Catalyst Fellowship Program - Final Report

September 10, 2023

Reza Samavi Ph.D., P.Eng.
Associate Professor
Department of Electrical, Computer, & Biomedical Engineering
Faculty of Engineering and Architectural Science
Toronto Metropolitan University

Secondary Appointment: Faculty Affiliate; Vector Institute, Toronto, Canada
Tertiary Appointment: Adjunct Professor; McMaster University
address: ENG 457, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada
email: samavi@torontomu.ca
web: `https://www.ee.torontomu.ca/~samavi`

Abstract

This is the final report on the Catalyst Research Fellowship program. The objective of this report is to provide an overview of my one-year appointment as a research fellow working along with a team of two other research fellows, three industry fellows and the Rogers Cybersecure academic director. The report emphasizes how the Catalyst fellowship program helped me extend my scientific collaboration network and briefly describes my achievements during my tenure. The first part focuses on individual achievements directly impacted by the program and the second part reports the collective achievements and my role as a research fellow. I conclude the report by providing a number of recommendations for improving the next cohort of the fellowship.

1 Individual Accomplishments

1.1 Project

The title of the proposed project was "Quantifying Machine Learning Model Trustworthiness." Machine learning (ML) algorithms are at the core of modern AI. The focus of ML researchers in the past decade has been mainly on improving the performance of ML by developing novel algorithms to outperform the prior algorithms and complete even more

complex tasks. The enthusiasm in the ML research community combined with the thirst of the ML consumers (e.g., self-driving car manufacturers) pushed the ML community to speed up the deployment phase of the developed algorithms. This fast-paced process of development and deployment resulted in a phenomenon called by Google researchers as *technical debt* for ML systems [5]. The concept of technical debt was originally coined by Ward Cunningham in 1992 expressing concerns about rapid software development cycles. The concept is analogous to financial debt and helps reason about the long-term cost incurred by moving fast from a prototypical algorithm to a live version of the algorithm in the wild[1]. While the authors in [5] define technical debt in terms of the cost associated with the maintainability of a deployed learning system, we argue that this concept of debt is also applicable to the risk associated with the trustworthiness of the deployed learning systems. The trustworthiness of a system is considered the degree of alignment between the trustor's (e.g., users of an ML system) expectations and the observed behaviour of a trustee's system (e.g., ML algorithms). An ML system, which in many cases deployed as a blackbox, incurs liability if the system doesn't meet the expected qualities, including users' expectation of accuracy, safety, security, privacy or fairness of the predicted outcomes.

The objective of this research project was to attract security and machine learning research communities and AI industry practitioners to discuss and explore the requirements for developing a common trust management framework for AI systems. For this project, I focused on security and trust in the medical AI setting. Since an AI system is considered a complex system with many components, I used this catalyst program opportunity to 1) identify determinants of trust in the ML pipeline and understand the roles and relationships of different determinants of trust for each component of the ML pipeline and AI as a whole, and 2) understand the interaction between human and ML models, the dynamics in which human and ML algorithms are parts of a joint cognitive system.

Throughout this Catalyst fellowship, in addition to the following publications related to this topic,

[C1] Yuting Liang and Reza Samavi. Advanced defensive distillation with ensemble voting and noisy logits. *Applied Intelligence*, 53(3):3069–3094, 2023. IF: 4.602.

[C2] Mini Thomas, Omar Boursalie, Reza Samavi, and Thomas E Doyle. Bayesian-based parameter estimation to quantify trust in medical devices. In *International Workshop on Health Intelligence @AAAI23 conference*, pages 95–108. Springer, 2023.

[C3] Thomas E Doyle, Victoria Tucci, Calvin Zhu, Yifei Zhang, Basem Yassa, Sajjad Rashidiani, Md Asif Khan, Reza Samavi, Michael Noseworthy,

[1] the term *wild* is used to express the idea of moving a model to an environment where the model has never been exposed to it before [1]

and Steven Yule. Artificial intelligence nomenclature identified from delphi study on key issues related to trust and barriers to adoption for autonomous systems. *arXiv preprint arXiv:2210.09086*, 2022.

[C4] Sajjad Rashidiani, Thomas Doyle, Reza Samavi, Laura Duncan, Paulo Pires, and Roberto Sassi. Textionnaire: An NLP-Based questionnaire analysis method for complex and ambiguous task decision support. In *The 21st IEEE International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, In Press, 2022.

[C5] Moe Sabry and Reza Samavi. ArchiveSafe LT: Secure long-term archiving system. In *Proceedings of the 38th Annual Computer Security Applications Conference*, pages 936–948, 2022. (IF: 4.46).

addressing different aspects of modern machine learning and archiving systems security challenges, I consider organizing the AAAI symposium as the major achievement that became possible by the support provided by the Catalyst fellowship program as described below.

In line with my Catalyst project objectives and in collaboration with researchers from McMaster University (Canada), Virginia Tech, Cornell University, and Mayo Clinic (USA), and the University of Edinburgh (UK), we organized a symposium on "Human Partnership with Medical AI: Design, Operationalization, and Ethics." At the heart of this proposal was topics related to the security of AI system. This summer series AAAI symposium was held at Singapore EXPO, Singapore from July 17-19, 2023.

The three-day symposium comprised keynote speakers, technical and position paper presentations, break-out sessions, and interactive expert panel discussions. The symposium brought together researchers and practitioners from academia and healthcare to discuss the challenges and opportunities of AI-human partnership, share their latest research and insights, and develop actionable strategies to create trustworthy, ethical, and secure AI systems.

During the symposium, participants explored several research areas about AI-assisted medical advisory systems, including, Human-AI partnership, Reliability and robustness, and Fairness and explainability of of AI systems. Among the themes, the reliability and robustness of AI systems were closely related to the objectives of this catalyst project. In this research theme, we intended to answer to what extent sources of errors, bias, uncertainty, violation of privacy and adversarial inputs can be captured, and the risk can be mitigated in clinical AI settings. What metrics should be used/developed for benchmarking AI systems' robustness and reliability? The guest speaker, Dr. Fang Liu (Department of Applied and Computational Mathematics and Statistics, University of Notre Dame), discussed formal privacy guarantees in practice. Researchers from Toronto Metropolitan University, the University of Notre Dame, and the Arctic University of Norway presented

and discussed uncertainty quantification of DNNs, privacy-preserving data synthesis, and ethical challenges in using health synthetic data, respectively.

At the end of the first two days of the symposium, a break-out session followed by a round table discussion covered the future of medical AI partnerships, enhancing trust in AI, and improving clinical adoption. On the third day, a summary of breakout session outcomes was shared with the participants, and then the panellists discussed the paths to success in addressing the core challenges of AI-assisted medical tools. In addition to the talks, the symposium also ran two surveys to better understand the ethical and security challenges of medical AI partnerships. The responses were discussed with the symposium participants for consensus and then ranked based on the complexity and importance of the concepts. The outcome is expected to provide the community with insight and research directions with the greatest impact in the pursuit of mitigating ethical risks of medical AI and promoting responsible AI.

The papers submitted to the symposium are published as AAAI summer series symposium proceedings and will become available soon.

1.2 Service

The fellowship also provided the opportunity to engage with the larger community of cybersecurity researchers and industry practitioners via a number of events including, one-on-one meetings with the directors of Rogers Cybersecure Catalyst, Mastercard Emerging Leaders Cyber Initiative reception, and celebrating the Candian Cyber Innovation event.

I also had a chance to engage in service activities with the program by providing a Cybersecurity Catalyst story for the Office of the Vice-President, Research and Innovation (OVPRI). The story is published at the OVPRI website at `https://www.torontomu.ca/research/publications/newsletter/`. I have also served as a member of the adjudication committee to select the new cohort of research and industry fellows for the year 2023-2024.

2 Team Accomplishments

Throughout the course of my one-year research fellowship, I had the opportunity to collaborate with several talented and dedicated fellow researchers and industry practitioners. Our teamwork yielded several notable achievements, which include:

1. **Structured Progress Tracking**: Our biweekly meetings provided a structured platform to track the progress of individual and collective research projects. This allowed us to stay on course, meet project milestones, and promptly address any roadblocks encountered along the way.

2. **Dynamic Problem-Solving**: Through brainstorming sessions during these meetings, we collectively tackled cybersecurity challenges, specifically designing and planning cybersecurity webinars. The exchange of ideas and diverse perspectives led to

the design of three important webinars and knowledge dissemination through panel discussions and publications as described below.

3. **Webinar Planning**: In collaboration with other research and industry Catalyst fellows and with the leadership of the fellowship program director, we were able to design three webinars focusing on three major aspects of Cybersecurity: 1) cybersecurity research, 2) cybersecurity education and 3) cybersecurity commercialization. The composition of the webinar was carefully selected by the team to include diverse experts from academia, industry, government and other branches of cybersecurity policymakers.

4. **Webinar Organizing**: With my fellow industry Catalyst fellow AJ Khan, I had the opportunity to organize the Cybersecurity Commercialization webinar. The objective of this webinar was to investigate the challenges cyber startups face in commercializing innovative Cyber technologies. These challenges include: a relatively small domestic market, risk-aversion and a lack of innovation mindset in Canada, struggles to secure early adopters, lack of cybersecurity expertise and connections among investors, and lack of perceived value of IP protection. Panellists were Heather Galt, Growth Coach and Mentor Startups & Scaleups from Communitech, Joanna Ma, Partner (Patent Lawyer & Agent), from Bereskin & Parr and Ian Paterson, CEO, Plurilock (TSXV: PLUR). The webinar was moderated by AJ Khan and myself.

5. **Webinar Panel Participating**: I had the opportunity to serve as the panellist in the Cybersecurity Research webinar, sharing my views on security research and how we can bridge the gap between academia and industry and how to generate synergy between application and research in cybersecurity. The recording of the webinar is available on the Catalyst website.

6. **Publication of Research Papers**: Our collaborative efforts led to the authorship of three papers, synthesizing cybersecurity challenges from three perspectives of research, education and commercialization. We authored and published three research papers in reputable cybersecurity conferences and journals. The paper reporting the findings of the first webinar entitled "Bridging the Bubbles: Connecting Academia and Industry in Cybersecurity Research," has been accepted to appear in the Proceedings of the 2023 IEEE Secure Development Conference. The paper on Cybersecurity Education challenges (the topic of Webinar #2) is being submitted to the Computers and Security Journal, and the paper of the third webinar (Cybersecurity Commercialization) is in progress and scheduled to be submitted by mid-October 2023. These papers contribute to the academic and industry community's understanding of emerging cybersecurity challenges.

3 Lessons Learned and a few Recommendations

The fellowship was well organized and the advantages are evident both at the individual level and collectively as a team, as described in Section 1 and 2, respectively.

Improving the next cohort of the fellowship is crucial for ensuring its continued success and impact of the program. Here are several recommendations to enhance the experience for future fellows:

- **Diverse Selection Criteria**: While compared to our cohort, I noticed the new cohort, specially in the academic side is more diverse, I think considering diversifying the selection criteria for fellows to encompass a wider range of backgrounds, including not only computer science but also law, ethics, and other related fields. This diversity can bring fresh perspectives and innovative approaches to cybersecurity challenges.

- **Mentorship Program**: Establish a mentorship program within the fellowship, where experienced researchers or industry professionals can guide and support junior fellows. This mentorship can help fellows navigate their research projects and career development.

- **Interdisciplinary Collaboration**: Encourage interdisciplinary collaboration among fellows by fostering connections with other research programs or departments. This can lead to a broader understanding of the complex and emerging issues, for instance, at the intersection of cybersecurity and machine learning.

- **Practical Training**: Incorporate practical training components, such as workshops, hackathons, or simulations, into the fellowship curriculum (e.g., practical workshops on Cyber Range). Hands-on experience can deepen fellows' understanding of cybersecurity challenges and their ability to develop effective solutions.

- **Ethical Considerations**: Include a strong focus on ethical considerations in research and practice. Fellows should be well-versed in the ethical implications of their work and equipped to make responsible decisions in the development of cybersecurity solutions.

- **External Partnerships**: Foster partnerships with industry, government agencies, and nonprofit organizations to provide fellows with opportunities for real-world application of their research. These partnerships can also offer valuable insights into industry needs and trends.

- **Flexibility and Adaptability**: Recognize that the field of cybersecurity is rapidly evolving. Ensure that the fellowship program remains flexible and adaptable to accommodate emerging trends and challenges. This is specifically related to the project described in Section 1.1.

Of course, not all recommendations can be implemented in one year, but I believe that having these recommendations in mind, the next cohorts of the fellowship can be better equipped to address the complex and evolving cybersecurity issues while fostering a diverse and inclusive community of researchers.

References

[1] Battista Biggio and Fabio Roli. Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84:317–331, 2018.

[2] Thomas E Doyle, Victoria Tucci, Calvin Zhu, Yifei Zhang, Basem Yassa, Sajjad Rashidiani, Md Asif Khan, Reza Samavi, Michael Noseworthy, and Steven Yule. Artificial intelligence nomenclature identified from delphi study on key issues related to trust and barriers to adoption for autonomous systems. *arXiv preprint arXiv:2210.09086*, 2022.

[3] Yuting Liang and Reza Samavi. Advanced defensive distillation with ensemble voting and noisy logits. *Applied Intelligence*, 53(3):3069–3094, 2023. IF: 4.602.

[4] Moe Sabry and Reza Samavi. ArchiveSafe LT: Secure long-term archiving system. In *Proceedings of the 38th Annual Computer Security Applications Conference*, pages 936–948, 2022. (IF: 4.46).

[5] Sajjad Rashidiani, Thomas Doyle, Reza Samavi, Laura Duncan, Paulo Pires, and Roberto Sassi. Textionnaire: An NLP-Based questionnaire analysis method for complex and ambiguous task decision support. In *The 21st IEEE International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, In Press, 2022.

[6] David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. *Advances in neural information processing systems*, 28:2503–2511, 2015.

[7] Mini Thomas, Omar Boursalie, Reza Samavi, and Thomas E Doyle. Bayesian-based parameter estimation to quantify trust in medical devices. In *International Workshop on Health Intelligence @AAAI23 conference*, pages 95–108. Springer, 2023.