



COLLEGE of ENGINEERING
AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

PhD Seminar 1

Friday April 12, 2024, at 10AM, online via Zoom (Remote)

Hedyeh Nazari

*Autonomous Detection and Response to Disinformation and
Misinformation Campaigns on Social Networks*

Advisor: Dr. Ali Dehghantanha

Co-Advisor: Dr. Fattane Zarrinkalam

Advisory: Dr. Gautam Srivastava (Brandon University)

Advisory: Dr. Natalia Stakhanova (University of Saskatchewan)

Abstract:

In the digital era, social media's rise has unfortunately facilitated the spread of disinformation and misinformation, challenging the fabric of public discourse and democracy. Existing methods in social cybersecurity, including troll detection, bot detection, cyberbullying, and account repurposing detection, have struggled to keep pace with the sophisticated tactics employed by malevolent actors. These traditional systems, primarily reliant on static, historical data for pattern recognition, are increasingly ineffective against the dynamic, evolving nature of cyber threats, highlighting a critical need for innovation in real-time threat identification and mitigation strategies.

Our research contributions address this gap by proposing an autonomous detection system that will be designed to counteract disinformation and misinformation campaigns on social networks. Leveraging deep reinforcement learning, our solution aims to overcome the limitations of traditional cybersecurity systems that rely heavily on historical data. This innovative approach allows for dynamic decision-making, adapting and evolving to counter novel threats effectively, with minimal human intervention. Focusing on three primary challenges—the detection of human-operated troll accounts, identification of malicious bots, and recognition of repurposed social network accounts—our research aims to introduce a robust framework capable of autonomous, real-time threat detection and mitigation.

Furthermore, our research proposes a secondary contribution by developing a simulated social network testing environment. This tool aims to overcome the limitations imposed by the absence of a versatile simulator in current research. By facilitating the creation and interaction of detailed social media personas, this environment allows for the exploration of new and unseen misinformation campaigns' impacts in a controlled setting.