BRIEFING NOTES

#BN-38-The role of AI-Feb2021

WHY SPECIALIZED MILITARY AI INSPECTORS ARE NEEDED?

Authors: Reza Bahrevar[1] and Kash Khorasani [2]
[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

SUMMARY

- One of the major problems in AI systems is the lack of specialized inspectors that can certify the safety and transparency of the developed AI systems. Currently, many academic AI institutes such as Mila are offering courses that generally familiarize students with problems such as bias and interpretability of AI systems.
- However, the main concern is that in such academic institutes there are very few courses and training opportunities that are specifically dedicated to recognizing the ethical or technical biases in the AI solutions, and even fewer that are specifically addressing the crucial application-centric challenges.
- Since AI applications are embedded into almost every domain of our lives, one needs to have specialized investigators that are capable of certificating the safety and reliability and trustworthiness of an AI system with the view that involves and incorporates a combination of data mining and expert knowledge on the potential ethical/technical biases.

CONTEXT

- Digital technologies such as AI, IoT, etc. have been increasingly employed in our every day's life as well as in industry. These devices and services rely on Internet to operate. On the other hand, inherited vulnerabilities associated with digital technologies and the growing number of cybercriminals as well as state sponsored hackers pose continuous threats to security of society and industry.
- In the last few years, the training of multidisciplinary experts in the AI Institutes is one of the strategies that Quebec and the AI community are promoting [1].
- In institutes such as Mila, a student will receive an education that helps him/her to know there can be biases in AI systems related to a data mining perspective or general concepts related to the idea of ethical problems in AI systems. Basically, one is trained on why an AI system should be interpretable, but not specifically what are the biases in different AI platforms when combined with the Internet-Of-Things (IoT) systems.
- The training these students/experts are to receive does not directly respond and address an answer to the bias with considerations to a specific application. Therefore, the view point is quite broad as opposed to being focused, whereas the industry do actually need a proper combination of both knowledge and training.
- The output produced by these Institutes will be students that can design an AI system for a specific AI platform but not necessarily become knowledgeable on how to inspect its biases since they are biased to design by considering potential biases and are not trained for the sole purpose of criticism.

- In this regard, Gartner discusses a critical factor for accomplishing a reliable AI system: "Promote people skills. Fill or hire people in key AI roles related to AI ethics, governance, and policy. Look for privacy/brand remediation and AI behavior forensic specialists who can explain models and perform investigations when AI fails to reduce risk."

CONSIDERATIONS

- Quebec AI established a steering committee through the University of Montreal in order to promote mathematical literacy and responsible AI among AI students and deliver specialized AI trainees to the industry [1]. However, a responsible AI has received less attention since the concept may seem more dubious.
- HumanIA in UQAM is a research-based laboratory that is focused on ethical and legal issues related to AI systems. While considering multidisciplinary issues in their defined objectives, not enough attention has been given to determining biases with regards to the domain-centric AI systems.
- Considering quality assurance of AI systems by AI developers is already being practiced in AI laboratories and institutes such as MIT-IBM Watson AI lab, Mila institute, Gerad institute, and HumanIA. However, the main focus in these institutes is on training responsible developers, and not much on the AI ethical/technical inspectors.

NEXT STEPS

- Military training AI inspectors will make one ready for the upcoming changes in the AI technology. It provides preparation for stable governance of these systems, and accelerates the process of integrating AI systems into other aspects of emerging technologies.
- A military trained inspector must have strong mathematical literacy to be able to recognize deficiencies and common biases specific to an AI algorithm. These military trainees must have a multi-disciplinary exposure that is specialized in a few essential ground models that recognize different types of biases of AI models. They should be able to perform adversarial tests, robustness tests, and security tests through purely mathematical and data mining knowledge and skill sets [2,4].
- A military trained person should be able to introduce adversarial inputs that can result in misclassification by AI systems [2]. They should be able to introduce these inputs through the available algorithms through purely data-mining knowledge, as well as experience-based inputs generated by the pre-defined simulators, or their knowledge about the specific application under consideration.
- The military trained person should be able to recognize what are the elements that are missing from the input/output of an AI system that can cause a legal, discriminatorily, or ethical gaps [3].

- The military should also be trained specifically with regards to particular applications in order to be able to challenge/question the types of defined inputs/outputs, and demand adding the neglected inputs/outputs by the developer.
- A military trained inspector should be aware of the IoT connected privacy, transparency, and security problems concerning the specific application of AI systems. Since every AI system can include various and multiple realm of science and engineering such as biology, computer science, and medicine, it is suggested that the military inspectors can be more valuable if they are selected to be domain-specific.

REFERENCES

[1] Strategy for the Development of Quebec's Artificial Intelligence Ecosystem. In *2018 A mandate from l'Économie, Science et Innovation Quebec.*

[2] Marino, D.L., Wickramasinghe, C.S. and Manic, M., 2018, October. An adversarial approach for explainable ai in intrusion detection systems. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society* (pp. 3237-3243). IEEE.

[3] C arter, S.M., Rogers, W., Win, K.T., Frazer, H., Richards, B. and Houssami, N., 2020. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. *The Breast*, *49*, pp.25-32.

[4] Gehr, T., Mirman, M., Drachsler-Cohen, D., Tsankov, P., Chaudhuri, S. and Vechev, M., 2018, May. Ai2: Safety and robustness certification of neural networks with abstract interpretation. In *2018 IEEE Symposium on Security and Privacy (SP)* (pp. 3-18). IEEE.