Generative AI-Empowered Simulation for Autonomous Driving in Vehicular Mixed Reality Metaverses

Minrui Xu, *Graduate Student Member, IEEE*, Dusit Niyato, *Fellow, IEEE*, Junlong Chen, Hongliang Zhang, *Member, IEEE*, Jiawen Kang, *Member, IEEE*, Zehui Xiong, *Member, IEEE*, Shiwen Mao, *Fellow, IEEE*, and Zhu Han, *Fellow, IEEE*

***Abstract*—In the vehicular mixed reality (MR) Metaverse, the discrepancy between physical and virtual entities can be overcome by fusing the physical and virtual environments with multi-dimensional communications in autonomous driving systems. Assisted by digital twin (DT) technologies, connected autonomous vehicles (AVs), roadside units (RSUs), and virtual simulators can maintain the vehicular MR Metaverse via simulations for sharing data and making driving decisions collaboratively. However, it is challenging and costly to enable large-scale traffic and driving simulation via realistic data collection and fusion from the physical world for online prediction and offline training in autonomous driving systems. In this paper, we propose an autonomous driving architecture, where generative AI is leveraged to synthesize unlimited conditioned traffic and driving data via simulations for improving driving safety and traffic control efficiency. First, we propose a multi-task DT offloading model for the reliable execution of heterogeneous DT tasks with different requirements at RSUs. Then, based on the preferences of AV's DTs and real-world data, virtual simulators can synthesize unlimited conditioned driving and traffic datasets for improved robustness. Finally, we propose a multi-task enhanced auction-based mechanism to provide fine-grained incentives for RSUs on providing resources for autonomous driving. The property analysis and experimental results demonstrate that the proposed mechanism and architecture are strategy-proof and effective.**

***Index Terms*—Autonomous driving, Metaverse, generative artificial intelligence, auction theory.**

Manuscript received 15 February 2023; revised 11 June 2023; accepted 25 June 2023. Date of publication 10 July 2023; date of current version 16 November 2023. The work of Shiwen Mao was supported by the NSF under Grant CNS-2148382. This work was supported in part by the National Key R&D Project of China under Grant 2022YFE0111900, in part by the NSFC under Grant 62102099, in part by the Pearl River Talent Recruitment Program under Grant 2021QN02S643, in part by the Guangzhou Basic Research Program under Grant 2023A04J1699, in part by the NSFC under Grants U22A2054 and 62101594, in part by the National Research Foundation, Singapore, in part by the Infocomm Media Development Authority through its Future Communications Research and Development Programme, in part by DSO National Laboratories through AI Singapore Programme under AISG Award AISG2-RP-2020-019, in part by Energy Research Test-Bed and Industry Partnership Funding Initiative, in part by Energy Grid (EG) 2.0 Programme, in part by DesCartes and the Campus for Research Excellence and Technological Enterprise (CREATE) Programme, MOE Tier 1 under Grant RG87/22, in part by the SUTD under Grant SRG-ISTD-2021165, in part by the SUTD-ZJU IDEA under Grant SUTD-ZJU (VP) 202102, in part by the Ministry of Education, Singapore, through its SUTD Kickstarter Initiative under Grant SKI 20210204, in part by the NSF under Grants CNS2107216, CNS-2128368, CMMI-2222810, and ECCS-2302469, and in part by the U.S. Department of Transportation, Toyota, and Amazon. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Angel Lozano. *(Corresponding author: Zehui Xiong.)*

Minrui Xu and Dusit Niyato are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore 639798 (e-mail: minrui001@e.ntu.edu.sg; dniyato@ntu.edu.sg).

Hongliang Zhang is with the School of Electronics, Peking University, Beijing 100871, China (e-mail: hongliang.zhang@pku.edu.cn).

Junlong Chen and Jiawen Kang are with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China (e-mail: 3121001036@mail2.gdut.edu.cn; kavinkang@gdut.edu.cn).

Zehui Xiong is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore, Singapore 487372 (e-mail: zehui_xiong@sutd.edu.sg).

Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA (e-mail: smao@ieee.org).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: zhan2@uh.edu).

Digital Object Identifier 10.1109/JSTSP.2023.3293650

I. Introduction

The vehicular MR Metaverse emerges as a promising solution for addressing the burgeoning demands of autonomous driving by amalgamating the physical and virtual transportation systems [1], [2]. The multi-dimensional communications among physical and virtual entities can combat the distance of "data islands" on roads which hinder efficient information exchange crucial for enhancing road safety, traffic control, and sustainability [3]. Assisted by digital twin (DT) technologies, autonomous vehicles (AVs) utilize advanced sensors, e.g., ultrasonic radars, cameras, and LiDAR, to collect data from their surrounding environments for constructing representations in the virtual space [4], [5]. Through artificial intelligence (AI) methods, these virtual models then inform driving decisions, enhancing AVs' response to dynamic road conditions. Despite these advancements, AVs still grapple with limited environment perception, as even high-class LiDAR and panoramic cameras can't fully compensate for occlusions and other limitations [6]. Consequently, there is a growing need for a system where AVs, roadside units (RSUs), and virtual simulators can collaboratively share and fuse sensing data, thereby achieving a holistic environmental perception. However, achieving such large-scale data collection and processing in the vehicular MR Metaverse for real-time driving simulation and training of AVs is latency-sensitive and resource-intensive.

To address this issue, much effort from academia and industry has been devoted to the development of virtual traffic and driving simulation platforms [7], [8]. These platforms utilize DT and MR technologies to create virtual representations of AVs, enabling the efficient collection of traffic and training data [4], [9], [10]. In this way, simulation and testing on rare scenarios, such as virtual traffic accidents and car collisions under realistic conditions, are performed in the virtual space [11]. Although traditional simulation platforms can generate a wealth of various traffic and driving simulations, they rely heavily on manual labor for data labeling, which impedes the fully autonomous driving [7]. In response, with the multi-modal generative AI [12], [13], [14], the labeled traffic and driving data can be synthesized directly by virtual simulators of training and evaluation in autonomous driving [15]. In this way, the process of using simulation platforms for autonomous driving training and evaluation is revolutionized by shifting from collecting and labeling data to directly synthesizing labeled data [15], [16], [17]. Consequently, simulation systems empowered by generative AI can generate large and diverse labeled traffic and driving datasets based on local road conditions and user preferences for online prediction and offline training in autonomous driving systems.

In the vehicular MR Metaverse, a coordinated effort is needed among connected AVs, RSUs, and virtual simulators to build efficient traffic and driving simulation platforms in the virtual space. To update the virtual representations in the virtual space, AVs continuously generate and offload multiple computation-intensive DT tasks to RSUs in online traffic simulation [4]. Specifically, these DT tasks range from simulation and decision-making to monitoring, each of which requires different degrees of computing and communication resources due to their resource requirements and latency constraints. In driving simulations, virtual simulators are capable of synthesizing controllable traffic and driving data for satisfying specific simulation requirements, e.g., passenger preferences and weather conditions, of the simulated driving tasks. In addition, the synthesized datasets also serve as valuable training material for the virtual representations of AVs, further enhancing their driving robustness. Nevertheless, these synchronization activities, encompassing DT task execution, traffic and driving simulations, and AV training, impose significant demands on the communication and computing resources of RSUs [18], [19]. Therefore, it is imperative to develop effective incentive mechanisms that motivate RSUs to optimize their resource allocation, thus promoting more efficient support for generative AI-empowered simulation platforms.

As illustrated in Fig. 1, in this article, we propose a novel DT-assisted autonomous driving architecture for the vehicular MR Metaverse, where generative AI is leveraged to synthesize massive and conditioned traffic and driving data for online and offline simulations. In particular, to improve reliability in DT task execution, we propose a multi-task DT offloading model where AVs can offload heterogeneous DT tasks with differing resource and deadline requirements to RSUs for remote execution. To improve the robustness in driving decision-making, virtual simulators empowered by generative AI can utilize the information in DTs, such as current location, historical trajectory, and user preferences, for intelligent traffic simulations [21], [22].

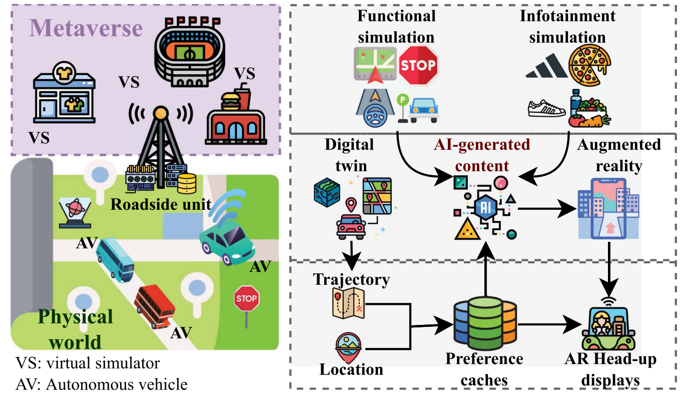


Fig. 1. The vehicular mixed reality Metaverse architecture of DT-assisted autonomous driving systems with traffic and driving simulations empowered by generative AI.

Moreover, by combining sensory data from the physical world with user preferences encapsulated in DTs, virtual simulators can synthesize diverse datasets for AV training, As a representative use case, we propose a diffusion model-based traffic sign generator, named TSDreamBooth, which is developed based on DreamBooth [23] fine-tuned using the Belgium traffic sign (BelgiumTS) dataset [24]. TSDreamBooth can generate virtual traffic sign images under varied local road conditions and user preferences. Finally, we propose a multi-task enhanced auction-based mechanism to satisfy multi-dimensional requirements (e.g., resource consumption and deadlines) of multiple DT tasks. Through rigorous analysis, the proposed mechanism not only ensures a strategy-proof operation but also mitigates the issue of adverse selection. The experimental results demonstrate that our proposed framework can enhance total social surplus by up to 150%, thus validating its effectiveness and efficiency. The main contributions of this work can be summarized as follows:

- To improve the safety and reliability of autonomous driving, we propose a novel DT-assisted MR Metaverse architecture with AI-generated simulations. In this architecture, connected AVs, RSUs, and virtual simulators build MR traffic and driving simulation platforms in the virtual space. The collaborative data collection, sharing, and utilization across physical and virtual transportation systems promise enhancements in driving safety, traffic control efficiency, and sustainability.
- In this architecture, we propose a reliable DT task offloading framework where AVs can continuously offload multiple DT tasks with distinct requirements from AVs to RSUs, ensuring timely updates of Digital Twins in the virtual environment.
- In traffic and driving simulations, we consider generative AI-empowered virtual simulators to synthesize new traffic and driving datasets for robustness improvement in decision-making and training for AVs.
- To incentivize RSUs for providing resources in supporting autonomous driving systems, we propose a multi-task enhanced auction-based mechanism to offer fine-grained

resource allocation and pricing for executing heterogeneous DT tasks with different deadlines. Upon rigorous property analysis, this mechanism is proven to be fully strategy-proof and immune to adverse selection.

The rest of this article is organized as follows. In Section II, we review the related works. In Section III, we discuss the proposed system architecture and present its system model. Then, in Section IV, we develop the auction-based mechanism. We present the experimental results in Section V, and provide conclusions in Section VI. The main notations used in this article are listed in Table I.

TABLE I
THE MAIN NOTATIONS AND DEFINITION IN THE ARTICLE

Notation	Definition
$I, \mathcal{I}$	The number and the set of AVs
$J, \mathcal{J}$	The number and the set of RSUs
$K, \mathcal{K}$	The number and the set of virtual simulators
B_j^u, B_j^d	The uplink and downlink bandwidth of RSU j
f_j^C, f_j^G	The CPU and GPU frequencies of RSU j
N_i	The number of DT tasks generated by AV i
$s_{i,n}^{DT}$	The size of DT data
$e_{i,n}^{DT}$	The number of CPU cycles required by per unit data
$d_{i,n}$	The deadline for completing task n of AV i
C_i	The size of preference of AV i
v_i	The private value of AV i
$m_{i,k}$	The match quality between i and k
$U_{i,k}$	The simulation value of virtual simulator k to AV i
$g_{i,j}$	The channel gain between AV i and RSU j
P_j, P_i	The transmit powers of RSU j and AV i
$R_{i,j}^d, R_{i,j}^u$	The downlink and the uplink transmission rate between AV i and RSU j
$t_{i,n,j}^{DT}$	The transmission latency for AV i to upload DT task n to RSU j
$l_{i,n,j}^{DT}$	The computation latency of AV i to process DT task n at RSU j
$G_{i,j,k}$	The generative score of RSU j for AV i and virtual simulator k
s_k^{SIM}	The data size of simulation of virtual simulator k
e_k^{SIM}	The required GPU cycles per unit data for offline simulation virtual simulator k
$h_{i,k}$	The number of hit preference caches of virtual simulator k for AV i
$Q_{i,n,j,k}$	The total amount of simulations generated by RSU j for virtual simulator k in task n of AV i
θ	The relative accuracy
$t_{i,j,k}^{SIM}$	The transmission latency of simulations of virtual simulator k for AV i at RSU j
$l_{i,j,k}^{SIM}$	The computation latency of simulations of virtual simulator k for AV i at RSU j
$g_{i,j}^{DT}$	The allocation variable of AV i for executing DT tasks at RSU j
$g_{i,j,k}^{SIM}$	The allocation variable of virtual simulator k for simulations of AV i at RSU j
$T_{i,n,j,k}^{total}$	The total latency required by RSU j to process DT task n of AV i and simulations of virtual simulator k
p_i^{DT}	The payment of AV i for executing DT tasks
p_k^{SIM}	The payment of virtual simulator k for processing simulations
S^{DT}	The expected surplus in online submarket for executing DT tasks
S_D^{SIM}	The expected surplus in offline submarket for executing driving simulations
S_T^{SIM}	The expected surplus in offline submarket for executing traffic simulations
b_i^{DT}	The bidding pricing of AV i
b_j^{SIM}	The bidding pricing of RSU j
Φ	The multi-task DT scoring rule
α	Pricing scaling factor

II. Related Works

A. The Vehicular Mixed Reality Metaverse

In the vehicular MR Metaverse, DT technologies empowered with driving data and AI algorithms serve as a pivotal component for significantly enhancing the accuracy and reliability of AVs in physical transportation systems [25], [26]. For instance, Niaz et al. in [11] develop an autonomous driving test framework via DT technologies. They emphasize the utilization of V2X communications to connect virtual space and physical space for driving safety and traffic control efficiency improvement. In addition, DT technologies can also improve resource utilization by efficiently managing resources in vehicular networks. Specifically, Li et al. in [27] propose a DT-driven computation offloading framework, where resource-intensive computation tasks are offloaded from AVs to RSUs for remote execution, for optimizing computation latency and service discontinuity.

Considering social influence in vehicular networks, Zhang et al. in [28] propose a DT-empowered content caching framework for improving caching scheduling in highly dynamic environments. In detail, they model vehicular networks as DTs and propose a learning-based caching algorithm to improve the system utility under dynamic content popularity, traffic density, and vehicle speed collaboratively. However, existing works on digital twin-based autonomous driving systems often restrict their focus to the single and homomorphic digital twin tasks. This limited perspective overlooks the inherent heterogeneity in autonomous driving tasks, such as complexity and urgency, necessary to adapt to the dynamic vehicle status and driving environments. Finally, while these frameworks employ AI algorithms to leverage the vehicle's digital twin data for online driving decision-making and offline training, the performance of these algorithms remains dependent on the available datasets.

B. Generative AI-Empowered Autonomous Driving Simulation

Through autonomous driving simulations, traffic and driving datasets can be synthesized for enhancing the inference and generalization capabilities of AI models [29]. The synthesized simulations and sensor data need to include not only diverse events but reflect realistic observations akin to real-world scenarios. Therefore, generative AI, such as generative adversarial networks (GANs) and diffusion models [30], emerges as the promising solution for synthesizing virtual traffic and driving datasets. For instance, Kim et al. propose the controllable simulation platform, named DriveGAN [15], which can generate high-resolution and diverse simulations based on user-defined conditions, e.g., weather conditions and locations of simulation objects.

Nevertheless, there is a gap between virtual and physical datasets data for AVs that are trained in simulation platforms [31]. The gap between virtual and physical datasets in autonomous driving refers to the discrepancies and unpredictability in real-world conditions that are difficult to accurately replicate in simulated environments, potentially leading to suboptimal or unsafe behaviors when AVs trained in virtual scenarios are deployed on real roads. Therefore, Yang et al. in [17] propose Surfelgan, which is a realistic sensor data synthesizing framework. The proposed data-driven camera generation scheme demonstrates that the generated data can not only be visualized as high-quality output but also valuable training datasets. Furthermore, Zhong et al. [21] propose a generative diffusion model to create controllable and realistic traffic simulations for autonomous driving systems. However, these simulation platforms fall short in synthesizing controllable and realistic simulations and sensory data. Therefore, they are restricted to synthesizing and then labeling traffic and driving simulations manually for utilization in autonomous driving, rather than directly synthesizing labeled datasets in response to the prompts.

C. Incentive Mechanisms in Connected Vehicular Networks

With the goal of improving resource utilization in connected vehicular networks, a variety of incentive mechanisms are being developed to motivate owners of vehicular infrastructure, i.e., RSUs, to offer resources to vehicles [32], [33]. For example, Sun et al. in [34] propose a preference-based incentive mechanism based on the Stackelberg game for resource allocation and scheduling in dynamic DT-assisted vehicular networks. Recognizing the value of storage resources in vehicular networks due to limited caching capacity and high deployment costs, Xing et al. in [35] propose a coalition game-based mechanism to motivate storage resource sharing in vehicular networks. Furthermore, Hui et al. in [4] consider sharing of computing and communication resources based on the auction game within DT-enabled connected autonomous driving systems. However, the previous works predominantly focus on optimizing resource allocation in the physical world or building simulation platforms in the virtual world, thereby overlooking the synergistic effects between the two. In this work, we propose a novel DT-assisted autonomous driving architecture empowered by generative AI which can synthesize diverse and conditioned datasets for traffic and driving simulations. Finally, we propose a multi-task enhanced second-score auction-based mechanism to offer fine-grained resource allocation for RSUs with incentives.

III. System Model

In this section, we first provide an overall description of the proposed architecture of the vehicular MR Metaverse in Section III-A. Then, we introduce the system model, including the network model in Section III-B, the DT task model in Section III-C, and the generative AI-empowered traffic and driving simulation model in Section III-D. Finally, we formulate the problems in Section III-E.

Fig. 2. The screenshots of our implemented driving simulation testbed [20] with synthetic traffic signs generated by the proposed generative diffusion model, named TSDreamBooth.

A. The Architecture of DT-Assisted Autonomous Driving

To enable autonomous driving in the vehicular MR Metaverse, connected AVs, RSUs, and virtual simulators collaborate to build digital simulation platforms that enable data sharing and AI-driven decision-making. In this architecture, RSUs with substantial communication and computing resources can provide online and offline simulation services for AVs and virtual simulators. Connecting RSUs with 5 G wireless communication, AVs can maintain their digital representations in the virtual space. Specifically, AVs continuously generate updates to digital representations during their trips, which are offloaded to RSUs for remote executions within required deadlines [36], [37]. During the execution of DTs, online simulations can be conducted to improve the performance of decision-making modules of autonomous driving. As illustrated in Fig. 3, virtual simulators can use the available resources and time from RSUs to fine-tune AI modules of AVs via offline simulations. During the offline simulations, virtual driving environments allow AVs to collect training data. Empowered by generative AI models, diverse and conditioned traffic and driving simulations can be synthesized and compiled into datasets. Therefore, virtual simulators can leverage a larger quantity of diverse, high-quality datasets to train AVs. Finally, the simulation results are relayed back to AVs for future use. In this architecture, online decision-making and offline training via traffic and driving simulations can improve driving safety, traffic control efficiency, and sustainability in autonomous driving systems.

In the system model, we consider three main roles in the vehicular MR Metaverse, i.e., AVs, RSUs, and virtual simulators. The set of I AVs is represented by $\mathcal{I} = \{1, \ldots, i, \ldots, I\}$, the set of J RSUs is denoted as $\mathcal{J} = \{1, \ldots, j, \ldots, J\}$, and the set of K virtual simulators is represented by $\mathcal{K} = \{0, 1, \ldots, k, \ldots, K\}$. Specifically, DTs of AVs provide real-time data about autonomous vehicles and their environment, which is used by virtual simulators to create testing scenarios for analyzing AV algorithms and decision-making under a variety of conditions. We consider the RSUs to own adequate communication and

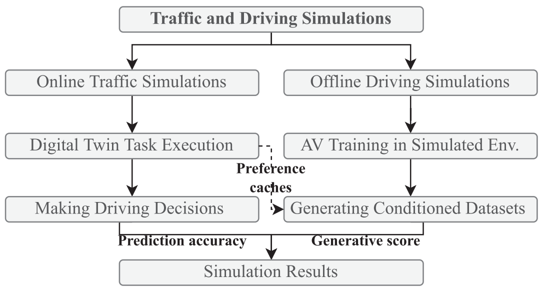


Fig. 3. The workflow of DT-assisted autonomous driving simulation platforms empowered by generative AI.

computing resources for enabling autonomous driving systems, i.e., the resources of RSUs are enough for executing all the computation tasks of AVs within deadlines. To facilitate autonomous driving systems, both uplink and downlink communication channels are allocated to upload DT tasks and stream simulation results. Therefore, the communication resources at RSU j consist of uplink bandwidth B_j^u and downlink bandwidth B_j^d. Moreover, to provide services such as executing DT tasks and simulating virtual traffic and driving, each RSU j is equipped with computing resources, including the CPU frequency f_j^C and the GPU frequency f_j^G. The vehicular MR Metaverse is constructed and maintained across a network of distributed RSUs through a coordinated resource and information sharing mechanism, where the DTs of AVs are continually updated with real-world data and RSUs synthesize traffic and driving data using generative AI models to create realistic virtual scenarios. These scenarios are used for decision-making and training to enhance AV performance and ensure a unified Metaverse representation.

In autonomous driving, AVs maintain the DTs in the virtual space and continuously update the DTs by executing DT tasks, e.g., simulation, decision-making, and monitoring. These DT tasks necessitate a heterogeneous range of resources and are subject to varying deadlines. Therefore, in the system model, N_i DT tasks are generated by AV i, which are represented as $DT_i = (< s_{i,1}^{DT}, e_{i,1}^{DT}, d_{i,1} >, \ldots, < s_{i,n}^{DT}, e_{i,n}^{DT}, d_{i,n} >, \ \ldots, < s_{i,N+i}^{DT}, e_{i,N_i}^{DT}, d_{i,N_i} >)$, where $s_{i,n}^{DT}$ is the size of DT data, $e_{i,n}^{DT}$ represents the number of CPU cycles required per unit data, and $d_{i,n}$ denotes the deadline for completing the task. As part of the DT data from AVs, there are preference caches that store passenger preferences, interests, and behaviors. This information is used to personalize users' experience with the DT and provide them with relevant and targeted content, services, and advertisements [38]. The size of the preference caches of AV i within the DT_i is C_i. Each AV $i \in \mathcal{I}$ has its private value v_i for executing its DT task DT_i, drawn from the probability distributions. The values of DT tasks can be interpreted as the characteristics of the AVs, such as the level of urgency to align with DT models [4], which may vary for each AV during its trips.

We consider two types of virtual simulators in the vehicular MR Metaverse, i.e., driving virtual simulators and traffic virtual simulators. Traffic virtual simulators, denoted by $1, \ldots, K$, facilitate online traffic simulation purposed to assist real-time decision-making process of AVs, including driving mode selection, information fusion, and motion planning [6]. The online decision-making module via traffic simulation uses data collected from AVs, RSUs, and virtual simulators to create a simulated driving environment. This environment is designed for testing and validating driving decisions, thereby enhancing the safety and reliability of autonomous driving systems. As a result, the feedback of these driving decisions is immediately returned and perceived by the AVs. The driving virtual simulator 0 delivers offline driving simulation to provide training simulation platforms to AVs. In the offline training module, the AI algorithms of AVs are trained by simulated driving practice. However, the performance improvement of AVs in the future still needs to be tested and validated, and thus the AVs cannot perceive immediate returns. The value of simulations for each simulation pair of AV i and virtual simulator k is $U_{i,k}$, which is the product of the common value v_i of AV i and the match quality $m_{i,k}$, i.e., $U_{i,k} = v_i m_{i,k}$. The common values for every virtual simulator k are gained from the provisioning of traffic simulation for AV i, which can be represented by AV i's private value v_i [39]. Additionally, the amount of personalized information determines the match quality $m_{i,k}$ of virtual simulator k. This way, the values of AVs and virtual simulators in autonomous driving systems are positively correlated. Then, in the resource allocation part, private values inform the prioritization of AVs, while the values of simulations measure the performance of virtual simulators with an AV's preferences and requirements. Finally, let $U_{\iota,(l)}$ and $m_{\iota,(l)}$ represent the l highest value and match quality for the AV ι, respectively.

B. Communication Model

In autonomous driving systems, vehicular-to-infrastructure communications with orthogonal frequency-division multiple access are utilized for updating DTs and streaming simulation results [40], [41], respectively. The channel gain between AV i and RSU j is represented by $g_{i,j}$, and the downlink transmission rate can be calculated as $R_{i,j}^d = B_j^d \log(1 + \frac{g_{i,j} P_j}{\sigma_i^2})$, where P_j is the transmit power of RSU j and σ^2 is the additive white Gaussian noise. Additionally, the transmit power of AV i is represented by p_i, and the uplink transmission rate can be calculated as $R_{i,j}^u = B_j^u \log(1 + \frac{g_{i,j} p_i}{\sigma_j^2})$, where P_i is the transmit power of AV i.

C. Multi-Task Digital Twin Model

In the multi-task DT model, AVs update their DTs in virtual space by executing multiple DT tasks for different functions with heterogeneous levels of complexity and urgency.

1) DT Task Execution: To maintain the digital representation with the vehicular MR Metaverse, physical entities, i.e., AVs, generate and offload DT tasks to RSUs for remote execution. Therefore, we consider the demands as tasks that are required to be accomplished by RSUs. The transmission latency $t_{i,n,j}^{DT}$ for AV i to upload its DT task $< s_{i,n}^{DT}, e_{i,n}^{DT}, d_{i,n} >$ to RSU j can

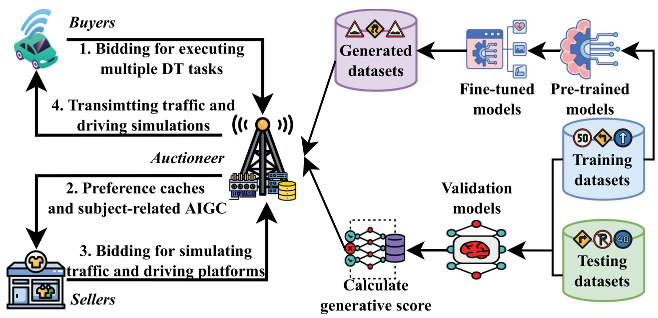


Fig. 4. The workflow of the proposed auction-based resource allocation for execution of DT tasks and generative AI-empowered simulations.

be calculated as [4] $t_{i,n,j}^{DT} = \frac{s_{i,n}^{DT}}{R_{i,j}^{u}}$, where $R_{i,j}^{u}$ is the downlink transmission rate between AV i and RSU j. After completing the upload of the DT task, RSU j uses its computing resources f_j^C to execute the received DT tasks. The computation latency in processing the DT task DT_i of AV i for RSU j can be calculated as $l_{i,n,j}^{DT} = \frac{s_{i,n}^{DT} e_{i,n}^{DT}}{f_j^C}$. In the proposed system, without loss of generality, we consider that each RSU can allocate enough virtual machines that the total execution latency for accomplishing DT tasks is within its required deadline [4], i.e., $t_{i,n,j}^{DT} + l_{i,n,j}^{DT} \leqslant d_{i,n}, \forall i \in \mathcal{I}, j \in \mathcal{J}, n = 1, \ldots, N$. In addition, virtual simulators can provide traffic and driving simulation services to AVs with the remaining available communication and computing resources of RSUs. After traffic and driving simulation, virtual simulators send the simulation results to AVs for further use.

D. Traffic and Driving Simulation Model

In this subsection, we first describe how we train and fine-tune the generative models to synthesize labeled data for traffic and driving simulations in vehicular networks. Then, we propose TSDreamBooth, a generative diffusion model that is fine-tuned using traffic sign datasets to synthesize new driving and traffic data based on user preferences. Finally, we present the simulation model of virtual simulators in terms of computation and communication latency of RSUs to generate datasets, train AVs, and return simulation results to AVs.

1) Generative AI-Empowered Simulation: As depicted in Fig. 4, the generative AI-based traffic and driving simulations encompass stages of training, fine-tuning, and inference stages. Initially, the low-resolution text-image model is fine-tuned using input images paired with a text prompt containing a unique identifier and the class name of the subject [23]. The model incorporates a class-specific prior preservation loss, designed to leverage the semantic prior that it has over the class and facilitate the creation of diverse instances associated with the subject's class. Subsequently, the super-resolution components of the model are fine-tuned using low and high-resolution image pairs from the input images. This method allows the model to retain high accuracy on the minute details of the subject while generating diverse instances in various scenarios. Consequently, this process of fine-tuning a text-image diffusion model using data from a virtual simulator enables the creation of more accurate and diverse driving simulations, thereby enhancing the development and testing of AVs. The virtual simulators adopt the Prior-Preservation Loss proposed in [23] to fine-tune the pre-trained models for customization of the local traffic signs.

2) Tsdreambooth: During the fine-tuning process of generative AI, virtual simulators use their local datasets for fine-tuning the generative models. Leveraging the specifics of the driving simulation, such as the class of traffic signs, the fine-tuned generative AI model for vehicular networks can effectively extract the features of these traffic signs. In this context, we propose the TSDreamBooth, fine-tuned on the traffic sign datasets. In the driving simulation, virtual simulators can use TSDreamBooth to generate a large amount of synthetic driving data based on local traffic signs using user preferences in AVs as input. In detail, RSUs extract preferences from DTs of AVs, known as preference caches. The preferences of AVs are collected by leveraging some user analytical equipment [16], such as eye-tracking devices. These preferences are then input into generative AI models as text prompts to produce diverse and conditioned simulation results. This enables virtual simulators to generate unlimited AV training experiments based on AV requirements similar to that are collected from realistic environments. As a result, the amount of simulations for offline training is no longer limited to the hit preference caches [20]. However, due to the limitations of generative AI models, some synthesized simulations may not satisfy the requirements. Such synthesized simulations can be identified using trained validation models, allowing for continuous improvement and refinement of the system, as the workflow shown in Fig. 4.

The models generated are based on probability distributions, and thus the results produced by TSDreamBooth are not deterministic. The results of TSDreamBooth may not be the same every time that the model is executed. Virtual simulators have the ability to capture variability in results and provide a better understanding of the uncertainties in the model's predictions by interrogating multiple results. Therefore, the validation models indicate the quality of generative AI models with a generative score $G_{i,j,k} \in [0, 1]$, as illustrated in Fig. 4. For each simulation result of virtual simulator k, the simulation task can be represented by $SIM_k =< s_k^{SIM}, e_k^{SIM} >$ [42], where s_k^{SIM} is the data size of each simulation and e_k^{SIM} is the required GPU cycles per unit data for offline simulation. Therefore, given the total number of virtual simulators $K + 1$, the match quality $m_{i,k}$ and hit preference caches $h_{i,k}$ are drawn independently from a set of distributions $m_{i,k} = h_{i,k} \sim F_{i,k}$. To explain further, given AV ι, the traffic virtual simulators $k = 1, \ldots, K$ can measure the match qualities $m_{\iota,k}$ of their traffic simulation. However, the driving virtual simulator 0 that provides driving simulation to AV ι cannot immediately measure its match quality $m_{\iota,0}$. Therefore, asymmetric information exists among virtual simulators that might result in adverse selection [39].

With the empowerment of generative AI models, the match quality $m_{i,k}$ is no longer constrained by hit preference caches $h_{i,k}$. As generative AI can generate countless and diverse simulation results based on user preferences and local information, virtual simulators can utilize computing resources and downlink transmission resources during the offline training process.

This enhancement not only increases the richness and variety of generated datasets but also enhances the overall training efficiency and output quality. During the remaining time of DT execution, the total amount of simulations $Q_{i,n,j,k}$ can be calculated as $Q_{i,n,j,k} = (d_{i,n} - t^{DT}_{i,n,j} - l^{DT}_{i,n,j})R^{SIM}_{i,j}/s^{SIM}_k$ for task n in DT_i of AV i and its RSU j. Then, the marginal generative AI-empowered match quality of AV i in simulator k via RSU j can be measured as

$$m_{i,n,j,k} = \frac{\log_2(1 + G_{i,j,k}Q_{i,n,j,k})h_{i,k}}{\theta(h_{i,k})}, \tag{1}$$

where $\theta(h_{i,k})$ is the relative accuracy among the original model w_i and the fine-tuned model $w_{i,k}$ for strongly convex objectives [43], [44]. Particularly, a value of $\theta(\cdot) = 1$ signifies that no improvement has been made in the training within simulation platforms. Conversely, $\theta(\cdot) = 0$ implies that the AI model has been trained to its optimal performance. These distinct conditions underscore the significant role of $\theta(\cdot)$ in evaluating the effectiveness of AI model training within simulation environments.

3) Simulation and Offline Training Model: The effective transmission latency in simulation and transmitting the driving simulation SIM_k to AV i for task n from RSU j can be calculated as

$$t^{SIM}_{i,j,k} = \frac{Q_{i,n,j,k}s^{SIM}_k}{R^d_{i,j}}, \tag{2}$$

where $R^d_{i,j}$ is the downlink transmission rate between AV i and RSU j. Moreover, the effective computation latency in completing the simulation SIM_k can be calculated as

$$l^{SIM}_{i,j,k} = \frac{Q_{i,n,j,k}s^{SIM}_k e^{SIM}_k}{f^G_j}, \tag{3}$$

which depends on the simulation latency in the GPUs of RSU j. (2) and (3) imply that the offline simulations in generative AI-empowered vehicular MR Metaverse can improve the resource utilization, performance, and sustainability of autonomous driving systems.

In autonomous driving systems, RSUs use their available computation and communication resources to provide real-time synchronization services for AVs and virtual simulators. However, it is imperative that the total latency does not exceed the requisite deadline specified by AV i. Let $g^{DT}_{i,j}$ be the allocation variable that AV i is associated with RSU j and $g^{SIM}_{i,j,k}$ be the allocation variable indicating that virtual simulator k is allocated by RSU j to match AV i. Consequently, the total latency $T^{total}_{i,j,k}$ required by RSU j to process both the DT task of AV i and the simulations of virtual simulator k must not exceed the predetermined deadline, which can be expressed as

$$\begin{aligned} T^{total}_{i,n,j,k} = {} & g^{DT}_{i,j} \cdot (t^{DT}_{i,n,j} + l^{DT}_{i,n,j}) \\ & + g^{SIM}_{i,j,k} \cdot (t^{SIM}_{i,n,j,k} + l^{SIM}_{i,n,j,k}) \leqslant d_{i,n}, \end{aligned} \tag{4}$$

$\forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}, n = 1, \ldots, N$. The driving simulation of virtual simulator k is running in the background of AV i during the processing of DT tasks at RSU j, and thus the expected duration of offline training can also be represented by $T^{total}_{i,n,j,k}$.

E. Problem Formulation

In the proposed system, a market for allocating resources of RSUs with incentives, consisting of online and offline submarkets, is instituted to incentivize RSUs to provide communication and computing resources for traffic and driving simulation for AVs and virtual simulators. The resource market in our architecture is divided into an online submarket and an offline submarket, both consisting of resources provided by Road Side Units (RSUs). The online submarket handles real-time tasks related to executing and updating DTs and assisting AVs with real-time decision-making, while the offline submarket handles tasks related to the offline simulation and training of AVs. This division allows for efficient and specialized use of RSU resources and caters to the different demands of autonomous driving. Participants in this market are considered to be risk-neutral, and their surpluses exhibit a positive correlation. Accordingly, the mechanism is expected to map the DT values $\mathbf{v} = (v_1, \ldots, v_I)$ and simulation values $\mathbf{U} = (U_{1,0}, \ldots, U_{I,K})$ to the payments of AVs $\mathbf{p}^{DT} = (p^{DT}_1, \ldots, p^{DT}_I)$ and the payments of virtual simulators $\mathbf{p}^{SIM} = (p^{SIM}_1, \ldots, p^{SIM}_K)$ with the allocation probabilities $\mathbf{g}^{DT} = (g^{DT}_1, \ldots, g^{DT}_I)$ and $\mathbf{g}^{SIM} = (g^{SIM}_0, \ldots, g^{SIM}_K)$.

After accomplishing the execution of DT tasks, the total expected surplus for RSUs from AV $i \in \mathcal{I}$ in the online submarket can be represented by $S^{DT}(\mathbf{g}^{DT}) = \mathbb{E}[\sum_{i=1}^{I} \mathcal{R}_{i,j}v_i\mathbf{g}^{DT}_{i,j}(\mathbf{v})]$. Based on the optimal reaction to the dominant strategies of the traffic virtual simulators, the driving virtual simulator can incentivize RSUs with the expected surplus of $S^{SIM}_D = \mathbb{E}[U_{i,0}g^{SIM}_{i,j,0}(Q_i)]$. In addition, the total expected surplus provided by traffic virtual simulators is defined by $S^{SIM}_T(\mathbf{g}^{SIM}) = \mathbb{E}[\sum_{k=1}^{K} U_{i,k}g^{SIM}_{i,j,k}(U_i)]$. In addition, we consider a cost-per-time payment model for simulation platforms, where users are billed per unit of time they utilize the simulation platform, e.g., per minute or per hour. Since AVs can only access the platforms provided by virtual simulators for a limited time T while driving, the cost-per-time payment model is a viable and flexible solution for renting virtual simulation platforms. In conclusion, the social surplus that RSU j can gain from the offline submarket can be given as $T \cdot (\gamma S^{SIM}_D(\mathbf{g}^{SIM}) + S^{SIM}_T(\mathbf{g}^{SIM}))$, where γ denotes the relative bargaining power of driving virtual simulator 0.

To maximize the social surplus in the market, the non-cooperative game among AVs, virtual simulators, and RSUs in the mechanism $\mathcal{M} = (\mathbf{g}^{DT}, \mathbf{g}^{SIM}, \mathbf{p}^{DT}, \mathbf{p}^{SIM})$ can be formulated as

$$\max_{\mathcal{M}} \quad S^{DT} + \sum_{n=1}^{N} T^{total}_{i,n,j,k} \cdot \left(\gamma S^{SIM}_D + S^{SIM}_T\right) \tag{5a}$$

$$\text{s.t.} \quad T^{total}_{i,n,j,k} \leqslant d_{i,n} \tag{5b}$$

$$h_{i,k} \leqslant C_i \tag{5c}$$

$$0 \leqslant b^{DT}_i \leqslant v^{DT}_i \tag{5d}$$

$$0 \leqslant p^{SIM}_k \leqslant U^{SIM}_{\iota,k} \tag{5e}$$

$$\sum_{i=1}^{I} g_{i,j}^{DT} \leqslant 1 \tag{5f}$$

$$\sum_{k=0}^{K} g_{i,j,k}^{SIM} \leqslant 1 \tag{5g}$$

$$g_{i,j}^{DT}, g_{i,j,k}^{SIM} \in \{0,1\} \tag{5h}$$

$$\forall i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}, n = 1, \ldots, N. \tag{5i}$$

Constraint (5b) ensures the reliability of each DT task that can be accomplished within the required deadline. Constraint (5c) guarantees that the number of hit preference caches is less than the size of preference caches. Pricing constraints (5d) and (5e) are listed to guarantee the individual rationality (IR) of traders. Allocation constraints (5f), (5g), (5h), and (5i) guarantee that each physical or virtual entity can be assigned by one and only one RSU.

In the auction-based resource allocation for online and offline submarkets, there are two issues, i.e., externalities and asymmetric information, causing adverse selection in the social surplus maximization problem formulated in (5a). First, adverse selection, as described in [18] refers to a market situation where participants with asymmetric information are only willing to pay the average market price. This often results in an inefficient distribution and pairing outcome in the market. In the context of traffic and driving simulations for autonomous driving, the physical and virtual entities (AVs and virtual simulators) have positively correlated surpluses for services. This means that the surplus of AVs in the online submarket can exert an influence on the surplus of virtual simulators in the offline submarket by affecting the common valuation of driving simulations. This correlation introduces externalities and asymmetric information for allocating physical and virtual entities in the service market.

- Externalities: The externalities are introduced to the online submarket as a result of influences from the offline submarket. The traffic and driving simulation results of virtual simulators have different match qualities for different physical AVs. However, during the allocation of AV in the physical market, the virtual simulator is unknown to the participants in the online submarket, which might affect the total processing latency in the online submarket. Therefore, AVs in the online submarket may prefer to prompt RSU to establish a predetermined execution latency threshold prior to the allocation of AVs.
- Asymmetric Information: There exists an asymmetry of information among virtual simulators regarding their traffic and driving simulations. The traffic simulation (e.g., movement predictions) can induce immediate responses from users. In contrast, driving simulations (e.g., training of traffic sign recognition models for AVs) may not be instantaneously discernible virtual simulators.

To ensure efficient allocation and pricing results, it is important to scrutinize the impact of this correlation and to find strategies to address asymmetric information and externalities.

IV. Multi-Task Enhanced Mechanism Design

To tackle the multi-task DT offloading problem in autonomous driving systems with traffic and driving simulations empowered by generative AI, we propose the multi-task enhanced second-score auction-based mechanism, named MTEPViSA, based on the EPViSA mechanism proposed in [20]. Incorporating aspects of the multi-dimensional auction [45] and the enhanced second-price auction [39], the MTEPViSA mechanism consists of four components, including the bidding process, the scoring rule, the allocation rules, and the pricing rules.

The MTEPViSA allocates and prices the winning AV in the online submarket by calculating the scoring rule. Therefore, we first define the AIGC-empowered scoring rule similar to [46], [47] as follows. Similar to the EPViSA mechanism, we apply several advanced techniques in auction theory [39] to enhance the auction-based mechanism by overcoming the externalities in the online submarket and the asymmetric information in the offline submarket described in Section IV-A. In the online DT tasks execution, AVs in the online submarket are allowed to submit their prices and required deadlines of DT tasks to the auctioneer. In addition, for the offline driving simulation, the auctioneer, e.g., the proxy of the RSUs, can determine the allocation rule according to the received bids from AVs with the AIGC scoring rule. Moreover, in the offline traffic and driving simulation, by adopting the price scaling factor $\alpha \geqslant 1$ in the offline submarket, the auctioneer can capture a significant fraction of the social surplus from both performance and brand virtual simulators. Finally, we analyze the properties of the MTEPViSA mechanism in Section IV-B.

A. Designing the MTEPViSA Mechanism

This subsection describes the workflow and property analysis of the multi-task enhanced second-score auction-based mechanism. To begin with, the definition of the multi-task DT scoring rule that is similar to [45] is provided as follows.

Definition 1 (Multi-task DT Scoring Rule): Let b_1^{DT} be any offered bidding price of AV i, the multi-task DT scoring rule $\Phi(b_i^{DT}, \mathbf{d}_i)$ under deadline requirement $\mathbf{d}_i$ for each task $n = 1, \ldots, N_i$ is defined as

$$\Phi(b_i^{DT}, \mathbf{d}_i) = b_i^{DT} + \sum_{n=1}^{N_1} \phi(d_{i,n}), \tag{6}$$

where $\mathbf{d}_i = (d_{i,1}, \ldots, d_{i,N_i})$ contains the submitted deadlines of AV i's DT tasks and $\phi(\cdot)$ is a non-decreasing function that $\phi(0) = 0$.

The scoring rule defined in (6) involves the deadlines of DT tasks and one element in the price vector. Therefore, for each AV i, N scores are calculated. Based on these scores, the marginal score sequence $\chi_i = \{\chi_{i,1}, \chi_{i,2}, \ldots, \chi_{i,N_i}\}$ can be calculated for each AV i. The marginal score indicates AV i's score increases when the total number of executed tasks increases. In addition, the n marginal score of AV i can be defined as

$$\chi_{i,n} = \begin{cases} \Phi(b_i^{DT}, \mathbf{d}_{i,1}), & n = 1, \\ \Phi(b_i^{DT}, \mathbf{d}_{i,n}) - \Phi(b_i^{DT}, \mathbf{d}_{i,n-1}), & 2 \leqslant n \leqslant N_i, \end{cases} \tag{7}$$

where $\mathbf{d}_{i,n} = (d_{i,1}, \ldots, d_{i,n})$. Then, we have the assumption on the property of marginal scores as follows [45].

Assumption 1 (Marginal Score): For any AV $i \in \mathcal{I}$, the marginal score sequence χ_i is non-negative and non-increasing in n, i.e., $\chi_{i,n} \geqslant \chi_{i,n+1}, n = 1, 2, \ldots, N_i - 1$.

The meaning of Assumption 1 is that performing additional simulations provides a higher score and the score is non-increasing with the performed simulations.

The auctioneer can calculate the scoring rule based on previous transaction results and current submitted bids and deadlines. In the online submarket, AVs submit their multi-dimensional bids $\mathbf{b}^{DT} = ((b_1^{DT}, \ldots, b_I^{DT}), \mathbf{d} = (\mathbf{d}_1, \ldots, \mathbf{d}_I))$ to the auctioneer. The auctioneer computes the scores $\Phi = \Phi(b^{DT}, \mathbf{d}) = (\Phi_1(b_1^{DT}, \mathbf{d}_1), \ldots, \Phi_I(b_I^{DT}, \mathbf{d}_I))$. Then, the auctioneer determines the winning AV in the online submarket for providing online simulation services according to the calculated scores. The auctioneer allocates the trader with the highest score as the winning physical entity as

$$g_i^{DT}(\Phi) = 1_{\{\Phi_i > \max\{\Phi_{-i}\}\}}, \tag{8}$$

while the payment that the winning AV needs to pay is the bidding price of the second-highest score, i.e.,

$$p_i^{DT}(\Phi) = g_i^{DT}(\Phi) \cdot b^{DT}_{\arg\max\{\Phi_{-i}\}}. \tag{9}$$

In the offline submarket, virtual simulators submit their bids $b^{SIM} = (b_0^{SIM}, b_1^{SIM}, \ldots, b_K^{SIM})$ to the auctioneer. In the MTEPViSA mechanism, the price scaling factor $\alpha \geqslant 1$ is utilized. First, the auctioneer determines the allocation probabilities for traffic virtual simulators as $g_k^{SIM}(b^{SIM}) = 1_{b_k^{SIM} > \alpha b_{-k}^{SIM}}$. Then, the allocation probability of the virtual simulator is calculated as $g_0^{SIM}(b^{SIM}) \leqslant 1 - \sum_{k=1}^{K} g_k^{SIM}(b^{SIM})$. Based on the price scaling factor α, the winning virtual simulator is required to pay

$$p_k^{SIM}(b^{SIM}) = g_k^{SIM}(b^{SIM}) \cdot \rho_k^{SIM}, \tag{10}$$

where

$$\rho_k^{SIM} = \begin{cases} T_{i,n,j,0}^{total} b_0^{SIM}, & k = 0, \\ T_{i,n,j,k}^{total} \alpha \max\{b_{-k}^{SIM}\}, & k = 1, \ldots, K. \end{cases} \tag{11}$$

By introducing the price scaling factor in the pricing rule in the offline submarket, the MTEPViSA mechanism can increase the expected social surplus of RSUs by providing offline simulation services compared with the traditional second-price auction. We then analyze the strategy-proofness of the MTEPViSA mechanism in Theorem 1.

These allocation and pricing rules are effective and efficient when the efficient scoring rule exists [45] and the price scaling factor is selected as $\alpha_\iota = \max(1, \gamma[Q_{\iota,0}]/\mathbb{E}[Q_{\iota,(2)}])$ [39], where ι is the winning AV in the online submarket. Finally, under the cost-per-time payment model of traffic and driving simulations and the efficient multi-task DT scoring rule, the MTEPViSA is fully strategy-proof and adverse-selection-free.

B. Property Analysis

To maximize its utility, each AV i can choose the deadline that can maximize the sum of its valuation v_i and externalities $\phi(d_{i,n})$ for the offline submarket. In Proposition 1, each AV can strategically select the optimal deadline to maximize its expected payoff.

Proposition 1 (Optimal Deadline): The optimal deadline bidding strategy for task n of AV i is given by

$$d_{i,n}^* = \arg \max_{d \in (0, d_{i,n}]} \left(v_i + \sum_{n=1}^{N_i} \phi(d) \right), \tag{12}$$

where $\phi(\cdot)$ is a non-decreasing function.

Proof: For any submitted multi-dimensional bid $(\bar{b}_i^{DT}, \bar{d}_i)$ of AV i, there invariably exists an alternative bid $(\hat{b}_i^{DT}, \hat{d}_i)$ capable of yielding an expected utility equal to or greater than that of AV i. This implies that the expected utility for the physical bidder i is at least as high as the utility obtained from the initial bid $(\bar{b}_i^{DT}, \bar{\mathbf{d}}_i)$. First, the deadline $\hat{\mathbf{d}}_i$ can be obtained from (12). Second, the deadline for the new bid $\hat{b}_i^{DT}$ can be determined by $\Phi(\hat{b}_i^{DT}, \hat{\mathbf{d}}_i) = \Phi(\bar{b}_i^{DT}, \bar{\mathbf{d}}_i)$, indicating that both bids result in the same score and allocation probability. In the event of a loss, the losing bidders would have utilities of zero with their scores. However, if bidder i is the winner, the new bid $(\hat{b}_i^{DT}, \hat{\mathbf{d}}_i)$ will yield the utility higher than or equal to the utility obtained from submitting other bids, i.e.,

$$\begin{aligned} & v_i - \hat{b}_i^{\mathrm{DT}} - \left(\max\{\Phi_{\mathcal{I}/\{i\}}\} + \sum_{n=1}^{N_i} \phi(\hat{d}_{i,n}) \right) \\ & \geqslant v_i - \bar{b}_i^{\mathrm{DT}} - \left(\max\{\Phi_{\mathcal{I}/\{i\}}\} + \sum_{n=1}^{N_i} \phi(\bar{d}_{i,n}) \right), \end{aligned} \tag{13}$$

where $\Phi_{\mathcal{I}/\{i\}} = (\Phi(b_1^{DT}, \mathbf{d}_1), \ldots, \Phi(b_{i-1}^{DT}, \mathbf{d}_{i-1}), \Phi(b_{i+1}^{DT}, \mathbf{d}_{i+1}), \ldots, \Phi(b_I^{DT}, \mathbf{d}_I))$, which holds true because the deadline is determined through the calculation of (12).

For the optimality of the selection of deadline, a similar proof of Proposition 1 can be found in [45]. Based on the bids submitted by AVs and the chosen optimal deadline, the auctioneer can maintain the efficient scoring rule, which can maximize social surplus to guide the allocation decisions in the online submarket as follows. Then, the efficient multi-task DT scoring rule can be defined as follows.

Definition 2 (Efficient Multi-task DT Scoring Rule): An efficient multi-task DT scoring rule can be expressed as

$$\Phi(b^{DT}, \mathbf{d}^*) = b^{DT} + T(\mathbf{d}^*)[\gamma S_D^{SIM}(\mathcal{M}) + S_T^{SIM}(\mathcal{M})], \tag{14}$$

where $T(\mathbf{d}^*)[\gamma S_D^{SIM}(\mathcal{M}) + S_T^{SIM}(\mathcal{M})]$ is the social surplus of virtual simulators by providing simulations and $T(\mathbf{d}^*)$ is the realized duration of AV training.

For a mechanism, strategy-proofness indicates that participants cannot increase their utility by altering their truthful bids. A mechanism is considered to be strategy-proof if and only if it can be characterized by a critical payment function ψ, such that a bidder n is deemed the winner if and only if their bid b_n surpasses the threshold price $\psi(b_{-n})$ when compared to the other competing bids b_{-n}. Once bidder n has won the auction, the payment charged by the auctioneer is the critical payment ψ. Adverse-selection free indicates that if the existence of market externalities and asymmetric information is independent of bidders' valuations, then under this mechanism, the factors of market externalities and asymmetric information are also independent of the allocation rules of the mechanisms. As a consequence, it should be highlighted that the MTEPViSA mechanism is fully strategy-proof and adverse-selection free, as demonstrated in the following theorem.

Theorem 1: The MTEPViSA mechanism is fully strategy-proof and adverse-selection-free in the market with the efficient multi-task DT scoring rule and the cost-per-time model of simulations.

Proof: To demonstrate that the MTEPViSA mechanism is fully strategy-proof, we must identify the critical payment functions for traders in both the online and offline submarkets to satisfy the conditions for a strategy-proof auction. To begin with, we show that there is a critical payment function $\psi^{on}(b_{-i}^{DT})$ for the MTEPViSA mechanism in the online submarket. If a bidder i in the online submarket submits a truthful bid, the score can be determined by the function $\Phi(b_i^{DT}, \mathbf{d}_i)$, given any deadlines $\mathbf{d}_i$. It is necessary to demonstrate that bidder i cannot increase the benefit by altering the bid. If the bidder were to submit a false bid $b_i' \neq b_i^{DT}$, and did not win the auction, the reward would be zero, regardless of the specified deadline $\mathbf{d}_i$ or the score calculation function $\Phi(b_i^{DT}, \mathbf{d}_i)$. However, if bidder i were to win the auction by submitting a false bid, the expected reward can be represented as follows:

$$\begin{aligned} S_i^{\mathrm{DT}} &= v_i - b_i' \\ &= v_i - \left(\max\{\Phi_{-i}\} + \sum_{n=1}^{N_i} \phi(d_{i,n})\right) \\ &= \Phi_i - \max\{\Phi_{-i}\}, \end{aligned} \tag{15}$$

where $\max \Phi_{-i}$ represents the highest score excluding the bid of bidder i. Hence, regardless of whether the bidder wins or loses under either Φ' or Φ_i, the utilities that bidder i receives will always be lower than or equal to the utility that it would receive if it submitted the truthful bid. The critical payment function in the online submarket can be represented as $\psi_{on}(b_i^{\mathrm{DT}}, \mathbf{d}_i) = \max\{\Phi_{-i}\} + \sum_{n=1}^{N_i} \phi(d_{i,n})$. Additionally, the auctioneer must compute the synchronization scores for bidders in the online submarket. As a result, all bidders in the online submarket are protected against false-name bidding.

The critical payment function for the MTEPViSA mechanism in the offline submarket is $\psi_{\mathrm{off}}(b_{-k}^{\mathrm{SIM}}) = \alpha \max\{b_{-k}^{\mathrm{SIM}}\}$, where $\alpha \geqslant 1$. With this critical payment function in place, the top-performing bidder can win by ensuring that $\psi_{\mathrm{off}}(b_{-k}^{\mathrm{SIM}}) \geqslant \max\{b_{-k}^{\mathrm{SIM}}\}$. Furthermore, the mechanism is proof against false-name bidding in the offline submarket if $\psi_{\mathrm{off}}(b_{-k}^{\mathrm{SIM}}) = \psi_{\mathrm{off}}(\max\{b_{-k}^{\mathrm{SIM}}\})$. Consider a set of bids b_k^{SIM} that result in $\psi_{\mathrm{off}}(b_{-k}^{\mathrm{SIM}}) \neq \psi_{\mathrm{off}}(\max\{b_{-k}^{\mathrm{SIM}}\})$, and let there be two bidders in the offline submarket. If one bidder has a higher valuation than $\psi_{\mathrm{off}}(b_{-k})$ and the other has a valuation of $\max\{b_{-k}^{\mathrm{SIM}}\}$, and $\psi_{\mathrm{off}}(b_{-k}^{\mathrm{SIM}}) < \psi_{\mathrm{off}}(\max\{b_{-k}^{\mathrm{SIM}}\})$, then the first bidder could submit a lower price while keeping the other bids in the set b_{-k}. This means that the mechanism is not winner false-name proof. Conversely, if $\psi_{\mathrm{off}}(b_{-k}^{\mathrm{SIM}}) > \psi_{\mathrm{off}}(\max\{b_{-k}^{\mathrm{SIM}}\})$, the losing bidder in the offline submarket could submit a higher bid compared to the winner's bid while maintaining the other bids in the set b_{-k}^{SIM}. As a result, the mechanism in the offline submarket is loser false-name proof.

To show that the mechanism is free from adverse selection, the critical payment function in the online submarket is quasi-linear and the critical payment function in the offline submarket is homogeneous of degree one [39]. In the online submarket, we consider two types of external effects from the offline submarket, $d \in \{0, \infty\}$, with a probability of $\Pr(\phi = 0) \in (0, 1)$ while keeping the other bidding prices v_{-i} constant. If $\phi = 0$, there are no external effects from the offline submarket, and we have $g_i^{\mathrm{DT}}(v + \phi(0)) = g_i^{\mathrm{DT}}(v) = 1_{\{v_i > \max v_{-i}\}} = 1_{\{v_i > \psi_{\mathrm{on}}(v_{-i}, 0)\}}$. If $\phi = \infty$, then $g_i^{\mathrm{DT}}(v + \phi(0)) = g^{\mathrm{DT}} i(v + \infty) = 1_{\{v_i > \psi_{\mathrm{on}}(v_{-i}, \phi(\infty))\}} = 0$, meaning no bidder can win in the online submarket. Thus, the proposed mechanism in the online submarket is free from adverse selection. In the offline submarket, suppose that $v \in 1, c$ with a probability of $\Pr(C = 1) \in (0, 1)$ while keeping the vehicular MR Metaverse simulation qualities constant. It can be shown that $g_0^{\mathrm{SIM}}(vm) = 1_{\{v=c\}}$. When $v = 1$, $g_k^{SIM}(vm) = g_k^{SIM}(vm) = 1_{\{m_i > \psi_{\mathrm{off}}(m_{-i})\}} = 1$, and therefore $g_0^{\mathrm{SIM}}(\mathrm{cm}) = 0$. When $v = c, g_k^{SIM}(\mathrm{cm}) = 1_{\{cm_i > \phi_{\mathrm{off}}(cm_{-i})\}} = 0$, meaning no top-performing bidder can win the auction, and then $g_0^{SIM}(\mathrm{cm}) = 1$. In conclusion, we have shown that the MTEPViSA mechanism is both strategy-proof for bidders in the online and offline submarkets and free from adverse selection by utilizing the efficient multi-task DT scoring rule and cost-per-time model.

From Theorem 1, it can be deduced that the proposed mechanism for AVs and virtual simulators is fully strategy-proof. This indicates that AVs and virtual simulators cannot manipulate their bids to achieve higher utility. Although we introduce the scoring rule and price scaling factors to eliminate the externalities and asymmetric information, these additional components may not provide additional information to AVs and virtual simulators when they develop their own strategy. In the MTEPViSA mechanism, the optimal strategy for AVs in the online submarket and virtual simulators in the offline submarket is to tell the truth. Moreover, due to the interaction of two submarkets leading to externalities and asymmetric information for traders, the proposed mechanism is free from adverse selection as all participants have sufficient motivation to join the market. Therefore, the social surplus achieved by the MTEPViSA mechanism is still efficient enough to avoid market failure.

Finally, we consider the implementation overhead of the proposed auction-based mechanism, where a centralized auctioneer collects bids, computes scoring rules, and determines allocation and pricing results. To begin with, I AVs and K virtual simulators submit their bids to the auctioneers. Let N be the number of DT tasks of each AV, the computation complexity to compute the scores is $O(IK \log(K)(N))$. Then, the computation complexity to sort the scores of AVs in the online market is $O(I\log(I))$. Finally, the computation complexity of determining and pricing the winning virtual simulator is $O(K)$. Overall, the computation complexity of the proposed MTEPViSA mechanism is $O(IKN\log(K) + I\log(I) + K)$.

V. Experimental Results

In this section, we implement the generative AI-empowered autonomous driving simulation system. First, we validate the DT-assisted movement prediction model in Section V-A1 and test the generative AI-empowered traffic and driving simulation model in Section V-B. Then, we evaluate the performance of the proposed mechanism in Section V-C.

A. Experimental Setups

In the simulation of the vehicular MR Metaverse, we consider an autonomous driving system with 30 AVs, 30 virtual simulators, and 1 RSU by default. For the RSU, 20 MHz uplink and 20 MHz downlink channels are allocated for DT task uploading and MR content streaming, respectively. In addition, the CPU frequency of RSU is set to 3.6 GHz, and the GPU frequency is set to 19 GHz. The channel gain between RSUs and AVs is randomly sampled from $U[0,1]$, where U denotes the uniform distribution. The transmit power of AVs is randomly sampled from $U[0,1]$ mW and the transmit power of RSUs is randomly sampled from $U[0,5]$ mW. The additive white Gaussian noise at AVs and RSUs is randomly sampled from $\mathcal{N}(0,1)$, where $\mathcal{N}$ denotes the normal distribution. For each DT task generated by AV, the data size is randomly sampled from $U[0,0.5]$ MB, the required CPU cycles per unit data are randomly sampled from $[0,2]$ Gcycles/MB, and the required deadline is randomly sampled from $U[1,1.5]$ seconds. For each simulation, the data size is randomly sampled from $U[0,2.5]$ MB and the required GPU cycles per unit data are randomly sampled from $U[0,5]$ Gcycles/MB. The valuation of AVs for accomplishing the DT tasks is randomly sampled from $U[0,1]$ and the number of preferences of AVs is sampled from $Zipf(2)$, where $Zipf$ denotes the Zipf distribution. The relative bargaining power of the offline virtual simulator is set to 1 while the default accuracy is 0.5. For digital twin-assisted vehicular movement prediction, we set the past P steps to 60 and the future F steps to 5. In addition, we set epoch e to 500, batch size to 40, and dropout to 0.05. The default local relative accuracy is set to 0.53 [43] and the default generative score is sampled from $U[0.4,0.6]$.

The simulation environment for the vehicular MR Metaverse was created using a 3D model of a few city blocks in New York City. This model was developed by Geopipe, Inc., which employed AI to build a digital replica from photographs captured throughout the city. The simulation depicts an autonomous car navigating through a road, as shown in Fig. 2, with artificially placed highway advertisements. Eye-tracking data was collected from human participants who were immersed in the virtual environment using the HMD Eyes addon from Pupil Labs. After the simulation, the participants were asked to complete a survey to assess their subjective opinion level of interest in each aspect of the simulation.

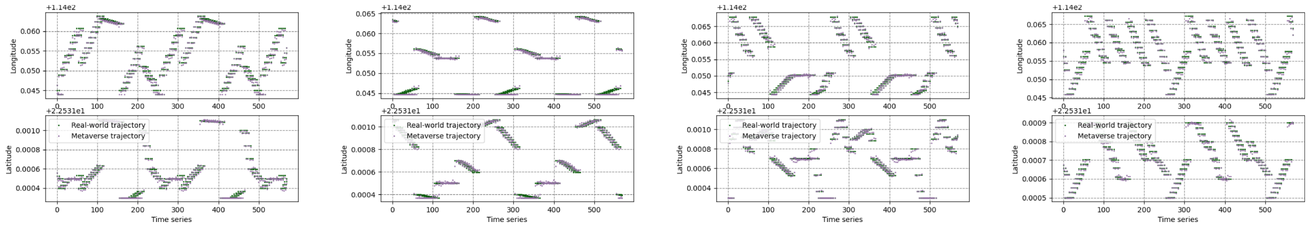
1) Digital Twin-Assisted Vehicular Movement Prediction: Through continuously updating DTs in the virtual space, AVs can leverage the results of online traffic simulations for improving driving safety and traffic control efficiency. Specifically, we use the historical trajectory data of AVs in DT to predict their future movements aiming to concretize the concept of DT-assisted autonomous driving. Let the location of AV i at time slot t be $p_i^t = (x_i^t, y_i^t)$, where x_i^t and y_i^t are longitude and latitude of AV i, respectively. The historical trajectory of AV i consists of the last P locations can be represented as $\tau_i^{past}(t) = (p_i^{t-P}, \ldots, p_i^{t-1}, p_i^t)$. When RSUs leverage AI models to predict the future movement of AVs denoted as $\mathcal{A}_j$ for RSU j. Then, the input of past trajectories into the AI model of RSU j predicts the movement $\tau_i^{pre}(t) = \mathcal{A}_j(\tau_i^{past}(t)) = (p_i^{t+1}, p_i^{t+2}, \ldots, p_i^{t+F})$ in the future F steps of the vehicles and simulates the movements in the virtual space. In the training module, the AI model is evaluated by the mean squared error (MSE), i.e., the training loss is calculated as $\mathbb{E}_{\tau_i^{past}(t), \tau_i^{true} \sim DT_i}(\tau_i^{pre}(t) - \tau_i^{true}(t))^2$. Finally, the model's efficacy is evaluated by the R2 score $\mathcal{R}_{i,j}$, which is 1 when the predicted movements are perfectly correlated with the true movements. The details of DT-assisted movement prediction are shown in Algorithm 1. In this way, instead of collecting the physical location of AVs, the virtual space can simulate real-time vehicular movement prediction. This virtual positioning method can be regarded as a supplement to physical positioning methods, e.g., positioning via a global navigation satellite system.

As shown in Fig. 5, we use four trajectory collections sampled from the dataset in [48] to demonstrate the effectiveness of movement prediction of AVs based on the current location and historical routes. For this task, we select the LSTM model for movement prediction. We use the R2 score to evaluate the performance of the movement prediction AI model. R2 is commonly understood as using the mean as the error benchmark to see if the prediction error is greater or less than the mean benchmark error. When R2 score $= 1$, the predicted and true values of the sample are exactly identical without any error. This indicates that the independent variable is an effective predictor of the dependent variable in the regression analysis. Based on this experiment, we compile the prediction results of the prediction in the system to calculate the simulation accuracy in the virtual space.

B. Virtual Traffic Sign Synthesizing in Generative AI-Empowered Simulation

Generative AI based on large text-to-image models, such as stable diffusion [49] and Dreambooth [23], is poised to revolutionize content creation in the MR Metaverse. Dreambooth is a personalized diffusion model that learns to preserve the features of the specific subject and subsequently generating new images based on this subject. To demonstrate the ability to generate diverse and high-quality images for the vehicular MR Metaverse, as illustrated in Fig. 6, this demonstration involves the manipulation of background color and the re-contextualization of traffic signs.

We first use the training set in the BelgiumTS dataset [24] to fine-tune Dreambooth to TSDreambooth. Then, we train a



(a) Trajectories 1, R2 score=0.9972. (b) Trajectories 2, R2 score=0.9939. (c) Trajectories 3, R2 score=0.9984. (d) Trajectories 4, R2 score=0.9983.

Fig. 5. Difference between the real-world trajectories and DT-assisted predicted trajectories.

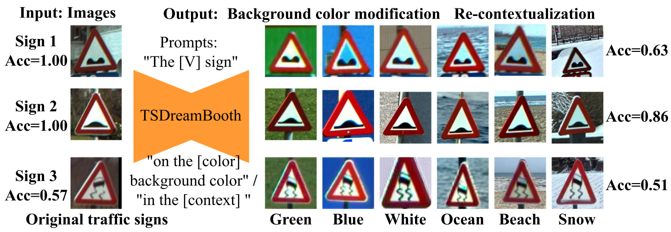


Fig. 6. Synthesized traffic signs generated by TSDreambooth for background color modification and re-contextualization.

Algorithm 1: DT-Assisted Movement Prediction.

Input: Digital Twin DT_i of AV i
Initialization: The movement prediction accuracy $\mathcal{L}_{i,j}$ and movement prediction model A_j
Obtain current location p_i^t, history trajectory $\tau_i^{past}(t)$, and future trajectory $\tau_i^{true}(t)$;
Training Process
for *each training epoch* e **do**
  Input the history trajectory $\tau_i^{past}(t)$ to trajectory generation model A_j^e and then obtain the predict trajectory $\tau_i^{pre}(t)$;
  Calculate the mean squared error loss $\mathcal{L}_{i,j}^{train} = \mathbb{E}_{\tau_i^{past}(t), \tau_i^{true} \sim DT_i}(\tau_i^{pre}(t) - \tau_i^{true}(t))^2$;
  Update the Trajectory generation model A_j^{e+1} based on $\mathcal{L}_{i,j}^{train}$ via back-propagation;
end
Evaluate its performance by the R2 score $\mathcal{R}_{i,j}$;
Output: The optimal movement prediction model A_j and its performance $\mathcal{R}_{i,j}$.

validation model based on the pre-trained GoogLeNet to fit the BelgiumTS dataset. The learning rate of the validation model is set to 0.1 and the number of epochs is set to 10. After optimizing via cross-entropy loss, the final recognition accuracy is 96% on the testing sets. Finally, we generate new images based on the testing set in BelgiumTS and evaluate the generative score using the validation model. We summarize the obtained generative score in Fig. 7 from the above experiments. As we can observe, the validation model performs almost perfectly in the real test dataset. However, for the generated dataset, the validation model can only recognize around 80% of the images generated by TSDreambooth. Then, the synthesized datasets are leveraged for fine-tuning the validation model. During the fine-tuning, the batch size is set to 30 for one iteration. Finally, we obtain the local relative accuracy for the entire generated dataset ($\theta = 0.82$), the background modification dataset ($\theta = 0.42$), and the recontextualization dataset ($\theta = 0.85$).

C. Performance Evaluation of Auction-Based Mechanisms

Then, we evaluate the performance of the proposed mechanism under different system settings compared with PViSA and EPViSA proposed in [20]. The PViSA mechanism consistently selects the AV with the highest valuation in the online submarket to synchronize thereby ignoring the potential surplus in the offline submarket.

1) Performance Evaluation Under Different System Parameters: Setting the number of tasks to 1, the achieved surplus in the system under different market parameters is illustrated in Fig. 8. With the enlargement of the market size, both the number of buyers in Fig. 8(a) and the number of sellers in Fig. 8(b), the surplus achieved by the proposed framework becomes increasingly higher. A larger market size will lead to more competition among traders in the market, resulting in a higher surplus for RSUs in providing services. Our findings indicate that generative AI-empowered simulations can increase surplus by at least 150% compared with the simulations without generative AI. In addition, the proposed MTEPViSA mechanism can improve more than half of the surplus compared with the PViSA mechanism. Furthermore, the generative score also has a substantial impact on the surplus in the simulation of the system, as shown in Fig. 8(c). Therefore, the mechanism can not only select AVs and virtual simulators with a high valuation but also fine-tune the AI models of AVs for higher accuracy.

2) Performance Evaluation Under Different System Settings: In Fig. 9, we evaluate the performance of the proposed mechanism under different system settings. As illustrated in Fig. 9(a), the total revenue of the virtual simulator exhibits an increase as the number of tasks increases. The proposed MTEPViSA mechanism can amplify the surplus twofold compared to PViSA. As the number of tasks becomes higher, the performance gap between METPViSA and EPViSA becomes larger. From Fig. 9(c), we can observe that the growth points of the surplus mainly rely on the surplus obtained from provisioning traffic simulation results. From Fig. 9(b), we find the reason for the inefficiency of PViSA. The PViSA mechanism always selects the AV with the highest valuation in the online submarket to synchronize while ignoring the potential surplus in the offline submarket. Finally, the MTEPViSA and PViSA mechanisms can obtain a higher surplus

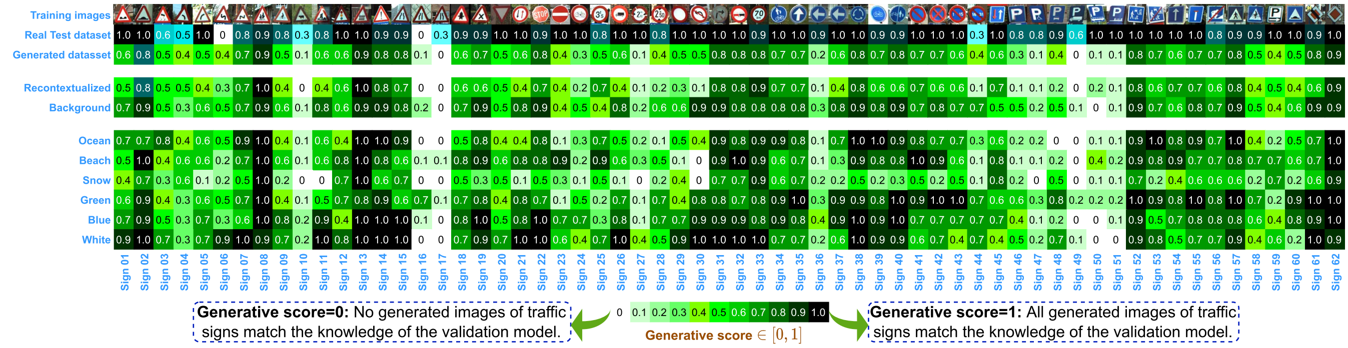
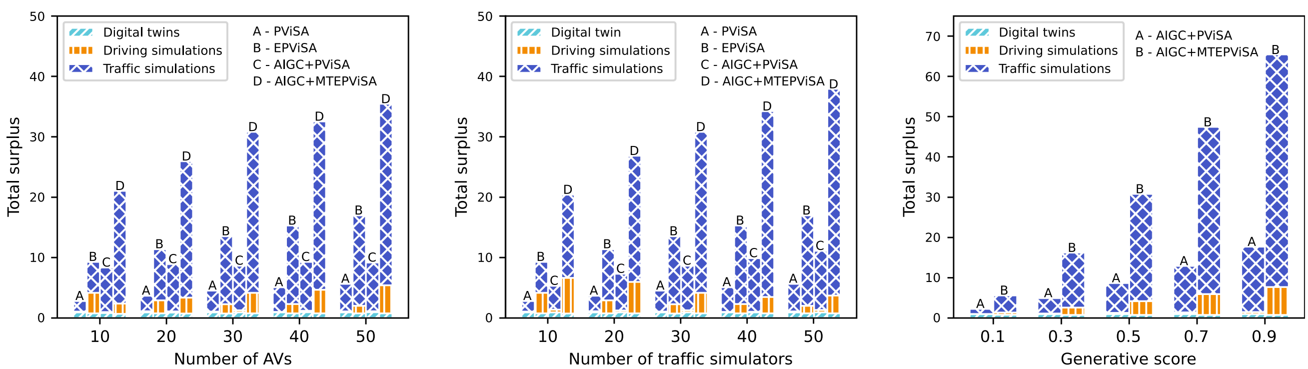
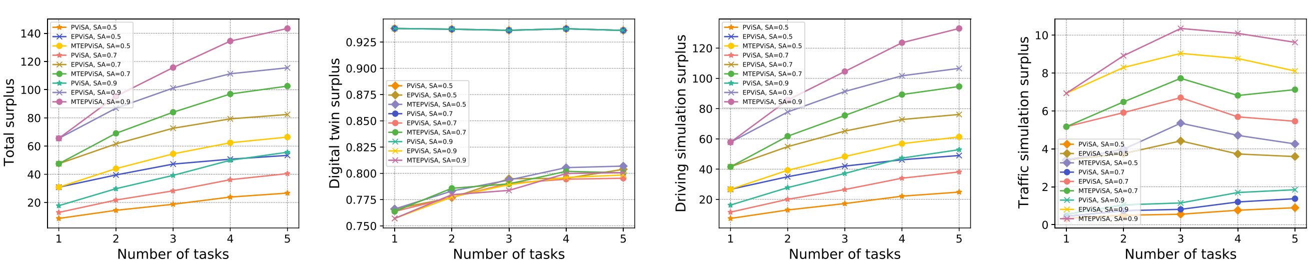


Fig. 7. The generative score of the TSBreamBooth fine-tuned on the BelgiumTS dataset.



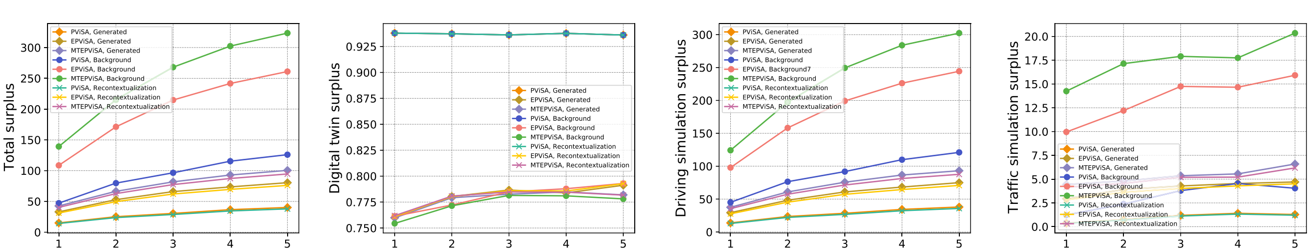
(a) Social surplus v.s. number of AVs. (b) Social surplus v.s. number of traffic simulators. (c) Social surplus v.s. generative score.

Fig. 8. Performance evaluation under different sizes of the market and generative scores.



(a) Total surplus v.s. number of tasks. (b) DT surplus v.s. number of tasks. (c) Traffic simulation surplus v.s. number of tasks. (d) Driving simulation surplus v.s. number of tasks.

Fig. 9. Performance evaluation of simulated experiments on different generative scores and the numbers of tasks.



(a) Total surplus v.s. number of tasks. (b) DT surplus v.s. number of tasks. (c) Traffic simulation surplus v.s. number of tasks. (d) Driving simulation surplus v.s. number of tasks.

Fig. 10. Performance evaluation of experiments under different datasets generated by TSDreamBooth and numbers of tasks.

in provisioning driving simulations by tackling the asymmetric information in the offline submarket. As illustrated in Fig. 10, using the synthesized datasets of TSDreambooth, we obtain the total, DT, traffic simulation, driving simulations surpluses for the generated, background modification, and re-contextualization datasets. We can see that the growth trend of the surplus for each mechanism in the figure is similar to that in the simulation. However, since the quality of the dataset in the real experiment is not as good as that in the simulation, the distribution of the results in the experiment is relatively uneven. This can be seen most clearly in Fig. 10(a) and (c). These results show that although the generated datasets improve the performance of AI models compared to the original datasets, the improvement is not homogeneous depending on the datasets generated by different preferences.

VI. Conclusion

In this article, we have proposed a generative AI-empowered autonomous driving architecture for the vehicular MR Metaverse. In this architecture, we have proposed the multi-task DT offloading model for reliably executing AVs' DT tasks with different requirements at RSUs. In addition, we have leveraged the generative AI models to synthesize diverse and labeled traffic and driving datasets for AVs' offline training. Finally, we have devised the multi-task enhanced auction-based mechanism to allocate resources of RSUs with incentives to support the autonomous driving systems. Through the property analysis, the proposed mechanism is validated as strategy-proof and adverse-selection free. The experimental results have shown that the proposed framework can bolster the social surplus by around 150%. In the future work, we will investigate how generative AI can integrate these human factors, such as fatigue and distraction, into the synthesis of traffic and driving datasets.

References

[1] P. Zhou et al., "Vetaverse: Technologies, applications, and visions toward the intersection of metaverse, vehicles, and transportation systems," Feb. 2022, *arXiv:2210.15109*.

[2] H. Zhang, S. Mao, D. Niyato, and Z. Han, "Location-dependent augmented reality services in wireless edge-enabled metaverse systems," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 171–183, Jan. 2023.

[3] M. Xu et al., "A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 656–700, Firstquarter 2023.

[4] Y. Hui et al., "Collaboration as a service: Digital twins enabled collaborative and distributed autonomous driving," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18607–18619, Oct. 2022.

[5] X. Yuan, J. Chen, N. Zhang, J. Ni, F. R. Yu, and V. C. M. Leung, "Digital twin-driven vehicular task offloading and IRS configuration in the internet of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24290–24304, Dec. 2022.

[6] Z. Xiao, J. Shu, H. Jiang, G. Min, H. Chen, and Z. Han, "Perception task offloading with collaborative computation for autonomous driving," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 457–473, Feb. 2023.

[7] S. Feng, X. Yan, H. Sun, Y. Feng, and H. X. Liu, "Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment," *Nature Commun.*, vol. 12, no. 1, Feb. 2021, Art. no. 748.

[8] Y. Feng, C. Yu, S. Xu, H. X. Liu, and H. Peng, "An augmented reality environment for connected and automated vehicle testing and evaluation," in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 1549–1554.

[9] L. Teng et al., "QoE driven vr 360° video massive MIMO transmission," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 18–33, Jan. 2022.

[10] H. Duan, W. Shen, X. Min, D. Tu, J. Li, and G. Zhai, "Saliency in augmented reality," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 6549–6558.

[11] A. Niaz, M. U. Shoukat, Y. Jia, S. Khan, F. Niaz, and M. U. Raza, "Autonomous driving test method based on digital twin: A survey," in *Proc. IEEE Int. Conf. Comput., Electron. Elect. Eng.*, 2021, pp. 1–7.

[12] R. Gozalo-Brizuela and E. C. Garrido-Merchan, "ChatGPT is not all you need a state of the art review of large generative ai models," 2023, *arXiv:2301.04655*.

[13] W. Xia et al., "Generative neural network channel modeling for millimeter-wave UAV communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9417–9431, Nov. 2022.

[14] A. Gilles, P. Gioia, R. Cozot, and L. Morin, "Computer generated hologram from multiview-plus-depth data considering specular reflections," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2016, pp. 1–6.

[15] S. W. Kim, J. Philion, A. Torralba, and S. Fidler, "DriveGAN: Towards a controllable high-quality neural simulation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 5820–5829.

[16] Z. Wang, C. Yang, and S. Mao, "Data augmentation for RFID-based 3D human pose tracking," in *Proc. IEEE 96th Veh. Technol. Conf.*, 2022, pp. 1–2.

[17] Z. Yang et al., "SurfelGAN: Synthesizing realistic sensor data for autonomous driving," in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit.*, 2020, pp. 11118–11127.

[18] Y. Zhang, L. Song, M. Pan, Z. Dawy, and Z. Han, "Non-cash auction for spectrum trading in cognitive radio networks: Contract theoretical model with joint adverse selection and moral hazard," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 3, pp. 643–653, Mar. 2017.

[19] C. Zhang, M. Zhao, L. Zhu, W. Zhang, T. Wu, and J. Ni, "FRUIT: A blockchain-based efficient and privacy-preserving quality-aware incentive scheme," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 12, pp. 3343–3357, Dec. 2022.

[20] M. Xu et al., "Epvisa: Efficient auction design for real-time physical-virtual synchronization in the metaverse," 2022, *arXiv:2211.06838*.

[21] Z. Zhong et al., "Guided conditional diffusion for controllable traffic simulation," 2022, *arXiv:2210.17366*.

[22] L. Yang, M. Xu, Y. Guo, X. Deng, F. Gao, and Z. Guan, "Hierarchical bayesian LSTM for head trajectory prediction on omnidirectional images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7563–7580, Nov. 2022.

[23] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," 2022, *arXiv:2208.12242*.

[24] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition–how far are we from the solution?," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2013, pp. 1–8.

[25] Y. Jiang et al., "Reliable distributed computing for metaverse: A hierarchical game-theoretic approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 1, pp. 1084–1100, Jan. 2023.

[26] A. Niaz, S. Khan, F. Niaz, M. U. Shoukat, I. Niaz, and J. Yanbing, "Smart city IoT application for road infrastructure safety and monitoring by using digital twin," in *Proc. IEEE Int. Conf. IT Ind. Technol.*, 2022, pp. 1–6.

[27] M. Li, J. Gao, C. Zhou, X. Shen, and W. Zhuang, "Digital twin-driven computing resource management for vehicular networks," in *Proc. IEEE Glob. Commun. Conf.*, 2022, pp. 5735–5740.

[28] K. Zhang, J. Cao, S. Maharjan, and Y. Zhang, "Digital twin empowered content caching in social-aware vehicular edge networks," *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 1, pp. 239–251, Feb. 2022.

[29] Q. Chao et al., "A survey on visual traffic simulation: Models, evaluations, and applications in autonomous driving," in *Proc. Comput. Graph. Forum*, 2020, pp. 287–308.

[30] M. Xu et al., "Unleashing the power of edge-cloud generative ai in mobile networks: A survey of aigc services," 2023, *arXiv:2303.16129*.

[31] A. Stocco, B. Pulfer, and P. Tonella, "Mind the gap! a study on the transferability of virtual vs physical-world testing of autonomous driving systems," *IEEE Trans. Softw. Eng.*, vol. 49, no. 4, pp. 1928–1940, Apr. 2023.

[32] L. U. Khan, Z. Han, W. Saad, E. Hossain, M. Guizani, and C. S. Hong, "Digital twin of wireless systems: Overview, taxonomy, challenges, and opportunities," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2230–2254, Fourthquarter 2022.

[33] L. U. Khan, W. Saad, D. Niyato, Z. Han, and C. S. Hong, "Digital-twin-enabled 6G: Vision, architectural trends, and future directions," *IEEE Commun. Mag.*, vol. 60, no. 1, pp. 74–80, Jan. 2022.

[34] W. Sun, P. Wang, N. Xu, G. Wang, and Y. Zhang, "Dynamic digital twin and distributed incentives for resource allocation in aerial-assisted internet of vehicles," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 5839–5852, Apr. 2022.

[35] R. Xing, Z. Su, Q. Xu, N. Zhang, and T. H. Luan, "Secure content delivery for connected and autonomous trucks: A coalition formation game approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 20522–20537, Nov. 2022.

[36] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 615–637, Mar. 2021.

[37] Y. Wu, X. Gao, S. Zhou, W. Yang, Y. Polyanskiy, and G. Caire, "Massive access for future wireless communication systems," *IEEE Wirel. Commun.*, vol. 27, no. 4, pp. 148–156, Aug. 2020.

[38] L. E. Chatzieleftheriou, M. Karaliopoulos, and I. Koutsopoulos, "Caching-aware recommendations: Nudging user preferences towards better caching performance," in *Prof. IEEE Conf. Comput. Commun.*, 2017, pp. 1–9.

[39] N. Arnosti, M. Beck, and P. Milgrom, "Adverse selection and auction design for internet display advertising," *Amer. Econ. Rev.*, vol. 106, no. 10, pp. 2852–66, Oct. 2016.

[40] T. Zhang and S. Mao, "Energy-efficient federated learning with intelligent reflecting surface," *IEEE Trans. Green Commun. Netw.*, vol. 6, no. 2, pp. 845–858, Jun. 2022.

[41] J. Gao, Y. Wu, S. Shao, W. Yang, and H. V. Poor, "Energy efficiency of massive random access in MIMO quasi-static rayleigh fading channels with finite blocklength," *IEEE Trans. Inf. Theory*, vol. 69, no. 3, pp. 1618–1657, Mar. 2023.

[42] P. Ren et al., "Edge AR X5: An edge-assisted multi-user collaborative framework for mobile web augmented reality in 5G and beyond," *IEEE Trans. Cloud Comput.*, vol. 10, no. 4, pp. 2521–2537, Oct–Dec. 2022.

[43] N. H. Tran, W. Bao, A. Zomaya, M. N. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. IEEE Conf. Comput. Communications,*, Paris, France, 2019, pp. 1387–1395.

[44] W. Zhang et al., "Optimizing federated learning in distributed industrial IoT: A multi-agent approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3688–3703, Dec. 2021.

[45] M. Tang, H. Pang, S. Wang, L. Gao, J. Huang, and L. Sun, "Multi-dimensional auction mechanisms for crowdsourced mobile video streaming," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2062–2075, Oct. 2018.

[46] M. Tang, S. Wang, L. Gao, J. Huang, and L. Sun, "MOMD: A multi-object multi-dimensional auction for crowdsourced mobile video streaming," in *Proc. IEEE Conf. Comput. Commun.*, May 2017, pp. 1–9.

[47] Y.-K. Che, "Design competition through multidimensional auctions," *RAND J. Econ.*, vol. 24, no. 4, pp. 668–680, Aug. 1993.

[48] G. Wang, X. Chen, F. Zhang, Y. Wang, and D. Zhang, "Experience: Understanding long-term evolving patterns of shared electric vehicle networks," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, pp. 1–12.

[49] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10684–10695.

Minrui Xu (Graduate Student Member, IEEE) received the B.S. degree from Sun Yat-Sen University, Guangzhou, China, in 2021. He is currently working toward the Ph.D. degree with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests interests include metaverse, deep reinforcement learning, and mechanism design.

Dusit Niyato (Fellow, IEEE) received the B.Eng. degree from the King Mongkuts Institute of Technology Ladkrabang, Bangkok, Thailand, in 1999 and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Winnipeg, MB, Canada, in 2008. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include the areas of sustainability, edge intelligence, decentralized machine learning, and incentive mechanism design.

Junlong Chen is currently working toward the B.S. degree with the Guangdong University of Technology, Guangzhou, China. His research interests include deep reinforcement learning, Internet of Things, and metaverse.

Hongliang Zhang (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Electrical Engineering and Computer Science, Peking University, Beijing, China, in 2014 and 2019, respectively. He is currently an Assistant Professor with the School of Electronics, Peking University. Prior to that, he was a Postdoctoral Associate with Princeton University, Princeton, NJ, USA, and the University of Houston, Houston, TX, USA. His research interests include reconfigurable intelligent surfaces, aerial access networks, optimization theory, and game theory.

Jiawen Kang (Member, IEEE) received the Ph.D. degree from the Guangdong University of Technology, Guangzhou, China, in 2018. From 2018 to 2021, he was a Postdoc with Nanyang Technological University, Singapore. He is currently a Professor with the Guangdong University of Technology. His research interests include blockchain, security, and privacy protection in wireless communications and networking.

Zehui Xiong (Member, IEEE) received the Ph.D. degree from Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Singapore University of Technology and Design, and also an Honorary Adjunct Senior Research Scientist with Alibaba-NTU Singapore Joint Research Institute, Singapore. He was the Visiting Scholar with Princeton University, Princeton, NJ, USA, and University of Waterloo, Waterloo, ON, Canada. His research interests include wireless communications, IoT, blockchain, edge intelligence, and metaverse.

Shiwen Mao (Fellow, IEEE) is currently a Professor, the Earle C. Williams Eminent Scholar, and the Director of the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interest include wireless networks, multimedia communications, and smart grid. He was the recipient of the IEEE ComSoc TC-CSR Distinguished Technical Achievement Award in 2019 and the NSF CAREER Award in 2010. He was a co-recipient of the 2021 Best Paper Award of Elsevier/KeAi Digital Communications and Networks Journal, 2021 IEEE Internet of Things Journal Best Paper Award, 2021 IEEE Communications Society Outstanding Paper Award, IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, 2004 IEEE Communications Society Leonard G. Abraham Prize in the field of communications systems, and several conference best paper/demo awards. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING and a Distinguished Lecturer of IEEE Communications Society and IEEE Council of RFID.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department and the Computer Science Department, University of Houston, Houston, TX, USA. His main research interests include the novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources. Dr. Han was the recipient of the NSF Career Award in 2010, Fred W. Ellersick Prize of IEEE Communication Society in 2011, EURASIP Best Paper Award for the Journal on Advances in Signal Processing in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (best paper award in IEEE JSAC) in 2016, and several best paper awards in IEEE conferences. Dr. Han is also the winner of the 2021 IEEE Kiyo Tomiyasu Award. Dr. Han was an IEEE Communications Society Distinguished Lecturer during 2015–2018, AAAS Fellow since 2019, and ACM distinguished Member since 2019. Dr. Han is a 1% highly cited Researcher since 2017 according to Web of Science.