FY22/23 DAGSI Research Topic

1. **Research Title:** Counter-Autonomy through Adversarial Reinforcement Learning
2. **Individual Sponsor:**

Dr. Corey Schumacher
AFRL/RYZA, Bldg 620
2241 Avionics Circle
WPAFB, OH 45433-7333
corey.schumacher@us.af.mil

3. **Academic Area/Field and Education Level**

Aerospace Engineering (MS or PhD level)
Mechanical Engineering (MS or PhD level)
Electrical Engineering (MS or PhD level)
Computer Science (MS or PhD level)
Mathematics (MS or PhD level)
Robotics (MS or PhD level)

4. **Objectives:** Research and develop methods to find exploitable flaws in an autonomous system's artificial intelligence capabilities, and then to improve the robustness of the AI system against adversarial exploitation.
5. **Description:** Autonomous decision systems based on artificial intelligence algorithms are potentially exploitable by adversarial machine learning approaches. Typically, adversarial AI approaches require full knowledge (white box access) of the victim AI, or unlimited access to query it (black box access) to train an adversarial AI agent to defeat the victim AI. Additionally, many published approaches assume the adversary can modify the input to which the victim AI is reacting. Recent methods then use these approaches to robustify Deep Reinforcement Learning (DRL) agents by then training against "adversarial" agents.
   Such methods are appropriate for studying AI robustness and vulnerability, and have valuable application in testing and improving the robustness of DRL AI's. However, existing methods tend to be ad hoc, lean heavily on unlimited access to train against a targeted AI agent, and lack a systematic or analyzable method. This goal of this project is to find more systematic, repeatable, and analyzable methods to find vulnerabilities in AI decision agents based on deep neural networks. Open-source game environments or simple simulations should be used as the development and test environment.
6. **Research Classification/Restrictions:** Unclassified
7. **Eligible Research Institutions:** All DAGSI-eligible institutions

PA Approval #: AFRL-2022-3955