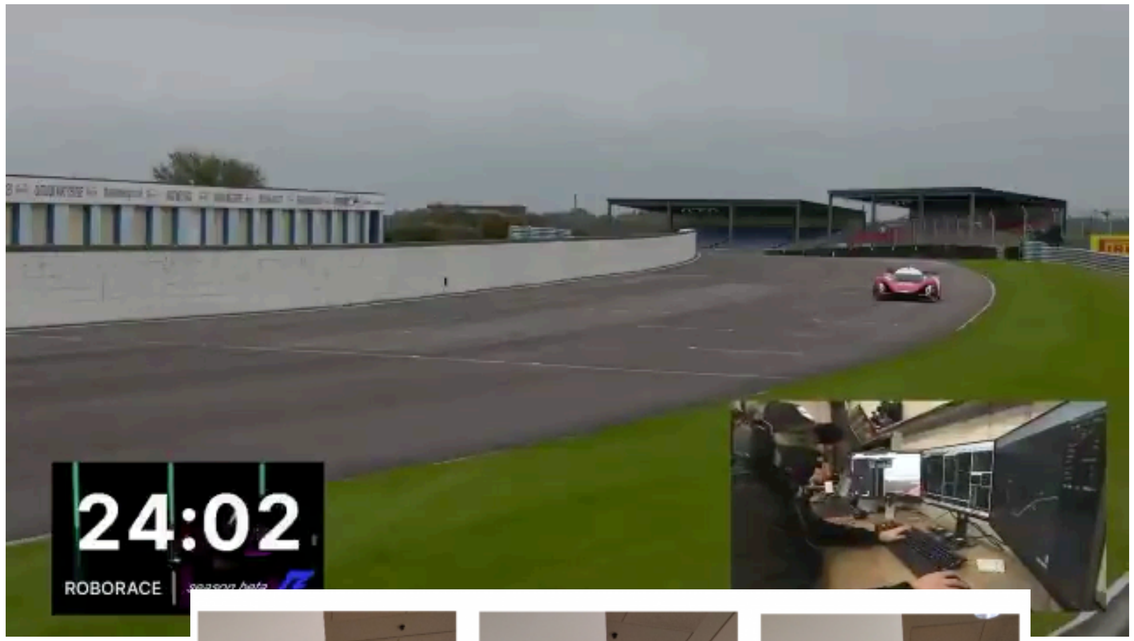
Perception Challenge: Autonomous Vehicles

Leilani H. Gilpin
with Adam Amos-Binks and Dustin Dannenhauer

Autonomous Vehicles are Prone to Failure

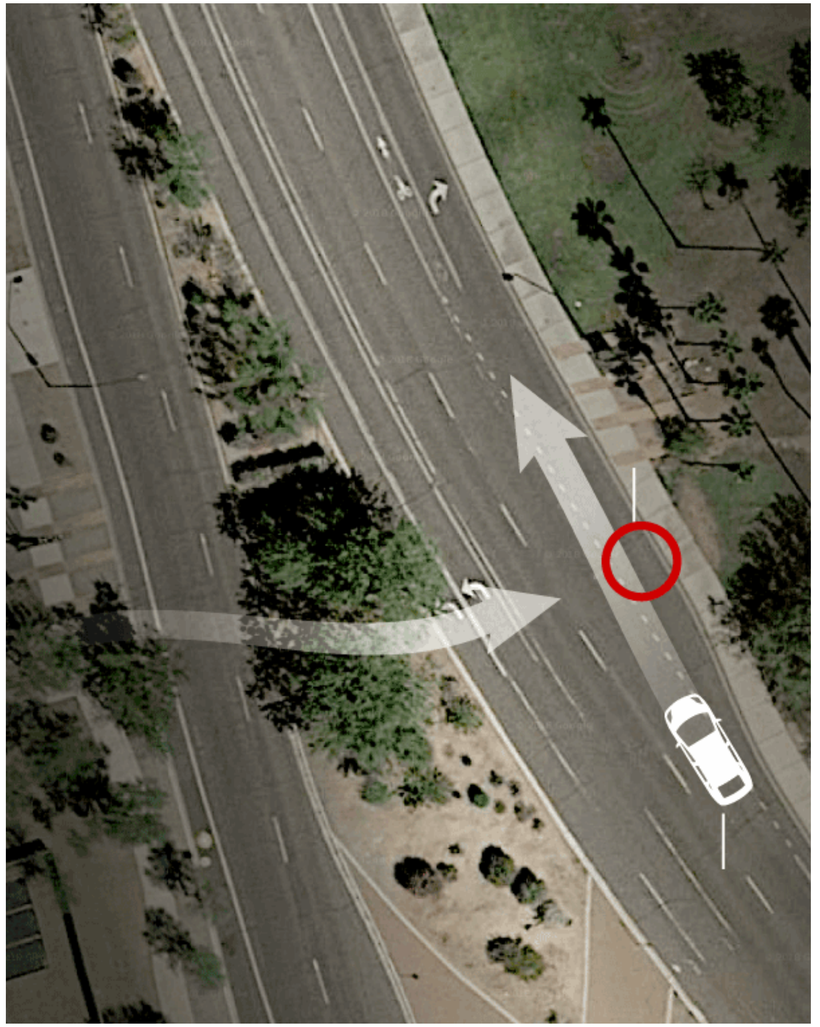


K. Eykholt et al. “Robust Physical-World Attacks on Deep Learning Visual Classification.”

Predictive Inequity in Object Detection

Benjamin Wilson [1] Judy Hoffman [1] Jamie Morgenstern [1]

Uber Example in my PhD Work



Uber Example in my PhD Work

L. H. Gilpin, V. Penubarthi and L. Kagal, "Explaining Multimodal Errors in Autonomous Vehicles," *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, 2021, pp. 1-10, doi: 10.1109/DSAA53316.2021.9564178.

Lack of Data and Challenges for AVs

- Existing Challenges
  - Targeted as optimizing a mission or trajectory and not safety.
  - Data is hand-curated
- Failure data is not available
  - Unethical to get it (cannot just drive into bad situations).
  - Want the data to be realistic (usually difficult in simulation).
- Develop a set of challenges and stress tests that **generate** new errors.

Existing Challenges and Benchmarks

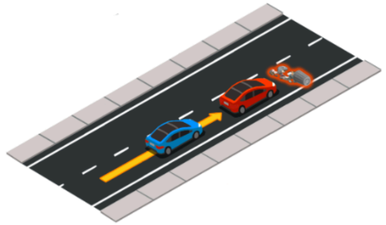
Not Focused on Out of Domain Errors

NHTSA-inspired pre-crash scenarios

We have selected 10 traffic scenarios from the **NHTSA pre-crash typology** to inject challenging driving situations into traffic patterns encountered by autonomous driving agents during the challenge.

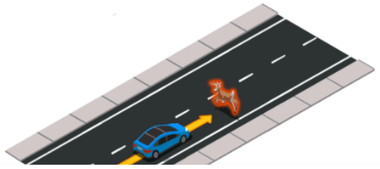
Traffic Scenario 01: Control loss without previous action

- **Definition**: Ego-vehicle loses control due to bad conditions on the road and it must recover, coming back to its original lane.



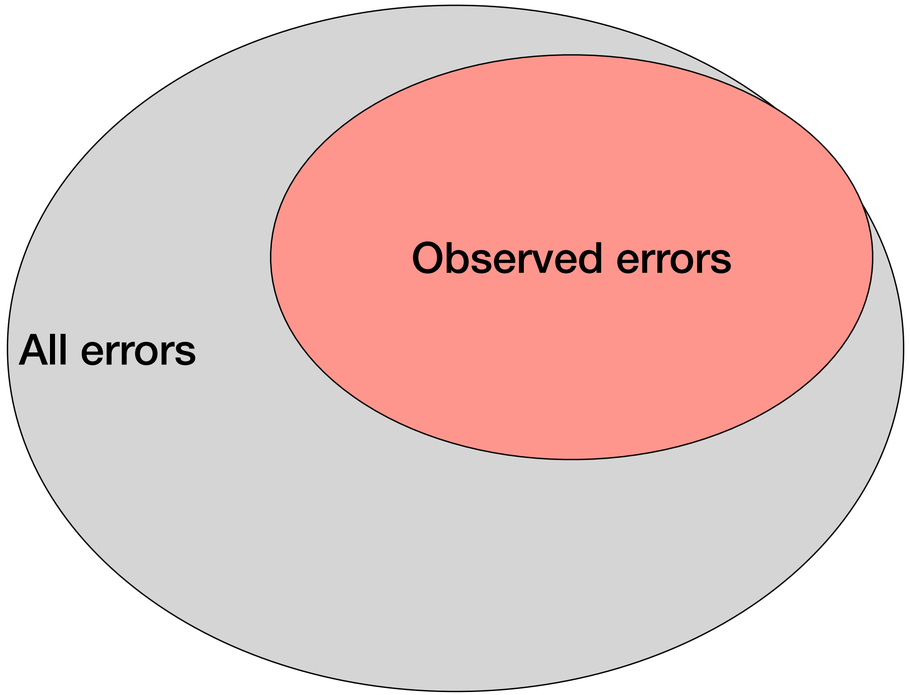
Traffic Scenario 02: Longitudinal control after leading vehicle's brake

- **Definition**: Leading vehicle decelerates suddenly due to an obstacle and ego-vehicle must react, performing an emergency brake or an avoidance maneuver.



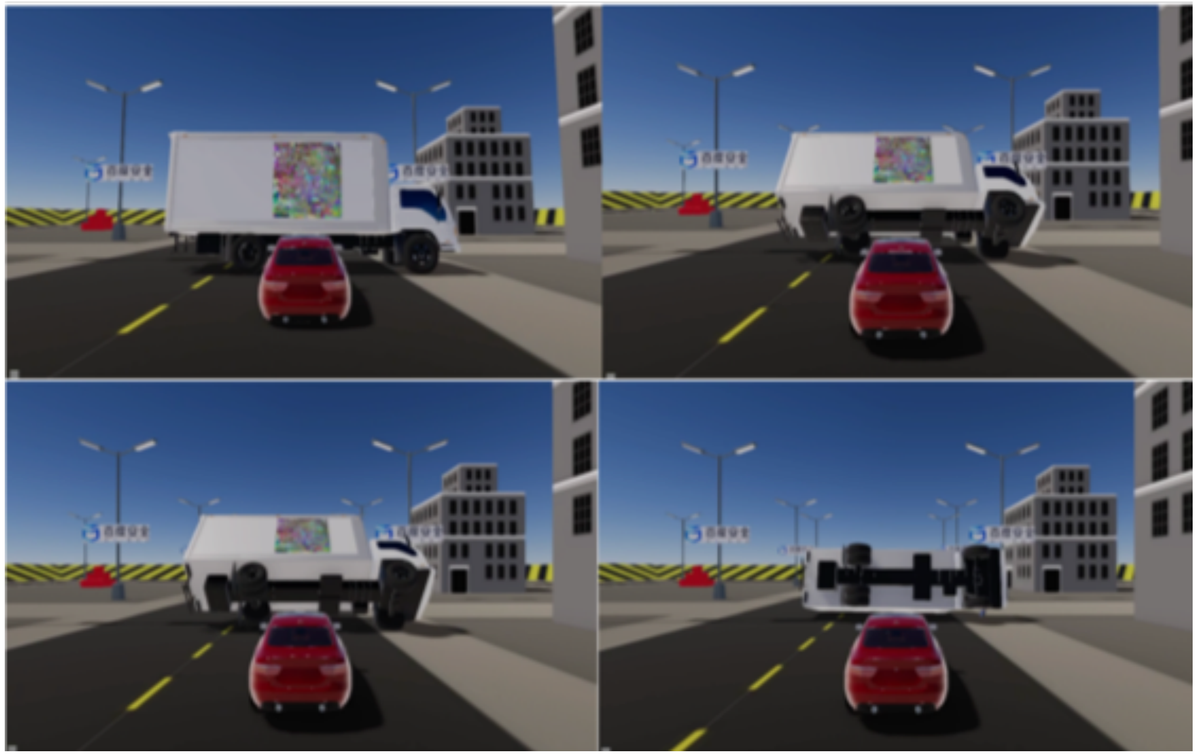
Traffic Scenario 03: Obstacle avoidance without prior action

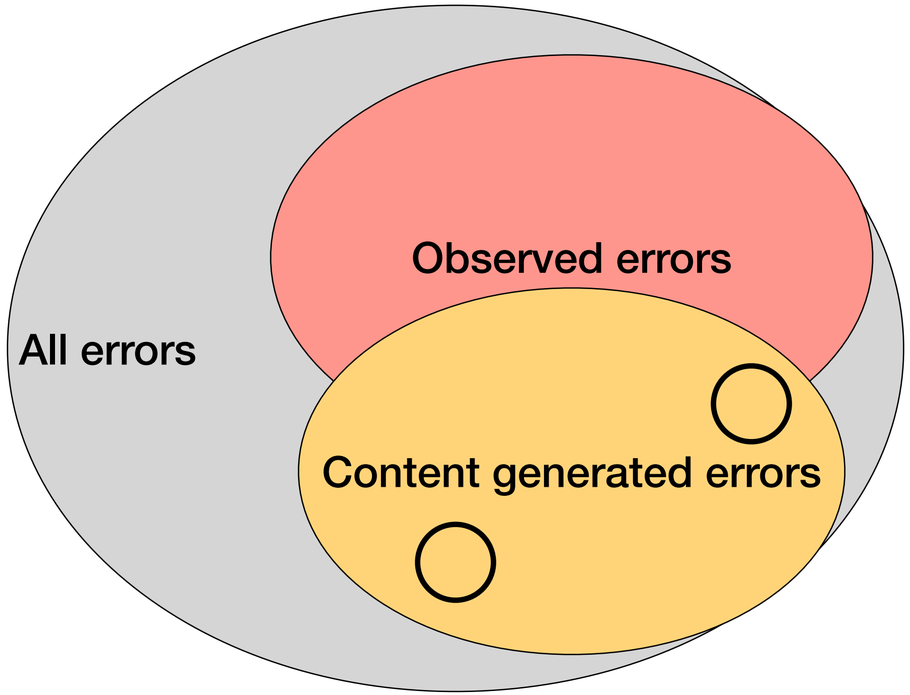
- **Definition**: The ego-vehicle encounters an obstacle / unexpected entity on the road and must perform an emergency brake or an avoidance maneuver.



Other Challenges Not Anticipatory

Not Focused on Error Detection





Autonomous Vehicle Limitations

- Complexity
  - Complex system build out of sensors, opaque software, and machinery.
  - It’s difficult to trace back what happened.
- Opaqueness
  - Proprietary mechanisms
  - Computer vision systems that are too opaque and dense to understand.

Current Approaches for Robust AVs

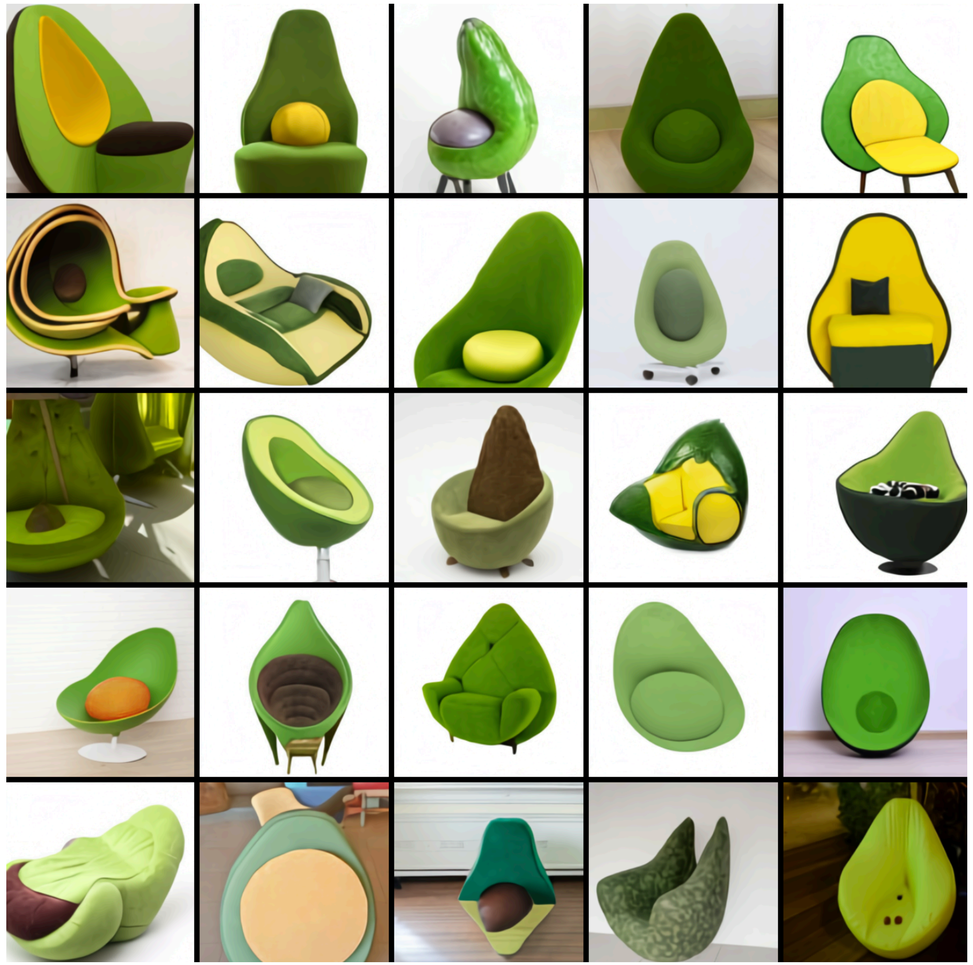
1. Error and failure analysis is **post mortem and reactive** instead of anticipatory

2. Explanations are a **post mortem** tool.

3. **Lack of Redundancy**: Unlike aircrafts (that purposely has components that are repeated), autonomous vehicles that rely entirely on a single system for perception (e.g., Tesla camera system) and it is prone to failure and error.

Approach: Content Generation

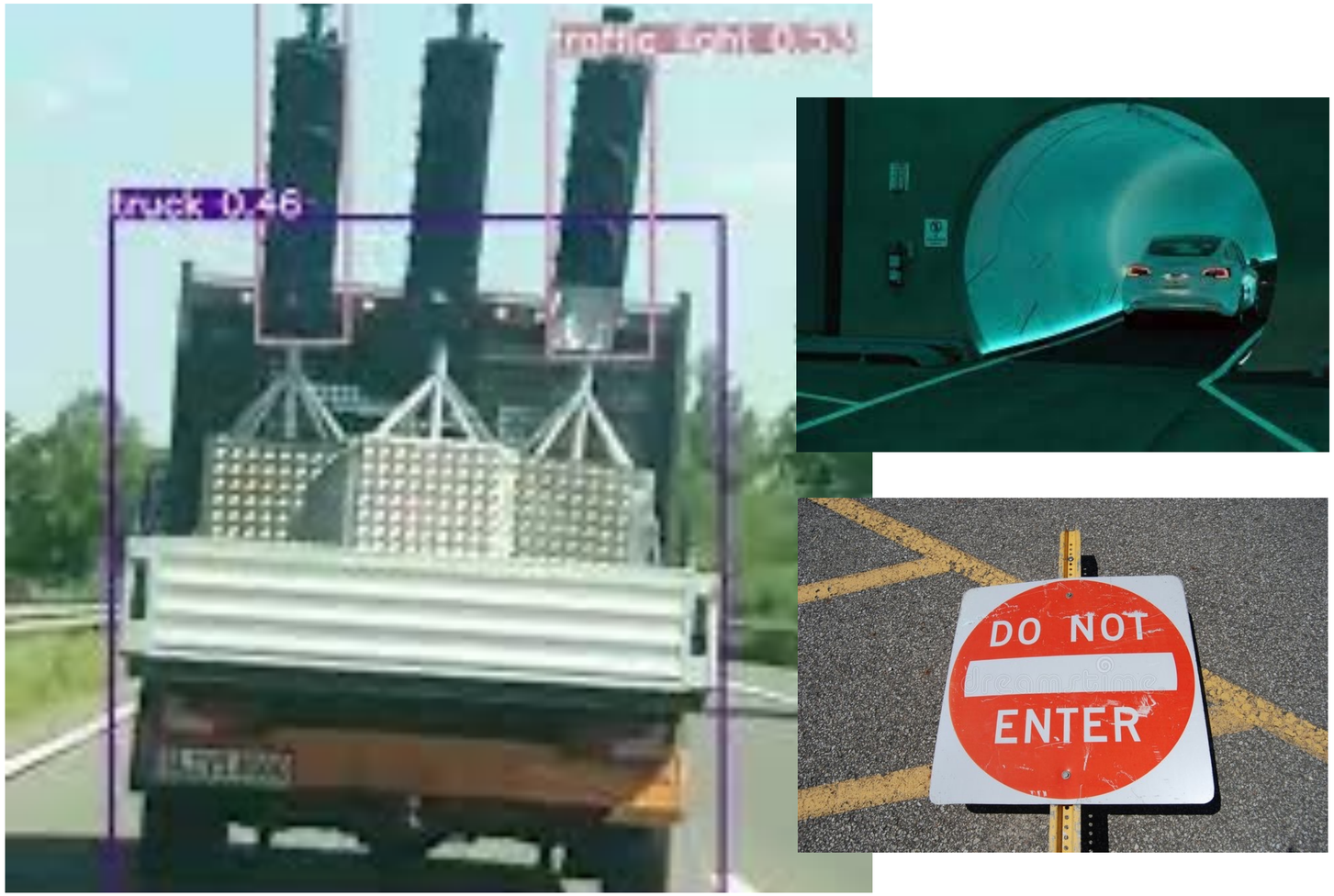
Anticipatory Thinking Layer for Error Detection



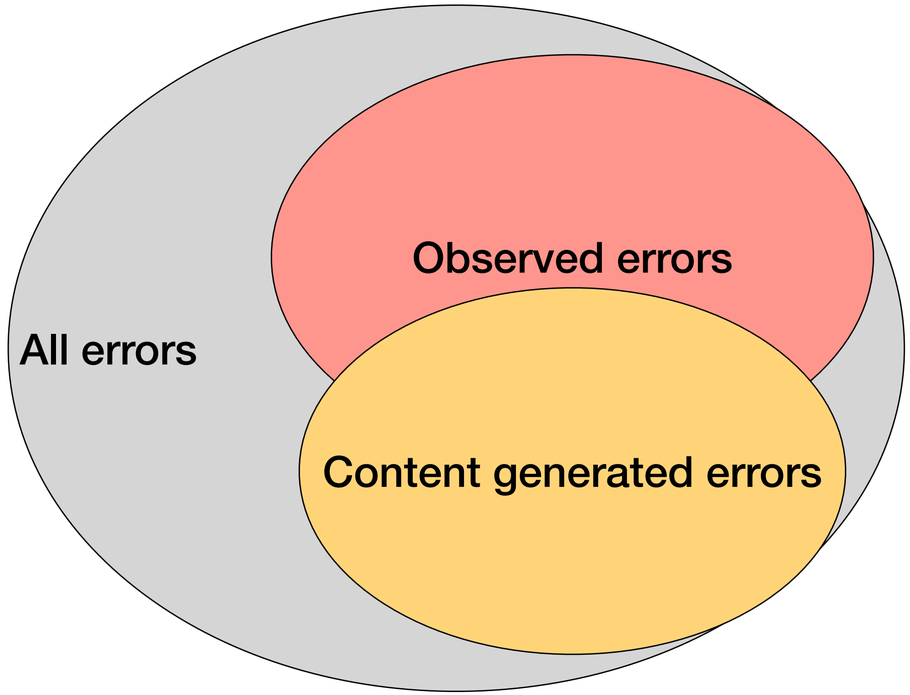
DALL-E Generates “A chair in the shape of an avocado”

Synthetic images produced by StyleGAN, a GAN created by Nvidia researchers.

Need for Context



“Realistic” Adversarial examples

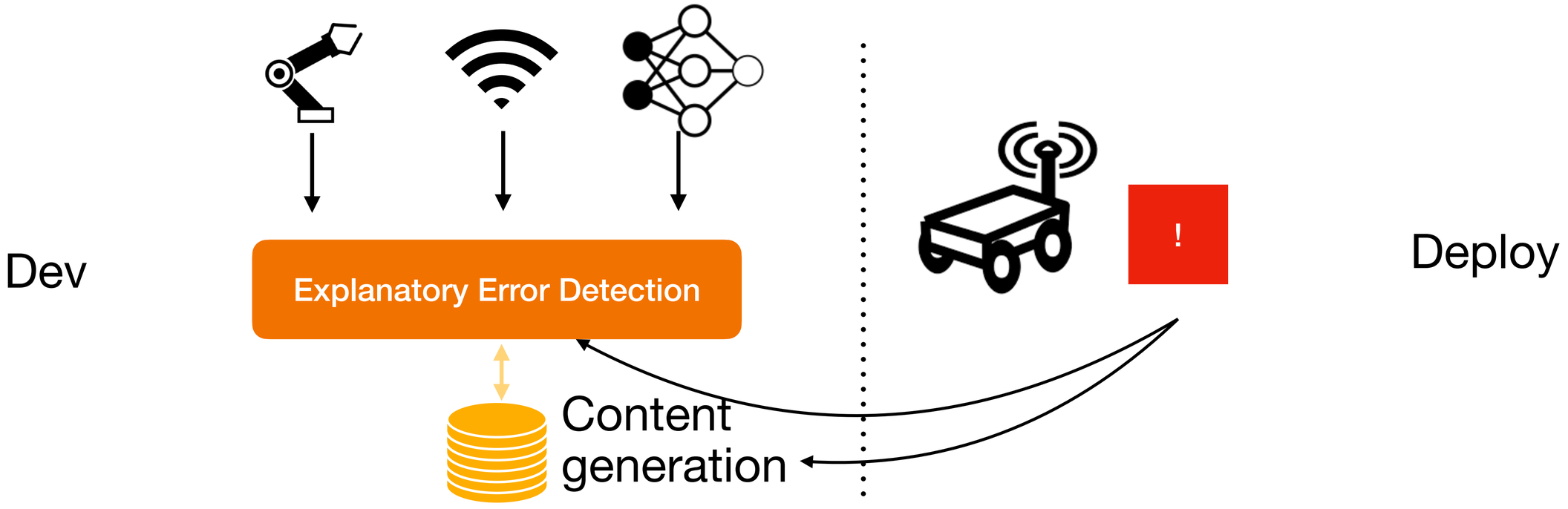


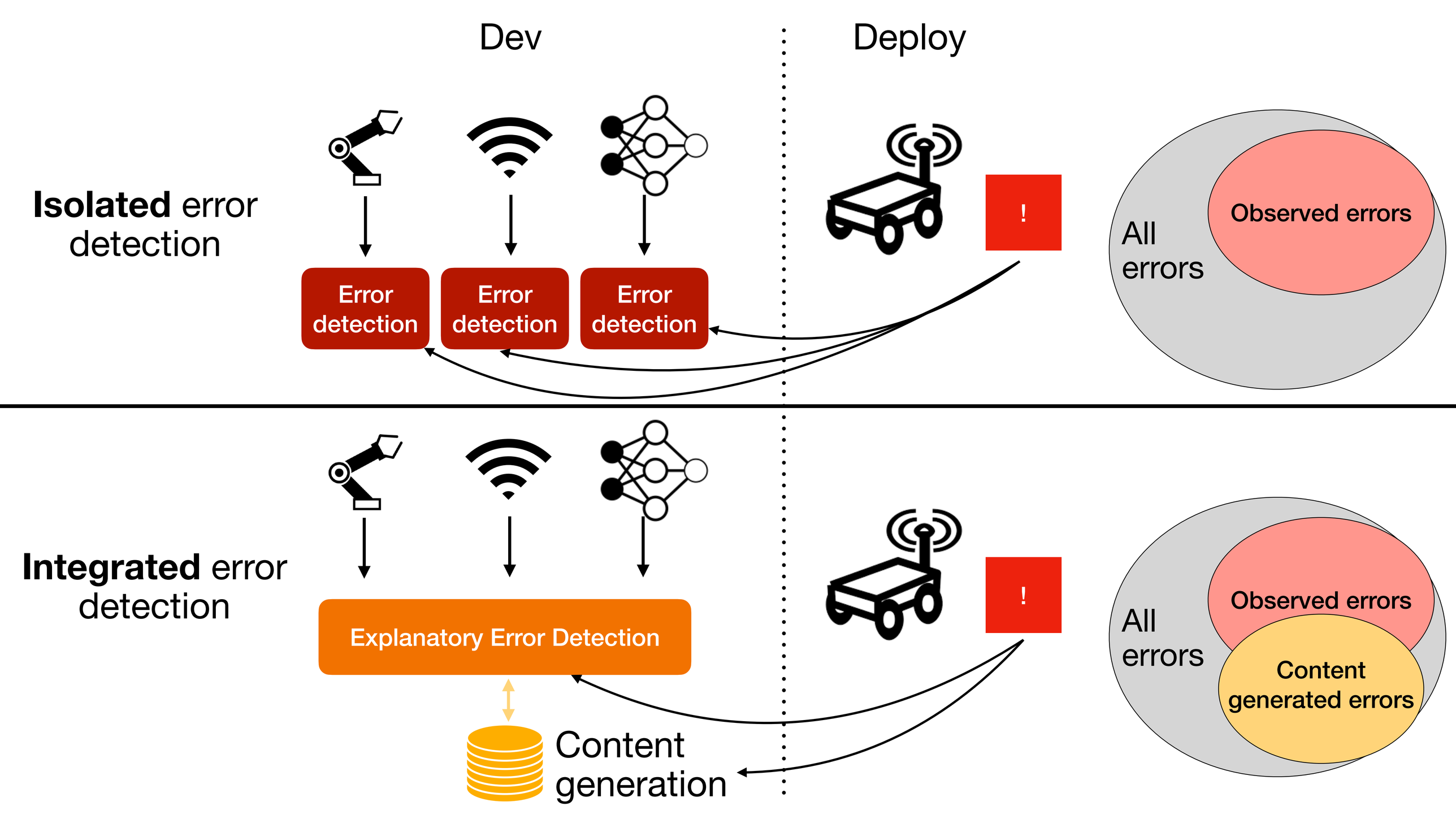
Approach: How it Works

Use Adversarial Images in Dev Testing

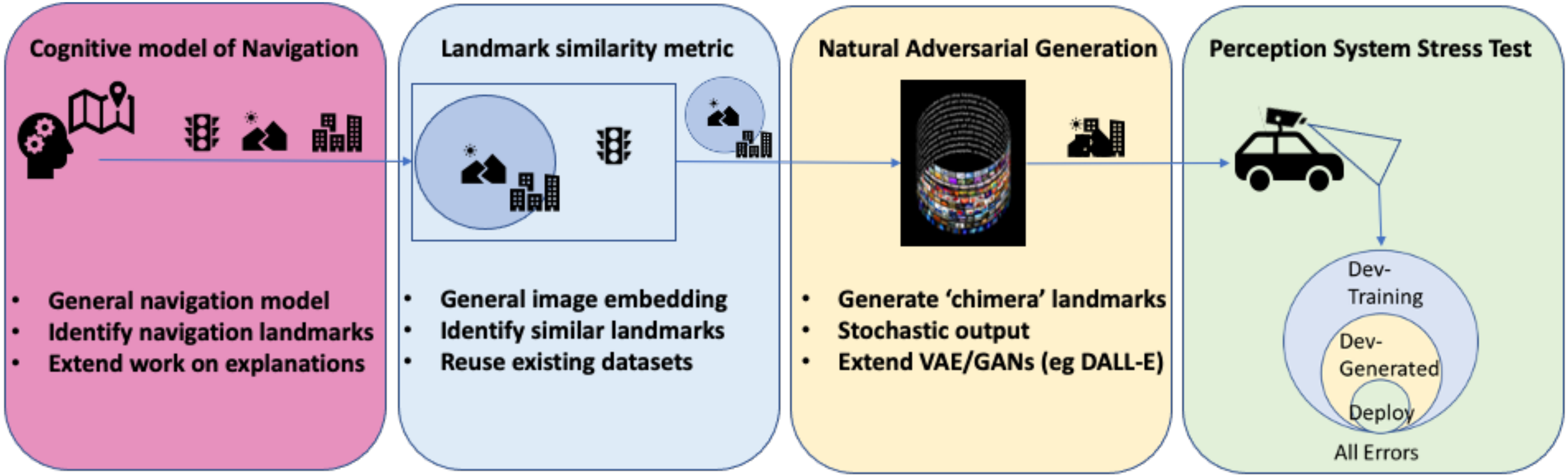
- Solution: Use a cognitive architecture that helps to anticipate and understand these failure cases.

- Assess autonomous vehicles for their risk management capabilities **before** being deployed and provide incident level risk management explanations in human readable form.




Dev
Deploy
Isolated error detection
Error detection
Error detection
Error detection
!
All errors
Observed errors
Integrated error detection
Explanatory Error Detection
Content generation
!
All errors
Observed errors
Content generated errors

Larger Approach



Impact

Anticipatory Thinking Layer for Error Detection

- Goal - Develop methods that a priori can explain an autonomous vehicle's ability to manage the risks stemming from errors in perceiving their environment.
- One possible solution is to explain why the autonomous behavior is safe (or risky, trustworthy, etc.) or not.
- Impact - Consumer confidence and safety features, appropriate legal and regulatory oversight.

.