Safety21
INNOVATING FOR SAFETY

US DOT National
University Transportation Center for Safety

Carnegie Mellon University

Cooperative Sensing of Vulnerable Road Users and Real-Time Response to Potential Collisions via Vehicle and Infrastructure Communication

Stephen F Smith (PI) - https://orcid.org/0000-0002-7053-3166
John Dolan (Co-PI) - https://orcid.org/0000-0003-2062-100X

Final Report – August 25, 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 numbers 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.
511		

4. Title and Subtitle Cooperative Sensing of Vulnerable Road Users and Real-time Response to Potential Collisions via Connected Vehicle and Infrastructure Communication	**5. Report Date** August 25, 2025
	6. Performing Organization Code
7. Author(s) Stephen F. Smith – https://orcid.org/0000-0002-7053-3166 John Dolan - https://orcid.org/0000-0003-2062-100X Hsu-Kuang Chiu - https://orcid.org/0000-0002-8643-017X Yiwei Lyu - https://orcid.org/0000-0002-3882-4246	**8. Performing Organization Report No.**
9. Performing Organization Name and Address Safety21 University Transportation Center, 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 U.S. Department of Transportation's University Transportation Centers Program (Research Applied) Washington DC	**13. Type of Report and Period Covered** Final Report (July 1, 2024 – June 30, 2025)
	14. Sponsoring Agency Code USDOT

15. Supplementary Notes

None

16. Abstract

This report summarizes progress made on two problems of central importance to achieving increased safety in autonomous urban driving: (1) the development of end-to-end frameworks for cooperative perception, tracking and planning, and (2) the development of real-time strategies for collision avoidance when collisions are predicted. With respect to the first problem, a new framework that utilizes a multi-modal large language model (MLLM) has been developed and shown to outperform previous approaches to end-to-end autonomous navigation on a contemporary open-source data set. Progress on the second problem has led to an analysis of the effectiveness of several alternative conflict mitigation strategies. The report ends with a discussion of next steps toward refining and integrating these sets of results.

17. Key Words	18. Distribution Statement
LLM-based Collaborative Sensing, End-to-end Autonomous Navigation, Real-time Collision Avoidance	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	14	

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

1. Overview

Intersection safety is critical for all traffic participants, but especially for vulnerable road users (VRU) such as pedestrians and cyclists. Recent autonomous driving advances allow mitigation of several human driver risk factors, such as fatigue and recklessness. However, state-of-the-art autonomous driving technology also has limitations. The perceptual field-of-view of an individual vehicle's sensors can be compromised by nearby occluding objects, greatly reducing detection accuracy with respect to the vehicle's immediate and short-term future state. The vehicle's sensors also have limited range, and distant objects can be difficult to detect. Both limitations introduce further challenges in the downstream tasks of object tracking, trajectory prediction, and motion planning.

To begin to address these limitations, this project has focused on two complementary technology objectives:

1. *Cooperative Sensing and Planning Pipeline* - Development and analysis of an end-to-end framework for cooperative perception, tracking, prediction and planning at intersections that incorporates data from the sensors of connected autonomous vehicles (CAVs) moving through the intersection, from infrastructure sensors mounted at the intersection, and from vulnerable road users connected to the intersection.
2. *CAV Collision Mitigation Strategies* - Development and analysis of CAV strategies for responding to likely collisions that have been identified by the cooperative sensing pipeline.

To address the first objective, we have combined ideas from prior work in cooperative perception, which has focused mainly on object detection to date, with more recent work in large language model (LLM) approaches to safe navigation of individual CAVs. To address the second objective, we have coupled the use of Control Barrier Functions with Reinforcement Learning to learn how to react safely in different circumstances. In both contexts, we start from the assumption of real-time connectivity among all travelers in the vicinity of an intersection.

In the sections below we summarize our accomplishments toward achieving each of these technology objectives and discuss next steps.

2. Cooperative Sensing and Planning Pipeline*

To address CAV detection errors due to occlusion of objects (vehicles, pedestrians, bicyclists, etc.) by other vehicles or buildings at the intersection, recent research has proposed a number of "cooperative perception" algorithms [CHE19,CHI24,WAN20, XU22a,XU22b,XU22c], wherein sensor information from multiple CAVs and/or infrastructure sensors is shared and then fused to produce better overall detection or tracking results. The performance of such algorithms has been verified and comparatively assessed using a growing number of publicly available datasets, both real and simulated, that capture traffic navigation scenarios. V2V4Real [XU23], for example, was the first worldwide available, real-world vehicle-to-vehicle cooperative perception dataset with perception benchmarks to support cooperative perception model training and evaluation.

Other, more recent research has demonstrated the promise of using LLMs as the basis for developing end-to-end perception and planning algorithms for a single CAV [CHE24,NIE24,SIM24,TIA24a,TIA24b,WAN24,WAN23,XU24]. In basic LLM-based approaches, data comprising the driving scene, object detection results and the ego-vehicle's state are first transformed into text input to the LLM, and then the LLM generates text output including the suggested driving action or the planned future trajectory. More sophisticated, multi-modal LLMs (MLLMs) are used to encode point clouds or images into visual features, which are subsequently projected to the language embedding space to enable the LLM to perform visual understanding and question-answering tasks.

Given the perceived potential of an LLM-based (and especially an MLLM-based) approach in a cooperative perception setting, we have focused in this project on developing an LLM-based framework for end-to-end cooperative perception and planning. Since prior research provides no example of an LLM-based cooperative perception framework, investigation of such an approach raises the additional need to develop a cooperative perception dataset for evaluating its performance. In the following two subsections we summarize these two accomplishments and contributions to the research community. Further details of the framework and the data set can be found in [CHI25a].

* This segment of the research has benefited from additional support provided by Nvidia, Inc. Through an internship awarded to and carried out by CMU Ph.D. student Hsu-Kuang Chiu at Nvidia over the summer of 2024 and extending on a part-time basis through the 2024-25 academic year, Nvidia has contributed GBU computing power and in-kind support for collaborating Nvidia researchers.

2.1 V2V-QA: A Dataset for Cooperative Autonomous Driving with LLMs

In contrast to datasets used for prior work in cooperative perception, our LLM-based problem setting involves both sharing of perception data from nearby vehicles to a single LMM situated at a given intersection and LLM support for answering perception and planning questions that are subsequently asked by any nearby vehicle (see Figure 1). Consequently, the dataset must provide the capability to benchmark performance of different pipeline models on both fusing perception information and answering safety-critical driving related questions. Creation of the Vehicle-to-Vehicle Question-Answering (V2V-QA) dataset has been driven by this requirement.

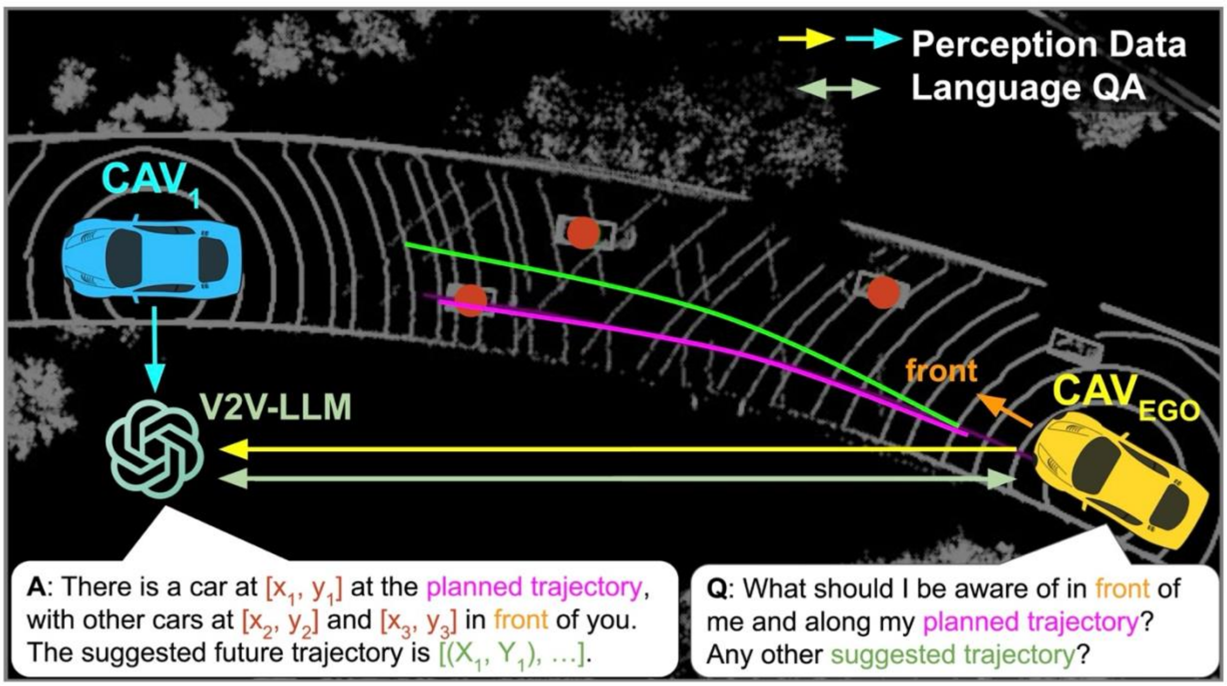

Figure 1: V2V-LLM Problem Setting – All CAVs share their perception information with a single LLM (located at the intersection). Any CAV can then ask the LLM a question to obtain useful information for driving safely.

The V2V-QA dataset that has been developed is comprised of two splits, V2V-split and V2X-split, that are composed respectively from the previously developed V2V4Real [XU23] and V2X-Real [XIA24] datasets. For each frame contained in the two reference datasets 5 different types of question-answer pairs are created, including 3 types of grounding questions, 1 type of notable object identification question and 1 type of planning question. Figure 2 gives examples of each of these 5 types of question-answer pairs, which are all designed for cooperative driving scenarios and include the following:

- Q1: Grounding at reference location
- Q2: Grounding behind a reference object at a location
- Q3: Grounding behind a reference object in a direction
- Q4: Notable object identification
- Q5: Planning

To generate instances of these question-answer pairs, we use V2V4Real and V2X-Real's ground-truth bounding box annotations, each CAV's ground-truth trajectories, and individual detection results as the source information. Then we use different manually designed rules based on the geometric relationship among the entities and text templates to generate our question-answer pairs. Please refer to [CHI25a] for descriptions of the generation rule used for each question-answer type and the text templates that were used.

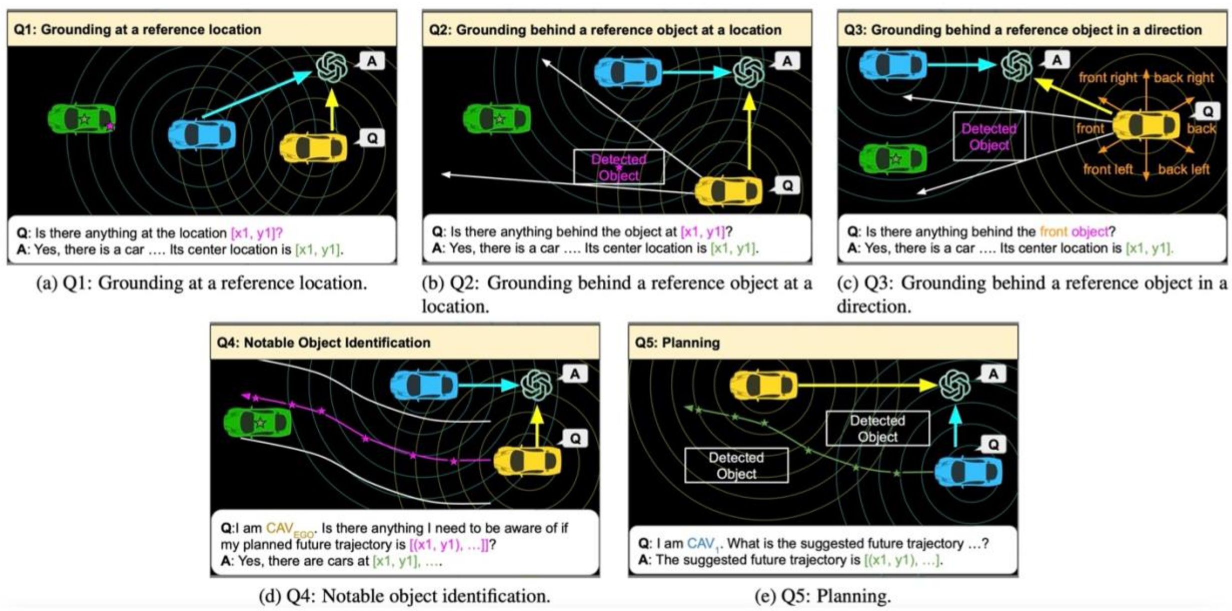

Figure 2: Examples of V2V-QA's 5 types of Question-Answer pairs. The arrows pointing at LLM indicate the perception data from CAVs.

Also associated with the V2V-QA data set is a set of benchmark algorithms for use in calibrating the performance of new LLM-based cooperative driving algorithms that might be proposed by future research. In addition to the strong benchmark provided by the V2V-LLM technique developed in this project and described in the next section, 3 additional benchmarks that rely on different approaches to fusion of perception information - labeled respectively as no fusion, early fusion, and intermediate fusion - have also been defined and incorporated (V2V-LLM adopts what has historically been referred to as a late fusion approach.) As a first contribution to the general automated driving R&D community, we are making the dataset and all benchmark code publicly available for future use. It will be accessible at https://eddyhkchiu.github.io/v2vllm.github.io/ .

2.2 V2V-LLM: Cooperative Perception and Planning

Figure 3 graphically depicts the V2V-LMM architecture for cooperative autonomous driving. We use a multi-modal LLM (MLLM) that takes the individual perception features of every CAV as the vision input, a question as the language input and generates an answer as the language output. For extracting the perception input features, each CAV applies a 3D object detection model to its individual LIDAR point cloud. We use PointPillars [LAN19] as the 3D object detector to remain consistent with what was used in prior work with V2V4Real and V2X-Real datasets and enable fair comparisons. We utilize LLaVA [LIU23] to develop our MLLM, given its superior performance on visual question-answering tasks.

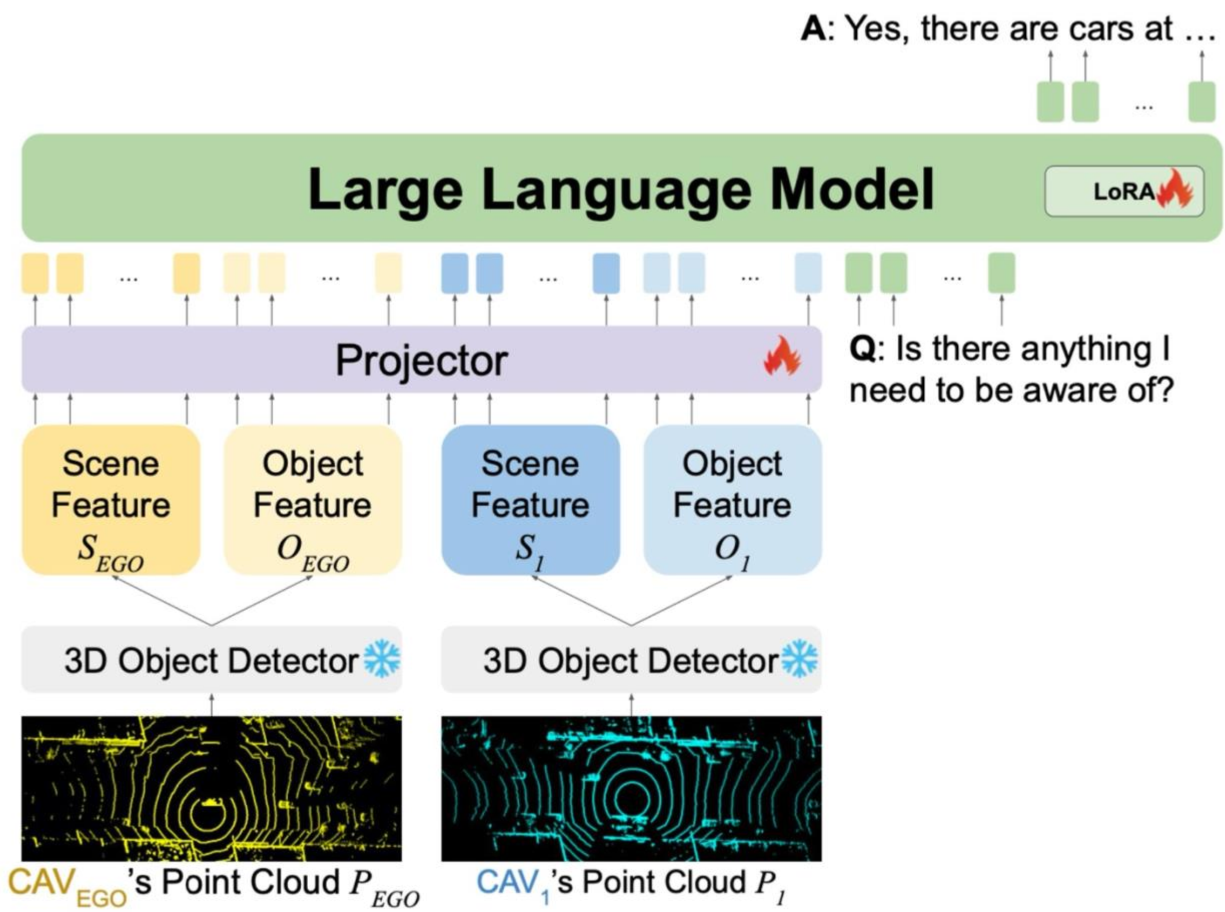

Figure 3: Model diagram of the V2V-LMM for cooperative autonomous driving.

To train our model, we first initialize by loading the pre-trained LLaVA-v1.5-7b checkpoint [LIU23]. We then freeze the LLM and the point cloud feature encoder, and finetune the projector and the LoRA [HU22] parts of the model. Complete details on all settings used to obtain the experimental results reported below can be found in [CHI25a]. We used 8 NVIDIA A100-80GB GPUs to train our model and the rest of the benchmarks we generated.

Table 1 shows the relative performance of V2V-LLM on the V2V-QA dataset in comparison to various benchmarks that rely on other strategies for perception information fusion, including some drawn from prior research [XU22a,XU22b, XU22c]. Overall, V2V-LLM is seen to outperform all other models in the notable object identification and planning tasks and produces competitive results in the grounding tasks. In terms of communication costs, V2V-LLM increases communication costs by just 1.5% in relation to other intermediate fusion methods and is much cheaper than the early fusion benchmark. Additional qualitative performance analysis and an ablation study can be found in [CHI25a].

Method	V2V-split							V2X-split							Comm(MB) ↓
	Q1	Q2	Q3	Q_{Gr}	Q4	Q5		Q1	Q2	Q3	Q_{Gr}	Q4	Q5		
	F1 ↑	F1 ↑	F1 ↑	F1 ↑	F1 ↑	L2 (m) ↓	CR (%) ↓	F1 ↑	F1 ↑	F1 ↑	F1 ↑	F1 ↑	L2 (m) ↓	CR (%) ↓	
No Fusion	66.6	22.6	17.2	35.5	47.3	6.55	4.57	55.7	21.4	25.2	34.1	64.4	2.31	9.21	**0**
Early Fusion	**73.5**	23.3	20.8	39.2	53.9	<u>6.20</u>	<u>3.55</u>	<u>59.7</u>	23.3	26.1	36.4	<u>67.6</u>	<u>2.12</u>	8.61	1.9208
Intermediate Fusion															
AttFuse [[XU22c]	70.7	26.4	18.4	38.5	56.9	6.83	4.12	58.9	23.9	<u>26.3</u>	36.4	65.9	2.19	<u>8.39</u>	<u>0.4008</u>
V2X-ViT [XU22b]	70.8	28.0	**22.6**	40.5	<u>57.6</u>	7.08	4.33	59.6	<u>24.2</u>	26.1	<u>36.6</u>	65.0	2.29	8.86	<u>0.4008</u>
CoBEVT [XU22a]	<u>72.2</u>	<u>29.3</u>	<u>21.3</u>	**40.9**	<u>57.6</u>	6.72	3.88	-	-	-	-	-	-	-	<u>0.4008</u>
LLM Fusion															
V2V-LLM (Ours)	70.0	**30.8**	21.2	<u>40.7</u>	**59.7**	**4.99**	**3.00**	**60.5**	**25.3**	**26.7**	**37.5**	**69.3**	**1.71**	**6.89**	0.4068

Table 1: V2V-LLM's testing performance in V2V-QA's V2V-split and V2X-split in comparison to other baseline methods. Q1: Grounding at a reference location. Q2: Grounding behind a reference object at a location. Q3: Grounding behind a reference object in a direction. Q_{Gr}: Average of grounding (Q1, Q2, and Q3). Q4: Notable object identification. Q5: Planning. L2: L2 distance error. CR: Collision rate. Comm: Communication cost. In each column, the best results are in boldface, and the second-best results are in underline.

3. CAV Collision Mitigation Strategies

The second technology objective of this project was the development of control strategies for safely responding to potential collisions predicted by our proposed end-to-end MLLM-based perception and planning pipeline.

In this work [LYU24], we introduced an innovative risk-aware behavior planning framework designed for autonomous driving, with the aim of fostering socially compliant vehicle behavior in diverse mixed-traffic scenarios. Our objective was to enable autonomous vehicles to exhibit behavior that aligns with societal norms, thus enhancing their acceptability among human drivers. We expanded the scope of Control Barrier Function-inspired risk assessment to encompass a heterogeneous spectrum of road participants, allowing us to explicitly model varying degrees of social influences between different classes of vehicles. We also derived a mathematical condition for accountability tracing, enabling the identification of responsible entities in situations where risks surge. We established social compliance conditions grounded in our unique risk concept, which seamlessly integrate with a wide range of existing safety-critical controllers, regardless of their type or design. By incorporating these conditions, which encode societal expectations, into existing safe controllers, we were able to demonstrate that autonomous vehicles can exhibit context-aware behavior without compromising the safety guarantees provided by existing controllers. This approach effectively excludes behaviors that may be safe but do not align with human intuition while guaranteeing the least interference with the existing controller.

We provide an illustrative example to show the validity and effectiveness of our proposed approach. The existing controller is set to satisfy a collision-free safety requirement so that the vehicle maintains a nominal travel speed whenever possible without collisions. We showcase the performance of our proposed approach in comparison to the existing controller in a two-vehicle scenario, as depicted in Figure 4. We have a scenario plot on the left showing the ego truck in the fast lane with a small passenger vehicle following it. The dashed line represents the decision space of the ego truck in our proposed approach on whether to change its lane and how fast it would like to travel. Distinct ego behavior is observed in three different setups (left, middle, right) on the right.

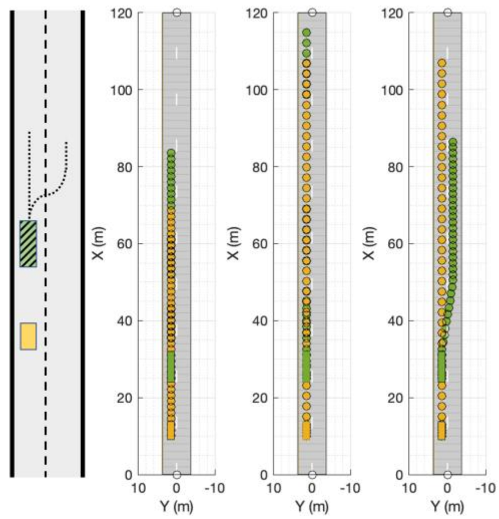

Figure 4: Scenario illustration and simulation plots. The green vehicle is the ego truck, and the yellow vehicle is a small passenger vehicle. The three subplots (left, middle, right) on the right introduce different scenario setups with waypoints plotted out for easier understanding. Left: The small passenger car maintains a steady speed and doesn't create any pressure or risk for the ego truck. Since no risk surge is detected, our behavior planning framework doesn't need to intervene. Consequently, the ego truck continues to follow its existing controller. Middle: The ego truck is solely relying on its existing controller without our behavior planning framework. At a certain point, the small passenger vehicle begins to accelerate, rapidly closing the gap between the two vehicles. Without considering the potential risks to both itself and the human passenger vehicle, the controller is left with no alternative but to instruct the ego truck to also accelerate in order to maintain a safe inter-vehicle distance. Right: Here the small passenger vehicle takes the same action as in the middle subplot by accelerating. However, the difference is that the ego truck is equipped with our proposed algorithm. As the gap between the two vehicles rapidly shrinks, the increased risk to the ego truck triggers the accountability trace, leading the ego truck to hold the small passenger vehicle accountable. Subsequently, the intervention mechanism is activated. In evaluating various choices available to the ego truck and considering their alignment with typical human expectations, it decides to temporarily deviate from its nominal controller. It does so by executing a lane change to the right lane, allowing the small passenger vehicle to pass first. This decision is a wiser one compared to the scenario in the middle subplot, as the ego truck chooses not to jeopardize the safety of both itself and the small passenger vehicle. At the same time, it strives to adhere to the task-related nominal controller to the greatest extent possible.

4. Next Steps

4.1 Cooperative Sensing and Planning Pipeline Extensions

Although the V2V-LLM end-to-end framework performs well on the LLM-QA dataset, there are a few important extensions to be considered:

1. It may be possible to further improve the performance of the pipeline by providing the LLM with an explicit reasoning plan. Recent experiments where a Chain-of-Thoughts reasoning graph was provided to the LLM have produced improved results on the V2V-QA dataset, albeit at increased computational cost. [CHI25b] Current work is aimed at strategies for more cost-effectively exploiting the reasoning graph by restricting attention to a portion of it.

2. One potential limitation of the V2V-LLM framework is its scalability. The LLM-QA dataset that we have tested on involves only 2 CAVs, and as the number of CAVs, infrastructure sensors, and other travelers (connected or not) moving through a given intersection increases, the ability of a centralized LLM framework to keep pace with execution becomes more challenging. The exploration of decentralized architectures, where CAVs are seen as LLM agents and the fusing of perceptual data can be limited to those CAVs in near proximity that a given agent needs to worry about, could provide a more scalable long-term solution.

3. Finally, from an application perspective, logical next steps are to extend the framework to enable communication of predicted collisions back to relevant CAVs and detected VRUs, develop mechanisms for determining when trajectory prediction and planning results warrant classification as a potential collision, and test these integrated capabilities in the field.

4.2 Better Collision Mitigation

Regarding strategies for mitigation of predicted collisions, next steps include:

1. Incorporation of human driver behavior models such as the one presented in [LUD24] could improve understanding and prediction of likely control behaviors of human-driven vehicles.

2. Incorporation of recent bank-of-models control techniques such as [ALS25] can improve vehicle modeling and safety under uncertainty regarding road surface conditions and vehicle tire wear.

References

[ALS25] Mohammed AL-Sunni, Hassan Almubarak, Katherine Horng, and John M. Dolan, "LLA-MPC: Fast Adaptive Control for Autonomous Racing", to appear in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (October 2025).

[CHE24] Long Chen, Oleg Sinavski, Jan Hunermann, Alice Karnsund, Andrew James Willmott, Danny Birch, Daniel Maund, and Jamie Shotton. "Driving with LLMs: Fusing object-level modality for explainable autonomous driving", In *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[CHE19] Qi Chen, Xu Ma, Sihai Tang, Jingda Guo, Qing Yang, and Song Fu. "F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3D point clouds", In *ACM/IEEE Symposium on Edge Computing (SEC)*, 2019.

[CHI24] Hsu-Kuang Chiu, Chien-Yi Wang, Min-Hung Chen, and Stephen F. Smith. "Probabilistic 3D multi-object cooperative tracking for autonomous driving via differentiable multi-sensor Kalman filter", In *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[CHI25a] Hsu-Kuang Chiu, Ryo Hachiuma, Chien-Yi Wang, Stephen F. Smith, Yu-Chiang Frank Wang, and Min-Hung Chen, "V2V-LLM: Vehicle-to-Vehicle Cooperative Autonomous Driving with Multi-Modal Large Language Models", arXiv PrePrint, arXiv:2502.09980v3, April 2025 (submitted to *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025)

[CHI25b] Hsu-Kuang Chiu, et al., Unpublished.

[HU22] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen, "LoRA: Low-rank adaptation of large language models", In *International Conference on Learning Representations (ICLR)*, 2022.

[LAN19] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom, "Pointpillars: Fast encoders for object detection from point clouds", In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[LIU23] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee, "Visual instruction tuning", In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

[LUD24] Nathan Ludlow, Yiwei Lyu, and John M. Dolan, “Hierarchical Learned Risk-Aware Planning Framework for Human Driving Modeling”, In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2223-2229 (May 2024).

[LYU24] Yiwei Lyu, Wenhao Luo, and John M. Dolan, “Risk-based Socially-Compliant Behavior Planning for Autonomous Driving”, In *American Control Conference (ACC)*, pp. 3827-3832 (July 2024).

[NIE24] Ming Nie, Renyuan Peng, Chunwei Wang, Xinyue Cai, Jianhua Han, Hang Xu, and Li Zhang, “Reason2drive: Towards interpretable and chain-based reasoning for autonomous driving”, In *European Conference on Computer Vision (ECCV)*, 2024.

[SIM24] Chonghao Sima, Katrin Renz, Kashyap Chitta, Li Chen, Hanxue Zhang, Chengen Xie, Ping Luo, Andreas Geiger, and Hongyang Li, “Drivelm: Driving with graph visual question answering. In *European Conference on Computer Vision (ECCV)*, 2024.

[TIA24a] Ran Tian, Boyi Li, Xinshuo Weng, Yuxiao Chen, Edward Schmerling, Yue Wang, Boris Ivanovic, and Marco Pavone, “Tokenize the world into object-level knowledge to address long-tail events in autonomous driving”, In *Conference on Robot Learning (CoRL)*, 2024.

[TIA24b] Xiaoyu Tian, Junru Gu, Bailin Li, Yicheng Liu, Yang Wang, Zhiyong Zhao, Kun Zhan, Peng Jia, Xianpeng Lang, and Hang Zhao, “Drivevlm: The convergence of autonomous driving and large vision-language models. arXiv preprint arXiv:2402.12289, 2024.

[WAN24] Shihao Wang, Zhiding Yu, Xiaohui Jiang, Shiyi Lan, Min Shi, Nadine Chang, Jan Kautz, Ying Li, and Jose M Alvarez, “OmniDrive: A holistic LLM-agent framework for autonomous driving with 3D perception, reasoning and planning”, arXiv:2405.01533, 2024.

[WAN20] [Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Bin Yang, Wenyuan Zeng, James Tu, and Raquel Urtasun. “V2VNET: Vehicle-to-vehicle communication for joint perception and prediction”, In *European Conference on Computer Vision (ECCV)*, 2020.

[WAN23] Wenhai Wang, Jiangwei Xie, ChuanYang Hu, Haoming Zou, Jianan Fan, Wenwen Tong, Yang Wen, Silei Wu, Hanming, Deng, Zhiqi Li, et al. “Drivemlm: Aligning multi-modal large language models with behavioral planning states for autonomous driving”, arXiv preprint arXiv:2312.09245, 2023.

[XIA24] Hao Xiang, Zhaoliang Zheng, Xin Xia, Runsheng Xu, Letian Gao, Zewei Zhou, Xu Han, Xinkai Ji, Mingxi Li, Zonglin Meng, Li Jin, Mingyue Lei, Zhaoyang Ma, Zihang He, Haoxuan Ma, Yunshuang Yuan, Yingqian Zhao, and Jiaqi Ma, “V2X-Real: a large-scale dataset for vehicle-to-everything cooperative perception”, In *European Conference on Computer Vision (ECCV)*, 2024.

XU22a[50] Runsheng Xu, Zhengzhong Tu, Hao Xiang, Wei Shao, Bolei Zhou, and Jiaqi Ma. “Cobevt: Cooperative bird’s eye view semantic segmentation with sparse transformers”, In Conference on Robot Learning (CoRL), 2022.

XU22b[51] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma, “V2x-vit: Vehicle-to-everything cooperative perception with vision transformer”, In *European Conference on Computer Vision (ECCV)*, 2022.

XU22c[52] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma, “Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication, In *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

XU23[53] Runsheng Xu, Xin Xia, Jinlong Li, Hanzhao Li, Shuo Zhang, Zhengzhong Tu, Zonglin Meng, Hao Xiang, Xiaoyu Dong, Rui Song, Hongkai Yu, Bolei Zhou, and Jiaqi Ma, “V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception”, In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1,

XU24[54] Zhenhua Xu, Yujia Zhang, Enze Xie, Zhen Zhao, Yong Guo, Kwan-Yee K. Wong, Zhenguo Li, and Hengshuang Zhao, “Drivegpt4: Interpretable end-to-end autonomous driving via large language model”, IEEE Robotics and Automation Letters, 2024.