SERC RESEARCH REVIEW 2023 | NOVEMBER 15, 2023

Trusted Artificial Intelligence Systems Engineering Challenge

WRT-1085

Sponsored by DEVCOM

PI: Peter Beling

NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

SERC Research Team

Virginia Tech

- Peter Beling (PI)
- Tyler Cody
- Stephen Adams

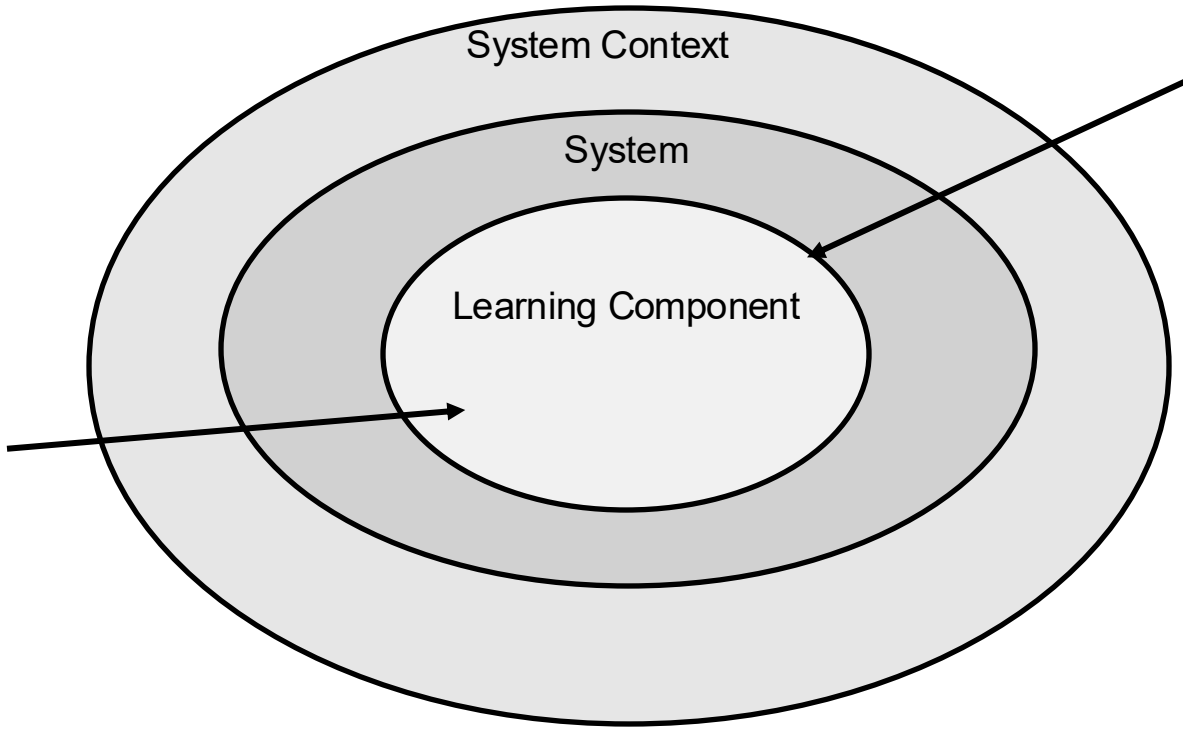
Stevens Institute of Technology

- Tom McDermott

Motivation

Premise

Focus on relationship between learning algorithms and their systems

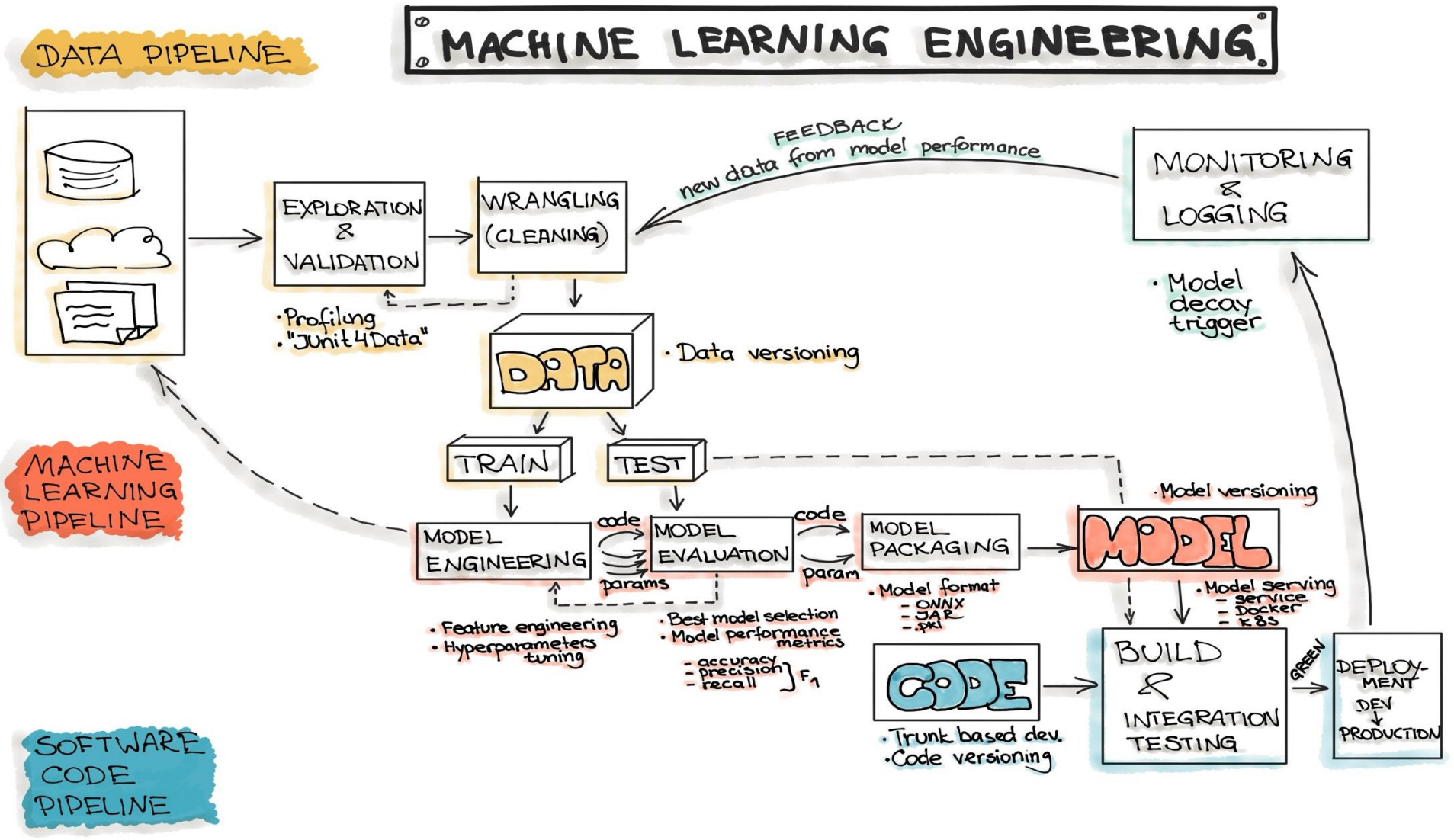


ML/AI focus on learning algorithms (and sometimes data)

Systems ≠ Problems

Proposition: Engineering intelligence requires focusing on **learning systems** not the **problems they solve**

ML Pipelines (And What They Miss)

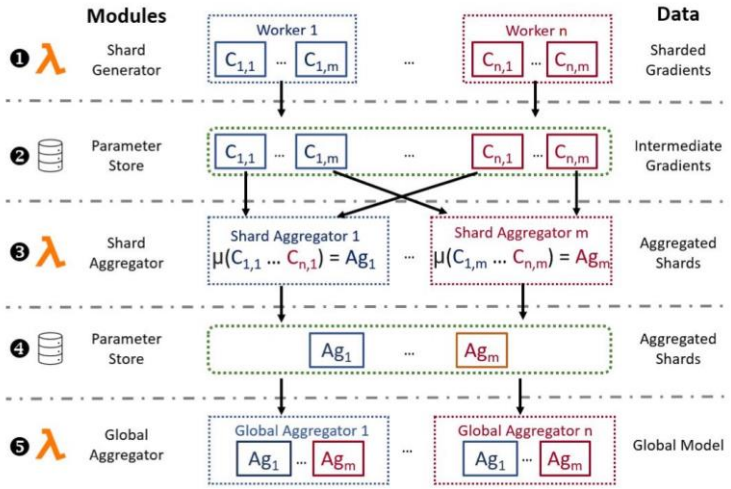


https://ml-ops.org/content/end-to-end-ml-workflow#data-engineering

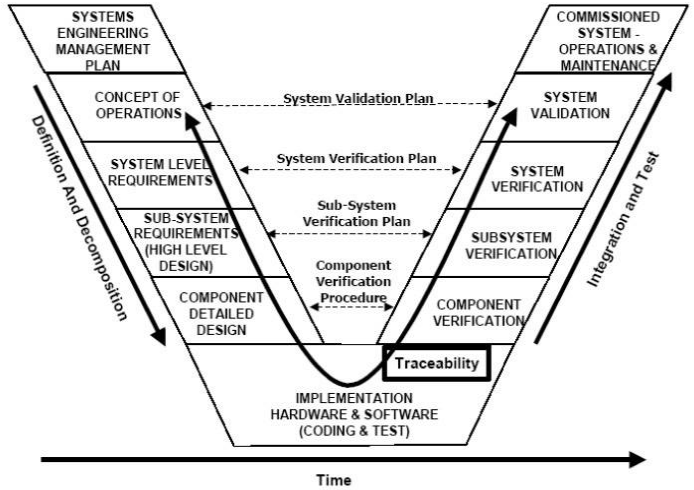
Developing Engineering Processes for Real-world AI

$$
\begin{aligned}
\textbf{minimize:} \quad & f(\mathbf{x}) \\
& \mathbf{x} \in \Re^n \\
\textbf{subject to:} \quad & \mathbf{g}_L \leq \mathbf{g}(\mathbf{x}) \leq \mathbf{g}_U \\
& \mathbf{h}(\mathbf{x}) = \mathbf{h}_t \\
& \mathbf{a}_L \leq \mathbf{A}_i\mathbf{x} \leq \mathbf{a}_U \\
& \mathbf{A}_e\mathbf{x} = \mathbf{a}_t \\
& \mathbf{x}_L \leq \mathbf{x} \leq \mathbf{x}_U
\end{aligned}
$$

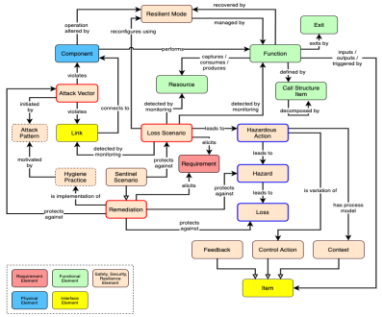
AI Research

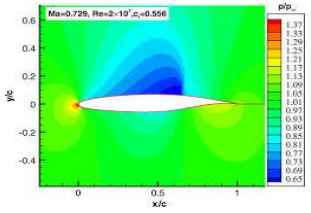


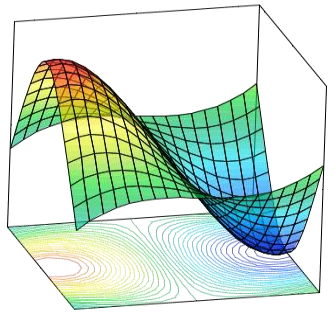
AI Engineering Research

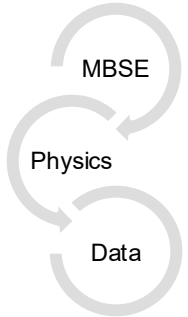


AI Systems Engineering Research





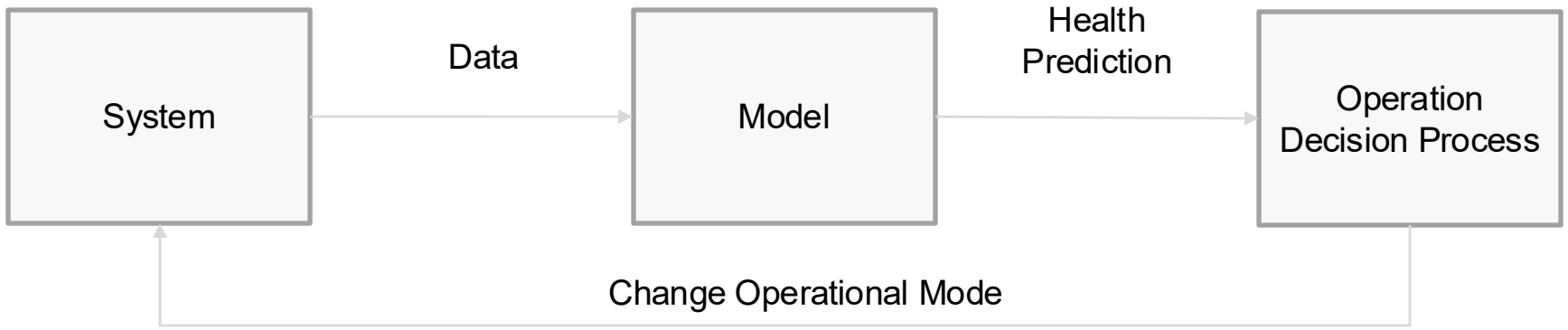




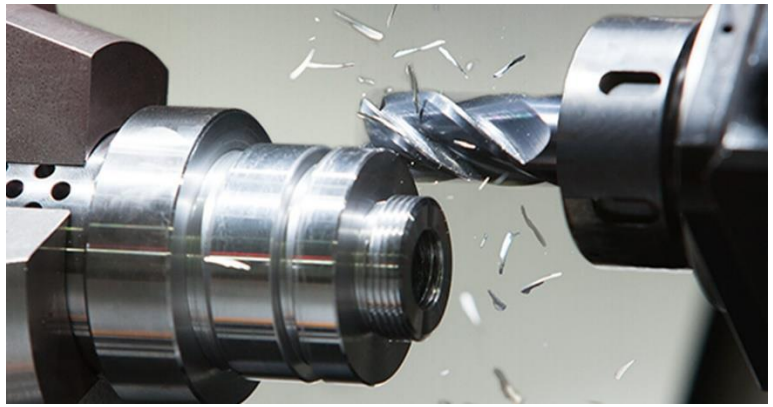
Treat AI as an engineered system rather than a problem-solving technique

Co-design system and algorithms to mitigate frailties of AI

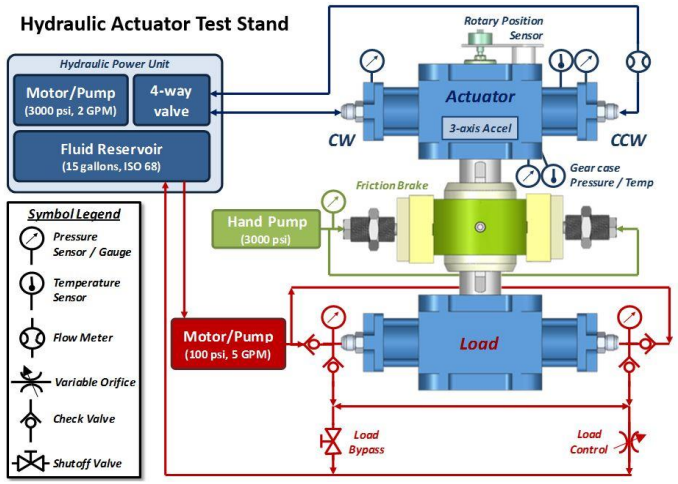
Prognostics and Health Management (PHM)



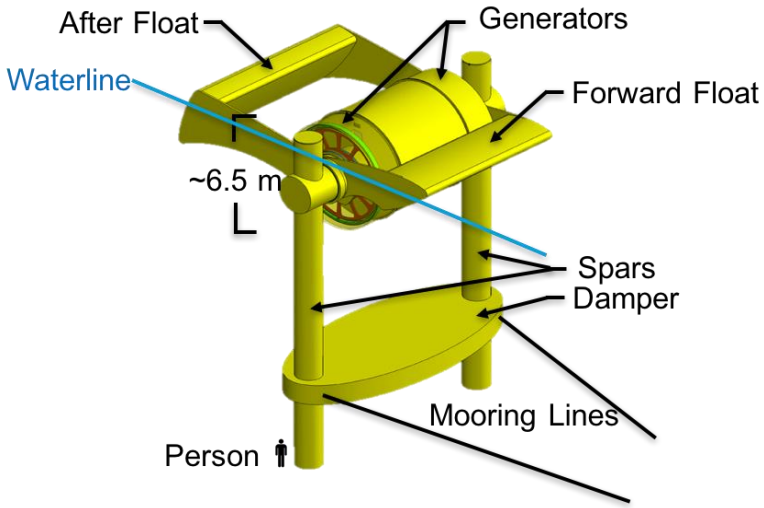
Tool Wear Prediction



Hydraulic Actuator CBM

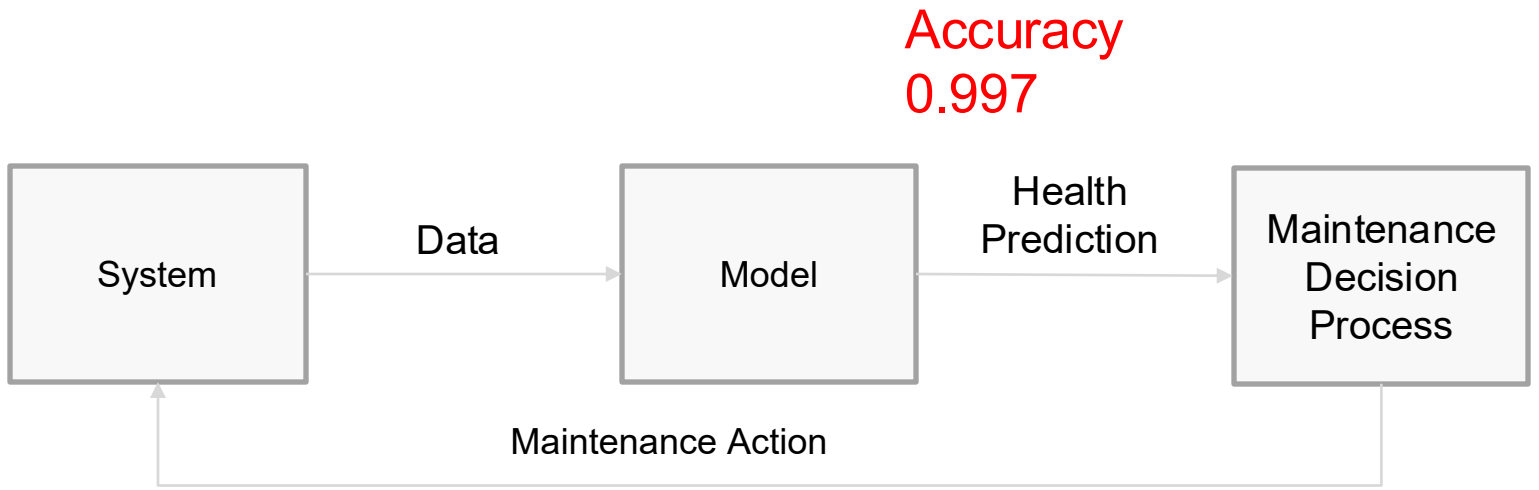


Structural Health Monitor for Wave Energy Converters



PHM on Machine as Produced by Factory

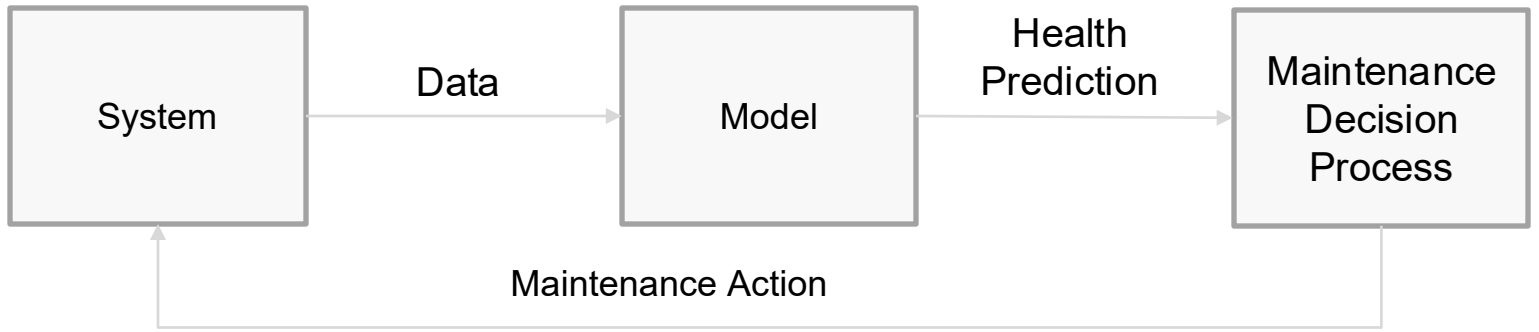
- Use knowledge/predictions about current and future health of a system to plan its maintenance
- Data-driven approaches use models to evaluate sensor readings



PHM on Field-repaired Machine

- Repair is successful !(?)
- Same mean time to failure
- Same operating characteristics

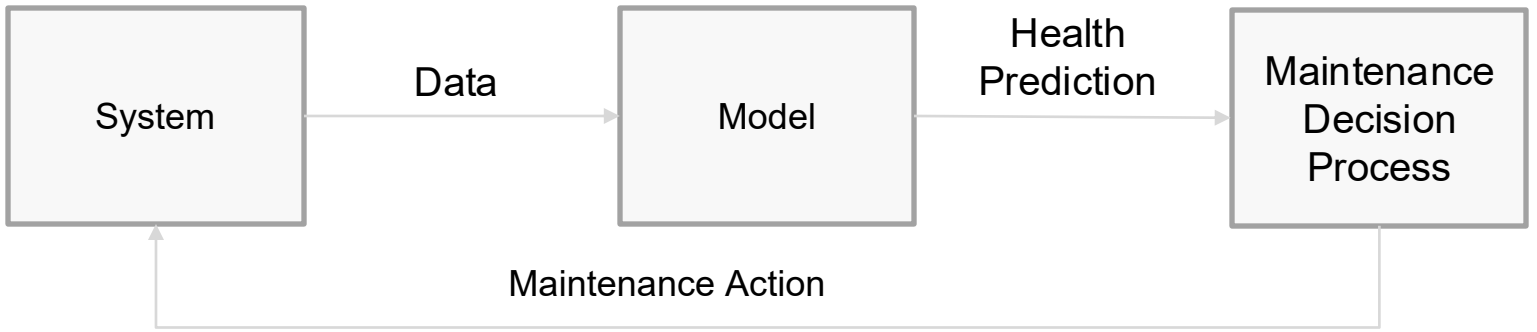
Accuracy 0.78 after **successful** maintenance



Transfer Learning after Field Repair

- Repair is successful !(?)
- Same mean time to failure
- Same operating characteristics

Accuracy 0.91 after transfer learning using healthy-case target data

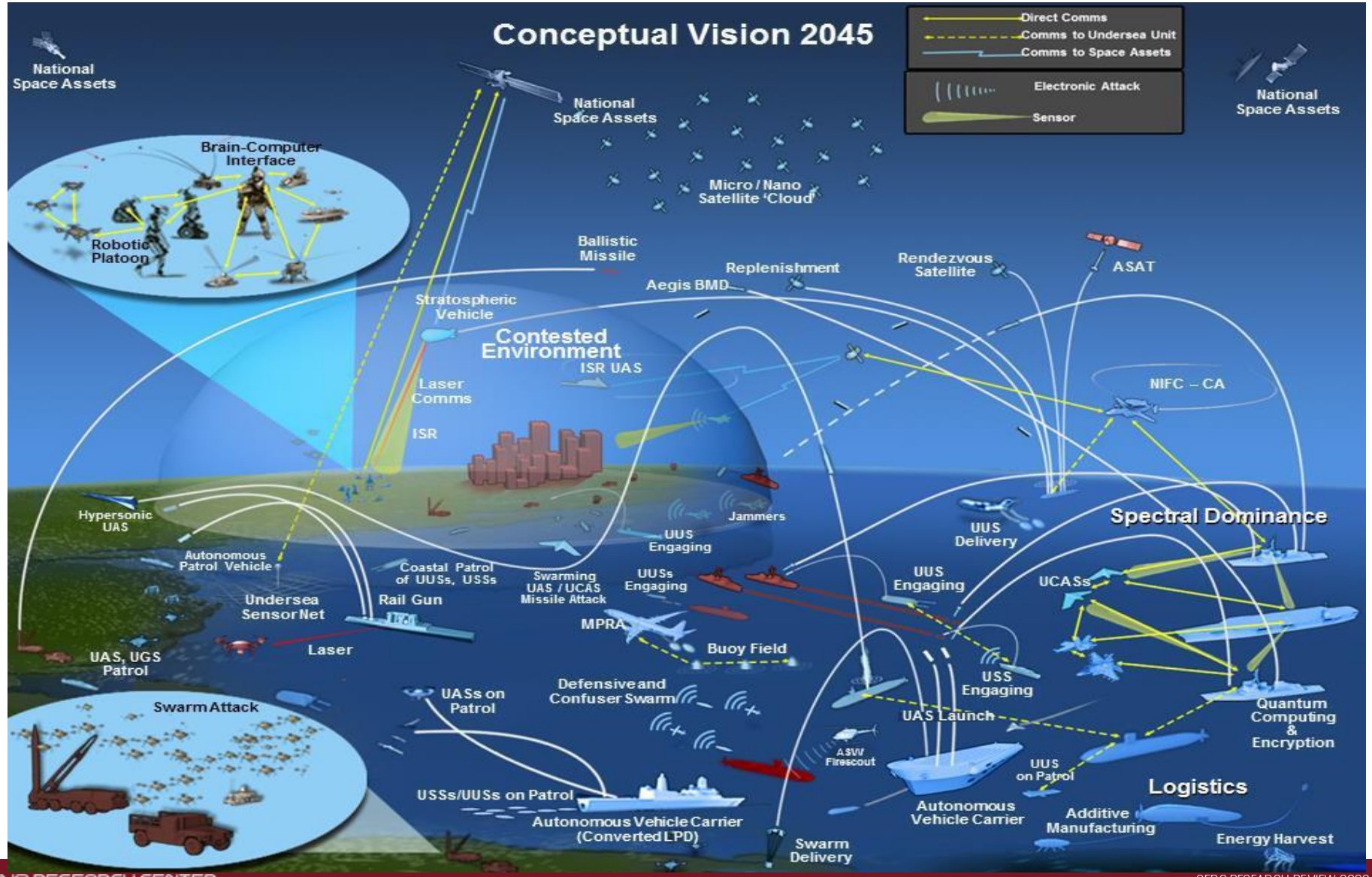


There is (essentially) no opportunity to learn from damage-case data in the target

Questions

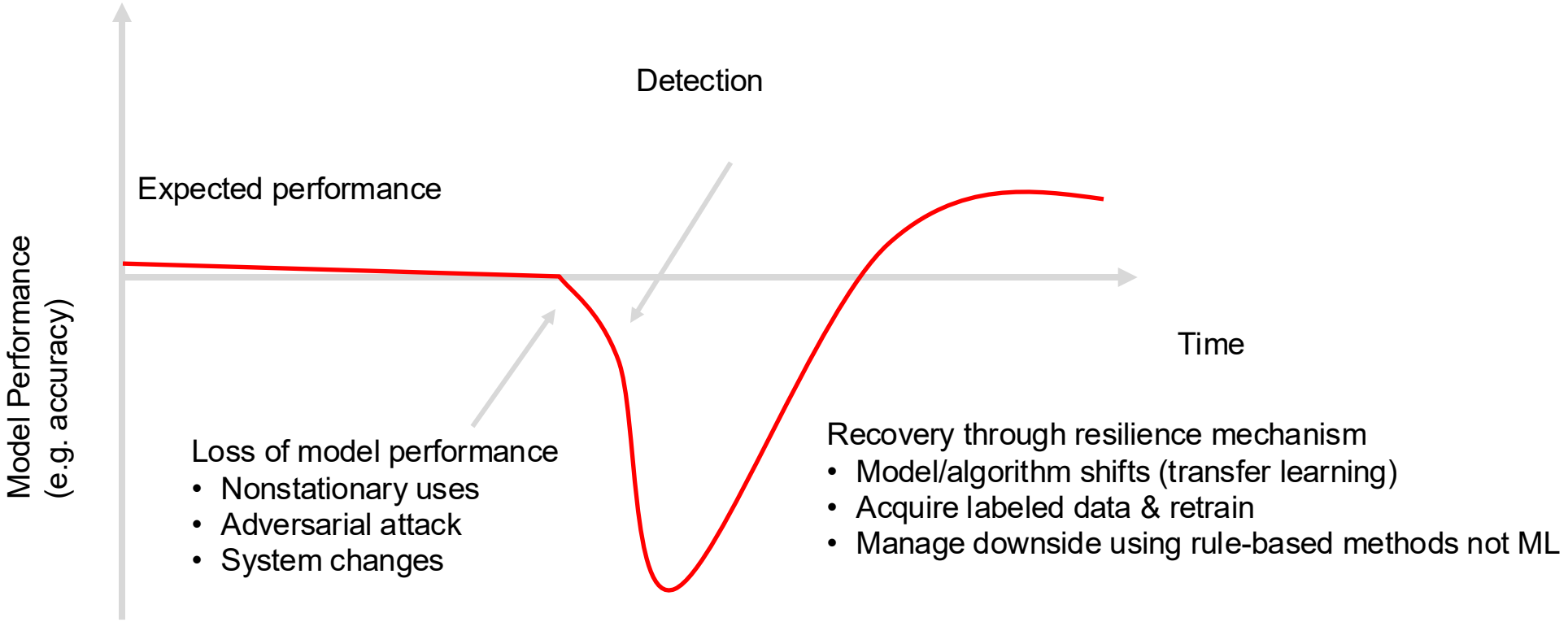
- What if 91% accuracy is not good enough?
  - Abandon condition-based maintenance and go back to a conservative, time-based scheme?
  - Wait for transfer learning methods to improve sufficiently?
- What if we could take advantage of the fact that we know what caused the change in the system?
- Could we alter the design and operation of the system?
  - Change maintenance process; e.g., torque bolts more uniformly, use OEM parts?
  - Are there any principles or tools that could guide system design and operation?
  - Would a systems view help? Is it enough?
- The life cycles of the physical and cognitive elements seem entangled….

But can't we just train on all system states?



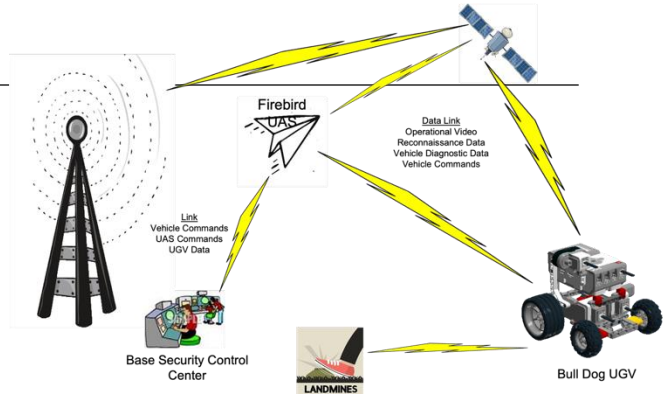
AI Will Break

What is the definition of resilience and what are notions of resilience mechanisms for AI-enabled systems?



Trusted AI Systems Engineering Challenge

Overview



- Teams engage in
  - Assured design of AI and autonomy into notional system
  - Risk-based monitoring and management of operational use of AI capabilities.

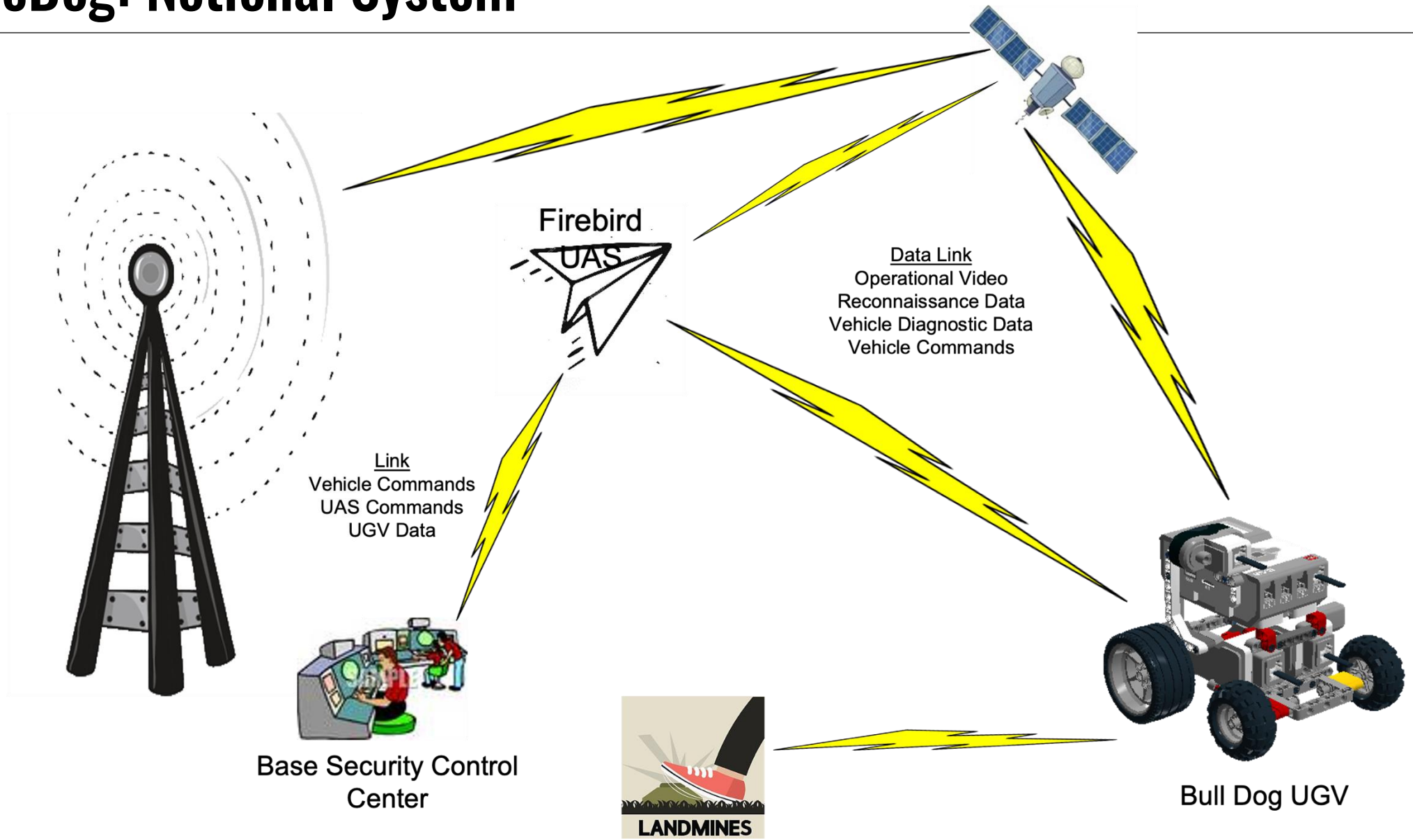
- Semester-long Stages:
  1. Explore performance of AI models over variety of operational scenarios.
  2. Design of the decision system; human-machine teaming, resilience.
  3. Operational simulation of mission scenarios.

- Teams judged on quantitative performance & SE approaches used to design and operate the system.

- Open to all SERC universities + HBCUs and MSIs

- Prizes!

FireDog: Notional System



Stage 1: Spring 2024

- Competitors will be provided with an operational concept and supporting MBSE models for FireDog
- Competitors will be provided with:
  - ML models for central classification problem (mine detection),
  - Input/meta data and ground truth across a variety of environmental conditions.
- Characterize operating envelopes of provided models and developing metrics to characterize confidence in model performance.
- Form initial ideas about relationships between model performance and system performance.
- Submissions will consist of a presentation and white paper describing approach and results. These submissions will be made available to all competitors in Stage 2.

Stage 2: Fall 2024

- Stage 2 centers on design of the decision system.
- Designs might consider
  - Architectures for human & machine decision making
  - Methods to increase operator trust
  - Resilience mechanisms based on estimates of model performance
- Competitors will be provided operational simulation models for FireDog.
- Teams may choose to use AI-based methods for system design, verifications, and validation or to explore modular/open architecture concepts or design for high-level system properties such as safety, security, and trust.
- Submissions will consist of a presentation and white paper describing approach and results. These submissions will be made available to all competitors in Stage 3 for their consideration.
- The top teams may be invited to make conference presentations.

Stage 3: Spring 2025

- Stage 3 centers on operational simulation of mission scenarios.
- Goal is to exercise system in scenarios not previously seen by the participants.
- Scoring will be done on quantitative measures of system performance as well as on the novelty and utility of the SE processes used to generate or validate the design.
- Submissions will consist of an oral presentation, accompanying slides, and final report.
- The top teams may be invited to present at SERC SSRR or the SE4AI/AI4SE workshop.

AI Systems Engineering Challenge: Prizes

- $250K in prizes in the form of SERC research awards

- Prizes by stage:
  1. $75K
  2. $75K
  3. $100K

- Teams judged on quantitative performance & SE approaches used to design and operate the system.

AI Systems Engineering Challenge: Eligibility

Competition open to student/professor teams from:

- SERC universities
- Ohio State University, North Carolina State University, University of Arizona, The George Washington University, and George Mason University
- HBCUs and MSIs

Thank you

Stay connected with SERC Online:

Email the presenter: Peter Beling

beling@vt.edu

Email the research team: VT National Security Institute

beling@vt.edu