Cybersecurity, Artificial Intelligence, and Autonomous Weapons: Critical Intersections

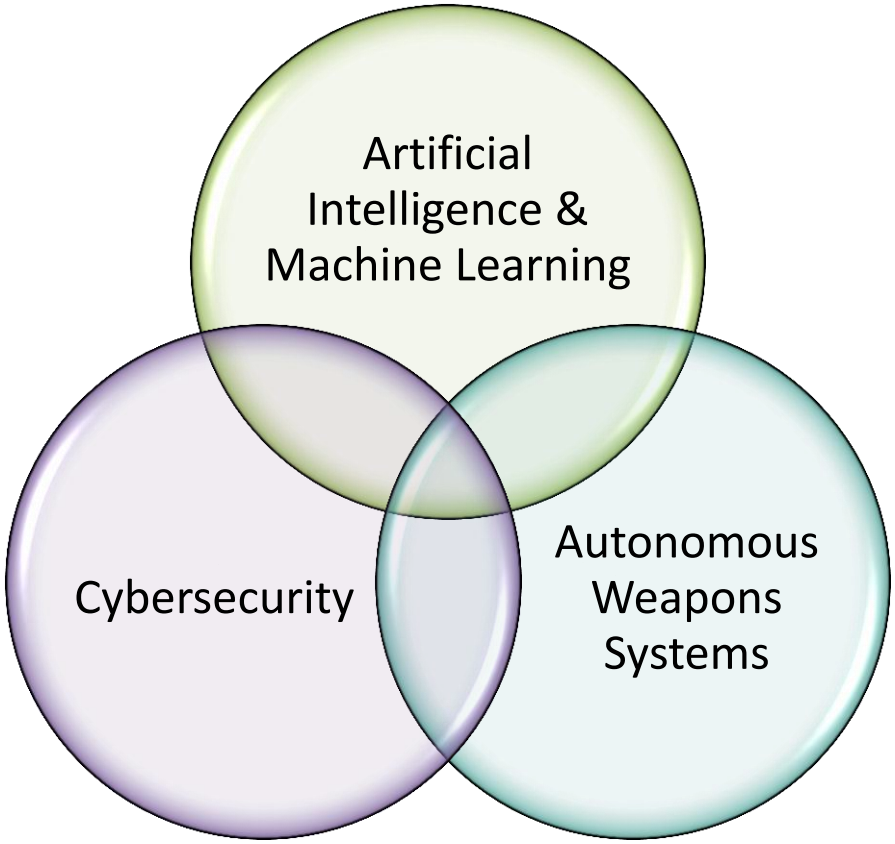
Heather M. Roff, Ph.D.

Senior Research Fellow
University of Oxford
Department of Politics & International Relations

Research Scientist
Arizona State University
Global Security Initiative

Cybersecurity Fellow
New America Foundation

Understanding the Relationships



Artificial Intelligence and Cyber Defense/Offense

- Defense of Critical Networks: real time, pattern finding, anomaly seeking
  - Must utilize machine learning algorithms to efficiently, and instantaneously respond to potential network threats

- Human on or out of the loop?

  - On the Loop:
    - Anomaly detection: human notified, IT analysis, response.

  - Out of the Loop:
    - Anomaly detection: AI decides best method of response: quarantine, honey pot monitoring, hack-back
    - This is an ***autonomous cyber weapon***.

Artificial Intelligence & Autonomous Weapons

- Autonomous weapons: a weapon that can select and engage a target without intervention by a human operator.

Are these machines artificially intelligent?

*No, Yes, and YES**

No: Present weapons systems are not capable of human level reasoning.

Yes: AI algorithms are presently employed to process sensor data, monitor system health, take and respond to vocal commands, manage data, navigate

YES*: Future Autonomous Weapons Systems will Require stronger AI to comply with ROE and IHL and be secure and operationally and cost effective. Moreover, self-aware autonomous cyber systems are crucial.

Cybersecurity Redux

What is cybersecurity?

A: the ability to control access to networked systems and the information they contain.

- Acts: prevent, detect, recover, react?
- Objects: people, process, technology
- Goals: confidentiality, integrity, availability

Resilience

Cyber Weapons

What is a cyber weapon?

- Malware? (viruses, trojans, zero-days, worms, ransomware, spyware, etc.)
- Does it require a particular objective? (military, paramilitary or intelligence)
- Does it require: physical harm? Functional harm or interruption? Mental harm?

A "weapon" presupposes that it is an object or tool. What about when it is an ***agent****?*

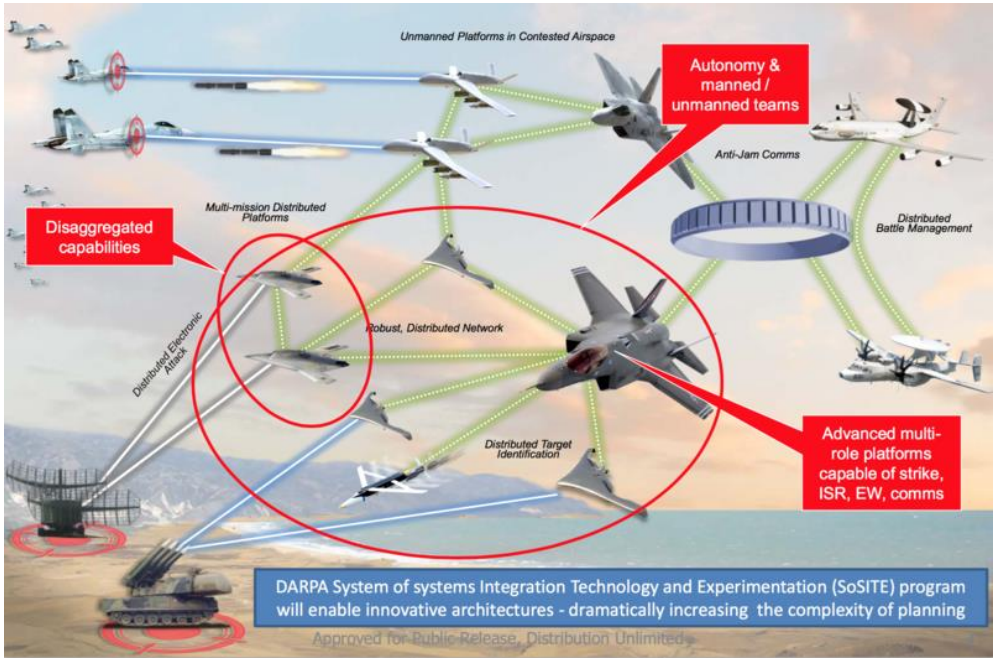
Systems of Systems

Platforms: weapons platforms: structures that launch weapons, i.e. jets, ships, vehicles. Platform + Weapon + Software Architecture = Weapons System

System of Systems: Platforms + Platforms

Systems of Systems: ISR, Processing and Platforms

$C^4I^{2:}$: System of Systems of Systems



Risks

Autonomy: the ability to problem solve; the power to act; the power to change course; ability to create a new goal. We cannot know a priori what an autonomous system will do.

Known Unknowns:

- Emergent behaviors
- Whatever system design we use, there will be cybersecurity problems arising from computational design/complexity
  - One can manipulate the system to act against itself
  - One can utilize traditional “cyber weapons” against the system
  - One can manipulate the system to lie to humans… but due to complexity there is no way to know if it is lying or not.
  - Bounded rationality: satisficing.

Unknown Unknowns:

- They are not anthropomorphic systems
- Learning, Reasoning, Communication
- “Self-Aware” systems