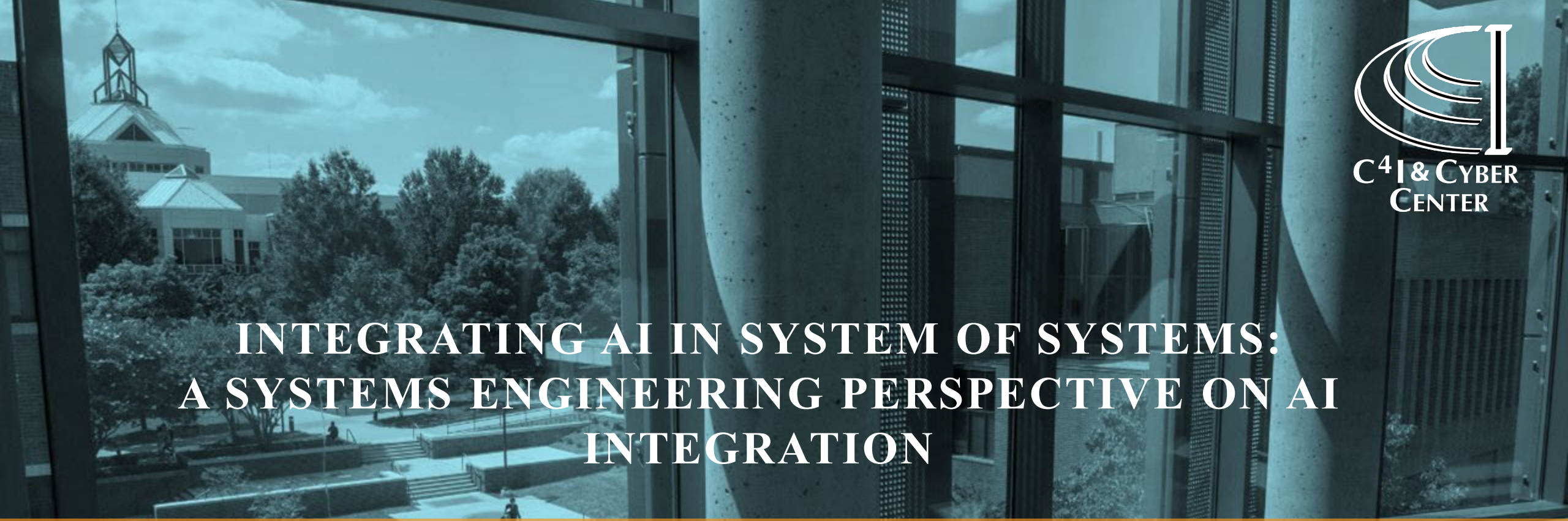


INTEGRATING AI IN SYSTEM OF SYSTEMS: A SYSTEMS ENGINEERING PERSPECTIVE ON AI INTEGRATION

INCOSE New England Chapter Fall Workshop
Worcester Polytechnique Institute (WPI)

Ali K. Raz

Assistant Professor Systems Engineering

Assistant Director of C4I and Cyber Center

Chair INCOSE AI Working Group

Chair AIAA Information, Command, and Control Systems Technical Committee

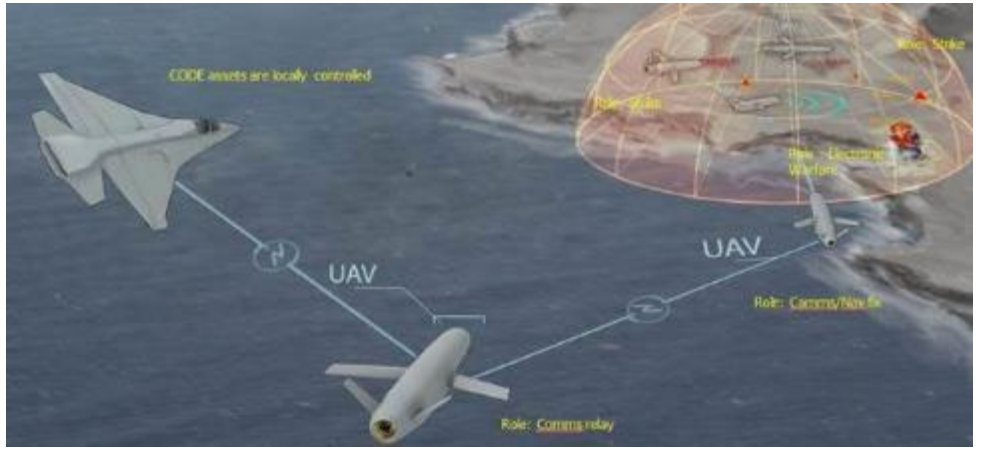
George Mason University

araz@gmu.edu

Research Motivation/Overview

Future operational concepts of complex systems in both civil and defense domains increasingly demand ***integration and interoperability*** of multiple ***intelligent systems***

https://utm.arc.nasa.gov/index.shtml



https://www.darpa.mil/program/collaborative-operations-in-denied-environment

Intelligent Systems require:
Artificial Intelligence and Machine Learning

Integration requires:
Systems Engineering and Systems of Systems Engineering

Interoperability requires:
Human/Machine Interpretability and Fusion of Information

This presentation covers System Engineering for AI/ML….

….and provides a snapshot of ongoing research in SoS/Fusion

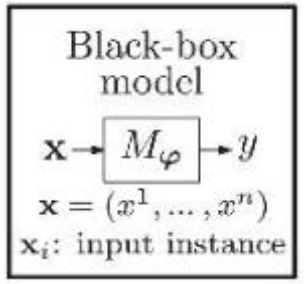
The Need for Systems Engineering of AI

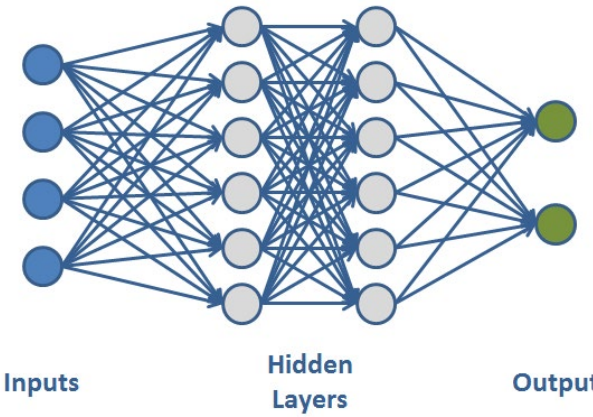
Deep Neural Networks (DNNs) are most common form of AI/ML Implementation

- ✓ State-of-the-art implementation for AI/ML algorithms *(Supervised, Unsupervised, Reinforcement Learning, Natural Language Processing etc.)*
- ✓ Well-established performance outcomes in a variety of applications *(Intuitive and non-intuitive outcomes)*
- ✓ Strong focus on algorithmic development, computational efficiency, and implementation
- ✓ Selective demonstration of test cases, mostly based on training data partitioning in training and validation sets

Common Challenges for DNNs

- ❑ **Trained DNNs are essentially backboxes to the designers and users**
- ❑ Limited characterization of performance bounds due to variations and uncertainties; limited Monte Carlo simulations and user selected variations
- ❑ Limited explanation of black-box decision-making logic
- ❑ Limited evaluation of acceptable and unacceptable performance regions





Systems Engineering Perspective Example SE Questions to Ask

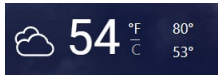
- ❑ What is the impact of variations in input data and environment?
- ❑ How does the input (i.e., observed state) influence DNNs decision making?
- ❑ Does training data considers edge cases?
- ❑ How does the DNNs respond to modeled (i.e., included in training) and unmodeled uncertainties?
- ❑ How does the DNN interact with other system components and external systems?

Why Establish Trust and have Explainable AI?

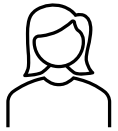
Let's try a thought experiment

Q: What will be the weather tomorrow?

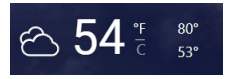
It will be chilly and cloudy.
Remember to put on a warm jacket



Q: How do you know what will be the weather tomorrow?

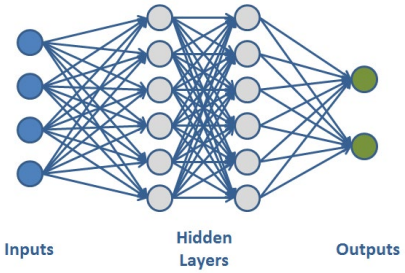


- I heard it on the radio
- I looked up on my phone
- Weather radar showed a cold front
- I love looking at NOAA models, you want to discuss variations in the barometric pressure!



Can we be okay with lack of trust and explainability?

Datasets/Models/ Rewards



Artificial Intelligence (AI) Models

- Create new materials
- Create new drugs
- Predict person's health/weight
- Predict a terrorist
- Reject loans

Example AI Uses

1. Why this action?
2. Why not another action?
3. When do I succeed/fail?
4. When can I trust the results?
5. How can I fix an error?

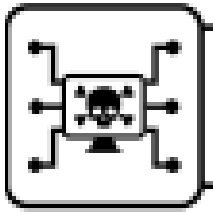
End User

Unsolved Problems in ML Safety*

Robustness Create models that are resilient to adversaries, unusual situations, and Black Swan events.

Monitoring Detect malicious use, monitor predictions, and discover unexpected model functionality.

Alignment Build models that represent and safely optimize hard-to-specify human values.



Systemic Safety Use ML to address broader risks to how ML systems are handled, such as cyberattacks.

Hendrycks, Dan, Nicholas Carlini, John Schulman, and Jacob Steinhardt. "Unsolved problems in ml safety." *arXiv preprint arXiv:2109.13916* (2021)

Systems Engineering for AI/ML

- Examine, develop, and build AI with the SE principles, concepts, and tools
  - SE life-cycle stages place particular emphasis on the unsolved problems for AI/ML at the outset

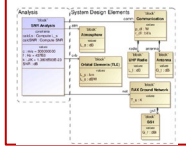
Alignment

SE Principles for Alignment

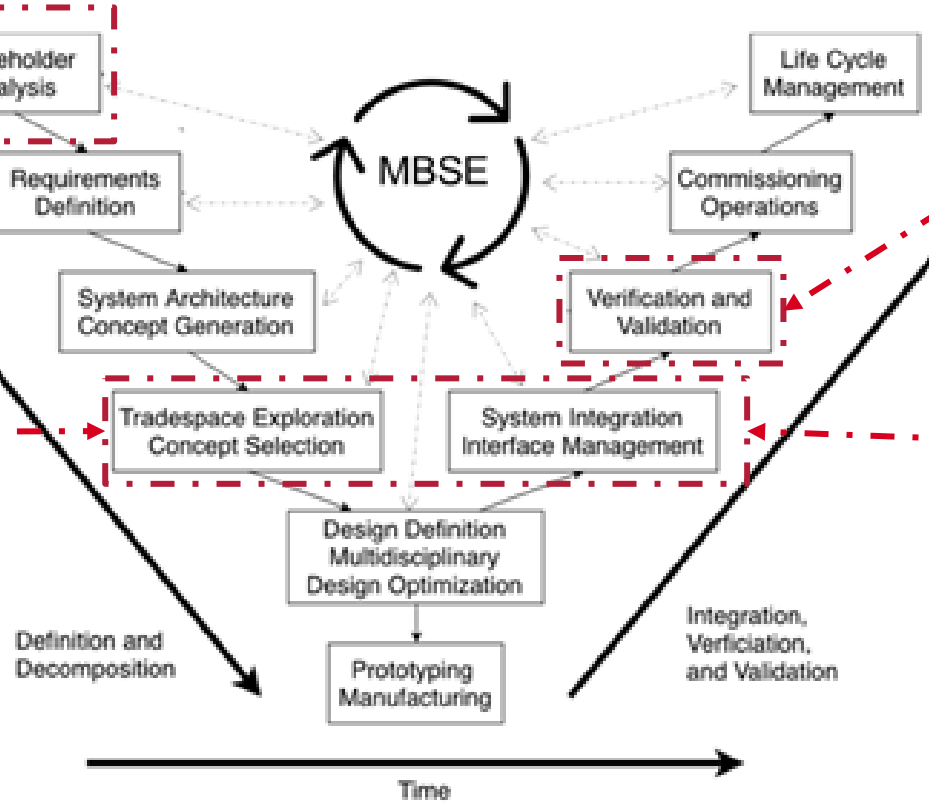
- Stakeholder Analysis including Tension
- Use Cases

Monitoring

SE Principles for Monitoring



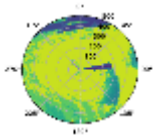
- Functional Interactional Analysis
- Physical/Functional Trade space



Vee Life-Cycle Model*
(MBSE – Model based Systems Engineering)

Robustness

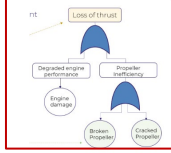
SE Principles for Robustness



- Detailed test cases, incl failures
- V&V Design and Planning

Systemic Safety

SE Principles for Syst. Safety



- Failure Mode Effect Analysis
- Probabilistic Hazard Analysis

Raz, Ali K., et al. "Explainable AI and counterfactuals for test and evaluation of intelligent engineered systems." *INCOSE International Symposium*. Vol. 33. No. 1. 2023.
de Weck, O. L. Vee-Model. MIT OCW. https://ocw.mit.edu/courses/aeronautics-and-astronautics/16-842-fundamentals-of-systems-engineering-fall-2015/lecture-notes/MIT16_842F15_Ses1SE_Ovr_vw.pdf.

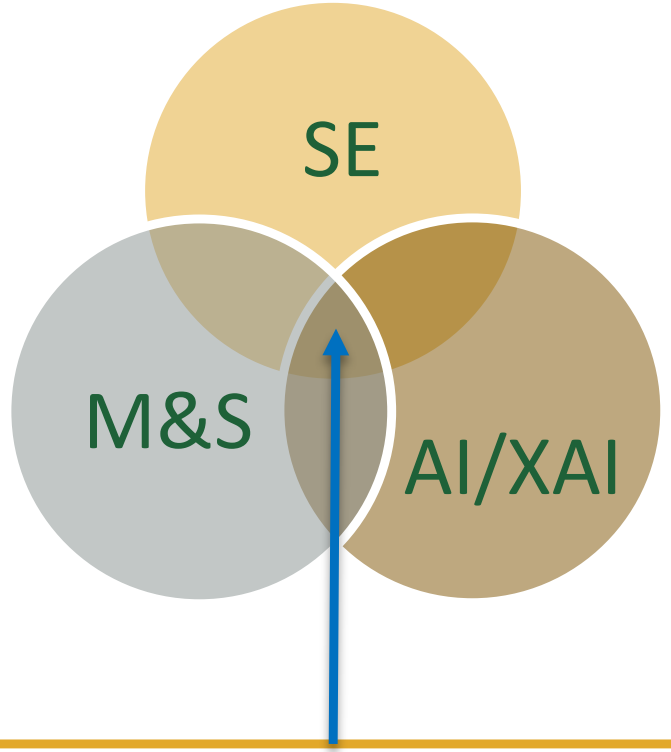
From AI/ML Algorithms to Systems Implementation

Systems Engineering

Complex systems are built on SE foundations

Modeling and Simulation

Use models as a basis for simulations to develop data for analysis, education, decision making, *etc*.



Artificial Intelligence with Explainability

AI built with SE principles in which the solution (including results, failure modes, rationale, justification, assumptions, *etc.*) can be interpreted and understood by humans.

Provide a window of opportunity to understand, analyze and validate the assumptions, theories, operations, and decision-making constructs of modern complex systems with embedded AI/ML-components.

Example: High Speed Aerospace Application with SE for AI

Problem Formulation:

- Provide guidance commands to high-speed aerospace vehicles

Why AI is Needed:

- Real-time trajectory generation is computationally prohibitive for high-speed missions
- Human intervention and guidance is not feasible beyond supersonic speeds

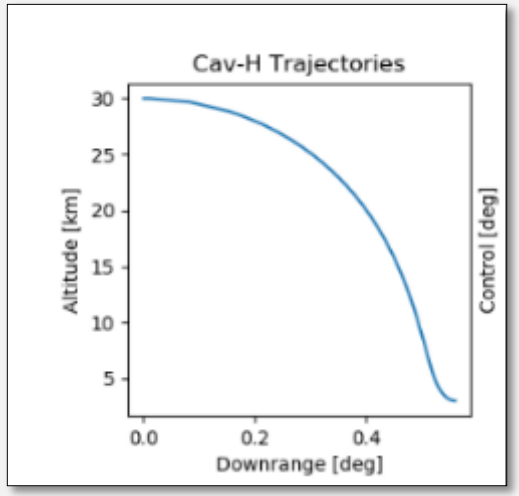
What AI Techniques can be used:

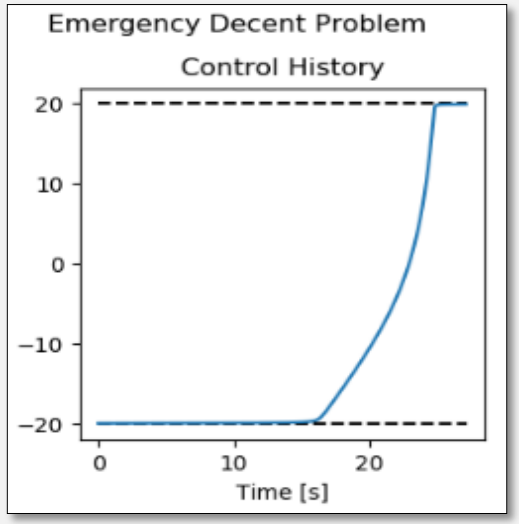
- Reinforcement Learning (RL) enables training of an artificial intelligence (AI) agent to operate in dynamic uncertain environments
- Impressive performance outcomes to learn nearly-optimal solutions in a variety of application domains

How SE can help:

- Provide Robustness examination for RL-based solutions
- Pair Explainable AI with System Modeling and Simulation to develop an understanding and logic behind RL solutions space
- Provide a mechanism to compare RL implementation to analytic solutions

Sample Problem



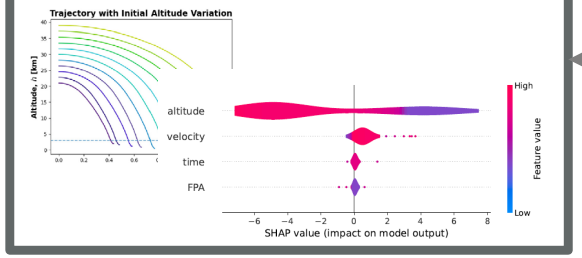


Emergency descent of vehicle from 30km altitude to 3km altitude in minimum time

Framework for Implementing RL in Aerospace Applications

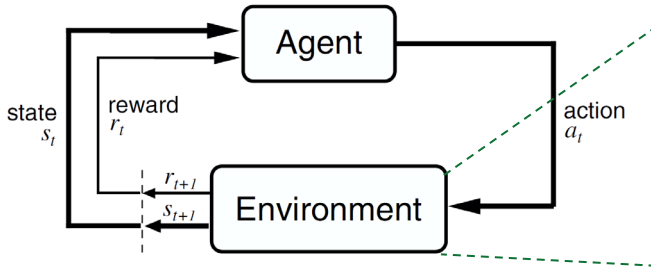
- **AI/RL Requirements:**
  - A model of the mission, the vehicle, and an interface that exposes a reward assessment
  - A supplied set of actions which correspond to state change of the vehicle
  - A method for understanding and validating agent behavior

Systems Engineering Test and Evaluation of RL
(Tool Set: SE/XAI)

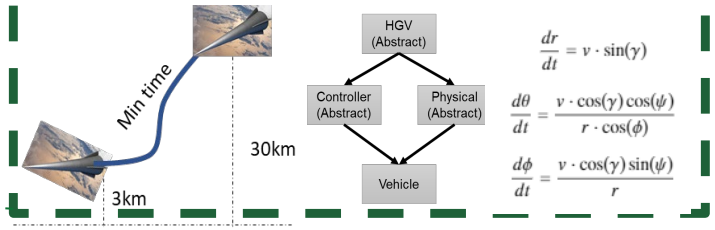


Test and Evaluation of RL

Reinforcement Learning
(Tool Set: Python/Keras)

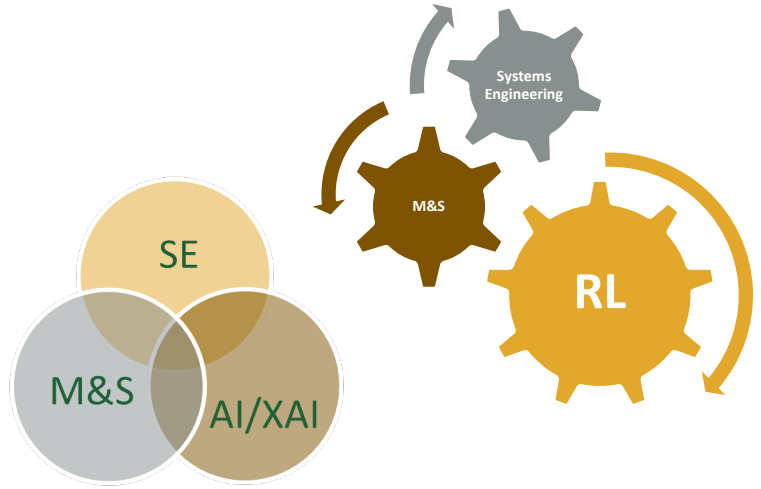


Agent-based Aerospace Vehicle Dynamics and Simulation Model
(Tool Set: Discrete Agent Framework – MATLAB/Python)



- **Modeling and Simulation:**
  - Provides agent-based abstraction of Equations of Motion and Vehicle Constraints
  - Provides agent-based abstraction of models and assessment of rewards
  - Provide atmospheric models

- **Systems Engineering based Test and Evaluation of RL**
  - Robustness Testing of RL
  - Explainable AI for RL
  - Validating Optimal Behavior



Raz, Ali K., et al. "Test and evaluation of reinforcement learning via robustness testing and explainable ai for high-speed aerospace vehicles." *2022 IEEE Aerospace Conference (AERO)*. IEEE, 2022.

Proof-of-Concept High-Speed Aerospace Mission

Vehicle Model Parameters

- **States:**
  h: altitude, θ: downrange angle,
  v: velocity, γ: flight path angle
- **Control:** α: angle of attack
- **Dynamics:**

$$\dot{x} = \begin{bmatrix} \dot{h} \\ \dot{\theta} \\ \dot{v} \\ \dot{\gamma} \end{bmatrix} = \begin{bmatrix} v \sin\gamma \\ \frac{v}{r}\cos\gamma \\ -\frac{D(\alpha)}{m} - \frac{\mu}{r^2}\sin\gamma \\ \frac{L(\alpha)}{mv} - \left(\frac{v}{r} - \frac{\mu}{vr^2}\right)\cos\gamma \end{bmatrix}$$

- **Objective:** $J = \min t_f = \int_0^{t_f} dt$
- **Initial Constraints:**

$$\Psi_0 = 0 = \begin{bmatrix} h & -30\ \text{km} \\ \theta & \\ v & -3\ \text{km/s} \\ \gamma & \end{bmatrix}_{t=t_0}$$

- **Path Constraint:**

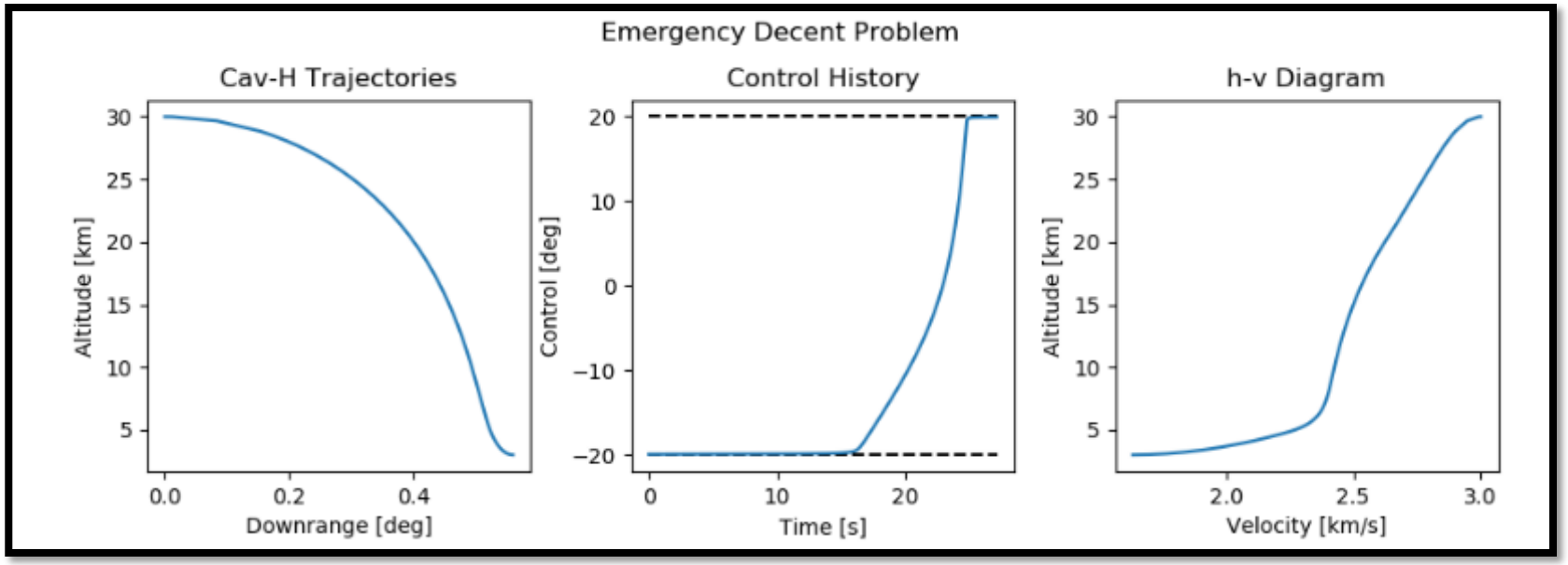
$$|\alpha| \leq 20°$$

- **Terminal Constraints:**

$$\Psi_f = 0 = \begin{bmatrix} h & -3\ \text{km} \\ \gamma & \end{bmatrix}_{t=t_f}$$

Emergency Descent Problem for High-Speed Glide Vehicle

- The vehicle at 30 km altitude and 3 km/s velocity needs to descend to level flight at a safe altitude of 3 km in minimum time
- Constraints must be satisfied



Raz, Ali K., et al. "Test and evaluation of reinforcement learning via robustness testing and explainable ai for high-speed aerospace vehicles." *2022 IEEE Aerospace Conference (AERO)*. IEEE, 2022.

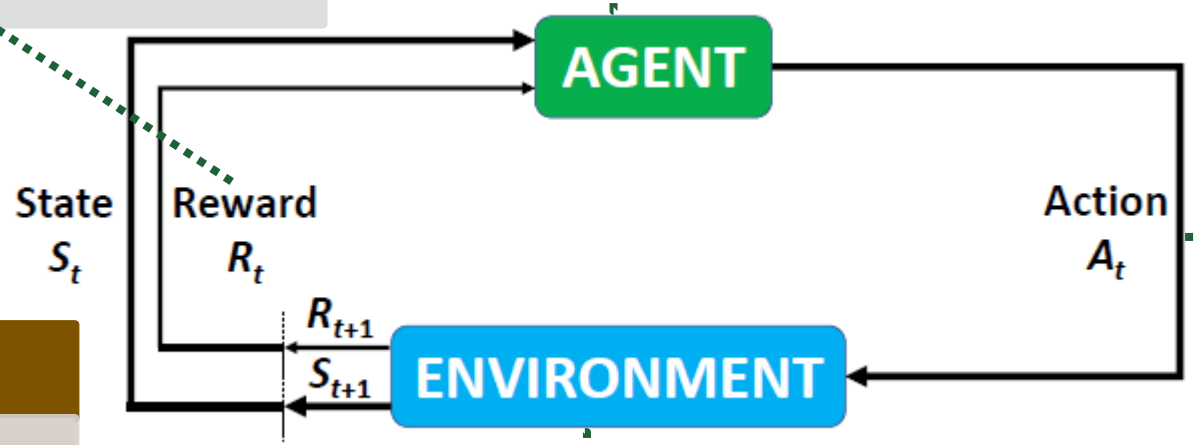
Reinforcement Learning Problem Formulation

Reward Function

- Designed to train the RL agent an emergency descent problem
- Reward structure based on distance to target and FPA

RL Agent

- RL trained from SB3 Python package
- Proximal Policy Optimization (PPO)
- RL training parameters (backup)



Action Space

- AoA command
- ±20º in variable increments of 2º

State

- Distance to target
- Altitude
- Velocity
- AoA
- FPA

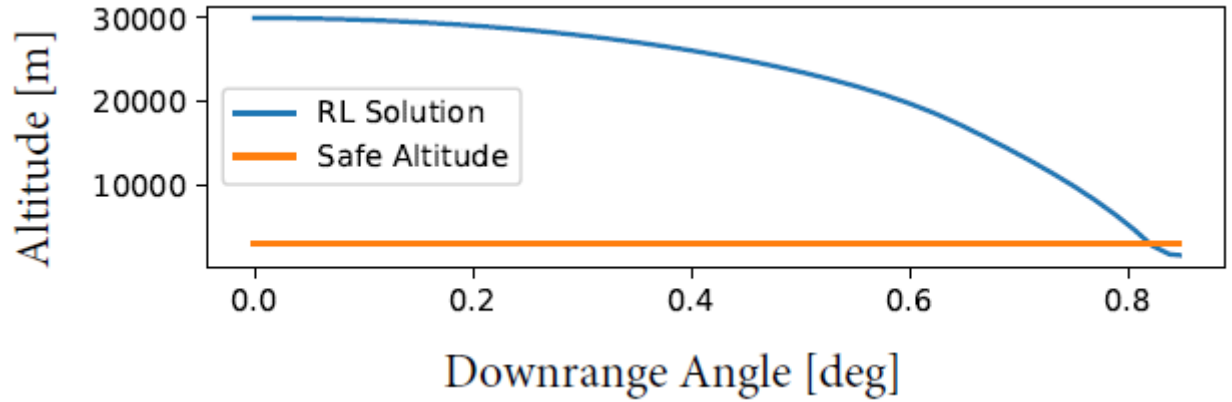
Environment

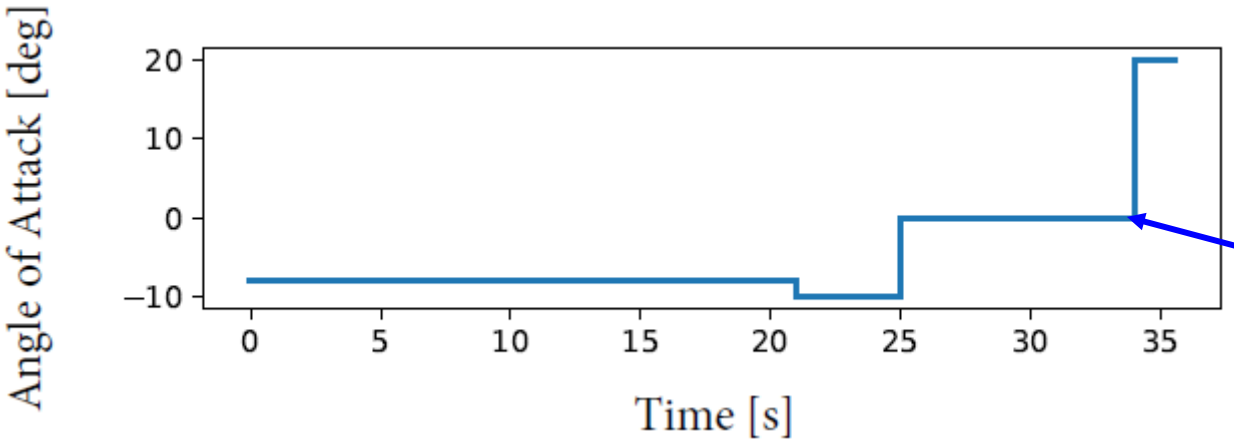
- Emergency descent problem space
- Atmosphere, simulation clock, scheduler, etc.

RL Results: Nominal Case (Vehicle Descent From 30 km to 3 km)

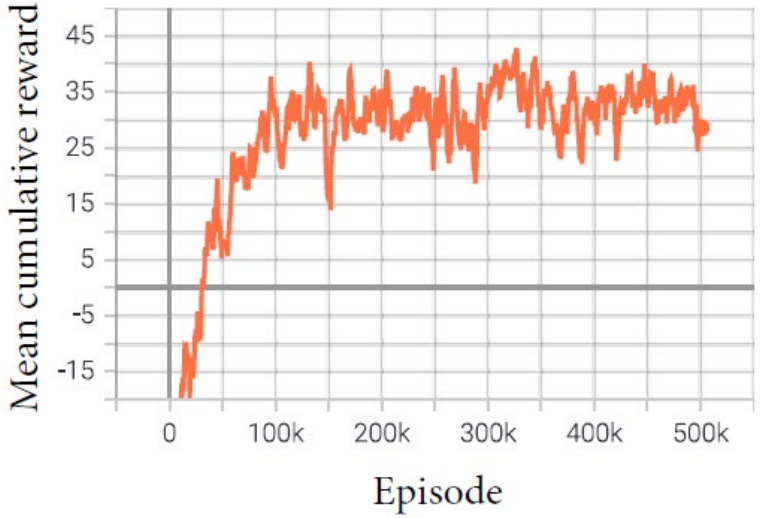
RL training:

- Provides AoA commands to guide the vehicle to a pre-determined safe altitude
- Included randomly sampling vehicle initial conditions
- Completed after 500k episodes





Sufficient cumulative reward of +30 to train policy



AoA commands issued by the RL agent

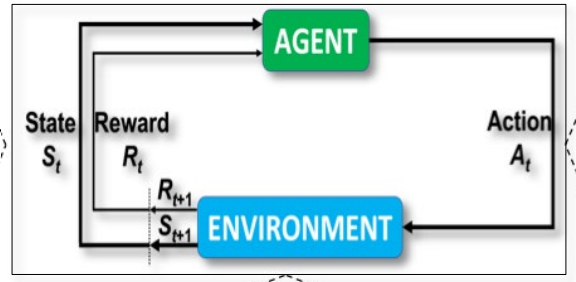
Three Part Test and Evaluation Framework for RL

Robustness Testing

Purpose:
Sensitivity analysis of variations in action space, enviorment, and state observation

Methodology:
Design of Experiments and Statistical Analysis

Value:
Performance bounds and characterization of uncertanities



Compare to Known Solutions

Purpose:
Evaluate RL performance to known and accepted solutions

Methodology:
Problem space dependent; closed form mathematical solutions.

Value:
Validate RL performance and robustness testing results

Explainable AI (XAI)

Purpose:
Determine influential features of trained RL decision-making logic

Methodology:
Post-hoc XAI method: Shapely Additive Explanations

Value:
Explain which state vector values contribute to RL decision and why sensitivities are present in robustness test

Raz, Ali K., et al. "Explainable AI and Robustness based Test and Evaluation of Reinforcement Learning." *IEEE Transactions on Aerospace and Electronic Systems* (2024).

Robustness Testing of RL Solutions

Purpose: Identify sources of variation in RL problem space and quantify the impact of variation on RL performance

General Sources of Variations in RL

Source	Nature of Variation	Modeling Approach for Robustness Testing
Environment	Initial Conditions	• Latin Hypercube Sampling • Monte Carlo Simulations • Design of Experiments
Action Space/ State Space	Tolerance and Sensitivity	Expected probability distribution with parameters
	Impulses and Hard Overs	Expected magnitude and time duration

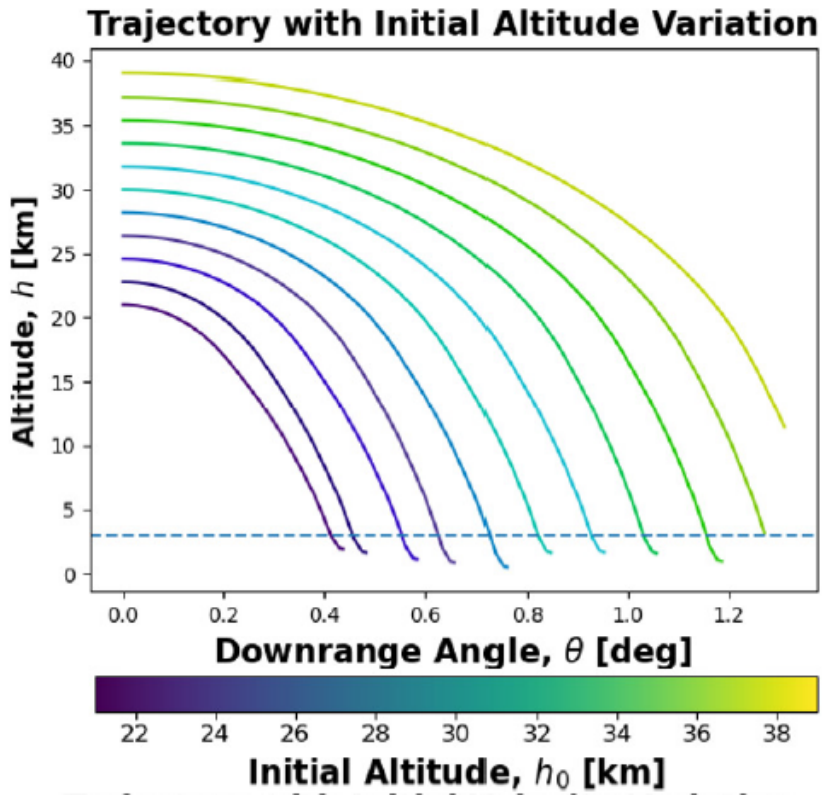
Derived Test Cases for High-Speed Vehicle RL Solution

Test Cases	Objective
TC-1	Individually vary environment Initial Conditions (ICs) (i.e., altitude, velocity, and FPA) to examine RL performance
TC-2	Quantify performance bounds on ICs variations with Latin Hypercube Sampling
TC-3	Sensitivity to impulses on the action space
TC-4	Sensitivity to random variations in the action space
TC-5	Sensitivity to impulses on the state space
TC-6	Sensitivity to random variations in the state space

Robustness Testing Results (TC-1 and 2)

TC-1 Modeling Approach:

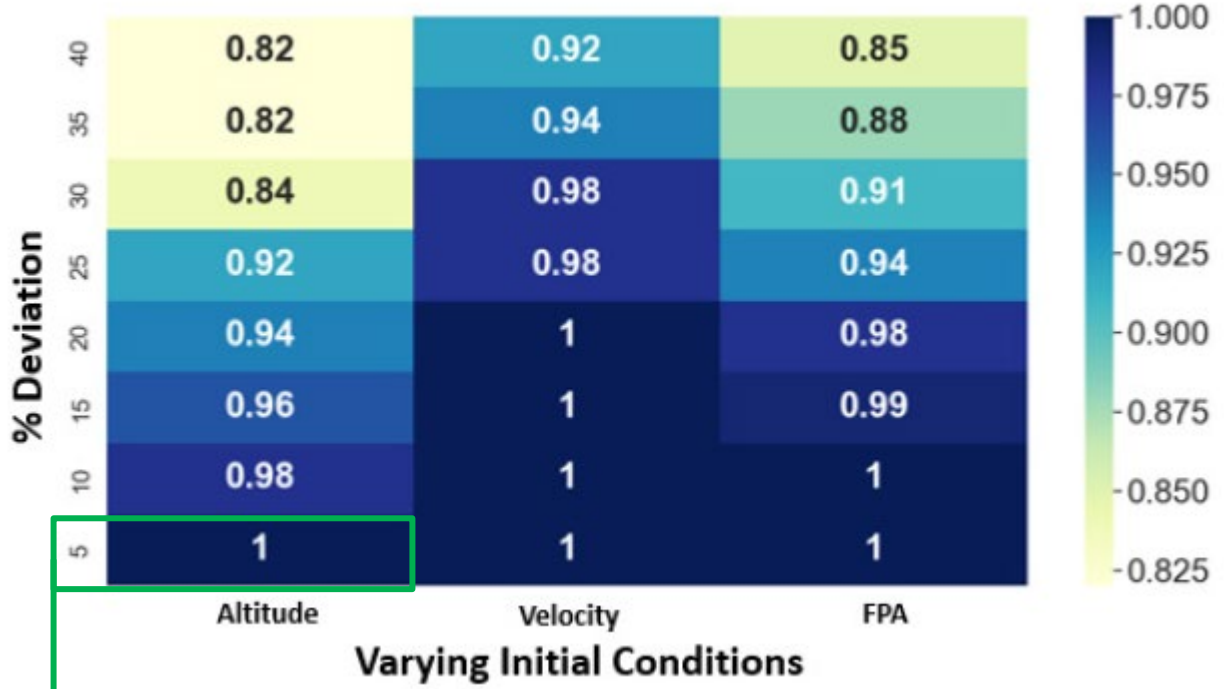
- Train RL agent by randomly sampling ICs with pre-defined range
- Results shown for 30% variations



Safe target altitude not reached from higher altitudes

TC-2 Modeling Approach:

- Utilize Latin Hypercube Sample to generate IC samples outside training bounds
- Results show fractions of successful trajectories per 50 samples



100% success only within 5% of altitude variation

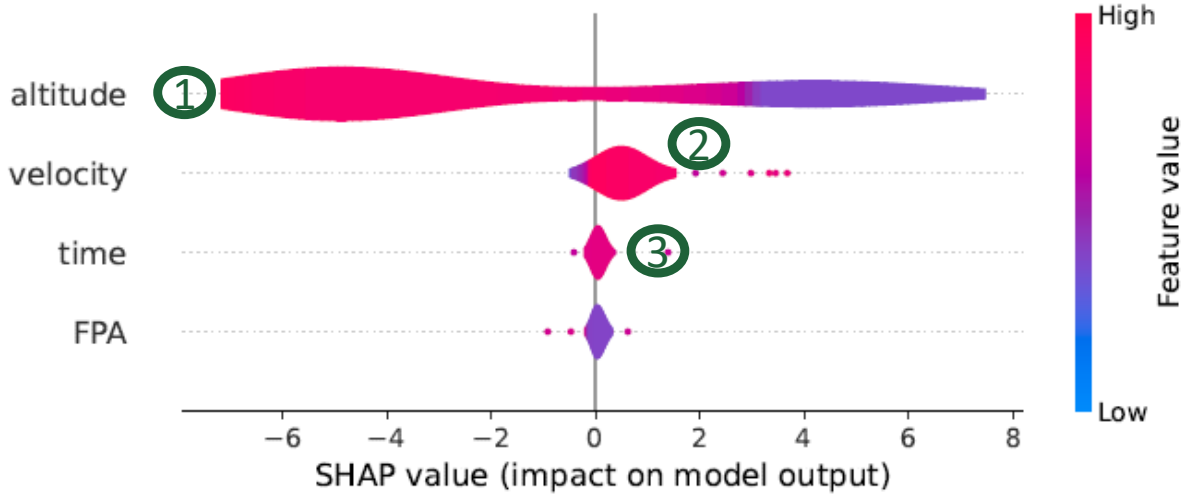
Examination Via Explainable AI (XAI) Techniques

Brief Introduction to Explainable AI

- Investigates trained Deep Neural Network (DNN) models with analytical techniques to extract decision making attributes
- SHapley Additive exPlanations (SHAP)
  - State of the art for reverse engineering the output of any predictive model
  - Yields importance of input features for a given prediction
  - Focuses on coalitions in cooperative game theory

SHAP Applied to RL Problem

- **Inputs:** Time, Altitude, Velocity, and Flight Path Angle
- **Output:** Angle of Attack (between -20˚ and 20˚)
- **Number of Trajectories:** 1000
- **Objective:** Reach a particular target in a minimum time

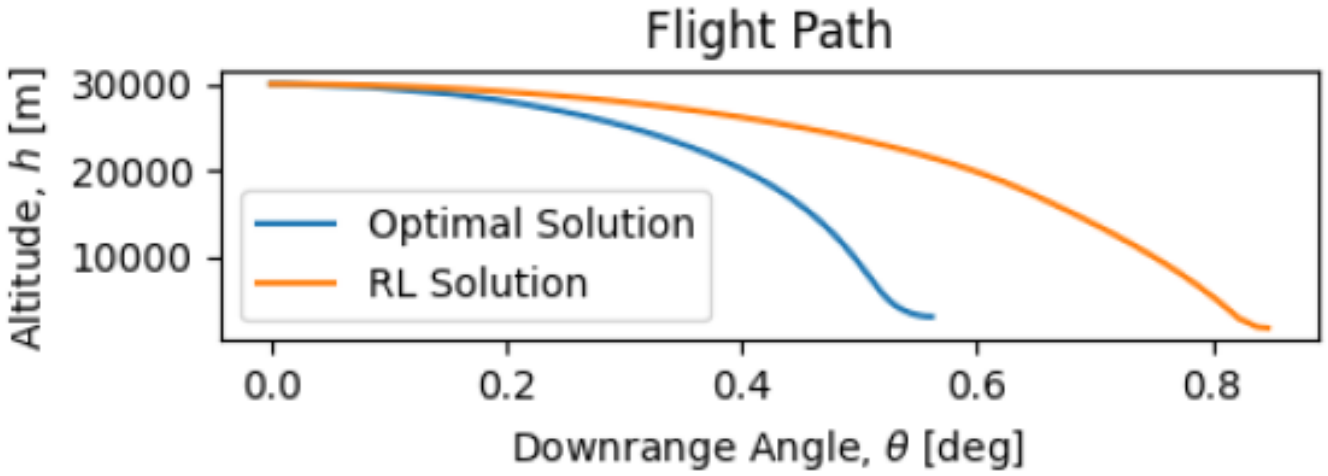


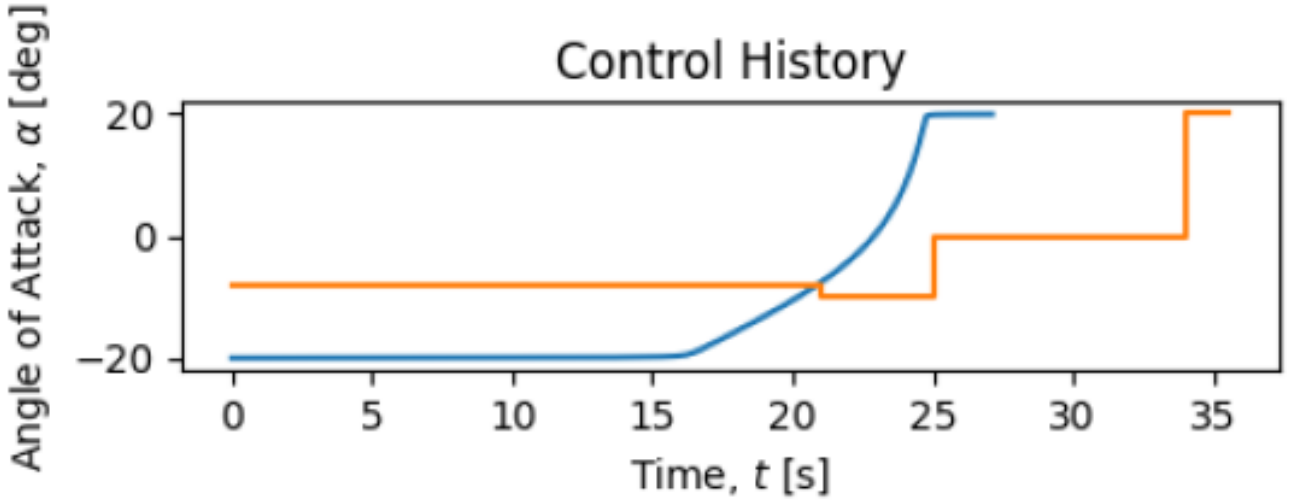
① Higher altitude values oppose a change in AoA whereas lower altitudes support it.

② Higher velocity values positively influence change in AoA

③ FPA and Time have least impact.

Validation With Optimal Control Solutions





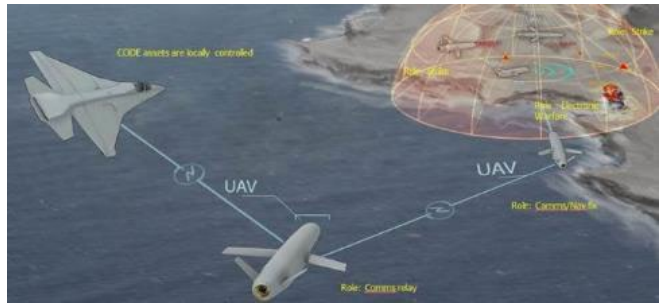
- RL Results with PPO algorithm
  - Training Options:
    - Varying initial conditions
    - 500K Episodes
- Optimal control solved by indirect methods

- RL agent approximates optimal solution
- Potential differences due to:
  - RL solution is discrete action space
  - OP solution is continuous action space
  - **Goal is not to exactly reproduce the optimal trajectory**

Systems Engineering of AI is needed to help address Unsolved problems in AI and transition AI into practical systems

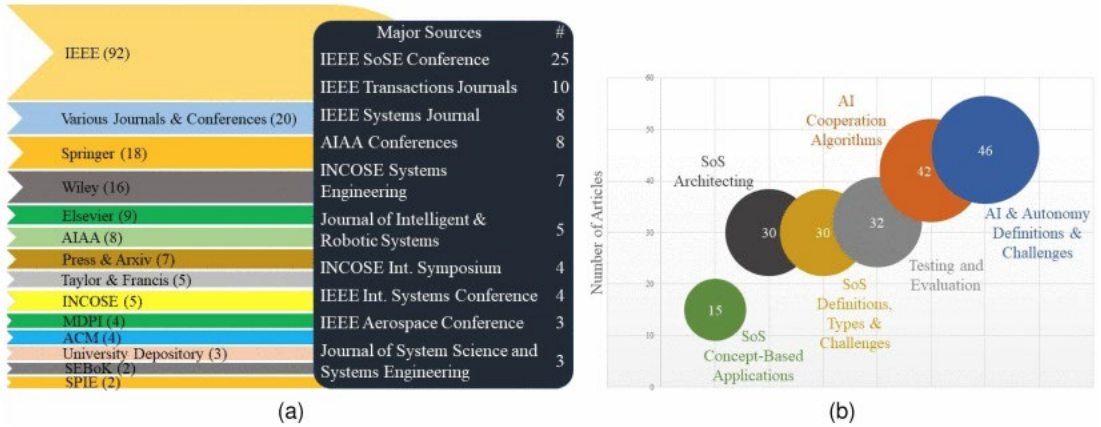
- Robustness Testing and Explainable AI expose DNNs decision-making and limitations
  - Characterize performance envelopes of the system; emergent behavior, and failure regions
  - Make DNNs interpretable to Subject Matter Experts

- However, further research is needed to:
  - Holistically explore and develop SE artifacts for AI/ML components
  - Help transition AI/ML in practical systems
  - **Build System of Systems with AI/ML components**



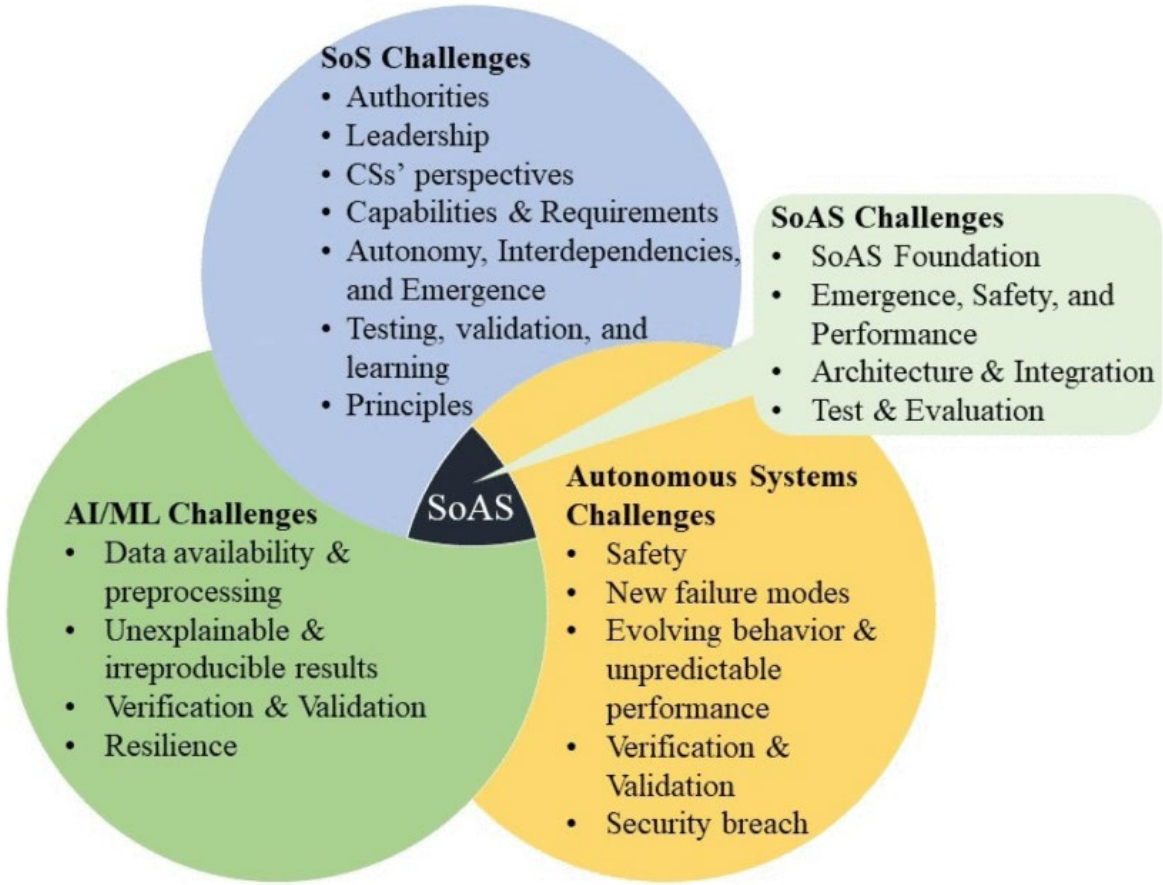
Snapshot of Ongoing Research At Mason (1)

Integrating Autonomy in System of Systems: Towards a System of Autonomous Systems



Performed a comprehensive literature survey to identify challenges of integrating autnomoy in System of System (~198 papers) from:

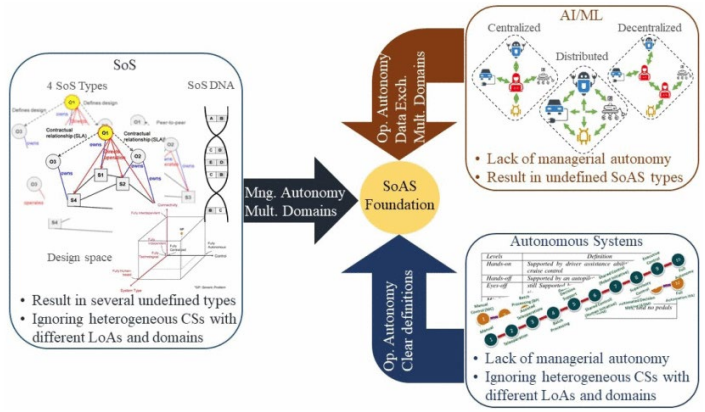
- AI/ML Literature
- Autnomous Systems Literature
- System of Systems Literature

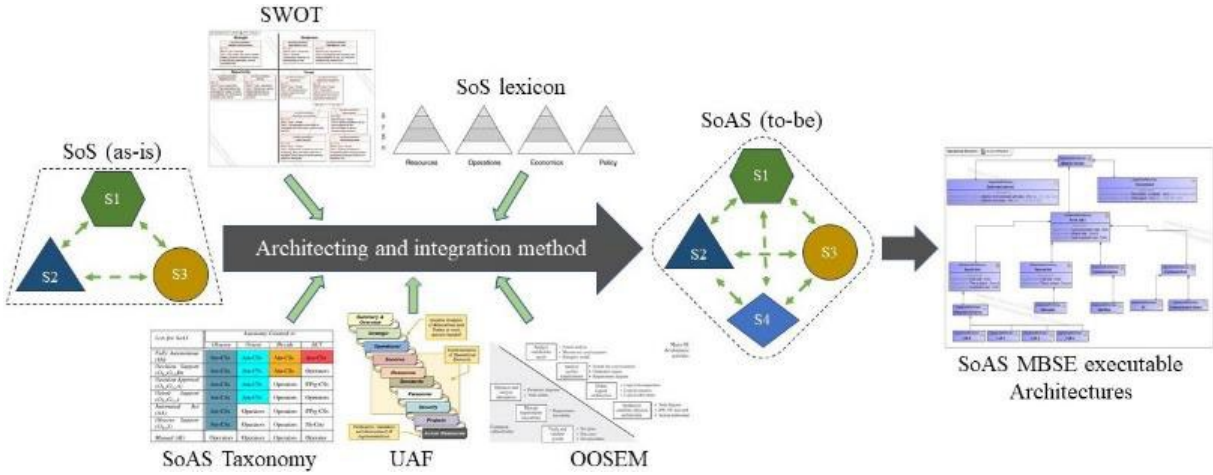


Torkjazi, Mohammadreza, and **Ali K. Raz**. "A Review on Integrating Autonomy into System of Systems: Challenges and Research Directions." *IEEE Open Journal of Systems Engineering* (2024).

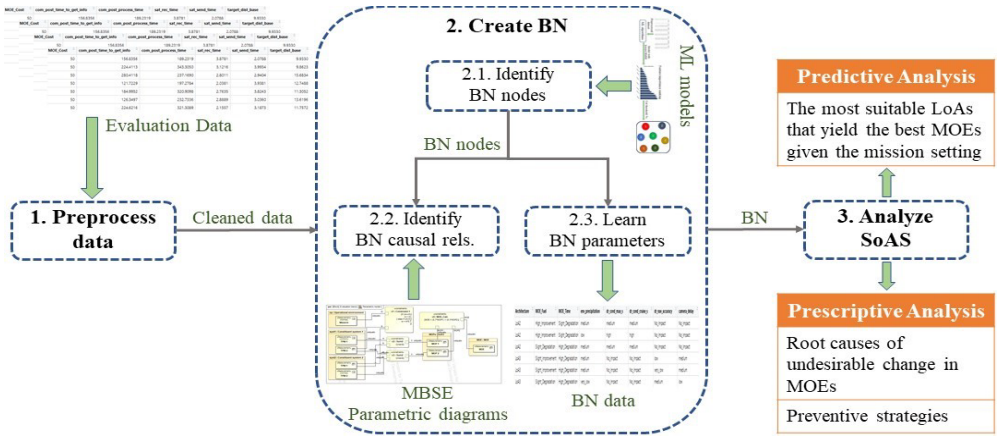
Snapshot of Ongoing Research At Mason (2)

Integrating Autonomy in System of Systems: Towards a System of Autonomous Systems





Currently building MBSE frameworks for Architecture, Integration, Test, and Evaluation of System of Autnomous Systems:



Key Research Questions:

1) How to identify the right **Level of Autonomy** in System of Autnomous Systems?
2) How to characterize **emergent behaviors** with different Level of Autnomoy?

Torkjazi, M., & **Raz, A. K**. (2024b). Model-Based Systems Engineering (MBSE) Methodology for Integrating Autonomy into a System of Systems Using the Unified Architecture Framework. *INCOSE International Symposium*, 34, 1051–1070. https://doi.org/10.1002/iis2.13195

Torkjazi, M., & **Raz, A. K**. (2024c). Predictive and Prescriptive Analyses of Autonomy Integration into the System of Systems. In A. Salado, R. Valerdi, R. Steiner, & L. Head (Eds.), *The Proceedings of the 2024 Conference on Systems Engineering Research* (pp. 213–228). https://doi.org/10.1007/978-3-031-62554-1_14

Snapshot of Ongoing Research At Mason (3)

Mission Engineering (Sponsor: DARPA)

Develop an integrated analytical process for mission engineering

Mission Engineering Design Space: Concept of operations requiring distributed complex adaptive systems and system of systems

Ontological Foundations

Provide a lexicon, taxonomy, and grammar for MMO

Formal Semantic Representation via Domain Specific Language

Common representation across a multidisciplinary ecosystem

Mathematical Formulation

Develop logically consistent, scalable, and mathematical representation

$$\mathcal{MW}: \mathcal{MN} \rightarrow \Omega_{MEDS}$$

$$\mathcal{MN}_{\mathcal{M}_i} = \{MP_{1\mathcal{M}_i}, MP_{2\mathcal{M}_i}, ..., MP_{l\mathcal{M}_i}\}$$

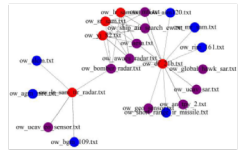
$$\forall \mathcal{M}_i \in \Omega_{MEDS} \exists \mathcal{MN}_i \in \mathcal{MN} | \mathcal{M}_i = \mathcal{MW}$$

Set-based Design + Allen's Interval Algebra

Machine interpretability of MEDS

Analytical Formalism

Provide an analytical reasoning and analysis engine for capability engineering



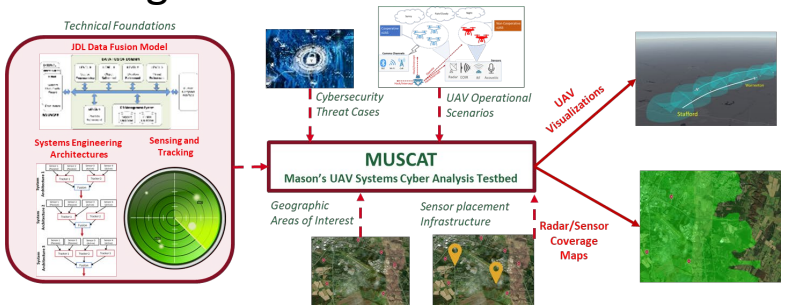
Graph Theory
Multi-Attribute Utility Theory

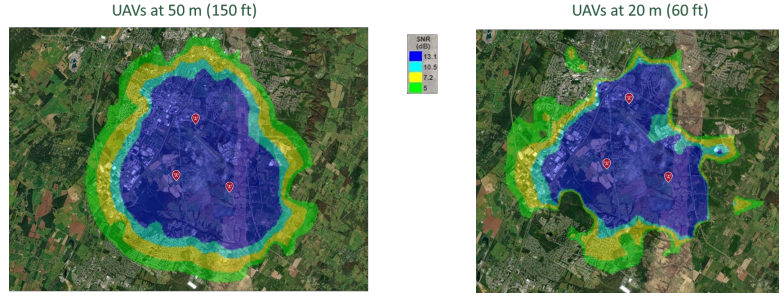
Metrics and methods for large interdependent design spaces

Flexible, dynamic, and feasible System of Systems configurations for meeting mission objectives

MUSCAT (Sponsor: AFCENT)

A system of systems modeling and simulation testbed for sensing, tracking, and data fusion of sUAVs to guide infrastructure-level decision making





Raz, Ali K., et al. "Conceptual, Mathematical, and Analytical Foundations for Mission Engineering and System of Systems Analysis." *IEEE Systems Journal* (2024).

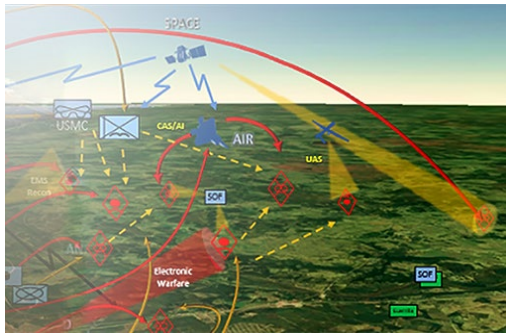
Raz, Ali K., and Marco Gamarra. "Mission Engineering: Analysis of Mission Threads with System of Systems Interdependence." *2024 IEEE International Systems Conference (SysCon)*. IEEE, 2024.

de Albuquerque, P. K., Ferrari, J. F., Hieb, M., Costa, P., Sherry, L., & **Raz, A**. (2023, October). Multi-Sensor Placement and Information Fusion Analysis to Enable Beyond Visual Line of Sight Operations for Small Uncrewed Aerial Vehicles. In *2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC)* (pp. 1-8). IEEE.

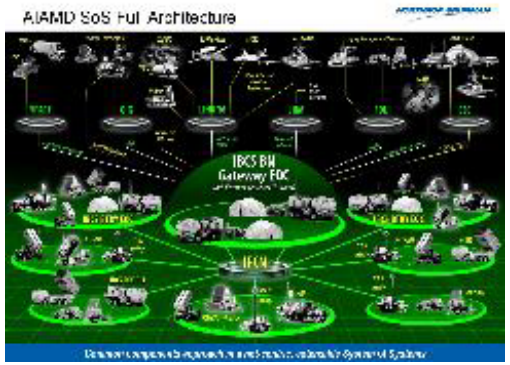
Future Vision and Concluding Remarks

Future collaborative autonomous systems and operational concepts across multiple domain need:

- ❑ Not just development of AI but transition of AI into practical systems
- ❑ Not just advanced sophisticated systems but interoperable systems
- ❑ Not just static system-of-systems but system-of-systems that can be dynamically configured for multiple missions



Multi-Domain C2 / JADC2



Army IBCS

Human Machine Integrated Formations with Trusted Autnomoy

Our research goal is to develop system-of-systems digital models, analytical methods, and systems engineering processes to enable collaborative autonomous systems

Thank You!

Dr. Ali K. Raz

Assistant Professor Systems Engineering

Assistant Director of C4I and Cyber Center

Chair INCOSE AI Working Group

Chair AIAA Information, Command, and Control Systems Technical Committee

George Mason University

araz@gmu.edu