Multi-Fidelity Testing and Evaluation of AI-Enabled Systems

September 17, 2024

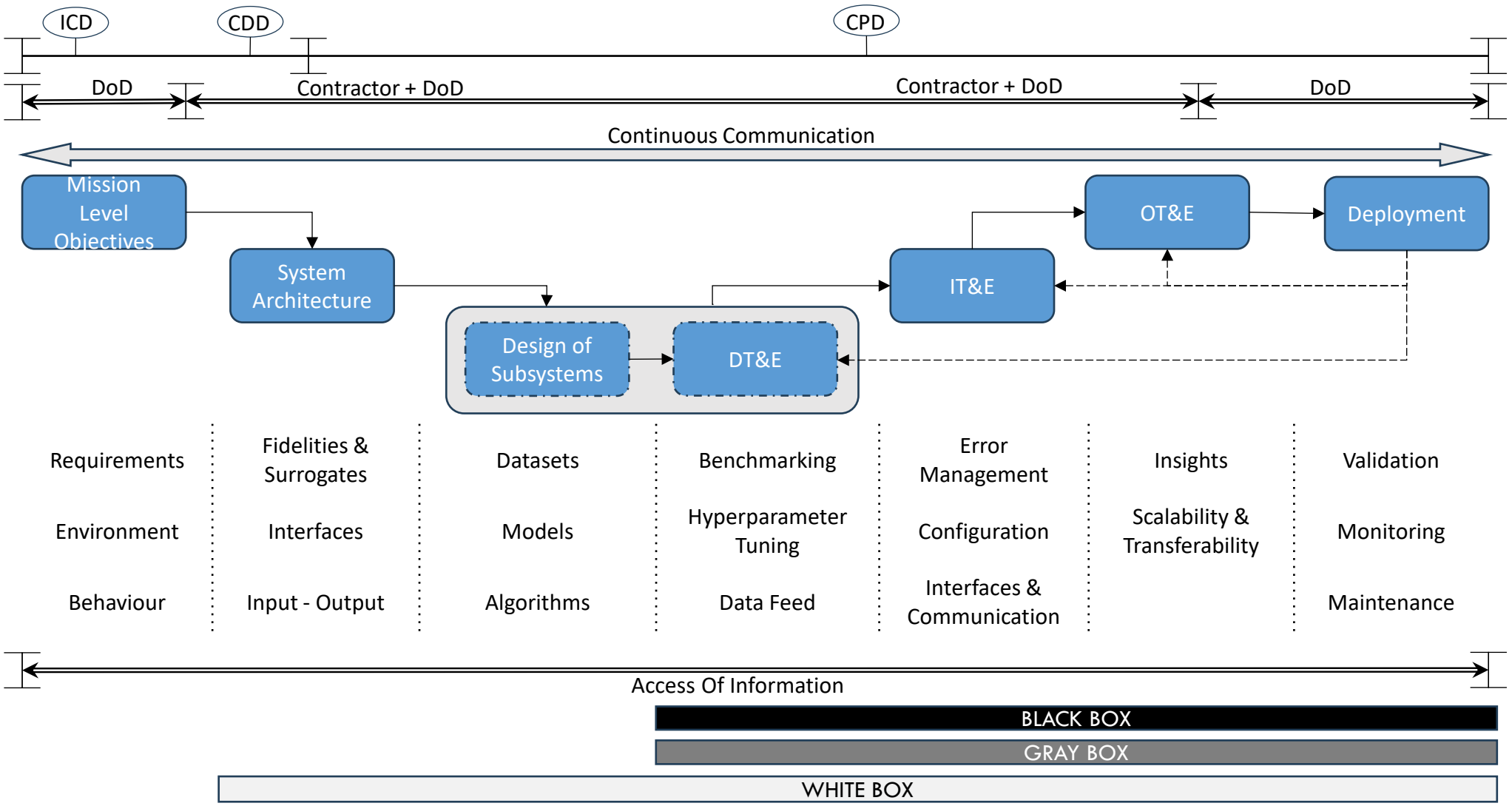
Robert J. Seif, Zichong Yang, Ziran Wang, Laura Freeman,
Jitesh H. Panchal

PURDUE UNIVERSITY

Design Engineering Laboratory @ Purdue
http://engineering.purdue.edu/DELP

Design Engineering Lab at Purdue

Motivation: Levels of Access of Information



Driving Questions Pertaining to T&E

- **Efficiency of T&E:** How to reduce the cost and time for T&E using evidence generated throughout the systems engineering process?

- **Access:** To what extent does the government need access to training data, algorithms, hardware, and software?
  - Reduced the need for testing (and increase in confidence assurance) resulting from an increase in data/model rights.

- **Cost of IP:** How much to pay for information (data, model parameters, etc.)?

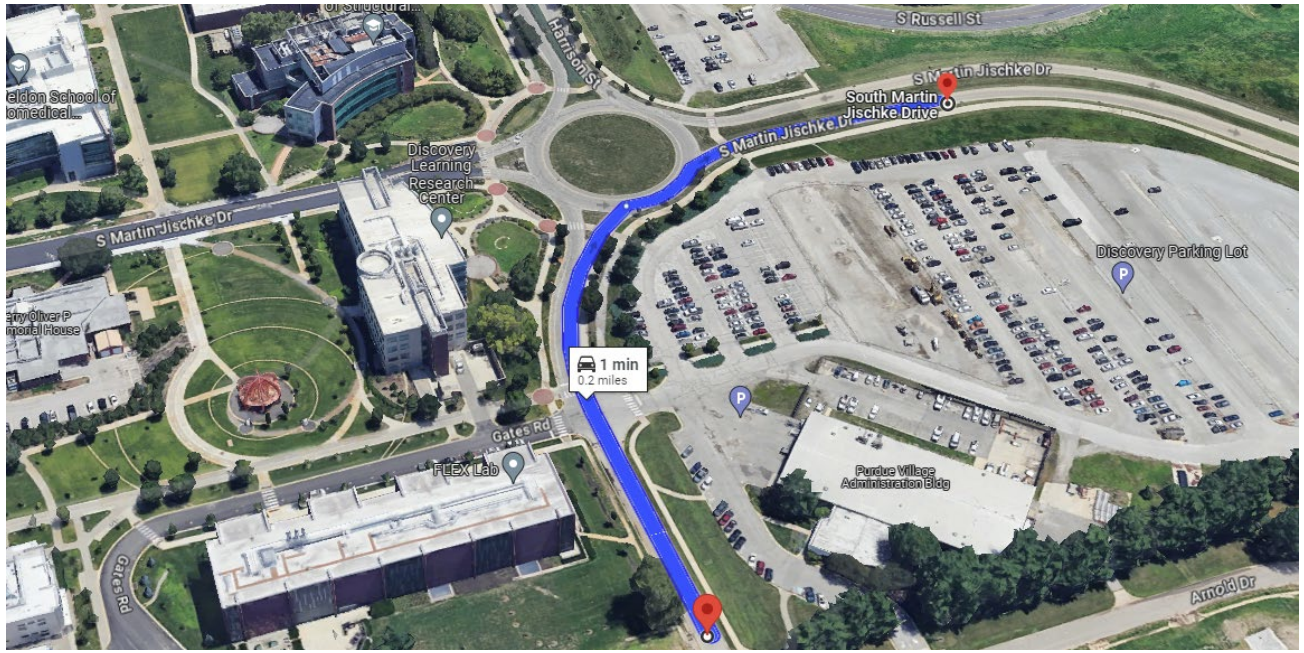
Autonomous Driving Use Case

- AI model: detection system on an autonomous vehicle
- Input: Video file from camera
- Output: Object location in each frame

[6] https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.alten.com%2Fnext-generation-camera-based-adas-development

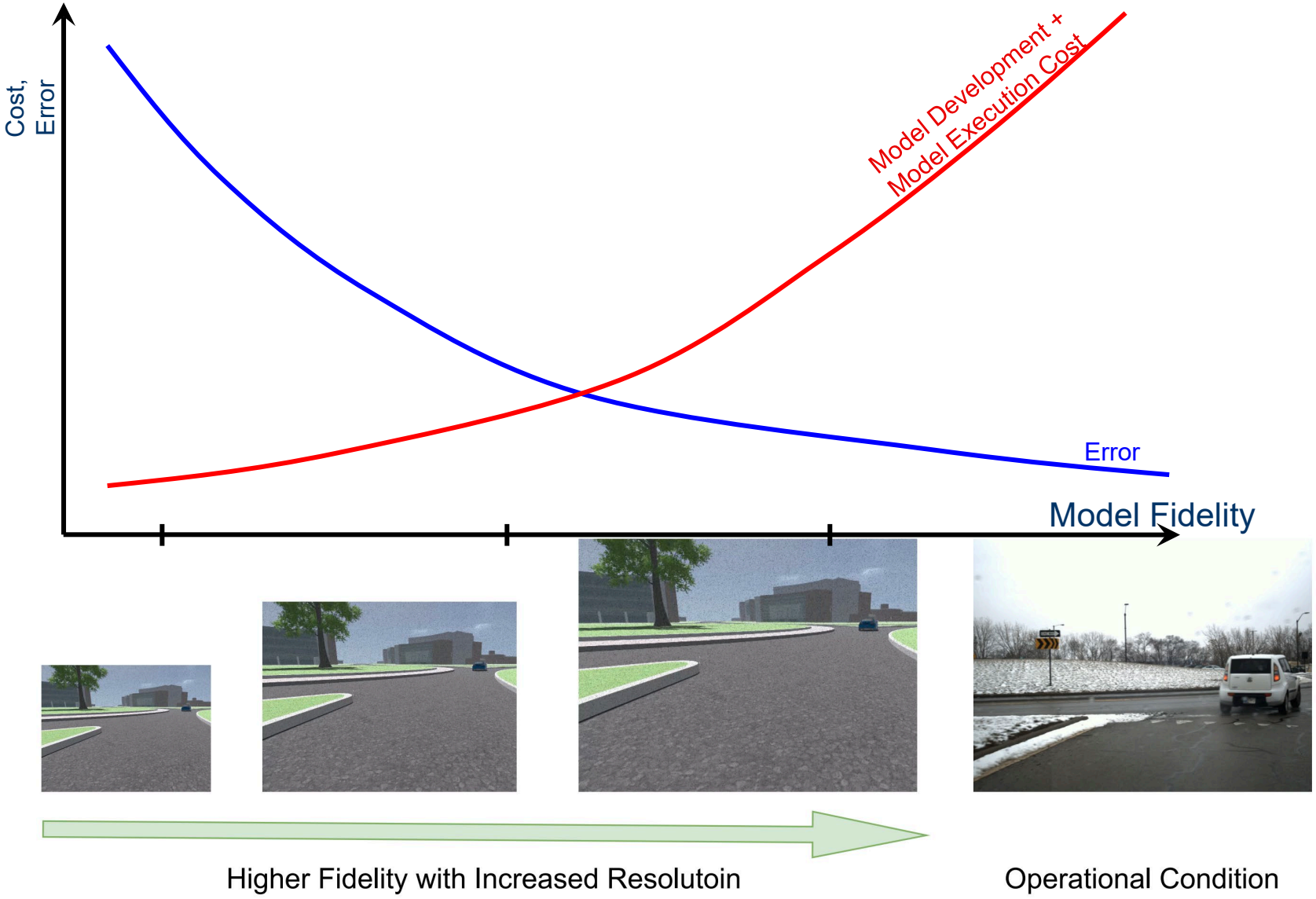
Autonomous Driving Use Case

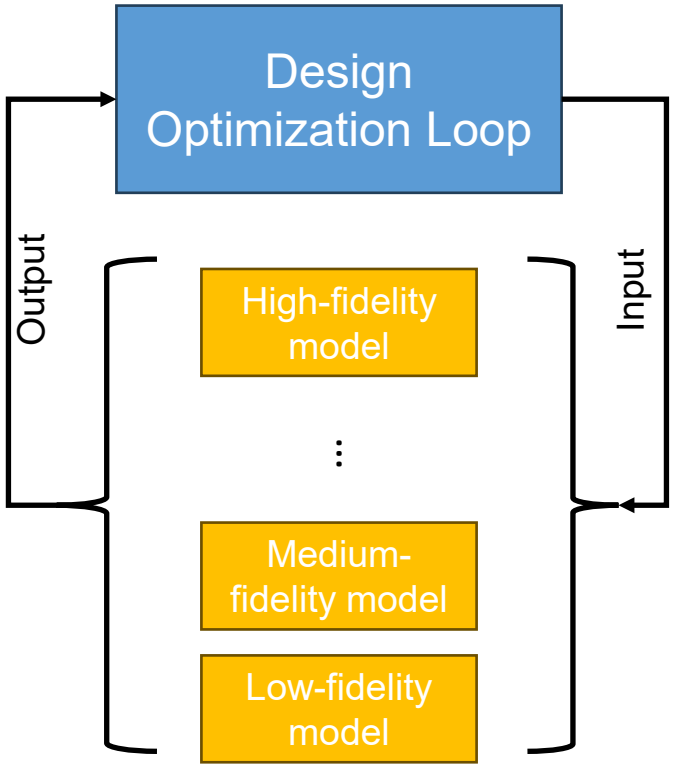
- **System tested:** YOLOv8 (object detection)
- **Setting:** Autonomous Vehicle driving through a roundabout
- **Simulation:** Unity simulation engine and occlusion post-processing



R	Requirements
r_1	System must operate on a 1080p HD video format
r_2	System must operate in direct overhead sunlight
r_3	System must operate in raining weather condition
r_4	System must operate under 25% random occlusion of vision field
r_5	System must operate with a 30 fps operational speed
r_6	System must have average confidence over 80% in vehicle prediction
r_i	…

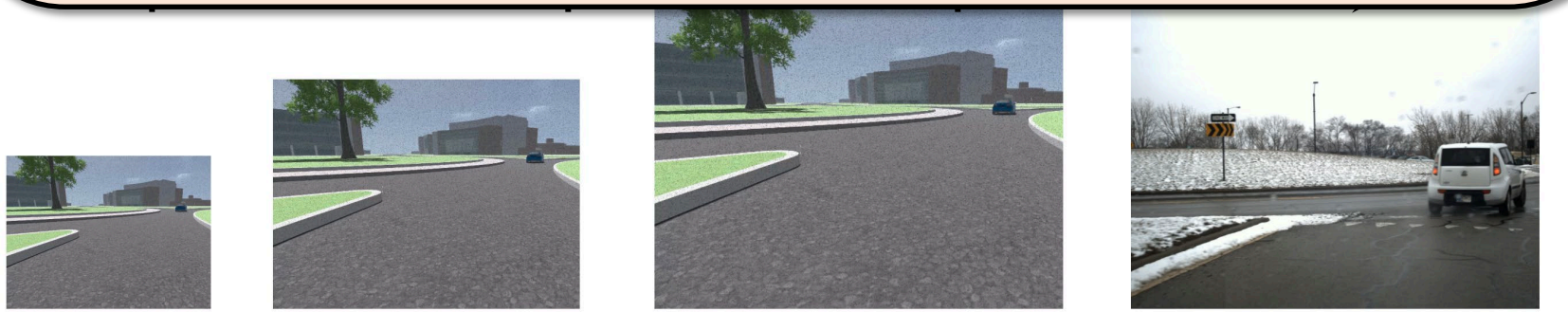
Models at Multiple Fidelities





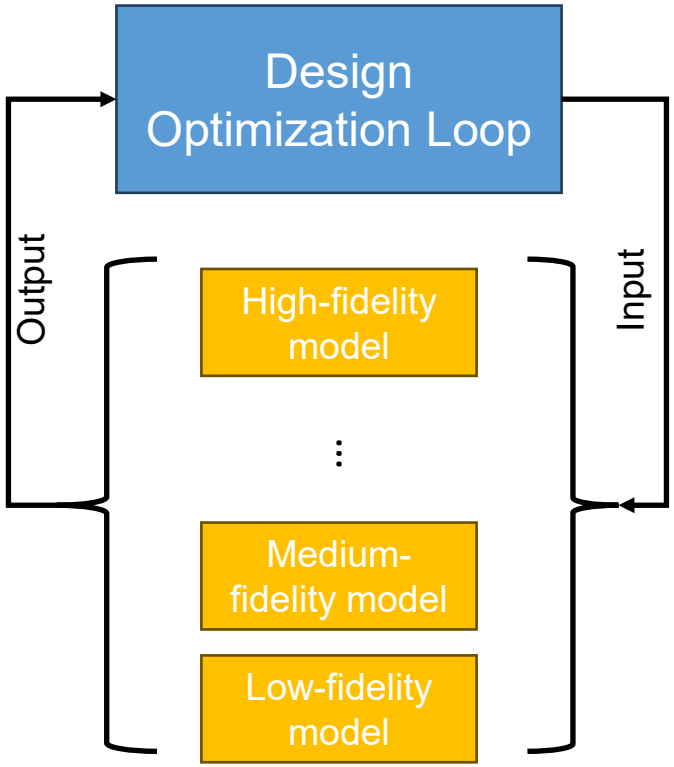
Models at Multiple Fidelities

Need: a method to **design test sequences** to most efficiently reach a **degree of certainty** about how well the system satisfies each requirement in the designated operational environment.



Higher Fidelity with Increased Resolutoin

Operational Condition



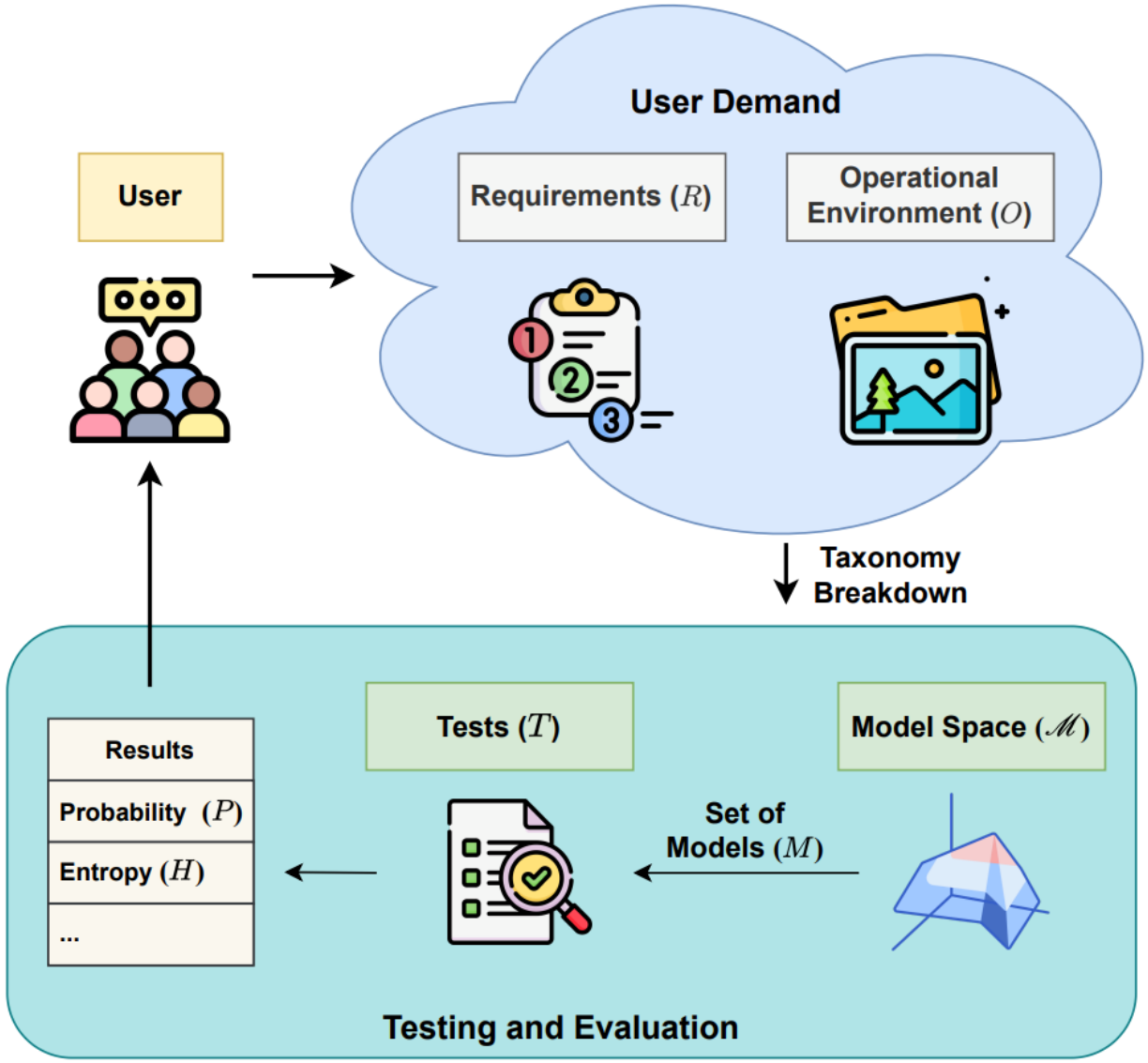
Proposed Approach

Acquisition (User) Side

- Define R & O
- Receive test results

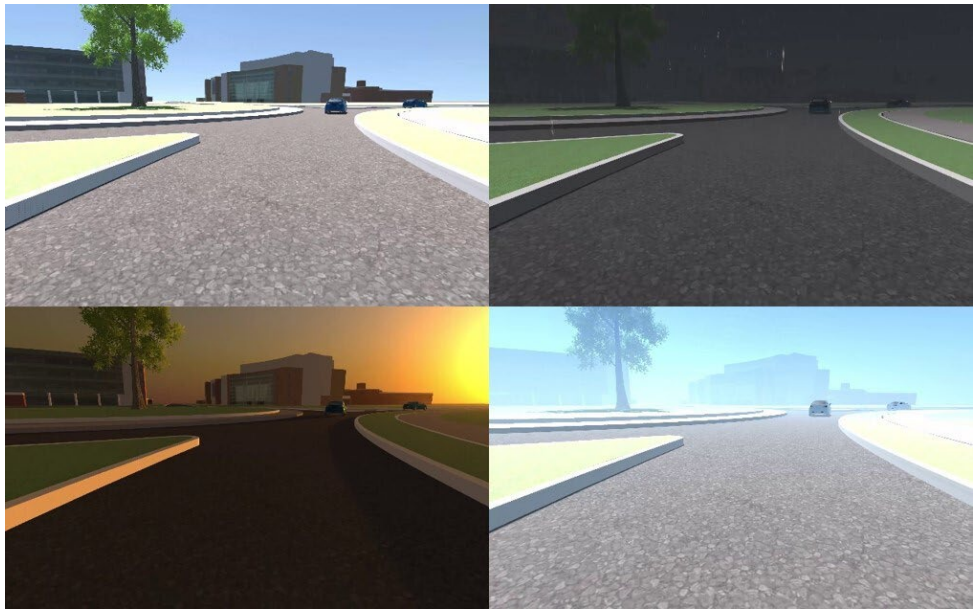
Testing and Evaluation Side

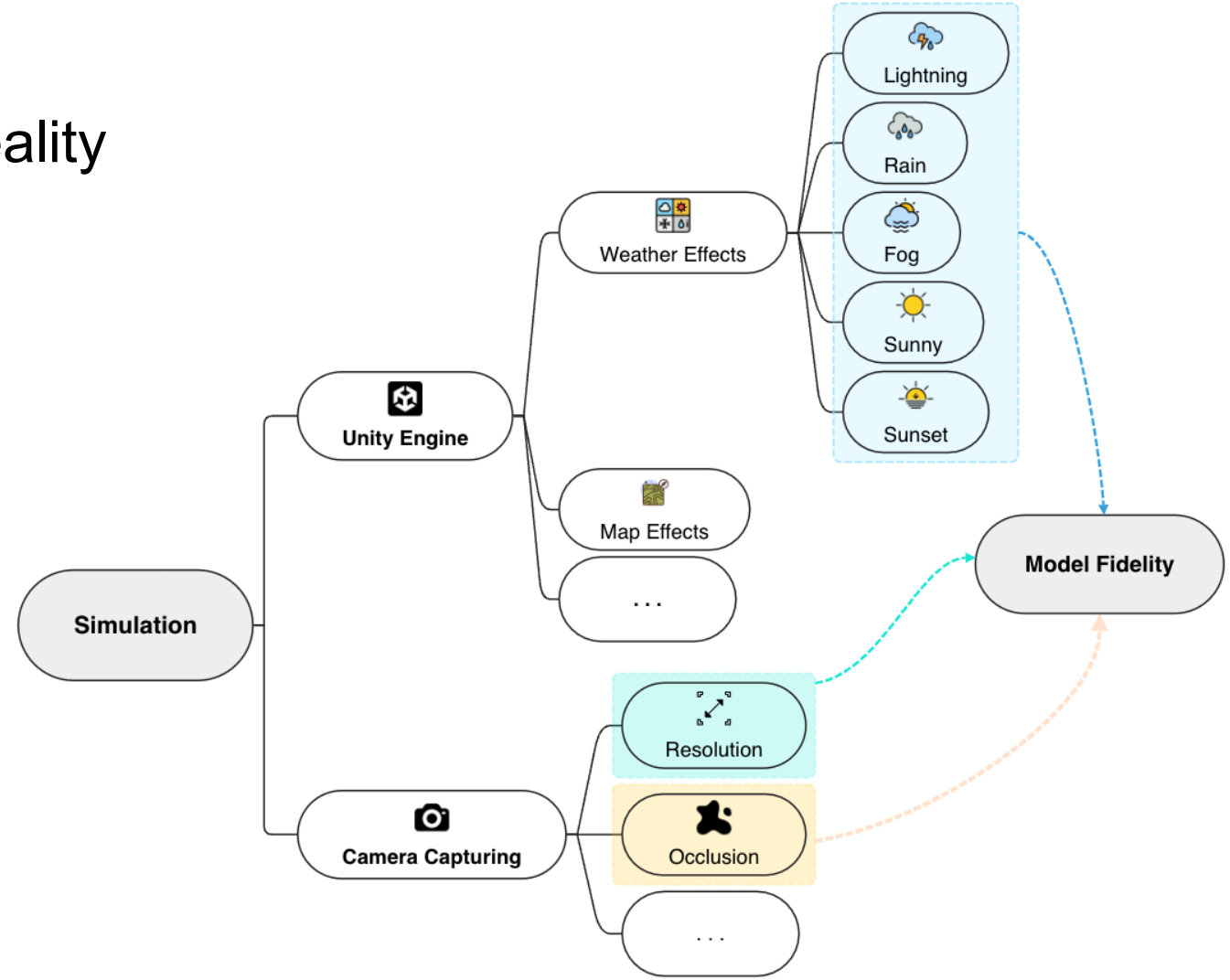
- Model Space Definition
  - Taxonomy Breakdown
- Sequential Test Selection
  - Bayesian Optimization + Entropy
- Updating Belief
  - Bayes' Rule



Constructing the Model Space

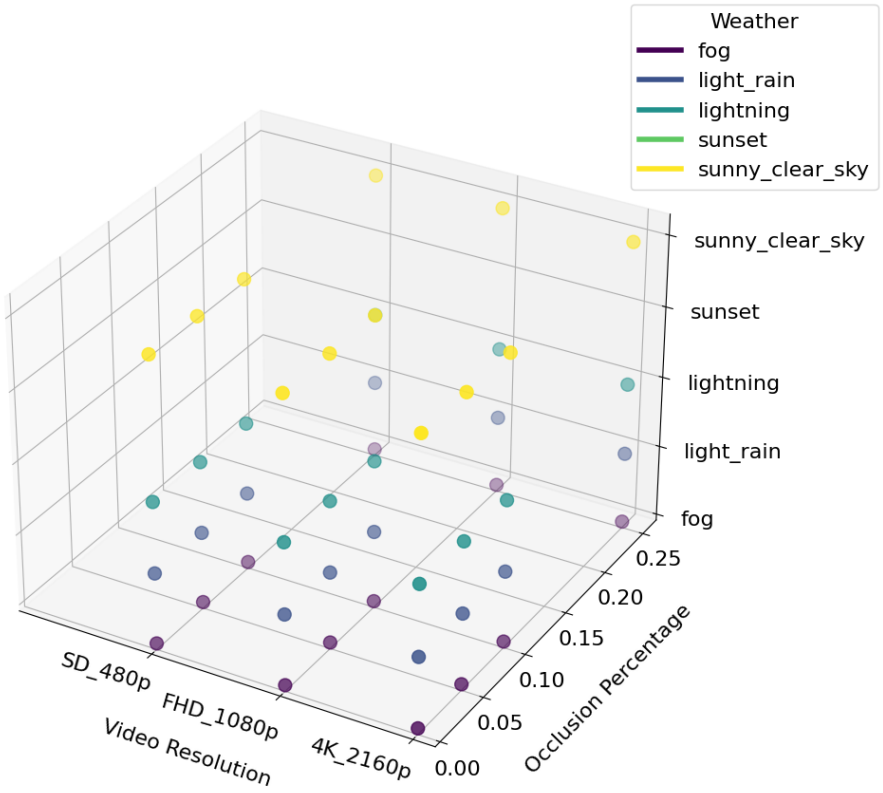
- ***Model(m_j):*** instance of simulation
- ***Fidelity:*** accuracy in representing reality
- ***Cost:*** $c = c_p \times c_l \times c_o$

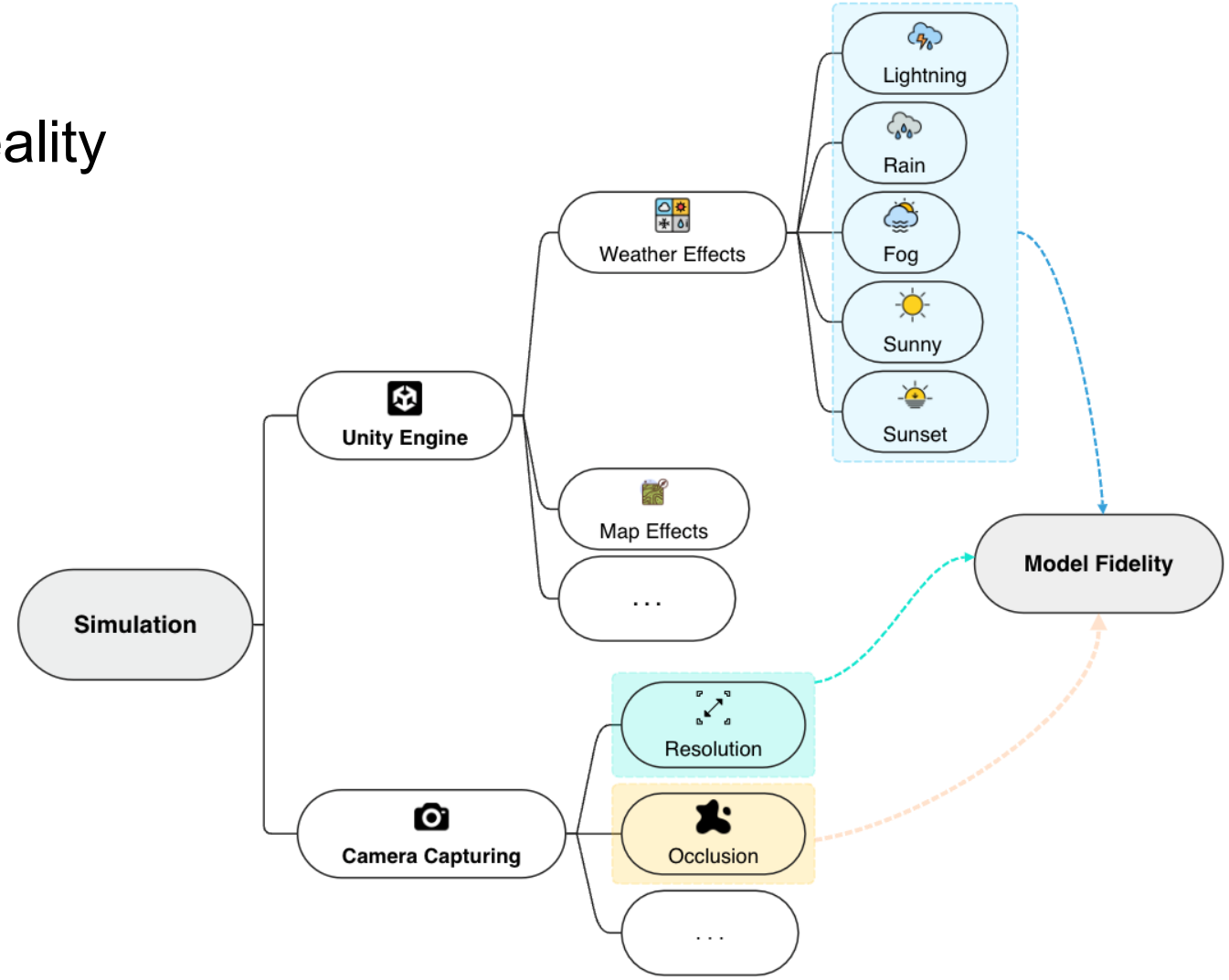




Constructing the Model Space

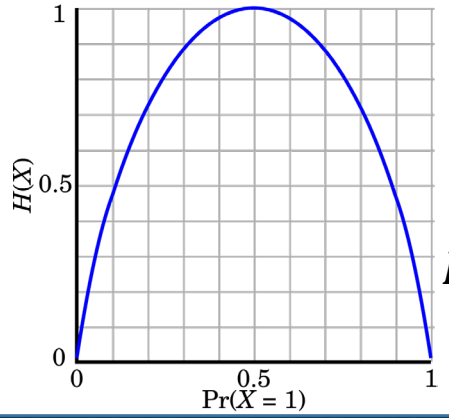
- ***Model(m_j):*** instance of simulation
- ***Fidelity:*** accuracy in representing reality
- ***Cost:*** $c = c_p \times c_l \times c_o$





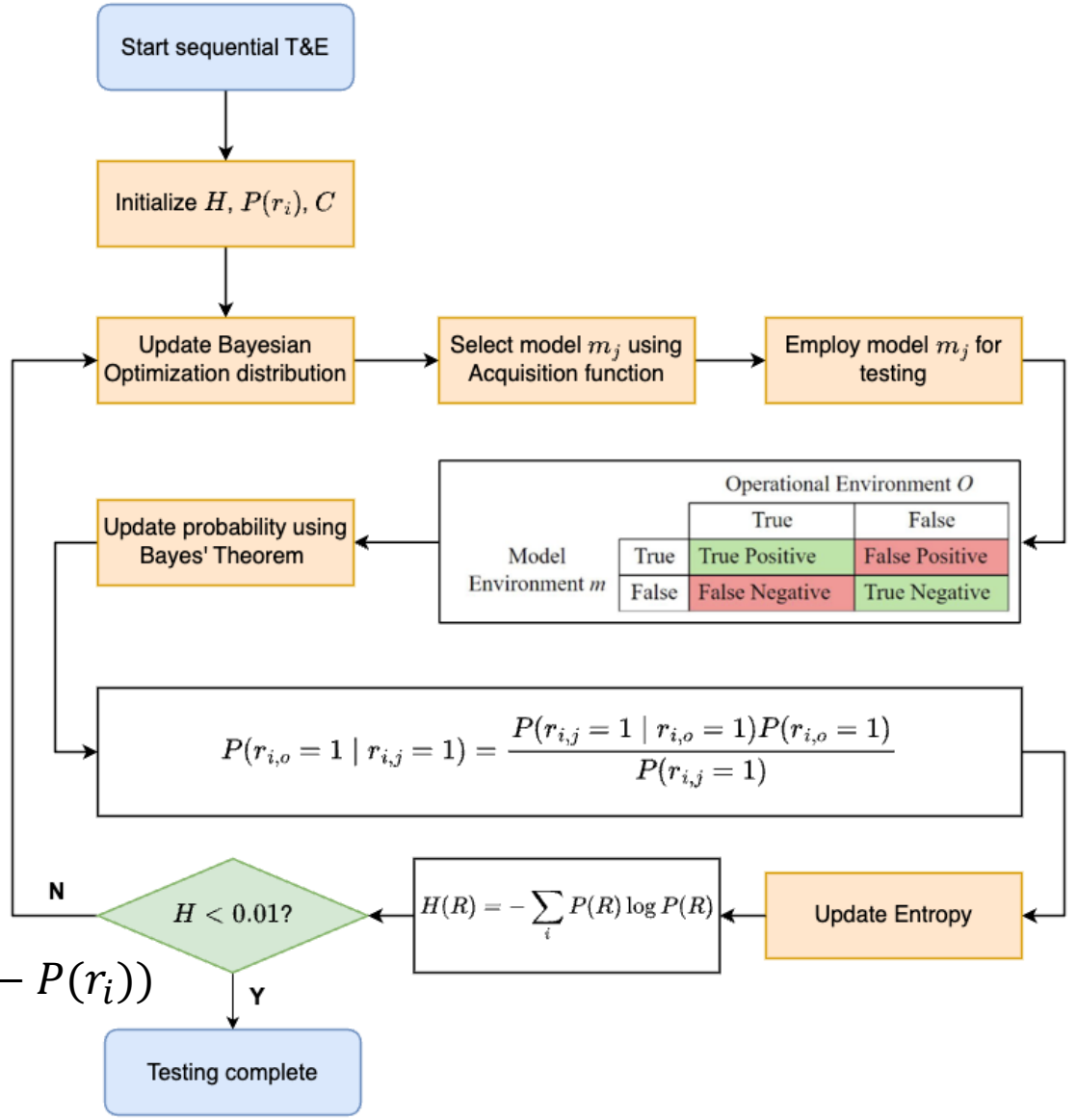
Sequential Model Selection

- **Random variables:** $r_{i,j}$, $r_{i,o}$ (probability of requirement r_i being satisfied)
- **Objectives of the T&E process:**
  - minimize uncertainty about whether the system satisfies the requirements in the operational environment
  - Minimize the cost of T&E
- **Model selection:** Acquisition Function
- **Belief update:** Bayes' Rule



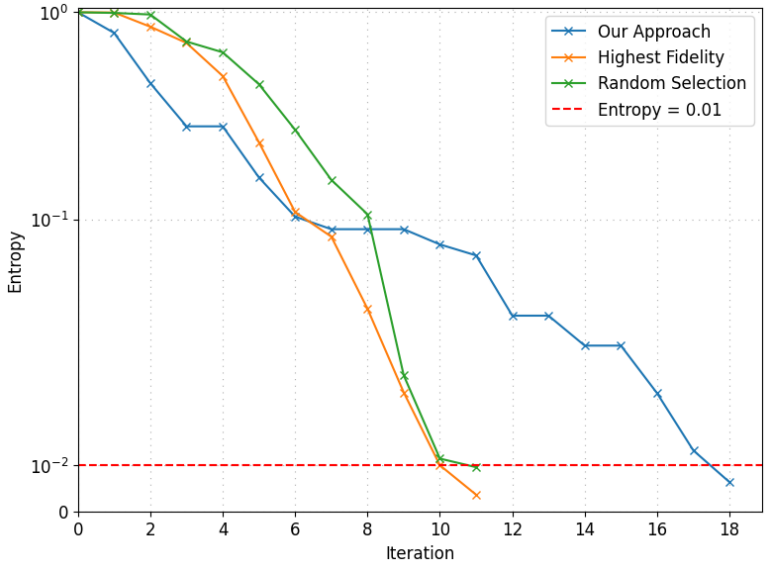
Uncertainty measure: Shannon Entropy

$$H(r_i) = -P(r_i)log_2P(r_i) - (1 - P(r_i))log_2(1 - P(r_i))$$

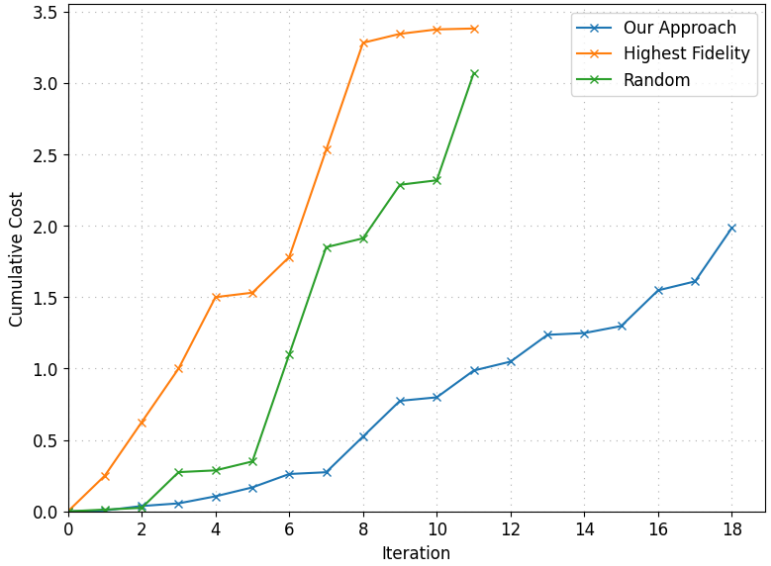


Experiment Results

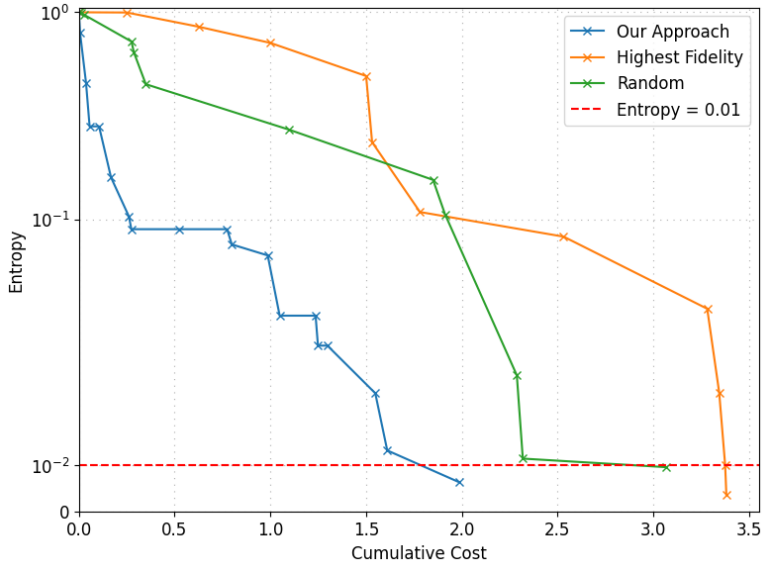
- Stop condition: entropy $H < 0.01$.
- Entropy and accumulated costs are recorded.
- Our approach: **41% reduced cost**, broader exploration of model space



entropy over number of tests



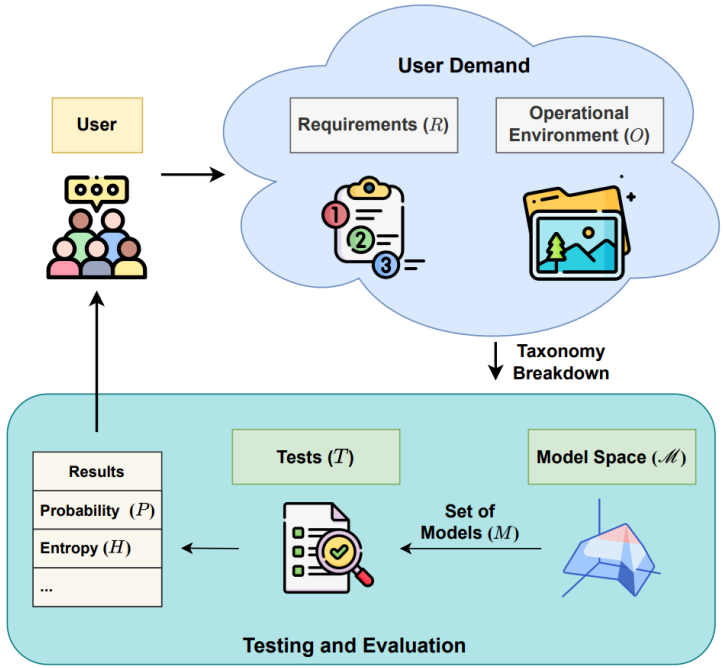
cumulative cost over number of tests

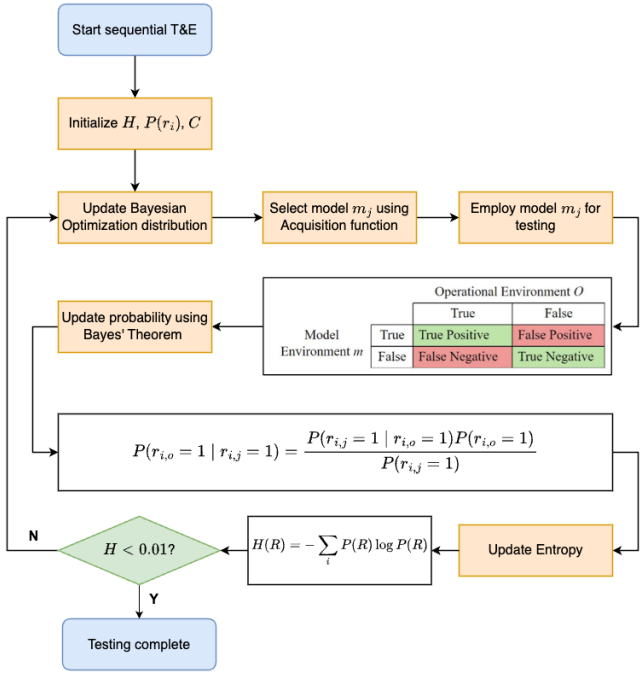


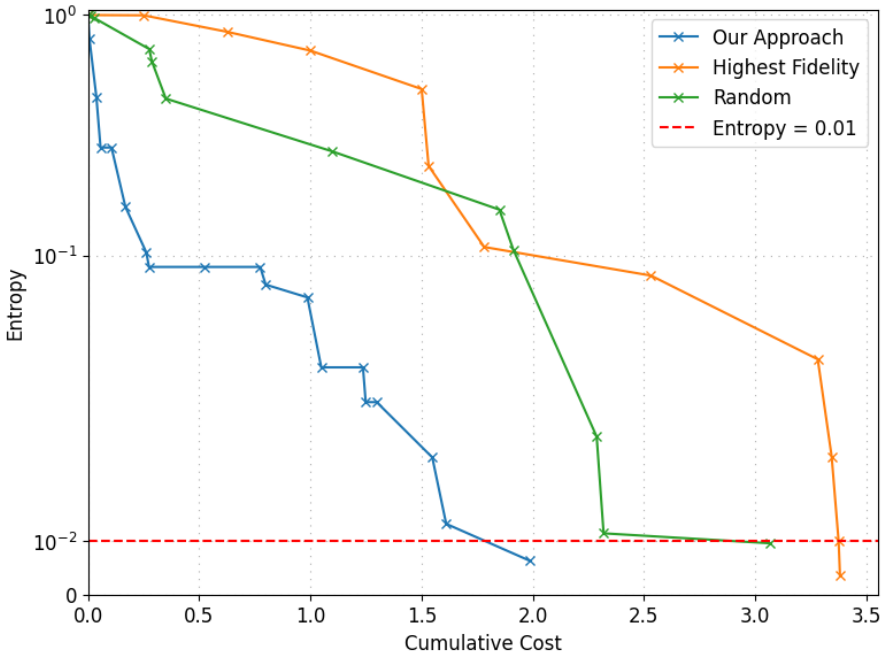
entropy vs. cumulative cost

Conclusions

- ***A method*** for designing a multi-fidelity test plan to verify whether the given AI-enabled system can satisfy the given list of requirements and corresponding failure modes.
- The proposed method can ***find the testing sequence with maximum utility*** by experimenting with testing the perception system of an autonomous vehicle.







Future Work

- Refine the use case to include evolving AI models
- Automation in test generation and model selection.
- Using high and low-fidelity models in parallel.
- Adaptive testing with evolving systems.

Video captured from Carla simulation engine

Thank you

Acknowledgment: This material is based upon work supported, in whole or in part, by the U.S. Department of Defense, Director, Operational Test and Evaluation (DOT&E) through the Office of the Assistant Secretary of Defense for Research and Engineering (ASD(R&E)) under Contract HQ003419D0003. Any views, opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense (specifically DOT&E and ASD(R&E)).