Synergizing Generative Intelligence: Advancements in Artificial Intelligence for Intelligent Vehicle Systems and Vehicular Networks

ARCHISMITA GHOSH[1], GADDAM PRATHIK KUMAR[2], PAARTH PRASAD[3], DHEERAJ KUMAR[4], SAMYAK JAIN[5], JATIN CHOPRA[6]
[1] *Software Engineer, LTIMindtree, Kolkata, India*
[2] *Graduate Student, (Computer Science Engineering), Sharda University, Uttar Pradesh, India*
[3, 6] *Undergraduate Student, (Computer Science Engineering), ADGIPS, Delhi, India*
[4] *Graduate Student, (Computer Science Engineering), GITAM, Andhra Pradesh, India*
[5] *Graduate Student, (Computer Science Engineering), WIT, Solapur, India*

Abstract*—This research paper presents a comprehensive exploration of generative artificial intelligence (AI) and its transformative impact on intelligent vehicles and vehicular networks. In the context of intelligent vehicles, the current state and future potential of generative AI technologies, emphasizing their applications in speech, audio, vision, and multimodal interactions are examined. The paper outlines critical future research areas, including domain adaptability, alignment, and multimodal integration, addressing associated ethical challenges. Simultaneously, recognizing the immense benefits of integrating generative AI into intelligent transportation systems, applications and challenges within vehicular networks are discussed. The integration of generative AI enhances various aspects, including navigation optimization, traffic prediction, and data generation, while facing challenges such as real-time processing and privacy concerns. To address these challenges, a multi-modality semantic-aware framework is proposed, leveraging text and image data to enhance generative AI service quality. A deep reinforcement learning (DRL)--based approach for resource allocation in generative AI-enabled vehicle-to-vehicle (V2V) communication is presented in a case study form. By synthesizing insights from both domains, this paper advocates for collaborative research efforts to unlock the full potential of generative AI, fostering transformative advancements in the driving experience and the evolution of intelligent vehicles and vehicular networks.

Indexed Terms*—Generative AI, Intelligent Vehicles, Vehicular Networks, Multi-modal Interactions, V2V, DRL

I. INTRODUCTION

Generative artificial intelligence (AI) within the automotive industry denotes the utilization of AI-driven generative techniques to conceive and advance various facets of vehicle design and development. This involves the generation of simulations depicting design concepts and the facilitation of optimization processes for individual components, thereby contributing to the development of vehicles characterized by heightened efficiency and technological advancement. The integration of generative AI technology assumes a pivotal role in expediting innovation and enhancing the overall efficacy of manufacturing and design procedures within the automotive domain. Furthermore, generative AI contributes to resource and cost optimization by proffering innovative solutions geared towards enhancing fuel efficiency, safety, and overall performance. In essence, it empowers the automotive sector to maintain a leading position in technological innovation, fostering the creation of vehicles that are not only smarter and greener but also inherently more capable, thereby shaping the trajectory of the industry toward a more advanced and sustainable future.

The potential integration of generative artificial intelligence (AI) within vehicular network systems holds the promise of substantial advancements in the realm of intelligent transportation systems (ITS). Through the automation of content creation processes tailored to individualized and proactive

AI requirements, there exists a notable opportunity to significantly enhance the efficacy of services reliant on vehicular networks [1]. By way of illustration, generative AI applications stand poised to autonomously generate pertinent and timely information for drivers, encompassing real-time updates on traffic conditions, weather forecasts, and proximate amenities. Moreover, the technology can be adeptly employed in the generation of specialized driving data, including simulations of accidents, thereby contributing to the augmentation of real-world operational efficiency. This integration of generative AI into vehicular networks not only exemplifies technological innovation but also underscores its practical utility in optimizing the operational dynamics of intelligent transportation systems. Such advancements have the potential to transcend traditional paradigms, presenting a transformative trajectory for the future of vehicular network-based services.

Concurrently, Waymo Corp. introduced the VectorNet model as part of their initiatives, aiming to prognosticate the conduct of traffic agents and their interrelations within the encompassing environment within the context of autonomous driving scenarios [2]. In this vein, the incorporation of generative artificial intelligence (AI) emerges as a pivotal mechanism with the potential to enhance the operational efficiency and user convenience of services reliant on vehicular networks. This augmentation is achieved through the automated facilitation of discrete facets within the content creation process, meticulously aligned with the personalized requisites of end-users.

Notwithstanding the manifold benefits intrinsic to the integration of generative artificial intelligence (AI) within vehicular networks, the implementation of such systems confronts several challenges and limitations. Foremost among these challenges is the constraint imposed by limited communication bandwidth, thereby engendering potential lapses in data transmission reliability and instigating concerns about communication security within the ecosystem of vehicular networks. Furthermore, the presence of latency in the communication interface between vehicles and infrastructure, concomitant with the intricate nature of data processing, constitutes a substantive impediment to the pervasive adoption of vehicular networks.

Initially, this study introduces diverse generative artificial intelligence (AI) techniques, subsequently delineating their potential applications and challenges within the context of vehicular network scenarios. Subsequently, in response to the identified challenges, an innovative generative AI-enabled framework for vehicular networks is presented. This framework adopts a multi-modal architecture, proficient in assimilating both textual and image data. Such integration is designed to furnish more dependable guidance to recipient vehicles. In furtherance of enhancing the dependability and efficiency of information transmission within the delineated framework, a deep reinforcement learning (DRL)-based approach is posited. This approach seeks to optimize the definite system quality of experience (QoE) while adhering to constraints imposed by transmission power budgets and the probability of successful transmission for each vehicle. The proposed approach concurrently optimizes both transmission communication resources and generative AI resources, with simulation results validating its efficacy.

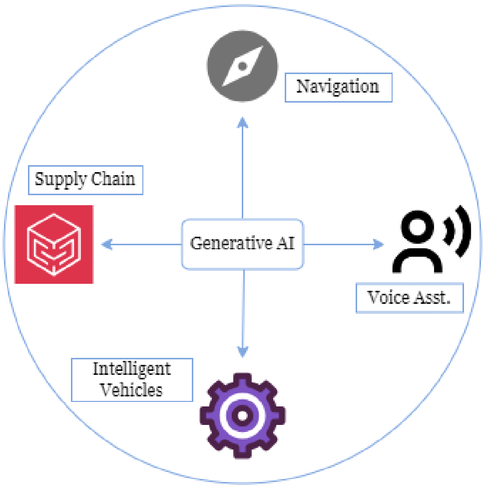

Fig | 1: Gen AI Applications for Automotive Manufacturing

AI-augmented functionalities within vehicular systems are anticipated to play a pivotal role in shaping the future of mobility, encompassing a broad categorization into perception and generation systems. Notably, extensive research has been devoted to perception models, which are instrumental in comprehending intra-vehicular interactions [10], [11], and prognosticating passengers' behaviors and emotional states [12]. In contrast, the exploration of generative models has

been comparatively less exhaustive. Noteworthy applications include leveraging Variational Autoencoders (VAEs) for unsupervised domain adaptation among automotive sensors to enhance visual classification [13]. Furthermore, conditional Generative Adversarial Networks (GANs) have been employed to bolster the capabilities of autonomous vehicles by simulating more realistic and diverse real-world scenarios, as demonstrated in the work of [14]. Within the domain of audio generation, the SoundsRide system [15] introduces an in-car audio system designed to synchronize music with real-time sound affordances along the journey, offering an immersive music experience. This integration holds implications for driving safety, with effects contingent on the mix and user engagement. In the context of enhancing dialogue and personalized voice characters for in-car speech interfaces, [16] investigates the design and impact, juxtaposing four distinct assistant personalities (friend, admirer, aunt, and butler). The findings posit that aligning the assistant's personality with that of the user engenders heightened likability and trust. Nevertheless, extant research within the realm of generative AI remains circumscribed both in terms of the breadth and depth of applications and in the domain of multimodal modeling. The existing body of research, while notable, underscores the need for further exploration and diversification to comprehensively comprehend and harness the potential of generative AI in the multifaceted landscape of vehicular systems.
The array of generative artificial intelligence (AI) approaches is primarily circumscribed by distinct training types, each exhibiting unique advantages in specific contextual applications, as elucidated in Table I. Unsupervised learning methodologies, exemplified by generative adversarial networks (GANs) and variational autoencoders (VAE), are adept at generating data samples. Notably, the StableDiffusion architecture integrates VAE with a reconstruction (diffusion) process, thereby facilitating the generation of photo-realistic images [17]. This capability extends beyond images to encompass audio generation without the prerequisite of labeled data, rendering it particularly valuable in scenarios where data annotations are scarce [18].

In the realm of supervised generative modeling, architectures such as conditional GANs (e.g., pix2pix) are instrumental. Trained with an abundance of labeled data, these models excel in generating data samples contingent upon specific attributes, proving particularly advantageous when adherence to predefined conditions is imperative [19]. The foundational models, which serve as the basis for diverse downstream tasks, predominantly employ self-supervised learning in their training techniques. This is exemplified by autoregressive transformer models like GPT-3 [20], which generate their supervision signal from input data, allowing them to harness extensive datasets. By predicting components of the input from the remaining information, such models prove advantageous in generating output that is both coherent and contextually relevant. In recent developments, reinforcement learning has garnered increased prominence for learning intricate behaviors. Notably, techniques such as fine-tuning models to align with human expectations using proximal policy optimization [21] have gained traction. In this paradigm, a supervised reward function is initially learned, subsequently serving as an evaluator for the generated data samples. This closed-loop reward system facilitates automatic refinement of the model outputs, enhancing their alignment with desired criteria.

This paper adopts a practical approach, seeking to develop a research agenda that explores the application of generative artificial intelligence technologies in intelligent vehicles. The role of AI in augmenting intelligent vehicle functions and the potential of generative artificial intelligence to facilitate multimodal interaction, encompassing audio, video, and speech in these systems, are examined. The research agenda is guided by key principles, such as model capabilities, ethical considerations, and the alignment of AI models.

II. RELATED BACKGROUND

TABLE I COMPREHENSIVE ANALYSIS OF MODEL-BASED GENERATIVE AI TECHNOLOGIES

GEN-AI TECHNOLOGIES	Advantages	Disadvantages

Diffusion-based model [1]	- Resistance to mode collapse and overfitting -Generative process based on denoising score matching	-Complexity in training and selecting hyperparameters
Variational Autoencoder [3]	- Unified framework for both inference and generation - Effective representation of data in a probabilistic latent space	-Difficulty in choosing likelihood functions - Limited expressive power of the prior distribution
Autoregressive model [4]	- Compatibility with both discrete and continuous data -Scalability achieved through parallelization techniques	-Sequential nature limits speed - Complexity increases with sequence length
Generative adversarial network [5]	- Ability to address mode collapse with advanced techniques - Exceptional quality of generated images	-Difficulty in evaluating model quality

TABLE 2 COMPREHENSIVE ANALYSIS OF DATA-BASED GENERATIVE AI TECHNOLOGIES

GEN-AI TECHNOLOGIES	Advantages	Disadvantages
Example-based synthesis [6]	- Flexibility in incorporating domain-specific constraints -Leveraging existing data for generation	- Computational complexity - Dependence on the quality and variety of input examples
Transformer-based model [7]	- Scalability with parallelization techniques - Effective modelling of long-range dependencies	- Dependence on large training datasets - Memory consumption in large-scale applications
Neural style transfer [8]	- Real-time style transfer with optimized algorithms - Ability to combine style and content from different images	- Difficulty in controlling the level of stylization - Dependence on high-quality style and content images
Pixel CNN/Pixel RNN [9]	- Autoregressive modelling of image pixels	-Sequential nature limits generation speed

- GENERATIVE AI-ENABLED VEHICULAR NETWORKS

Generative artificial intelligence (AI) technology, as an emerging subfield within the broader domain of AI, is dedicated to the autonomous generation of diverse content or data, spanning images, text, audio, video, and even system designs. Analogous to conventional AI technologies, such as discriminative AI, generative AI technologies can be categorically delineated into two overarching classes: model-based technologies and data-based technologies. In the context of data-based generative AI, learning algorithms are employed to autonomously generate novel content derived from existing datasets. Conversely, model-based generative AI harnesses predefined models and simulations, manipulating model parameters to generate new outputs. These methodologies are adept at discerning patterns and structures inherent in existing datasets, yielding outputs that closely approximate real-world samples. Furthermore, these technologies facilitate the emulation of human creativity, imagination, and problem-solving capacities through their proficiency in generating content that mirrors the intricacies of human-like output.

Attaining a profound comprehension of generative AI techniques tailored to specific applications within vehicular networks is imperative for the optimization of performance and user experience. Notably, the deployment of Generative Adversarial Networks (GANs) for data augmentation emerges as a strategic approach, enhancing the efficacy of AI models in tasks such as object detection or traffic prediction systems [5]. Concurrently, Variational Autoencoders (VAEs) assume significance in the realm of dimensionality reduction and feature extraction for sensor data, thereby facilitating more streamlined processing and analysis [3]. The discernment of both the strengths and limitations inherent in each generative AI technique is fundamental for developers tasked with the implementation of the aptest strategy for particular applications within vehicular networks. This judicious selection of techniques contributes substantively to the enhancement of overall system performance and user satisfaction. To facilitate a comprehensive overview, the advantages and disadvantages of diverse generative AI technologies are succinctly delineated in Table I, thereby providing a structured framework to harness the full potential of generative AI-enabled vehicular networks and their associated applications.

POTENTIAL APPLICATIONS OF GENERATIVE AI IN INTELLIGENT VEHICLES

1. Audio:

Generative audio artificial intelligence (AI) holds the transformative potential to redefine the auditory landscape, affording users an unprecedented degree of personalization and immersion. An illustrative application involves the generation of tailored welcome melodies upon the user's entry into vehicles, eliciting a sense of familiarity and inclusion [22], [23]. As the prevalence of semi-autonomous vehicle functionalities increases, facilitating heightened driver engagement, the prospect of individualized auditory warning signals and a guidance system tailored for individuals with disabilities in fully autonomous driving scenarios becomes feasible. Additionally, electric vehicles stand to gain from AI-generated in-car sound modeling, which can simulate engine noises or offer customized audio experiences [24] for passengers. This application adheres to safety considerations by avoiding an impact on exterior noise levels.

2. Vision:

The domain of visual interactions presents a diverse spectrum of engaging features, further augmented by generative artificial intelligence (AI) applications. Notably, AI-generated persona avatars [25] assume an interactive role, actively listening, empathizing, and responding with emotions to establish a deeper connection between users and the vehicle's assistant. This visual representation can also tailor projections to intended recipient screens, such as the driver's head-up display, thereby individualizing the user experience. This expanded interaction surpasses the constraints inherent in conventional voice-only communication. Moreover, the navigation and information system can harness generative AI to dynamically adapt its visual appearance [26], taking into account contextual factors such as ambient light sensors, weather data, user preferences, interior design, and road type. This adaptability extends to the creation of custom album art, artist portraits, or visualizations based on the audio content being played. Additionally, AI-generated visual content can be deployed for the creation of personalized LED night sky displays and interactive animations projected from the door handle, thereby enhancing the in-car entertainment experience with an element of wonder and fascination. Furthermore, generative AI contributes to the efficiency of accident reporting by facilitating the creation of visual summaries, which can be transmitted to emergency services, potentially streamlining response times. As evidenced in [27], generative AI holds promise in enhancing training data for driver assistance systems, generating realistic, diverse, and high-quality synthetic data. The adaptability of visual style transfer, involving the translation of signage characteristics between different countries through typical colors and symbols, is also within the purview of generative AI. However, it is crucial to acknowledge that the practical implementation of such models in safety-critical domains faces challenges due to the stringent regulatory environment, as previously noted.

3. Multimodal:
The ongoing progression of generative artificial intelligence systems is distinguished by a paradigm shift toward multimodal models [28], designed to emulate human cognition through the seamless processing and integration of multiple modalities. This holistic approach affords a more profound understanding of user requirements, thereby cultivating a more enriched and intuitive user experience. Consider, for instance, the scenario of vehicle issue diagnosis: a multimodal AI system could adeptly comprehend a user's verbal description of a problem while concurrently analysing visual cues from the vehicle's sensors or user-provided images. This synergistic processing of information holds the potential to yield expeditious and precise diagnoses, accompanied by tailored recommendations.

4. Speech:
Generative voice artificial intelligence (AI) holds considerable potential in enhancing the vocal capabilities of intelligent vehicles, offering users a more immersive and interactive experience. Voice interaction emerges as an optimal modality for drivers to access information and receive decision support without diverting their attention from the road, thereby concurrently enhancing user experience and safety. The efficacy of voice interaction is contingent upon the system's ability to meet drivers' expectations and in still trust. By imbuing generative models within vehicle assistants with the capacity to engender more humanlike dialogues [29], the system can furnish context-aware and personalized responses to user queries and requests, ensuring seamless and efficient interactions. Moreover, a productivity assistant, as depicted in Figure 1, has the potential to assist users in composing and editing emails, preparing presentations, and executing other tasks through advanced voice interactions, thereby contributing to a more productive in-car environment. Generative AI systems also facilitate the generation of emotional voice [30] and personality, enabling the voice assistant to express emotions and modulate its tone to align with brand attributes, user preferences, or specific contextual scenarios. This personalized approach fosters a heightened connection between the user and the vehicle. Furthermore, the integration of generative artificial intelligence introduces the capability for real-time translation services [31] for radio traffic warnings or conversations between drivers and passengers. This functionality mitigates language barriers, enhancing communication during foreign travel or interactions with passengers speaking diverse languages.

III. PROPOSED FRAMEWORK/MODEL

The integration of Vehicle-to-Vehicle (V2V) communication within vehicular networks presents a prospective avenue for enhancing road safety. An envisioned application involves leveraging multiple cameras in vehicles to capture real-time images of road conditions. In the event of a car accident, for instance, an adjacent vehicle could capture an image of the incident and transmit it to the respective driver through a smart assistant. This mechanism serves the dual purpose of alerting other drivers and mitigating the likelihood of subsequent accidents. However, the practicality of transmitting high-resolution images in real-time confronts challenges attributable to stringent latency requirements and the constrained bandwidth resources inherent in vehicular networks.

Semantic communication technology emerges as a viable remedy, offering a solution by facilitating the extraction of pertinent information from images, thereby significantly diminishing transmitted data and mitigating communication delays [11]. Furthermore, the transmission of this distilled information serves the additional purpose of safeguarding sensitive details, such as faces and license plates, present in images. As depicted in Figure, conventional generative artificial intelligence (AI)-enabled vehicles can employ semantic communication technology to extract information from images at the transmitter end, subsequently restoring it through the utilization of generative AI technology at the receiver end. This procedural approach engenders a reduction in transmitted data, thereby diminishing communication delays and augmenting the efficiency of Vehicle-to-Vehicle (V2V) communication within vehicular networks. Nevertheless, a challenge persists with existing generative AI models reliant solely on essential text information. Such models encounter uncertainty and unreliability, impeding their ability to furnish precise road information to smart assistants in other vehicles. For instance, when a receiving vehicle encounters a message like "Two vehicles have been involved in a traffic accident near a green belt," despite the generative AI generating a high-quality image of the accident, substantial disparities between the generated image and the actual accident scene may ensue. This unreliability poses the risk of vehicles receiving inaccurate road information, a scenario incompatible with the safety-centric objectives of vehicular networks. Consequently, there arises a need for a more dependable and efficient generative AI framework within vehicular networks to address these challenges, providing accurate road information to facilitate informed decision-making and enhance overall safety.

In this section, we present a multi-modality semantic-aware framework designed for generative AI-enabled vehicular networks. The primary objective of this framework is to mitigate the substantial discrepancies observed between transmitted and recovered images, thereby addressing a pertinent challenge in the context of image transmission within vehicular networks.

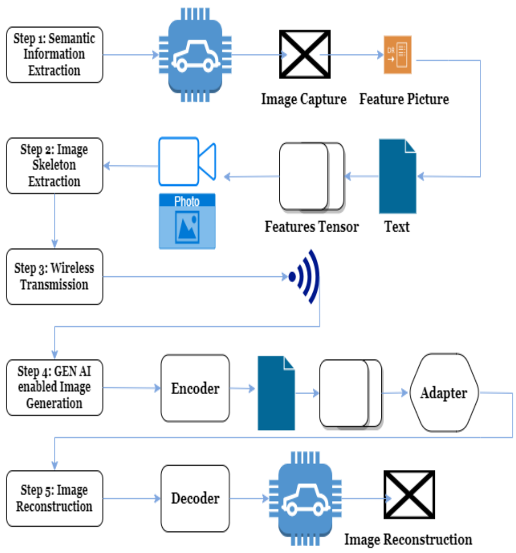

Fig | 2: Multi-modality Semantic-aware Framework for generative AI-enabled vehicular networks.

Step 1: Semantic Information Extraction
In the initial stage of our proposed framework for generative AI-enabled vehicular networks, akin to conventional frameworks, semantic information extraction is conducted from real-time road images. Employing computer vision techniques, this process involves the identification and classification of objects within the images, encompassing vehicles, pedestrians, road signs, and traffic lights. Attributes such as speed, direction, and location of these identified objects are extracted, contributing to an enhanced understanding of the prevailing road environment. It is noteworthy that this extraction process can be optimized, taking into consideration factors such as wireless transmission resources and image reconstruction algorithms.

Step 2: Image Skeleton Extraction
The subsequent step involves the extraction of the image skeleton, which plays a pivotal role in providing a fundamental structure for the image and serves as the foundation for more advanced semantic analyses. This process entails the identification of edges and contours within the road condition image, resulting in the creation of a simplified representation of these features. The resultant image skeleton, capturing essential features without unnecessary details, is utilized for subsequent semantic analyses, including object recognition, lane detection, and traffic sign identification. The extraction of the image skeleton

is thus a critical step in generating reliable and pertinent semantic information from road condition images.

Step 3: Wireless Transmission
Following the extraction of the image skeleton and text information, these components are amalgamated into a compact data package, which is subsequently wirelessly transmitted to the receiving vehicle through Vehicle-to-Vehicle (V2V) communication. Notably, the data volume of the skeleton information, approximately 0.5 megabytes, is considerably smaller than that of the original captured image, approximately 6.7 megabytes. This approach conserves bandwidth resources, ensuring that the receiving vehicle obtains fundamental information for the reliable reconstruction of the road condition image.

Step 4: Generative AI-Enabled Image Generation
The generative AI model utilizes the extracted image skeleton and semantic information to generate a road condition image closely mirroring the actual scene. The image skeleton provides the foundational structure, while semantic text information augments the details, encompassing road signs, vehicles, and pedestrians. Leveraging both components, the generative AI model creates a more dependable road condition image, facilitating informed decision-making for vehicles on the road. Additionally, the model is capable of generating images tailored to diverse scenarios, such as varying weather conditions, times of day, or traffic situations, thereby offering a comprehensive understanding of road conditions. Optimization of this generation process is achievable based on factors such as diffusion steps.

Step 5: Image Reconstruction
The generated road condition image is reconstructed on the receiving end and provided to the intelligent assistant, which, in turn, alerts the driver to potential hazards. The intelligent assistant analyses the reconstructed image, comparing it with the current road condition to identify potential hazards such as accidents or roadblocks. In the event of hazard detection, the assistant issues an alert to the driver through audio or visual cues, prompting necessary precautions. This systematic process ensures that the vehicle remains well-informed about road conditions, facilitating appropriate actions to uphold the safety of occupants and other road users. By adhering to these sequential steps, the proposed framework effectively mitigates data transmission requirements while enhancing the reliability of road condition images within vehicular networks. Moreover, the framework contributes to heightened driving safety and a reduction in accident occurrences, establishing its noteworthy significance in the domain of vehicular networks.

IV. CASE STUDY: GENERATIVE AI-ENABLED V2V RESOURCE ALLOCATION

In the context of generative AI-enabled Vehicle-to-Vehicle (V2V) systems, where transmission rate and image similarity constitute critical metrics, we propose the adoption of a novel Quality of Experience (QoE) metric grounded in the Weber-Fechner law, aiming to integrate these indicators within the overarching system optimization objective [13]. The Weber-Fechner law delineates the correlation between the magnitude of a physical stimulus and the perceived intensity, with our application associating it with the perceived quality of transmitted images. By incorporating this law into the devised QoE metric, a comprehensive consideration of both transmission rate and image similarity is achieved in a unified manner. The optimization objective for resource allocation in generative AI-enabled V2V systems is articulated as follows: Maximize the system QoE while adhering to constraints imposed by the transmission power budget and the probability of successful transmission for each vehicle. This optimization necessitates the fine-tuning of several parameters, including channel selection, transmission power allocation for each vehicle, and the determination of diffusion steps for the insertion of the skeleton. This approach ensures the efficient allocation of resources to individual V2V links, thereby maximizing overall system performance. Simultaneously, it guarantees the reliable and consistent sharing of safety information among vehicles within the vehicular networks.

A. Proposed Double Deep Q-Network (DDQN)-Based Approach and Simulations:

In addressing the resource allocation challenges inherent in generative AI-enabled Vehicle-to-Vehicle (V2V) communication, we advocate for a novel approach rooted in double deep Q-network

(DDQN) techniques. DDQN, a form of deep reinforcement learning, leverages two neural networks [14], namely the online network and the target network, to approximate the Q-value function. The online network is responsible for action selection based on the current state, while the target network estimates the Q-value for the subsequent state. The utilization of a DDQN-based approach enhances training stability and mitigates the risk of Q-value overestimation.

1) Actions: The action space encompasses optimizing parameters, specifically the selectable channel, transmit power, and diffusion steps. Binary variables within the action space indicate the selection or non-selection of sub-channels. For transmit power and diffusion steps, practical circuit limitations necessitate quantization into multiple values to facilitate learning.
2) States: The definition of the state space aims to encapsulate pertinent environmental information for the considered problem [15]. In our proposed DDQN-based approach, the state space comprises two components: current information and previously selected actions. The current information encompasses channel details for each V2V link, transmission rates for each link, and the payload of the generated image. The previously selected action refers to the action undertaken in the preceding time step.

V. RESULTS & DISCUSSION

Training Stage Analysis: In the training phase, as illustrated in Fig. 3(a), a comparison of the average rewards is conducted across four distinct resource allocation strategies over increasing iteration numbers. The curves are subjected to smoothing through a sliding window to enhance clarity and delineate the overall trend. Notably, the proposed DDQN-based approach consistently outperforms the greedy-based and random-based strategies in terms of achieving higher rewards and converging more expediently. Moreover, while the rewards obtained by the proposed DDQN and DQN during training exhibit similarity, the DDQN-based approach demonstrates superior performance post-convergence, with diminished fluctuations. This outcome is attributed to the DDQN's efficacy in mitigating overestimation by decoupling the Q-target.

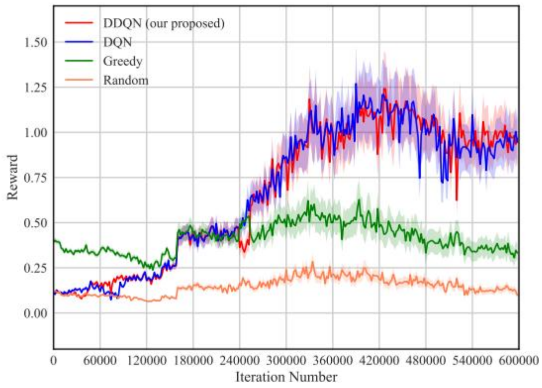

Fig | 3. a) The obtained reward versus the number of iteration number with the different approaches.

Testing Stage Evaluation: Moving to the testing stage, Fig. 3(b) portrays the Quality of Experience (QoE) achieved by the four resource allocation strategies concerning the size of the image payload. The results highlight the superiority of the proposed DDQN-based approach over the comparative benchmarks. Additionally, the figure underscores the positive impact of increasing image payload size on system QoE. A larger payload implies a more substantial transmission of semantic information, thereby enhancing overall system QoE.

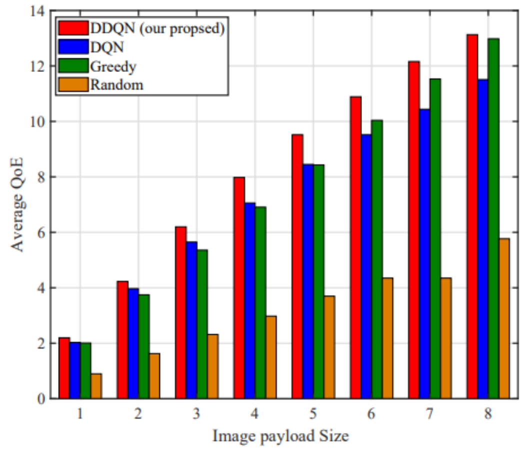

Fig | 3. b) The average QoE versus the size of the image payload with the different approaches.

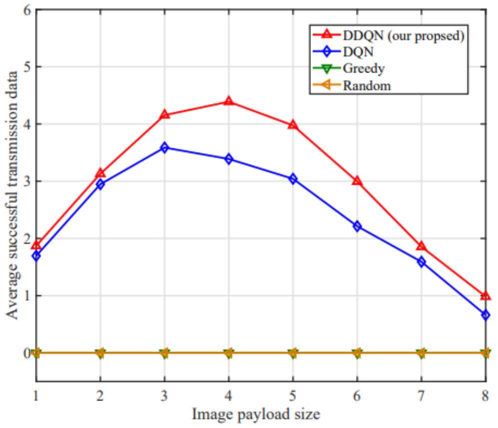

Fig | 3. c) The average successful transmission data with the different approaches.

Impact of Image Payload Size on Received QoE: To further explore the influence of image payload size on received QoE, Fig. 3(c) is presented. The findings reaffirm the superiority of the proposed DDQN-based approach over alternative benchmarks, with random and greedy strategies consistently yielding zero data transmission due to their inability to guarantee the stipulated successful transmission outage probability for each vehicle. Furthermore, an intriguing observation is made that, as the image payload increases, the average amount of data for successful image transmission initially rises before experiencing a subsequent decline. This phenomenon is attributed to potential congestion within vehicular networks and an elevated likelihood of packet loss under high payload conditions. Consequently, an increase in retransmissions occurs, leading to a reduction in the overall transmission success rate.

CONCLUSION

The integration of generative artificial intelligence (AI) technologies into the automotive industry heralds a transformative era for intelligent vehicles, promising enhanced in-car experiences characterized by immersion, intuition, and personalization. This paper has meticulously surveyed contemporary applications and future research trajectories in the domain of generative AI and intelligent vehicles. The exploration underscores the potential of these technologies to redefine user interactions and spur innovation within the sector. Future research imperatives encompass multimodal integration, model optimization, personalization, reliability, and architectural considerations. Ethical dimensions, notably concerning user privacy, data security, and the prevention of potential misuse, emerge as crucial focal points. Collaborative efforts to address these challenges will unlock the full potential of generative AI, reshaping the driving experience and shaping the trajectory of intelligent vehicles. While the current perspective centres on driving as the primary task, acknowledging a potential shift in priorities is imperative. The evolution of generative AI suggests a future where intelligent vehicles cater to diverse user needs, potentially shifting interaction modalities and augmenting use-cases to combat passive fatigue. As generative AI advances, the development of intelligent vehicles attuned to user preferences will play an increasingly pivotal role, fundamentally altering our interaction with the surrounding world. In tandem with this, the research introduces the concept of generative AI-enabled vehicular networks, elucidating the role of generative AI technologies and their application scenarios. The proposed multi-modality semantic-aware framework aims to elevate the service quality of generative AI, incorporating both semantic and multi-modal technologies. Addressing the challenge of system transmission efficiency in generative AI-enabled Vehicle-to-Vehicle (V2V) communication, a reinforcement learning-based approach is presented. Simulation results validate the efficacy of the proposed approach in improving system performance. Finally, potential avenues for future research in the domain of generative AI-enabled vehicular networks are outlined, foreseeing continued advancements and broader applications in the evolving landscape of intelligent transportation systems.

The incorporation of generative artificial intelligence (AI) in the automotive sector transcends a mere technological progression; rather, it signifies a paradigmatic transformation that fundamentally redefines the core aspects of vehicle design, production, and operation. Its far-reaching impact extends to safety augmentation, unparalleled customization capabilities, streamlined design and manufacturing processes, and the evolution of autonomous driving systems. The amalgamation of human creativity with machine intelligence facilitated by generative AI portends a future wherein automobiles transcend their conventional role as mere modes of transportation. Instead, they embody a harmonious synthesis of innovative concepts and personalized technologies, heralding a future characterized by experiential richness on the road, surpassing the utilitarian function of point-to-point transportation. The ongoing development of advanced driver-assistance systems and the imminent realization of fully autonomous vehicles underscore the pivotal role of AI in reshaping the trajectory of transportation. As technological advancements persist, the driving experience is poised for further enhancement, with a concomitant reduction in accidents and a contributory role in fostering a more sustainable and interconnected automotive milieu. The automotive industry's proactive integration of AI, marked by collaborative initiatives with AI development entities, represents a decisive stride towards the realization of safer, more efficient, and intelligently orchestrated mobility solutions.

REFERENCES

[1] C. Mou, X. Wang, L. Xie, Y. Wu, J. Zhang, Z. Qi, Y. Shan, and X. Qie, "T2I-Adapter: Learning adapters to dig out more controllable ability for text-to-image diffusion models," 2023.

[2] Z. Yang, Y. Chai, D. Anguelov, Y. Zhou, P. Sun, D. Erhan, S. Rafferty, and H. Kretzschmar, "Surfelgan: Synthesizing realistic sensor data for autonomous driving," in Proc. IEEE CVPR, June 2020.

[3] W. Jin, R. Barzilay, and T. Jaakkola, "Junction tree variational autoencoder for molecular graph generation," in Proc. ICML, 2018. [Online]. Available: https://proceedings.mlr.press/v80/jin18a.html

[4] H. Du, Z. Li, D. Niyato, J. Kang, Z. Xiong, Xuemin, Shen, and D. I. Kim, "Enabling AI-generated content (AIGC) services in wireless edge networks," 2023.

[5] S. W. Kim, J. Philion, A. Torralba, and S. Fidler, "DriveGAN: Towards a controllable high-quality neural simulation," in Proc. IEEE CVPR, June 2021.

[6] M. Fisher, D. Ritchie, M. Savva, T. Funkhouser, and P. Hanrahan, "Example-based synthesis of 3D object arrangements," ACM Trans. Graph., vol. 31, no. 6, 2012. [Online]. Available: https://doi.org/10. 1145/2366145.2366154

[7] E. A. van Dis, J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting, "ChatGPT: Five priorities for research," Nature, vol. 614, no. 7947, pp. 224–226, 2023.

[8] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, "Neural style transfer: A review," IEEE Trans. Vis. Comput. Graph., vol. 26, no. 11, pp. 3365–3385, 2020.

[9] S. Kong and C. C. Fowlkes, "Recurrent pixel embedding for instance grouping," in Proc. IEEE CVPR, June 2018.

[10] Lukas Stappen, Georgios Rizos, and Bjorn Schuller. X-aware: Context- ¨ aware human-environment attention fusion for driver gaze prediction in the wild. In Proceedings of the 2020 International Conference on Multimodal Interaction, ICMI '20, pages 858—-867, New York, NY, USA, 2020. ACM.

[11] Lukas Stappen, Xinchen Du, Vincent Karas, Stefan Muller, and ¨ Bjorn W Schuller. Domain adaptation with joint learning for generic, ¨ optical car part recognition and detection systems (go-card). arXiv preprint arXiv:2006.08521, 2020.

[12] Sebastian Zepf, Javier Hernandez, Alexander Schmitt, Wolfgang Minker, and Rosalind W Picard. Driver emotion recognition for intelligent vehicles: A survey. ACM Computing Surveys (CSUR), 53(3):1–30, 2020.

[13] Simon Reiß, Alina Roitberg, Monica Haurilet, and Rainer Stiefelhagen. Deep classification-driven domain adaptation for cross-modal driver behavior recognition. In 2020 IEEE Intelligent Vehicles Symposium (IV), pages 1042–1047, 2020.

[14] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8798– 8807, 2018.

[15] Mohamed Kari, Tobias Grosse-Puppendahl, Alexander Jagaciak, David Bethge, Reinhard Schutte, and Christian Holz. Soundsride: ¨ Affordance-synchronized music mixing for in-car audio augmented reality. In The 34th Annual ACM Symposium on User Interface Software and Technology, UIST '21, pages 118–133, New York, NY, USA, 2021. ACM.

[16] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, and Florian Alt. At your service: Designing voice assistant

personalities to improve automotive user interfaces. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI ’19, page 1–11, New York, NY, USA, 2019. ACM.
[17] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent ¨ diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10684–10695, 2022.
[18] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8798– 8807, 2018.
[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1125–1134, 2017.
[20] Luciano Floridi and Massimo Chiriatti. Gpt-3: Its nature, scope, limits, and consequences. Minds and Machines, pages 1–14, 2020.
[21] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35:27730–27744, 2022.
[22] Christine Payne. Musenet. OpenAI Blog, 3, 2019.
[23] Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. Jukebox: A generative model for music. arXiv preprint arXiv:2005.00341, 2020.
[24] Andrea Agostinelli, Timo I Denk, Zalan Borsos, Jesse Engel, Mauro ´ Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. Musiclm: Generating music from text. arXiv preprint arXiv:2301.11325, 2023.
[25] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8110–8119, 2020.
[26] Pamela Mishkin, Lama Ahmad, Miles Brundage, Gretchen Krueger, and Girish Sastry. Dall·e 2 preview - risks and limitations. 2022.
[27] Zijie Guo, Rong Zhi, Wuqaing Zhang, Baofeng Wang, Zhijie Fang, Vitali Kaiser, Julian Wiederer, and Fabian Flohr. Generative model based data augmentation for special person classification. In 2020 IEEE Intelligent Vehicles Symposium (IV), pages 1675–1681, 2020.
[28] Ruben Villegas, Mohammad Babaeizadeh, Pieter-Jan Kindermans, Hernan Moraldo, Han Zhang, Mohammad Taghi Saffar, Santiago Castro, Julius Kunze, and Dumitru Erhan. Phenaki: Variable length video generation from open domain textual description. arXiv preprint arXiv:2210.02399, 2022.
[29] OpenAI. Gpt-4 technical report, 2023.
[30] Jonathan Shen, Ruoming Pang, Ron J Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, Rj Skerrv-Ryan, et al. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 4779– 4783. IEEE, 2018.
[31] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via largescale weak supervision. arXiv preprint arXiv:2212.04356, 2022.
[32] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the people: The role of humans in interactive machine learning. AI Magazine, 35(4):105–120, 2014.
[33] Advait Sarkar. Confidence, command, complexity: metamodels for structured interaction with machine intelligence. In PPIG, page 3, 2015.
[34] Taina Bucher. The algorithmic imaginary: Exploring the ordinary affects of facebook algorithms. Information, communication & society, 20(1):30–44, 2017.
[35] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. Bringing

transparency design into practice. In 23rd International Conference on Intelligent User Interfaces, pages 211–223, 2018.