consequences, and create trust between human operators and the AI system. AI systems have to provide their users with evidence that is understandable, technically transparent to experts, and comprehensible to consumers and citizens so that they can meaningfully satisfaction or challenge their use.

CONSIDERATIONS

Recommendations: To express, implement, and develop the aforementioned policies and principles, the following recommendations are considered by the RAAF [1]:

- **Ethical AI for Defense Checklist:** A checklist is required to define ethical AI for defense. This checklist contains:
  - ❖ Give a description of the military context in which AI technology will be used.
  - ❖ Give an explanation of what types of decisions the AI supports.
  - ❖ Provide an explanation of how the AI is combined into human operators to assure effective human-machine communication and ethical decision-making in the context of countermeasures to defend against potential abuse.
  - ❖ Provide a brief description of the framework or frameworks that will be used.
  - ❖ Assistance and help from relevant specialists. Experts in the relevant fields can guide the development of AI.
  - ❖ Reducing risk by using suitable verification and validation methods.
- **Ethical AI Risk Matrix:** For each project activity, an ethical AI risk matrix is required to define ethical AI for defense. This risk matrix contains:
  - ❖ Decide what kind of activity will be used and describe it.
  - ❖ Provide a detailed description of the ethical aspect of the activity.
  - ❖ Determine the risks to project objectives if the issue is not resolved.
  - ❖ Determine the specific actions that will support the activity.
  - ❖ Set a timetable for the activities.
  - ❖ Describe the results of an action or activity.
  - ❖ Determine which party or parties is (are) responsible.
  - ❖ Give a status report on the activity.
- **Legal and Ethical Assurance Program Plan (LEAPP):** A general legal and ethical plan should be given for AI programs with ethical risk estimation above a specific threshold. Under a contractor's LEAPP, the contractor ensures that the software acquired under the contract meets the Commonwealth's legal and ethical assurance (LEA) provisions. The LEAPP:
  - ❖ Allows Defense to see how its contractors are doing regarding ethical and legal matters.
  - ❖ Monitors progress and assess risks.
  - ❖ Contributes information and data into Defense’s internal planning, containing weapons survey.

- ❖ Defense and industry stakeholders will have the opportunity to review the draft Data Item Description (DID) and provide their feedback and comments before it is considered for Defense contracts.

NEXT STEPS

More policy implications and decision-making recommendations, observations, and guidelines by the defense policymakers and defense decision-makers will be examined in the subsequent report. In addition, the mentioned cases will be studied and researched in the field of superpower countries.

References

[1] Board, Devitt, Kate, Michael Gan, Jason Scholz, and Robert Bolia. "A method for ethical AI in Defence."

[2] The Australian Government. (1903). *The Defence Act*. (C1903A00020 No. 20). Retrieved from https://www.legislation.gov.au/Series/C1903A00020.