A Survey on Large Language Models and Their Applications in Intelligent Transportation Systems

www.surveyx.cn

Abstract

This survey paper explores the transformative potential of Large Language Models (LLMs), Vision-Language Models (VLMs), and Multimodal Large Language Models (MLLMs) in Intelligent Transportation Systems (ITS). LLMs, with their advanced natural language processing capabilities, enhance decision-making and resource allocation, thereby improving operational efficiency in transportation. VLMs integrate visual and textual data to enhance autonomous systems' interpretative capabilities, crucial for tasks such as object detection and scene understanding, ultimately improving safety and reliability. MLLMs further extend these capabilities by adeptly integrating diverse data modalities, essential for managing the complexities of modern transportation systems, ensuring intelligent and adaptive networks. The paper details the applications of these models in autonomous vehicles, focusing on navigation, obstacle avoidance, and human-vehicle interaction, and highlights their role in traffic management through real-time monitoring and congestion prediction. Despite their potential, the deployment of these models raises significant challenges and ethical considerations, including computational resource demands and data privacy concerns. Future research directions emphasize enhancing model efficiency, refining ethical frameworks, and improving interpretability to ensure responsible integration into ITS. By leveraging these advanced models, transportation systems can achieve greater efficiency, safety, and adaptability, driving innovation and efficiency across various sectors.

1 Introduction

1.1 Overview of LLMs, VLMs, and MLLMs

Large Language Models (LLMs), Vision-Language Models (VLMs), and Multimodal Large Language Models (MLLMs) are at the forefront of artificial intelligence, each offering unique capabilities that enhance various technological applications. LLMs are designed for processing and generating human language, with applications in text generation, translation, and data augmentation [1]. They have been instrumental in automating complex tasks, such as evaluating responses in medical Question and Answer (QA) systems [2] and extracting structured information from unstructured text [3]. However, challenges remain, particularly the generation of factually incorrect responses, known as hallucinations [4]. Ongoing research aims to enhance LLMs' decision-making capabilities in complex environments through optimization algorithms [5].

VLMs build on LLMs by incorporating visual data, facilitating a comprehensive interpretation of both visual and textual information. This integration is crucial for tasks requiring nuanced visual context understanding, such as object detection and scene comprehension [6]. In healthcare, VLMs have demonstrated significant potential by merging computer vision with natural language processing [7]. Furthermore, VLMs enhance our understanding of the world by leveraging the interplay between language and vision [8].

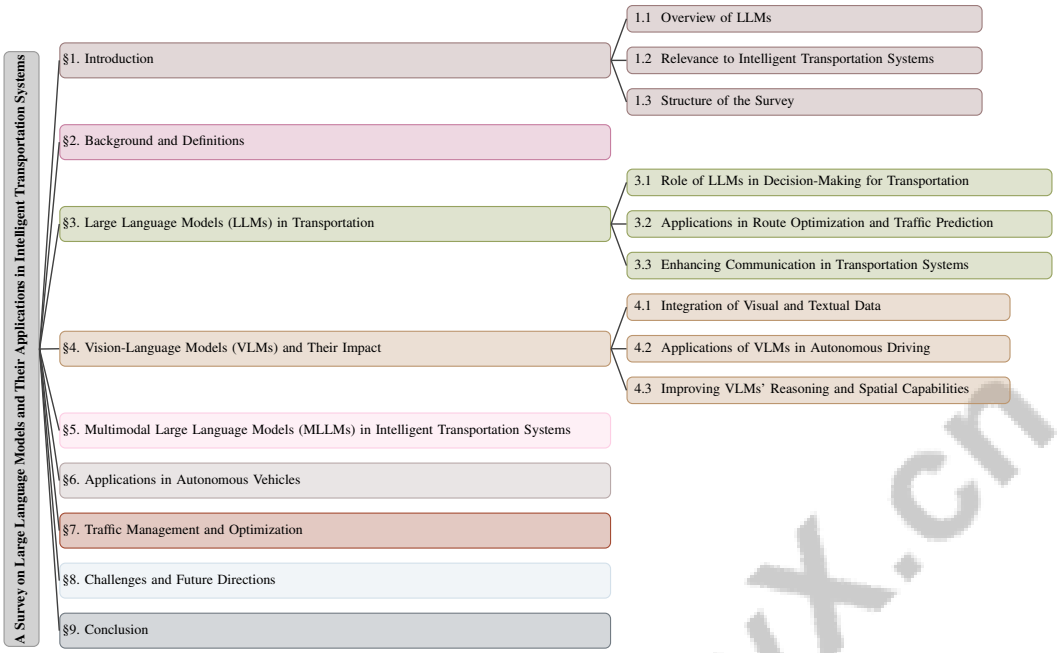

Figure 1: chapter structure

MM-LLMs represent a significant advancement, adeptly integrating multiple data modalities to improve input understanding and output generation. They address the limitations of LLMs and VLMs by providing a holistic approach to data processing [6]. The transition from LLMs to MLLMs has implications across various domains, including medical practice [9]. Collectively, LLMs, VLMs, and MLLMs are pivotal in advancing intelligent systems, applicable to autonomous vehicles, traffic management, and enhancing human-machine interactions. They also play a critical role in knowledge integration tasks like Ontology Matching (OM) and Ontology Learning (OL), fostering data interoperability and knowledge representation across diverse domains. The integration of LLMs such as ChatGPT into clinical applications, including automated dental diagnosis and treatment planning, further illustrates their expanding role in healthcare [10].

1.2 Relevance to Intelligent Transportation Systems

The integration of LLMs, VLMs, and MLLMs into Intelligent Transportation Systems (ITS) represents a transformative shift in addressing complex transportation challenges. LLMs enhance the analysis of user feedback by converting unstructured data into actionable insights, thereby improving public transit services and aligning with technology acceptance models [3, 11]. They also show promise in emergency management decision-making, indicating a parallel impact on ITS [12].

VLMs are essential for autonomous vehicle operations, particularly in object detection and scene interpretation [13]. Their ability to interpret complex scenes enhances human-robot interaction, promoting effective collaboration in transportation environments [14]. However, deploying these models in robotics raises safety concerns due to their vulnerability to adversarial inputs, highlighting the need for further research to mitigate risks [4].

MLLMs extend the capabilities of LLMs and VLMs by integrating diverse data modalities, which is crucial for addressing ontology matching challenges and ensuring data interoperability within ITS [9, 15]. Developing MLLMs that can seamlessly process and deliver content across various modalities is vital for achieving human-level AI, essential for advancing ITS [16].

Moreover, LLMs are critical in addressing ethical challenges associated with their deployment in ITS, including the need for improved benchmarks to evaluate their performance and limitations [2]. Their capabilities in enhancing semantic routing and intent-based networking within 5G core networks further emphasize their relevance to ITS, particularly in improving communication and operational efficiency [9]. Additionally, automating table processing tasks using LLMs and VLMs can significantly enhance data handling efficiency in transportation systems [17].

Collectively, LLMs, VLMs, and MLLMs provide innovative solutions to longstanding transportation challenges, underscoring their transformative potential in ITS. Their integration enhances operational efficiency and safety while accelerating the development of intelligent, adaptive transportation networks. The applicability of LLMs in improving clinical outcomes in dentistry suggests similar benefits for patient care and treatment efficiency in transportation systems [10].

1.3 Structure of the Survey

This survey is meticulously structured to provide a comprehensive examination of LLMs, VLMs, and MLLMs within Intelligent Transportation Systems (ITS). The paper begins with an that establishes the context by detailing the significance of LLMs in enhancing modern transportation systems, particularly through analyzing social media data for customer feedback, identifying emerging issues, and improving service quality. Leveraging advanced techniques such as Retrieval-Augmented Generation (RAG) and novel sentiment analysis methods, this section highlights the transformative potential of LLMs in fostering responsiveness and operational excellence within transit agencies [18, 19, 20, 21, 22]. It includes an **Overview of LLMs, VLMs, and MLLMs**, emphasizing their core functionalities and applications, followed by a discussion on their **Relevance to Intelligent Transportation Systems**.

The second section, **Background and Definitions**, offers foundational knowledge by defining key concepts such as LLMs, VLMs, MLLMs, ITS, and their related technologies. It also traces the **Evolution and Advancements** in these fields, underscoring their growing importance in transportation.

In the third section, **Large Language Models (LLMs) in Transportation**, the focus shifts to the role of LLMs in processing natural language data within transportation systems. This section delves into their contributions to **Decision-Making**, **Route Optimization and Traffic Prediction**, and **Enhancing Communication** between vehicles and infrastructure.

The fourth section, **Vision-Language Models (VLMs) and Their Impact**, explores how VLMs integrate visual and textual data to improve transportation systems. It covers their applications in **Autonomous Driving**, focusing on **Object Detection**, **Scene Understanding**, and enhancing **Reasoning and Spatial Capabilities**.

The fifth section, **Multimodal Large Language Models (MLLMs) in Intelligent Transportation Systems**, discusses the integration of various data types by MLLMs to enhance transportation efficiency and safety. It provides insights into **Multimodal Data Fusion**, **Dynamic and Adaptive Processing Techniques**, and **Safety and Efficiency Enhancements**.

In the sixth section, **Applications in Autonomous Vehicles**, the paper details the use of LLMs, VLMs, and MLLMs in autonomous vehicles, focusing on **Navigation**, **Obstacle Avoidance**, and **Human-Vehicle Interaction**.

The seventh section, **Traffic Management and Optimization**, analyzes the impact of these models on traffic management systems. It discusses their role in **Real-Time Traffic Monitoring**, **Congestion Prediction**, and **Multimodal Data Integration** for improved decision-making.

The eighth section, **Challenges and Future Directions**, identifies the current challenges in implementing these models in transportation systems. It explores **Challenges and Ethical Considerations**, **Future Research Directions**, and the importance of **Ethical and Interpretability Concerns**.

In the **Conclusion**, the survey synthesizes key findings regarding the transformative potential of LLMs, VLMs, and MLLMs in revolutionizing transportation systems. It underscores their capacity to enhance service quality through advanced data analysis and customer feedback mechanisms, improve safety and robustness in robotics applications, and facilitate effective communication in crisis scenarios. By leveraging these models, the future of intelligent transportation systems is positioned to be more responsive, efficient, and capable of addressing complex challenges in real-time [23, 24, 25, 20].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Definitions of Key Concepts

Large Language Models (LLMs) are advanced AI systems designed to understand, generate, and manipulate human language using extensive datasets, playing a crucial role in natural language processing and decision-making in complex scenarios such as emergency management [12]. These models can be categorized based on architecture and training methods, including task-dependent versus task-agnostic frameworks, as well as pretraining, transfer learning, and in-context learning [26]. Despite their capabilities, LLMs face challenges like producing factually incorrect outputs, known as hallucinations, which can undermine trust in AI systems.

Vision-Language Models (VLMs) enhance LLM capabilities by integrating visual data with textual information, enabling richer contextual understanding essential for tasks like visual reasoning and robotic manipulation [8]. These models are particularly beneficial in healthcare, improving practices through automated report generation and visual question answering (VQA) [9]. However, VLMs encounter difficulties in fine-grained concept identification and syntactic encoding, leading to potential mislocalized explanations and overconfident predictions.

Multimodal Large Language Models (MLLMs) represent a significant AI advancement, incorporating multiple data modalities to enhance input understanding and output generation. These models are critical in scenarios requiring comprehensive data fusion and processing across various modalities [9]. Their training demands substantial computational resources, posing ongoing challenges.

Intelligent Transportation Systems (ITS) use sophisticated applications to improve transportation efficiency, safety, and decision-making through real-time data analysis and predictive modeling. By leveraging LLMs, VLMs, and MLLMs, ITS optimizes transportation networks and manages complex traffic incidents [3]. The integration of these models addresses the challenge of aligning diverse ontologies to ensure semantic interoperability among various knowledge systems [15]. Developing benchmarks for evaluating model performance in resource-constrained environments remains essential [27].

Integrating language understanding with multimodal sensory inputs further enhances these models' abilities to navigate complex transportation scenarios, enabling semantic trajectory data mining and classification of Points of Interest (POIs) from GPS trajectory data [17]. Benchmarks simulating real-world network operations tackle the challenge of accurately extracting user intents in 5G core network management [10].

2.2 Evolution and Advancements

The development of Large Language Models (LLMs), Vision-Language Models (VLMs), and Multimodal Large Language Models (MLLMs) has been significantly influenced by advancements in computational power and large-scale datasets. Early LLMs struggled with complex tasks due to limited size and less sophisticated architectures [5]. The introduction of advanced pretraining objectives and transfer learning has expanded their capabilities, enabling applications such as analyzing complex patient data and assisting in clinical workflows [10]. These advancements have been pivotal in improving decision-making processes across various domains, including emergency management.

Despite these enhancements, challenges like hallucinations and biases persist, affecting LLM reliability. Developing robust benchmarks and optimization algorithms is critical to mitigating these issues and strengthening their theoretical foundations and practical applications [5].

VLMs have also progressed, achieving significant improvements in integrating visual and textual data for applications like image captioning and visual question answering. However, multimodal data fusion presents technical challenges, especially in healthcare where ethical considerations are paramount. Innovative approaches, such as memory-augmented techniques, have emerged to address the limitations of current large multimodal models, particularly regarding complex, long-duration tasks. For instance, the MA-LMM model uses an online processing method with a memory bank to store and reference historical video content, enhancing performance in video understanding tasks like long-video analysis, video question answering, and video captioning, thereby addressing challenges faced by existing LLM-based multimodal systems [9, 28].

MLLMs signify a crucial advancement in AI, characterized by their capacity to integrate diverse data modalities, enhancing both input comprehension and output generation. The evolution of these models underscores the need for comprehensive benchmarks incorporating situational awareness and quantitative abilities, which earlier models have not fully explored. Systems like Galois, facilitating structured data retrieval through SQL queries on LLMs, exemplify efforts to overcome traditional data limitations and enhance practical applications [5].

The continuous advancements in LLMs, VLMs, and MLLMs are increasingly vital in fostering innovation and enhancing operational efficiency across various sectors. These models are evolving from text-based frameworks to sophisticated multimodal systems, significantly impacting fields such as healthcare, where they integrate and analyze multiple data types—text, images, and audio—to support clinical decision-making and patient engagement. Their application in crisis scenarios further illustrates the potential of LLMs and MLLMs to improve machine translation capabilities for low-resource languages, highlighting the importance of community-driven efforts in creating specialized datasets. Collectively, these developments emphasize the integral role of intelligent systems in addressing complex challenges and driving progress across various domains [23, 24, 9].

3 Large Language Models (LLMs) in Transportation

Incorporating Large Language Models (LLMs) into transportation systems is increasingly pivotal due to the complexity of modern networks, which demand sophisticated decision-making tools. This section explores the multifaceted roles of LLMs in enhancing decision-making processes, contributing to operational efficiency and adaptability in transportation networks. Figure 2 illustrates these roles by categorizing LLM applications into several key areas: enhancing decision-making, emergency management, route optimization frameworks, user feedback analysis, central processing in communication, and data transformation. This figure emphasizes how LLMs facilitate improved traffic prediction and communication, ultimately highlighting their significant contributions to the overall efficiency and adaptability of transportation systems.

3.1 Role of LLMs in Decision-Making for Transportation

LLMs significantly enhance decision-making in transportation by automating workflows and optimizing resource allocation. For instance, the CALM framework employs cross-attention mechanisms to facilitate interactions between anchor and augmenting models [29]. LLMs also improve the safety of service robots in transportation through Embodied Robotic Control Prompts (ERCPs) and Embodied Knowledge Graphs (EKGs) [14]. The Memory-Augmented Large Multimodal Model (MA-LMM) is instrumental in dynamically referencing historical video data, crucial for informed decision-making [28], while the GameVLM framework enhances task planning and evaluation through multiple decision agents [30].

In emergency management, LLMs integrate structured emergency knowledge with reasoning capabilities applicable to transportation [12]. The Galois method combines traditional database processing with LLM reasoning to improve decision-making [3]. Furthermore, integrating optimization algorithms with LLMs addresses decision-making challenges in dynamic environments, enhancing responses to transportation issues [5]. The Automated Medical QA Evaluation using LLMs (AMQE-LLM) exemplifies their utility in evaluating responses based on predefined metrics, highlighting their potential in decision-making assessment [2].

The incorporation of LLMs into transportation decision-making processes fosters innovative, data-driven strategies essential for navigating modern transportation complexities. By leveraging advanced natural language processing and optimization techniques, LLMs automate tasks, enhance data analysis, and improve real-time decision-making, resulting in more responsive transportation solutions [31, 24, 23, 5, 32].

3.2 Applications in Route Optimization and Traffic Prediction

LLMs are crucial in optimizing routes and predicting traffic conditions. The ExpeL framework showcases this capability by learning from experiences without requiring parameter updates, providing an efficient alternative to traditional fine-tuning methods [33]. This adaptability is essential in environments with fluctuating traffic patterns. The REAL framework enhances mission planning

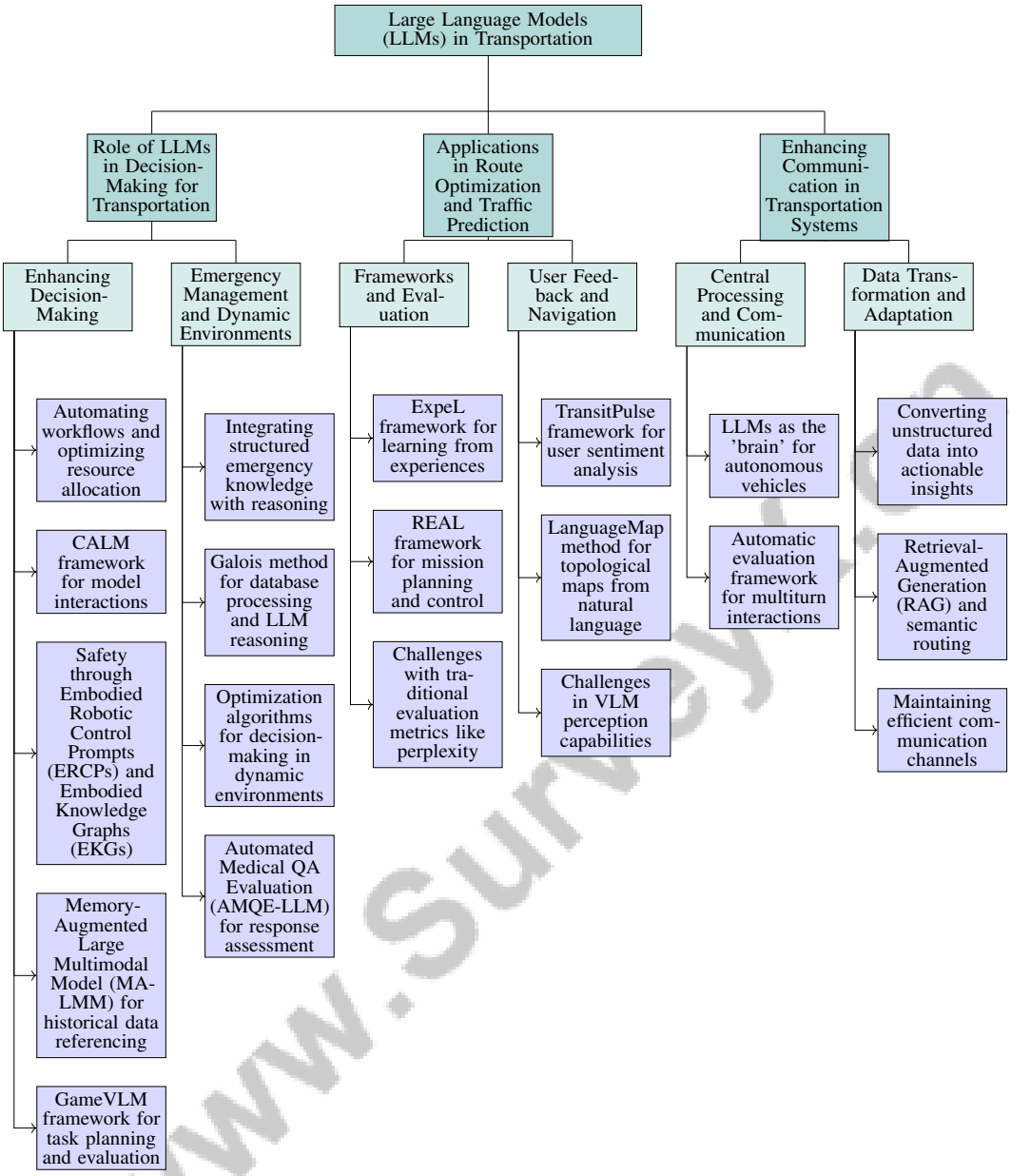

Figure 2: This figure illustrates the multifaceted roles of Large Language Models (LLMs) in transportation, focusing on decision-making, route optimization, traffic prediction, and enhancing communication. It categorizes LLM applications into enhancing decision-making, emergency management, route optimization frameworks, user feedback analysis, central processing in communication, and data transformation, highlighting their contributions to operational efficiency and adaptability in transportation systems.

and control systems, improving the resilience and adaptability of autonomous robots, applicable to real-time route planning and traffic flow optimization [34].

Despite their potential, traditional evaluation metrics like perplexity may not correlate with performance in downstream tasks such as route optimization and traffic prediction, as highlighted by Hu's study [35]. This calls for the development of more effective evaluation methods. LLMs also analyze user sentiment and identify issues in public transit systems, as demonstrated by the TransitPulse framework, which transforms user feedback into actionable insights for service enhancement and route optimization [20]. Additionally, the LanguageMap method's development of topological maps

from natural language path descriptions illustrates LLM utility in route optimization, streamlining navigation processes [36].

Challenges persist, particularly regarding LLM perception capabilities, essential for real-time driving decisions. Current Vision-Language Models (VLMs) face difficulties in analyzing sequential image frames and accurately identifying complex scenarios in autonomous driving, highlighting ongoing challenges in performance enhancement in dynamic environments [37, 38, 21, 6].

3.3 Enhancing Communication in Transportation Systems

LLMs enhance communication within transportation systems by acting as a central processing unit that integrates various data inputs, enabling seamless interactions between vehicles and infrastructure. As proposed by Cui et al., LLMs serve as the 'brain' of autonomous vehicles, processing environmental data through sensory tools for informed decision-making and efficient communication pathways [39]. This capability is vital for developing intelligent transportation networks reliant on real-time data exchange for optimized traffic flow and safety.

The automatic evaluation framework for multiturn interactions discussed by Liao et al. provides a benchmark for assessing LLM capabilities in complex communication scenarios within transportation systems [40]. This framework enhances understanding of LLMs' practical applications in facilitating effective communication.

LLMs improve communication by converting unstructured data into actionable insights, enabling vehicles to interpret and respond to changing traffic conditions. Techniques such as Retrieval-Augmented Generation (RAG) and semantic routing enhance this capability, allowing transit agencies to swiftly identify emerging issues and adapt to dynamic environments, ultimately fostering service excellence [23, 20, 32, 29]. This transformation is essential for maintaining efficient communication channels that support adaptive traffic management strategies and enhance the overall responsiveness of transportation systems.

4 Vision-Language Models (VLMs) and Their Impact

4.1 Integration of Visual and Textual Data

Vision-Language Models (VLMs) enhance the interpretative and decision-making capabilities of autonomous transportation systems by integrating visual and textual data. This multimodal approach provides a comprehensive understanding of complex environments, crucial for traffic scene analysis, object detection, and navigation, thereby improving efficiency and safety [8]. A cognitive architecture framework categorizes VLM components into visual processing, prior knowledge, and reasoning, optimizing configurations to minimize inference costs [8, 41]. GameVLM exemplifies VLMs' utility in robotic task planning within transportation contexts [30]. However, challenges remain in accurately interpreting responses due to limited training data incorporating 3D spatial knowledge, affecting spatial reasoning [42]. Benchmarks like MM-SY offer standardized evaluation methods for VLMs, assessing susceptibility to sycophantic behavior [43]. Integrating Large Language Models (LLMs) with VLMs, as in DVDet, optimizes class descriptors through evolutionary strategies, enhancing region-text alignment and open-vocabulary detection [44].

4.2 Applications of VLMs in Autonomous Driving

VLMs significantly enhance autonomous driving systems by integrating visual and textual data for object detection and scene understanding, crucial for safe and efficient vehicle operation. DVDet utilizes fine-grained descriptors to improve visual-textual alignment, enhancing open-vocabulary object detection [44]. The ScVLM framework combines supervised and contrastive learning to improve understanding of driving videos, enabling nuanced interpretations essential for real-time decision-making [45]. Mobility VLA demonstrates VLMs' effectiveness in executing complex navigation tasks in real-world environments [46]. VLM-HOI enhances Human-Object Interaction detection, improving vehicles' understanding and interaction with their environments [47]. Incorporating explicit reasoning processes into VLMs, as suggested by Uehara et al., further improves interpretability and reliability [48]. The MM-SY benchmark standardizes VLM evaluation across visual understanding tasks, providing insights into their effectiveness in autonomous driving [43].

MobileVLM V2's superior performance in accuracy and inference speed highlights the benefits of optimized training strategies [49].

4.3 Improving VLMs' Reasoning and Spatial Capabilities

Advancements in VLMs have enhanced their reasoning and spatial capabilities, crucial for transportation applications. Iterative optimization processes incorporating visual feedback enable more accurate class descriptor generation [50]. The ScVLM framework employs a hybrid approach for event classification and conflict identification, reducing hallucinations and improving reliability [45]. The Chain-of-Reasoning (CoR) method simulates human-like reasoning, enhancing accuracy [48]. VLM-HOI leverages robust visual and linguistic understanding for improved Human-Object Interaction predictions [47]. CVR-LLM integrates visual and textual reasoning, improving accuracy in complex tasks [41]. Integrating object detection into VLMs, as proposed by He, advances object recognition and scene understanding [51]. DVDet demonstrates leveraging VLMs' alignment capabilities for enhanced object detection [44]. These advancements are vital for developing intelligent transportation systems, enhancing safety, efficiency, and reliability. ScVLM improves comprehension of safety-critical events, while V2X-VLM integrates multimodal data for better situational awareness in cooperative autonomous driving [45, 52, 25, 51].

5 Multimodal Large Language Models (MLLMs) in Intelligent Transportation Systems

5.1 Multimodal Data Fusion in Complex Scenarios

Multimodal Large Language Models (MLLMs) are pivotal for decision-making in complex transportation scenarios through effective multimodal data fusion. The -UMi framework illustrates a modular architecture where specialized Large Language Models (LLMs) focus on distinct learning aspects, optimizing multimodal data processing [53]. GameVLM employs Vision-Language Models (VLMs) and zero-sum game theory to enhance decision-making in robotic task planning, showcasing the benefits of multimodal integration for operational efficiency [30]. Niu categorizes multimodal data fusion strategies into converters, perceivers, tools assistance, and data-driven approaches, providing a comprehensive overview [9].

Despite advancements, MLLMs can generate hallucinated responses, leading to inaccuracies [54]. Addressing this is crucial for reliable MLLM deployment in transportation. The Multi-Modal LLM AI System (MMLLM) integrates diverse data types, automating processes like dental diagnosis, highlighting MLLMs' versatility [10]. This integration is applicable to transportation, where multimodal data fusion enhances decision-making and efficiency.

MLLMs' ability to fuse multimodal data is essential for managing modern transportation complexities. Techniques like Retrieval-Augmented Generation (RAG) and dynamic input scaling enhance decision-making, enabling transit agencies to swiftly address issues via social media insights [20, 55]. This integration enhances safety, efficiency, and responsiveness in transportation infrastructures.

5.2 Dynamic and Adaptive Processing Techniques

Method Name	Operational Efficiency	Low-Latency Solutions	Spatial Reasoning
VisToG[56]	Optimize Computational Efficiency	Reduce Inference Time	-
LLM-OP[57]	Dynamic Processing Techniques	Efficient Model Deployment	Enhancing Graph Reasoning
SVLM[58]	-	-	Spatial Reasoning Capabilities

Table 1: Comparison of methods for enhancing operational efficiency, low-latency solutions, and spatial reasoning in multimodal large language models (MLLMs) for Intelligent Transportation Systems. The table outlines the specific contributions of VisToG, LLM-OP, and SVLM in optimizing computational efficiency, reducing inference time, and improving spatial reasoning capabilities.

Dynamic and adaptive processing techniques in MLLMs are crucial for operational efficiency in Intelligent Transportation Systems (ITS). Frameworks like VisToG optimize performance with fewer visual tokens, reducing computational resources and inference times [56]. This is vital for real-time decision-making in transportation scenarios requiring rapid adaptation. Table 1 presents a

comparative analysis of various methods employed to enhance the operational efficiency, low-latency solutions, and spatial reasoning in multimodal large language models, emphasizing their application in Intelligent Transportation Systems.

The deployment of these models faces challenges due to high computational and memory demands, necessitating low-latency solutions [24]. Techniques like LLM-based Online Prediction (LLM-OP) enhance decision-making by aggregating information from neighboring nodes [57]. Future research should focus on adaptive token processing and refining token compression algorithms to improve performance across applications [59]. These advancements are essential for managing computational demands and deployment costs on edge devices [60].

Systems like SpatialVLM, generating large-scale spatial VQA datasets, enhance VLMs' spatial reasoning capabilities, crucial for complex transportation tasks [58]. Such techniques ensure MLLMs effectively process and respond to dynamic transportation systems, contributing to intelligent and adaptive infrastructures.

5.3 Safety and Efficiency Enhancements

MLLM deployment in transportation systems significantly enhances safety and efficiency by integrating diverse data modalities for complex decision-making. The VisToG framework reduces inference time by 27

The Spider framework improves user experience by generating multiple modalities in a single response, reducing cognitive load [61]. This is beneficial in air traffic management, where the CHATATC system uses MLLMs to summarize complex data, enhancing situational awareness and reducing workload [62]. LLMs in semantic trajectory data mining improve classification of Points of Interest (POI) from incomplete datasets, enhancing transportation system robustness [63].

However, integrating LLMs into robotic systems introduces security risks, particularly in navigation tasks requiring precision [64]. Addressing these concerns is vital for safe MLLM deployment in transportation.

Strategic MLLM implementation enhances transportation safety and efficiency by optimizing data processing, improving situational awareness, and managing complex environments. Models like ScVLM enhance safety-critical event understanding in driving scenarios, reducing hallucinations and ensuring key safety feature recognition. Integrating MLLMs with emergency decision-making systems like E-KELL enables evidence-based responses in critical situations, crucial for effective emergency management. These advancements contribute to safer and more efficient transportation systems [45, 12, 25].

6 Applications in Autonomous Vehicles

To explore the multifaceted applications of autonomous vehicles, it is essential to examine key areas where technological advancements have been pivotal. A significant domain is navigation and decision-making, where the integration of advanced models has transformed operational capabilities. This section delves into methodologies and frameworks that enhance these aspects, providing insights into their implications for autonomous vehicle technology.

6.1 Navigation and Decision-Making

The integration of Vision-Language Models (VLMs) and Large Language Models (LLMs) is crucial for advancing navigation and decision-making in autonomous vehicles, utilizing the synergy between visual and textual data to optimize processes for safer navigation. The YOLOS detection network exemplifies enhanced decision-making in complex driving environments [51]. Processing multi-view images through advanced tokenization methods and aligning outputs with LLMs generate precise driving recommendations, significantly improving analysis capabilities in intricate driving environments through AI technologies [39, 65]. However, adversarial attacks on VLMs pose safety risks, necessitating robust frameworks to address vulnerabilities.

Overconfidence in model predictions, manifesting as high calibration errors with an average rate exceeding 21

The deployment of VLMs and LLMs significantly enhances navigation and decision-making by integrating multimodal data, improving safety, and facilitating personalized driving experiences. Recent advancements, particularly through LLMs and VLMs integration, enhance the intelligence and adaptability of autonomous systems, enabling vehicles to engage in human-like interactions. Models like WiseAD have shown substantial reductions in critical accidents, while tools like GPT-4V improve recognition of complex traffic situations [39, 66, 65, 67].

6.2 Obstacle Avoidance and Safety

Integrating LLMs and VLMs into autonomous vehicle systems significantly enhances obstacle avoidance and safety measures. Agyei et al.'s LLM-based decision-making framework generates navigation decisions compliant with International Regulations for Preventing Collisions at Sea (COLREGs), improving safety in complex scenarios through real-time risk assessments [68]. Yang et al.'s VLM-based pipeline effectively detects challenging cases in motion prediction, enhancing training efficiency via strategic data selection, crucial for autonomous driving reliability [69]. The WiseAD framework highlights knowledge-augmented trajectory planning's importance in enhancing decision-making and safety in autonomous vehicles [67].

The ScVLM framework addresses accurately identifying and understanding safety-critical events (SCEs) like crashes and near-crashes, essential for traffic safety and automated driving systems [45]. Recent advancements in LLMs and VLMs are crucial for improving obstacle avoidance and safety, enhancing multimodal sensory data integration for accurate real-time decision-making. However, deploying these models raises safety concerns due to vulnerabilities to adversarial inputs leading to critical failures. Research indicates minor input data alterations can degrade performance, underscoring the need for robust safety protocols [45, 25, 51, 66, 70].

6.3 Human-Vehicle Interaction

Human-vehicle interaction is significantly enhanced by integrating advanced models facilitating seamless communication. The WiseAD framework incorporates fundamental driving knowledge into decision-making and trajectory planning, enabling knowledge-aligned decisions surpassing existing methods [67]. This integration allows autonomous vehicles to interpret and respond to human inputs effectively, enhancing the overall driving experience.

Nwankwo et al.'s dual-modality framework integrates vocal and textual commands, allowing users to communicate with autonomous systems more naturally [71]. This framework improves user experience and enhances autonomous vehicles' responsiveness to human commands.

Advanced multimodal models enhance human-vehicle interaction by integrating diverse data types, such as natural language and visual inputs, allowing for accurate interpretation and responsive actions. Employing techniques from LLMs and deep learning, these systems analyze complex user preferences and contextual information, leading to personalized and effective interactions [72, 48, 73, 38]. Enhancing communication pathways between humans and autonomous systems contributes to developing more intelligent and user-friendly transportation systems.

7 Traffic Management and Optimization

7.1 Enhancing Traffic Management with LLMs

Large Language Models (LLMs) revolutionize traffic management by employing advanced natural language processing to interpret vast unstructured datasets, thereby enhancing real-time data analysis and predictive modeling. This integration optimizes traffic flow and mitigates congestion. For instance, LLMs in semantic trajectory data mining classify Points of Interest (POI) from incomplete datasets, bolstering traffic management system robustness [63]. They also automate user feedback and social media data interpretation, translating insights into actionable strategies for public transit improvements, which is crucial for aligning with technology acceptance models [20, 11]. Processing natural language data allows LLMs to predict congestion patterns and optimize traffic signals, reducing delays.

LLMs address ontology alignment challenges, ensuring semantic interoperability among diverse knowledge systems, which is vital for integrating varied data sources and improving decision-making

in traffic management [15]. Establishing benchmarks for model performance in resource-constrained environments is also critical [27]. Combining LLMs with optimization algorithms facilitates adaptive traffic management strategies, enhancing safety and efficiency [5]. Their role in semantic routing and intent-based networking within 5G core networks underscores their significance in optimizing communication and operational efficiency [9].

Integrating LLMs into traffic management systems enhances navigation in complex environments, ensuring safer, more efficient traffic flow. By leveraging advanced processing capabilities, LLMs are pivotal in developing intelligent, adaptive traffic management systems that employ semantic routing and intent-based networking to optimize urban mobility and alleviate congestion [31, 24, 32, 29].

7.2 Real-Time Traffic Monitoring and Congestion Prediction

The application of LLMs in real-time traffic monitoring and congestion prediction marks a significant advancement in intelligent transportation systems. These models employ natural language processing to analyze extensive data from diverse sources, offering real-time traffic insights and enabling proactive congestion management strategies. Continuous traffic pattern monitoring facilitated by LLMs allows networks to swiftly adapt to changing conditions [63].

LLMs excel in processing unstructured data, such as social media posts and sensor data, to develop a comprehensive understanding of current traffic conditions, essential for predicting congestion and implementing timely interventions [20]. By converting qualitative data into quantitative insights, LLMs enhance decision-making processes within traffic management systems, enabling more accurate traffic flow predictions [15].

Moreover, LLMs integrate multimodal data sources, ensuring real-time traffic information remains accurate and actionable. Their ability to align various ontologies and maintain semantic interoperability among data streams is critical for reliable traffic predictions and optimized management strategies [5]. LLMs also employ advanced machine learning algorithms to identify patterns in historical traffic data, allowing authorities to anticipate congestion events and implement preemptive measures, improving traffic flow and reducing travel times.

Integrating LLMs into real-time monitoring and congestion prediction systems enhances transportation networks' responsiveness to dynamic conditions, optimizing traffic flow. By leveraging capabilities like sentiment detection from social media, LLMs provide transit agencies with actionable insights that surpass traditional analysis methods. Techniques like Retrieval-Augmented Generation (RAG) further enhance information extraction and system performance, making LLMs essential for modern traffic management [20, 32].

7.3 Multimodal Data Integration for Improved Decision-Making

Multimodal data integration is vital for enhancing decision-making in traffic management systems. By utilizing Multimodal Large Language Models (MLLMs), transportation networks can efficiently process and analyze diverse data types—including textual, visual, and sensor data—yielding comprehensive insights into traffic conditions. These models enable transit agencies to extract valuable information from social media, detect emerging issues, and enhance service quality, thereby improving system responsiveness and decision-making capabilities [20, 9, 55].

MLLMs excel in synthesizing and interpreting information from various sources, empowering traffic management systems to analyze complex scenarios and implement adaptive strategies effectively [74, 75, 23, 55, 9]. Combining visual data from traffic cameras with textual data from social media and sensor readings provides a holistic understanding of traffic dynamics, enhancing congestion prediction accuracy and intervention strategy effectiveness.

The ability of MLLMs to align and integrate diverse data modalities is essential for maintaining semantic interoperability among data streams, ensuring traffic management decisions are based on consistent, reliable information. This capability is particularly advantageous in urban environments with intricate, fluctuating traffic patterns necessitating sophisticated, adaptable decision-making frameworks. Advanced models like GPT-4V and frameworks such as CityLLaVA enhance understanding and prediction of complex traffic scenarios, enabling more effective real-time responses and improving overall traffic safety and management [76, 20, 65].

Integrating multimodal data, including visual, textual, and social media inputs, enhances predictive model development, allowing accurate traffic congestion forecasting and traffic signal timing optimization through large vision-language models (VLMs) like GPT-4V. This comprehensive approach provides a nuanced understanding of complex traffic situations, enabling transit agencies to respond effectively to emerging issues and improve overall traffic management [38, 65, 72, 20, 9]. By analyzing historical and real-time data, MLLMs empower transportation authorities to implement proactive measures that mitigate congestion and enhance overall traffic efficiency.

The integration of multimodal data through MLLMs significantly improves decision-making capabilities in traffic management by enabling the synthesis of diverse information types—such as text, images, and audio—providing comprehensive insights and enhancing the accuracy of traffic-related analyses [54, 9, 73, 55]. By delivering a comprehensive understanding of traffic conditions and enabling adaptive responses to dynamic environments, MLLMs contribute to developing intelligent and responsive transportation infrastructures, ensuring safer and more efficient urban mobility.

8 Challenges and Future Directions

8.1 Challenges and Ethical Considerations

Integrating Large Language Models (LLMs), Vision-Language Models (VLMs), and Multimodal Large Language Models (MLLMs) into Intelligent Transportation Systems (ITS) presents several challenges and ethical dilemmas. A primary challenge is the significant computational resources required, which necessitate extensive datasets and high-performance computing, potentially restricting access for resource-limited researchers and organizations, thus impacting reproducibility and inclusivity [5, 26]. Additionally, processing complex queries and ensuring reliable LLM outputs remain challenging due to difficulties in handling intricate data formats and schemas, compounded by MLLMs' reliance on language biases over image content [54]. VLMs face specific challenges in reasoning and map-based question answering, with existing benchmarks like LLaVA-1.5 requiring broader validation for robustness [43]. The quality of fine-grained descriptors from LLM interactions also critically affects model effectiveness, underscoring the need for high-quality input data [44].

Ethical considerations are paramount, particularly regarding data privacy, bias in training datasets, and model output interpretability. Ensuring data privacy and addressing biases are crucial for public trust and fair, transparent model operations [10]. The interpretability and robustness of optimization algorithms further necessitate comprehensive ethical frameworks that advocate for accountability and inclusivity [5].

8.2 Future Directions

Future research on LLMs, VLMs, and MLLMs in ITS is poised to advance across several dimensions. Enhancing LLM efficiency through unsupervised learning and addressing related ethical concerns are critical for improving adaptability in dynamic transportation environments [26]. For MLLMs, refining hallucination benchmarks and exploring reinforcement learning (RL)-based strategies for model alignment are vital, alongside improving dataset quality and diversity to enhance performance across transportation applications [54]. The integration of LLMs with optimization algorithms offers promising opportunities for improved data utilization and addressing robustness and interpretability challenges [5].

In the VLM domain, exploring synergies between LLMs and VLMs for open vocabulary dense prediction tasks presents rich innovation opportunities. Enhancing descriptor generation will further improve model alignment and applicability in transportation contexts [44]. Additionally, refining domain-specific knowledge integration and enhancing data privacy measures, especially in clinical transportation applications, are crucial for ethical compliance and model reliability [10].

Future research should focus on enhancing LLM training techniques, exploring new prompting strategies, and developing diverse datasets reflecting real-world transportation scenarios [17]. Addressing these areas will ensure that LLMs, VLMs, and MLLMs evolve to meet the dynamic demands of modern transportation systems, contributing to more intelligent, efficient, and safe infrastructures.

8.3 Ethical and Interpretability Concerns

The deployment of LLMs, VLMs, and MLLMs in ITS raises significant ethical and interpretability challenges that must be addressed for responsible integration. A major concern is the potential for LLMs to generate incorrect or biased information, underscoring the need for model interpretability [77]. This complexity can lead to insufficient stakeholder understanding and potential misapplication [19].

To address these issues, dynamic auditing systems and tailored ethical frameworks that adapt to evolving LLM capabilities are urgently needed [78]. These frameworks should involve interdisciplinary collaboration to tackle the unique ethical challenges posed by these technologies. Regulatory frameworks are also necessary to balance LLM productivity benefits with potential labor market disruptions, ensuring ethical and sustainable ITS deployment [79].

Interpretability is crucial for building trust and ensuring stakeholders can comprehend and verify LLM outputs. Enhancing transparency and accountability will help address critical ethical challenges, such as privacy, bias, and misinformation, arising from LLM deployment in ITS. This proactive approach fosters understanding among diverse stakeholders and ensures responsible LLM integration into transportation networks, promoting innovation while mitigating associated risks. Strategies like model reporting, evaluation publishing, and uncertainty communication can significantly enhance the ethical and effective application of LLMs in this context [19, 24, 78, 80].

9 Conclusion

The exploration of Large Language Models (LLMs), Vision-Language Models (VLMs), and Multimodal Large Language Models (MLLMs) within Intelligent Transportation Systems (ITS) reveals their pivotal role in advancing the field. These models significantly enhance the functionality of transportation networks by improving efficiency, safety, and adaptability. LLMs excel in automating complex decision-making processes and optimizing resource distribution, thereby elevating operational performance across diverse transportation scenarios. VLMs contribute by integrating visual and textual information, which is vital for tasks like object detection and scene comprehension, thus boosting the safety and reliability of autonomous systems.

The introduction of MLLMs into ITS underscores their transformative impact, as they adeptly combine various data modalities to elevate transportation efficiency and security. This multimodal integration is crucial for navigating the intricacies of contemporary transportation systems, fostering the development of more intelligent and responsive networks. The CoDrivingLLM framework exemplifies the advancements in interaction and learning for connected and autonomous vehicles (CAVs), surpassing traditional methods in terms of safety, efficiency, and adaptability.

Additionally, LLMs' influence extends to transforming business models through enhanced automation and optimized information resource management, underscoring their significance in transportation systems. Their application in regulatory analysis and compliance highlights their potential to decrease manual effort and increase precision, essential for building intelligent transportation infrastructures. Furthermore, LLMs' role in spectrum regulation illustrates their ability to streamline administrative tasks and improve decision-making, crucial for the effective governance of transportation networks.

References

[1] Bosheng Ding, Chengwei Qin, Ruochen Zhao, Tianze Luo, Xinze Li, Guizhen Chen, Wenhan Xia, Junjie Hu, Anh Tuan Luu, and Shafiq Joty. Data augmentation using large language models: Data perspectives, learning paradigms and challenges, 2024.

[2] Jack Krolik, Herprit Mahal, Feroz Ahmad, Gaurav Trivedi, and Bahador Saket. Towards leveraging large language models for automated medical qa evaluation, 2024.

[3] Mohammed Saeed, Nicola De Cao, and Paolo Papotti. Querying large language models with sql, 2023.

[4] Ernests Lavrinovics, Russa Biswas, Johannes Bjerva, and Katja Hose. Knowledge graphs, large language models, and hallucinations: An nlp perspective, 2024.

[5] Sen Huang, Kaixiang Yang, Sheng Qi, and Rui Wang. When large language model meets optimization, 2024.

[6] Akash Ghosh, Arkadeep Acharya, Sriparna Saha, Vinija Jain, and Aman Chadha. Exploring the frontier of vision-language models: A survey of current methodologies and future directions, 2024.

[7] Iryna Hartsock and Ghulam Rasool. Vision-language models for medical report generation and visual question answering: A review, 2024.

[8] Allison Chen, Ilia Sucholutsky, Olga Russakovsky, and Thomas L. Griffiths. Analyzing the roles of language and vision in learning from limited data, 2024.

[9] Qian Niu, Keyu Chen, Ming Li, Pohsun Feng, Ziqian Bi, Lawrence KQ Yan, Yichao Zhang, Caitlyn Heqi Yin, Cheng Fei, Junyu Liu, Benji Peng, Tianyang Wang, Yunze Wang, Silin Chen, and Ming Liu. From text to multimodality: Exploring the evolution and impact of large language models in medical practice, 2024.

[10] Hanyao Huang, Ou Zheng, Dongdong Wang, Jiayi Yin, Zijin Wang, Shengxuan Ding, Heng Yin, Chuan Xu, Renjie Yang, Qian Zheng, and Bing Shi. Chatgpt for shaping the future of dentistry: The potential of multi-modal large language model, 2023.

[11] Alejandro Tlaie. Exploring and steering the moral compass of large language models, 2024.

[12] Minze Chen, Zhenxiang Tao, Weitong Tang, Tingxin Qin, Rui Yang, and Chunli Zhu. Enhancing emergency decision-making with knowledge graphs and large language models, 2023.

[13] Xu Huang, Weiwen Liu, Xiaolong Chen, Xingmei Wang, Hao Wang, Defu Lian, Yasheng Wang, Ruiming Tang, and Enhong Chen. Understanding the planning of llm agents: A survey. *arXiv preprint arXiv:2402.02716*, 2024.

[14] Yong Qi, Gabriel Kyebambo, Siyuan Xie, Wei Shen, Shenghui Wang, Bitao Xie, Bin He, Zhipeng Wang, and Shuo Jiang. Safety control of service robots with llms and embodied knowledge graphs, 2024.

[15] Rudolf Laine, Bilal Chughtai, Jan Betley, Kaivalya Hariharan, Jeremy Scheurer, Mikita Balesni, Marius Hobbhahn, Alexander Meinke, and Owain Evans. Me, myself, and ai: The situational awareness dataset (sad) for llms, 2024.

[16] Swapnaja Achintalwar, Adriana Alvarado Garcia, Ateret Anaby-Tavor, Ioana Baldini, Sara E. Berger, Bishwaranjan Bhattacharjee, Djallel Bouneffouf, Subhajit Chaudhury, Pin-Yu Chen, Lamogha Chiazor, Elizabeth M. Daly, Kirushikesh DB, Rogério Abreu de Paula, Pierre Dognin, Eitan Farchi, Soumya Ghosh, Michael Hind, Raya Horesh, George Kour, Ja Young Lee, Nishtha Madaan, Sameep Mehta, Erik Miehling, Keerthiram Murugesan, Manish Nagireddy, Inkit Padhi, David Piorkowski, Ambrish Rawat, Orna Raz, Prasanna Sattigeri, Hendrik Strobelt, Sarathkrishna Swaminathan, Christoph Tillmann, Aashka Trivedi, Kush R. Varshney, Dennis Wei, Shalisha Witherspooon, and Marcel Zalmanovici. Detectors for safe and reliable llms: Implementations, uses, and limitations, 2024.

[17] Weizheng Lu, Jing Zhang, Ju Fan, Zihao Fu, Yueguo Chen, and Xiaoyong Du. Large language model for table processing: A survey, 2024.

[18] Dmitry Scherbakov, Nina Hubig, Vinita Jansari, Alexander Bakumenko, and Leslie A. Lenert. The emergence of large language models (llm) as a tool in literature reviews: an llm automated systematic review, 2024.

[19] Q. Vera Liao and Jennifer Wortman Vaughan. Ai transparency in the age of llms: A human-centered research roadmap, 2023.

[20] Jiahao Wang and Amer Shalaby. Transit pulse: Utilizing social media as a source for customer feedback and information extraction with large language model, 2024.

[21] Hugo Laurençon, Andrés Marafioti, Victor Sanh, and Léo Tronchon. Building and better understanding vision-language models: insights and future directions, 2024.

[22] Banghao Chen, Zhaofeng Zhang, Nicolas Langrené, and Shengxin Zhu. Unleashing the potential of prompt engineering in large language models: a comprehensive review, 2024.

[23] Séamus Lankford and Andy Way. Leveraging llms for mt in crisis scenarios: a blueprint for low-resource languages, 2024.

[24] Baolin Li, Yankai Jiang, Vijay Gadepally, and Devesh Tiwari. Llm inference serving: Survey of recent advances and opportunities, 2024.

[25] Xiyang Wu, Souradip Chakraborty, Ruiqi Xian, Jing Liang, Tianrui Guan, Fuxiao Liu, Brian M. Sadler, Dinesh Manocha, and Amrit Singh Bedi. Highlighting the safety concerns of deploying llms/vlms in robotics, 2024.

[26] Rajvardhan Patil and Venkat Gudivada. A review of current trends, techniques, and challenges in large language models (llms). *Applied Sciences*, 14(5):2074, 2024.

[27] Yannis Bendi-Ouis, Dan Dutartre, and Xavier Hinaut. Deploying open-source large language models: A performance analysis, 2025.

[28] Bo He, Hengduo Li, Young Kyun Jang, Menglin Jia, Xuefei Cao, Ashish Shah, Abhinav Shrivastava, and Ser-Nam Lim. Ma-lmm: Memory-augmented large multimodal model for long-term video understanding, 2024.

[29] Rachit Bansal, Bidisha Samanta, Siddharth Dalmia, Nitish Gupta, Shikhar Vashishth, Sriram Ganapathy, Abhishek Bapna, Prateek Jain, and Partha Talukdar. Llm augmented llms: Expanding capabilities through composition. *arXiv preprint arXiv:2401.02412*, 2024.

[30] Aoran Mei, Jianhua Wang, Guo-Niu Zhu, and Zhongxue Gan. Gamevlm: A decision-making framework for robotic task planning based on visual language models and zero-sum games, 2024.

[31] Hao Zhou, Chengming Hu, Ye Yuan, Yufei Cui, Yili Jin, Can Chen, Haolun Wu, Dun Yuan, Li Jiang, Di Wu, Xue Liu, Charlie Zhang, Xianbin Wang, and Jiangchuan Liu. Large language model (llm) for telecommunications: A comprehensive survey on principles, key techniques, and opportunities, 2024.

[32] Dimitrios Michael Manias, Ali Chouman, and Abdallah Shami. Semantic routing for enhanced performance of llm-assisted intent-based 5g core network management and orchestration, 2024.

[33] Andrew Zhao, Daniel Huang, Quentin Xu, Matthieu Lin, Yong-Jin Liu, and Gao Huang. Expel: Llm agents are experiential learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19632–19642, 2024.

[34] Andrea Tagliabue, Kota Kondo, Tong Zhao, Mason Peterson, Claudius T. Tewari, and Jonathan P. How. Real: Resilience and adaptation using large language models on autonomous aerial robots, 2023.

[35] Yutong Hu, Quzhe Huang, Mingxu Tao, Chen Zhang, and Yansong Feng. Can perplexity reflect large language model's ability in long text understanding?, 2024.

[36] Hideki Deguchi, Kazuki Shibata, and Shun Taguchi. Language to map: Topological map generation from natural language path instructions, 2024.

[37] Li Liu, Diji Yang, Sijia Zhong, Kalyana Suma Sree Tholeti, Lei Ding, Yi Zhang, and Leilani H. Gilpin. Right this way: Can vlms guide us to see more to answer questions?, 2024.

[38] Jesse Atuhurra, Iqra Ali, Tatsuya Hiraoka, Hidetaka Kamigaito, Tomoya Iwakura, and Taro Watanabe. Constructing multilingual visual-text datasets revealing visual multilingual ability of vision language models, 2024.

[39] Can Cui, Yunsheng Ma, Xu Cao, Wenqian Ye, and Ziran Wang. Drive as you speak: Enabling human-like interaction with large language models in autonomous vehicles, 2023.

[40] Yusheng Liao, Yutong Meng, Hongcheng Liu, Yanfeng Wang, and Yu Wang. An automatic evaluation framework for multi-turn medical consultations capabilities of large language models, 2023.

[41] Zhiyuan Li, Dongnan Liu, Chaoyi Zhang, Heng Wang, Tengfei Xue, and Weidong Cai. Enhancing advanced visual reasoning ability of large language models, 2024.

[42] Srija Mukhopadhyay, Abhishek Rajgaria, Prerana Khatiwada, Vivek Gupta, and Dan Roth. Mapwise: Evaluating vision-language models for advanced map queries, 2024.

[43] Shuo Li, Tao Ji, Xiaoran Fan, Linsheng Lu, Leyi Yang, Yuming Yang, Zhiheng Xi, Rui Zheng, Yuran Wang, Xiaohui Zhao, Tao Gui, Qi Zhang, and Xuanjing Huang. Have the vlms lost confidence? a study of sycophancy in vlms, 2024.

[44] Sheng Jin, Xueying Jiang, Jiaxing Huang, Lewei Lu, and Shijian Lu. Llms meet vlms: Boost open vocabulary object detection with fine-grained descriptors, 2024.

[45] Liang Shi, Boyu Jiang, Tong Zeng, and Feng Guo. Scvlm: Enhancing vision-language model for safety-critical event understanding, 2025.

[46] Mobility vla: Multimodal instruc.

[47] Donggoo Kang, Dasol Jeong, Hyunmin Lee, Sangwoo Park, Hasil Park, Sunkyu Kwon, Yeongjoon Kim, and Joonki Paik. Vlm-hoi: Vision language models for interpretable human-object interaction analysis, 2024.

[48] Kohei Uehara, Nabarun Goswami, Hanqin Wang, Toshiaki Baba, Kohtaro Tanaka, Tomohiro Hashimoto, Kai Wang, Rei Ito, Takagi Naoya, Ryo Umagami, Yingyi Wen, Tanachai Anakewat, and Tatsuya Harada. Advancing large multi-modal models with explicit chain-of-reasoning and visual question generation, 2024.

[49] Xiangxiang Chu, Limeng Qiao, Xinyu Zhang, Shuang Xu, Fei Wei, Yang Yang, Xiaofei Sun, Yiming Hu, Xinyang Lin, Bo Zhang, and Chunhua Shen. Mobilevlm v2: Faster and stronger baseline for vision language model, 2024.

[50] Songhao Han, Le Zhuo, Yue Liao, and Si Liu. Llms as visual explainers: Advancing image classification with evolving visual descriptions, 2024.

[51] Linfeng He, Yiming Sun, Sihao Wu, Jiaxu Liu, and Xiaowei Huang. Integrating object detection modality into visual language model for enhanced autonomous driving agent, 2024.

[52] Junwei You, Haotian Shi, Zhuoyu Jiang, Zilin Huang, Rui Gan, Keshu Wu, Xi Cheng, Xiaopeng Li, and Bin Ran. V2x-vlm: End-to-end v2x cooperative autonomous driving through large vision-language models, 2024.

[53] Weizhou Shen, Chenliang Li, Hongzhan Chen, Ming Yan, Xiaojun Quan, Hehong Chen, Ji Zhang, and Fei Huang. Small llms are weak tool learners: A multi-llm agent. *arXiv preprint arXiv:2401.07324*, 2024.

[54] Elmira Amirloo, Jean-Philippe Fauconnier, Christoph Roesmann, Christian Kerl, Rinu Boney, Yusu Qian, Zirui Wang, Afshin Dehghan, Yinfei Yang, Zhe Gan, and Peter Grasch. Understanding alignment in multimodal llms: A comprehensive study, 2024.

[55] Yonghui Wang, Wengang Zhou, Hao Feng, and Houqiang Li. Adaptvision: Dynamic input scaling in mllms for versatile scene understanding, 2024.

[56] Minbin Huang, Runhui Huang, Han Shi, Yimeng Chen, Chuanyang Zheng, Xiangguo Sun, Xin Jiang, Zhenguo Li, and Hong Cheng. Efficient multi-modal large language models via visual token grouping, 2024.

[57] Dayu Qin, Yi Yan, and Ercan Engin Kuruoglu. Llm-based online prediction of time-varying graph signals, 2024.

[58] Boyuan Chen, Zhuo Xu, Sean Kirmani, Brian Ichter, Danny Driess, Pete Florence, Dorsa Sadigh, Leonidas Guibas, and Fei Xia. Spatialvlm: Endowing vision-language models with spatial reasoning capabilities, 2024.

[59] Kevin Y Li, Sachin Goyal, Joao D Semedo, and J Zico Kolter. Inference optimal vlms need only one visual token but larger models. *arXiv preprint arXiv:2411.03312*, 2024.

[60] Yizhang Jin, Jian Li, Yexin Liu, Tianjun Gu, Kai Wu, Zhengkai Jiang, Muyang He, Bo Zhao, Xin Tan, Zhenye Gan, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. Efficient multimodal large language models: A survey, 2024.

[61] Jinxiang Lai, Jie Zhang, Jun Liu, Jian Li, Xiaocheng Lu, and Song Guo. Spider: Any-to-many multimodal llm, 2024.

[62] Sinan Abdulhak, Wayne Hubbard, Karthik Gopalakrishnan, and Max Z. Li. Chatatc: Large language model-driven conversational agents for supporting strategic air traffic flow management, 2024.

[63] Yifan Liu, Chenchen Kuai, Haoxuan Ma, Xishun Liao, Brian Yueshuai He, and Jiaqi Ma. Semantic trajectory data mining with llm-informed poi classification, 2024.

[64] Wenxiao Zhang, Xiangrui Kong, Conan Dewitt, Thomas Braunl, and Jin B. Hong. A study on prompt injection attack against llm-integrated mobile robotic systems, 2024.

[65] Xingcheng Zhou and Alois C. Knoll. Gpt-4v as traffic assistant: An in-depth look at vision language model on complex traffic events, 2024.

[66] Sonda Fourati, Wael Jaafar, Noura Baccar, and Safwan Alfattani. Xlm for autonomous driving systems: A comprehensive review, 2024.

[67] Songyan Zhang, Wenhui Huang, Zihui Gao, Hao Chen, and Chen Lv. Wisead: Knowledge augmented end-to-end autonomous driving with vision-language model, 2024.

[68] Klinsmann Agyei, Pouria Sarhadi, and Wasif Naeem. Large language model-based decision-making for colregs and the control of autonomous surface vehicles, 2024.

[69] Yi Yang, Qingwen Zhang, Kei Ikemura, Nazre Batool, and John Folkesson. Hard cases detection in motion prediction by vision-language foundation models, 2024.

[70] Xiyang Wu, Ruiqi Xian, Tianrui Guan, Jing Liang, Souradip Chakraborty, Fuxiao Liu, Brian M Sadler, Dinesh Manocha, and Amrit Bedi. On the safety concerns of deploying llms/vlms in robotics: Highlighting the risks and vulnerabilities. In *First Vision and Language for Autonomous Driving and Robotics Workshop*, 2024.

[71] Linus Nwankwo and Elmar Rueckert. Multimodal human-autonomous agents interaction using pre-trained language and visual foundation models, 2024.

[72] Rajat Chawla, Arkajit Datta, Tushar Verma, Adarsh Jha, Anmol Gautam, Ayush Vatsal, Sukrit Chaterjee, Mukunda NS, and Ishaan Bhola. Veagle: Advancements in multimodal representation learning, 2024.

[73] Jiahao Tian, Jinman Zhao, Zhenkai Wang, and Zhicheng Ding. Mmrec: Llm based multi-modal recommender system, 2024.

[74] Sheng Jin, Xueying Jiang, Jiaxing Huang, Lewei Lu, and Shijian Lu. Llms meet vlms: Boost open vocabulary object detection with fine-grained descriptors. *arXiv preprint arXiv:2402.04630*, 2024.

[75] Tuan Bui, Oanh Tran, Phuong Nguyen, Bao Ho, Long Nguyen, Thang Bui, and Tho Quan. Cross-data knowledge graph construction for llm-enabled educational question-answering system: A case study at hcmut, 2024.

[76] Zhizhao Duan, Hao Cheng, Duo Xu, Xi Wu, Xiangxie Zhang, Xi Ye, and Zhen Xie. Cityllava: Efficient fine-tuning for vlms in city scenario, 2024.

[77] Yufan Chen, Arjun Arunasalam, and Z. Berkay Celik. Can large language models provide security privacy advice? measuring the ability of llms to refute misconceptions, 2023.

[78] Junfeng Jiao, Saleh Afroogh, Yiming Xu, and Connor Phillips. Navigating llm ethics: Advancements, challenges, and future directions, 2024.

[79] Qin Chen, Jinfeng Ge, Huaqing Xie, Xingcheng Xu, and Yanqing Yang. Large language models at work in china's labor market, 2023.

[80] Shabnam Hassani. Enhancing legal compliance and regulation analysis with large language models, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.