Trustworthy AI Solutions for Cyberbiosecurity Challenges in Water Supply Systems

Wan-Yi Mao[1], **Mehmet Yardimci**[1], **Minh T. Nguyễn**[1], **Dan Sobien**[1], **Laura Freeman**[1], **Abdul Rahman**[2], **Vinita Fordham**[2], **Feras A. Batarseh**[1]

[1]Virginia Polytechnic Institute and State University (Virginia Tech), Arlington, VA

[2]Deloitte Touche Tohmatsu Limited, Arlington, VA

wanyi@vt.edu, oguzy@vt.edu, mnguyen0226@vt.edu, sdan8@vt.edu, laura.freeman@vt.edu, abdulrahman@deloitte.com, vfordham@deloitte.com, batarseh@vt.edu,

Abstract

The recent increased adoption of AI into cyber-physical systems requires that AI models perform as intended and without security blunders. However, for such critical AI deployments, serious adversaries and data poisoning issues are still persistent. The digital (data-driven) infrastructure of the bioeconomy is progressively being utilized by state and private sector entities; especially through leveraging models and data sets for real-time decision support. Models however, can be easily targeted by adversaries or be prone to unintentional quality issues. Accordingly, our research is dedicated to defining methods that mitigate different types of attacks and their undesired consequences. We are building an AI framework that can detect anomalies and malicious acts (in water supply systems) as well as analyze cause-effect measures in multiple scenarios. The proposed defense strategies include assessing the potential vulnerabilities that can be exploited in the convergence of AI and physical systems. Such outliers can cause extensive civilian harm, produce damaging bio-security incidents, and threaten agricultural and food production. The new framework will be applied to multiple physical systems with the aim of creating meta-learning outcomes for supporting the wide-scale adoption of trustworthy AI.