Problem Statement

As robots become increasingly autonomous and are deployed in complex, real-world environments, their ability to adapt to novel, unforeseen situations becomes critical. Traditional robotics systems rely heavily on hard-coded rules or pre-defined paths, which are insufficient for dynamic and unstructured environments. Recent advancements in artificial intelligence—particularly reinforcement learning (RL) and imitation learning (IL)—enable robots to learn from interactions and adjust their behaviors without explicit programming. This shift toward adaptive behavior significantly enhances functionality but also introduces new security vulnerabilities and trust concerns.

Autonomous robots must process large volumes of sensor data, make real-time decisions, and interact with the physical world. If these systems are compromised—through sensor spoofing, adversarial attacks on learning models, or unauthorized command injections—the consequences can be severe, affecting both safety and performance. Furthermore, the black-box nature of many learning-based systems undermines user trust and raises questions about transparency, accountability, and resilience under attack.

This research addresses the urgent need to secure dynamic robotic behavior by developing a dual framework that:

1. Enables robots to learn and adapt to unexpected events and environments, and
2. Embeds robust cybersecurity mechanisms to defend against manipulation, ensuring the system's integrity, reliability, and trustworthiness.