



Assurance Issues for Remote Operations: an AI/ML Cybersecurity Perspective

Dr. Natasha Neogi

NASA Langley Research Center

Panel on Cybersecurity for Remote Operations

AIAA SciTech Forum and Exposition

January 6, 2025

Remote operations in a nutshell...

- Range of operational paradigms including remote assistance, remote management, remote control, etc.

Characteristics

- A system is monitored and controlled by operators located at such a distance that they cannot physically intervene to correct issues that may arise
- Reliable communication between operator and system is critical
 - increasing autonomy & AI/ML implications
- The operator needs to maintain situational awareness
 - Systems sensors
 - External sensors (ground based RADAR)
 - Line of sight



NASA has a rich history with remote operations (air & space).

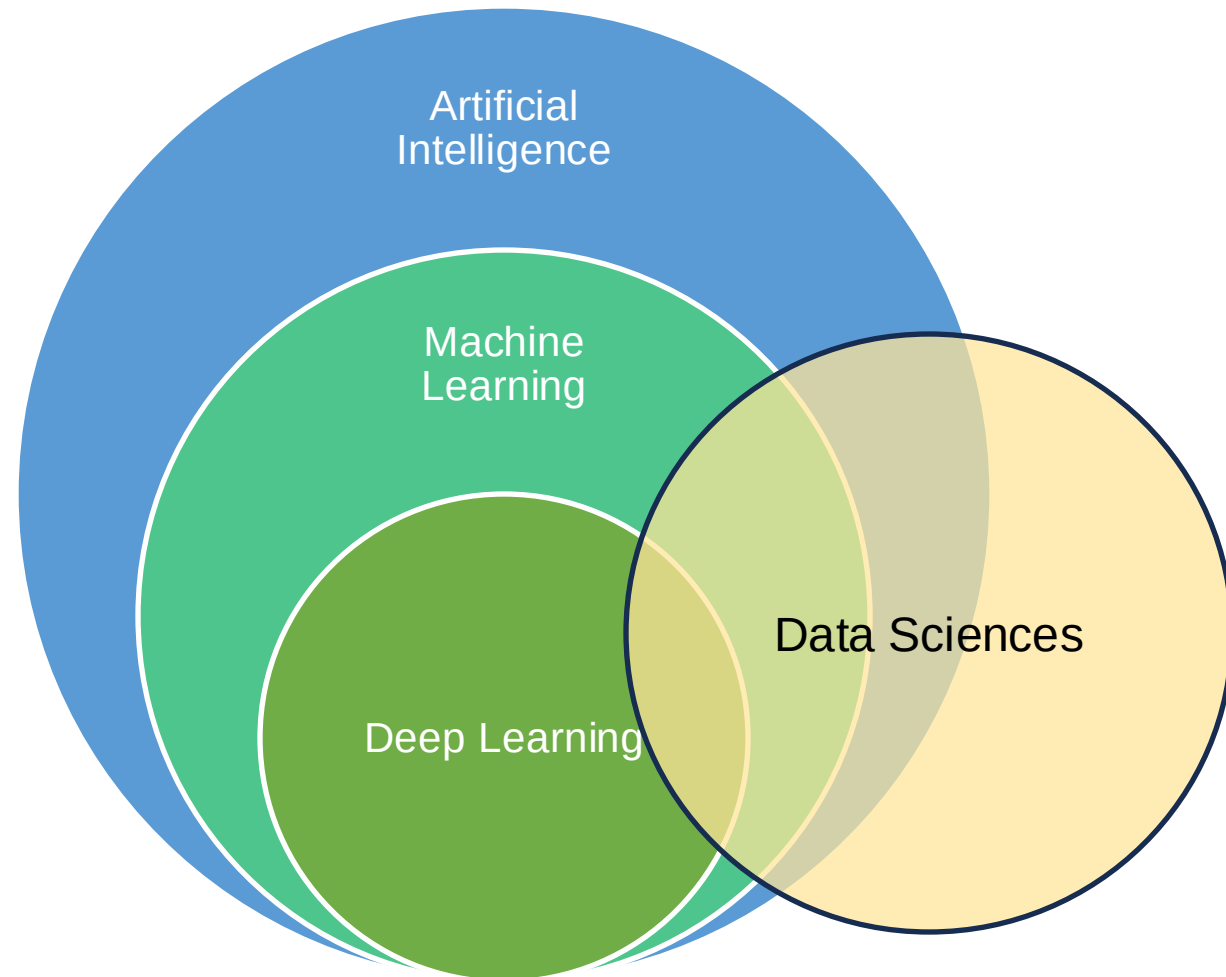
Hazards and Threats

- Hazard and threat analysis needs to be thorough
- Critical components will break and do so in malicious ways
 - Byzantine faults vs crash faults
- Communication will fail
 - How are you going to handle loss of communication and control
- There will be malicious actors who will attack your system
 - Conventional cybersecurity threats
 - Cybersecurity threats related to AI/ML



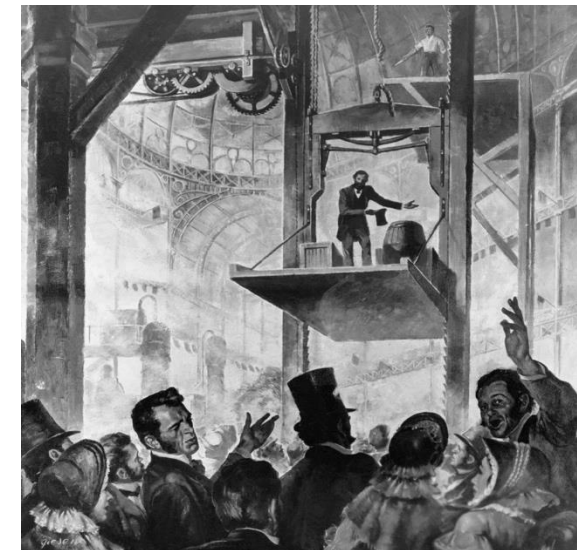
What do we mean by AI and ML?

- AI: “AI is the science of making machines do things that would require intelligence if done by men.”
 - Marvin Minsky, 1956
- ML: Use of statistical techniques to analyze data and create algorithms that can generalize to unseen data without explicit programming
 - <https://gradml.mit.edu/intro/>





Elevators carry two billion passengers a day over hundreds of millions of vertical miles in over 200 nations.



Courtesy of
Unknown author -
Copie de gravure
ancienne, Public
Domain,
<https://commons.wikimedia.org/w/index.php?curid=30135037>

Problem and Goal

Problem

- The inability to establish appropriate assurance for AI/ML components leaves us unable to effectively manage their risks and benefits.
 - Drives cost of development uneconomically high
 - Delays adoption of AI/ML at scale in safety critical systems
 - Results in unknown and unmanageable risks

Goal

- Discover and define what constitutes sufficient scientific-based evidence to substantiate a safety claim related to an AI/ML component performing a safety-critical function.

Towards standards for AI/ML...

- Standards require a stakeholder consensus on a driving need, commitment to support development, and subsequent application.
 - Broad sector of stakeholders should be involved, or uptake will suffer
 - SME contributions from all relevant or impacted stakeholders (e.g. aircraft OEMs, avionics OEMs, pilots, controllers, researchers, regulators, etc.)
- Standards development activities require a keen and deep understanding of the problem being solved and the technologies deployed in its reference implementation.
 - Understand mechanisms and limits of fundamental, underlying science of implementation and verification technologies
- NASA's Role is to:
 - Provide SME(s) to help create the standard;
 - Provide relevant findings from NASA R&D activities;
 - Learn of gaps or challenges that need to be addressed, then start up (or change) R&D activities to help fill these gaps or address the challenges.



A few key questions to answer...

How do we know when an ML component's behavior meets its requirements?

- Sufficient representation and size of training dataset, accuracy vs. generalizability, what constitutes an actionable specification, etc.

What are the limits of current processes and metrics currently used in developing and evaluating both traditional and ML systems?

- How do you use testing (i.e., creating logical based oracles, etc.), simulation (i.e., model validity), (formal) analysis (i.e., scalability), runtime verification frameworks, etc. in assurance?

How should information assurance be handled for ML components in order to yield (composable) safe systems?

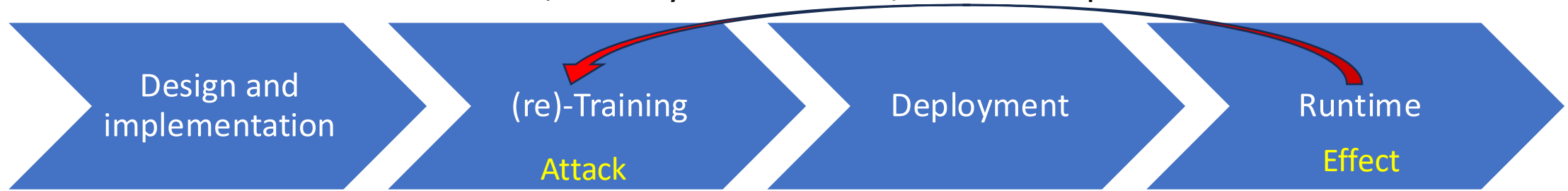
- Data fusion; information synthesis; data collection, curation, and assurance; **data poisoning**; etc.

How can change be managed in systems with ML components in order to preserve assurance?

- Configuration management, version control, database management, etc.
- Full recertification, continuous authorization to operate, etc.

Data poisoning and ML

- Adversary or malicious insider injects intentionally corrupted, false, misleading, or incorrect samples into training or fine-tuning datasets.
 - Attack can add to dataset; modify dataset and/or deletes portion of dataset.



- Potential impacts of data poisoning include
 - introducing biases;
 - creating erroneous outputs;
 - reducing precision, accuracy and/or recall;
 - introducing vulnerabilities (i.e., backdoors) and potential for system failure; and/or
 - influencing the decision-making or predictive capabilities of the ML component.
- Can occur subtly over time, making it challenging to identify until significant damage has already been inflicted.

Types of data poisoning

- Targeted
 - attempts to manipulate the model's behavior with respect to a specific scenario
 - does not change overall model performance
- Non-targeted
 - attempts to manipulate the dataset to negatively impact the overall performance of the model

Types of Attacks

- Mislabelling attack
- Data injection/manipulation
- Backdoor poisoning
- Availability attacks
- Model inversion attacks
- Stealth attacks
- Supply chain attacks

Combating data poisoning

Detection

- Model degradation
- Unintended outputs
- Increase in false positives/negatives
- Biased results
- Breaches/security events
- User Awareness
- Baseline misconfiguration

Mitigation

- Use of diverse data sources
- Data validation (including cross validation)
- Provenance tracking
- Monitoring, detecting, and auditing (including anomaly detection)
- Robust learning
- Secure data handling
- Adversarial Training

Combating data poisoning requires (1) classical cybersecurity knowledge, (2) understanding of ML fundamentals, and (3) foundational assurance principles.

Summary considerations for cybersecurity of remote operations

- **Provide highly reliable and secure communications:** Both safety and security must be considered in the communication link requirements.
- **Provide rich information:** The information sent must be rich and in real time, and available to all agents to encourage self-checking pairs or other strategies for redundancy.
- **Establish fault tolerance and redundancy in technical systems and/or human resources, as well as warnings:** If redundancy is missing in safety or security critical elements, the property will be compromised. Establish a culture that encourages use of redundancy (e.g., data, computational, communications, human, etc.) when applicable.
- **Consider the existence of a ‘quiescent state’ and the pace of the operations when designing safety and security into the system:** The ability to control, mitigate, and eliminate hazards and/or threats will depend on whether operations can be interrupted with no safety-, security- or mission-critical effects.
- **Focus on safety and/or security in function allocation:** A structured methodology should be used to allocate functions between onboard and offboard (i.e., ground/cloud-based) to improve safety and security as well as increase resilience.

Questions?

natasha.a.neogi@nasa.gov

Backups

When I say AI, I mean...

Machine Learning Systems

- Reinforcement Learning, Supervised Learning, Unsupervised Learning, Generative Systems...

Rule Based Learning Systems

- Production Systems, Expert Systems, Fuzzy Logic,...

Search and Optimization Techniques

- Uninformed search, Informed Search, Parallel Search,...

Decision Making under Uncertainty

- Bayesian Inference, Parameter Learning, Structured Learning,...

Evolutionary Strategy

- Evolutionary Algorithms, Swarm Based Algorithms,...

⋮

⋮

But wait! AI/ML is already here...

- Use of optimization tools during design of Boeing 787
 - Wing-Body Design
 - Composite Material Design
- Initial efforts at NASA to deploy AI/ML techniques in aviation contexts
 - Using Large Language Models to look for positive contributions to safety in ASRS reports
 - Anomaly Detection/Vulnerability Discovery
- and others...



Courtesy of Timo Breidenstein –
[http://www.airliners.net/photo/United-Airlines/ Boeing-787-822-Dreamliner/2142634/L/](http://www.airliners.net/photo/United-Airlines/Boeing-787-822-Dreamliner/2142634/L/), GFDL 1.2,
<https://commons.wikimedia.org/w/index.php?curid=20544763>

Takeaways

- Deployment and standardization efforts should proceed methodically and with a justifiable basis, thereby enabling safe adoption of ML applications in aviation.
- Premature efforts to (deploy and) standardize may damage paths to transition for ML technologies, engender technical debt, or set back the entire aviation industry.