Contribution ID: **99** Type: **Lightning 5 min talk + poster**

AI Red Teaming for Science

Tuesday 15 October 2024 16:10 (5 minutes)

AI Red Teaming, an offshoot of traditional cybersecurity practices, has emerged as a critical tool for ensuring the integrity of AI systems. An under explored area has been the application of AI Red Teaming methodologies to scientific applications, which increasingly use machine learning models in workflows. I'll highlight why this is important and how AI Red Teaming can highlight vulnerabilities unique to ML-based systems used in scientific research. This approach not only protects against malicious actors but enhances the routine functioning of AI systems in scientific research. I will also briefly introduce FABRIC, an NSF testbed for optimizing science cyberinfrastructure, and show how it might be used for AI Red Teaming.

Focus areas

Author: NIKOLICH, Anita (UIUC)

Presenter: NIKOLICH, Anita (UIUC)

Session Classification: Lighting talks