Citation: LI Jian, ZHU Guojun, WANG Ao, XIA Qiang, ZHOU Xujun, LI Yizhe. Knowledge-Enhanced Large Language Models for Urban Transportation: Modeling and Applications[J], Urban Transport of China, 2025 (2).

Knowledge-Enhanced Large Language Models for Urban Transportation: Modeling and Applications

LI Jian[1,2], ZHU Guojun[1], WANG Ao[1], XIA Qiang[1,2], ZHOU Xujun[1], LI Yizhe[1]

1. Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China;
2. Urban Mobility Institute, Tongji University, Shanghai 201804, China

Abstract: Large language models (LLMs) have become a hot topic of discussion because of their powerful semantic understanding and generation ability. Although the large language models perform well in dealing with general knowledge-based questions and answers, there is still a widespread phenomenon of "hallucinations" in industries involving complex decision-making, and problems such as interpretability and credibility are prominent. On the basis of combining current domestic and abroad research, this paper puts forward a knowledge-enhanced system architecture for large language models in urban transportation starting from the perspective of integrating knowledge graphs with large language models. Furthermore, the paper explores the technologies of prompt word engineering, retrieval enhancement generation, model integration, and agent construction. A knowledge-enhanced LLM for urban transportation (TransKG-LLM) is developed. Practical explorations are conducted from four dimensions: data enhancement, knowledge enhancement, model enhancement, and task enhancement. The results indicate that the proposed model can alleviate the "hallucinations" phenomenon of the general large language model, and help to improve the scientific, refined, and intelligent level of urban transportation management ability.**DOI:** 10.13813/j.cn11-5141/u.2024.0046-en

Keywords: urban transportation; generative artificial intelligence; knowledge-augmented generation; knowledge graph; large language model

0 Introduction

Urban transportation constitutes an important habitat for new technologies and emerging industries. It is a typical knowledge- and technology-intensive sector that encompasses multiple disciplines, including transportation, urban and rural planning, sociology, economics, and management [1], with diverse and highly differentiated application scenarios. Consequently, practitioners' cognitive capacity to address complex problems and their ability to master new technologies are crucial for cultivating new productive forces. At present, China's urban transportation sector faces challenges such as increasing complexity, the need for rapid decision-making responses, and heightened demands for scientific decision-making [2]. These challenges impose higher requirements on practitioners' overall competence and underscore the urgent need for innovative theoretical and methodological systems to enhance the sector's new productive forces.

Large Language Models (LLMs), as a major branch of Generative Artificial Intelligence (GAI), represent a key technology for advancing new productive forces. Since the launch of ChatGPT by OpenAI at the end of 2022, LLMs have attracted widespread attention from both academia and industry worldwide [3]. Numerous institutions, both domestic and international, have successively introduced their own LLMs [4], such as Google's Gemini [5], Facebook's LLaMA [6], Mistral AI's Mistral-7B [7], OpenAI's text-to-video model Sora [8], and more recently, OpenAI's O1, which demonstrates advanced reasoning capabilities for complex tasks [9]. In China, Baidu's ERNIE Bot, Alibaba's Tongyi Qianwen, and Zhipu AI's ChatGLM are advancing in parallel, while the release of DeepSeek-R1 by DeepSeek in early 2025 has drawn particular attention.

Despite their strong performance in general-purpose question answering, LLMs still face prominent limitations in industry domains requiring complex decision-making. These include hallucinations, as well as deficiencies in trustworthiness, interpretability, and controllability. Hallucinations refer to the phenomenon whereby LLMs generate inaccurate or fabricated information, often due to incomplete domain knowledge coverage, biases in training data, the absence of fact-verification mechanisms, or mismatches in reasoning processes. Moreover, the enormous parameter scale of LLMs, coupled with the lack of explicit decision pathways, hampers interpretability and controllability. As a result, tracing the reasoning logic behind generated content is difficult, which, in application domains emphasizing causal reasoning, undermines user trust. In

Received: 2024-05-09
First author: LI Jian, PhD, associate professor. His research interests include resilience and emergency management of integrated transportation systems, decision theory and methods, data mining and artificial intelligence. Email: jianli@tongji.edu.cn
Corresponding author:

recent years, research in Knowledge Engineering (KE) has sought to alleviate these challenges by effectively integrating LLMs with domain knowledge graphs. This approach has proven effective in mitigating hallucinations, enhancing adaptability in industry-specific contexts, and providing a crucial technological pathway for practical LLM applications. The deep integration of LLMs and knowledge graphs has emerged as a new research frontier. Sectors such as healthcare and finance have already engaged in active exploration, developing domain-specific LLMs for medical [10] and financial [11] applications. By contrast, exploration in the urban transportation domain remains in its early stages.

Based on a review of current studies, this study proposes a system architecture for a knowledge-enhanced LLM in urban transportation, grounded in the integration of knowledge graphs and LLMs. It further explores enabling techniques such as prompt engineering, retrieval-augmented generation, model fusion, and agent construction. Based on these advances, the study develops the Transportation Knowledge-Enhanced LLMs (TransKG-LLM) and conducts practical investigations across four dimensions: data enhancement, knowledge enhancement, model enhancement, and task enhancement.

1 Literature review

LLMs can be categorized into general-purpose LLMs and domain-specific LLMs (also referred to as vertical LLMs). General-purpose LLMs are trained on massive datasets and designed for broad applicability. Domain-specific LLMs are typically fine-tuned from general-purpose models or trained on domain-specific datasets to better address specialized problems. This section first reviews the developmental trajectory and technical pathways of both general-purpose and domain-specific LLMs, with a particular focus on applications in the medical, financial, and transportation sectors, and concludes with an overall assessment of current research.

1.1 General-purpose LLMs

From a historical perspective, the evolution of language models can be divided into four stages: statistical language models, neural network language models, pre-trained language models, and large language models [4] (see Fig. 1). Early statistical models often employed probabilistic methods to predict word sequence probabilities, such as the N-gram model [12]. Neural network language models, such as Word2vec [13], transformed words into vectors for prediction through neural architectures. A milestone was reached in 2017 when A. Vaswani et al. [14] introduced the Transformer model, which leveraged self-attention mechanisms and parallel computation to better capture long-range dependencies and improve training and inference efficiency. In 2020, OpenAI released GPT-3, with 175 billion parameters, whose strong emergent abilities signaled the advent of the large language model era.

With the launch of ChatGPT in late 2022 [15], LLMs entered the public spotlight. In recent years, development has gradually shifted from text-based models toward multimodal integration, encompassing text, images, and video (see Fig. 2). Such multimodal large language models (MLLMs) substantially enhance models' capacity to understand and generate content aligned with the real world, as demonstrated by models such as LLaVA [16], GPT-4 multimodal [17], and the text-to-video model Sora. Nonetheless, due to limitations in textual knowledge structures, update frequency, and training costs, the performance of LLMs in vertical domains such as medicine, finance, and transportation remains constrained, with hallucinations particularly pronounced.

1.2 Domain-specific LLMs

Domain-specific LLMs are generally optimized from general-purpose models through two technical pathways: internal optimization and external optimization.

1) Internal optimization: This involves modifying model parameters (e.g., weights) to improve performance on domain-specific tasks. Two principal approaches are: Pre-training, which learns domain knowledge from large-scale unsupervised datasets to build parameterized models capable of performing real-world tasks [14]; Fine-tuning, which employs domain-labeled data for efficient parameter adjustment and incremental training, thereby enhancing performance in specialized tasks [18].

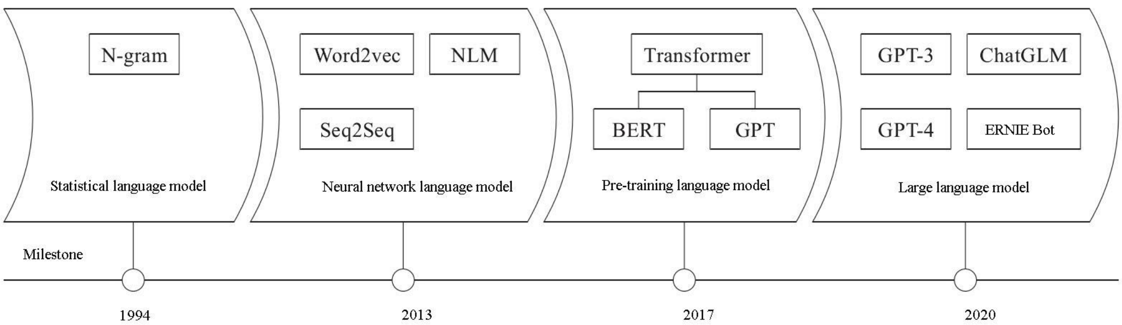

Fig. 1 Development of language models

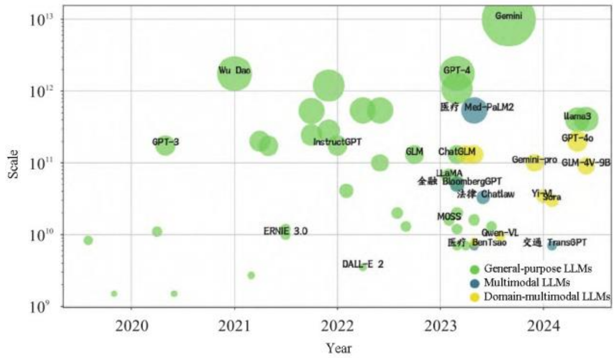

Fig. 2 Development of large language models

2) External optimization: Instead of directly updating model parameters, external optimization improves performance in vertical domains by incorporating external knowledge bases or tools. Major approaches include: Prompt engineering, which guides models toward more accurate outputs through carefully designed prompts [19]; Retrieval-Augmented Generation (RAG), which leverages external knowledge repositories with re-ranking strategies during retrieval to support the generation of highly relevant content [20]; Model fusion, which combines the generalization ability of LLMs with the domain-task proficiency of smaller models; Agent, which are complex systems built upon LLMs that typically integrate planning, memory, feedback, and execution modules to handle more sophisticated tasks [21–22].

In summary, internal optimization enhances the performance of general-purpose LLMs in vertical domains but faces challenges such as large data requirements, high costs, and slow knowledge updates. External optimization, by contrast, offers faster knowledge updates, stronger interpretability, and lower costs, making it particularly suited to leveraging accumulated domain knowledge. Consequently, it has attracted considerable attention in recent years in the development of domain-specific LLMs. A comparative analysis of internal and external optimization approaches across aspects such as domain data, construction costs, data security, and controllability of outputs is shown in Fig. 3.

	External optimization				Internal optimization	
	Prompt	RAG	Model fusion	Agent	Fine-turning	Pre-training
Domain data	Few	Moderate	Moderate	Moderate	More	Many
Construction costs	Low	Lower	Moderate	Higher	Higher	High
Data security	Low	Moderate	Moderate	Higher	High	High
Controllability of outputs	Moderate	Better	Good	Good	Better	Better

Fig. 3 Comparison of internal and external optimization methods for large domain language models

1.2.1 LLMs in the medical domain

The medical domain benefits from relatively strong data and knowledge-graph foundations and has been among the earliest to explore applications of domain-specific LLMs. In the internal optimization aspect: Google's Med-PaLM2 [23] combines fine-tuning on medical datasets with prompt optimization to significantly improve intelligent question answering in medicine. PMC-LLaMA [24], built on LLaMA, incorporates medical scholarly articles and classic textbooks together with knowledge graphs for fine-tuning, achieving strong performance on multiple medical QA benchmarks. Similarly, ChatDoctor [25], also based on LLaMA, uses doctor–patient dialogue datasets and external knowledge bases for fine-tuning, creating a QA system better aligned with real-world medical contexts. In the external optimization aspect: MedDoc-Bot [26] integrates multiple LLMs to build a medical document-based conversational system. Microsoft has demonstrated that, through prompt optimization alone, ChatGPT can achieve expert-level performance in medical QA without further fine-tuning [27]. Despite these advances, medical LLMs still face challenges such as insufficient knowledge training and immature long-context reasoning technologies [28]. Further in-depth research and exploration remain necessary.

1.2.2 LLMs in the financial domain

The financial sector is characterized by large-scale data computation, coupled with stringent requirements for privacy and security. In recent years, efforts have been made to integrate domain-specific expertise with general-purpose LLMs, achieving promising results in applications such as intelligent question answering and work-order assistants. In terms of internal optimization, leading enterprises in the sector have leveraged vast proprietary datasets and combined general-purpose LLMs with financial knowledge to build industry-focused models. For example, based on the open-source BLOOM model, BloombergGPT [29] was trained on decades of financial data, resulting in a finance-specialized LLM that demonstrated superior performance in knowledge-based QA and reasoning tasks. For external optimization, approaches mainly involve combining financial knowledge with internal data to construct applications that meet strict security standards. One example is the ChatABC system developed by the Agricultural Bank of China, which is deployed as a QA assistant and work-order response assistant. Additionally, some applications combine financial tool libraries and model repositories to build LLM-driven systems capable of handling complex tasks. FinGPT [30], for instance, is an open-source LLM framework specifically designed for finance, aimed at improving adaptability in this domain. Given the sector's rigorous privacy and security constraints, most enterprises currently apply LLMs primarily for cost reduction and efficiency improvement. Nonetheless, persistent challenges remain, particularly in interpretability and data privacy protection.

1.2.3 LLMs in the urban transportation domain

Urban transportation is a multidisciplinary field that emphasizes system-wide complexity, dynamism, and uncertainty [31]. Current internal optimization of LLMs in this domain primarily builds upon open-source models. For example, TransGPT [32], fine-tuned on domain-specific textual and conversational data based on ChatGLM, has been developed for QA tasks, thereby improving accuracy in professional knowledge responses. Similarly, TrafficSafetyGPT, built on LLaMA, was trained on traffic safety instruction sets to overcome the limitations of prompt engineering [33]. However, most research has focused on external optimization. Some scholars have employed prompt engineering to assist LLMs in addressing complex traffic problems, such as signal control [34] and mobility behavior prediction [35]. Others have combined prompt optimization with multi-agent collaboration for monitoring and management tasks [36]. To enhance the transferability of transportation-specific models, researchers have also explored hybrid approaches that fuse domain models with LLMs, thereby improving their capacity to handle complex tasks [37]. Moreover, to adapt to multimodal scenarios in urban transportation, a multimodal LLM framework (MT-GPT) has been proposed, adopting a hierarchical "point-line-plane" structure to provide data-driven support for complex decision-making in transportation systems [38]. Currently, urban transportation LLMs face challenges in terms of domain knowledge completeness, adaptability to complex scenarios, and interpretability. Future research should focus on strengthening cross-domain knowledge integration, enhancing reasoning capabilities for complex tasks, and improving model interpretability.

1.3 Summary

Current studies on domain-specific LLMs have largely concentrated on two pathways: (1) internal optimization, which fine-tunes general-purpose LLMs to improve adaptability in industry-specific contexts; and (2) external optimization, which employs methods such as Retrieval-Augmented Generation (RAG) and large-small model integration to better adapt general-purpose LLMs to domain-specific applications. Internal optimization requires strong technical expertise, extensive datasets, and substantial computational resources, enabling intellectual property protection and deep customization of LLMs' capabilities. However, it is costly and technically demanding. By contrast, external optimization emphasizes the integration of industry-specific data and knowledge resources, facilitating rapid iterative development tailored to practical scenarios, thereby accelerating the deployment of domain-specific LLMs.

Given the complexity and dynamism of urban transportation, as well as the resource and expertise demands of training LLMs, external optimization presents a more feasible technical pathway. This approach relies primarily on open-source general-purpose LLMs, augmented by corpora and knowledge graphs constructed from disciplinary knowledge, scientific literature, and industry reports. Through prompt engineering, RAG, model fusion, and agent-based methods, domain-specific LLMs can be developed to enhance understanding of urban transportation and to address complex problems. In turn, such systems can provide more scientific and refined decision support for urban transportation governance.

2 Architecture of LLMs in the urban transportation domain

2.1 Causes of existing problems in urban transportation LLMs

When applied to urban transportation, general-purpose LLMs face the dual challenges of "hallucinations" and inadequate reasoning capabilities. The underlying causes are as follows:

1) Lack of high-quality corpora in the transportation domain. Compared with medicine and finance, the accumulation of textual data resources in urban transportation is limited, particularly in the form of open-source corpora. This constrains LLMs' ability to acquire a deep understanding of domain-specific terminology, standards, and regulatory evolution. For example, transportation planning must adhere to the latest urban planning, ecological protection, and public safety regulations. However, due to incomplete knowledge coverage, LLMs often fail to provide accurate or compliant content.

2) Insufficient integration with domain knowledge graphs. Urban transportation involves complex scenarios requiring the comprehension and analysis of multidimensional spatial semantics. Without integration with knowledge graphs, LLMs are unable to capture the underlying logic of decision-making in the transportation domain, and thus lack reasoning capabilities for business-specific tasks. For instance, when prompted with "assess traffic congestion during heavy rainfall and propose mitigation strategies," an LLM may fail to generate logically coherent content because it cannot recognize the causal chain linking rainfall → reduced road capacity → congestion.

3) Lack of integration with transportation-specific models. Even if LLMs can identify domain-specific problems and analyze their underlying logic, they remain incapable of conducting effective quantitative decision-making. For example, when asked to "predict and optimize traffic volume distribution for the next peak period based on current congestion and weather conditions," LLMs cannot dynamically invoke professional tools such as traffic flow models, meteorological forecasting models, and signal control models for simulation and strategy evaluation. As a result, it cannot provide optimized schemes for specific road segments, time periods, or signal timings, thereby limiting its practical utility.

4) Insufficient capacity for executing complex domain tasks. Although knowledge-based QA technologies have been widely applied across multiple domains, their application in tasks such as planning scheme evaluation and automated industry report generation remains at an exploratory stage. In particular, during transportation planning processes that require soliciting opinions from multiple stakeholders, unresolved challenges include: how to simulate and interpret the diverse perspectives and behavioral patterns of different participants, and how to conduct iterative, multi-round evaluation and feedback for complex tasks. Addressing these challenges requires the design of a scientifically grounded and rational system of evaluation metrics.

2.2 Decision-making paradigms and decision support systems integrating LLMs

For professionals in the urban transportation sector, scientifically understanding transportation dynamics and formulating feasible solutions to practical problems are of paramount importance. Urban transportation is a typical interdisciplinary domain involving sociology, economics, management, and engineering, where decision-making must balance diverse and often competing interests. Quantitative decision analysis methods are therefore critical to advancing decision-making toward greater scientific rigor and refinement. In practice, quantitative models for decision support are widely applied in transportation planning and governance, including tasks such as problem diagnosis, feature analysis, and scheme evaluation [39]. With China's urbanization entering a new stage of ecological civilization, the connotation and extension of urban transportation have undergone a profound transformation. Urban transportation is shifting from serving as a pioneer of urban development to acting as a catalyst for spatial organization. This shift poses new challenges for supporting coordinated development and refined governance in the context of high-quality development. Against this backdrop, the complexity of urban transportation decision-making is increasing. How to rationally apply digital and intelligent technologies to enhance the scientific, refined, and intelligent nature of decision-making has become a focal issue in the field.

Since the concept was introduced in the 1970s [40], decision support systems (DSSs) have been widely applied across industries. Based on their driving elements, the evolution of DSSs can be divided into four stages: experience-driven, model-driven, data-driven, and AI-driven (see Tab. 1). In the experience-driven stage, decisions largely depended on the expertise of decision-makers, supplemented by relevant cases and rules [41]. The model-driven stage reduced reliance on personal experience by incorporating models and simulations for quantitative evaluation of strategies or plans. The data-driven stage emphasized integrating data analysis into the decision-making process [42], focusing on generating evidence for decision-making and promoting evidence-based decision analysis. The AI-driven stage leverages the strengths of LLMs in human-computer interaction, knowledge reasoning, and model integration, thereby further improving the effectiveness of decision support.

Tab. 1 Development stages of the decision support system

Development stage	Decision mode	Decision method	Characteristics of decision-information environment
Experience-based judgment	Subjective experience + limited cases	Experience + rules	Supported by case libraries and rule bases
Model-driven	Subjective experience + model-based assessment	Qualitative judgment + quantitative evaluation	Theoretical models + computational simulation
Data-driven	Evidence-based analysis	Analytics	Big data analytics + model-based assessment
AI-driven	Evidence-based analysis + human–machine interaction	Knowledge reasoning	Supported by large models and other AI technologies

The integration of LLMs into decision support systems brings notable changes and improvements in both decision-making paradigms and system performance (see Fig. 4).

1) Decision-making paradigms. LLMs' capabilities in complex context understanding, reasoning, and solution generation can significantly reduce the workload of decision analysts. The current decision-making paradigm is structured as a three-tier chain: "decision-maker → decision analyst → decision object." With the incorporation of LLMs, simpler and more procedural decision tasks can be directly handled through human-model interaction. For more complex problems, the three-tier paradigm remains applicable, but LLMs can effectively augment the work of decision analysts. Consequently, the decision-making paradigm evolves from a three-tier to a "two-and-a-half-tier" structure. As LLMs are

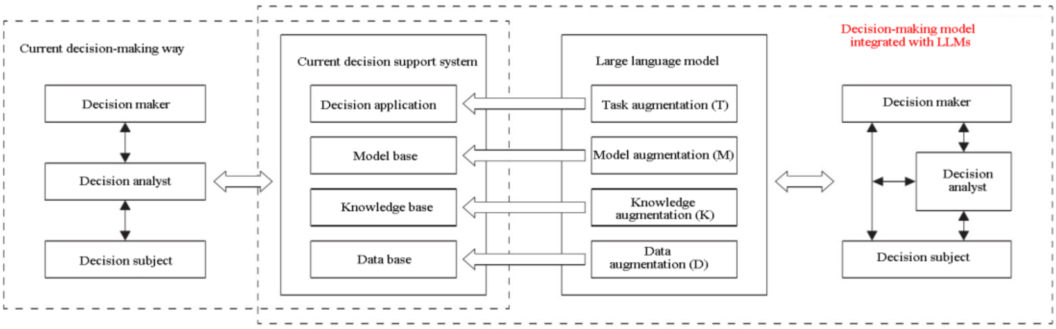

Fig. 4 The position of large language models in decision support systems

progressively optimized through internal and external enhancements, they can gradually assume roles as co-creators, executors, feedback providers, and evaluators of decision information, thereby greatly improving the capacity and efficiency of decision support.

2) Decision support system architecture. Compared with traditional DSSs built on databases, case libraries, model repositories, and task libraries, the integration of LLMs substantially enhances system performance: Data Enhancement: Incorporating large volumes of textual data compensates for the limitations of traditional relational databases. By curating domain-specific datasets, LLMs help construct foundational knowledge resources for urban transportation. Knowledge Enhancement: Building knowledge graphs structures domain knowledge systematically. Supported by internal and external knowledge bases, LLMs can broaden retrieval coverage and improve the relevance of generated content. Model Enhancement: By integrating professional transportation models, LLMs strengthen their ability to understand, execute, and provide feedback in domain-specific scenarios. Task Enhancement: LLMs can employ chain-of-thought reasoning to decompose complex tasks into subproblems, guiding step-by-step reasoning that improves both the accuracy and interpretability of decisions. Moreover, LLM-based multi-agent systems can simulate real-world multi-stakeholder decision processes. Through dialogue-based coordination, these agents iteratively enhance the interpretability and trustworthiness of decision outcomes.

2.3 System architecture of the knowledge-enhanced LLM for urban transportation

The system architecture of the knowledge-enhanced LLM for urban transportation (TransKG-LLM) is shown in Fig. 5. The successful deployment of LLMs in urban transportation hinges on effective data utilization, information retrieval, and improved reasoning, together with the integration of professional models and real-time tools to form productivity-enhancing agents. On the data side, the focus is on processing textual data and fusing it with existing structured data. On the knowledge side, building on prior knowledge-graph work and combining knowledge representation and evaluation enables generation of trustworthy and interpretable results. Model enhancement emphasizes the fusion of LLMs with existing industry models. Task enhancement centers on agent design and coordination for scenario-specific problems.

1) Data enhancement.

Curate and organize multidisciplinary, multi-scenario textual data within the domain.

(i) Data collection: Gather textual sources such as standards and specifications, industry reports, scientific literature, survey reports, and social media in urban transportation.

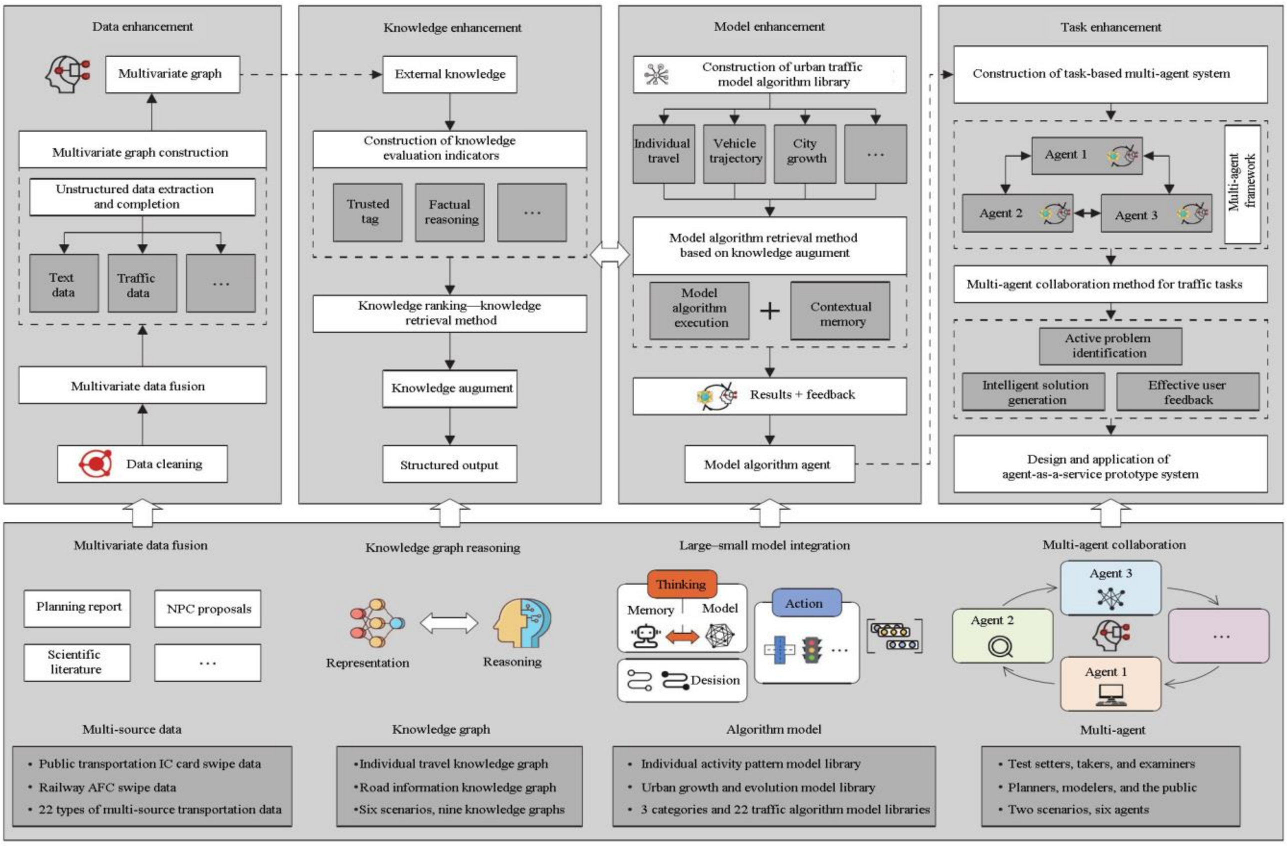

Fig. 5 System architecture of knowledge-enhanced LLM in urban transportation domain

(ii) Empirical analysis: For major transportation problems, screen and assess materials using empirical methods to ensure accuracy and reliability.

(iii) Instruction-set construction: Using supervised learning, create instruction-labeled text pairs (high-quality questions and expected responses) to support subsequent training of LLMs for the urban transportation domain.

2) Knowledge enhancement.

Use an urban-transportation knowledge graph to systematically organize entities and their complex relationships, employing a structured semantic network to make the generation process controllable and the outputs trustworthy.

(i) Offline text processing: Segment large-scale textual data and generate text embeddings; store segments and indices in a vector database to enable efficient vector-similarity retrieval. In parallel, perform named entity recognition and relation extraction; align and complete extracted triples to improve the completeness of the knowledge graph.

(ii) Online RAG: Upon receiving a user query, generate a query-specific subgraph from the knowledge graph and conduct vector-similarity retrieval. Perform topic-aware pre-selection and re-ranking to increase retrieval hit rate and the relevance of generated text, and finally return LLM-generated responses to the user.

3) Model enhancement.

Fuse domain-specific LLMs with transportation professional models to augment domain expertise on top of generalization capabilities.

(i) Static management of transportation models: Specify key metadata (e.g., primary functions, interface parameters), adapt interfaces to ensure interoperability, and manage/query services via a unified registry. Derive prompt templates from model parameters to guide LLM invocation.

(ii) Dynamic interaction with transportation models: Based on user needs, produce an initial response plan and identify required internal/external resources; route requests to external transportation models to complete scenario-specific tasks; rigorously evaluate model outputs and have the LLM render human-interpretable results.

4) Task enhancement.

Implement a multi-agent coordination framework enabling information exchange and task linkage among agents, thereby supporting agile governance in urban transportation.

(i) Define task objectives: Analyze application scenarios and business requirements to specify the problems to be solved and goals to be achieved, guiding agent creation and task orchestration.

(ii) Create agents and orchestrate tasks: Design role-differentiated agents according to scenario tasks and LLM capability boundaries; allocate tasks and resources to ensure efficient collaboration and completion of complex tasks.

(iii) Iterative evaluation: Continuously optimize collaborative execution; after each iteration, evaluate against predefined rules to determine the extent to which the desired conditions have been met.

3 Application exploration

Based on the above technical pathway and prior large-scale knowledge-graph development, this study explores prompt engineering, RAG, model fusion, and agent technologies to propose a construction method for a knowledge-enhanced LLM in urban transportation. Practical explorations are conducted along four dimensions: data enhancement, knowledge enhancement, model enhancement, and task enhancement. (i) Data enhancement: construct a knowledge graph for the urban transportation domain by integrating textual data to strengthen data-fusion capabilities. (ii) Knowledge enhancement: implement domain knowledge QA by integrating a knowledge graph with a general-purpose LLM to improve domain understanding. (iii) Model enhancement: develop an large-small model integration–based urban travel-planning assistant to strengthen problem identification and prediction. (iv) Task enhancement: build a multi-agent teaching assistant to improve instructional efficiency and effectiveness. In addition, we further analyze the development difficulty of domain LLM applications across scenarios (see Fig. 6), with the aim of informing the intelligent transformation of the urban transportation sector.

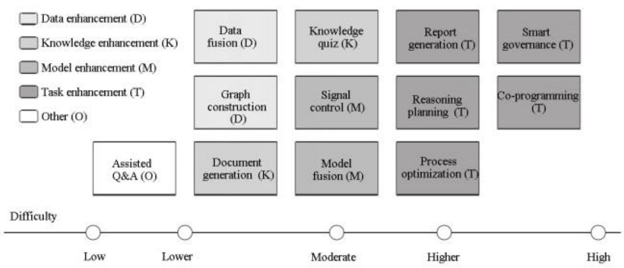

Fig. 6 Comparison of application scenarios and development challenges for LLM in urban transportation domain

3.1 Data enhancement: constructing a knowledge graph for urban transportation via multi-source data fusion

Multi-source data fusion is key to breaking data silos and unlocking data value. Knowledge graphs, with strong capabilities for knowledge organization and representation, offer a new technical pathway for data fusion in urban transportation. The technical workflow comprises ontology design, knowledge extraction, knowledge fusion, and knowledge reasoning (see Fig. 7).

(i) Ontology design: as the core structure of a knowledge graph, the ontology defines basic domain concepts and entities, typically including hierarchical design, attribute design, and relation design.

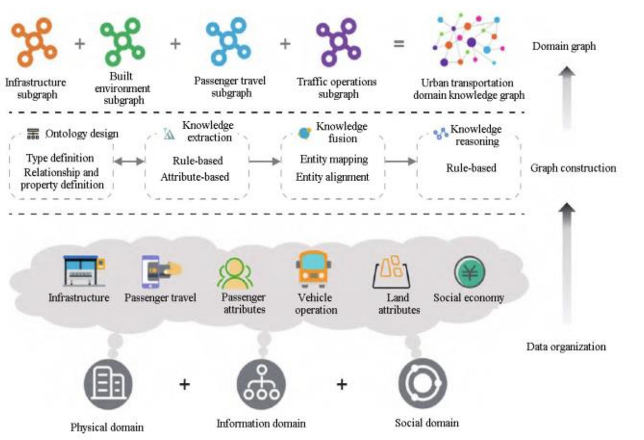

Fig. 7 Construction of knowledge graphs in urban transportation domain

(ii) Knowledge extraction: extract entities, relations, and attributes from unstructured or semi-structured texts (e.g., infrastructure, built environment, passenger travel, and transport operations).

(iii) Knowledge fusion: align and map knowledge extracted from different sources to form a unified, high-quality knowledge system.

(iv) Knowledge reasoning: apply rule-based inference to support diverse intelligent decision-making tasks with accuracy and interpretability.

We collected and curated transportation-domain texts, linking urban transportation scenario information on the basis of disciplinary knowledge. Sources included standards and specifications, industry reports, scientific literature, social media, disciplinary textbooks, and proposals submitted by National People's Congress deputies. Following a top-level → domain → task methodology, we designed the ontology for the urban transportation knowledge graph. Using LLMs and rule-learning methods, we achieved intelligent extraction and fusion of textual data. Through knowledge extraction and fusion, we constructed a domain knowledge graph encompassing tens of thousands of entities. While consolidating core knowledge elements in urban transportation, this graph also provides the foundational knowledge base for research and practice with domain LLMs.

3.2 Knowledge enhancement: domain knowledge QA by integrating a knowledge graph with a general-purpose LLM

Knowledge enhancement systematically organizes and presents complex relationships among entities in the domain via a knowledge graph, employing a structured semantic network to make the generation process controllable and the outputs trustworthy. To address the challenge of mitigating "hallucinations" when applying general-purpose LLMs in domain settings, we build a knowledge-graph–integrated QA system that combines a domain knowledge graph with a general-purpose LLM [43], thereby serving industry QA needs more effectively. The system comprises four components: knowledge representation, knowledge retrieval, knowledge output, and knowledge updating (see Fig. 8).

(i) Knowledge representation: introduce credibility labels to ensure the reliability and authoritativeness of sources.

(ii) Knowledge retrieval: leverage the LLM's natural language understanding to identify user intents and extract keywords, then perform retrieval within the relevant subgraph.

(iii) Knowledge output: include knowledge ranking (ordering content by credibility and relevance) and knowledge reasoning (generating content on that basis).

(iv) Knowledge updating: continuously update, correct, and supplement the domain knowledge graph through feedback.

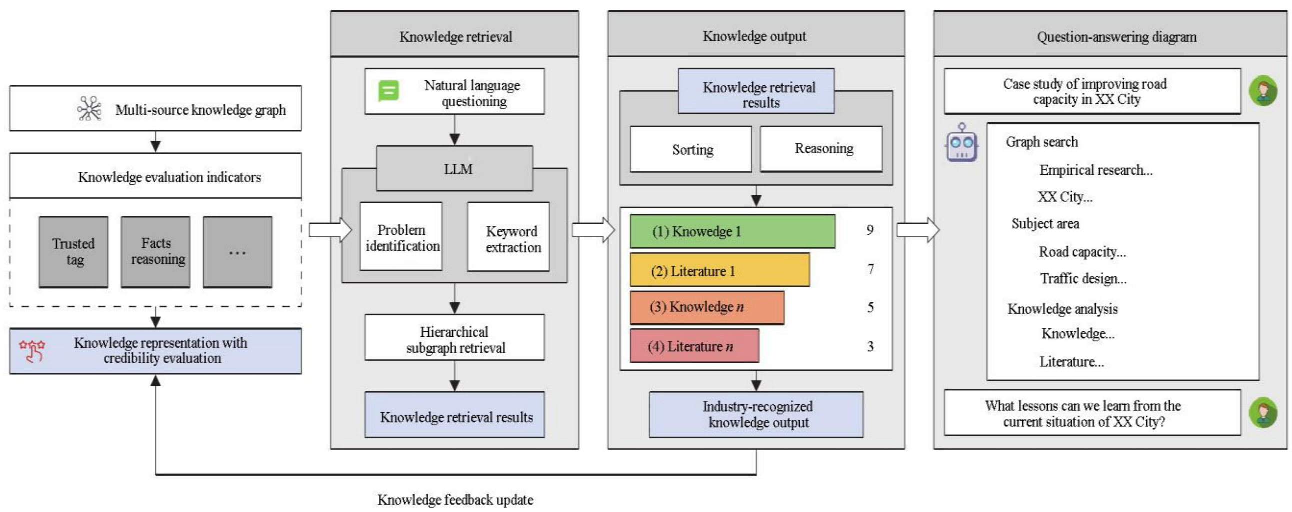

Fig. 8 Knowledge-based question-answering systems in urban transportation domain

To meet the scientific decision-making needs of transportation governance, and using low-carbon travel behavior interventions as the focal scenario, we constructed an evidence-based knowledge graph grounded in scientific literature [44] to enable LLMs to produce trustworthy scientific evidence. The study proposes methods for evidence classification and grading and supports meta-analysis. Decision-makers and decision analysts can query the evidence-based knowledge graph, tailored to specific contexts and objectives and taking city-specific characteristics into account to obtain evidence, including candidate evidence, evidence grades, and associated attributes (see Fig. 9). Building on this foundation, and leveraging an LLM-based QA system, decision-makers or analysts can pose queries targeted to particular experimental settings and populations, thereby rapidly and effectively retrieving credible scientific evidence.

3.3 Model enhancement: an intelligent assistant for urban travel planning

Model enhancement builds on the generalization capability of LLMs by embedding the domain knowledge and business understanding of small models, enabling LLMs to analyze and act within specific transportation scenarios. Using travel-characteristics analysis in transportation planning as an example, this study employs an LLM–small-model fusion approach to develop an intelligent assistant for urban travel planning (see Fig. 10). Its main components are:

1) Database. Serving as the center for information storage and processing, it integrates large-scale urban transportation data, including—but not limited to—road network structures, trajectories, and travel demand.

2) Model repository. Providing the capability for autonomous planning by the assistant, it integrates both online and local algorithm/model libraries to automate urban travel-planning tasks.

3) Planner. Building on domain understanding, it decomposes complex tasks into executable subtasks and routes them to the appropriate algorithmic services, while continually evaluating execution and performance during task processing.

Using the above method for a travel-behavior analysis scenario, we constructed a database containing hundreds of thousands of trajectory records and a model repository comprising more than ten small models for travel-behavior analysis. With model enhancement, we achieved large-scale spatial analysis and pattern recognition of travel behavior, as well as small-scale mode-choice prediction and trajectory prediction. Based on this repository, we tested mainstream open-source and closed-source models. Under the proposed model enhancement approach, with prompting, the hit rates of small-parameter models exceeded 80%; without prompting, mainstream large-parameter LLMs also achieved hit rates above 80% (see Tab. 2).

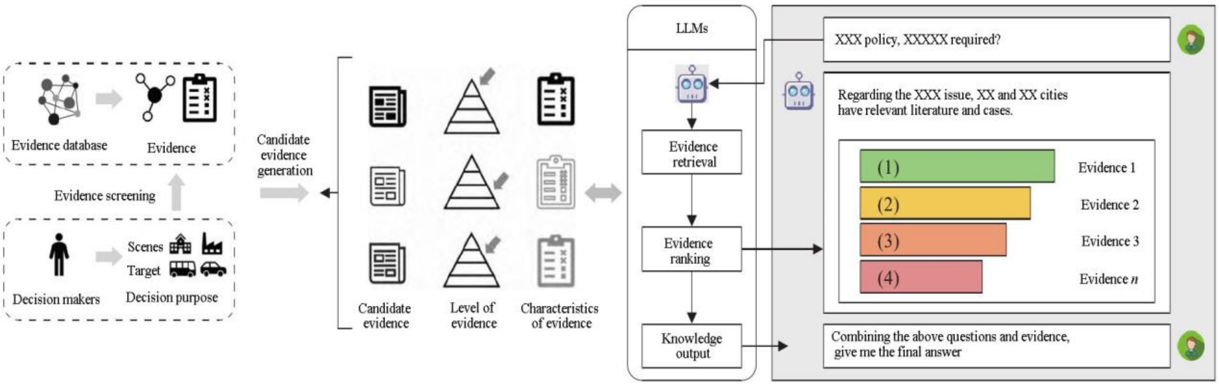

Fig. 9 Question and answer system for evidence-based decision- making of low-carbon travel behavior interventions

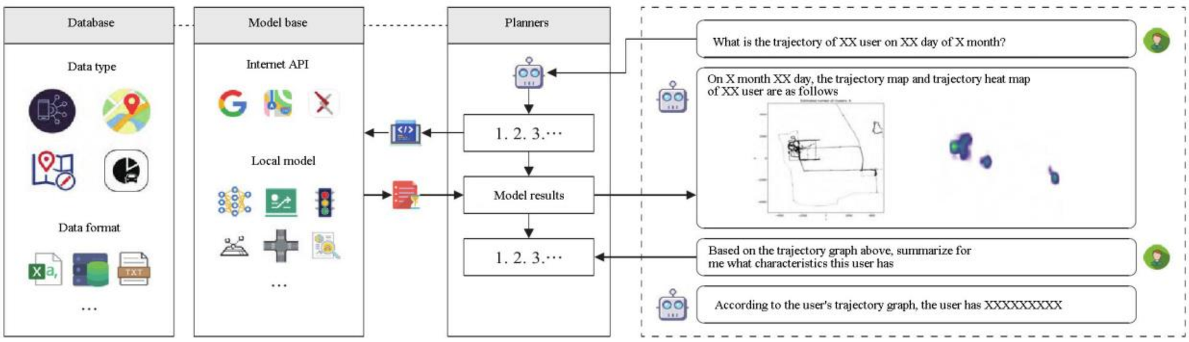

Fig. 10 Intelligent assistant system for urban travel planning

Tab. 2 Results of knowledge enhanced LLM %

Model type	Model name	Hit rate	
		Without prompt	With prompt
Open-source	qwen2-7b	30	80
	qwen2-57b	60	90
	qwen2-72b	98	99
Closed-source	gpt-4	90	95
	gpt-3.5	80	90
	qwen-max	80	95
	chatglm-4	95	99

3.4 task enhancement: a multi-agent intelligent teaching assistant

Task enhancement is task-oriented: through a multi-agent collaboration framework built on LLMs, it enables information exchange and task linkage among multiple agents to solve domain-specific problems. Using a transportation engineering course as a case, we design teaching and assessment agents and their coordination mechanisms, and develop a course intelligent teaching assistant (see Fig. 11), comprising:

1) Course knowledge-graph construction. Perform knowledge extraction, fusion, storage, and evaluation over teaching resources such as textbooks, slides, exercises, and exam papers.

2) Teaching agent. Consists of three modules—navigation, Q&A, and feedback. The navigation module leverages the knowledge graph and LLMs to generate customizable study syllabi; the Q&A module uses natural language processing to answer students' questions in real time; the feedback module produces personalized learning reports to improve assessment of learning outcomes.

3) Assessment agent. Comprises item generation, grading, and scoring modules. The item-generation module automatically creates questions using the knowledge graph and item-bank data; the grading module combines OCR with natural language processing for automatic marking and evaluation; the scoring module assesses performance using a multidimensional scoring framework.

Using this method, we tested the system with a university's final examination in transportation engineering. Taking Q&A items on professional knowledge points as the test set, we compared the results of the knowledge-enhanced LLM with those of a general-purpose LLM. The experiments show that knowledge enhancement effectively mitigates hallucinations and improves stability and accuracy in professional knowledge understanding, fine-grained detail recognition, and numerical computation. Tab. 3 reports the exam evaluation results: the knowledge-enhanced LLM achieved significant gains in accuracy across all question types, outperforming the general-purpose LLM. For example, in Q&A concerning traffic capacity categories and their ordering, the knowledge-enhanced LLM correctly distinguished theoretical, actual, and design capacities and their relative magnitudes, whereas the general-purpose LLM provided only generic responses; and for yellow interval (amber) time calculations, the knowledge-enhanced LLM produced correct computations, while the general-purpose LLM did not.

Tab. 3 Evaluation results of the final examination paper of transportation engineering in a university

Question type	Maximum score	Score of the knowledge-enhanced LLM	Score of the general-purpose LLM
True/false questions	10	8	7
Multiple-choice questions	30	24	16
Short-answer questions	30	30	22
Calculation problems	30	18	13
Total score	100	80	58

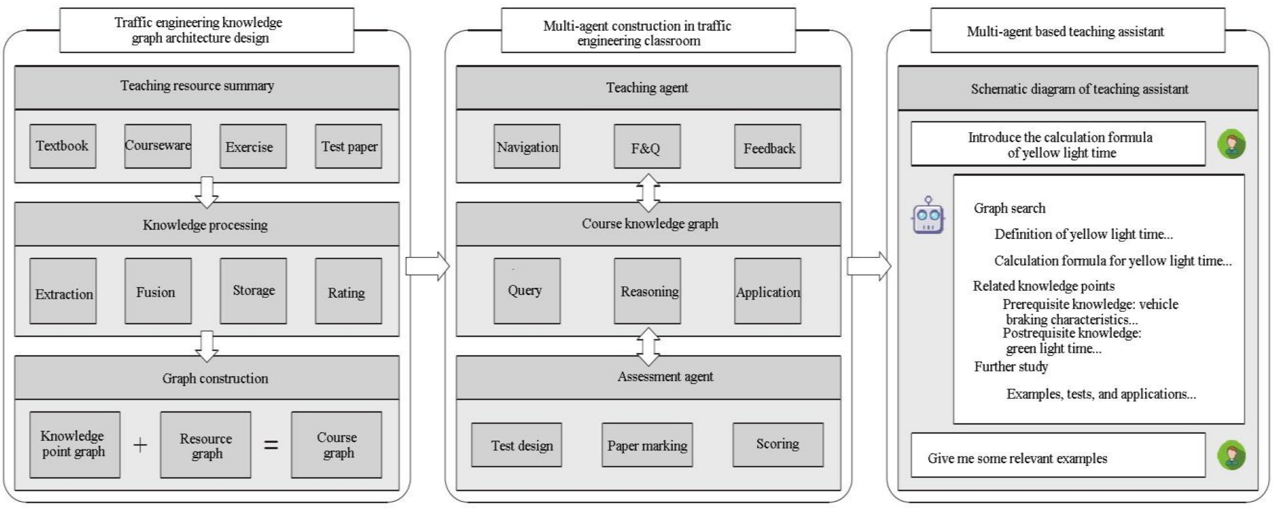

Fig. 11 Schematic diagram of intelligent teaching assistant

4 Conclusions

As a major branch of AI, LLMs exhibit broad application prospects in urban transportation. Their capabilities in semantic understanding, text generation, and logical reasoning provide robust support for management and decision-making in this sector. From a technical-pathway perspective, an external optimization approach—leveraging open-source general-purpose LLMs and constructing corpora and knowledge graphs from existing disciplinary knowledge, scientific literature, and industry reports, while exploring prompt engineering, RAG, model fusion, and agent-based methods—offers a pragmatic route for advancing LLM applications in urban transportation. However, despite substantial accumulations of structured data and quantitative analysis models, research remains limited in industry knowledge-graph construction, LLM–industry-model fusion, and agent design. These gaps hinder the effective alignment between the generalization capacity of LLMs and domain-specific specialization needs. Looking ahead, it is necessary to refine knowledge graphs and knowledge-enhanced LLMs for concrete application scenarios and, through industry–academia–research collaboration, to build an open-source ecosystem for LLMs in urban transportation. Such efforts will enhance the scientific rigor, refinement, and intelligence of governance capabilities in the field.

References

[1] WANG G T. Concept and objectives of urban transportation management[J]. Urban transport of China, 2018, 16(1):1-6.
[2] LI J, WU Z X, YANG F, et al. Highlights of China Urban Transportation Planning 2023 Annual Conference[J]. Urban transport of China, 2023, 21(6):102-113.
[3] WU T, HE S, LIU J, et al. A brief overview of ChatGPT: the history, status quo and potential future development[J]. IEEE/CAA journal of automatica sinica, 2023, 10(5): 1122-1136.
[4] ZHAO W X, ZHOU K, LI J, et al. A survey of large language models[J]. arXiv preprint arXiv, 2023, 2303.18223.
[5] TEAM G, ANIL R, BORGEAUD S, et al. Gemini: a family of highly capable multimodal models[J]. arXiv preprint arXiv, 2023, 2312.11805.
[6] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: open and efficient foundation language models[J]. arXiv preprint arXiv, 2023, 2302.13971.
[7] JIANG A Q, SABLAYROLLES A, MENSCH A, et al. Mistral 7B[J]. arXiv preprint arXiv, 2023, 2310.06825.
[8] LIU Y, ZHANG K, LI Y, et al. Sora: a review on background, technology, limitations, and opportunities of large vision models[J]. arXiv preprint arXiv, 2024, 2402.17177.
[9] ZHONG T, LIU Z, PAN Y, et al. Evaluation of OpenAI o1: opportunities and challenges of AGI[J]. arXiv preprint arXiv, 2024, 2409.18486.
[10] THIRUNAVUKARASU A J, TING D S J, ELANGOVAN K, et al. Large language models in medicine[J]. Nature medicine, 2023, 29(8): 1930-1940.
[11] LI Y, WANG S, DING H, et al. Large language models in Finance: a survey[C]//Association for Computing Machinery. ICAIF'23: proceedings of the fourth ACM international conference on AI in finance. 2023: 374-382.
[12] CAVNAR W B, TRENKLE J M. N-gram-based text categorization[C]//US Department of Energy. 3rd annual symposium on document analysis and information retrieval. 1994: 14.
[13] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv, 2013, 1301.3781.
[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 1706.03762.
[15] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback[J]. Advances in neural information processing systems, 2022, 35: 27730-27744.
[16] LIU H, LI C, LI Y, et al. Improved baselines with visual instruction tuning[C]//IEEE. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle, WA, USA: IEEE, 2024: 26296-26306.
[17] ACHIAM J, ADLER S, AGARWAL S, et al. Gpt-4 technical report[J]. arXiv preprint arXiv, 2023, 2303.08774.
[18] HU E J, SHEN Y, WALLIS P, et al. Lora: low-rank adaptation of large language models[J]. arXiv preprint arXiv, 2021, 2106.09685.
[19] SAHOO P, SINGH A K, SAHA S, et al. A systematic survey of prompt engineering in large language models: techniques and applications[J]. arXiv preprint arXiv, 2024, 2402.07927.
[20] GAO Y, XIONG Y, GAO X, et al. Retrieval-augmented generation for large language models: a survey[J]. arXiv preprint arXiv, 2023, 2312.10997.
[21] XI Z, CHEN W, GUO X, et al. The rise and potential of large language model based agents: a survey[J]. arXiv preprint arXiv, 2023, 2309.07864.
[22] ZHOU Y Z, LUO J R, GU X Q, et al. Survey on intelligent planning methods from large language models perspective[J]. Journal of system simulation, 2024: 1-19[2024-09-08]. https://doi.org/10.16182/j.issn1004731x.joss.23-1468.
[23] SINGHAL K, TU T, GOTTWEIS J, et al. Towards expert-level medical question answering with large language models[J]. arXiv preprint arXiv, 2023, 2305.09617.
[24] WU C, ZHANG X, ZHANG Y, et al. Pmc-llama: further finetuning LLaMA on medical papers[J]. arXiv preprint arXiv, 2023, 2304.14454.
[25] LI Y X, LI Z H, ZHANG K, et al. Chatdoctor: a medical chat model fine-tuned on LLaMA model using medical domain knowledge[J]. arXiv preprint arXiv, 2023: 2303, 14070.
[26] JABARULLA M Y, OELTZE-JAFRA S, BEERBAUM P, et al. MedDoc-Bot: a chat tool for comparative analysis of large language models in the context of the pediatric hypertension guideline[J]. arXiv preprint arXiv, 2024, 2405.03359.
[27] NORI H, LEE Y T, ZHANG S, et al. Can generalist foundation models outcompete special-purpose tuning? case study in medicine[J]. arXiv preprint arXiv, 2023, 2311.16452.
[28] WU S, IRSOY O, LU S, et al. BloombergGPT: a large language model for finance[J]. arXiv preprint arXiv, 2023, 2303.17564.
[29] RUAN T, BIAN Y A, YU G Y, et al. A review on research and application of medical large language models[J]. Chinese journal of health informatics and management, 2023,20(6):853-861.
[30] YANG H, LIU X Y, WANG C D. Fingpt: open-source financial large language models[J]. arXiv preprint arXiv, 2023, 2306.06031.
[31] LI J. Agile governance for new forms of transportation[J]. Urban transport of China,2023, 21(3):9-10.
[32] WANG P, WEI X, HU F, et al. TransGPT: multi-modal generative pretrained transformer for transportation[J]. arXiv preprint arXiv, 2024, 2402.07233.
[33] ZHENG O, Abdel-Aty M, WANG D, et al. TrafficSafetyGPT: tuning a pre-trained large language model to a domain-specific expert in transportation safety[J]. arXiv preprint arXiv, 2023, 2307.15311.
[34] LAI S, XU Z, ZHANG W, et al. Large language models as traffic signal control agents: capacity and opportunity[J]. arXiv preprint arXiv, 2023, 2312.16044.
[35] ZHANG Z, SUN Y, WANG Z, et al. Large language models for mobility in transportation systems: a survey on forecasting tasks[J]. arXiv preprint arXiv, 2024, 2405.02357.
[36] WANG B, KARIM M M, LIU C, et al. Traffic performance GPT (TP-GPT): real-time data informed intelligent ChatBot for transportation surveillance and management[J]. arXiv preprint arXiv, 2024, 2405.03076.
[37] ZHANG S Y, FU D C, LIANG W Z, et al. TrafficGPT: viewing, processing and interacting with traffic foundation models[J]. Transport policy, 2024, 150: 95-105.
[38] ZHOU Z, GU Z Y, QU X B, et al. Urban multimodal transportation generative pretrained transformer foundation model: hierarchical techniques and application scenarios of spot-corridor-network decomposition[J]. China journal of highway and transport, 2024, 37(2):253-274.

[39] Editorial Department of China Journal of Highway and Transport. Review on China's traffic engineering research progress: 2016[J]. China journal of highway and transport, 2016, 29(6):1-161.
[40] EOM S, KIM E. A survey of decision support system applications (1995–2001)[J]. Journal of the operational research society, 2006, 57(11): 1264-1278.
[41] ZHANG Y, BU F L. Design and implementation of emergency decision support system based on CBR and RBR[J]. Software guide. 2019, 18(2):55-59.
[42] YU H, HE D N, WANG G Y, et al. Big data for intelligent decision making[J]. Acta automation sinica, 2020, 46(5):878-896.
[43] WEN S, QIAN L, HU M D, et al. Review of research progress on question-answering techniques based on large language models[J]. Data analysis and knowledge discovery, 2024, 8(6):16-29.
[44] LI J, XU C. Evidence-based practices in sustainable travel behavior intervention: a knowledge graph-based systematic review[J]. Journal of traffic and transportation engineering (English edition), 2024, 11(2): 293-311.

