CENTER FOR CONNECTED AND
AUTOMATED TRANSPORTATION

Final Report #82
February 2026

CAV Pilot Development and Deployment in Midwest Winter

Yupeng Zhou
Can Cui
Zichong Yang
Juntong Peng
Ziran Wang

PURDUE
UNIVERSITY®

CENTER FOR CONNECTED AND AUTOMATED TRANSPORTATION

Report No. 82
Report Date: January 2026
Project Start Date: June 2024
Project End Date: May 2025

CAV Pilot Development and Deployment in Midwest Winter

Yupeng Zhou
Graduate Researcher
Can Cui
Graduate Researcher
Zichong Yang
Graduate Researcher
Juntong Peng
Graduate Researcher
Ziran Wang
Professor
Purdue University

CENTER FOR CONNECTED
AND AUTOMATED
TRANSPORTATION

DISCLAIMER

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant No. 69A3551747105 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. 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 under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Suggested APA Format Citation:

Zhou, Y., Cui, C., Yang, Z., Peng, J., and Wang, Z. (2026). CAV Pilot Development and Deployment in Midwest Winter, CCAT Report No. 82, The Center for Connected and Automated Transportation, Purdue University, W. Lafayette, IN.

Contacts

For more information:

Ziran Wang
550 Stadium Mall Drive
West Lafayette, IN 47907
Phone: 765-494-2154
Email: ziran@purdue.edu

CCAT
University of Michigan Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48152
uumtri-ccat@umich.edu
(734) 763-2498

Technical Report Documentation Page

1. Report No. 82

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
CAV Pilot Development and Deployment in the Midwest Winter

5. Report Date
January, 2026

6. Performing Organization Code
Enter any/all unique numbers assigned to the performing organization, if applicable.

7. Author(s)
Yupeng Zhou https://orcid.org/0009-0000-3355-0955
Can Cui https://orcid.org/0009-0009-7082-3444
Zichong Yang https://orcid.org/0009-0005-8019-3539
Juntong Peng https://orcid.org/0009-0007-1142-3067
Ziran Wang, Ph.D. https://orcid.org/0000-0003-2702-7150

8. Performing Organization Report No.

9. Performing Organization Name and Address
Center for Connected and Automated Transportation
University of Michigan Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48109

Purdue University
550 Stadium Mall Drive
West Lafayette, IN 47907

10. Work Unit No.

11. Contract or Grant No.
Contract No. 69A3551747105

12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
1200 New Jersey Avenue, SE
Washington, DC 20590

13. Type of Report and Period Covered
Final Report (June 2024-May 2025)

14. Sponsoring Agency Code
OST-R

15. Supplementary Notes
Conducted under the U.S. DOT Office of the Assistant Secretary for Research and Technology's (OST-R) University Transportation Centers (UTC) program.

16. Abstract

This project addresses the critical challenges of deploying Connected and Automated Vehicles (CAVs) in adverse Midwest winter conditions by developing a novel hierarchical framework that integrates Large Vision-Language Models for robust, context-aware decision-making. Real-world pilot experiments demonstrate that this architecture significantly enhances vehicle safety and personalization, effectively translating abstract human commands into precise control adjustments on low-friction, snow-covered surfaces. Furthermore, a specialized high-fidelity dataset capturing unique environmental edge cases was established to bridge the domain gap in existing resources and support the future development of resilient all-weather autonomous technologies.

17. Key Words
Autonomous vehicles, connected vehicles, and artificial intelligence

18. Distribution Statement
No restrictions.

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
27

22. Price

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

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

In recent years, significant progress has been made in the development of autonomous driving systems, particularly in advances in perception, planning, control, and connectivity. These emerging breakthroughs are enabling large-scale commercialization in the robotaxi industry, with deployments in cities such as Phoenix, San Francisco, and Austin. There are extensive validation miles accumulated by industry leaders such as Waymo, Zoox, and Tesla. However, some common characteristics emerge across these deployments, particularly large-scale environmental features: a moderate, warm-temperate climate and predictable weather without extreme conditions. In the American Midwest, by contrast, winter storms, snow, and icy roadways are the norm for several months of the year. This extreme weather potentially creates a number of edge cases that bring serious challenges for connected and automated vehicles (CAVs) and threaten their safety and robustness. The inability of current CAV perception and control stacks to operate reliably in snow, sleet, and freezing rain threatens to create a technological divide. If CAVs remain operational only during favorable weather, the benefits of autonomy will be limited to a small area, hindering the Midwest’s modernization of its transportation infrastructure.

The primary challenge to deploying CAVs in the Midwest is the dramatic performance degradation of perception and control stacks under winter conditions. Standard autonomous driving perception pipelines rely on deep learning techniques trained primarily on clear-weather data, resulting in significant performance degradation when the operational design domain (ODD) shifts to winter conditions, particularly for snow and ice [1]. At the same time, adverse weather fundamentally alters the physics of sensing; LiDAR wavelengths suffer from backscatter on snowflakes, creating "phantom obstacles," while cameras lose the high-frequency edge information required for lane keeping due to whiteout conditions and low contrast [2]. Furthermore, the loss of semantic landmarks (where lane lines and curbs are buried under snow) causes high-definition (HD) map-matching algorithms to fail, leaving the vehicle without a localized reference frame. This perception uncertainty is coupled with vehicle motion planning and control, as the stable, limited handling region decreases [3]. In detail, low road-surface temperature and friction reduce tire performance, resulting in a loss of stability even during normal maneuvers. The slim safety region requires more sophisticated, precise, and robust decision-making, planning, and control to address this situation.

Recent advances in Vision Language Models (VLMs) reveal great potential to address multiple tasks in autonomous driving, including perception, planning, high-level semantic understanding, decision-making, data generation, and End-to-End autonomous driving [4][5]. Compared with traditional autonomous driving systems, which are trained on limited datasets, VLM-integrated CAVs can acquire depth in real-world knowledge for robust driving behavior, particularly in unfamiliar or uncommon winter conditions. As a result, VLM-integrated CAVs can make appropriate decisions in unfamiliar or uncommon situations, potentially increasing safety. VLMs are also effective at understanding abstract human instructions and the feelings behind them. By utilizing the contextual command, the CAV provides personalization and alignment with human-like driving performance.

This project focuses on deploying a connected and automated vehicle system equipped with innovative VLMs to navigate complex winter scenarios in the Midwest. This approach aims to improve safety by reducing weather-related incidents, enhance mobility through shorter travel time, and promote equity by improving human-autonomy teaming effectiveness for diverse populations, particularly in extreme weather conditions. Real-world demonstrations are conducted by a fully equipped CAV on a professional test track, generating vital data and experience to inform future CAV development and deployment in the Midwest, especially under extreme weather conditions.

CHAPTER 2. PLATFORM DESIGN

To incorporate large vision language models, we propose a hierarchical real-world platform specialized in VLM-integrated autonomous driving systems, as shown in Fig. 1. Specifically, our platform features have: a lightweight, structured, and low-latency middleware pipeline specialized for seamless VLM integration; a hierarchical modular architecture enabling flexible substitution between conventional and VLM-based autonomy components, providing exceptional deployment flexibility for rapid experimentation.

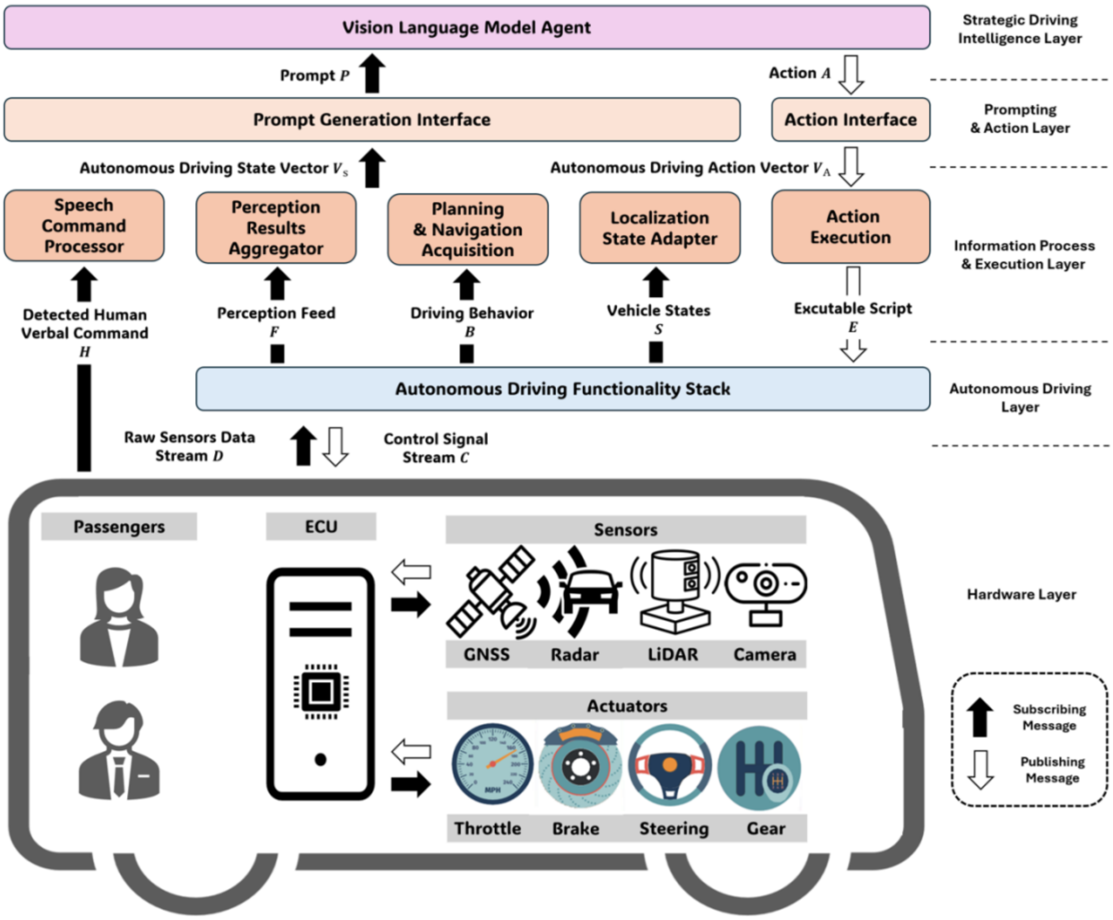


Fig. 1. An overview of the proposed hierarchical architecture for VLM-integrated autonomous driving in the testing vehicle.

2.1 Strategic Driving Intelligence Layer

This layer functions as the cognitive core of the system, hosting the VLM to manage high-level decision-making. Designed for flexibility, it supports both onboard and cloud-based VLM deployments, allowing for versatile experimentation. The primary input is a structured prompt that synthesizes multi-modal information, including perception results, vehicle states, real-time driving behaviors, and human feedback. By processing this rich context, the VLM determines appropriate driving strategies and outputs high-level action decisions, which are subsequently encoded into an action vector. This architecture effectively offloads complex strategic reasoning from lower-level modules—which continue to handle stable perception and control tasks—while offering a modular design that allows researchers to easily replace models or adjust prompt/action interfaces for different testing requirements.

2.2 Prompting & Action Interface Layer

Acting as the intermediary between the high-level intelligence and mid-level processes, this layer manages a bidirectional workflow. On the input side, the Prompting Interface aggregates relevant data regarding the vehicle's state and environment to construct structured prompts for the VLM. On the output side, the Action Interface translates the VLM's abstract reasoning into a concrete autonomous driving action vector containing specific parameters and instructions executable by downstream modules.

The workflow proceeds sequentially: the system first processes raw data into a state vector via the prompt interface; this is passed to the VLM to generate a decision; finally, the decision is converted into an action vector through the action interface. This design allows for independent tuning of prompt structures and action formats, enabling researchers to optimize the interaction between the VLM and the control system and explore how variations in prompt design influence vehicle behavior.

2.3 Information Process & Execution Layer

This layer implements a dual-phase approach to interpret and act upon VLM directives:

- Information Processing Phase: Dedicated modules transform autonomous driving statistics and messages into a unified Autonomous Driving State Vector. This includes a Speech Command Processor to capture verbal instructions from passengers, as well as sub-processors that aggregate perception feeds (such as object trackers or BEV images), planned trajectories, and vehicle states (speed, position, and map data). Researchers can dynamically configure which data points are included in this vector to suit specific testing scenarios.
- Action Execution Phase: This phase converts the VLM’s action vector into

executable maneuvers using specialized scripts. These scripts interact with exposed ports in the lower functional stack to modify inputs, outputs, or parameters within the perception, planning, and control modules. For instance, if the VLM commands a lane change, the execution module selects the appropriate trajectory and adjusts speed parameters accordingly. Crucially, this module optimizes and validates all VLM decisions against safety constraints before engaging the motion control layer.

2.4 Autonomous Driving Layer

The foundational layer houses the core autonomous driving functionalities, including object detection, prediction, decision-making, trajectory planning, and motion control. To facilitate command execution at higher layers, we developed a comprehensive set of interfaces and ports for the perception, planning, and control modules. These interfaces expose various parameters and data streams, offering developers the scalability and flexibility to create or fine-tune settings tailored to specific testing objectives. This allows the upper layers to dynamically influence the vehicle's behavior while relying on the robust, fundamental capabilities of the core stack.

During our real-world deployment, Autoware.AI [6] was implemented as the primary autonomous driving functionality backbone due to its open-source flexibility, comprehensive software stack, and robust ROS-based [7] architecture. We deploy YOLOv5 [8] as the primary perception module for object detection and recognition. Given an input image collected by the front camera, the model outputs the bounding boxes, class labels, and confidence scores of detected objects. These perception results may be processed and integrated by the subsequent module. The LiDAR point-cloud data is utilized for localization. Particularly, the Normal Distributions Transform (NDT) matching algorithm [9], which estimates the vehicle's pose by aligning the current LiDAR scan with a pre-built map. The NDT algorithm determines the vehicle's pose by optimizing an objective function that evaluates the alignment between the current scan and the map. Specifically, the algorithm seeks a spatial transformation that maximizes the probability of the current scan points fitting within the statistical covariance distributions of the pre-built map points. To further enhance localization accuracy and robustness, additional sensors, including GNSS and IMU, are integrated to provide precise measurements of the vehicle's position, velocity, acceleration, and orientation.

For vehicle motion planning and control, we adopt a decoupled control strategy, using a PID controller for longitudinal control and Model Predictive Control (MPC) for lateral control [10]. The upper planner computes trajectories consisting of a sequence of waypoints, defining the vehicle's states and speed profile to ensure precise navigation through the environment. The PID controller adjusts the throttle and brake commands based on the error between the desired and actual vehicle speed, and the MPC controller for lateral control generates the optimal steering signal.

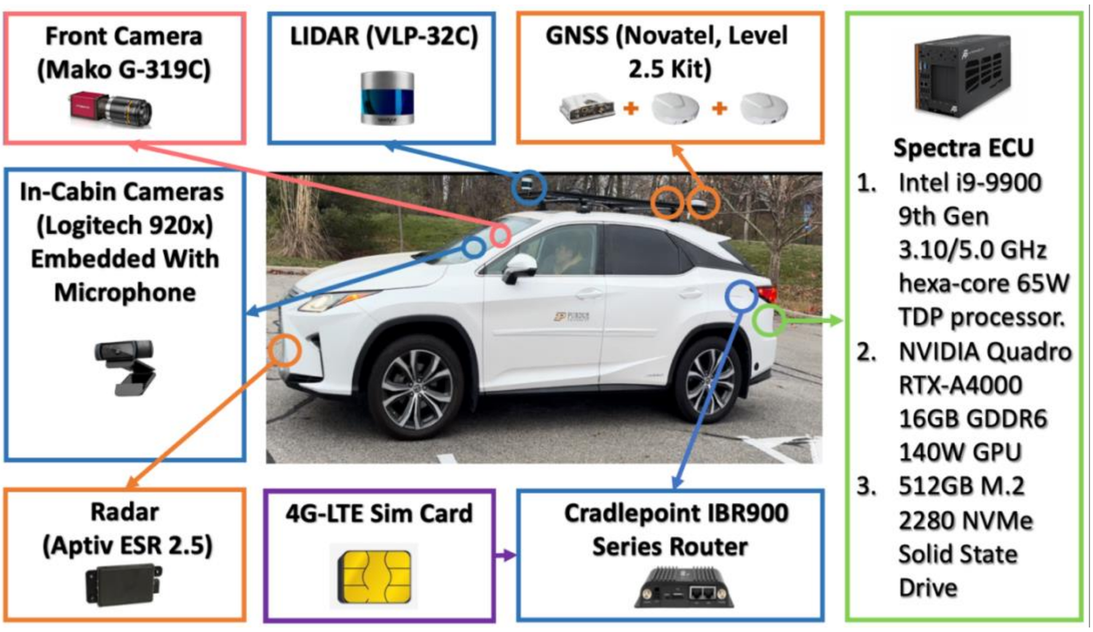


Fig. 2. The hardware layer setup for the CAV.

2.5 Hardware Layer

An overview of the vehicle hardware setup is shown in Fig. 2. This modified LX 450h connected autonomous driving vehicle is equipped with comprehensive subsystems, including a power distribution system, connection router, sensor suite, PACMod drive-by-wire system, and a high-performance onboard Electronic Control Unit (ECU).

Sensors configuration and management. The sensor suite on the vehicle includes a Velodyne VLP-32C LiDAR, a radar (Aptiv ESR 2.5 24V), two front-view cameras (Mako G-319C), and a real-time kinematic (RTK) correction-enabled GNSS Positioning Kit with an inertial measurement unit kit (Novatel Level 2.5 kit).

Connectivity & Computing. A Cradlepoint IBR900 Series Router embedded with a 4G-LTE AT&T SIM card is installed to provide a cellular connection for the vehicle. It offers a ping latency of 30-50 ms and bandwidths of 50-70 Mbps for uploads and 50-60 Mbps for downloads. A high-performance ECU is installed in the vehicle's trunk. The core computational configuration is below: Intel i9-9900 9th Gen 3.10/5.0 GHz Hexa-core 65W processor with eight cores and 16 threads, 64GB RAM, NVIDIA Quadro RTX-A4000 16GB GPU, and 512GB NVMe solid-state drive.

By-Wire Control. The onboard By-Wire control is a PACMod 3.0 System provided by AutonomouStuff [5]. This Drive-by-Wire system consists of two main subsystems: Speed and Steering Control (SSC) and the PACMod system.

CHAPTER 3. REAL-WORLD EXPERIMENT AND CASE STUDY

3.1 Real-world Experiment Setup

To evaluate the performance of the CAV performance in realistic driving scenarios, we design and deploy three independent testing scenarios in the heart of the Midwest, Indiana to conduct experiments: a three-lane highway, a two-lane intersection, and a parking lot track, as shown in Fig. 3. The highway and intersection tests are carried out at a dedicated test track located at 950 S 450 W, Columbus, IN, USA. This controlled environment allowed for the safe testing of high-speed maneuvers and complex traffic interactions. The parking lot track experiments are conducted at the North Stadium Parking Lot in West Lafayette, IN, USA, which provides a more confined space to assess the vehicle's low-speed maneuvering capabilities with more uncertainty.

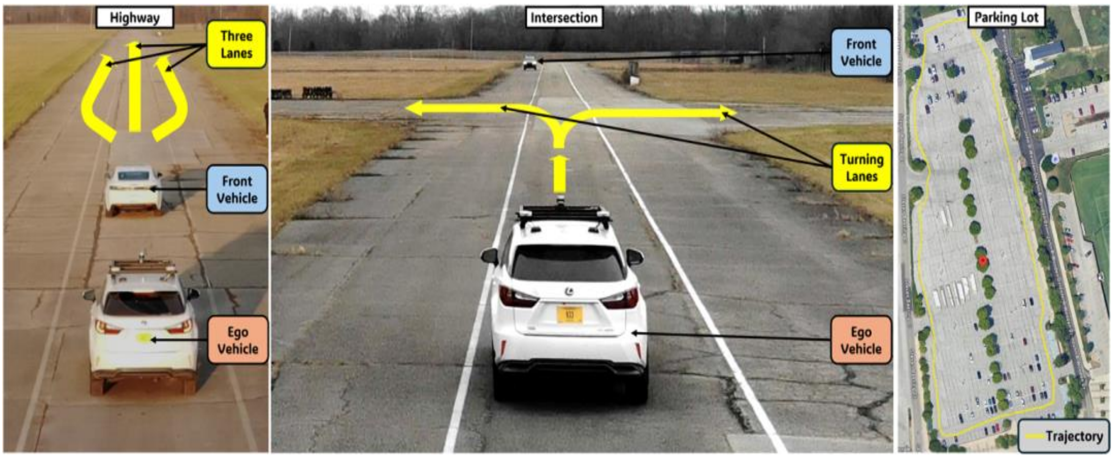


Fig. 3. Illustration of the real-world test environments.

We design a series of test scenarios for each environment, covering common driving situations. These test cases included:

- Highway: Maintaining a constant speed, performing lane changes, reacting to sudden slowdowns in traffic, and merging onto the highway from an on-ramp.
- Navigating or stopping, yielding to the opposite vehicle, and making unprotected left turns.
- Parking lot: Navigating narrow lanes, avoiding static and dynamic obstacles, and parallel parking.

3.2 Case I Personalized Autonomous Driving with Large Language Model

To address the high variability in autonomous driving performance under winter conditions, we present the Talk2Drive framework as a critical case study in personalized autonomous driving, as shown in Fig. 4. This pilot demonstrates the first successful integration of LLMs into a real-world autonomous vehicle to accurately interpret abstract human verbal commands and provide long-term, high-level personalization. This capability is particularly significant for CAV deployment in the Midwest, where extreme winter conditions demand high-level contextual awareness from the passengers. By integrating real-time weather data and historical user preferences, Talk2Drive can intelligently generate appropriate decisions for CAVs to maintain safety and comfort.

In the case study, a human's spoken instructions (I) are processed by cloud-based LLMs, which synthesize contextual data (C) from weather, traffic conditions, local traffic information, and perception results from the local end. Simultaneously, the system message (S) and the historical data (H) are sent to LLMs. Subsequently, the LLMs generate executable Language Model Programs (LMPs) and transmit them to the vehicle's Electronic Control Unit (ECU). These LMPs operate the actuation of vehicle controls, ensuring that the human's intent is translated into safe and personalized driving actions. A memory module archives every command (I), its resultant LMPs (P), and subsequent user feedback (F), ensuring continuous refinement.

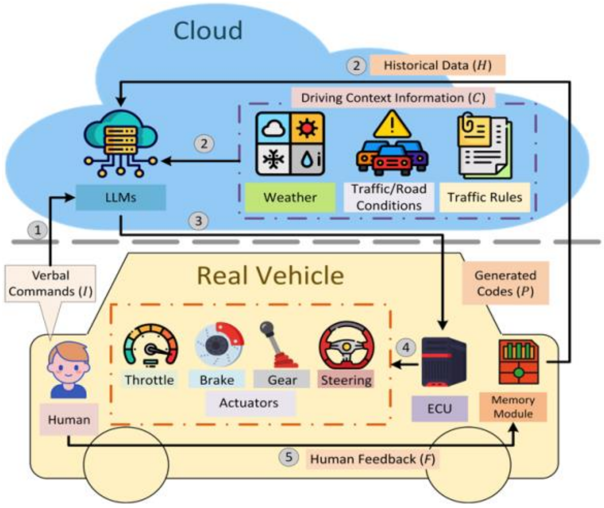
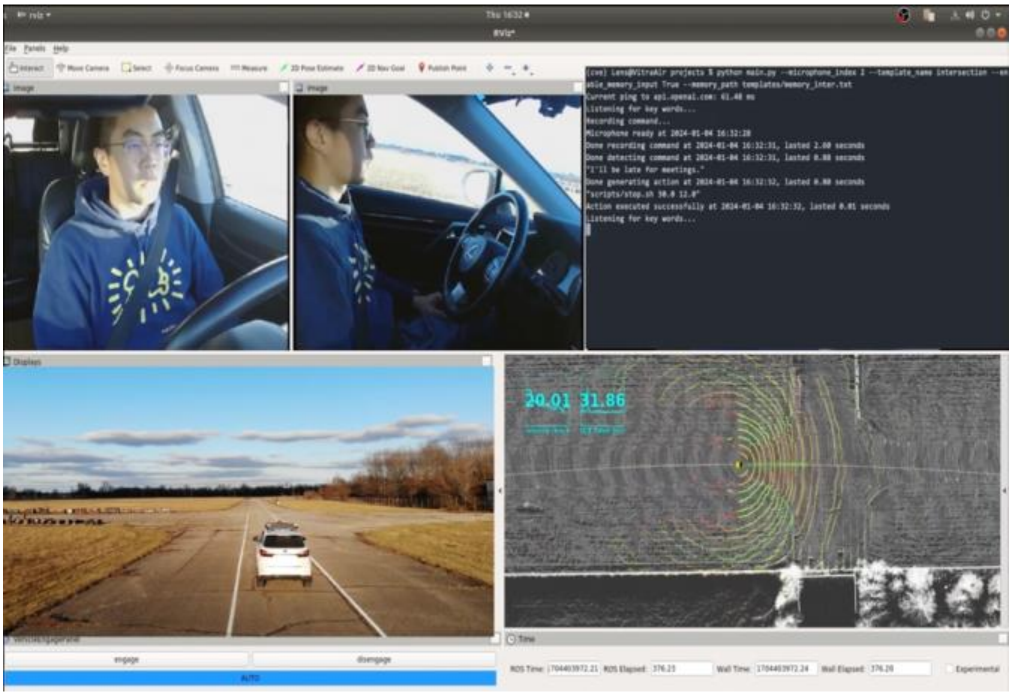


Fig. 4. (a) The Talk2Drive framework architecture for personalized autonomous driving using a large language model.



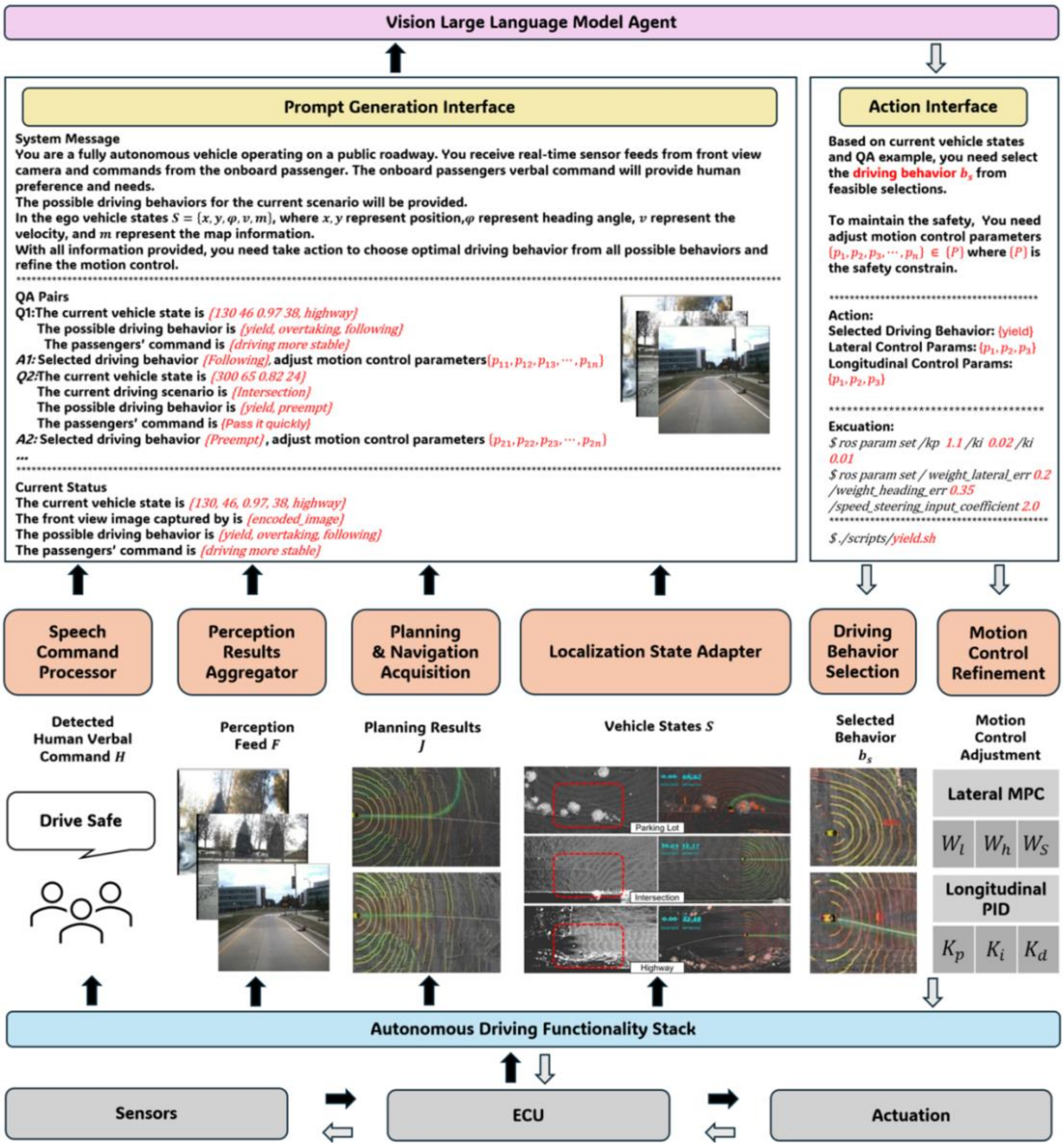
```
(cav) Lens@VitraAir projects $ python main.py --microphone_index 1 --template_name intersection --en
able_memory_input True --memory_path templates/memory_inter.txt
Current ping to api.openai.com: 41.48 ms
Listening for key words...
Recording command...
Microphone ready at 2024-01-04 16:32:28
Done recording command at 2024-01-04 16:32:31, lasted 2.00 seconds
Done detecting command at 2024-01-04 16:32:31, lasted 0.88 seconds
"I'll be late for meetings."
Done generating action at 2024-01-04 16:32:32, lasted 0.80 seconds
"scripts/stop.sh 30.0 12.0"
Action executed successfully at 2024-01-04 16:32:32, lasted 0.01 seconds
Listening for key words...
```

Fig. 5. (b) The experiment visualization for personalized autonomous driving using a large language model.

To evaluate the performance of the proposed Talk2Drive framework, comprehensive experiments are conducted on the real-world CAV across the three scenarios. In detail, the field experiments utilized a diverse number of participants, comprising approximately 61.4% male and 28.6% female drivers with varying levels of driving experience, to evaluate the Talk2Drive system under real-world conditions. To assess the framework's ability to interpret human intent, verbal instructions were categorized into three distinct levels of directness: Level I (direct and conventionally indirect commands), Level II (non-conventionally indirect with strong hints), and Level III (non-conventionally indirect with mild hints). The evaluation of the system was quantified through comprehensive metrics focused on safety, comfort, and personalization. Safety was assessed using Time to Collision (TTC) and speed variance, while ride comfort was measured through mean absolute acceleration and jerk. A critical component of the experiment was the measurement of the takeover rate, which served as the primary metric for quantifying human trust and the effectiveness of the system's personalization. The end-to-end latency is also measured for the entire workflow.

3.3 Case II VLM-Integrated CAV Pilot Deployment in the Winter Conditions

The initial Talk2Drive framework successfully demonstrated the capability to understand verbal commands along with driving contextual information to appropriate decision-making for the CAV. How to interpret the winter environment remains important, especially for the deployment in the Midwest. To address this, we propose a VLM-integrated framework that extends the original Talk2Drive architecture by adding a visual modality to the reasoning process. This approach allows the autonomous agent to function as a holistic decision-maker that correlates visual scene understanding with human intent, as shown in Fig. 6.



System Message
You are a fully autonomous vehicle operating on a public roadway. You receive real-time sensor feeds from front view camera and commands from the onboard passenger. The onboard passengers verbal command will provide human preference and needs.
The possible driving behaviors for the current scenario will be provided.
In the ego vehicle states $S = \{x, y, \varphi, v, m\}$, where x, y represent position, φ represent heading angle, v represent the velocity, and m represent the map information.
With all information provided, you need take action to choose optimal driving behavior from all possible behaviors and refine the motion control.

QA Pairs
Q1:The current vehicle state is *{130 46 0.97 38, highway}*
The possible driving behavior is *{yield, overtaking, following}*
The passengers' command is *{driving more stable}*
A1: Selected driving behavior *{Following}*, adjust motion control parameters $\{p_{11}, p_{12}, p_{13}, \cdots, p_{1n}\}$
*Q2:*The current vehicle state is *{300 65 0.82 24}*
The current driving scenario is *{Intersection}*
The possible driving behavior is *{yield, preempt}*
The passengers' command is *{Pass it quickly}*
A2: Selected driving behavior *{Preempt}*, adjust motion control parameters $\{p_{21}, p_{22}, p_{23}, \cdots, p_{2n}\}$
...

Current Status
The current vehicle state is *{130, 46, 0.97, 38, highway}*
The front view image captured by is *{encoded_image}*
The possible driving behavior is *{yield, overtaking, following}*
The passengers' command is *{driving more stable}*

Based on current vehicle states and QA example, you need select the **driving behavior b_s** from feasible selections.

To maintain the safety, You need adjust motion control parameters $\{p_1, p_2, p_3, \cdots, p_n\} \in \{P\}$ where $\{P\}$ is the safety constrain.

Action:
Selected Driving Behavior: {yield}
Lateral Control Params: $\{p_1, p_2, p_3\}$
Longitudinal Control Params: $\{p_1, p_2, p_3\}$

Excuation:
```
$ ros param set /kp 1.1 /ki 0.02 /ki 0.01
$ ros param set / weight_lateral_err 0.2 /weight_heading_err 0.35 /speed_steering_input_coefficient 2.0
```
```
$ ./scripts/yield.sh
```

Fig. 6. An illustration of the VLM integrated CAV framework in Case II.

To validate the proposed VLM-integrated framework under realistic winter conditions, we designed a series of comparative experiments across highway, intersection, and parking lot scenarios on the test track. We demonstrated several trials under clear weather and environments to rigorously test its understanding and adaptation capabilities. During each trial, the on-cloud VLM was queried at 3-second intervals with a structured prompt integrating real-time vehicle states, feasible planning trajectories, visual perception feeds, and passenger verbal commands.

The system's effectiveness was evaluated based on its ability to synthesize these multimodal inputs into high-level decisions, such as yielding or overtaking. To ensure robust quantitative analysis, comprehensive driving logs recording the vehicle States, safety matrices, comfort matrices, and VLM reasoning outputs were captured throughout the testing process.

3.4 Winter Dataset

In the first Talk2Drive case study, while the framework successfully demonstrated personalized autonomous driving by interpreting verbal commands into high-level decisions, its performance relied on language modality across the different models. In the second VLM-Integrated CAV case study, with the vision, Vision-Language Models can understand the physical environment and adapt to winter conditions like snowy weather. However, the robustness of the VLM is currently constrained by zero-shot [14] action without additional training on the data from real-driving scenarios. These limitations suggest that substantial latent potential within Vision-Language Models (VLMs) remains to be explored, contingent upon the availability of specialized training data. Thus, a specialized dataset for this Midwest Winter data set is necessary.

On the other hand, current open-source datasets, such as Waymo Open Dataset [15] or nuScenes [16], predominantly feature data collected in fair weather or mild urban environments (e.g., California, Singapore). They severely lack the "long-tail" edge cases characteristic of the Midwest, such as "whiteout" visibility, salt-covered lane markings, and ice-glazed sensors. Thus, the OOD problem still remains. To bridge this gap and solve the critical domain shift problem, where models trained on sunny data fail in snowy conditions, we collected a high-fidelity dataset specifically targeting these adverse conditions.

We utilized the same platform in our case studies introduced previously to conduct extensive data collection runs. The primary collection route involved a round trip from West Lafayette, IN, to Columbus, IN, traversing the I-65 corridor.

- **Route Coverage**: The dataset covers a round-trip distance of approximately 220 miles. This route provides a diverse mix of operational domains, ranging from the complex urban environments of the Purdue University campus (featuring roundabouts, crosswalks, and dense pedestrian traffic) to high-speed interstate driving on I-65, where vehicles frequently encounter heavy truck traffic and blowing snow from open agricultural fields.
- **Data Volume**: The campaign has amassed over 800 GB of synchronized data. This dataset focuses on adverse winter weather, providing density data that can be used for perception, planning, and E2E autonomous driving models.
- **Sensor Suite and Modalities**: With the comprehensive sensors and modalities of our real-world platform, the dataset provides a holistic, multimodal view of the environment. Two Mako G-319C cameras mounted on the front windshield capture a high-resolution video stream at up to 1080p at 10 Hz. A roof-mounted Velodyne VLP-32C LiDAR provides dense 3D point clouds for depth perception, while a front-bumper Delphi ESR 2.5 Radar offers robust object detection and tracking in front of the vehicle. A visualization of the different data is shown in Fig. 7.
- **Vehicle States:** Unlike many datasets that only provide trajectory data, we log raw CAN bus data (including steering angle, throttle position, and wheel speed) and high-precision location data from a Novatel GNSS/INS system. This allows researchers to correlate environmental perception directly with the vehicle's physical stability and control response on low-friction surfaces.

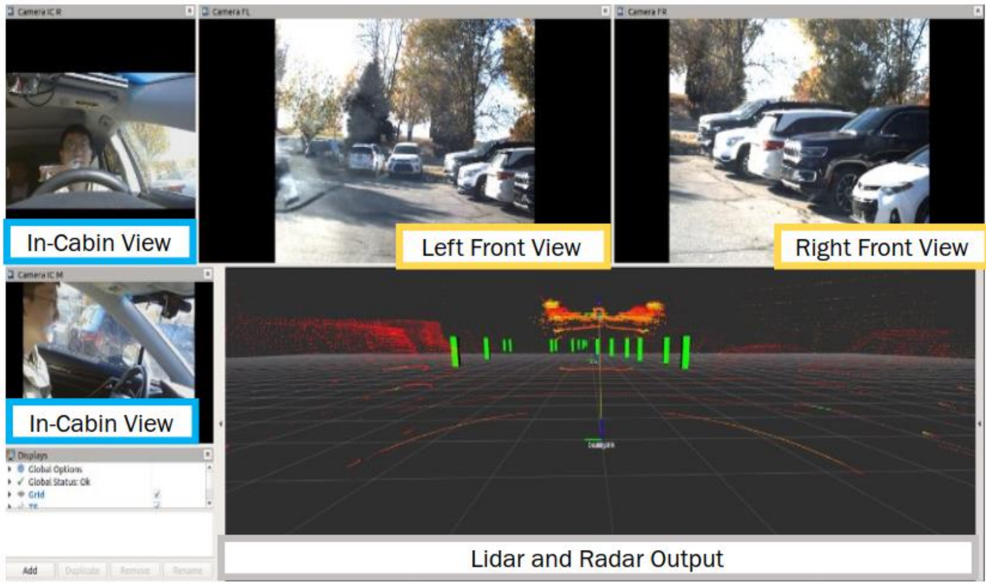


Fig. 7. A visualization of the sensor data for the Midwest winter autonomous driving dataset collection

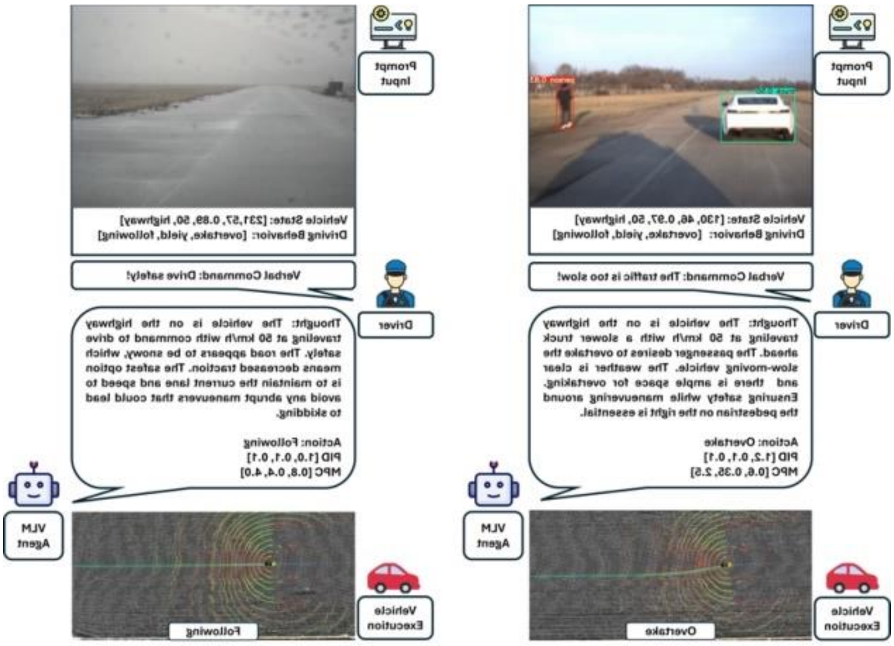
This dataset serves as a critical resource for the next generation of CAV development and deployment in the Midwest area. Its primary applications include robust perception training for winter extreme conditions and large VLM fine-tuning to enhance safety in extreme weather.

CHAPTER 4. FINDINGS

For Case I, personalized autonomous driving with LLMs, the field experiments demonstrated that the Talk2Drive framework significantly enhances autonomous vehicle performance across safety, comfort, and personalization metrics. Safety evaluations indicated that the system maintained a Time to Collision (TTC) well above the 1.5-second critical threshold, ensuring sufficient reaction time comparable to human drivers. In terms of comfort, the system's mean absolute acceleration and jerk remained within "Very Good" ride experience levels ($\leq 0.56\, m/s^2$ for acceleration and $\leq 2.94\, m/s^3$ for jerk), effectively matching or exceeding the baseline scores for human-driven performance. Furthermore, the end-to-end system latency, primarily driven by the GPT-4 API call, remained stable at approximately 1.6 seconds, which proved acceptable for the non-urgent maneuvers tested in highway, intersection, and parking scenarios.

The most notable finding was the substantial reduction in the takeover rate, which serves as a proxy for human trust and personalization. Compared to a keyword-trigger baseline, the Talk2Drive system reduced manual interventions by 56.3% to 100% across all command directness levels, with the highest reduction occurring in parking and highway following tasks. The integration of the memory module was particularly effective, further decreasing the takeover rate by up to 65.2% compared to configurations without historical data. These results confirm that the model successfully learns individual driver preferences over time, such as specific comfort thresholds for speed and proximity, thereby delivering a more reliable and personalized autonomous driving experience.

In Case II, as showing the Fig. 8, the experimental results demonstrate that the VLM-integrated CAV has a robust capability to adapt driving performance between normal winter driving conditions and adverse, snowy environments across the three tested scenarios. In highway trials, when the CAV operated under snowy conditions with reduced visibility compared to clear weather, the VLM was able to shift from an overtaking action to a conservative car-following action by decreasing the longitudinal PID proportional gain and increasing lateral MPC weights to prioritize stability. This decision significantly enhanced safety, increasing the TTC from 1.81 s to 3.40 s while reducing speed variance and jerk. Similarly, in the intersection scenario under snowy conditions, the system optimized the control parameters to minimize lateral acceleration, effectively eliminating the risk of slip during turns. Throughout all tests, including constrained maneuvering in the parking lot, the system maintained real-time responsiveness with consistent latency and required zero manual takeovers, validating its reliability in managing the challenging conditions in the Midwest winter.



Vehicle State: [231,57, 0.89, 50, highway]
Driving Behavior: [overtake, yield, following]

Verbal Command: Drive safely!

Thought: The vehicle is on the highway traveling at 50 km\h with command to drive safely. The road appears to be snowy, which means decreased traction. The safest option is to maintain the current lane and speed to avoid any abrupt maneuvers that could lead to skidding.

Action: Following
PID [1.0, 0.1, 0.1]
MPC [0.8, 0.4, 4.0]

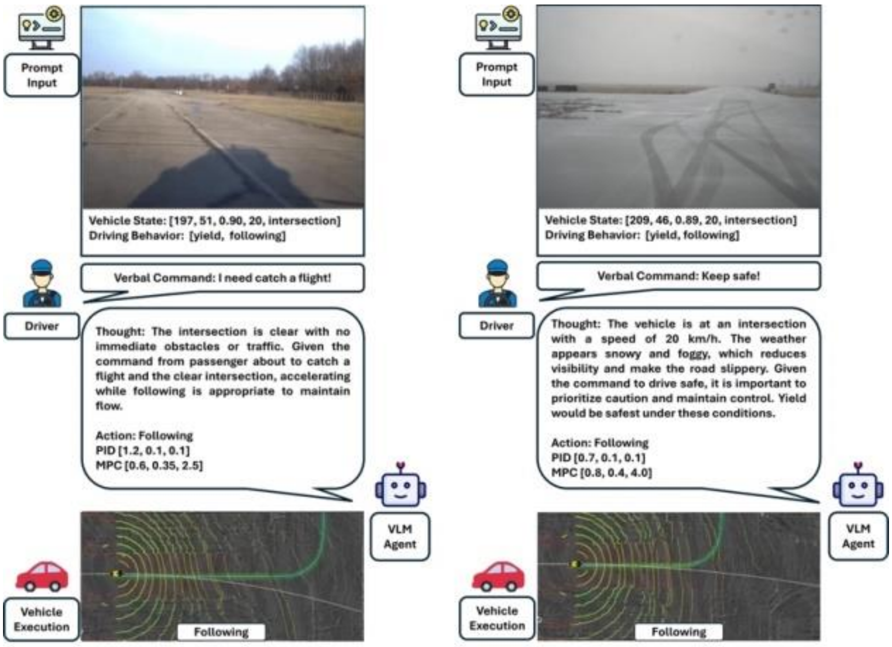
Vehicle State: [130, 46, 0.97, 50, highway]
Driving Behavior: [overtake, yield, following]

Verbal Command: The traffic is too slow!

Thought: The vehicle is on the highway traveling at 50 km\h with a slower truck ahead. The passenger desires to overtake the slow-moving vehicle. The weather is clear and there is ample space for overtaking. Ensuring safety while maneuvering around the pedestrian on the right is essential.

Action: Overtake
PID [1.2, 0.1, 0.1]
MPC [0.6, 0.35, 2.5]

(a) The highway scenario for Case Study II.



Vehicle State: [197, 51, 0.90, 20, intersection]
Driving Behavior: [yield, following]

Verbal Command: I need catch a flight!

Thought: The intersection is clear with no immediate obstacles or traffic. Given the command from passenger about to catch a flight and the clear intersection, accelerating while following is appropriate to maintain flow.

Action: Following
PID [1.2, 0.1, 0.1]
MPC [0.6, 0.35, 2.5]

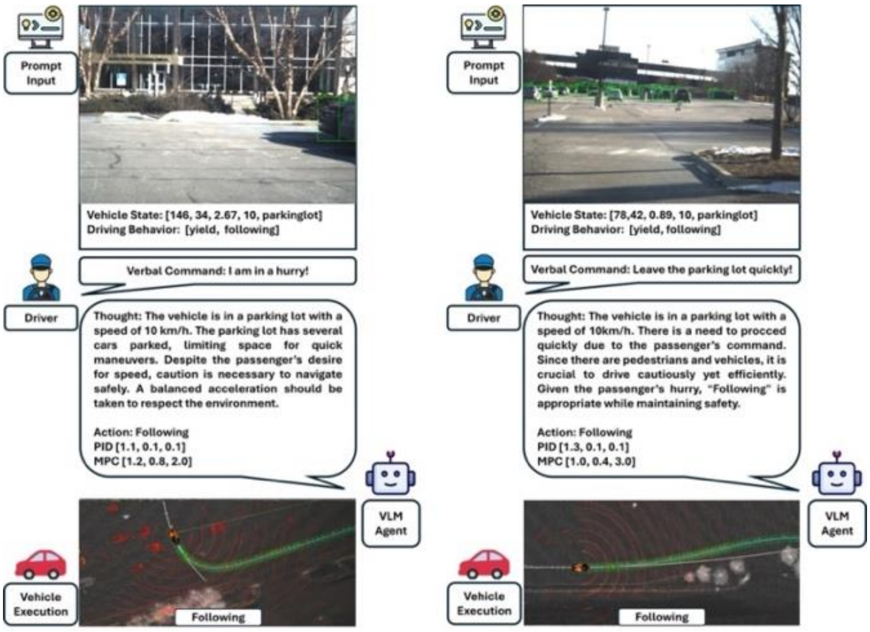
Vehicle State: [209, 46, 0.89, 20, intersection]
Driving Behavior: [yield, following]

Verbal Command: Keep safe!

Thought: The vehicle is at an intersection with a speed of 20 km/h. The weather appears snowy and foggy, which reduces visibility and make the road slippery. Given the command to drive safe, it is important to prioritize caution and maintain control. Yield would be safest under these conditions.

Action: Following
PID [0.7, 0.1, 0.1]
MPC [0.8, 0.4, 4.0]

(b) The intersection scenario for Case Study II.



Vehicle State: [146, 34, 2.67, 10, parkinglot]
Driving Behavior: [yield, following]

Verbal Command: I am in a hurry!

Thought: The vehicle is in a parking lot with a speed of 10 km/h. The parking lot has several cars parked, limiting space for quick maneuvers. Despite the passenger's desire for speed, caution is necessary to navigate safely. A balanced acceleration should be taken to respect the environment.

Action: Following
PID [1.1, 0.1, 0.1]
MPC [1.2, 0.8, 2.0]

Vehicle State: [78,42, 0.89, 10, parkinglot]
Driving Behavior: [yield, following]

Verbal Command: Leave the parking lot quickly!

Thought: The vehicle is in a parking lot with a speed of 10km/h. There is a need to procced quickly due to the passenger's command. Since there are pedestrians and vehicles, it is crucial to drive cautiously yet efficiently. Given the passenger's hurry, "Following" is appropriate while maintaining safety.

Action: Following
PID [1.3, 0.1, 0.1]
MPC [1.0, 0.4, 3.0]

(c) The intersection scenario for Case Study II.

Fig. 8. The experiment visualization for the VLM-integrated CAV case study.

CHAPTER 5. RECOMMENDATIONS

While the proposed Talk2Drive framework successfully validated the interpretation of verbal commands and the VLM framework demonstrated effective parameter tuning for specific snowy conditions, both systems were evaluated in controlled, closed-track environments with a limited range of variability. This data scarcity creates a "domain shift" challenge, in which models may struggle to generalize to the diverse, unpredictable, and extreme winter weather patterns characteristic of the broader Midwest region. Furthermore, both frameworks currently depend on cloud-based models, which introduces significant latency, approximately 1.6 seconds for Talk2Drive and between 2.01 and 3.53 seconds for the VLM-integrated system. While this response time proved acceptable for the low-speed, non-urgent maneuvers tested in our controlled experiments, it poses a substantial safety risk for high-speed public road deployment, where millisecond-level reaction times are essential for emergency collision avoidance. Therefore, the recommended next stage of research is two-fold: expanding the data foundation and optimizing the inference architecture to more advanced End-to-End (E2E) [11] [12] [13] autonomous driving architectures.

CHAPTER 6. OUTPUTS, OUTCOMES, AND IMPACTS

In this project, we proposed a novel hierarchical framework to incorporate with VLMs with connected autonomous vehicles. This architecture enables a high-level VLM agent to interact seamlessly with the connected autonomous vehicle platform, facilitating robust real-world deployment across a range of driving scenarios characterized by Midwest winter conditions. Through two detailed case studies, we validated the framework's ability to interpret abstract high-level commands and semantically understand the physical environment, demonstrating a significantly safer, more reliable, and personalized autonomous driving experience.

Furthermore, to advance autonomous driving development in the Midwest winter-specific operational domain, we collect a specialized dataset that serves as a foundation for exploring CAV performance under extreme weather conditions. As a pioneer in autonomous driving development and deployment in the Midwest, this specialized dataset will serve as a foundation for exploring the advanced performance of these CAV technologies under extreme weather conditions. These pilot development and deployment efforts not only reveal the significant potential for VLM-integrated CAVs to enhance safety and align with human preferences but also underscore the broader impact on the Midwest region by improving mobility and autonomous traffic management.

The following outputs were generated during the performance of this project:

Conference Proceeding:
C. Cui et al., "Personalized Autonomous Driving with Large Language Models: Field Experiments," 2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC), Edmonton, AB, Canada, 2024, pp. 20-27, doi: 10.1109/ITSC58415.2024.10919978.

Journal Publication:
Yupeng Zhou, Can Cui, Juntong Peng, Zichong Yang, Juanwu Lu, Jitesh Panchal, Bin Yao, and Ziran Wang. 2025. A Hierarchical Test Platform for Vision Language Model

(VLM)-Integrated Real-World Autonomous Driving. ACM Trans. Internet Things Just Accepted (September 2025). https://doi.org/10.1145/3769867.

An open-source code: Talk2Drive GitHub Repository.

REFERENCES

[1] Zhang, Y., et al. (2021). "Autonomous Driving in Adverse Weather Conditions: A Survey." IEEE Transactions on Intelligent Transportation Systems.

[2] Dreissig, M., et al. (2023). "Survey on LiDAR Perception in Adverse Weather Conditions." IEEE Intelligent Vehicles Symposium (IV).

[3] T. P. Weber, R. K. Aggarwal and J. C. Gerdes, "Human-Inspired Autonomous Racing in Low Friction Environments," in *IEEE Transactions on Intelligent Vehicles*, vol. 10, no. 6, pp. 3684-3696, June 2025, doi: 10.1109/TIV.2024.3462253

[4] Chenbin Pan, Burhaneddin Yaman, Tommaso Nesti, Abhirup Mallik, Alessandro G Allievi, Senem Velipasalar, and Liu Ren. 2024. VLP: Vision Language Planning for Autonomous Driving. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, 14760–14769. doi:10.1109/CVPR52733.2024.01398

[5] Hexagon AutonomouStuff, "Autonomy research and development platforms," [Online]. Available: https://autonomoustuff.com. [Accessed: Jan. 08, 2026].

[6] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada, "An Open Approach to Autonomous Vehicles," IEEE Micro, vol. 35, no. 6, pp. 60–68, Nov. 2015.

[7] M. Quigley et al., "ROS: an open-source Robot Operating System," in ICRA Workshop on Open Source Software, Kobe, Japan, 2009, vol. 3, no. 3.2, p. 5.

[8] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[9] P. Biber and W. Strasser. 2003. The normal distributions transform: a new approach to laser scan matching. In Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), Vol. 3. IEEE, Las Vegas, Nevada, USA, 2743–2748 vol.3. doi:10.1109/IROS.2003.1249285

[10] Zhou, Y., 2024. Design and Deployment of a Real-World Autonomous Driving Test Platform (Master's thesis, Purdue University).

[11] Cui, Can, et al. "A survey on multimodal large language models for autonomous driving." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024.

[12] Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T. and Chen, E., 2024. A survey on multimodal large language models. National Science Review, 11(12), p.nwae403.

[13] Chen, L., Wu, P., Chitta, K., Jaeger, B., Geiger, A. and Li, H., 2024. End-to-end autonomous driving: Challenges and frontiers. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[14] Wang, W., Zheng, V.W., Yu, H. and Miao, C., 2019. A survey of zero-shot learning: Settings, methods, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), pp.1-37.

[15] Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B. and Vasudevan, V., 2020. Scalability in perception for autonomous driving: Waymo open dataset. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2446-2454).

[16] Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G. and Beijbom, O., 2020. nuscenes: A multimodal dataset for autonomous driving. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11621-11631).