Human-level learning for autonomous driving: Learn to drive with Large Multimodal Foundation Models (LMFM)

Baha Zarrouki[1,2] and Johannes Betz[2]

Abstract*— Current autonomous driving technologies are resource-intensive in their development, and deploying them on public streets remains technologically challenging and financially unsustainable, contributing to the economic strain in Germany's and Europe's automotive sector [1]. Large Multimodal Foundation Models (LMFM) are emerging as a framework to tackle challenges in robotics due to their advanced context understanding and capturing causalities across different modalities being trained on various tasks. Fine-tuning LMFMs for specific tasks or domains, such as autonomous driving, promises higher performance. However, despite their potential, current state-of-the-art methods struggle to master the complexities of the driving task. They are often resource-intensive and sample-inefficient, lagging behind the efficiency and effectiveness of human learning processes. In this research, we propose a novel approach to mimic human-level learning for driving, by introducing a high-level multi-task fine-tuning curriculum that divides the driving task into four phases, mirroring the structured progression found in human driving school curricula. Our approach will gradually enhance task complexity while reducing the reliance on expert guidance. Being successful in this approach, we will deliver an end-to-end autonomous driving system capable of mapless navigation while adhering to Road Traffic Regulations (StVO). To increase the reasoning and decision-making capabilities of our base LMFM and bridge the gap between simulation and real-world driving, we introduce a novel continuous fine-tuning technique termed online Iterative Reinforcement Driving Learning from Driving Instructor Feedback/Suggestion (RDL-DIF). Leveraging our extensive experience in autonomous vehicle algorithm development across public roads [2][3][4] and racing environments [5][6][7], we aim to achieve a significant 92x reduction in the time required to master the real-world driving task, while simultaneously cutting resource requirements by at least 96x compared to existing state-of-the-art approaches. By streamlining resource demands, we seek to pioneer the next generation of autonomous vehicle software, making scalable and profitable autonomous vehicle systems a reality.

I. INTRODUCTION

In late September 2024, our institute demonstrated an autonomous passenger vehicle, EDGAR (see Fig. 1), as a shuttle service between Theresienwiese and Munich Main Station during Oktoberfest [3][4][2]. This challenging deployment required navigating around large crowds, including intoxicated pedestrians, showcasing our algorithm's robustness. The project involved three professors and around 25 PhDs actively over 2.5 years. Built on the Autoware open-source stack [8], this solution, though advanced, is highly specific to the route, limited in speed, and lacks full scalability for diverse environments. Despite these high investments, even industry leaders like Alphabet-backed Waymo, with 2,500 employees, can only operate in selected urban areas [9]. These classic modular software architectures [10][11][12][13][14][15][16] come with inherent complexities, consuming substantial time and resources during development. The emergence of Large Multimodal Foundation Models (LMFMs) presents a promising alternative, leveraging effective world-model prediction capabilities [17][18][19][20]. Recent work has introduced GAIA-1, the first generative world model tailored for self-driving systems [21]. Despite its potential, this approach demands extensive training time, utilizing around 4,700 hours of driving data from Microsoft-backed Wayve's UK corpus (385 employees) and exclusively relying on cameras, omitting lidars and radars. In contrast, human learning typically requires only 14 theory units (each lasting 90 minutes) and 30 hours of practical driving—a total of 51 hours, making it 92 times faster than Wayve's approach. In this research, we propose a highly time-efficient methodology to attain human-level driving proficiency, requiring 96 times fewer resources than Wayve.

Fig. 1. TUM research platform (EDGAR) operating as an autonomous shuttle service during Oktoberfest [3].

This research delivers an autonomous driving learning framework (RDL-DIF) and an end-to-end trained LMFM model that:

- **Accelerates learning by 92x**, drastically reducing the time needed to master driving tasks compared to state-of-the-art methods,
- **Achieves human-level proficiency** in robotic task learning, matching human standards in both performance quality and learning efficiency,
- **Offers a streamlined alternative** to modular au-

[1] Chair of Automotive Technology, Technical University of Munich
[2] Professorship of Autonomous Vehicle Systems, TUM School of Engineering and Design, Technical University of Munich, 85748 Garching, Germany; Munich Institute of Robotics and Machine Intelligence (MIRMI), {baha.zarrouki, johannes.betz}@tum.de

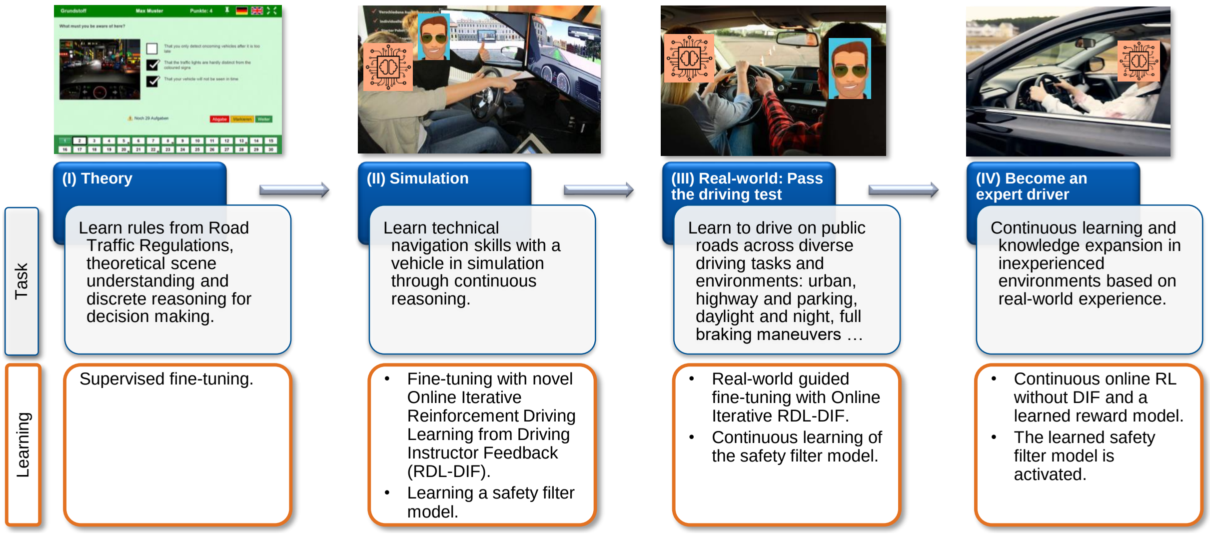

Fig. 2. Learning the driving task through a fine-tuning curriculum split into four distinct subtasks.

tonomous driving (AD) software stacks, reducing the time, cost, and complexity typically associated with AD development.

Further sub-goals are:

- Online sample-efficient fine-tuning based on real-time DIF feedback in voice form and continuous unsupervised learning on our research vehicle EDGAR [22] (Fig.1).
- Enhanced Safety: rule-aware learning and a deployed predicitve safety filter.
- Exploration of real-time capabilities of LMFMs on our research vehicle EDGAR [22] (Fig.1).

II. Methodology

To address the identified challenges and meet our objectives, we propose a high-level multi-task learning curriculum (Fig. 2). **This approach breaks down the learning-to-drive task into four distinct subtasks/phases, mirroring the structured progression of a human-driving school curriculum.** Our goal is to enhance and facilitate the agent's learning experience by gradually increasing task complexity while reducing the level of expert guidance. Our starting point is an already trained LMFL with a general understanding of a world model, and we will expand its scene understanding, reasoning, and decision-making capabilities for the driving task on public streets through fine-tuning for specific tasks according to the learning curriculum (Fig. 2). Experiments will be carried out on our research vehicle EDGAR [22].

In **phase (I)**, our model initially learns theoretical driving rules (StVO) through supervised fine-tuning, utilizing a predefined scene understanding catalog of questions.

In **phase (II)**, we leverage a sophisticated simulation environment to learn technical navigation skills, guided by Driving Instructor Feedback/Suggestion (DIF) in voice form. We introduce a novel fine-tuning concept termed online iterative RDL-DIF, illustrated in Figure 3. This phase includes the iterative update of the Reward Model (RM) based on DIF and a Predictive Safety Filter Model incorporating DIF and simulation penalties ($r_{\mathrm{p,sim}}$), such as those for crashes. The safety filter dynamically monitors the RL policy operation and can intervene by switching to a fallback safety policy to prevent failures.

In **phase (III)**, the agent transitions to driving on public roads guided by DIF in a voice form and Driving Instructor Intervention (DII) in form of steering, brake and acceleration interventions. We push our agent to bridge the sim-2-real gap through continuous online iterative RDL-DIF fine-tuning of the RDL policy, RM-DIF model, and Safety Filter Model. Simulation penalties ($r_{\mathrm{p,sim}}$) are no longer available, replaced by DII penalties ($r_{\mathrm{p,DII}}$).

In **phase (IV)**, after the agent passes the practical driving license test, it starts collecting experience without DIF. Now, we apply the trained Safety Filter Model instead. The agent autonomously improves its driving skills in new inexperienced environments through continuous RL training without DIF. In this phase, RM-DIF (r_{DIF}) and penalties (r_p) are not applicable. **Our novel online iterative RDL-DIF approach distinguishes itself from current SOTA Reinforcement Learning from Human Feedback (RLHF) [23] in several key aspects:** (1) our reward model is trained in tandem with the fine-tuned policy during operation, eliminating the need for resource-intensive offline pretraining, (2) is based real-time voice feedback and action suggestions from humans, offering a more convenient and interactive approach, (3) our reward structure adapts dynamically to the learning phase, (4) we enhance safety of the fine-tuned output through an online-trained safety model, (5) our approach handels multimodal inputs.

References

[1] McKinsey & Company. Autonomous vehicles moving forward: Perspectives from industry leaders, 2023.

[2] Der Spiegel. Oktoberfest in München: Selbstfahrendes Shuttle im Wiesn-Härtetest, 2024.

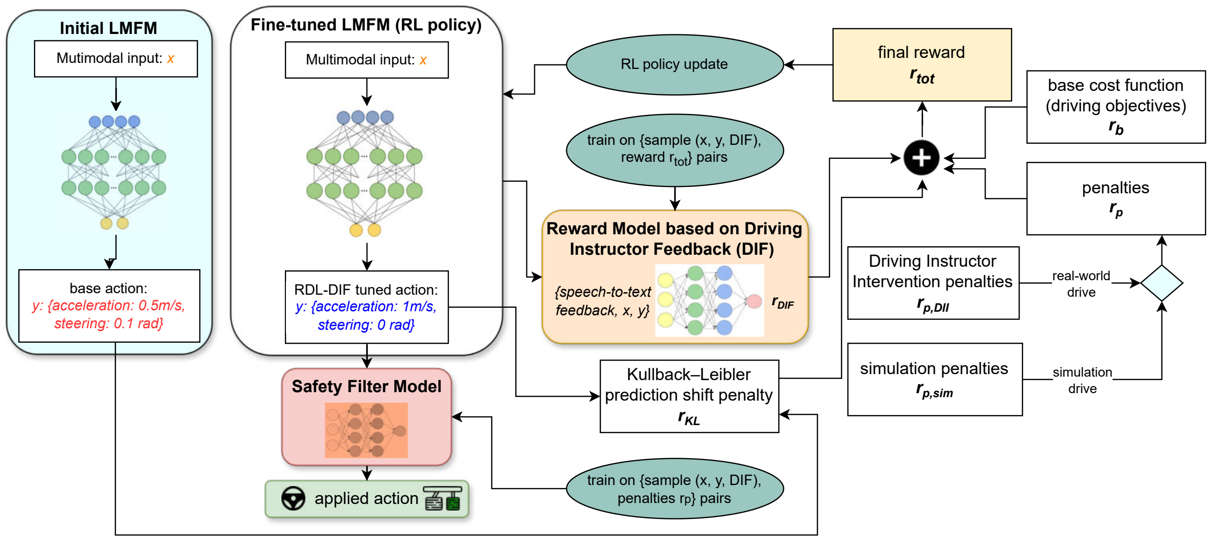

Fig. 3. Our Online Iterative Reinforcement Driving Learning from Driving Instructor Feedback/Suggestion (RDL-DIF). This architecture is valid for the fine-tuning curriculum Phases (I-IV).

[3] Süddeutsche Zeitung. Autonomes fahren zur wiesn: Kein chauffeur aus fleisch und blut – sondern aus software und kameras, 2024.

[4] Die Zeit. Autonomes Fahren: Autos und Technik auf dem Oktoberfest, 2024.

[5] A2RL. TUM Races to Victory at ASPIRE's Inaugural Abu Dhabi Autonomous Racing League at Yas Marina Circuit, 2024.

[6] Indy Autonomous Challenge. TUM Autonomous Motorsport Wins the Indy Autonomous Challenge Powered by Cisco at the Indianapolis Motor Speedway and the 1 Million Grand Prize, 2021. Accessed: 2024-11-07.

[7] Indy Autonomous Challenge. TUM Autonomous Motorsports Wins Third Annual Autonomous Challenge @ CES, 2024.

[8] Autoware Foundation. Autoware universe: An open-source autonomous driving system, 2024. Accessed: 2024-11-04.

[9] Waymo. Scaling Waymo One Safely Across Four Cities This Year, 2024.

[10] Johannes Betz, Hongrui Zheng, Alexander Liniger, Ugo Rosolia, Phillip Karle, Madhur Behl, Venkat Krovi, and Rahul Mangharam. Autonomous vehicles on the edge: A survey on autonomous vehicle racing. *IEEE Open Journal of Intelligent Transportation Systems*, 3:458–488, 2022.

[11] Johannes Betz, Tobias Betz, Felix Fent, Maximilian Geisslinger, Alexander Heilmeier, Leonhard Hermansdorfer, Thomas Herrmann, Sebastian Huch, Phillip Karle, Markus Lienkamp, et al. Tum autonomous motorsport: An autonomous racing software for the indy autonomous challenge. *Journal of Field Robotics*, 40(4):783–809, 2023.

[12] Felix Nobis, Maximilian Geisslinger, Markus Weber, Johannes Betz, and Markus Lienkamp. A deep learning-based radar and camera sensor fusion architecture for object detection. In *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–7. IEEE, 2019.

[13] Alexander Heilmeier, Alexander Wischnewski, Leonhard Hermansdorfer, Johannes Betz, Markus Lienkamp, and Boris Lohmann. Minimum curvature trajectory planning and control for an autonomous race car. *Vehicle System Dynamics*, 2020.

[14] Rainer Trauth, Korbinian Moller, Gerald Würsching, and Johannes Betz. Frenetix: A high-performance and modular motion planning framework for autonomous driving. *IEEE Access*, 2024.

[15] Baha Zarrouki, Chenyang Wang, and Johannes Betz. A stochastic nonlinear model predictive control with an uncertainty propagation horizon for autonomous vehicle motion control. In *2024 American Control Conference (ACC)*, pages 5466–5473. IEEE, 2024.

[16] Baha Zarrouki, Marios Spanakakis, and Johannes Betz. A safe reinforcement learning driven weights-varying model predictive control for autonomous vehicle motion control. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 1401–1408, 2024.

[17] Can Cui, Zichong Yang, Yupeng Zhou, Yunsheng Ma, Juanwu Lu, and Ziran Wang. Large language models for autonomous driving: Real-world experiments. *arXiv preprint arXiv:2312.09397*, 2023.

[18] Hao Shao, Yuxuan Hu, Letian Wang, Steven L Waslander, Yu Liu, and Hongsheng Li. Lmdrive: Closed-loop end-to-end driving with large language models. *arXiv preprint arXiv:2312.07488*, 2023.

[19] 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.

[20] Bu Jin, Xinyu Liu, Yupeng Zheng, Pengfei Li, Hao Zhao, Tong Zhang, Yuhang Zheng, Guyue Zhou, and Jingjing Liu. Adapt: Action-aware driving caption transformer. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7554–7561. IEEE, 2023.

[21] Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, and Gianluca Corrado. Gaia-1: A generative world model for autonomous driving. *arXiv preprint arXiv:2309.17080*, 2023.

[22] Phillip Karle, Tobias Betz, Marcin Bosk, Felix Fent, Nils Gehrke, Maximilian Geisslinger, Luis Gressenbuch, Philipp Hafemann, Sebastian Huber, Maximilian Hübner, et al. Edgar: An autonomous driving research platform–from feature development to real-world application. *arXiv preprint arXiv:2309.15492*, 2023.

[23] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.