Optimizing Autonomous Driving for Safety: A Human-Centric Approach with LLM-Enhanced RLHF

Yuan Sun
Rutgers University, WINLAB
Rutgers University
Piscataway, NJ , USA
ys820@soe.rutgers.edu

Salami Pargoo, Navid
Department of Electrical and Computer Engineering
Rutgers University
Piscataway, NJ , USA
navid.salamipargoo@rutgers.edu

Peter J. Jin
Civil and Environmental Engineering
Rutgers, The State University of New Jersey
Piscataway, NJ , USA
peter.j.jin@rutgers.edu

Jorge Ortiz
Electrical Computer Engineering
Rutgers University
Piscataway, NJ , USA
jorge.ortiz@rutgers.edu

ABSTRACT

Reinforcement Learning from Human Feedback (RLHF) is popular in large language models (LLMs), whereas traditional Reinforcement Learning (RL) often falls short. Current autonomous driving methods typically utilize either human feedback in machine learning, including RL, or LLMs. Most feedback guides the car agent's learning process (e.g., controlling the car). RLHF is usually applied in the fine-tuning step, requiring direct human "preferences," which are not commonly used in optimizing autonomous driving models. In this research, we innovatively combine RLHF and LLMs to enhance autonomous driving safety. Training a model with human guidance from scratch is inefficient. Our framework starts with a pre-trained autonomous car agent model and implements multiple human-controlled agents, such as cars and pedestrians, to simulate real-life road environments. The autonomous car model is not directly controlled by humans. We integrate both physical and physiological feedback to fine-tune the model, optimizing this process using LLMs. This multi-agent interactive environment ensures safe, realistic interactions before real-world application. Finally, we will validate our model using data gathered from real-life testbeds located in New Jersey and New York City.

CCS CONCEPTS

• Human-centered computing → Ubiquitous computing; **• Computer systems organization → Robotic control**; **• Computer systems organization → Embedded and cyber-physical systems**;

KEYWORDS

Autonomous Driving ; RLHF

UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-1058-2/24/10
https://doi.org/10.1145/3675094.3677588

ACM Reference Format:
Yuan Sun, Salami Pargoo, Navid, Peter J. Jin, and Jorge Ortiz . 2024. Optimizing Autonomous Driving for Safety: A Human-Centric Approach with LLM-Enhanced RLHF. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '24), October 5–9, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3675094.3677588

1 INTRODUCTION

Reinforcement Learning (RL) and Large Language Models (LLMs) play a crucial role in the development of autonomous driving systems. RL, a branch of machine learning, focuses on enabling agents to make optimal decisions over time by learning from their mistakes and experiences. Numerous studies have demonstrated the application of RL in autonomous driving. For instance, [5] proposed using RL to map sensor observations to control outputs in simulated environments. Similarly, [9] explored deep RL for continuous control tasks, which are applicable to autonomous driving scenarios. Additionally, [15] introduced ASAP-RL, an efficient RL algorithm that utilizes motion skills and expert priors to enhance learning efficiency and driving performance in dense traffic. Moreover, [6] presented a method for making decisions such as lane changing, accelerating, and braking on highways, employing Deep Q Networks (DQN) to train their model and predict optimal actions. Recent research has increasingly incorporated Large Language Models (LLMs) into autonomous driving systems, leveraging their capabilities for decision-making, reasoning, and interaction. For instance, [1, 4, 10] integrated LLMs to enhance commonsense reasoning and high-level decision-making. In another study, [2] utilized LLMs to help autonomous driving models mimic human behavior, thereby improving end-to-end driving performance. [16] employed GPT to extract crucial information from NHTSA accident reports using a QA approach, enabling the generation of diverse scene codes for simulation and testing. Additionally, [14] demonstrated the use of LLMs as powerful interpreters that translate user text queries into structured specifications of map lanes and vehicle locations for traffic simulation scenarios. RLHF is a fundamental component in the training of Large Language Models (LLMs) and is regarded as an essential element of the modern LLM training pipeline [7, 18, 19]. RLHF is particularly well-suited for LLMs [13], as it involves RL agents learning from human preference feedback. This type of feedback is considered more intuitive for human users, better aligned with human values, and easier to obtain in various applications [7].

Most works apply RLHF to optimize LLMs because the scenarios in which humans use LLMs are more suitable for tracking human preferences. This alignment with human values ensures that the models perform more intuitively in real-world applications. In contrast, for autonomous driving scenarios, it is impractical for humans to provide preference feedback on a frame-by-frame basis. Consequently, RLHF is seldom used in autonomous driving contexts.

In this research, we creatively apply RLHF to autonomous driving by modeling human preferences through various sensor feedback from our environment. We incorporate both physical and physiological feedback into the simulation to optimize the RL training loop for autonomous driving. Additionally, the LLM agent facilitates interaction within our multi-agent system. We posit that training autonomous car agents alongside human-driven cars in the simulation can significantly enhance safety and allow the agents to learn human preferences concurrently. This approach aligns the autonomous car models more closely with real-world scenarios, ultimately making the application of autonomous driving models safer in real-life contexts.

2 METHODOLOGY

Our system (figure 2) is designed as a multi-agent framework, centering on human interaction while incorporating both LLM agents and autonomous car agents. The system includes human drivers, human pedestrians, and an LLM agent that mimics their behavior to generate training interactions for the car agent. Humans control the car using physical controllers such as steering wheels and pedals. Additionally, they wear various wearable sensors, including VR headsets, wristbands to collect and monitor physiological signals in real-time, and smart glasses to track eye gaze. A camera is also set up in the environment to record human reactions. This collected multimodal data is sent to the simulation. Before integration, the LLM assists the car agent in understanding this data and aids the human agent in adapting to the simulation environment. The autonomous vehicle model learns from human feedback within the RL loop, with the LLM helping to interpret human data into "preferences" to optimize the model in the RL loop.

2.1 RLHF

In traditional reinforcement learning, the objective of the agent is to develop a policy, a function that dictates its actions. This policy is optimized to maximize rewards provided by a distinct reward function based on the agent's performance in a given task [11]. However, defining a reward function that accurately reflects human preferences is challenging. To address this, RLHF aims to train a "reward model" directly from human feedback [20].

However, in our scenario, it is typically difficult to obtain "preference" feedback directly on a frame-by-frame basis. For instance, when training an autonomous car model using RL, the car might simply reward itself with a positive score for avoiding collisions. However, the car might execute a rapid lane change that frightens the user. This type of feedback—reflecting user comfort and safety preferences—is crucial but challenging to capture. Additionally, scenarios such as aggressive braking, abrupt acceleration, or failing to yield to pedestrians can all contribute to negative user experiences, which are important "preferences" in the autonomous driving RL loop. At the same time, human driving behavior provides valuable feedback to the system, as it reflects real-world driving preferences and responses to various driving conditions.

In our autonomous driving RLHF framework, the objective function[12] is defined as follows:

$$\text{objective}(\phi) = \mathbb{E}_{(x,y)\sim D_{\pi_\phi^{RL}}} \left[r_\theta(x,y) - \beta \log \left(\frac{\pi_\phi^{RL}(y|x)}{\pi^{SFT}(y|x)} \right) \right] \quad (1)$$

where ,in our work, x represents the input data from various sensors, including physical sensors, physiological sensors, and simulation data such as LiDAR and camera inputs. The output y denotes the actions taken by the autonomous vehicle. The reward function $r_\theta(x,y)$ evaluates the quality of the action y given the sensor inputs x, guided by human feedback. The term $\beta \log \left(\frac{\pi_\phi^{RL}(y|x)}{\pi^{SFT}(y|x)} \right)$ introduces a KL divergence penalty to ensure the new policy remains close to the initial supervised model, balancing learning from new data while retaining useful information from the initial model.

2.2 LLM Integration in Autonomous Driving

The primary functions of the LLM in our work include acting as an agent in the simulation, facilitating interaction between humans and the simulation system, and optimizing the RL training loop.

2.2.1 LLM Agent in the Simulation. In this section, we emphasize our multi-agent system with two key use cases. First, the LLM agent can mimic human behavior to interact with the car agent when a human is not available. Second, when a human is in the loop, the LLM agent can act as another agent, such as a car or pedestrian, to increase the system's complexity. For instance, human feedback differs when there is one car versus multiple cars on the road. The feedback in such a complex scenario is more representative of real-life situations.

2.2.2 Enhanced Human-Simulation Interaction Based on LLM. When the simulation interacts with the human agent, the LLM can enhance the interaction. Firstly, when sending data collected from humans to the system, the LLM can help interpret the data. For example, if a driver is skilled, the LLM might adjust the simulation weather to foggy conditions. Conversely, if the driver is less skilled, the LLM can help the user adapt to the environment before the car agent begins its training.

2.2.3 LLM-Enhanced RLHF. In the RLHF loop, "preferences" are not as straightforward as yes or no answers. The LLM can translate physical and physiological data into preference formats, which are then incorporated into the objective function. For example, if a driver's heart rate increases significantly during a particular maneuver, the LLM can interpret this physiological response as a negative preference for that action. Similarly, if sensor data indicates smooth and confident handling of the vehicle, this can be translated into a positive preference. By integrating these nuanced preferences

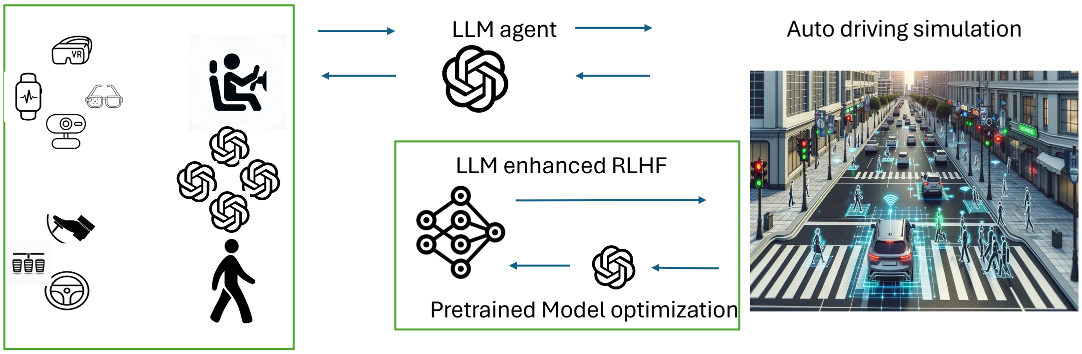

Figure 1: Overview of our human-centric multi-agent LLM-enhanced RLHF system framework. During the fine-tuning of an autonomous car model, human agents and proliferated LLM agents mimicking multiple human behaviors are incorporated into the environment to align with real-world human preferences.

Figure 2: Simulation room with VR headset, steering controls, and monitors for real-time multimodal data collection and autonomous driving optimization.

into the RL training loop, the autonomous driving model can be better aligned with human comfort and safety standards.

3 HARDWARE SETUP

The sensors utilized in our multi-agent LLM-enhanced RLHF system are categorized into two primary modalities: vehicle and physiological[17]. The vehicle sensors, primarily sourced from the CARLA simulator and Logitech hardware, include acceleration, rotation (gyroscope), speed, brake, steering, throttle, and reverse, all sampled at approximately 60 Hz. These sensors capture the dynamic state and control inputs of the autonomous vehicle. The physiological sensors, provided by Empatica, measure various physiological signals such as blood volume pulse (64 Hz), heart rate (1 Hz), interbeat interval (varying), electrodermal activity (4 Hz), wrist acceleration (32 Hz), and body temperature (4 Hz). Additionally, the gaze modality employs an Adhawk sensor to track coordinates on the screen at a sampling rate of 125 Hz, capturing the human agent's visual focus

Figure 3: Example of GPT-4o imitating a human agent in the CARLA simulation. The LLM agent is attempting to overtake the car in front in a human-like manner.

and attention. These sensors monitor the human agent's physical and emotional responses during the simulation. Except for the monitor, we also have a VR headset to create an immersive environment. Additionally, a Raspberry Pi camera in the simulation room observes the human reaction to the simulation. This multi-modal data integration is crucial for fine-tuning the autonomous driving model, providing comprehensive feedback to align the model's performance with human preferences and ensuring realistic and safe interactions in the simulation environment.

4 INITIAL IMPLEMENTATION

Our initial implementation demonstrates the integration of the LLM with the car simulation system using the GPT-4 interface. The LLM agent can imitate human driving behavior, especially when interacting with a car agent in front (figure 3). It also assists the car agent in managing situations such as avoiding collisions(figure 4). Additionally, the LLM agent helps human users by instructing them on how to effectively navigate and use the simulation system(figure 5).

Furthermore, we plan to implement our experiments in our real-life city test bed, located in Harlem[8], NYC, and New Brunswick[3], NJ. The real-life data collected from these locations will be used to test the robustness of our algorithm. Additionally, we can import real-life data into the simulation as a cross-validation method. Figure 6 shows an example of importing real-life road data from New Brunswick into the CARLA system.

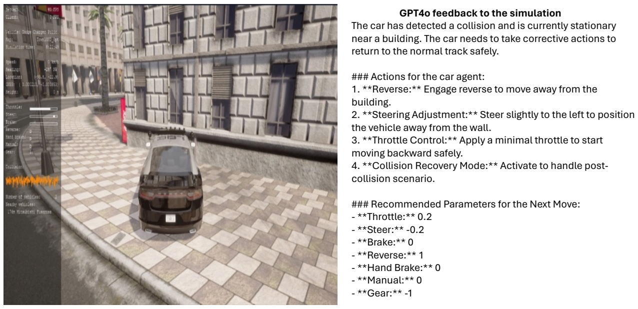

Figure 4: Example of the LLM agent guiding the autonomous car agent to reverse away from a collision with a building.

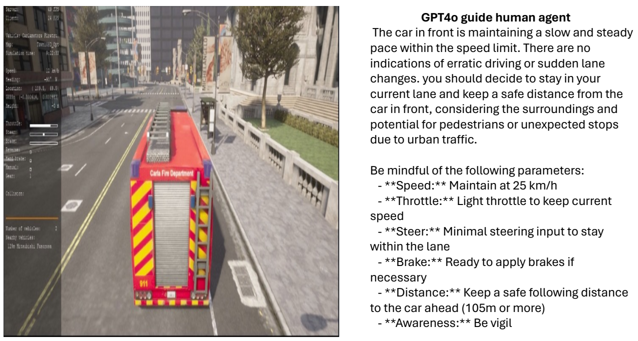

Figure 5: Example of the LLM agent assisting the human agent in using the simulation.

5 CONCLUSION AND FUTURE WORK

In this work, we introduce a novel framework that integrates RLHF and LLMs to optimize autonomous driving models. We define human preferences within the RLHF framework and build a simulation-to-reality system based on this concept. Our method simulates a multi-agent environment for training the car agent, allowing it to learn human behaviors through multimodal sensory data. The LLM agent can proliferate multiple human agents by mimicking human behavior and facilitating interactions between the car agent and other agents on the road within the simulation. When optimizing the model, the LLM agent also interprets human data to enhance the model via RLHF. Our system incorporates both physical and simulation sensors. The initial implementation demonstrates various scenarios where LLMs are applied to the framework. This preliminary work establishes the foundational infrastructure for our experiments and discusses the theoretical feasibility of the framework.

However, there is still much work to be done in the next stage of our plan. Firstly, the GPT-4 interface has rate limits; we may need to explore different interfaces for this study. The machine learning model for autonomous driving should be evaluated across different types of multimodal models to prove the robustness of our method. We will apply more real-life data to our research to improve the robustness of our method. We will recruit subjects with diverse backgrounds and varying driving skills for human evaluation. Individuals with good driving skills and those with less experience present different challenges in our study. We plan to provide a comprehensive evaluation of how different levels of background influence the RLHF autonomous driving framework.

Figure 6: Example of importing real-life road data from New Brunswick testbed into the CARLA simulation.

Finally, we hope that through our study, we can eventually propose a safe driving model that can help autonomous vehicles navigate real-life roads and contribute to the overall road safety of society.

6 ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) as part of the Center for Smart Streetscapes, under NSF Cooperative Agreement EEC-2133516. DataCity Smart Mobility Testing Ground is jointly funded by Middlesex County Resolution 21-821-R, New Jersey Department of Transportation and Federal Highway Administration Research Project 21-60168.

REFERENCES

[1] Yaodong Cui, Shucheng Huang, Jiaming Zhong, Zhenan Liu, Yutong Wang, Chen Sun, Bai Li, Xiao Wang, and Amir Khajepour. 2023. DriveLLM: Charting the path toward full autonomous driving with large language models. *IEEE Transactions on Intelligent Vehicles* (2023).

[2] Yiqun Duan, Qiang Zhang, and Renjing Xu. 2024. Prompting Multi-Modal Tokens to Enhance End-to-End Autonomous Driving Imitation Learning with LLMs. *arXiv preprint arXiv:2404.04869* (2024).

[3] Center for Advanced Infrastructure and Transportation. 2024. *DataCity Smart Mobility Testing Ground.* https://cait.rutgers.edu/datacity/ Accessed: 2024-05-26.

[4] Daocheng Fu, Wenjie Lei, Licheng Wen, Pinlong Cai, Song Mao, Min Dou, Botian Shi, and Yu Qiao. 2024. LimSim++: A Closed-Loop Platform for Deploying Multimodal LLMs in Autonomous Driving. *arXiv preprint arXiv:2402.01246* (2024).

[5] Christopher Galias, Adam Jakubowski, Henryk Michalewski, Błażej Osiński, and Paweł Zięcina. 2019. Simulation-based reinforcement learning for autonomous driving. (2019).

[6] Carl-Johan Hoel, Krister Wolff, and Leo Laine. 2018. Automated speed and lane change decision making using deep reinforcement learning. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2148–2155.

[7] Robert Kirk, Ishita Mediratta, Christoforos Nalmpantis, Jelena Luketina, Eric Hambro, Edward Grefenstette, and Roberta Raileanu. 2023. Understanding the effects of rlhf on llm generalisation and diversity. *arXiv preprint arXiv:2310.06452* (2023).

[8] COSMOS Lab. 2024. *COSMOS Lab.* https://cosmos-lab.org/ Accessed: 2024-05-26.

[9] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971* (2015).

[10] Jimuyang Zhang Zanming Huang Arijit Ray and Eshed Ohn-Bar. [n. d.]. Feedback-Guided Autonomous Driving. ([n. d.]).

[11] Stuart J Russell and Peter Norvig. 2016. *Artificial intelligence: a modern approach.* Pearson.

[12] Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. 2020. Learning to summarize with human feedback. *Advances in Neural Information Processing Systems* 33 (2020), 3008–3021.

[13] Hao Sun. 2023. Reinforcement learning in the era of llms: What is essential? what is needed? an rl perspective on rlhf, prompting, and beyond. *arXiv preprint arXiv:2310.06147* (2023).

[14] Shuhan Tan, Boris Ivanovic, Xinshuo Weng, Marco Pavone, and Philipp Kraehenbuehl. 2023. Language conditioned traffic generation. *arXiv preprint arXiv:2307.07947* (2023).

[15] Letian Wang, Jie Liu, Hao Shao, Wenshuo Wang, Ruobing Chen, Yu Liu, and Steven L Waslander. 2023. Efficient reinforcement learning for autonomous driving with parameterized skills and priors. *arXiv preprint arXiv:2305.04412* (2023).

[16] Sen Wang, Zhuheng Sheng, Jingwei Xu, Taolue Chen, Junjun Zhu, Shuhui Zhang, Yuan Yao, and Xiaoxing Ma. 2022. ADEPT: A testing platform for simulated autonomous driving. In *Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering*. 1–4.

[17] Tong Wu, Navid Salami Pargoo, and Jorge Ortiz. 2023. Multi-sensor Fusion for In-cabin Vehicular Sensing Applications. In *Proceedings of the 22nd International Conference on Information Processing in Sensor Networks*. 332–333.

[18] Chen Zheng, Ke Sun, Hang Wu, Chenguang Xi, and Xun Zhou. 2024. Balancing Enhancement, Harmlessness, and General Capabilities: Enhancing Conversational LLMs with Direct RLHF. *arXiv preprint arXiv:2403.02513* (2024).

[19] Banghua Zhu, Michael Jordan, and Jiantao Jiao. 2023. Principled reinforcement learning with human feedback from pairwise or k-wise comparisons. In *International Conference on Machine Learning*. PMLR, 43037–43067.

[20] Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593* (2019).