2021

Weapon Systems Safety When Deploying AI Technology

Berzins, Valdis A.

Monterey, California: Naval Postgraduate School

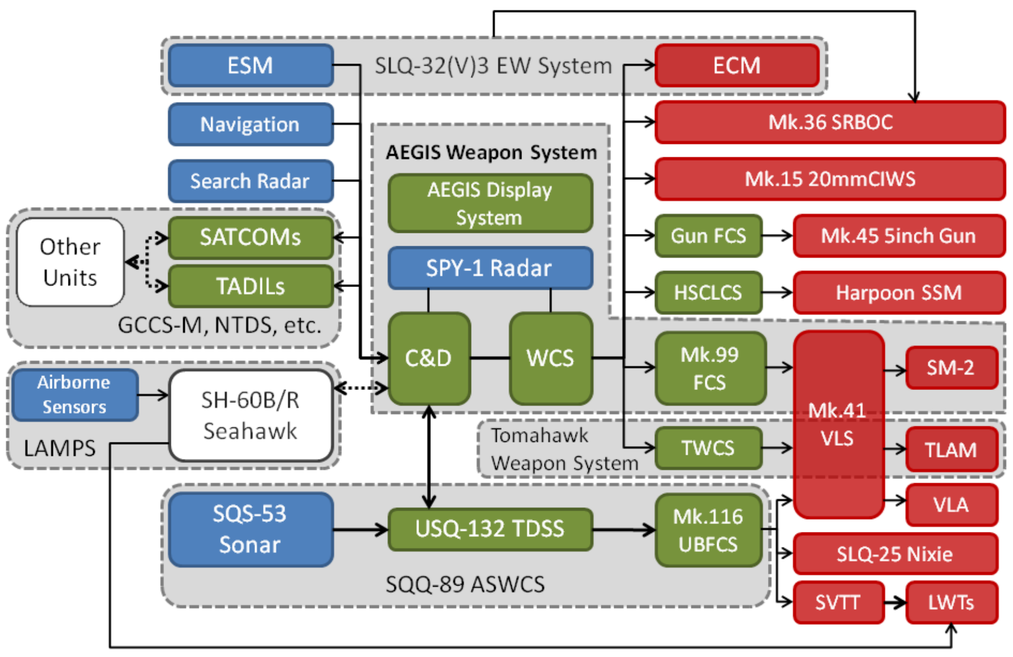
Weapon Systems Safety When Deploying AI Technology

Naval Postgraduate School

Project Description:

Safety concerns arise when deploying AI technology in weapon systems, autonomous systems carrying weapons/sensors/comms, and mission planning and decision aids. When AI technology is deployed in a weapon system, robot, or planning system, there is a likelihood that unwanted events occur. Naval Ordnance Safety and Security Activity (NOSSA) is responsible for understanding that likelihood and making risk decisions on naval employment.

Artificial Intelligence (Altmann, 2019)



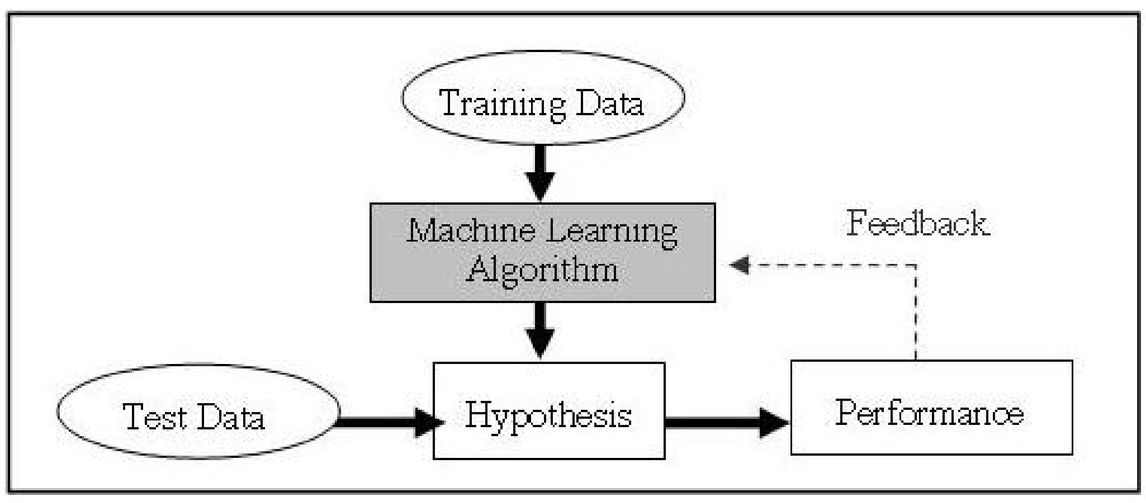
Aegis Weapons System (Wikimedia, 2010)

Research Questions:

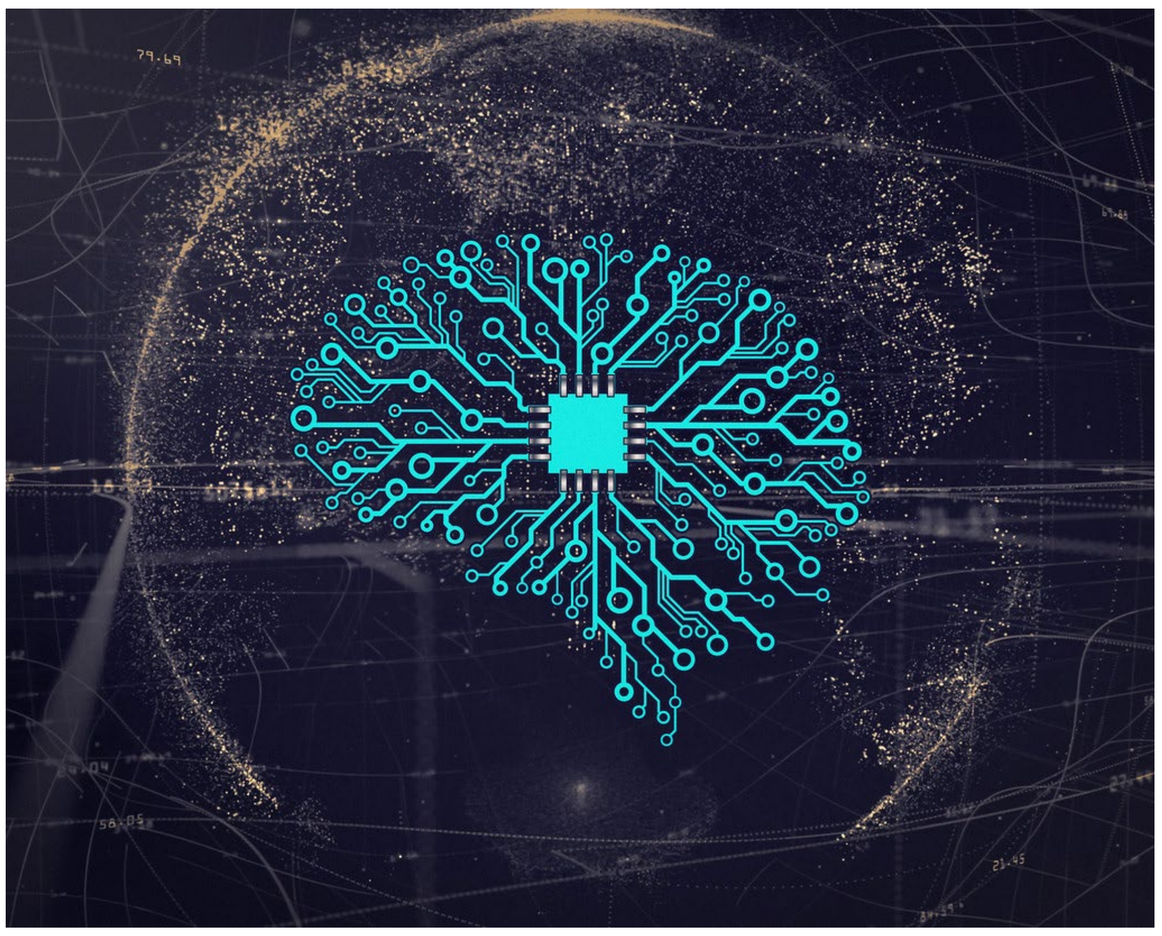
- What are the key factors to ensure system safety is addressed within the design and development of an AI deployed weapon system?
- What process is needed to understand how the design and development of a weapon system might cause an unwanted event?

Findings:

- The current state of knowledge in the area does not yet provide effective mitigations for several of these hazards.
- There is no known effective defense against adversarial examples.
- We developed a new defense against adversarial examples that improves performance over all previously known methods.



Machine Learning Technique (MacKenzie, 2018)



Machine Learning (Wikimedia, 2012)

This research is supported by funding from the Naval Postgraduate School, Naval Research Program (PE 0605853N/2098).

NRP Project ID: NPS-21-N387-A
Technical Report: http://hdl.handle.net/10945/68624

Researchers: Prof. Valdis Berzins,
Asst. Prof. A. Barton and Asst. Prof. J. Kroll | Computer Science, GSOIS
Topic Sponsor: Naval Information Warfighting Development Center (NIWDC)