US DOT National
University Transportation Center for Safety

Certification of Connected and Automated Vehicles for Vulnerable Road Users

Ding Zhao

Final Report – 7/30/2025

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under [grant number 69A3552344811 / 69A3552348316] from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Certification of Connected and Automated Vehicles for Vulnerable Road Users		**5. Report Date** July 30, 2025
		6. Performing Organization Code
7. Author(s) Miao Li, Haohong Lin, Ding Zhao		**8. Performing Organization Report No.**
9. Performing Organization Name and Address Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213		**10. Work Unit No.**
		11. Contract or Grant No. Federal Grant # 69A3552344811 / 69A3552348316
12. Sponsoring Agency Name and Address US DOT		**13. Type of Report and Period Covered** Final Report (July 1, 2024 – June 30, 2025)
		14. Sponsoring Agency Code USDOT
15. Supplementary Notes Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.		

16. Abstract			
This project aims to build a platform that can generate and deploy realistic critical traffic scenarios involving VRUs for physical testing. We seek to address this requirement in two stages. The first stage explores the design of an automatic toolkit, which automatically reconstructs critical scenarios from real-world traffic accidents. The second stage explores the design of the physical evaluation platform—designing and deploying robots that faithfully replay those scenarios.			
17. Key Words AI, autonomous vehicles, safety, evaluation		**18. Distribution Statement** No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161. Enter any other agency mandated distribution statements. Remove NTIS statement if it does not apply.	
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 7	**22. Price**

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Problem

Artificial intelligence (AI) has become a key powerhouse for the development of autonomous vehicles (AVs) and connected autonomous vehicles (CAVs). As one limitation of AI systems is unstable performance in corner cases, it is essential to establish a comprehensive evaluation scheme. For AI-based AVs and CAVs, a typical group of corner cases involves traffic scenarios with vulnerable road users (VRUs), such as pedestrians, motorists, and cyclists, particularly when they exhibit unpredictable behaviors (for example, a child unaware of traffic rules). Evaluating AVs and CAVs under these scenarios is challenging due to the scarcity of real-world examples and the lack of methods to reproduce these critical scenarios in physical tests. Therefore, the generation of critical scenarios and their physical evaluation has been a focus of industry and academia over the past decade.

This project aims to build a platform that can generate and deploy realistic critical traffic scenarios involving VRUs for physical testing. We seek to address this requirement in two stages. The first stage explores the design of an automatic toolkit, which automatically reconstructs critical scenarios from real-world traffic accidents and generates relevant scenarios that carry similar critical factors. The second stage explores the design of the physical evaluation platform, enabling robots to faithfully replay those scenarios. The primary objectives include:

1. Establish a simulation environment that is compatible with both commonly used scenario script formats and our custom configuration.
2. Develop a critical scenario generation toolkit that can generate a group of reconstructed scenarios based on given real-world traffic accidents, as well as create counterfactual scenarios and augmentations.
3. Design the intelligent robot, plus a standardized VRU robot platform to mimic the generated behaviors.
4. Design a testing platform that combines these components and enables both virtual and physical AV and CAV evaluation, particularly focusing on the safety of VRUs.

Methodology and Findings

For the first objective, we set up the environment using esmini. The environment supports parallel scenario replay with a capacity for over 3,000 concurrent threads. This environment is used not only for scenario rendering and computation, but also for replaying the scenario and calculating the route and trajectory for robot execution.

For the second objective, we created the crash agent. We leverage the spatial reasoning ability of vision–language models (VLMs) and the instruction-following capabilities of large language models (LLMs) to generate critical scenarios based on real-world traffic accidents. We use the National Highway Traffic Safety Administration (NHTSA) Crash Investigation Sampling System dataset [1] as our source of realistic critical scenarios. It includes textual descriptions, structured numerical data, and image representations. We identified two major challenges, namely, limited grounding and spatial reasoning in VLMs, which can cause misidentification of vehicle positions and irrelevant scenario generation, and incomplete understanding of complex parameter spaces of LLMs, which leads to scenario scripts missing desired parameters.

To address the first challenge, we proposed a visual-tree-of-thought method. Motivated by the observation that VLMs excel when interpreting simple images but falter on complex scenes, we hierarchically segment each accident sketch image, first by road layout, then by lane edges, VRU positions, and VRU behaviors. At the bottom of each tree, each segment contains only the minimal objects needed for accurate content extraction. Processing these segments in a bottom-up sequence along with the segmentation information allows us to aggregate reliable spatial information at each level until the full scene is processed, thereby exploiting the VLM's content-understanding strengths to compensate for its spatial-reasoning weaknesses.

For the second challenge, we proposed a road agent and a scenario agent. We observed that LLMs lack domain-specific knowledge of complex parameter spaces and struggle to follow verbose parameter instructions, which leads to incorrect parameter assignments. To address this, we developed the three complementary toolboxes, namely the road layout toolbox, scenario toolbox, and parameter search toolbox, to streamline scenario generation. The road layout and scenario toolboxes supply a rich set of traffic primitives, allowing the LLM to assemble detailed critical scenarios without wrestling directly with low-level parameters. The parameter search toolbox then uses a goal-conditioned genetic algorithm to fine-tune those parameters until the scenario produces the desired outcome, either a collision or a successful avoidance maneuver. By orchestrating these toolboxes, we built a road agent and a scenario agent that autonomously generate both the geometric layout and behavioral scripts needed for realistic, parameter-correct critical scenarios.

In addition to generating critical scenarios based on the real-world crash dataset, we also use the crash agent to identify which road user is responsible for the traffic accident. This information can be used as guidance to generate more critical scenarios that are dominated by the same critical factors and counterfactual scenarios that contain no collisions.

To summarize our work targeting the second objective, we identify two challenges in the critical scenario generation process and propose a crash agent to generate semantically realistic scenarios based on the real-world crash report, and to assess accident liability, identifying the responsible road user.

For the third objective, we surveyed standard robot platforms and selected a quadrupedal robot capable of carrying cost-effective dummies so it can faithfully reenact generated scenarios in our physical test environment. Drawing on our expertise in dynamic locomotion, autonomous navigation, and VLM-based tool use, we designed the robot execution pipeline, including positioning and interaction routines to mimic vulnerable road-user behaviors with high fidelity.

Conclusion

We propose a unified framework that leverages the strengths of VLMs and LLMs to build a crash agent capable of generating realistic critical scenarios from real-world accident records, which is summarized in an under-review conference paper. Once generated, these scenarios are designed to be executed by our quadrupedal-robot platform to reenact traffic accidents in a controlled physical testing environment. Our work makes two primary contributions. First, we implement a novel method for transforming unstructured crash reports into parameterized test scenarios and use this method to build a VRU-centric crash dataset. Second, we develop an end-to-end testing framework that enables AV and CAV evaluation both in high-fidelity simulation and through physical experimentation.

References

[1] https://crashviewer.nhtsa.dot.gov/