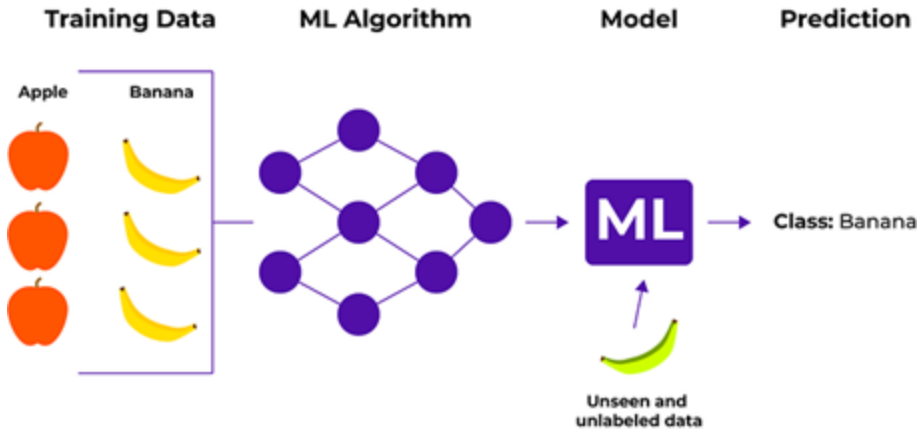
Reinforcement Learning Qualification Process (RLQP):
A Framework for Evaluating Safety and Robustness in Reinforcement Learning

Steven Senczyszyn	MTU
Tim Havens	MTU
Jason Summers	ARiA
Benjamin Werner	DEVCOM AC
Benjamin Schumeg	DEVCOM AC

9/17/2025

What is Reinforcement Learning?

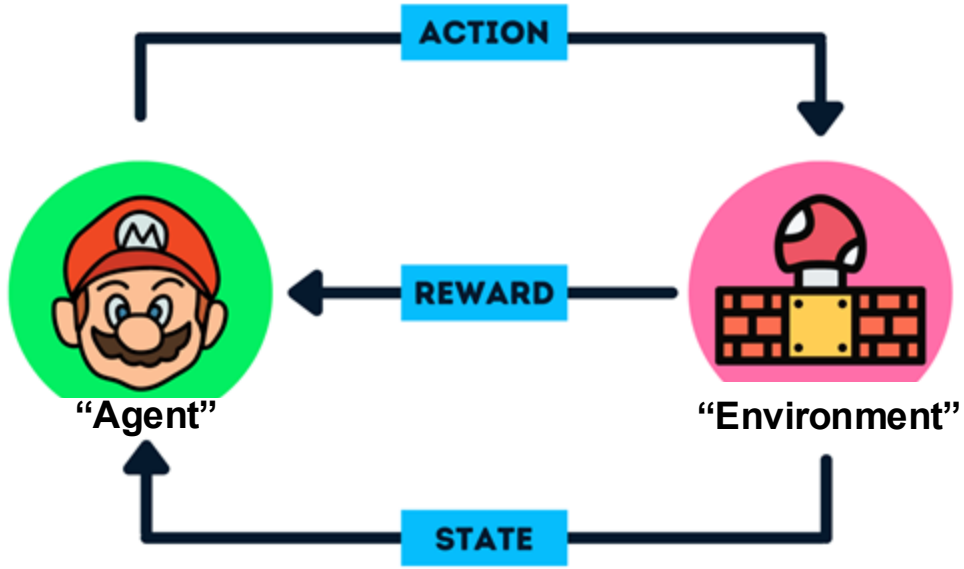
Supervised Learning



"Train and Test"

https://medium.com/ai-for-product-people/what-is-supervised-learning-fa8e2276893e

Reinforcement Learning



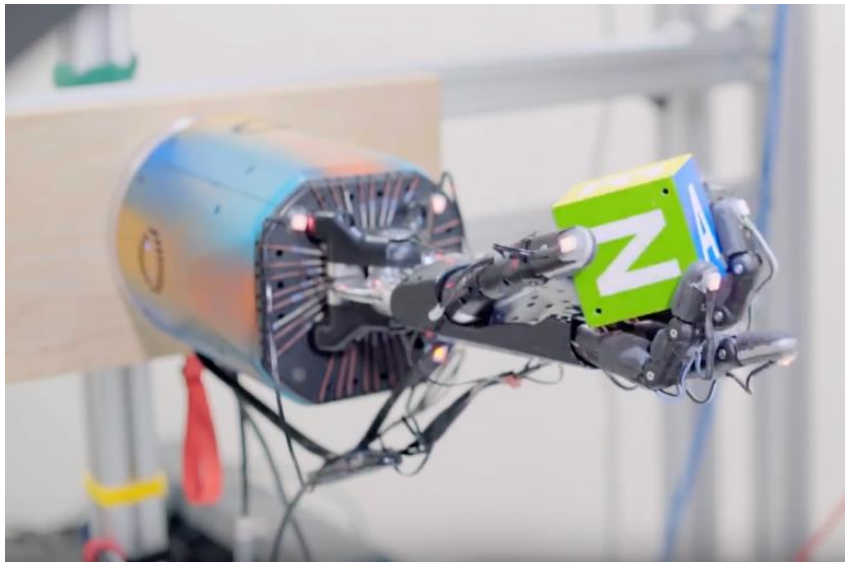
"Trial and Error"

https://www.kdnuggets.com/2022/05/reinforcement-learning-newbies.html

RL Can Learn Complicated Tasks

Competitive Drone Racing

https://www.science.org/doi/abs/10.1126/scirobotics.adg1462

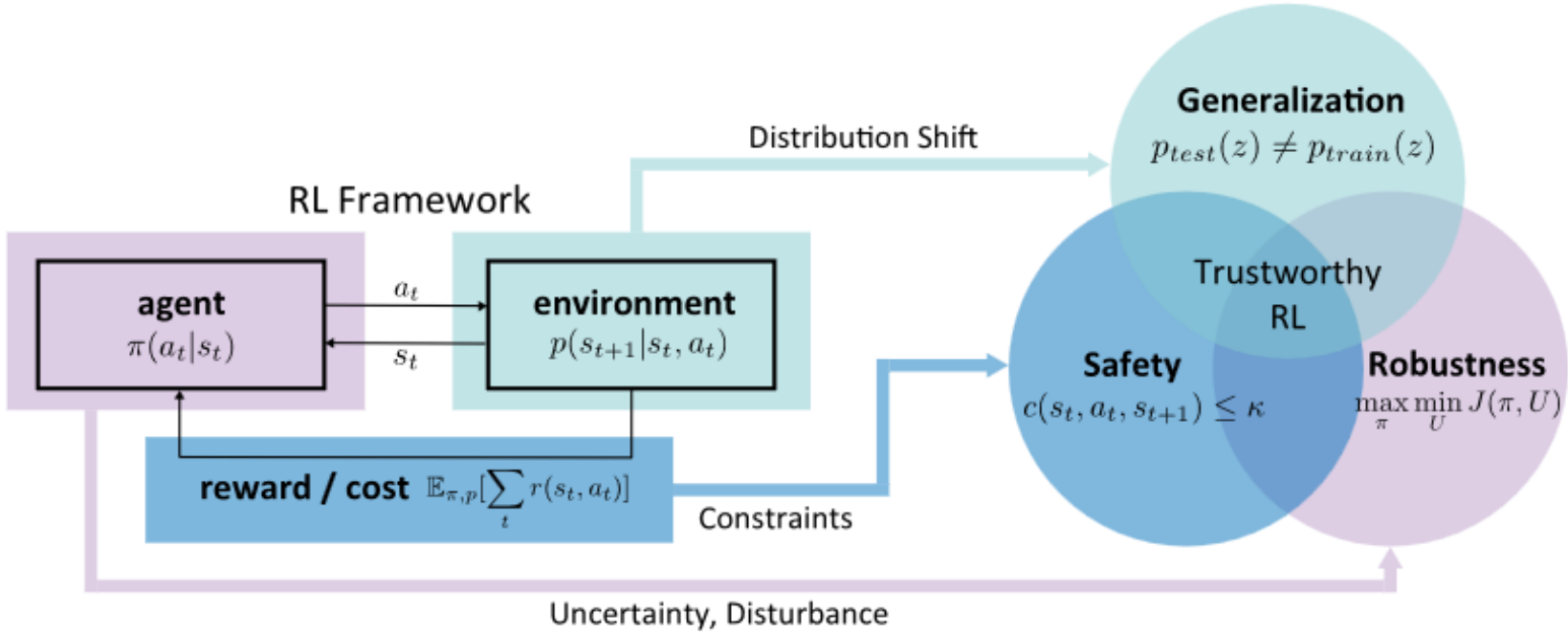


Complex Robotic Manipulation

https://openai.com/index/learning-dexterity/

Michigan Tech

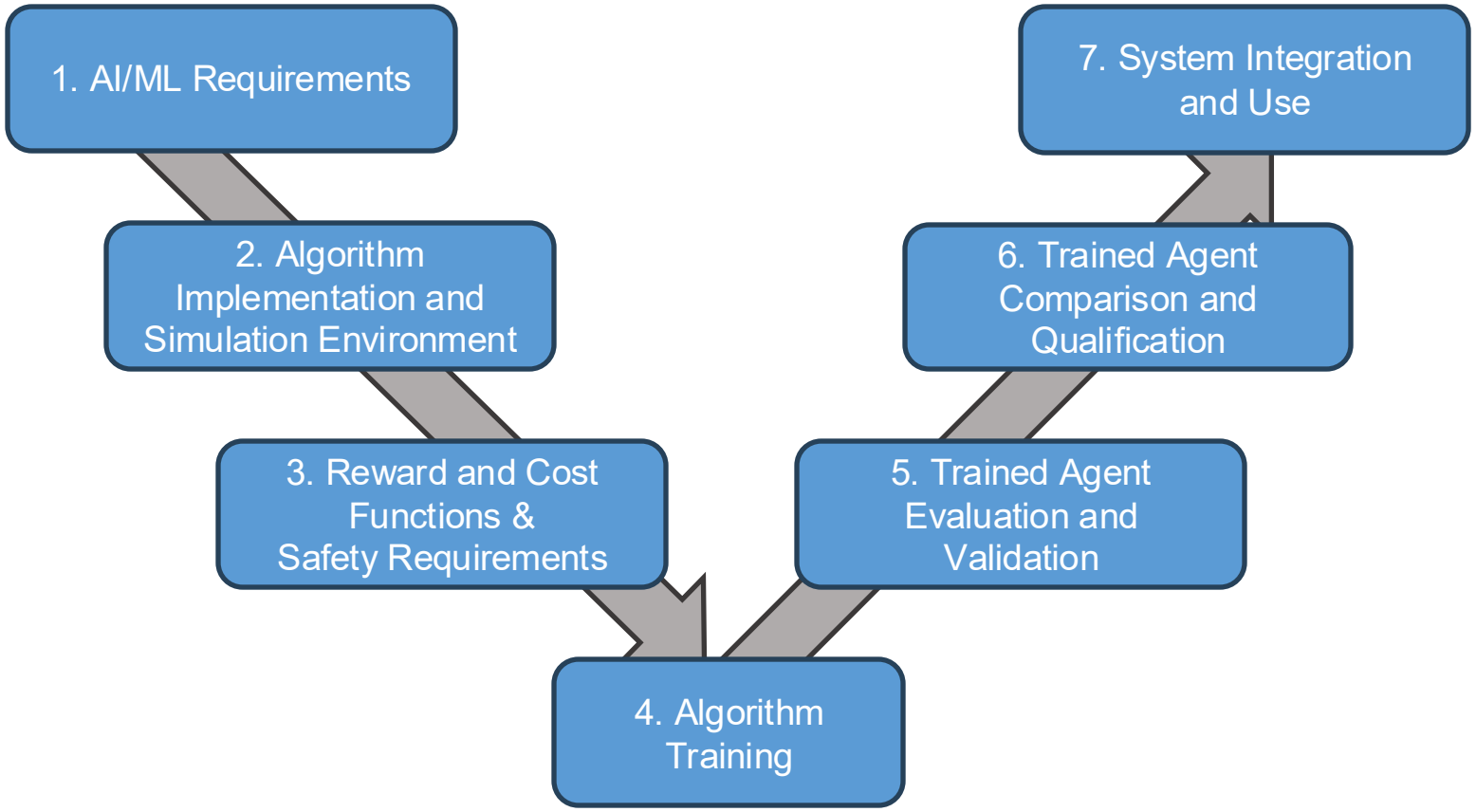
Considerations Must be Taken for Safety-Critical Deployments



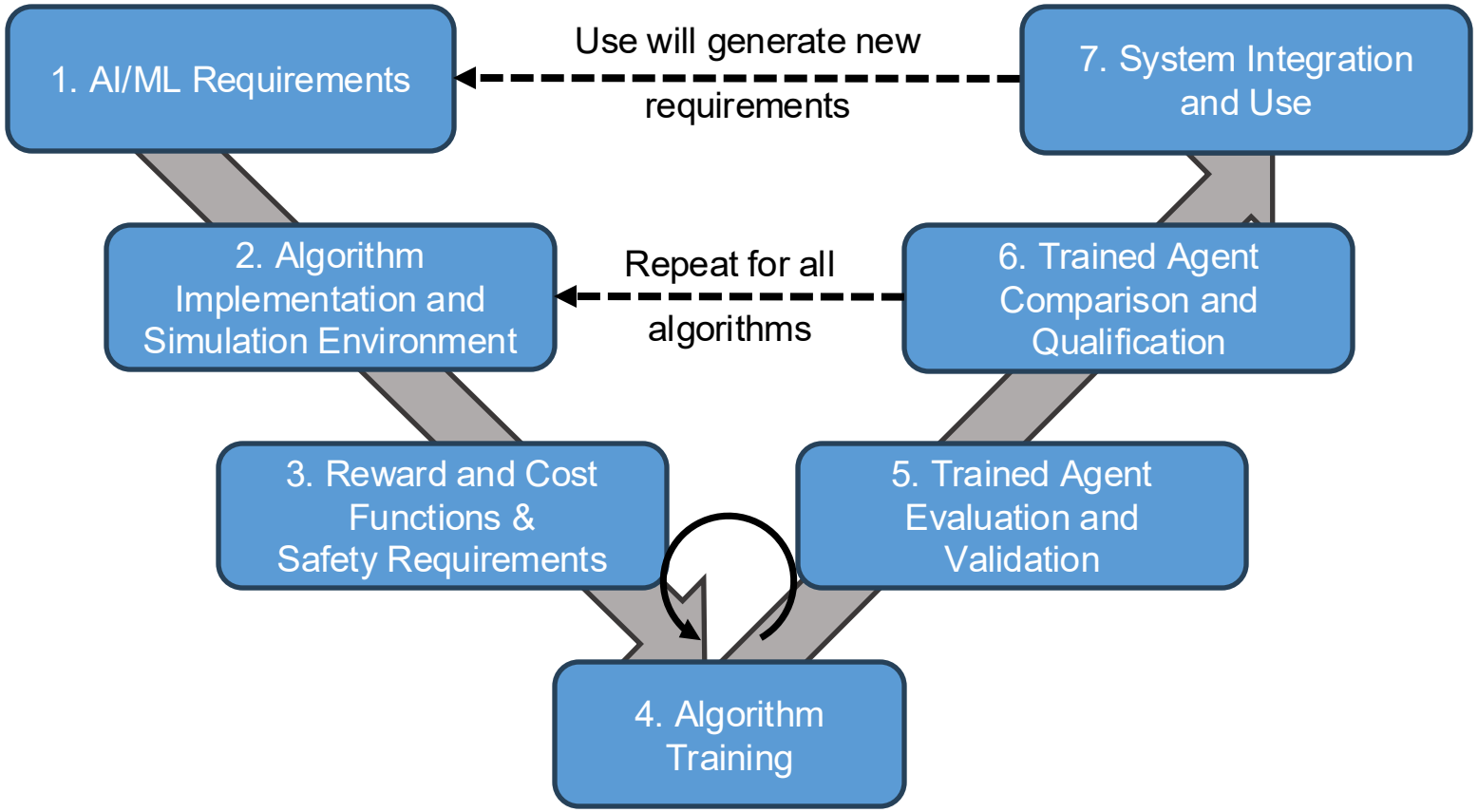
https://arxiv.org/abs/2209.08025

RLQP Helps Ensure Safety and Reliability

RLQP V-Model

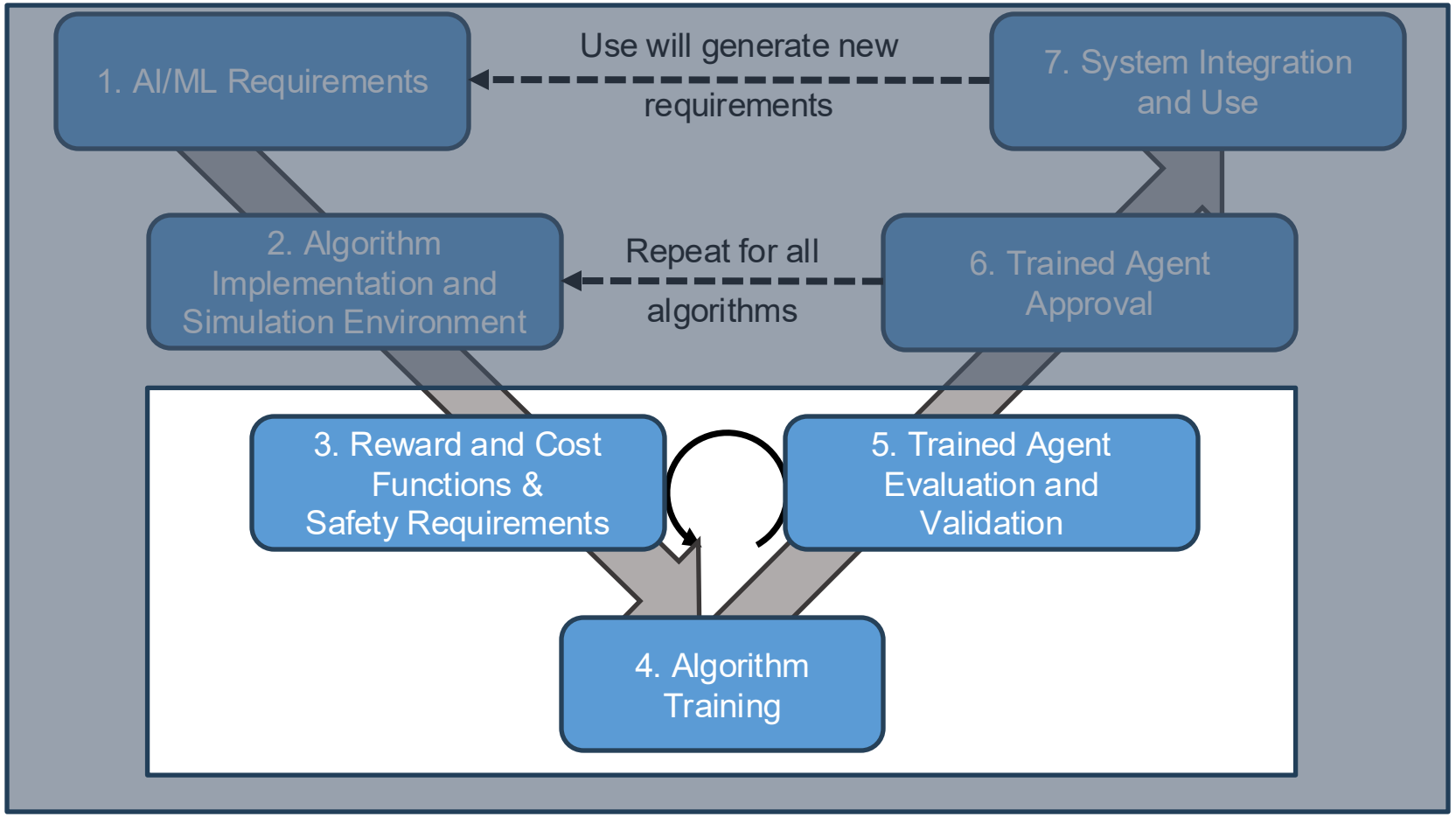


RLQP Helps Ensure Safety and Reliability



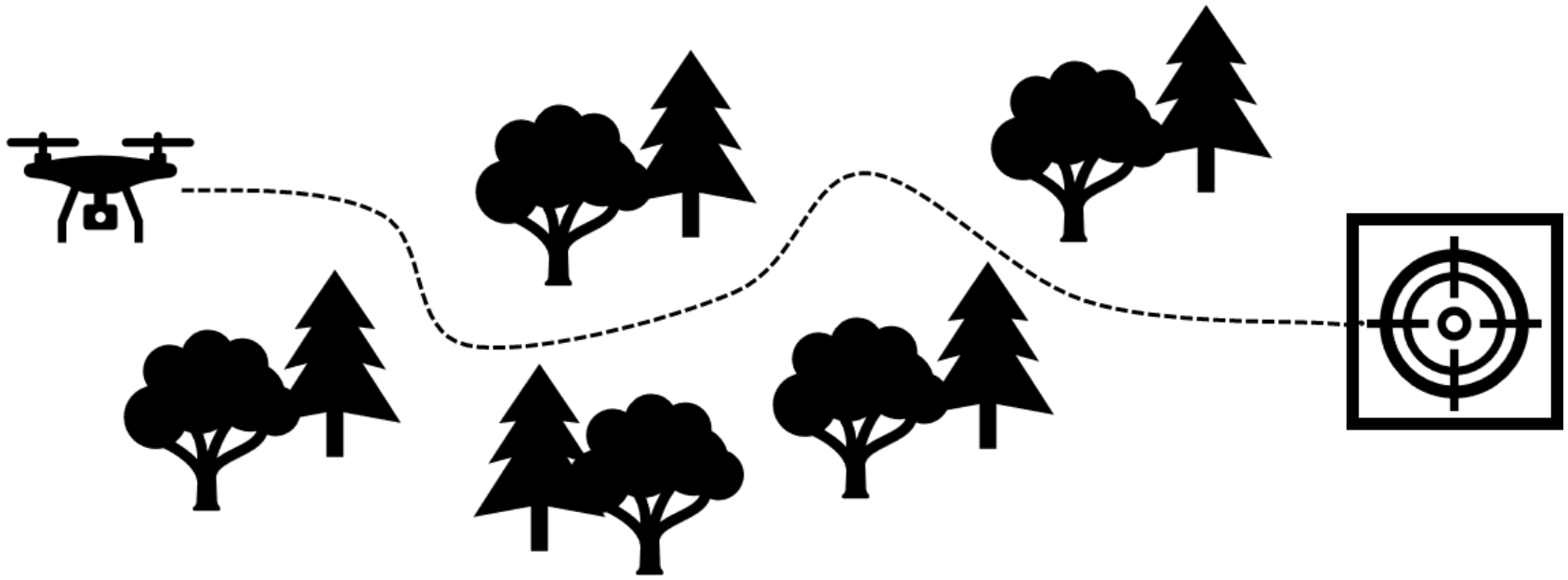
Michigan Tech

RL-STPA for Hazard Analysis and Safety

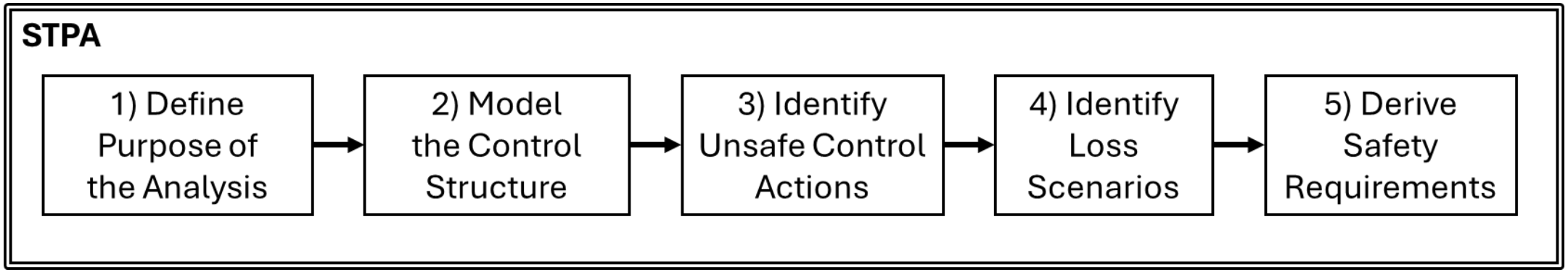


Michigan Tech

RLQP and RL-STPA are Demonstrated for Autonomous Drone Navigation

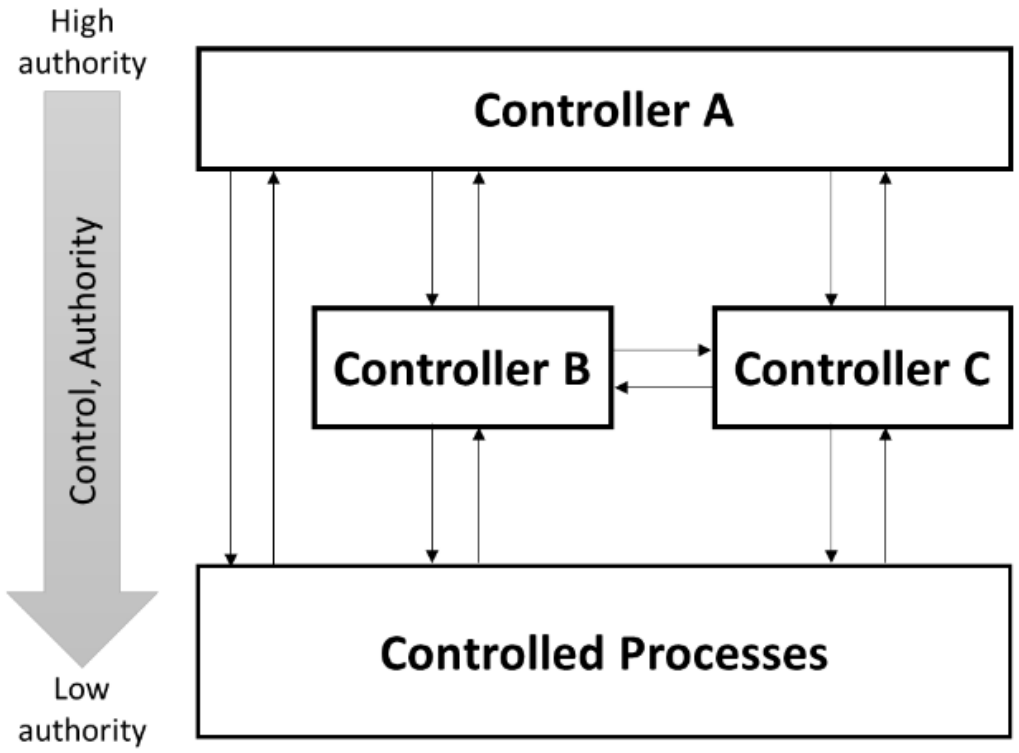


System-Theoretic Process Analysis is not Sufficient for RL



STPA Handbook, Leveson and Thomas 2018

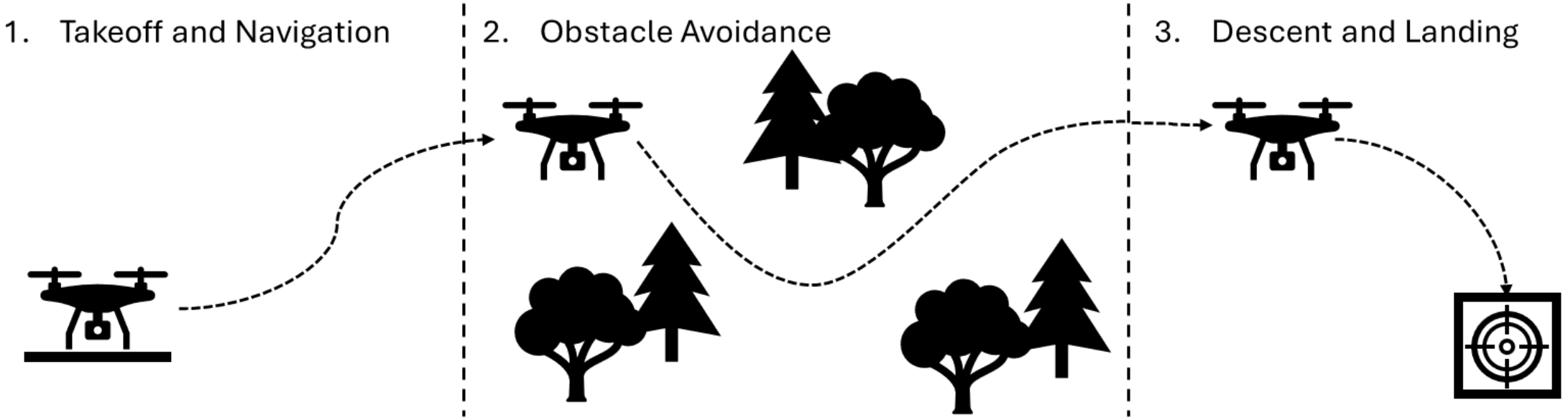
How Do We Model Control Structure in RL?



Reinforcement Learning is an **end-to-end** process, unlike traditional control

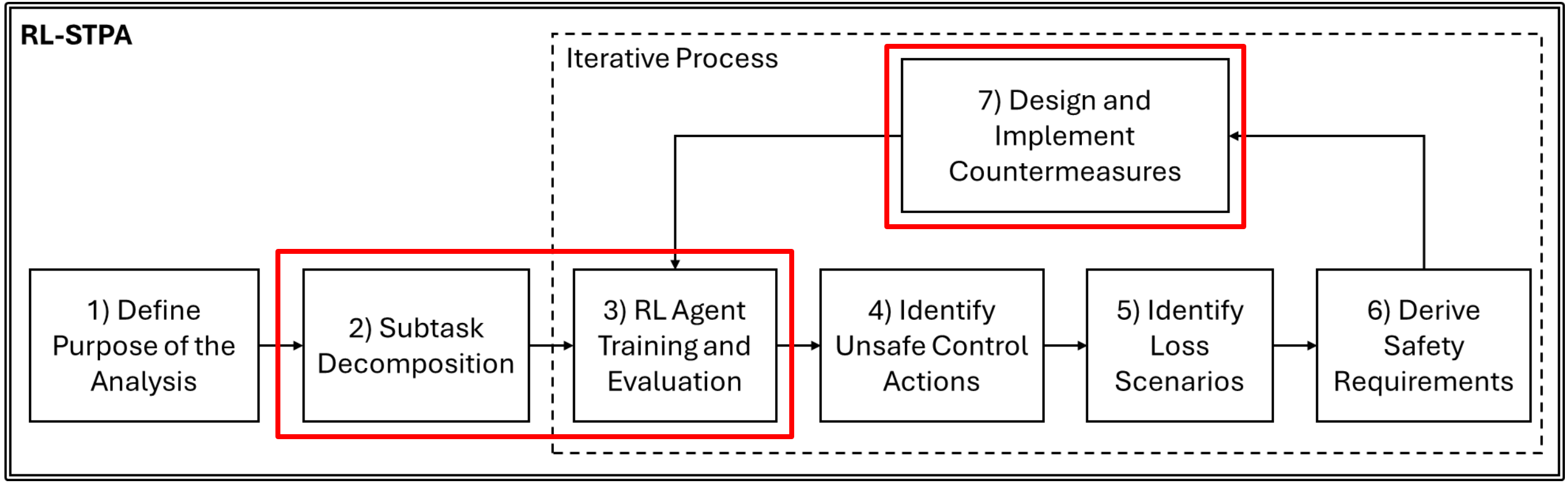
STPA Handbook, Leveson and Thomas 2018

Subtask Decomposition for End-to-End Learning

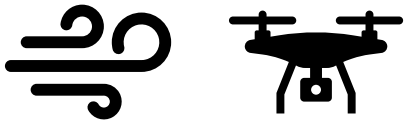


Subtasks allow for control elements to be extracted from an end-to-end process

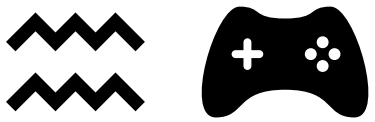
RL-STPA Reworks STPA for RL



Perturbation Testing for Evaluation



Environmental Perturbations



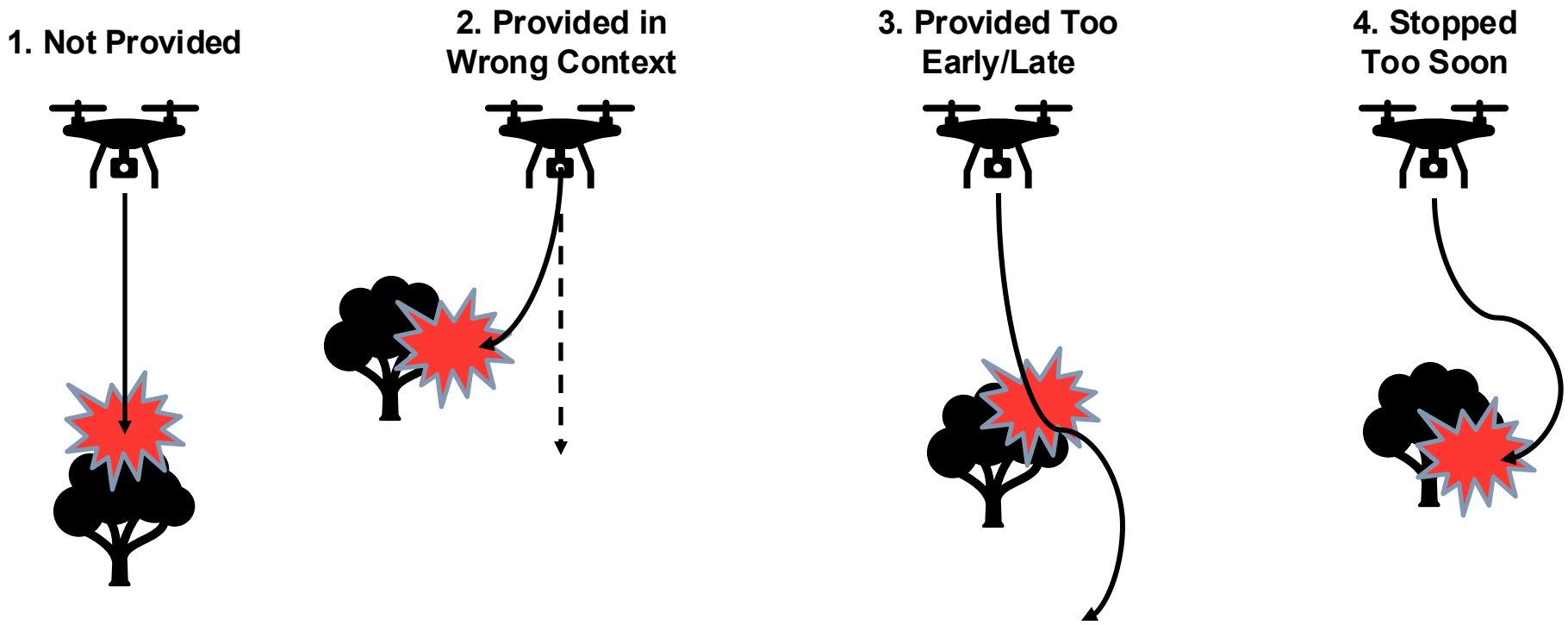
Sensor and Action Noise

Partial Observability

Input Delay

Michigan Tech

Unsafe Control "AGENT" Actions are Identified

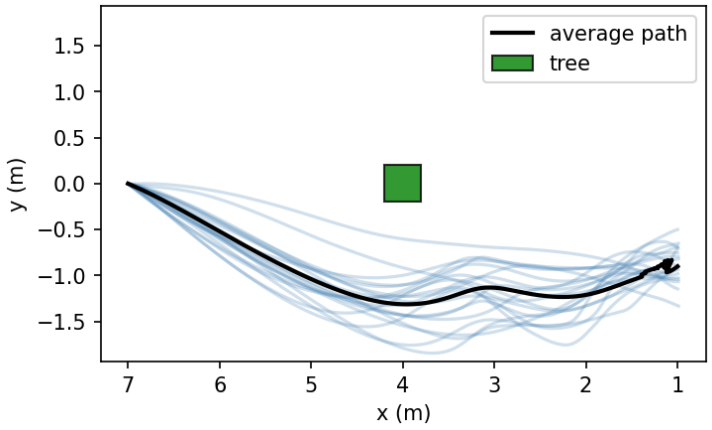


Unsafe Control Actions lead to hazards and loss scenarios

Michigan Tech

Perturbation Testing Exposes Loss Scenarios and Failure Rates

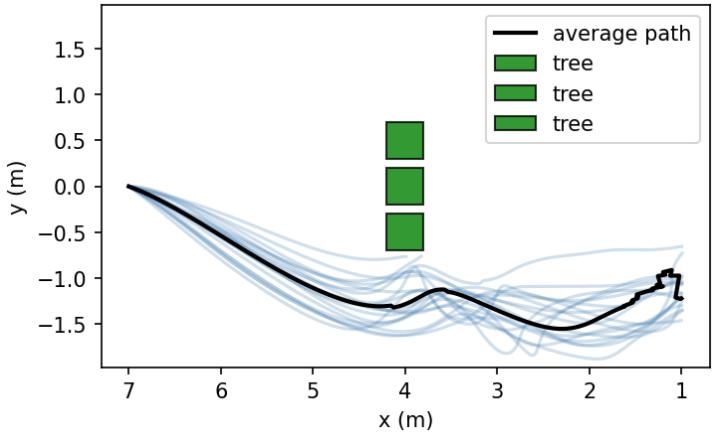
Baseline Condition



Minimum Separation Maintained

100% Success Rate

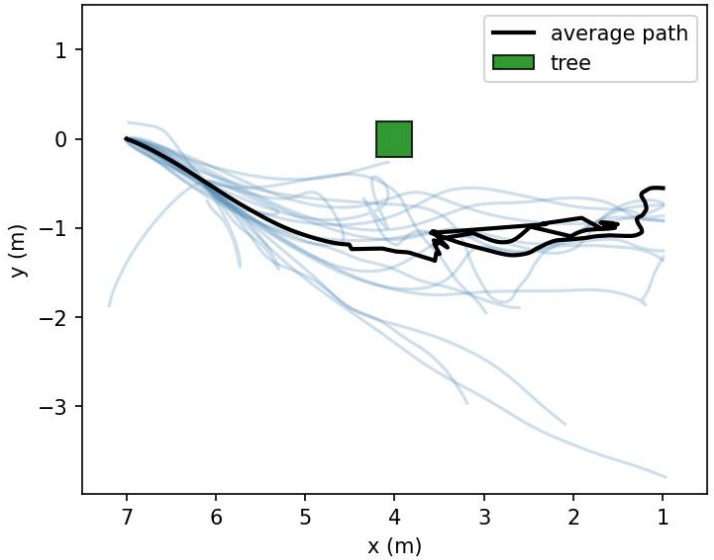
New Obstacle Type



Minimum Separation Violated

90% Success Rate

Heavy Wind



Minimum Separation Violated

55% Success Rate

Michigan Tech

Countermeasures are Implemented and Safety Requirements Derived

Curriculum Learning

Improved Reward Function

Functional Safety Limitations

By introducing wind into the training curriculum, the success rate improved from **55% to 95%**

Summary

- Reinforcement learning can learn complex control structures through trial and error
- **RLQP**: Framework for implementing safety-critical RL agents
- **RL-STPA**: Hazard analysis methodology designed for end-to-end learning systems
- Perturbation testing is critical for improving safety, robustness, and performance

Challenges and Future Work

- RL-STPA requires manual designation of the subtasks
- Automatic subtask discovery
- Combine with formal verification techniques
- Active perturbation selection
- Online adaptation for runtime monitoring

Questions?

The work presented here is funded by the Army through STTR-A22B-T002 “Metrics and Methods for Verification, Validation, Assurance and Trust of Machine Learning Models & Data for Safety-Critical Applications in Armaments Systems” (Contract #W15QKN-24-C-0038)

Contact: sasenczy@mtu.edu