Enhancing Situation Awareness of Adversary ML in Human-AI Collaboration for Safe Implementation of Automated Driving Systems

- Yiqi Zhang, College of Engineering (Department of Industrial and Manufacturing Engineering)
- Aiping Xiong, College of Information Sciences and Technology

Abstract: Artificial intelligence systems in automated driving systems (ADS) are known to be vulnerable to adversarial attacks, making ADS susceptible to safety- and security-critical errors that pose significant road hazards and fatalities. Recent research exploring human drivers' understanding of physical-world attacks indicates a lack of awareness of AI vulnerability, which may pose significant risks to driver safety. This project aims to establish a conceptual framework for driver-AI collaboration in response to adversarial attacks by addressing driver situation awareness. To this end, we will carry out behavioral studies to (1) examine the impact of AI system capabilities and reliability on driver situation awareness, trust, and takeover performance across various adversarial attack scenarios; and (2) investigate central/distributed human-machine interface design to promote driver situation awareness restoration and effective human-AI collaboration. The proposed research will advance socially responsible AI in transportation by deepening our understanding of driver situation awareness of adversarial attacks to enhance AI system resilience. The project's findings could guide Level 3 automated vehicle design optimization, promoting emerging AV technologies and enhancing driver safety. This project's outcomes will be used to obtain external research funding and build a long-term, multidisciplinary research program at Penn State in the area of human-AI collaboration. Additionally, this research will contribute to the Larson Transportation Institute's national leadership in testing automated vehicles.