# RESILIENT ARTIFICIAL INTELLIGENCE

## MONDAY: TRACK 3

## SESSION CHAIR

Char Sample, Idaho National Laboratory

## SESSION ABSTRACT

Artificial intelligence (AI) and machine learning (ML) are gaining a larger presence in cyber security and as such are being explored as new attack vectors into secure areas. The attack vectors are numerous and in the early stages of exploration, thus providing an opportunity to establish methods and techniques to improve the resilience and reliability of AI/ML. Attacking AI/ML has resulted in growing research areas in the malicious use of artificial intelligence (MUAI) and adversarial machine learning (AML). As MAUI/AML mature, the lessons learned present challenges to creating and maintaining trustworthy AI. Some examples of topics that challenge trustworthiness of AI are data manipulation, weight manipulation, algorithm biases, and even the supply chain. This workshop seeks to bring the discussion points to the forefront through discussions of the following topics.

- Training data manipulation
    - Machine learning classifiers
    - AI algorithm weights
- How AI biases can be identified and repaired
- AI/ML applications in supply chain resilience identification and repair
- MUAI
- AML
- Explainable AI
- Creating Trustworthy AI