HiVeGPT: Human-Machine-Augmented Intelligent Vehicles With Generative Pre-Trained Transformer

Junping Zhang, *Senior Member, IEEE*, Jian Pu, *Member, IEEE*, Jianru Xue, *Member, IEEE*, Ming Yang, *Member, IEEE*, Xin Xu, Xiao Wang, and Fei-Yue Wang, *Fellow, IEEE*

***Abstract*—Recently, a chat generative pre-trained transformer (ChatGPT) attracts widespread attention in the academies and industries because of its powerful conversational ability with human and its astonishing emergence ability such as admit mistakes, reject inappropriate problems. However, it is not easy to generalize ChatGPT into the field of intelligent vehicles because of its high computational cost, uncertain answers and decisions, and the difficulty of scenario generation for intelligent vehicles. To address these issues, we propose a novel framework, human-machine-augmented intelligent vehicles with generative pre-trained transformer. Under this framework, we discuss the potential, prospects, limitations and several typical applications of HiVeGPT in the domain of intelligent vehicles.**

***Index Terms*—ChatGPT, HiVeGPT, reinforcement learning, parallel learning, intelligent vehicles.**

I. Introduction

As a highly promising field, intelligent vehicles (IVs) have made significant progress in recent years. IVs are becoming smarter, more active and more convinced by human in the realm of intelligent transportation systems in recent decades [1]. IV-related technology has the potential to reduce traffic accidents, save human lives and properties, and alleviate psychological and physical burdens on human drivers.

Manuscript received 27 February 2023; accepted 1 March 2023. Date of publication 14 March 2023; date of current version 27 April 2023. This work was supported in part by the National Natural Science Foundation of China under Grants 62176059, 61825305, 62036008, 62173228, and 62173329, in part by the Shanghai Municipal Science and Technology Major Project under Grant 2018SHZDZX01, and in part by the Zhangjiang Laboratory (ZJLab) and the Shanghai Center for Brain Science and Brain-Inspired Technology, STI 2030-Major Projects under Grant 2021ZD0201300. *(Corresponding author: Fei-Yue Wang.)*

Junping Zhang is with the Shanghai Key Laboratory of Intelligent Information Processing, School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: jpzhang@fudan.edu.cn).

Jian Pu is with the Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai 200433, China (e-mail: jianpu@fudan.edu.cn).

Jianru Xue is with the National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Institute of Artificial Intelligence and Robotics, Xian Jiaotong University, Shaanxi 71009, China (e-mail: jrxue@mail.xjtu.edu.cn).

Ming Yang is with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: mingyang@sjtu.edu.cn).

Xin Xu is with the College of Intelligence Science, National University of Defense Technology, Changsha 410073, China (e-mail: xinxu@nudt.edu.cn).

Xiao Wang is with the School of Artificial Intelligence, Anhui University, Hefei 230039, China (e-mail: xiao.wang@ahu.edu.cn).

Fei-Yue Wang is with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: feiyue.wang@ia.ac.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIV.2023.3256982.

Digital Object Identifier 10.1109/TIV.2023.3256982

However, generating safe, robust, agile, and dexterous driving behaviors in open, dynamic traffic environments still face great challenges.

From the aspect of intelligent vehicles, the challenging can be summarized into three problems as follows.

1) *Deep understanding for navigation:* Current self-localization and navigation technologies still encounter difficulties in reasoning about free-space, obstacles and the topology of the environment, especially in adopting common sense rules and heuristics. A reason is that IVs lack of holistic understanding of scene elements and geometric layout, traffic participants and their motion patterns [2], [3]. The development of situation reasoning including clarifying relationships among traffic participants in terms of 3D object models, scene layout elements and traffic rules are still at their early stages.
2) *Situation prediction for motion planning:* Motion planning [4], [5] heavily depends on the prediction, which includes future spatial-temporal relationships among traffic participants including inferring and predicting physical and temporal relationships, and embedding traffic common sense knowledge and causal inference [6], [7].
3) *Human-like driving policy:* Complexity, uncertainty, and ambiguity are intrinsic problems in dynamic decision-making [8], [9]. Current driving models demonstrate unsatisfying performance in capacities of adaptivity, reliability, and learning to learn in real traffic scenarios.

In order to address the aforementioned challenges, research efforts towards human-like driving are important frontiers by utilizing artificial intelligence-related techniques [4].

In the early days of 2023, fortunately, Chat Generative Pre-trained Transformer (ChatGPT) gains widespread attention because of its powerful ability to converse with humans [10], [11]. Its success highlights the importance of several factors, including the volume of data, the number of GPUs, and powerful pre-trained models. These factors have proven crucial in advancing from traditional learning frameworks to the current ChatGPT-like learning framework [12].

Since it has not been extensively studied in the domain of Intelligent Vehicles, integrating ChatGPT-inspired learning into IV technology seems a natural and straightforward way for further advancements of IVs.

However, existing IV techniques fall short of meeting people's expectations on a large scale due to several bottlenecks. These bottlenecks include: 1) a lack of efficient methodologies and tools to analyze the vast amounts of data collected from IVs; 2) limited computational sources due to the limited space within each vehicle; 3) an absence of high-level intelligent behavior to deal with some emergent accidents, such as those involving human beings [13], [14]. These bottlenecks result in that it is necessary to redesign some basic modules and introduce new functions so that the potenital of ChatGPT can also be served for IVs.

Therefore, this paper proposes a framework of Human-Machine Augmented Intelligent Vehicles with Generative Pretrained Transformer (HiVeGPT), aiming to identify critical areas where the idea of ChatGPT can be utilized to improve IV technology by combing it with human-machine augmented intelligence [15], specifically in algorithm design, perception ability, the internet of vehicles, and hardware limitations.

In the remainder of this paper, we will introduce the history and preliminary of ChatGPT in Section II, and discuss the potential and prospects of HiVeGPT in Section III. We will raise some limitations of HiVeGPT in Section IV, and propose some potential applications in Section V. Finally, we will conclude this paper in Section VI.

II. The Preliminary of ChatGPT

While ChatGPT makes a huge achievement in 2023, its history can be traced back to the 1960 s in at least three areas, i.e., knowledge collection, conservation ability, and the algorithmic design of natural language processing.

One of ChatGPT's key features is its embedded knowledge graph system, which improves its ability of answering user questions and making decision. Actually, the earliest example of an expert system that automated decision-making and exhibited problem-solving behavior was DENDRAL, which was developed in 1965 and successfully implemented in 1973. Almost at the same time, another famous expert system called MYCIN is proposed. Then CYC, an AI project initiated in the 1980 s, attempted to collect vast amounts of knowledge but was initially deemed unsuccessful. Still, this project provided insight into how to build a knowledge graph and led to Google's proposal of the semantic graph, which has become a popular way of utilizing knowledge.

Besides the knowledge graph, ChatGPT is closely related to the field of multi-turn dialogue through Chain of Thought (CoT) [16]. Because of CoT, people tends to recognize that ChatGPT answers each problem in a human-like thinking manner step by step. In fact, Eliza, a groundbreaking chatbot developed in the 1960s, was regarded as a milestone because it managed to convince some people that it could understand people's emotions, even though it actually lacked this capability. Eliza paved the way for subsequent chatbot applications, such as SHRDLU and Microsoft's Tay, which gained popularity since its release in 2014.

Furthermore, Natural Language Processing (NLP) made a breakthrough in 2017 with the introduction of the transformer model [17], which enabled the development of Large Language Models (LLMs) capable of understanding human language. Since then, a series of LLMs have been developed, including GPT-1, BERT, GPT-2, GPT-3 [18], InstructGPT [19], [20], and ChatGPT. Note that in GPT-3, three crucial abilities are introduced. They are: 1) prompt learning, a natural human-machine-interactive way based on language models; 2) In-context learning (ICL) that allows users to construct models for a new case without the need of fine-tuning and storing new parameters for each new task; 3) World knowledge including factual knowledge and commonsense. Different from the previous LLMs, in addition, ChatGPT incorporates Reinforcement Learning from Human Feedback (RLHF) [21], resulting in a significant improvement in performance. It forms some astonishing emergent abilities [22] like admitting mistakes, rejecting illegal answers [23], and considering ethical concerns, that make it stand out from the other models. A detailed discussion on the pros and cons of ChatGPT can also be seen in Zhou et al.'s comment paper [11].

Such emergent abilities are useful in avoiding some traffic accidents. However, ChatGPT has its disadvantages. Specifically, its math and logic inference is weak. For example, a simple addition '2+2222' can be given a ridiculous answer 2244 by ChatGPT. However, it is obvious that IV requires an exact result for such a calculation.

Therefore, we propose an alternative framework of ChatGPT, i.e., HiVeGPT, by analyzing its potential and prospects around intelligent vehicles, which is the goal of this paper.

III. The Framework of HiVeGPT in IV

In this section, we will introduce how to apply HiVeGPT into intelligent vehicles. We will focus our attention on four parts that are important to the development of intelligent vehicles. They are perception and cognitive computing, decision, human-in-the-loop, and parallel learning. For better understanding our proposed HiVeGPT and its functions, we show its structure as in Fig. 1.

A. Perception, Cognitive Computing, and HiVeGPT

Perception is one of the essential components for intelligent vehicles with the camera, LiDAR, and Radar sensors. Current research efforts in environment perception for intelligent vehicle systems are pursuing adaptability to real traffic scenarios mixed with pedestrians, motor cyclists and other traffic participants [24], [25]. The core of this long-standing challenge problem is developing effective dynamic scene pattern representation and prediction technologies [7].

Traditional approaches to understanding and prediction of dynamic scenarios for intelligent driving generally adopt a modular pipeline computing architecture, which spans from traffic participants' detection, tracking to motion prediction [26]. The introduction of deep learning algorithms has continuously improved the performance of each module, but the limitations of modular pipeline computing architecture are becoming increasingly prominent: each problem is solved independently and sequentially, uncertainty is difficult to handle. In particular,

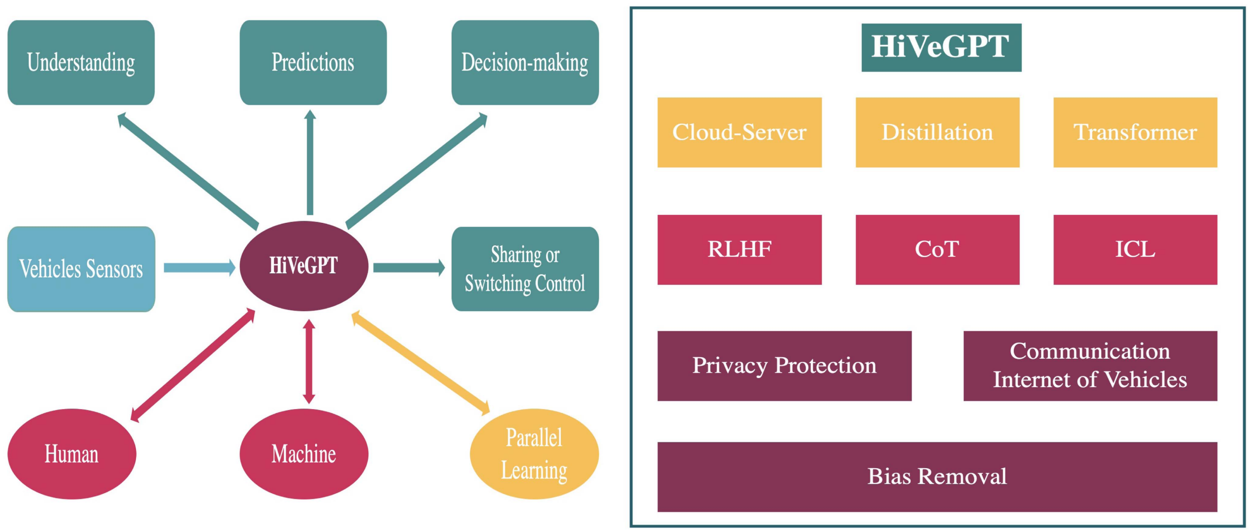

Fig. 1. The structure of HiVeGPT. Left: Human-machine-augmented intelligent vehicles with generative pre-trained model for perceiving and generating data, as well as dealing with several hard issues in the field of intelligent vehicles. Here bi-directional arrows mean that there exists interaction between two connected modules. Right: Several crucial modules of HiVeGPT to be included in it.

correlation and feedback among modules are too hard to implement, which make the perception system unable to understand complex traffic situations. In recent years, the end-to-end deep neural networks from sensing to prediction has overcome the limitations of modular pipeline computing architecture to a certain extent, but it is overly dependent on training data, and still difficult to perform spatiotemporal association reasoning in terms of traffic participants unique to dynamic traffic scenarios.

With the development of deep learning methods, foundation models are also introduced into the perception tasks [24]. People hope to find a feasible and efficient way to implement the human-like cognitive mechanisms of human attention, memory and prediction of complex traffic scenarios in intelligent vehicle [27]. However, designing a loss function to capture these human-like cognitive attributes seems intractable. Most neural network models designed for this goal are still trained with one or more simple prediction loss functions. Some complex measures for perception are hard to optimize directly, for instance, mAP (mean Average Precision) for object detection and PQ (Panoptic Quality) for panoptic segmentation, the misalignment between the prediction and the requirement of autonomous driving are noted [28]. To compensate for the shortcoming of the loss itself, we need to define metrics that are designed to better capture cognitive mechanism of human. The same as ChatGPT did by using RLHF, one promising way of constructing HiVeGPT is to introduce the feedback as a reward [9], [28], whether from humans or from ground truth labels, to finetune or instruct tune the perception module to meet the requirements of autonomous driving.

Furthermore, mastering world knowledge should be an indispensable part of perception systems. It has been found that ChatGPT can learn some factual knowledge and also have the ability to understand the language text related to the knowledge. However, the current research lacks in-depth research work on the knowledge cognitive ability of ChatGPT and HiVeGPT. We thus need to explore the ways in which HiVeGPT and large-scale knowledge graphs can help and enhance each other from two aspects including dynamically acquiring the latest knowledge, and making targeted enhancements for long-tail facts of driving situations [26].

B. Decision-Making, Human-in-The-Loop and HiVeGPT

The driving decision-making module is at the core of an IV system, requiring a high degree of accuracy and a low tolerance for errors. The design of driving policies also demands caution and careful consideration, which directly determine how intelligent IVs act in traffic situations.

Decision-making in real traffic situations should consider two types of uncertainties: uncertainty about the effect of action execution (hereinafter referred to as action uncertainty), and uncertainty in the perception (hereinafter referred to as perceptual uncertainty). Traditionally, it is often assumed that the configured sensor can measure the full state of the environment, but the sensor suite of an IV in an open, dynamic environment cannot meet the conditions for the state to be fully observable. Therefore, decision-making requires comprehensively assess to the uncertainty of the current situation.

To address the aforementioned uncertainties, the majority of research efforts of the past decade has been put into the applications of reinforcement learning methods [9], [29]. Reward-is-enough hypothesis is the idea that has an algorithm that maximizes a fundamental reward is sufficient to develop all of the specific behaviours that we observe in complex environments [30]. However, the reward functions are hard to be well defined to improve the learning efficiency of RL-based decision-making systems [31]. To solve this issue, one feasible way is to utilize Reinforcement Learning from Human Feedback (RLHF), which is an important module in optimizing the conversational performance of the natural language processing of ChatGPT. The reason is that by using the principle of RLHF, human experiences will be helpful to iteratively reduce the difficulty in designing reward functions.

Besides, there are alternative approaches to utilizing this model in the decision task without compromising the system's safety. The development of the decision module requires data on human reactions under different scenarios [32]. Previously, collecting this type of data needed high-precision instruments, but ChatGPT may simplify this process by allowing a wider value deviation range and the integration of different format data with less effort. Similar to ChatGPT, HiVeGPT should have a higher-tolerant ability because of the introduction of human-in-the-loop where a human decision can provide higher safety for IV.

Human-in-the-loop refers to the concept of human involvement in a system as a crucial role. In the context of IV, human-in-the-loop is essential during both the development and exploitation stages.

- *Development Stage:* During the development process, annotation, which involves adding metadata or labels to data sets, is an inevitable and resource-consuming task. This process is indispensable in training learning-based algorithms to recognize patterns and make accurate predictions [33]. While the development phase relies heavily on human input, we can resort to the idea of ChatGPT to help HiVeGPT better mine online data related to traffic scenarios. The reason is that ChatGPT can extract descriptive features of scenarios from subjective and casual texts, and then convert them into formal expressions. Consequently, HiVeGPT enables better exploitation of massive online texts with high variation.
- *Exploitation Stage:* When it comes to the exploitation stage, ChatGPT's outstanding performance among the existing dialogue system makes it become a potential choice for bridging the communication between the passengers and the IV system [34]. With mature speech recognition technology, ChatGPT can efficiently and accurately recognize passengers' commands and intentions, improving the overall user experience.

Furthermore, there exists another kind of human-in-the-loop, where human drivers will play a significant role in the driving process. In this case, human drivers steer an intelligent vehicle with a driving assistance system. The driving assistance system can provide the driver with enhanced sensing, decision-making, and sharing control functions. During the human-machine cooperative driving process, HiVeGPT can thus be used to promote the interaction and understanding between the human driver and the driving assistance system. This kind of human-machine cooperation will be beneficial to increase driving safety and reduce the working load of human drivers [35].

C. Scenario Generation, Parallel Learning and HiVeGPT

Under the framework of HiVeGPT, a large number of challenging scenarios can generate through prompt learning, soft-prompt learning, and even parallel learning.

For example, corner cases are the most challenging problem for high-level autonomous driving. According to Zipf's law, the scenarios and object categories follow a long-tailed distribution [36]. Such a phenomenon is also called the long tail effect, which is characterized by the highly frequent occurrence of normal scenarios and the scarce appearance of extreme scenarios. As a result, the data scarcity, imbalance, and lacking labels further limit the performance of machine learning or deep learning methods, which need a large number of data for training.

However, obtaining sufficient data for each scenario with rare objects is extremely expensive, and even dangerous. Alternatively, scenario generation is used to tackle the issue by generating interpretable, controllable, and diversified scenarios with rare objects.

To this end, the parallel vision method is proposed [37]. After then, researchers further explored background subtraction [38], scene-specific pedestrian detection [39], the realistic traffic image generation algorithms based on parallel vision. Wang et al. [40] studied how to regularize the long-tail scenarios under the infrastructure of parallel vision, which is supported by their research of the IVFC (Intelligent Vehicle Future Challenge of China) autonomous driving test.

As an alternative way, ChatGPT and HiVeGPT are also able to automatically generate a large number of diverse configuration files for existing simulation tools and generate expected scenarios. Moreover, when HiVeGPT is combined with more sophisticated generative methods, for instance, diffusion models [41], we would like to generate scenarios for autonomous driving from prompt learning or human-machine-augmented strategy.

Note that although virtual data generated by synthesizing or simulation have attracted increasing attention in recent years, how to make sure the generated data appropriately represent the data distribution in our physical world and the quality of data labeling are still of great challenge. To address this issue and provide better serving for intelligent vehicles, we can build a parallel learning system, which constructs a closed-loop virtual-to-real and real-to-virtual architecture for data interactions and policy deployment in HiVeGPT [42].

With such a parallel system, we can generate unprecedented traffic data related to intelligent vehicles. Meanwhile, human feedback can be replaced by feedback from virtual robotics. For example, a large amount of steering actions can be generated through parallel learning for handling the unmodeled dynamics and external disturbance once given the previewed trajectory points from a virtual world [43].

Although the success of ChatGPT stems from the collection of a large amount of data from the Internet and from the use of human feedback, there exist two problems. One is that the amount of data is still limited and the other one is human feedback is time-consuming and usually slow. In contrary, parallel learning system will learn and imitate human drivers' behaviors automatically, alleviating the human resources involved.

What's more, any virtual world can be sped up so that we can train an intelligent vehicle system much faster than the training speed that ChatGPT can achieve. Since traffic accidents are dangerous that can result in causalities, it is more feasible to utilize a parallel virtual system to simulate and generate such accidents for helping intelligent vehicles improve their performance in the real world.

IV. Limitations

In this section, we will discuss some limitations of HiVeGPT from five aspects including computational cost, data privacy and biases, Moravec's paradox and ethical issues, as well as other potential risks for intelligent vehicles.

A. Computational Costs

Generally speaking, the size of intelligent vehicles is the same as that of commonly-used vehicles. Therefore, the remained space can be used for intelligent computation is limited. As we know, however, ChatGPT is estimated of using 10,000 A100 GPUs and has 175 billion parameters. It is impossible to allow such a huge computational platform like ChatGPT is deployed in a real vehicle.

The costs are also a big problem. ChatGPT actually costs 12 million dollars for training. It may significantly increase the cost of each intelligent vehicle to be sold.

To solve these problems, we can consider some lightweight-level pre-trained transformers or distillation learning so that the cost of HiVeGPT can be reduced and the requirement of hardware can be decreased. The other way is to utilize cloud server-based computational platform and internet of vehicles so that the computation can be distributively implemented. With this way, we can deploy the HiVeGPT into intelligent vehicles without the significant increase of costs.

B. Data Privacy and Biases

The data of ChatGPT used mainly come from internet. Different from internet, intelligent vehicles keep more private space for drivers and passengers, which shouldn't been accessed by ChatGPT or HiVeGPT or similar models. A compromise between sharing data and keeping privacy is important to the development of intelligent vehicles.

Furthermore, biases should be avoided since they will influence the fairness of each intelligent vehicle. However, it seems a difficult problem that is hard to address according to the conditions of current ChatGPT or other AI products. We can see a lot of biases in it, including data biases, statistical biases, model designing biases and so on.

C. Moravec's Paradox

Moravec's paradox, proposed at the early 1970 s, pointed out that hard problem for AI is easy for human beings, whereas easy problem for AI is hard for human beings [44]. This paradox is often seen in a scenario that a man can smoothly deal with complex traffic scenario at a crossroad, for example, signal-free intersections [45], whereas self-driving car will stuck in a rut at the same scenario because of low fault-tolerant ability. Meanwhile, in a long-distance highway, driver may have a dangerous status of fatigue driving whereas self-driving car haven't such a status and thus leverage the safety of driving car.

A similar paradox can also be observed in ChatGPT. For example, it cannot give a correct math computation because of lacking a real computational module, even though such a computation is a simple addition or subtraction. Meanwhile, it has a powerful emergence abilities such as inference, correction and has a huge-volume data to help itself answer complex and continuous problems from users.

As a result, only depending on human driver or self-driving intelligent ChatGPT is not safe. Human-machine augmented intelligence is one feasible way to address this paradox by introducing either human-in-the-loop, or cognitive computing, or both when we design a HiVeGPT-based model.

D. Ethical Issues

Actually, ethical issues are always accompanied with the development of intelligent vehicles. For example, recently, the classical 'trolley problem' in ethics is reconsidered seriously as the study of self-driving car becomes hot. Furthermore, when a traffic accident happened, an ethical issue raises around that intelligent operate system or human or the manufacturer made the car, which one should be responsible for this accident. Since these are closely related to the right of making decision by human or machine, we should consider a good balance between HiVeGPT and human for obtaining safety guarantee, especially in uncertainty environment where the trajectories of surround vehicles are difficult to predict [46], [47]. Note that we should also have the ability of detecting driver abnormalities [14] such as driver distraction [48], [49], road rage, extreme fatigue and paroxysmal disease, as well as abnormal longitudinal driver behavior [50] so that machine can take over the control right to intelligent vehicle for avoiding unnecessary traffic accidents, and vice versa when machine makes fatal decision to drivers. Furthermore, some up-to-dated reinforcement learning strategies including reinforcement from human feedback, driver behavior cloning [51] and hierarchical interpretable imitation learning [5] can more or less alleviate the risks stemmed from uncertainty environments [52] or unexpected pedestrian intrusion [13].

E. Other Potential Risks

Although ChatGPT has shown remarkable success in conversational systems, relying on it for the driving decision-making module of an IV is impractical and risky. ChatGPT is not proficient in logic induction and may produce factual fallacies, which could lead to catastrophic consequences. Therefore, integrating ChatGPT directly into the decision-making process should be avoided. The same issue should be considered by HiVeGPT.

We should also note that while ChatGPT lacks human-like common sense, it is compensated with a stronger "imagination". In future studies, the researchers may ask HiVeGPT for proposals and shift the task from scenario generation to scenario selection, which could improve the model development workflow, making it more comprehensive, efficient, and easier.

V. Potential Applications

In this section, we will discuss three typical applications of HiVeGPT in the intelligent vehicle field. Note that we won't enumerate all the possible applications of HiVeGPT for saving pages.

A. Automated Parking

Automated parking is an important function of self-driving car, reflecting the intelligent level in planning the parking trajectories [53]. Due to the complicated and changeable parking environments, it is hard to make self-driving car have a strong adaptive parking capacities without human guidance. A feasible way is to collect some experienced drivers' parking data and generate a large amount of similar simulated data through parallel learning [54], [55], so that HiVeGPT can be trained with the mixture of huge-volume data based on a good generative pre-trained transformer.

B. Automatic Car-Following and Adaptive Cruise Control

Automatic Car-following and adaptive cruise control (ACC) are two crucial indices related to the success of intelligent vehicle [4]. However, it is dangerous to only rely on intelligent algorithm to make decision under all road geometry characteristics (curvature, slope and superelevation), especially in a particular sharp curve [56]. If we can integrate human-like decision in which car-following model is often natural and smooth with data-driven modeling [57], [58], then intelligent vehicles will have better passenger experiences and the potential threaten of longitudinal and mixed vehicle platooning, such as destabilizing attacks that cause vehicle collisions [59] can be greatly mitigated. To achieve this function, one feasible way is to utilize human-machine augmented intelligence [25], [47] to make HiVeGPT deal with different car-following scenarios robustly.

C. Intelligent Cockpit

The intelligent cockpit is a product of integrating modern sensors, new materials, artificial intelligence, visualization, and interactions, which have developed rapidly in recent years. It can usually be divided into three layers: perception, cognition and decision, and interaction [60]. Apart from the autonomous driving system's perception of the exterior environment, the intelligent cockpit should also be able to perceive the driver and passengers in the cockpit. The well-known perception of the driver is the driver monitoring systems (DMS) [61]. For passengers, perception of tactile information, linguistic speech, and nonlinguistic speech is usually obtained. Then the perception is sent to the cognition and decision layer to detect human behavior, predict human intention, and generate decision strategies. Finally, following the guidance of the cognition and decision layer, the interaction layer provides service to the interaction with both humans in and outside the cockpit.

An immediate application of HiVeGPT is to refine the cognition and decision layer in a comprehensive dialogue way to answer follow-up questions, admit mistakes, challenge incorrect premises, and reject inappropriate and hazardous requests. Moreover, it would make the intelligent vehicle smart and comfortable if combined HiVeGPT with advanced perception and interaction methods.

VI. Conclusion

In this paper, we propose a new framework for the development of intelligent vehicles, i.e., HiVeGPT. It introduces the basic modules of ChatGPT, and then optimizes these modules by merging the characteristics of intelligent vehicles, and proposes? several novel strategies including parallel learning, human-machine-augmented strategies. It should more or less reflect the potential developed direction of intelligent vehicles. In the future, how to make HiVeGPT achieve mutually trustworthy human-machine augmented intelligence in the field of intelligent vehicles deserve further studying [15], [62].

References

[1] F.-Y. Wang, "A new phase of ieee transactions on intelligent vehicles: Being smart, becoming active, and believing intelligent vehicles," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 3–15, Jan. 2023.

[2] D. Cui, J. Xue, and N. Zheng, "Real-time global localization of robotic cars in lane level via lane marking detection and shape registration," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1039–1050, Apr. 2016.

[3] X. Chen, J. Xue, and S. Pang, "Sparse semantic map-based monocular localization in traffic scenes using learned 2D-3D point-line correspondences," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 11894–11901, Oct. 2022.

[4] C. Wei, E. Paschalidis, N. Merat, A. S. Crusat, and F. Hajiseyedjavadi, "Human-like decision making and motion control for smooth and natural car following," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 263–274, Jan. 2023.

[5] S. Teng, L. Chen, Y. Ai, Y. Zhou, Z. Xuanyuan, and X. Hu, "Hierarchical interpretable imitation learning for end-to-end autonomous driving," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 673–683, Jan. 2023.

[6] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, "SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12077–12086.

[7] P. Zhang, J. Xue, P. Zhang, N. Zheng, and W. Ouyang, "Social-aware pedestrian trajectory prediction via states refinement LSTM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2742–2759, May 2022.

[8] X. Tang et al., "Prediction-uncertainty-aware decision-making for autonomous vehicles," *IEEE Trans. Intell. Veh.*, vol. 7, no. 4, pp. 849–862, Apr. 2022.

[9] Y. Pan, J. Xue, P. Zhang, W. Ouyang, J. Fang, and X. Chen, "Navigation command matching for vision-based autonomous driving," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 4343–4349.

[10] J. Schulman et al., "ChatGPT: Optimizing language models for dialogue," in *OpenAI Blog*, 2022.

[11] J. Zhou, P. Ke, X.-P. Qiu, M.-L. Huang, and J. Zhang, "ChatGPT: Potential, prospects and limitations," *Front. Inf. Technol. Electron. Eng.*, 2023. [Online]. Available: https://doi.org/10.1631/FITEE.2300089

[12] W. Fei-Yue, M. Qinghai, L. Quan, W. Xingxia, and L. Yilun, "What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence," *IEEE/CAA J. Automatica Sinica*, vol. 10, no. 3, pp. 575–579, 2023.

[13] Z. Shi, S. He, J. Sun, T. Chen, J. Chen, and H. Dong, "An efficient multi-task network for pedestrian intrusion detection," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 649–660, Jan. 2023.

[14] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, "Driver anomaly quantification for intelligent vehicles: A contrastive learning approach with representation clustering," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 37–41, Jan. 2023.

[15] J. Xue, B. Hu, L. Li, and J. Zhang, "Human-machine augmented intelligence: Research and applications," *Front. Inf. Technol. Electron. Eng.*, vol. 23, pp. 1139–1141, 2022.

[16] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022.

[17] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[18] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 1877–1901.

[19] J. Wei et al., "Finetuned language models are zero-shot learners," in *Proc. 10th Int. Conf. Learn. Representations*, 2022.

[20] L. Ouyang et al., "Training language models to follow instructions with human feedback," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022.

[21] N. Stiennon et al., "Learning to summarize with human feedback," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 3008–3021.

[22] J. Wei et al., "Emergent abilities of large language models," *Trans. Mach. Learn. Res.*, 2022.

[23] A. Glaese et al., "Improving alignment of dialogue agents via targeted human judgements," 2022, *arXiv:2209.14375*.

[24] J. Shao et al., "INTERN: A new learning paradigm towards general vision," 2021, *arXiv:2111.08687*.

[25] J. Xue, D. Wang, S. Du, D. Cui, and Y. Huang, "A vision-centered multi-sensor fusing approach to self-localization and obstacle perception for robotic cars," *Front. Inf. Technol. Electron. Eng.*, vol. 18, pp. 122–138, 2017.

[26] J. Xue et al., "BLVD: Building a large-scale 5D semantics benchmark for autonomous driving," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2019, pp. 6685–6691.

[27] J. Fang, D. Yan, J. Qiao, J. Xue, and H. Yu, "DADA: Driver attention prediction in driving accident scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 4959–4971, Jun. 2022.

[28] A. S. Pinto, A. Kolesnikov, Y. Shi, L. Beyer, and X. Zhai, "Tuning computer vision models with task rewards," 2023, *arXiv:2302.08242*.

[29] B. R. Kiran et al., "Deep reinforcement learning for autonomous driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 4909–4926, Jun. 2021.

[30] D. Silver, S. Singh, D. Precup, and R. S. Sutton, "Reward is enough," *Artif. Intell.*, vol. 299, 2021, Art. no. 103535.

[31] W. Yuan, Y. Li, H. Zhuang, C. Wang, and M. Yang, "Prioritized experience replay-based deep Q learning: Multiple-reward architecture for highway driving decision making," *IEEE Robot. Automat. Mag.*, vol. 28, no. 4, pp. 21–31, Apr. 2021.

[32] F. Hauer, T. Schmidt, B. Holzmüller, and A. Pretschner, "Did we test all scenarios for automated and autonomous driving systems?," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2019, pp. 2950–2955.

[33] Y. Qian, X. Wang, Z. Chen, C. Wang, and M. Yang, "Hy-Seg: A hybrid method for ground segmentation using point clouds," *IEEE Trans. Intell. Veh.*, vol. 8, no. 2, pp. 1597–1606, 2023.

[34] P. A. M. Ruijten, J. M. B. Terken, and S. N. Chandramouli, "Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior," *Multimodal Technol. Interact.*, vol. 2, no. 4, 2018, Art. no. 62.

[35] C. He, D. Sun, Y. Li, M. Zhao, and W. Liu, "Hierarchical human-vehicle collaboration control strategy for intelligent vehicle under human-cyber-physical system architecture," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst.*, 2022, pp. 1605–1610.

[36] X. Gu, T.-Y. Lin, W. Kuo, and Y. Cui, "Open-vocabulary object detection via vision and language knowledge distillation," in *Proc. Int. Conf. Learn. Representations*, 2022.

[37] K. Wang, C. Gou, N. Zheng, J. M. Rehg, and F.-Y. Wang, "Parallel vision for perception and understanding of complex scenes: Methods, framework, and perspectives," *Artif. Intell. Rev.*, vol. 48, no. 3, pp. 299–329, 2017.

[38] W. Zhang, K. Wang, Y. Liu, Y. Lu, and F.-Y. Wang, "A parallel vision approach to scene-specific pedestrian detection," *Neurocomputing*, vol. 394, pp. 114–126, 2020.

[39] W. Zheng, K. Wang, and F.-Y. Wang, "A novel background subtraction algorithm based on parallel vision and bayesian GANs," *Neurocomputing*, vol. 394, pp. 178–200, 2020.

[40] J. Wang et al., "Parallel vision for long-tail regularization: Initial results from IVFC autonomous driving testing," *IEEE Trans. Intell. Veh.*, vol. 7, no. 2, pp. 286–299, Feb. 2022.

[41] F. A. Croitoru, V. Hondru, R. T. Lonescu, and M. Shah, "Diffusion models in vision: A survey," 2022, *arXiv:2209.04747*.

[42] Q. Miao, Y. Lv, M. Huang, X. Wang, and F. Y. Wang, "Parallel learning: Overview and perspective for computational learning across Syn2Real and Sim2Real," *IEEE/CAA J. Automatica Sinica*, vol. 10, no. 3, pp. 603–631, 2023.

[43] F. Tian, Z. Li, F.-Y. Wang, and L. Li, "Parallel learning-based steering control for autonomous driving," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 379–389, Jan. 2023.

[44] H. Moravec, *Mind Children*. Cambridge, MA, USA: Harvard Univ. Press, 1988.

[45] H. Pei, J. Zhang, Y. Zhang, X. Pei, S. Feng, and L. Li, "Fault-tolerant cooperative driving at signal-free intersections," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 121–134, Jan. 2023.

[46] T. Brüdigam, M. Olbrich, D. Wollherr, and M. Leibold, "Stochastic model predictive control with a safety guarantee for automated driving," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 22–36, Jan. 2023.

[47] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, "A survey on trajectory-prediction methods for autonomous driving," *IEEE Trans. Intell. Veh.*, vol. 7, no. 3, pp. 652–674, Mar. 2022.

[48] S. Jha and C. Busso, "Estimation of driver's gaze region from head position and orientation using probabilistic confidence regions," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 59–72, Jan. 2023.

[49] C. Gou, Y. Zhou, Y. Xiao, X. Wang, and H. Yu, "Cascade learning for driver facial monitoring," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 404–412, Jan. 2023.

[50] G. Sidorenko, A. Fedorov, J. Thunberg, and A. Vinel, "Towards a complete safety framework for longitudinal driving," *IEEE Trans. Intell. Veh.*, vol. 7, no. 4, pp. 809–814, Apr. 2022.

[51] L. Wang, C. Fernandez, and C. Stiller, "High-level decision making for automated highway driving via behavior cloning," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 923–935, Jan. 2023.

[52] J. Wu, Z. Huang, and C. Lv, "Uncertainty-aware model-based reinforcement learning: Methodology and application in autonomous driving," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 194–203, Jan. 2023.

[53] B. Li, L. Fan, Y. Ouyang, X. Wang, D. Cao, and F.-Y. Wang, "Online competition of trajectory planning for automated parking: Benchmarks, achievements, learned lessons, and future perspectives," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 16–21, Jan. 2023.

[54] L. Li, N. Zheng, and F.-Y. Wang, "Parallel learning: A perspective and a framework," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 3, pp. 389–395, Mar. 2017.

[55] X. Li, K. Wang, Y. Tian, L. Yan, F. Deng, and F. -Y. Wang, "The paralleleye dataset: A large collection of virtual images for traffic vision research," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2072–2084, Jun. 2019.

[56] M. Waqas and P. Ioannou, "Automatic vehicle following under safety, comfort, and road geometry constraints," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 531–546, Jan. 2023.

[57] J. Zhang, F. Y. Wang, K. Wang, W. H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Apr. 2011.

[58] J. Zhan, Z. Ma, and L. Zhang, "Data-driven modeling and distributed predictive control of mixed vehicle platoons," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 572–582, Jan. 2023.

[59] E. Khanapuri, T. Chintalapati, R. Sharma, and R. Gerdes, "Learning based longitudinal vehicle platooning threat detection, identification and mitigation," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 290–300, Jan. 2023.

[60] W. Li et al., "CogemoNet: A cognitive-feature-augmented driver emotion recognition model for smart cockpit," *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 3, pp. 667–678, Mar. 2021.

[61] A. Koesdwiady, R. Soua, F. Karray, and M. S. Kamel, "Recent trends in driver safety monitoring systems: State of the art and challenges," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4550–4563, Jun. 2016.

[62] G. B. J. J. Z. Fei-Yue and W. J. GUO, "Mutually trustworthy human-machine knowledge automation and hybrid augmented intelligence: Mechanisms and applications of cognition, management, and control for complex systems," *Front. Inf. Technol. Electron. Eng.*, vol. 23, no. 8, pp. 1142–1147, 2022.