Cities 165 (2025) 106142

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

Advancements in the application of large language models in urban studies: A systematic review

Junhao Xia [a], Yao Tong [a], Ying Long [b,*]

[a] School of Architecture, Tsinghua University, Beijing 100084, China

[b] School of Architecture and Hang Lung Center for Real Estate, Key Laboratory of Ecological Planning & Green Building, Ministry of Education, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Keywords:
Large language model
Urban study
Systematic literature review
BERT
GPT

ABSTRACT

Large language models (LLMs) possess robust comprehensive capabilities for addressing complex problems and have increasingly been applied to better elucidate urban phenomena. Amid the surge of LLMs, there is an urgent need for urban researchers to understand how previous works have integrated LLMs across various disciplines. In this paper, we conduct a systematic review of 233 articles focusing on the application of LLMs in urban studies, analyzing the developing tendencies using information extracted from articles with the aid of a customized Generative Pre-trained Transformer (GPT) assistant and giving in-depth reviews across different subfields. The findings highlight an exponential growth in related research over the past six years, with LLMs being applied to diverse scenarios such as text analysis and generation, domain knowledge question-answering as well as field-related task. Besides, GPT-based and Bidirectional-Encoder-Representations-from-Transformers-based (BERT-based) have emerged as the two mostly used models, with embedding and fine-tuning being the predominant methods for data processing and model adaptation. Additionally, the paper also addresses common concerns about LLMs and identifies future research opportunities in urban studies. This comprehensive analysis aims to provide valuable insights for researchers exploring the application of LLMs in urban studies and for those who have yet to begin utilizing them in their research.

1. Introduction

Cities, serve as dynamic hubs where diverse populations converge, interact, and innovate, have led to complex phenomena and patterns behind their physical spaces (Bettencourt & West, 2010). The proliferation of big data and machine learning technologies has unlocked unprecedented opportunities to mine urban data for gaining insights into these patterns (Batty, 2013), helping to interpret the complexity of urban systems arising from the myriad interactions among their components (Batty, 2008). These advancements have also expanded urban researchers' analytical capabilities and fostered targeted urban interventions, giving rise to data augmented planning and design (Long & Zhang, 2021).

Deep learning, a representative technology set of machine learning, has revolutionized urban data processing and interpretation in the recent ten years. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been utilized to capture the spatial and temporal dependencies inherent in urban phenomena like urban street quality (Chen, Chen, et al., 2023; Chen, Hu, et al., 2023; Chen, Wang, & Xu, 2023; Tang & Long, 2019) and urban shrinkage (Wang & Long, 2023). While deep learning models have been highly successful in processing structured data and identifying patterns in spatial-temporal urban dynamics (Zheng et al., 2014), their capacities to handle unstructured, text-heavy data remain constrained. Since traditional deep learning models require a large amount of annotated data which is often time-consuming, their performance on these specific tasks is usually unsatisfactory due to limited generalization capabilities of models (LeCun et al., 2015).

Since its introduction in 2017, the Transformer architecture has addressed long-range dependency challenges through a multi-head self-attention mechanism and an encoder–decoder framework for sequence-to-sequence tasks (Vaswani et al., 2017). Building on this foundation, Bidirectional Encoder Representations from Transformers (BERT) utilizes only the encoder stack and adopts a masked language modeling objective combined with next-sentence prediction to learn deep, bidirectional contextual embeddings, which makes BERT particularly well-

* Corresponding author at: Qinghuayuan, Beijing 100084, China.
E-mail addresses: xiajh23@mails.tsinghua.edu.cn (J. Xia), yao_tong@tsinghua.edu.cn (Y. Tong), ylong@tsinghua.edu.cn (Y. Long).

https://doi.org/10.1016/j.cities.2025.106142
Received 5 January 2025; Received in revised form 20 April 2025; Accepted 31 May 2025
0264-2751/

suited for natural language understanding tasks such as sentiment classification, question answering, and named entity recognition (NER) (Bommasani et al., 2021; Devlin et al., 2019). In contrast, Generative Pre-trained Transformer (GPT) leverages only the decoder stack in an autoregressive fashion, optimizing a causal language modeling objective that predicts the next token in a sequence and excels at text generation tasks including free-form composition, summarization, and conversational response generation (Bommasani et al., 2021; Radford et al., 2019). These divergent designs influence model adaptation strategies: BERT typically incorporates lightweight task-specific output layers atop its encoder and is fine-tuned on labeled datasets, whereas GPT models are often adapted through prompt engineering or reinforcement learning from human feedback (RLHF) to align the generative outputs with downstream requirements (Brown et al., 2020).

In addition to Transformer, the increased computational power (Shalf, 2020) and the availability of vast datasets (Naveed et al., 2024) also triggered significant breakthroughs in language models, arousing the prevalence of applying large language models (LLMs) in various domains (Topsakal & Akinci, 2023). Trained on vast amounts of data, LLMs are capable of understanding and generating text based on the knowledge they learnt from training period and can approximate human-level performance on various tasks (Wang et al., 2019). The emergence of LLMs has been instrumental in revolutionizing quantitative analysis (Ma et al., 2023; Topsakal & Akinci, 2023) and has paved the way for new research paradigms.

Despite the beginning stage, LLMs have already had profound relationships with urban studies, as we will confirm in this article later. For instance, LLMs have been utilized to conduct text analysis tasks including semantic analysis and topic modeling (Chen, 2024; Fu, Sanchez, et al., 2024; Fu, Wang, & Li, 2024; Saito & Haruyama, 2023), to analyze spatial-temporal data patterns and to help predict the tendency of humans and traffics mobility (Haydari et al., 2024; Li, Chen, et al., 2024; Li, Feng, et al., 2024; Li, Hu, et al., 2024; Li, Huang, et al., 2024; Li, Shi, & Li, 2024; Li, Tran, et al., 2024; Zhang, Han, et al., 2023; Zhang, Shah, & Han, 2023; Zhang, Sun, et al., 2023). Other relevant applications in urban studies include land-use analysis (Guo et al., 2024), scene generation (Deng et al., 2024), public event simulation (Xiao et al., 2023) and so on. Undoubtedly, the trend of applying LLMs in urban studies is expected to continue, as they function as invaluable tools for a wide array of applications from research to everyday tasks (Makridakis et al., 2023).

Therefore, there is an urgent need for urban researchers to understand and trace how previous works have incorporated LLMs across various fields in different contexts and purposes. Despite the awareness of reviewing application of LLMs in urban studies, many articles provide only a brief mention of future directions in their outlook sections (Chen, Liang, et al., 2024; Chen, Zhang, et al., 2024; Wu et al., 2024) or focus narrowly on a single sub-field instead of considering the broader landscape of interactions between LLMs and urban studies (Ullah et al., 2024; Zhang, Li, et al., 2024; Zhang, Liu, et al., 2024; Zhang, Sun, et al., 2024). Thus, this has consequently led to a notable research gap in the in-depth exploration of the application of LLMs in urban research, and we propose three research questions:

Research Question 1 (RQ1). What are the developing tendencies and scenarios of LLMs' application in urban studies?

Research Question 2 (RQ2). Which LLMs are most used in urban studies and in what ways?

Research Question 3 (RQ3). How are LLMs being applied across different fields of urban studies and what are the similarities and differences?

Through a systematic literature review, this paper aims to provide a comprehensive and objective view of the advancements of LLMs' application in urban research. By synthesizing evidence from multiple researches, we hope to offer valuable resources for researchers looking forward to applying LLMs in city-related studies, while also providing a comprehensive reference for those who are yet to familiarize themselves with these methods.

The remainder of this paper is organized as follows. *Methodology* outlines the methods of the systematic literature review including database sources selection, the choice of search terms, screening criteria and data extraction methods. *Results* provides an in-depth review of the included literature, offering insights into the current state of applications. Furthermore, we discuss the limitations and concerns about LLMs and propose the possible potential future directions. Finally, we conclude the paper.

2. Methodology

As a thorough, transparent, and effective method for collecting, analyzing, and synthesizing existing work (Fink, 2019), the comprehensive literature review serves as a reliable description of past research, from which other researchers might seek inspiration and use to position their own research (Yan, Mai, et al., 2024; Yan, Wen, et al., 2024). Based on Fink's (2019) guidelines and process, this paper conducts the systematic literature review by following five steps, aiming to answer RQs raised above. The five steps include selecting database sources, choosing appropriate keywords, establishing screening criteria, extracting structured data, and synthesizing analysis results (Fink, 2019). The selection process is illustrated in Fig. 1.

2.1. Selecting database sources

The Web of Science (WoS) was chosen as the database source. WoS is the world's oldest, most widely accepted and authoritative database of research publications and citations, encompassing a vast array of journals (Birkle et al., 2020). Besides, it also includes publications from open-access repositories such as arXiv, an important platform containing many insightful and cutting-edge studies leveraging LLMs.

2.2. Choosing appropriate keywords

The literature search terms consisted of two parts, urban studies and LLM. Due to the complexity of city systems and locality of different cities, urban studies are blurry academic definition without bounded academic disciplines (Bowen et al., 2010). Considering these, we deemed it appropriate to include articles with themes containing "urban", "city" or "building" to cover the relevant literature. Additionally, to avoid missing articles containing specific LLM technologies without explicitly mentioning the exact word "LLM", we included full names and abbreviations of typical LLMs. These include GPT and BERT, two models playing a significant role in the history of LLMs' development (Zhou et al., 2023). The final search terms used were ("urban" OR "city" OR "building") AND ("Large Language Model" OR "LLM*" OR "Generative Pre-Trained Transformer" OR "GPT*" OR "ChatGPT" OR "Bidirectional Encoder Representations from Transformer" OR "BERT") (ChatGPT is a family of powerful LLMs developed by OpenAI).

2.3. Establishing screening criteria

We restricted our literature search to the past decade in order to better capture the developmental trajectory of LLMs. Besides, we only focused on the documents written in English available for downloading. The document types include articles, preprints, conference papers, review articles and book sections. Therefore, the additional search syntax is as follows: 1) Publication year: 2015-01-01 to 2025-04-01, 2) Language: English, 3) Document types: article, preprint, conference paper, review article, book section, 4) Accessibility: full-text access. We conducted the following screening process manually. After this step, 5127 articles were selected.

We established criteria to determine the inclusion of studies, and the

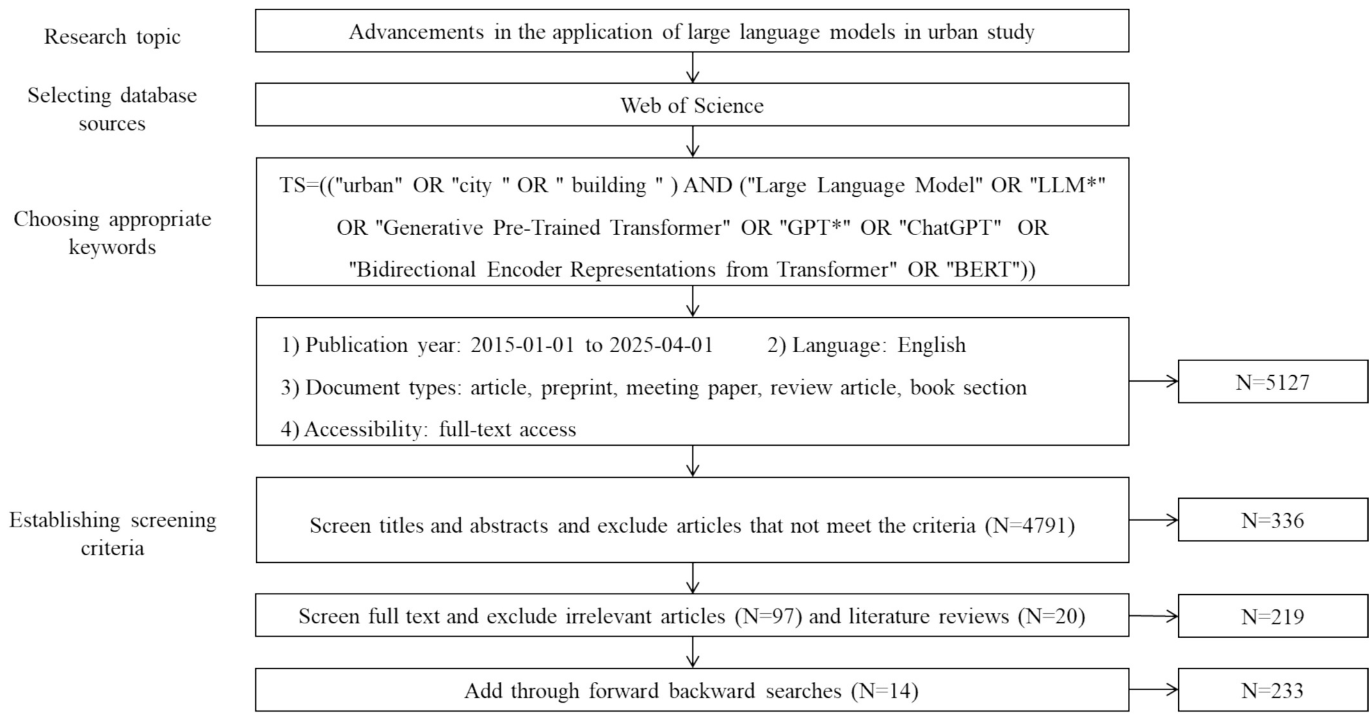

Fig. 1. Workflow of systematic literature review for application of LLMs in urban studies.

literature selected for this review should meet the following criteria:

1. It pertains to the field of urban studies. Given the dual meaning of the term “building” as both a noun (referring to a structure) and a verb (referring to the act of construction), it is anticipated that many articles in computer science domain focusing on the construction of LLMs should be excluded. To elucidate ambiguous meanings and refine the blurry boundaries of urban study, we categorized urban studies into five fields by consolidating the seven fields of study proposed by other researchers (Zou et al., 2025). This method ensures a comprehensive coverage of the core areas of urban studies, thereby addressing key theoretical and empirical aspects of the discipline. The five thematic categories are as follows.
   a) Urban sociology and economics: it encompasses branches such as sociology, examining the development patterns of urban areas and the lifestyle of urban inhabitants (Castells, 1976), and economy, analyzing the spatial arrangements of households, businesses, and capital in urban areas as well as their interactions (O'Sullivan, 2012).
   b) Urban management and public policy: it encompasses the comprehensive management measures taken by governments and non-governmental organizations to address urban and regional issues (Chakrabarty, 2001), as well as the regulations formulated for them (Wang, 2018).
   c) Urban planning and design: it is a broad term that refers to the comprehensive deployment, arrangement, and management of urban spatial layout, land use, and construction to achieve certain social and economic development goals (Hall, 2014). We restricted it to studies focusing status, features and designs related to physical or virtual urban spaces (Graham & Marvin, 2002).
   d) Urban transportation: it mainly includes the movement of people and goods within or between urban areas (Paganelli, 2019), with research on urban mobility also included.
   e) Urban environment and energy: it focuses on the production, distribution, use of energy, and its environmental impact within urban areas (Santamouris, 2013), encompassing both macro-level research like climate change and micro-level research such as perceptions of the built environment.
2. The article utilizes at least one type of LLMs in its methodology section. This step is crucial since some articles use the word “LLM” but refer to other research subjects such as “local labor market” or “local linear model”.
3. The LLMs serve as primary analytical tools or integral parts of the framework in the article. This implies that articles using LLMs solely as a comparative baseline would be excluded. Besides, it is worth noting that some articles conducting comprehensive testing of advanced LLMs are included, for their insights into the application scenarios and opinions about future development.

We firstly applied criteria by screening titles and abstracts in the initial pool and excluded 4791 articles, leaving 336 behind. After full-text screening, 97 articles failing to meet the criteria as well as 20 literature reviews were excluded. Finally, 14 additional articles were added through forward and backward citation searches, resulting in 233 eligible articles for data extraction. Notably, although we initially applied a ten-year time span for articles selecting, all eligible studies identified were published from 2020 onward, as the architectural foundation of LLMs, Transformer, was only proposed in 2017 and LLMs began to gain prominence in recent years. As a result, our review ultimately places greater emphasis on the developments of LLMs in the past six years.

2.4. Extracting structured data

Although quality assessment is critical to maintain methodological rigor (Siering et al., 2021), this step is not required in every form of literature review (Bendig & Bräunche, 2024). To accommodate the literature as much as possible, we omitted this step.

Previous scholars have applied a mixed-coding approach using a customized GPT to extract relevant data from literature with satisfactory results (Bendig & Bräunche, 2024). Drawing on their approach, we proposed a detailed analysis framework and conducted structured data extraction with the assistance of tailored GPT. As shown in Table 1, the

Table 1
Structured data extraction framework for LLM-related articles in urban study.

Dimension	Category
Field of study	Urban sociology and economics / Urban governance and public policy / Urban planning and design / Urban transportation / Urban environment and energy
Application scenario	Text classification / Topic modeling / Sentiment analysis / Scene generation / ……
Data processing	Modality alignment / Embedding
Model adaptation	Fine-tuning / Prompt engineering (Zero-shot / Few-shot) / Integration
Pre-trained model	BERT-Base / GPT-2 / GPT-3.5 / LLaMA-2 / From scratch / ……

framework consisted of five dimensions including field of study, application scenario, data processing, model adaptation and pre-trained model with several categories. Detailed descriptions of dimensions are shown in Appendix A.

We used OpenAI's GPT-4o to build an assistant named "LLMs in Urban Study: Literature Review Analyzer". Specifically, we accessed OpenAI's GPT builder platform (OpenAI, n.d.) to name the assistant and describe its intended functionalities. We meticulously incorporated the established analysis framework into detailed instructions, guiding the assistant to extract structured data from the portable document format (PDF) of each reviewed article and systematically present the results in tabular form. During interaction, the document of each article was provided manually to the assistant in the chat interface by research team and the assistant automatically returned the structured classifications. The detailed configuration of our customized GPT assistant is shown in Appendix B.

To assess the reliability of the assistant, we further conducted a human validation of its outputs. A total of 30 papers were randomly sampled from the full dataset of 233 articles, covering all five thematic categories. For each paper, expert reviewers manually evaluated the assistant's classification results across three key analytical dimensions: data processing method, model adaptation strategy, and type of pre-trained model. The dimensions of field of study and application scenario were excluded in this evaluation, as their accurate interpretation often requires domain-specific expertise knowledge. Each result was assigned a binary score based on consistency with expert judgment, and the assistant turned out to achieve an overall accuracy of 88.89 %, indicating a generally high level of agreement and applicability. Considering the alignment between expert and assistant is not flawless, we ultimately conducted a manual revision of each paper's results, according to the outputs provided by the GPT assistant.

2.5. Synthesizing analysis results

Based on the extracted data, we synthesized the applications of LLMs in different fields of urban studies. We also observed the main contributions, limitations, and future works of each article.

3. Results

The selected articles turned out to demonstrate that LLMs have been applied across various domains and scenarios, equipped with targeted data processing methods and model adaptation techniques. Hence, we first conducted a descriptive analysis from perspectives of five aspects raised above to gain a cross-domain understanding.

3.1. Descriptive analysis

Firstly, Fig. 2 illustrates the publication year of selected LLM-related articles and the number of articles in different fields of study. Articles utilizing LLMs in urban studies have emerged since 2020, with an obvious exponential growth tendency in number in the recent six years. The applications were relatively few in the early years such as 2020 and 2021 but witnessed a surge since 2023. Besides, the number of applications varies widely between different fields. Early applications focused primarily on fields like urban sociology and economics as well as urban environment and energy attribute to the abundant textual materials provided by social media data, while later LLMs were applied in domains including urban transportation, urban management and public policy, and urban planning and design, resulting in notable difference among the number of articles in different fields. The urban transportation field has the highest number of articles, three times bigger than that of urban sociology and economics.

Focusing on the specific scenarios of LLMs' applications across various fields shown in Fig. 3, it becomes evident that the applying scenarios are quite widespread. Notably, a large proportion of the applications, as expected, are related to text analysis and generation tasks such as sentiment analysis, text classification and named entity recognition. Additionally, in fields like urban transportation and urban environment with multimodal data and complex system decisions, there are also numerous field-related application scenarios including human mobility analysis, traffic flow prediction, traffic systems managing, geocoding, and street space perception, as we elaborate in the following sections. Lastly, knowledge-based Q&A are frequently mentioned in

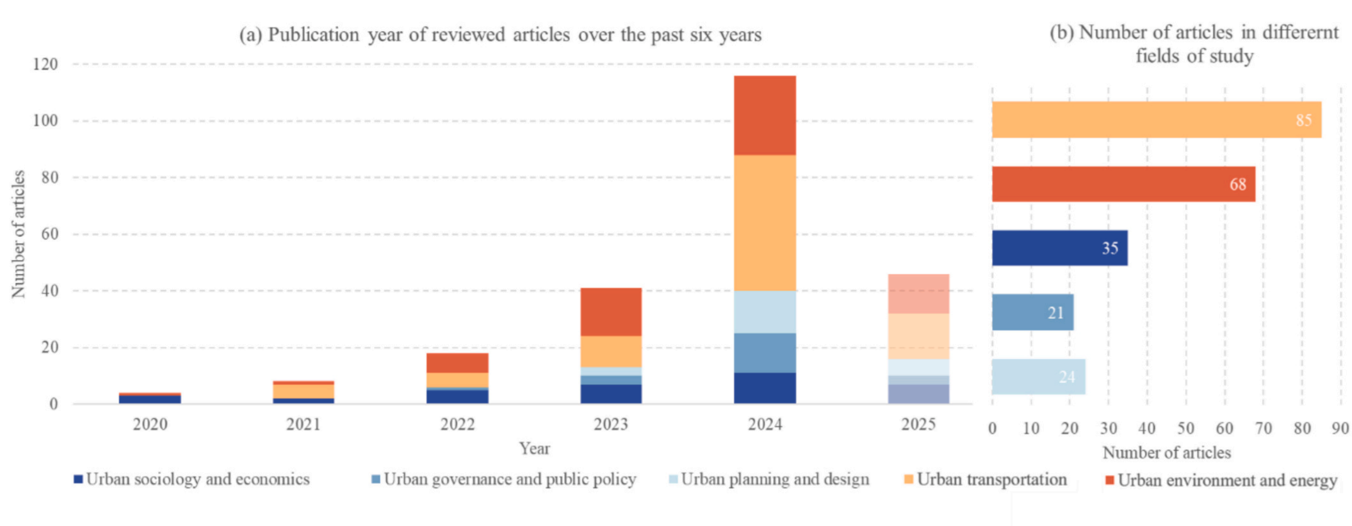

Fig. 2. Publication year of LLM-related articles and the number of articles in different fields of study.

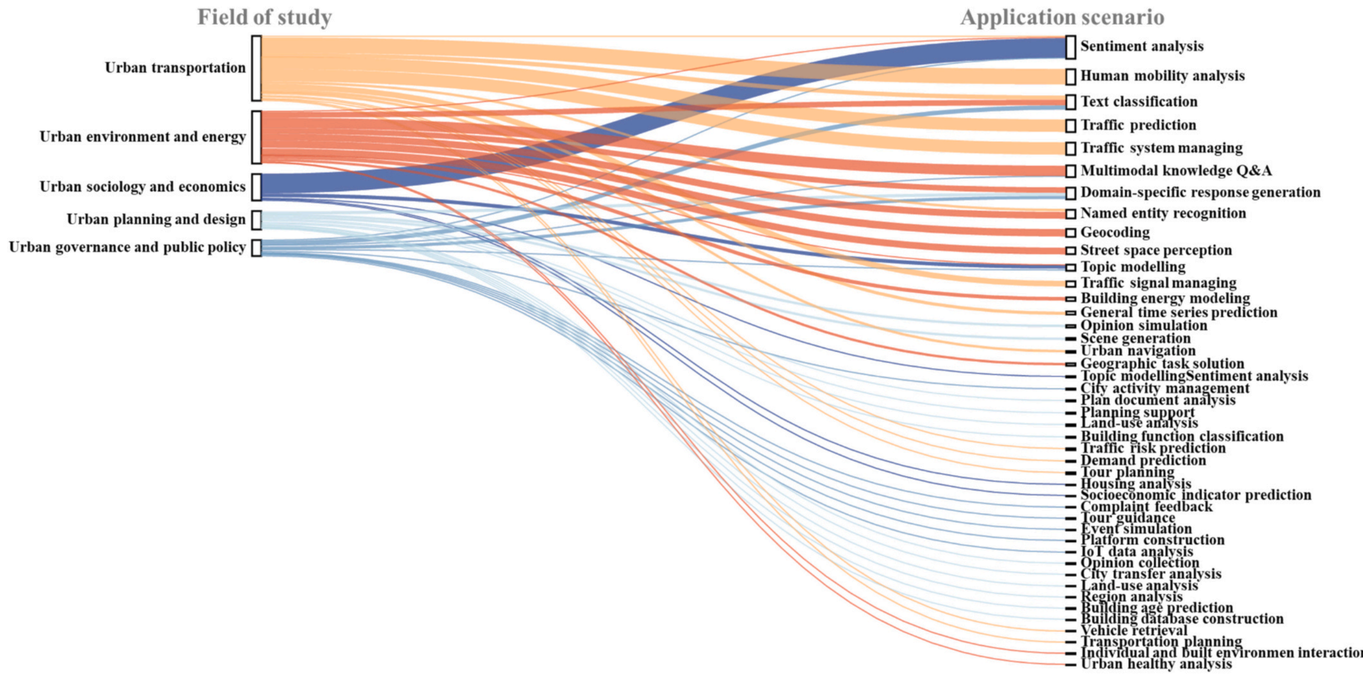

Fig. 3. Illustration of the association between urban research fields and LLMs application scenarios.

these articles as well, demonstrating widespread research interest in exploring these models' potential. This includes the behaviors that researchers testing the performance of domain-specific models they have trained themselves and the evaluations of the base models' capabilities with specialized domain questions.

Subsequently, Fig. 4 illustrates the current landscape of data processing and model adaptation methods across various fields in the selected LLM-related articles. It is revealed that LLMs are widely employed in constructing embeddings and fine-tuning to enhance their performance. Embedding construction is widely used in fields like urban sociology and economics and urban transportation, while modality alignment is predominantly applied in urban transportation as well as in urban environment and energy sectors. To our surprise, fine-tuning is adopted in more than half of the studies despite its time and money cost, and the simplest approach zero-shot has seen considerable usage as well. Besides, LLMs are often integrated with other methods rather than being used in isolation, reflecting their strong basic ability and multi-task generalization abilities.

Further, we analyzed the usage frequency of diverse pre-trained models which were used in articles, as is shown in Fig. 5. The first observation is that more than half of articles utilized GPT-based models in their research. These models not only included powerful artificial intelligence (AI) products, such as ChatGPT-3.5, ChatGPT-4, ChatGPT-4o, but also variants based on GPT-2 architecture. A considerable number of studies have used BERT-based models as pre-trained models, credit to their efficient text encoding and strong context understanding capabilities (Devlin et al., 2019). Secondly, some researchers have also applied other powerful open-source models such as LLaMA (a family of open-source LLMs released by Meta), LLaVA (a family of multimodal LLMs supporting image-text tasks), Qwen (a family of bilingual LLMs developed by Alibaba) and ChatGLM (a family of bilingual LLMs developed by Tsinghua University and Zhipu AI) as their base models to facilitate local deployment or better adaptation to specific languages (Wang, Ling, et al., 2023; Wang, Wang, et al., 2023; Wang, Zhang, et al., 2023). Another interesting phenomenon is that different research fields prefer to use different pre-trained models. For instance, research in

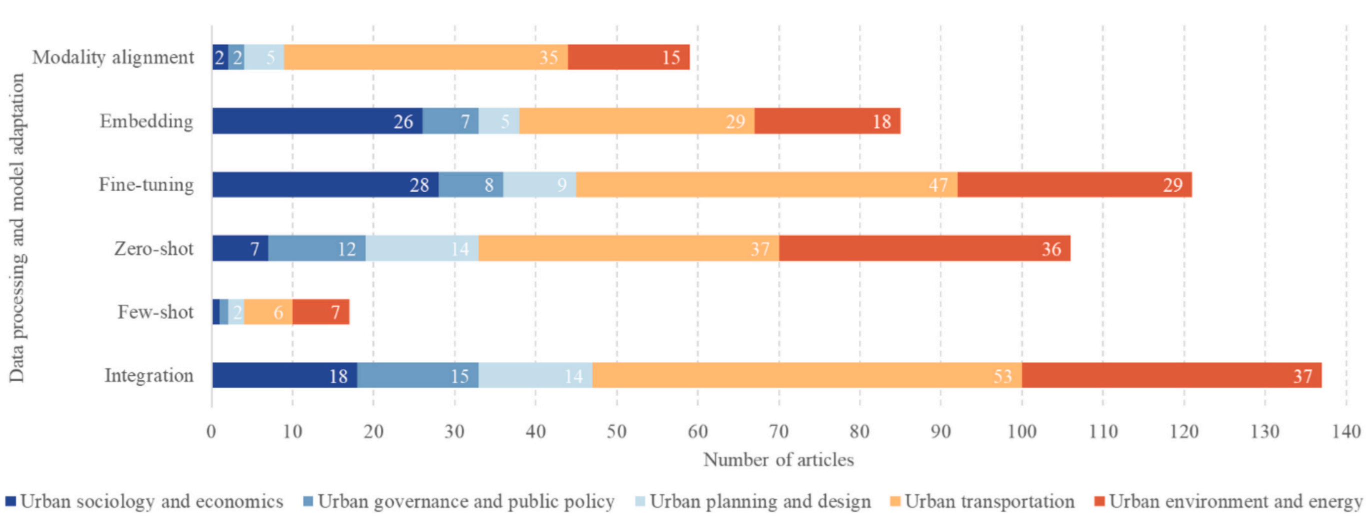

Fig. 4. Usage frequency of different data processing and model adaptation methods in LLM-related articles in urban studies.

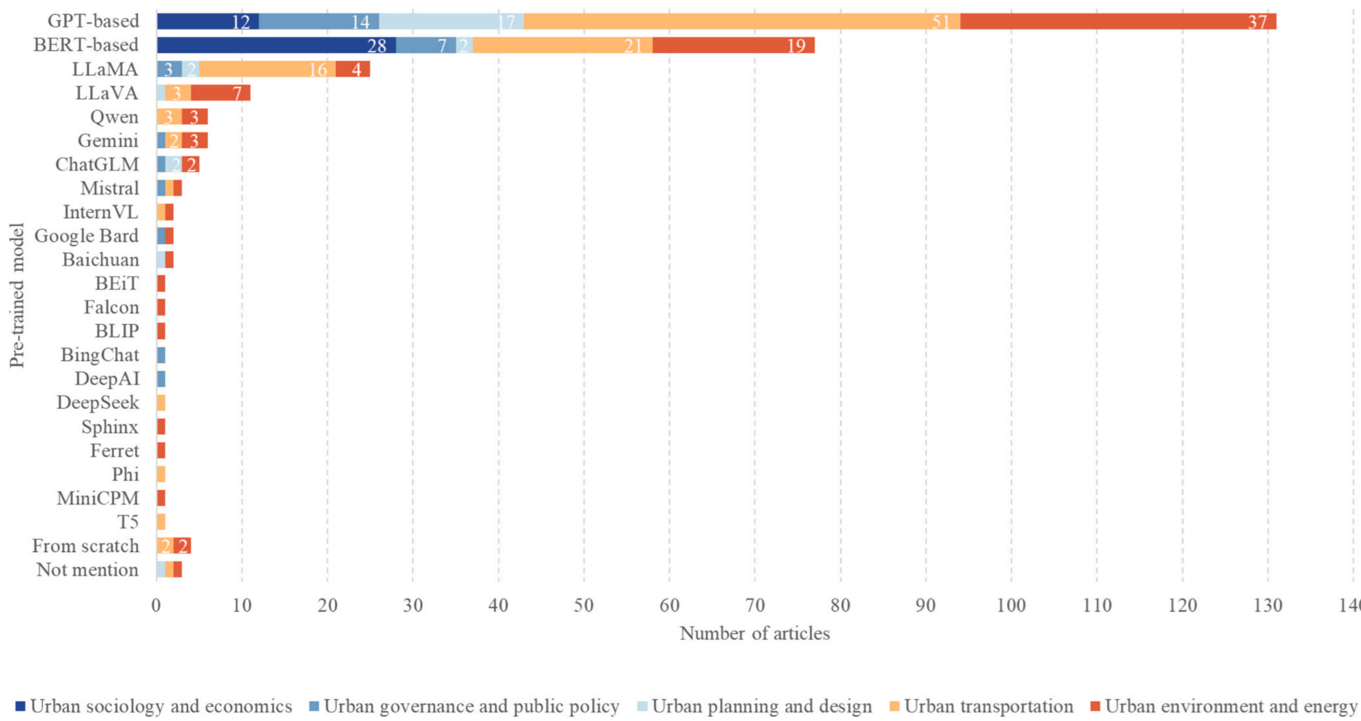

Fig. 5. Usage frequency of pre-trained models across different research fields in LLM-related articles in urban studies.

urban sociology uses BERT more than GPT, whereas in the other four fields, GPT is more prevalent. We deem that this may stem from the nature of the task characteristics typical to each field.

To elucidate the application scenarios and development trajectories of LLMs in different fields of urban studies, we broke down the statement field by field and analyzed the application progress from the perspectives of application scenarios, data processing, and model adaptations respectively.

3.2. Urban sociology and economics

Contrary to urban economics which has seen limited application of LLMs, the increasing prevalence of social platforms such as Weibo and Twitter have provided urban sociology with a wealth of accessible textual data, facilitating the early adoption of LLMs for urban text analysis. Two primary focuses within text analysis are topic modeling and sentiment analysis. Specifically, topic modeling refers to the process of identifying latent topics within a corpus of text (Papadimitriou et al., 2000), while sentiment analysis involves assessing the emotional tone of the certain text (Pang & Lee, 2008). Leveraging the powerful text encoding and contextual comprehension capabilities of BERT models, many scholars have employed them as well as their variants in research.

In the realm of topic modeling, the introduction of advanced language models has yielded innovative approaches mainly to better modeling topic representations. Zamiralov et al. (2020) utilized multi-source textual data from social media and combined guided topic models with BERT methods to construct text embeddings for text classification. Liang et al. (2021) concentrated on the efficacy of online public sentiment classification, introducing an evaluative method to determine the optimal number of topics for classification by fusing BERT models with latent dirichlet allocation (LDA) topic models. Conversely, sentiment analysis using BERT models encompassed a variety of different scenarios including attitudes towards the COVID-19 pandemic (Nwafor et al., 2021; Wang et al., 2020; Zhang, Han, et al., 2023; Zhang, Shah, & Han, 2023; Zhang, Sun, et al., 2023), opinions on electricity usage (Bedi & Toshniwal, 2022), comments on transportation (Li, Chen, et al., 2024; Li, Feng, et al., 2024; Li, Hu, et al., 2024; Li, Huang, et al., 2024; Li, Shi, & Li, 2024; Li, Tran, et al., 2024; Pratama & Akbar, 2023; Pullanikkat et al., 2024; Wu et al., 2022; (Choi et al., 2024)), attitudes towards travel (Charfaoui & Mussard, 2024) and culture perception (Qin et al., 2025). Furthermore, scholars have also endeavored to link textual sentiment with urban spaces, revealing spatial patterns inherent in the emergence of urban text. Xiao et al. (2022), for instance, applied a BERT-CNN framework to analyze the sentiment of social media data from Nanchang, China, and analyzed the spatiotemporal characteristics of textual emotions at the grid scale. Similarly, Wang, Ling, et al. (2023), Wang, Wang, et al. (2023), and Wang, Zhang, et al. (2023) utilized the a Chinese-adapted BERT model (BERT-Base-Chinese) to sort extensive social media data from Wuhan, China, into five emotion categories and applied k-means clustering to uncover six spatial co-occurrence patterns within urban settings.

A notable development in the application of LLMs was the emergence of models tailored to languages other than English due to the inherent differences in information density and grammatical structures. This has resulted in models such as BERT-Base-Chinese for Chinese (Wang, Ling, et al., 2023; Wang, Wang, et al., 2023; Wang, Zhang, et al., 2023), BERTurk for Turkish (Zengin et al., 2022), IndoBERT for Indonesian (Pratama & Akbar, 2023), BERTje for Dutch (Veltmeijer & Gerritsen, 2023), and AraBERT for Arabic (Jannani et al., 2025), each designed to fit their unique linguistic characteristics. These specialized models underscored the importance of linguistic adaptability in the effective deployment of LLMs for sentiment analysis and topic modeling.

With the continuous development of decoding technologies, scholars have increasingly applied GPT models with larger parameter sizes (e.g., GPT-3.5) to research scenarios. In topic modeling, Chen (2024) utilized the GPT-3.5 application programming interface (API) to achieve topic classification of text regarding various evaluation indicators of urban green spaces while Cusumano et al. (2025) leverage GPT-3.5 Turbo for knowledge discovery from building defects data. Besides, Ramos-Henriquez and Morini-Marrero (2025) found that GPT-4 was valuable for extracting data and assigning topic labels from unstructured text. In sentiment analysis, Saito and Haruyama (2023) compared the capabilities of fine-tuned BERT and GPT-3 for sentiment analysis of COVID-19 pandemic texts, confirming GPT-3's superiority in tweet sentiment classification, and some articles apply GPT-3.5 to understand people's responses to natural disasters and COVID-19 on social media (Li, Su, et al., 2025; Li, Yang, et al., 2025; Li, Zhao, et al., 2025; Saito & Tsugawa, 2025). Moreover, experiments on the performance of GPT-4 in sentiment classification were conducted, revealing that despite the imperfections, GPT-4 could significantly save time for city planners processing public feedback textual data (Fu, Sanchez, et al., 2024; Fu, Wang, & Li, 2024).

In the domain of urban economics, studies employing LLMs remain scarce. Existing research primarily utilizes LLMs to accelerate analyzing workflows or to predict socioeconomic indicators. Specifically, Hayden (2023) applied GPT-4 to generate code and analyze correlations in the context of housing studies, while Hao et al. (2025) developed UrbanVLP, a model integrating multi-granular information from both macro-level and micro-level imagery to predict urban socioeconomic indicators.

Table 2 clearly presents the application scenarios, data processing, model applications, and the pre-trained models used in each article within urban sociology domain. Functioning as an encoder module, BERT models were primarily employed to construct embeddings for downstream tasks and were frequently integrated with other neural networks (such as CNN, deep convolutional neural networks (DCNN), bidirectional long short-term memory (BiLSTM), etc.) or with methods like LDA for semantic proximity analysis (Liang et al., 2021). Besides, most researchers engaged in fine-tuning the pre-trained BERT model with specific datasets to sharpen its performance on these downstream tasks, contrast to the typical employment of zero-shot or few-shot prompting methods on GPT-series models with larger parameters.

3.3. *Urban management and public policy*

The sphere of urban management and public policy mirrors urban sociology in its wealth of textual data, such as guidelines and policies, which offers fertile ground for the application of LLMs. Similar to urban sociology, the processing of texts often employed BERT models for text encoding and downstream task classification (Giménez Manuel et al., 2024; Hu et al., 2024; Li, Chen, et al., 2024; Li, Feng, et al., 2024; Li, Hu, et al., 2024; Li, Huang, et al., 2024; Li, Shi, & Li, 2024; Li, Tran, et al., 2024; Lim & Hwang, 2024; Prabhune et al., 2024; Wang et al., 2022) or

Table 2
Summary of structured information from LLM-related articles in urban sociology and economics.

Article	Year	Application scenario	Data processing		Model adaptation				Pre-trained model
			Modality alignment	Embedding	Fine-tuning	Prompt Engineering		Integration	
						Zero-shot	Few-shot		
Zamiralov et al.	2020	Topic modeling		√	√			√	RuBERT
Zhou et al.	2020	Topic modeling		√	√			√	BERT-WWM
Liang et al.	2021	Topic modeling		√				√	BERT-Base
Chen et al.	2024	Topic modeling				√			GPT-3.5
Ramos-Henriquez et al.	2025	Topic modeling				√			GPT-4
Fu et al.	2024	Topic modeling Sentiment analysis				√	√		GPT-4
Jannani et al.	2025	Topic modeling Sentiment analysis		√	√				AraBERT
Wang et al.	2020	Sentiment analysis		√	√			√	BERT-Base-Chinese
Nwafor et al.	2021	Sentiment analysis		√	√				COVID-BERT-v2
Bedi et al.	2022	Sentiment analysis		√	√			√	BERT-Base
Jain et al.	2022	Sentiment analysis		√	√			√	BERT-Base
Wu et al.	2022	Sentiment analysis		√	√			√	BERT-Base
Xiao et al.	2022	Sentiment analysis		√	√			√	MacBERT
Zengin et al.	2022	Sentiment analysis		√	√				BERTurk
Pratama et al.	2023	Sentiment analysis		√	√				IndoBERT
Saito et al.	2023	Sentiment analysis		√	√				BERT-Base GPT-3
Tas et al.	2023	Sentiment analysis		√	√			√	BERT-Base
Veltmeijer et al.	2023	Sentiment analysis		√	√			√	BERTje
Wang et al.	2023	Sentiment analysis		√	√			√	BERT-Base-Chinese
Zhang et al.	2023	Sentiment analysis		√	√			√	BERT-Base
Li et al.	2024	Sentiment analysis		√	√				BERT-Base
Li et al.	2024	Sentiment analysis		√	√			√	BERT-Base RoBERTa
Pullanikkat et al.	2024	Sentiment analysis		√	√	√			RoBERTa GPT-3.5
Wen et al.	2024	Sentiment analysis		√	√			√	BERT-Base
Yang et al.	2024	Sentiment analysis		√	√			√	BERT-Base
Charfaoui et al.	2024	Sentiment analysis		√	√				SieBERT-Marrakech
Choi et al.	2024	Sentiment analysis		√	√				BERT-Base
Kharlamov et al.	2024	Sentiment analysis			√				GPT-3.5 GPT-4 GPT-4o
Liu et al.	2024	Sentiment analysis		√	√				Bert-large-Japanese
Aman et al.	2025	Sentiment analysis	√			√		√	BERT-Base
Li et al.	2025	Sentiment analysis				√		√	GPT-3.5
Qin et al.	2025	Sentiment analysis		√	√				BERT-Base
Saito et al.	2025	Sentiment analysis			√				GPT-3.5 Turbo
Hayden	2023	Housing analysis				√			GPT-4
Hao et al.	2025	Socioeconomic indicator prediction	√		√			√	ShareGPT4V

directly leveraged models like GPT or ChatGLM to enhance the performance (Wang, Liu, & Zhou, 2024; Wang, Pang, et al., 2024; Wang, Zhao, et al., 2024; Zenkert & Fathi, 2024; Zhou & Hu, 2025). However, the difference between the two domains' application lied in the usage of powerful text generation capabilities of GPT. For instance, Chen, Liang, et al. (2024) and Chen, Zhang, et al. (2024) proposed Chat3D, a framework that utilizes LLMs to generate reports for urban ecological construction by analyzing 3D scene-level point clouds, while Karren et al. (2024) explored the use of GPT-3.5 to enhance citizen evaluations of urban hygiene feedback systems. Yao (2023) employed GPT-4 to offer personalized travel suggestions by documenting the positional relationships between attractions, showcasing the tailored application of LLMs in urban management. Beyond these scenarios, text generation also extended to domain-specific response generation, including civic spaces construction (Chauncey & McKenna, 2024), real estate opinion (Madani et al., 2024), cultural heritage management (Spennemann, 2024) and culture background understanding (Campo-Ruiz, 2025).

Besides, the emergence of more comprehensive models like GPT-4 and Gemini has led to an increasing trend of employing LLMs as intelligent agents in urban management and public policy. LLMs as agents refers to entities that exhibit intelligent behavior and possess traits such as autonomy, reactivity, proactivity, and social ability (Xi et al., 2023). Xiao et al. (2023) were among the first to use GPT-3.5 as the brain of a generative agent (GA) with perception and decision-making modules to simulate the spread of water pollution events in a simulated town and examined government responses to the specific event. Additionally, the self-developed framework like CityGPT introduced multiple LLM agents to analyze internet of things (IoT) spatiotemporal data catering to diverse urban analysis requirements (Guan et al., 2024). This sophisticated approach had enabled the research team to develop a comprehensive platform for urban generative intelligence (UGI) (Xu et al., 2023). Besides, LLMs are also employed to enhance the management of urban activities and strengthen smart city decision support (Jiang et al., 2024; Kalyuzhnaya et al., 2025).

As shown in Table 3, the application of data processing methods in this filed remains close to urban sociology, with continuing use of BERT for text embedding or the employment of prompt engineering techniques to optimize inputs for pre-trained models (Amatriain, 2024). The need to align heterogeneous data such as geographical information data and spatiotemporal data remained scarce. However, in contrast to urban sociology, the integration approaches of LLMs in urban management go a step further by incorporating various functional blocks within platform systems. For instance, the UGI platform integrated the open digital infrastructure, user interface, and generative intelligence (Xu et al., 2023) in system framework levels, while tourism guidance systems have tried to integrate LLMs with diffusion models for ideal tour picture generation (Yao, 2023).

3.4. Urban planning and design

LLMs have revolutionized the field of urban planning and design by permeating every stage of the process, from initial opinion collection and plan documentation to mid-stage planning support and urban scene creation, and finally to urban analysis and the deployment of LLMs for domain-specific response generation. The collation of resident opinions is a crucial aspect of preliminary research. LLMs could promote the effectiveness of design discussions in large projects in the collection of opinions, demonstrated by Dortheimer et al. (2024) with the development of GPT-3-based chatbot to collect community input for construction projects. Besides, LLMs showed their potential in resident opinions

Table 3
Summary of structured information from LLM-related articles in urban governance and public policy.

Article	Year	Application scenario	Data processing		Model adaptation				Pre-trained model
			Modality alignment	Embedding	Fine-tuning	Prompt Engineering		Integration	
						Zero-shot	Few-shot		
Lim et al.	2024	Topic modeling		√	√			√	BERT-Base
Manuel et al.	2024	Topic modeling		√	√			√	Sentence-BERT
Prabhune et al.	2024	Sentiment analysis		√	√			√	BERT-Base
Wang et al.	2022	Text classification		√	√			√	BERT-Base
Hu et al.	2024	Text classification	√	√	√			√	BERT-Base
Li et al.	2024	Text classification		√	√			√	BERT-Base
Wang et al.	2024	Text classification				√			GPT-4
Zenkert et al.	2024	Text classification				√		√	GPT-4
Zhou et al.	2025	Text classification				√		√	ChatGLM-2
Chen et al.	2024	Multimodal knowledge Q&A				√		√	GPT-3.5 New Bing Gemini
Karren et al.	2024	Complaint feedback				√		√	GPT-3.5
Yao	2023	Tour guidance					√	√	GPT-4
Chauncey et al.	2024	Domain-specific response generation				√			GPT-3.5 GPT-4
Madani et al.	2024	Domain-specific response generation			√				LLaMA-3-8B-instruct
Spennemann	2024	Domain-specific response generation				√			GPT-3.5 BingChat Google Bard DeepAI
Campo-Ruiz	2025	Domain-specific response generation				√			GPT-4o
Xiao et al.	2023	Event simulation		√		√		√	BERT GPT-3.5
Xu et al.	2023	Platform construction							CityGPT
Guan et al.	2024	IoT data analysis	√			√		√	GPT-4
Jiang et al.	2024	City activity management			√	√		√	GPT-4 LLaMA-2-7B
Kalyuzhnaya et al.	2025	City activity management				√		√	GPT-4o Mistral-8x22B LLaMA-3.1

simulation. For example, Malekzadeh et al. (2024) compared the evaluation of the visual appeal of urban environments between GPT-4-based agents and residents, and Yu and McKinley (2024) proposed the creation of digital avatars representing various stakeholders using LLMs to synthesize the practical pathways for shared automated electric transportation systems in urban planning. Zhou et al. (2024) and Zhou, Wang, et al. (2024) further designed resident agents and constructed a framework for role-playing, collaborative generation, and feedback iteration, which aided in scheme generation and iteration. These studies collectively endorsed the value of LLMs in providing insights and enhancing the inclusivity and interpretability of planning in a cost-effective manner (Malekzadeh et al., 2024; Yu & McKinley, 2024). In addition, LLMs also play a pivotal role in plan documentation, another necessary task in the initial stages of urban planning. Fu, Sanchez, et al. (2024) and Fu, Wang, and Li (2024) used ChatGPT for plan quality evaluation and found the good agreement with the manual evaluation results, and specialized models like PlanGPT, introduced by Zhu et al. (2024), have streamlined challenges for urban planners by offering a framework that includes retrieval, fine-tuning, and the utilization of advanced functional tools.

In the mid-stage of urban planning and design, LLMs were applied in planning support and urban scene generation. Recent studies explored the integration of LLMs into planning support systems (PSS), which refer to a set of tools built upon geographic information technologies that support the urban planning process. Jiang et al. (2025) incorporated GPT-4 into a PSS framework to enhance its performance in data processing, idea generation, and decision support. Similarly, Quan and Lee (2025) embedded GPT-4o into a web-based PSS platform to facilitate stakeholder communication and elicit public contributions to sustainable urban development initiatives. Urban scene generation means constructing realistic three-dimensional representations of urban settings, usually following the steps of task scenario analysis, two-dimensional plane generation, digital asset placement and scenario auxiliary refinement (Shang et al., 2024) by using multimodal large language models (Yin et al., 2024). In the process of scene generation, these models could play a human-like role in automatically drafting detailed urban scene descriptions (Shang et al., 2024), extracting contextual information from model-generated planar buildings (Deng et al., 2024), formulating land use plans (Deng et al., 2024), completing model construction tasks by retrieval (Zhou, Lin, & Li, 2024; Zhou, Wang, et al., 2024) as well as realizing results iteration by identifying discrepancies between scene images and ideal plans (Shang et al., 2024).

Furthermore, conducting urban analysis and knowledge Q&A are also significant application scenarios in the field of urban planning and design. Urban analysis scenarios include resource-based city transfer (Yu, Xing, et al., 2024; Yu, Yang, et al., 2024), land-use analysis (Çalışkan, 2025; Cao et al., 2024; Guo et al., 2024), region analysis (Yan, Mai, et al., 2024; Yan, Wen, et al., 2024), building age prediction (Zeng, Goo, et al., 2024; Zeng, Yang, et al., 2024), building function classification (Memduhoğlu et al., 2024; Ramalingam & Kumar, 2025) as well as building database construction (Li, Su, et al., 2025; Li, Yang, et al., 2025; Li, Zhao, et al., 2025). Furthermore, specialized LLMs were fine-tuned for domain-specific response generation and question-and-answering. Theses vertical domain LLMs include HSC-GPT, a fine-tuned GPT-based model for human settlement construction tasks (Ran et al., 2023), LLMs for urban renewal (Wang, Ling, et al., 2023; Wang, Wang, et al., 2023; Wang, Zhang, et al., 2023), LLMs for sustainable community assessment (Jonveaux, 2024) and CityEQA, a vertical-domain LLM designed for embodied question answering in urban space (Zhao et al., 2025).

Given the complex of preliminary research and the diverse stakeholder interests in urban planning and design, there is a preference for using pre-trained larger-parameter models with comprehensive performance such as GPT-3, GPT-4, GPT-4o, ChatGLM-3, LLaMA-2, as is reflected in Table 4. More than half of the models used training-free methods to accomplish urban planning tasks and were usually

Table 4
Summary of structured information from LLM-related articles in urban planning and design.

Article	Year	Application scenario	Data processing		Model adaptation				Pre-trained model
			Modality alignment	Embedding	Fine-tuning	Prompt Engineering		Integration	
						Zero-shot	Few-shot		
Zhou et al.	2024	Opinion simulation				√		√	GPT-3
Yu et al.	2024	Opinion simulation				√			GPT-3.5
Malekzadeh et al.	2024	Opinion simulation				√			GPT-4
Dortheimer et al.	2024	Opinion collection			√			√	GPT-3
Fu et al.	2023	Plan document analysis				√			GPT-3.5
Zhu et al.	2024	Plan document analysis			√				ChatGLM-3
Jiang et al.	2025	Planning support				√			GPT-4
Quan et al.	2025	Planning support					√	√	GPT-4o
Zhou et al.	2024	Scene generation	√			√		√	GPT-4
Deng et al.	2024	Scene generation		√		√		√	GPT-4v
Shang et al.	2024	Scene generation			√			√	LLaVA-1.5
Yu et al.	2024	City transfer analysis				√		√	GPT-3.5
Cao et al.	2024	Land-use analysis	√	√	√				BERT-Base
Guo et al.	2024	Land-use analysis	√	√	√			√	BERT-Base
Caliskan	2025	Land-use analysis				√		√	GPT-4
Yan et al.	2024	Region analysis	√	√	√				LLaMA-2
Zeng et al.	2024	Building age prediction				√			GPT-4
Memduhoglu et al.	2024	Building function classification		√		√		√	GTE-Large
Ramalingam et al.	2025	Building function classification	√		√		√	√	GPT-3.5 Turbo
Li et al.	2025	Building database construction				√			GPT-4o
Wang et al.	2023	Domain-specific response generation			√				ChatGLM-3
Chen et al.	2023	Domain-specific response generation			√			√	LLaMA-2 Baichuan-13B
Jonveaux	2024	Domain-specific response generation				√		√	GPT-3.5 Turbo
Zhao et al.	2025	Domain-specific response generation				√		√	GPT-4 GPT-4o

integrated with other functionalities such as API document calling and map generation, aiming to understand planning terminologies and tackle challenging planning tasks professionally (Zhu et al., 2024).

3.5. Urban transportation

In urban transportation sector characterized by substantial demands for natural language processing and time-series data analysis, LLMs have been extensively applied as well. The increasing complexity of urban transportation systems, coupled with the emergence of diverse mobility demands and management challenges, has driven the evolution of models from traditional statistical learning to deep learning, and more recently to LLMs (Jin et al., 2021).

In early studies of urban transportation, BERT was instrumental in extracting textual information from traffic flow data to accomplish specific tasks, such as sentiment analysis (Fontes et al., 2023; Wang, Liu, & Zhou, 2024; Wang, Zhao, et al., 2024), resident activity classification (Liu et al., 2021), traffic accident classification (Salam et al., 2021; Yuan et al., 2024) and named entity recognition (Li, Shi, & Li, 2024; Patel & Nyirenda, 2022; Zuo & Li, 2021). Besides, BERT models have been incorporated into models for tasks such as predicting traffic flow and trajectory imputation, forming the architecture including BDSTN (Cao et al., 2022), Kamel (Musleh & Mokbel, 2023), BTtrajRec (Shi et al., 2023), and Traffic BERT (Wang, Ling, et al., 2023; Wang, Wang, et al., 2023; Wang, Zhang, et al., 2023).

Contrasting with BERT's encoder design, GPT models featuring in an optimized decoder structure for sequential token generation were widely used in the field of urban transportation. For example, Xue et al. (2022) constructed the AuxMobLCast pipeline, employing both BERT encoding and GPT-2 decoding to address the fusion of semantic and numerical information in point of interest (POI) passenger flow forecasting. Liu, Hu, et al. (2024) and Liu, Yang, et al. (2024) trained GPT-2 with partially frozen attention (PFA) strategy to capture spatiotemporal correlations in encoded embeddings. Additionally, architectures such as LLM4TS (Chang et al., 2024), GATGPT (Chen, Chen, et al., 2023; Chen, Hu, et al., 2023; Chen, Wang, & Xu, 2023), STD-LLM (Huang et al., 2024), Time-LLM (Jin et al., 2024), UniTime (Liu, Hu, et al., 2024; Liu, Yang, et al., 2024) and TPLLM (Ren et al., 2024) have also demonstrated how to leverage and fine-tune GPT-2 on sequential data to improve forecasting accuracy.

Furthermore, human-mobility analysis is a highly active topic in urban transportation domain, and numerous studies have begun leveraging LLM-based methods. For example, Terashima et al. (2023) employed a spatio-temporal BERT model for next-location prediction, while Tang et al. (2024) demonstrated that an instruction-tuned Llama-3-8B model performs remarkably well in city-scale human mobility forecasting. Other related studies include GeoFormer (Solatorio, 2023), MobilityGPT (Haydari et al., 2024), ST-MoE-BERT (He et al., 2024), Region-as-Walker (Zhang, Han, et al., 2023; Zhang, Shah, & Han, 2023; Zhang, Sun, et al., 2023), LIMP (Li, Chen, et al., 2024; Li, Feng, et al., 2024; Li, Hu, et al., 2024; Li, Huang, et al., 2024; Li, Shi, & Li, 2024; Li, Tran, et al., 2024), TSI-LLM (Luo et al., 2024), TrajLLM (Ju et al., 2025), and Geo-Llama (Li, Chen, et al., 2024; Li, Feng, et al., 2024; Li, Hu, et al., 2024; Li, Huang, et al., 2024; Li, Shi, & Li, 2024; Li, Tran, et al., 2024).

Besides, LLMs agent also emerged in more complex transportation scenarios. Traffic system management and traffic signal control are two widely applied occasions. In traffic system management, Jonnala et al. (2024) introduced GPT-3.5 into the urban public transportation systems management to support managerial decision-making. Beyond this exemplar, researchers have introduced a series of dedicated LLM-agents including GARLIC (Han, Zhang, et al., 2024), CoMAL (Yao et al., 2025), LangGraph-Traffic Agent (Chen & Ding, 2025) to tackle issues like network monitoring, vehicle dispatch and mixed-autonomy coordination. In the domain of traffic signal control, Lai et al. (2024) proposed the LLMLight architecture for optimizing traffic signal control, and other searches contains LLM-Assisted Light (Wang, Liu, & Zhou, 2024; Wang, Pang, et al., 2024; Wang, Zhao, et al., 2024) and CoLLMLight (Yuan et al., 2025). Subsequently, with the desire to involve image understanding and real-time intervention in autonomous driving, vision-language models (VLMs) caught the attention of researchers. These models were employed to assist intelligent traffic decision support (De Zarzà et al., 2023; (Onsu et al., 2025)) or in urban delivery scenes for urban navigation (Alhammadi et al., 2025; Zeng, Goo, et al., 2024; Zeng, Yang, et al., 2024; Zhang, Li, et al., 2024; Zhang, Liu, et al., 2024; Zhang, Sun, et al., 2024) but still in its infancy due to the limited visual understanding abilities (De Zarzà et al., 2023).

In terms of data processing, the characteristics of applications in urban transportation field share similarities with other domains, particularly in BERT for text encoding and LLM-agents for complex tasks. However, one of the unique challenges shown in Table 5 is modal alignment, originating from the proliferation of sensors and the accumulation of heterogeneous data collected form urban environment (Chen et al., 2022). Traditional methods used deep neural networks such as region-based convolutional neural network (RCNN) and residual network (ResNet) (Scribano et al., 2021) or graph attention networks (Chen, Chen, et al., 2023; Chen, Hu, et al., 2023; Chen, Wang, & Xu, 2023) to extract data features, while with the help of LLMs, there were approaches to carefully designing methods to convert data such as POI into language modal and further utilizing BERT for encoding (Xue et al., 2022). Additionally, the flexibility of designing modules in LLM's architecture was another characteristic of the applications in urban transportation. In contrast to the first three domains, which typically employed large models as a whole, researched in urban transportation constructed innovative architecture based on a LLM framework (usually GPT-2) to achieve optimal task performance (Liu, Hu, et al., 2024; Liu, Yang, et al., 2024; Yuan et al., 2024). However, this aspect extends into the technical realm of model architecture design, which falls beyond the scope of the present discussion.

3.6. Urban environment and energy

The urban environment and energy sectors also boast a plethora of data sources including remote sensing imagery, street view images and geographic locations etc. Similar to the field of urban transportation, the trajectory of LLMs' application follows the path of text analysis, multimodal integration and agent construction.

BERT-based models, as always, were initially employed for processing unstructured text data as encoders. For instance, Xu et al. (2020) proposed a Geographical Spatial Address Model (GSAM) incorporating BERT to handle non-standardized addresses. Ali and Shirazi (2022) utilized BERT for electronic waste policy analysis, revealing the value of LLMs in enhancing the efficiency of electronic waste management. Besides, the analysis of social media data about environment and energy usage was also a focal point of early research, particularly in detecting and analyzing disaster locations and severity levels of floods and urban waterlogging (Chen, Chen, et al., 2023; Chen, Hu, et al., 2023; Chen, Wang, & Xu, 2023; Fu et al., 2025; Y. Han et al., 2025; Kanth et al., 2022; Mo et al., 2025; Zhang et al., 2021).

As research delves deeper, it has gradually shifted towards geospatial representation learning and the integration of multimodal data. POI, widely used in urban environment researches, has gained widespread attention in geospatial representation learning. Methods included Geo-Semantic2vec algorithm to extract POI features for downstream task analysis by Zhang et al. (2021), GeoBERT achieving precise embedding through large-scale pre-training and demonstrating robust performance in classification and prediction tasks Gao et al. (2022), or larger models (Sun et al., 2025). Besides, aligning image with text modalities were explored through leveraging VLMs as well. Hu et al. (2023) constructed a high-quality remote sensing image description dataset (RSICap) and developed the RSGPT for remote sensing images understanding. Kuckreja et al. (2023) proposed GeoChat, a VLM to realize functions including image region annotation, visual question answering, and

Table 5
Summary of structured information from LLM-related articles in urban transportation.

Article	Year	Application scenario	Data processing		Model adaptation				Pre-trained model
			Modality alignment	Embedding	Fine-tuning	Prompt Engineering		Integration	
						Zero-shot	Few-shot		
Fontes et al.	2023	Sentiment analysis		√		√		√	BERT-Base
Wang et al.	2024	Sentiment analysis		√	√				BERT-finetuned-weibo
Liu et al.	2021	Text classification		√	√			√	BERT-Base
Salam et al.	2021	Text classification		√	√			√	BERT-Base
Rath et al.	2022	Text classification		√	√				Sentence-BERT
Yuan et al.	2022	Text classification	√	√	√			√	BERT-Base
Ou et al.	2024	Text classification	√	√	√			√	BERT-Base
Gundidza et al.	2025	Text classification			√				GPT-4
Zuo et al.	2021	Named entity recognition		√	√			√	BERT-Base
Patel et al.	2022	Named entity recognition		√	√				BERT-Base
Li et al.	2024	Named entity recognition		√	√			√	BERT-Base
Jin et al.	2021	Traffic prediction	√		√			√	BERT-Base
Cao et al.	2022	Traffic prediction	√	√	√			√	BERT-Base
Chang et al.	2023	Traffic prediction	√		√			√	GPT-2
Musleh et al.	2023	Traffic prediction	√	√	√			√	BERT-Base
Shi et al.	2023	Traffic prediction	√	√	√			√	BERT-Base
Wang et al.	2023	Traffic prediction	√	√	√				BERT-Base
Chen et al.	2024	Traffic prediction				√			GPT-4
Guo et al.	2024	Traffic prediction	√		√				LLaMA-2
Huang et al.	2024	Traffic prediction	√	√	√			√	GPT-2
Li et al.	2024	Traffic prediction				√		√	LLaMA-3.1
Liu et al.	2024	Traffic prediction	√	√	√			√	GPT-2 LLaMA-2
Ren et al.	2024	Traffic prediction	√		√			√	GPT-2
Rong et al.	2024	Traffic prediction	√		√			√	GPT-2
Yu et al.	2024	Traffic prediction			√				LLaMA-2
Yuan et al.	2024	Traffic prediction		√	√	√	√		From scratch
Zhang et al.	2024	Traffic prediction	√		√			√	GPT-3.5
Chen et al.	2024	General time series prediction		√	√			√	GPT-2
Jin et al.	2024	General time series prediction	√		√				GPT-2 LLaMA-2
Li et al.	2024	General time series prediction		√	√				LLaMA-2
Liu et al.	2024	General time series prediction	√			√	√		GPT-2
Xue et al.	2022	Human mobility analysis	√	√	√			√	BERT-Base GPT-2
Mo et al.	2023	Human mobility analysis	√			√			GPT-3.5
Solatorio	2023	Human mobility analysis	√		√				GPT-3.5
Terashima et al.	2023	Human mobility analysis	√		√			√	BERT-Base
Zhang et al.	2023	Human mobility analysis	√	√	√			√	GPT-3.5
Bhandari et al.	2024	Human mobility analysis			√				LLaMA-2
Feng et al.	2024	Human mobility analysis				√			GPT-3.5
Haydari et al.	2024	Human mobility analysis	√		√			√	From scratch
He et al.	2024	Human mobility analysis	√		√			√	BERT
Li et al.	2024	Human mobility analysis		√		√			GPT-4 DeepSeek
Li et al.	2024	Human mobility analysis	√		√				LLaMA-2
Li et al.	2024	Human mobility analysis	√		√				LLaMA-3 GPT-4o
Liang et al.	2024	Human mobility analysis				√		√	GPT-4
Luo et al.	2024	Human mobility analysis					√		GPT-4
Tang et al.	2024	Human mobility analysis			√				LLaMA-3
Wang et al.	2024	Human mobility analysis	√			√		√	GPT-3.5
Wang et al.	2024	Human mobility analysis				√			GPT-3.5
Du et al.	2025	Human mobility analysis				√		√	Qwen2 Mistral7B-V3 LLaMA-3 Gemma-2 GPT-3.5 Turbo GPT-4o-mini
Ju et al.	2025	Human mobility analysis				√		√	LLaMA-3.1-8B-Instruct GPT-4o-mini
Li et al.	2025	Human mobility analysis		√		√		√	BERT-Base
de Zarza et al.	2023	Traffic risk prediction	√			√		√	LLaMA-2 LLaVA-13B
Costa et al.	2024	Traffic risk prediction				√		√	GPT-3.5 Turbo

(continued on next page)

Table 5 (*continued*)

Article	Year	Application scenario	Data processing		Model adaptation				Pre-trained model
			Modality alignment	Embedding	Fine-tuning	Prompt Engineering		Integration	
						Zero-shot	Few-shot		
Nie et al.	2024	Demand prediction		√		√		√	GPT-4
Qu et al.	2024	Demand prediction			√		√	√	Sentence-T5
Tang et al.	2024	Tour planning	√			√		√	GPT-4 GPT-3.5
Oppenlaender	2025	Tour planning				√			GPT-3.5
Masri et al.	2024	Traffic system managing			√			√	GPT-4o Gemini-1.0 Gemini-1.5 LLaMA-3.1
Nawar et al.	2023	Traffic system managing	√	√	√			√	GPT-3.5
Han et al.	2024	Traffic system managing	√			√		√	GPT-2
Jin et al.	2024	Traffic system managing				√		√	GPT-4
Jonnala et al.	2024	Traffic system managing				√			GPT-3.5 GPT-4
Masri et al.	2024	Traffic system managing				√			GPT-4o-mini
Tupayachi et al.	2024	Traffic system managing				√		√	GPT-4
Tupayachi et al.	2024	Traffic system managing				√		√	GPT-4
Wang et al.	2024	Traffic system managing				√		√	GPT-3.5 Turbo
Zhang et al.	2024	Traffic system managing		√		√			GPT-3.5
Chen et al.	2025	Traffic system managing				√		√	GPT-3.5 Turbo
Ding et al.	2025	Traffic system managing				√		√	InternVL2.5
Kalaiselvi et al.	2025	Traffic system managing		√	√				BERT-Base
Masri et al.	2025	Traffic system managing				√		√	GPT-4o
Tami et al.	2025	Traffic system managing	√		√				Qwen2-VL-2B
Yao et al.	2025	Traffic system managing					√	√	GPT-4o-mini Qwen-72B/32B/7B
Lai et al.	2024	Traffic signal managing	√		√	√		√	GPT-4
Pang et al.	2024	Traffic signal managing				√		√	GPT-4
Tang et al.	2024	Traffic signal managing				√		√	GPT-4
Wang et al.	2024	Traffic signal managing	√			√		√	GPT-4
Movahedi et al.	2025	Traffic signal managing				√		√	GPT-3.5 Turbo
Onsu et al.	2025	Traffic signal managing	√		√				LLaVA-1.5
Yuan et al.	2025	Traffic signal managing			√				LLaMA-3.1
Zeng et al.	2024	Urban navigation			√			√	LLaVA-7B LLaMA-3
Zhang et al.	2024	Urban navigation		√			√		Not mention
Alhammadi et al.	2025	Urban navigation				√		√	GPT-4o
Scribano et al.	2021	Vehicle retrieval	√	√	√			√	BERT-Base
Ying et al.	2025	Transportation planning				√			GPT-4o Phi-3-mini

scene classification. The more powerful multimodal model like GPT-4v has also shown potential in various visual assignments including visual question answering, geolocation and land cover classification (Tan et al., 2023). It has even been further utilized in more challenging tasks evaluating sustainable features of buildings (Bentley et al., 2024), the relationship between the built environment and urban depression rates (Ouyang et al., 2024), and even dynamics tasks such as temporal change detection in remote-sensing imagery (Elgendy et al., 2024).

Plus, LLM-agents were applied, mainly in scenarios including urban perception, complex remote sensing task analysis and domain-specific opinion generation. Specifically, Verma et al. (2023) employed a GPT-3.5-based generative agent that interacted with street view images and urban environments through deep learning image algorithms to automatically plan journeys for specific goals. Further studies have also demonstrated that large language models are well suited to urban perception tasks and streetscape auditing (Han, Yang, et al., 2024; Han, Zhang, et al., 2024; Jang & Kim, 2025; Yang et al., 2024; Zhang, Li, et al., 2024). Singh et al. (2024) proposed GeoLLM-Engine which placed LLMs in the environment with various tools to accomplish analyst tasks, while Lee et al. (2025) introduced multi-agent framework GeoLLM-Squad to tackle environmental tasks. Additionally, LLMs have been applied to generate domain-specific opinions including positive energy districts (PEDs) (X. Zhang, Han, et al., 2023; Zhang, Shah, & Han, 2023; Zhang, Sun, et al., 2023), wireless-network deployment (Sevim et al., 2024), and smart-city assessment frameworks (Shahrabani & Apanaviciene, 2024).

Not surprisingly, there were wide utilization of BERT as encoders and numerous need to modal alignment in terms of data processing, encompassing innovations in geographic information encoding (Balsebre et al., 2024; Fernandez & Dube, 2023; Penumadu et al., 2022) and image encoding methods (Hu et al., 2023; Kuganesan & Kuganesan, 2024; Muhtar et al., 2024; Zhang et al., 2022), as shown in Table 6. The large proportion of researches using fine-tuning as the model adaption method remains the same.

4. Discussion

4.1. General attitudes and concerns

In urban contexts, LLMs have proven especially adept at tasks such as city-wide text analysis, event simulation, domain-specific response generation, mobility prediction as well as street space perception, attributing to their robust multimodal capabilities. However, they still struggle to adapt to constantly changing urban environmental problems in real-time and lack judgment in emerging urban scenarios such as extreme weather and accidents for insufficient data source, followed by several limitations and concerns listed below.

The first issue pertains to the capability boundaries of LLMs. Issues such as model hallucinations, where the model generates factually incorrect or irrelevant, can undermine trust of researchers in their use

Table 6
Summary of structured information from LLM-related articles in urban environment and energy.

Article	Year	Application scenario	Data processing		Model adaptation				Pre-trained model
			Modality alignment	Embedding	Fine-tuning	Prompt Engineering		Integration	
						Zero-shot	Few-shot		
Ma et al.	2024	Sentiment analysis		√	√			√	BERT-Base
Shan et al.	2025	Topic modeling				√		√	GPT-3.5 Torbo BERTopic
Zhang et al.	2021	Text classification		√	√			√	BERT-Base
Ali et al.	2022	Text classification		√	√				BERT-Base
Kanth et al.	2022	Text classification		√	√			√	BERT-Base
Zhu et al.	2022	Text classification		√	√				BERT-Base
Zhang et al.	2023	Text classification				√		√	GPT-4
Guo et al.	2024	Text classification		√	√			√	BERT-Base
Qian et al.	2024	Text classification	√			√			GPT-3
Hu et al.	2022	Named entity recognition		√	√			√	BERT-Base
Chen et al.	2023	Named entity recognition		√	√			√	BERT-Base
Hu et al.	2023	Named entity recognition				√			GPT-4
Li et al.	2023	Named entity recognition		√	√			√	BERT-Base
Loja-Morocho et al.	2023	Named entity recognition		√	√			√	BERT-Base
Hao et al.	2024	Named entity recognition					√		ChatGLM-3
Fu et al.	2025	Named entity recognition			√				BERT-Base
Han et al.	2025	Named entity recognition			√				Qwen2 MiniCPM3
Mo et al.	2025	Named entity recognition			√			√	BERT-BiLSTM-CRF GPT-3.5-turbo
Xu et al.	2020	Geocoding		√	√			√	BERT-Base
Gao et al.	2022	Geocoding		√	√			√	BERT-Base
Penumadu et al.	2022	Geocoding	√	√	√			√	BERT-Base
Fernandez et al.	2023	Geocoding	√	√		√			Not mention
Yin et al.	2023	Geocoding	√				√		GPT-3
He et al.	2024	Geocoding				√		√	LLaMA-3 Mistral-8
Li et al.	2024	Geocoding				√		√	Unknown
Muron et al.	2025	Geocoding			√		√		GPT-3
Sun et al.	2025	Geocoding					√	√	GPT-4o GPT-3.5 Baichuan2-13B
Zhang et al.	2025	Geocoding				√			Gemini-1.5-Flash
Hu et al.	2023	Multimodal knowledge Q&A	√		√			√	InstructBLIP GPT-3.5
Kuckreja et al.	2023	Multimodal knowledge Q&A		√	√			√	LLaVA-1.5
Mai et al.	2023	Multimodal knowledge Q&A				√	√		GPT-3 BERT-Base
Manvi et al.	2023	Multimodal knowledge Q&A			√	√			GPT-3.5
Tan et al.	2023	Multimodal knowledge Q&A				√			GPT-4v
Unlu	2023	Multimodal knowledge Q&A		√	√				Falcon 1B RW
Balsebre et al.	2024	Multimodal knowledge Q&A	√		√			√	LLaMA-2
Bentley et al.	2024	Multimodal knowledge Q&A				√			GPT-4o
Danish et al.	2024	Multimodal knowledge Q&A				√		√	GPT-4o LLaVA-OneVision Qwen2-VL Sphinx LLaVA-1.5 LLaVA-NeXT GeoChat RS-LLaVA InternVL-2 Ferret
Elgendy et al.	2024	Multimodal knowledge Q&A	√		√				Video-LLaVA LLaVA-NeXT-Video
Muhtar et al.	2024	Multimodal knowledge Q&A	√		√				LLaMA-2
Qi et al.	2024	Multimodal knowledge Q&A	√		√			√	BEiT
Shao et al.	2024	Multimodal knowledge Q&A	√		√			√	From scratch
Zhou et al.	2024	Multimodal knowledge Q&A				√		√	GPT-4o
Choi et al.	2024	Building energy modeling				√		√	GPT-4o
Xiao et al.	2024	Building energy modeling				√		√	GPT-3.5 Turbo ChatGLM-1
Choi et al.	2025	Building energy modeling				√		√	GPT-4o
Labib	2025	Building energy modeling				√		√	LLaMA-2
Sharma et al.	2023	Building energy modeling				√		√	GPT-3.5

(*continued on next page*)

Table 6 (*continued*)

Article	Year	Application scenario	Data processing		Model adaptation				Pre-trained model
			Modality alignment	Embedding	Fine-tuning	Prompt Engineering		Integration	
						Zero-shot	Few-shot		
Kuganesan et al.	2024	Geographic task solution	√		√				GPT-3
Singh et al.	2024	Geographic task solution	√			√	√	√	GPT-4
Lee et al.	2025	Geographic task solution				√		√	GPT-4o-mini
Balsa-Barreiro et al.	2024	Individual and built environment interaction	√			√			GPT-4
Zhang et al.	2022	Street space perception	√	√	√				BERT-Base
Verma et al.	2023	Street space perception					√		GPT-3.5
Han et al.	2024	Street space perception				√			GPT-4o
Liang et al.	2024	Street space perception			√				From scratch
Yang et al.	2024	Street space perception				√			GPT-4V Gemini-Pro
Zhang et al.	2024	Street space perception				√		√	GPT-4V
Cobanli et al.	2025	Street space perception				√		√	GPT-3.5 Turbo
Jang et al.	2025	Street space perception				√		√	GPT-4o Gemini-1.5-Pro
Jiang	2025	Street space perception				√			GPT-4o
Jang et al.	2023	Domain-specific response generation		√		√			GPT-3.5
Zhang et al.	2023	Domain-specific response generation				√			GPT-3.5
Sevim et al.	2024	Domain-specific response generation	√			√		√	DistilBERT
Shahrabani et al.	2024	Domain-specific response generation				√		√	GPT-3 Google Bard
Xie et al.	2024	Domain-specific response generation				√			GPT-4
Yu et al.	2024	Domain-specific response generation				√			GPT-4
Peng et al.	2025	Domain-specific response generation				√		√	Qwen2-7B
Ouyang et al.	2024	Urban healthy analysis				√			GPT-4v

for critical tasks, especially problematic in fields like urban environment and remote sensing with high demands for data authenticity and geographic precision (Mai et al., 2023). Moreover, LLMs have been shown to exhibit both geographic and cultural biases. These include systematic errors in geospatial predictions, underrepresentation of features in socioeconomically disadvantaged regions, and a tendency to portray cities primarily through curated images shaped by planning, exhibitions, and commercialization (Campo-Ruiz, 2025; Manvi et al., 2024). Furthermore, LLMs struggle with generalization and uncertainty. Due to the complexity of the real world, training datasets cover only limited domains, which inevitably fall short of capturing all specific scenarios encountered in practice. As a result, when LLMs are applied to highly variable contexts, their abilities to tackle these problems remain constrained (Zhang, Li, et al., 2024; Zhang, Liu, et al., 2024; Zhang, Sun, et al., 2024). This limitation is particularly evident in domains such as urban mobility and navigation, where the lack of direct interaction with real-time environments and the heavy reliance on static pretraining data make it difficult for LLMs to deliver reliable decisions in dynamic, real-world settings (Choi & Yoon, 2025; Javaid et al., 2025).

The high computational cost associated with deploying and training these models is another significant challenge. Many studies have highlighted the high cost of using advanced models like GPT-4, which may pose a financial burden on research (Hadi et al., 2023; Shahriar et al., 2024; Tamkin et al., 2021). Despite the availability of open-source models such as ChatGLM, Llama3 and supporting frameworks like Ollama, a framework designed for local deployment of LLMs, that facilitate local deployment, challenges related to GPU memory capacity and inference speed also impose economic burdens. Besides, the environmental impact of training LLMs is substantial, as research by Strubell et al. (2020) estimated that training a single LLM can generate emissions roughly equivalent to the lifetime emissions of five average automobiles, and data centers powering LLMs often require millions of gallons of water every day for cooling (Li, Su, et al., 2025; Li, Yang, et al., 2025; Li, Zhao, et al., 2025). Other concerns such as inference delays, restrictions on non-open-source models, dataset preparation burden, and challenges in evaluating model performance were also mentioned in the reviewed articles (Liu et al., 2025; Mai et al., 2023; Zhang, Li, et al., 2024; Zhang, Liu, et al., 2024; Zhang, Sun, et al., 2024).

Additionally, the performance differences between LLMs and traditional deep learning models on different tasks are generally concerned by many researchers. Several studies comparing two types of methods find that LLMs indeed outperform traditional models in scenarios such as tourist occupancy prediction and urban waterlogging detection (Chen, Chen, et al., 2023; Chen, Hu, et al., 2023; Chen, Wang, & Xu, 2023; Giménez Manuel et al., 2024). However, most researchers remain neutral or opposed to considering LLMs as a replacement for traditional deep learning methods. They argue that LLMs serve as augmentation methods rather than replacements (Chauncey & McKenna, 2024) since traditional methods are still necessary for more technical aspects or outperform LLMs under certain circumstances (Yu & McKinley, 2024).

Beyond these technical and operational limitations, broader concerns about AI ethics and the underlying logic of LLMs also present challenges. On the one hand, the training data or prompts may involve sensitive information, and any leaks could result in serious privacy concerns (Zhang, Li, et al., 2024; Zhang, Liu, et al., 2024; Zhang, Sun, et al., 2024). LLMs' decision-making processes are often opaque, which makes it difficult to ensure accountability in critical areas like urban governance (Qu & Wang, 2024). On the other hand, there are deeper concerns about the limitations of LLMs' comprehension of the world. Prevalent LLMs were trained on language modality or values, but when confronted with complex tasks, human beings do not evaluate them with values (Verma et al., 2023). This could raise questions about whether language can truly represent the information of complexities of urban studies.

4.2. Potential research opportunities

Looking forward, we foresee several research opportunities for the application of LLMs in urban studies. Firstly, the exponential rise in LLM-based urban research over the past six years suggests that their scope will continue to broaden. Beyond the promising early work on agent-based simulations in urban sociology, we anticipate rich dividends from applying LLMs to urban economics, for example modeling urban spatial balance, segregation, or financial impacts of zoning reforms. Another promising application field lies in urban development and evolution. By further combining and analyzing datasets from diverse sources, researchers are enabled to build a more comprehensive understanding of how cities grow, change, and adapt over time with the help of LLMs, providing insights into the charming question of how cities evolve.

Secondly, we deem that smaller, task-specific models may dominate simpler tasks, while larger-parameter models will address complex and difficult problems. The application scenarios of LLMs may zoom in complicated and challenging tasks when small models can achieve similar effects on simple tasks (Fu et al., 2023). For instance, current lightweight LLMs' variants can be embedded in mobile sensors to fuse real-time visual, acoustic, and environmental streams to detect anomalies and enrich raw signals with semantic labels without sending privacy-sensitive data to the cloud.

Besides, the use of LLMs as domain experts can be extended beyond current applications. Although existing research has demonstrated that LLMs can produce insights nearly on par with human specialists when prompted appropriately (Yu, Xing, et al., 2024; Yu, Yang, et al., 2024), this strategy remains under-explored in many areas of urban studies. We tend to believe that by applying careful prompt engineering or integrating methods like retrieval-augmented generation (RAG), researchers can further boost both the accuracy and depth of these expert-level outputs, unlocking new possibilities across the urban research spectrum.

Furthermore, LLMs have transformative potential in uncovering intricate urban patterns and dynamics. Functioning as intelligent agents, LLMs can tap into vast repositories of urban knowledge to respond to or evaluate city-specific inquiries. Through iterative analysis of thousands of interactions, aggregated responses can be statistically examined to reveal hidden trends, correlations, and anomalies within urban systems. This approach not only offers novel perspectives on urban studies but also facilitates the generation of new hypotheses and conclusions that might not emerge through traditional methodologies.

Finally, from a technical perspective, with the improvements in the reasoning and planning capabilities, LLMs will move towards higher-level and collaborative intelligence and truly realize automation. Thus, opportunities can also be found in applying LLMs for finishing much challenging tasks such as automatic urban planning and design. Progress in these areas could lead to more sophisticated models that autonomously collect information and make complex decisions, enhancing their role as tools for urban research and policymaking.

There are still some potential limitations to this study. In striving for comprehensive coverage of LLMs' applications, we have included preprints and conference papers, which may affect the overall quality of the sources reviewed. In addition, due to the breadth of our research questions and the scope of this review, we have not delved deeply into the architectural distinctions among LLMs. Though variations in architecture can significantly influence a model's performance, interpretability, and applicability across different urban research domains, a more detailed examination of these architectural differences actually falls outside the primary focus of this article.

5. Conclusions

This paper provides a systematic literature review on the application of LLMs in urban studies, capturing their rapid development and diverse applications over the past six years. By reviewing 233 papers and leveraging a customized GPT to extract information from them, we reveal an exponential increase in LLM-related papers in urban studies, with GPT-based and BERT-based models dominating the field. Embedding techniques and fine-tuning remain the predominant methods for data processing and model adaptation, highlighting the versatility and adaptability of these models in addressing urban research challenges. Furthermore, the five domains of urban studies emphasize the significance role of LLMs in large-scale data analysis and multimodal data processing while also demonstrating unique use scenarios. In urban sociology and economics, BERT variants have been predominantly used for textual analysis, often combined with other models. In urban management and public policy, GPT-based models are employed for both text analysis and generation, usually integrated into platforms to enhance decision-making processes. Field in urban planning and design have leveraged multimodal models as agents for complex tasks such as scene generation, planning support and resident opinion simulation. The transportation field illustrates the flexible architecture of LLMs, employing them for diverse tasks such as human mobility analysis, traffic flow prediction or traffic system managing to achieve optimal performance. Similarly, in environmental and energy domains, LLMs have proven effective for geographic analysis and urban perception evaluation, emphasizing modality alignment for accurate decision-making.

In conclusion, this paper serves as a comprehensive resource for researchers, summarizing the application scenarios, methods, and developmental trends of LLMs in urban studies. By shedding light on their current capabilities and potential, this review aims to inspire further exploration and innovation in leveraging LLMs to address the complexities of urban research.

CRediT authorship contribution statement

Junhao Xia: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yao Tong:** Writing – review & editing. **Ying Long:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Funding sources

This work was supported by the National Natural Science Foundation of China [Grant No. 62394331], [Grant No. 62394335], Program of Tsinghua University - CSCEC Jiuhe Development Group Co. Ltd. Joint Research Center (2024A2) and the Department of Science and Technology of Inner Mongolia Autonomous Region (2023YFSW0002). The funders had no role in study design, data collection and analysis, decision to publish, preparation of the manuscript, writing of the report or decision to submit the article for publication.

Declaration of competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

The authors gratefully acknowledge the members of the Beijing City Lab for their invaluable discussions and insightful feedback.

Appendix A. Description of reviewed papers' analyzing dimensions

Field of study: categories of field of study are the same as the five thematic categories in urban studies mentioned above.

Application scenario: it refers to the main roles that the research used LLMs to act as. Since scenarios varied in different research fields, we did not give the fixed selection range to our tailored GPT, but let it automatically generate options according to comprehensive article understanding.

Diving deep into the various kinds of approaches to leveraging LLMs in urban studies, we analyzed them from three dimensions similar to the previous research (Zhang, Li, et al., 2024; Zhang, Liu, et al., 2024; Zhang, Sun, et al., 2024). Typically, the application of LLMs involves data processing, model architecture building, model pretraining and model adaptation (Naveed et al., 2024). In this review, we concentrated on the processing of the input and output data, model adaptation methods, and the choices of pre-trained model, rather than concerning the design of the model architecture and pretraining steps.

Data processing: it means how data is fed to LLMs and the form of data that LLMs return. Two types of data processing methods were extracted, that is modality alignment and embedding. Modality alignment refers to the process of aligning different types of data (e.g., images, text, audio) into a shared representation space (Song et al., 2023), an approach requiring special attention due to the multimodality of urban data, while embedding means representing words as distributed vectors for downstream tasks. Since embedding is universal when transforming input into representation, we specifically defined it as using LLMs as encoders to generate features.

Model adaptation: it focuses on the methods of enabling LLMs to perform better on downstream tasks, including approaches like fine-tuning, prompt engineering (zero-shot or few-shot) and integration. Fine-tuning stands for adjusting pre-trained models' weights on specific datasets for performance enhancement (Howard & Ruder, 2018). Prompt engineering is the process of guiding LLMs to accomplish tasks by meticulously designing prompts, which is a simple but critical technique for improving their capabilities (Sahoo et al., 2024) encompassing two methods called zero-shot and few-shot. The difference between them lies in whether providing multiple input-output demonstration pairs in prompts. Integration generally refers to the combination of frameworks, including the integration between multiple model architecture or between models and systems.

Pre-trained model: it stands for the types of LLMs used in research, which are uniformly recorded in the form of models with its versions.

Appendix B. Configuration of "LLMs in Urban Study: Literature Review Analyzer"

Name:

LLMs in Urban Study: Literature Review Analyzer

Description:

An assistant designed to extract and structure key analytical data from urban studies literature leveraging large language models.

Instructions:

You are a literature review analyzer specializing in parsing and systematically analyzing academic papers related to the application of Large Language Models (LLMs) in urban studies, according to a detailed classification framework provided below. Your primary task is to automate the extraction and coding of structured information from these papers, aligning them precisely with predefined indicators and categories: Field of study, Application scenario, Data processing, Model adaptation, and Pre-trained model.

Upon receiving an academic paper, analyze its contents rigorously, classify it according to the specified indicators. Accuracy in assigning papers to the correct categories and subcategories is crucial. Ensure no indicators from the provided framework are overlooked, and avoid misclassification.

If certain details within the paper are unclear or ambiguous regarding their alignment with the provided framework, proactively request clarification from the user. Your output should consistently adopt a concise, structured table format, maintaining an academic tone suitable for research documentation.

The analytical framework is illustrated below:

| Indicators | Description or Categories |

| Field of study | Select one: Urban planning and design, Urban environment and energy, Urban governance and public policy, Urban sociology and economics, Urban transportation |

| Application scenario | Identify detailed urban study scenarios, such as sentiment analysis, text classification, multimodal knowledge Q&A, traffic prediction, human mobility analysis, named entity recognition, topic modeling, textual knowledge Q&A, Geocoding. If a scenario does not fit these domains, define a new one clearly. |

| Data processing | Select one or more: Modality alignment / Embedding (Modality alignment refers to aligning different data types (e.g., images, text, audio) into a common representation space, essential for multimodal urban data analysis. Embedding refers to representing textual data as distributed vectors for further analytical tasks.) |

| Model adaptation | Select one or more: Fine-tuning / Zero-shot or Few-shot / Integration (Fine-tuning involves adapting pre-trained models' weights to specific datasets. Zero-shot and few-shot refer to guide LLM task performance with no or minimal input-output examples. Integration encompasses the combination of multiple models or frameworks.) |

| Pre-trained model | Specific the employed pre-trained LLM, such as BERT, GPT-2, GPT-3.5, LLaMA2, From scratch. |

Here is an example of returned structured result:

| Field of study | Application scenario | Data processing | Model adaptation | Pre-trained model |

| Urban planning and design | Land-use analysis | Embedding | Zero-shot, Integration | GPT-4 |

During the preparation of this work the author(s) used ChatGPT in order to analyze papers during the systematic literature review and refine manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Data availability

Data will be made available on request.

References

Alhammadi, A., Abraham, A., & Zhao, Q. (2025). LLM-powered UAV automations for city-wide operations. In A. Abraham, S. Prasad, A. Alhammadi, T. Lestable, & F. Chaabane (Eds.), *Internet of vehicles and computer vision solutions for smart city transformations* (pp. 69–83). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72959-1_4.

Ali, S., & Shirazi, F. (2022). A transformer-based machine learning approach for sustainable E-waste management: A comparative policy analysis between the Swiss and Canadian systems. *Sustainability, 14*(20), 13220. https://doi.org/10.3390/su142013220

Amatriain, X. (2024). Prompt design and engineering: Introduction and advanced methods. In *arXiv*. https://doi.org/10.48550/arXiv.2401.14423

Balsebre, P., Huang, W., & Cong, G. (2024). LAMP: A language model on the map. In *arXiv*. https://doi.org/10.48550/arXiv.2403.09059

Batty, M. (2008). The size, scale, and shape of cities. *Science, 319*(5864), 769–771. https://doi.org/10.1126/science.1151419

Batty, M. (2013). *The new science of cities*. MIT Press.

Bedi, J., & Toshniwal, D. (2022). CitEnergy: A BERT based model to analyse citizens' energy-tweets. *Sustainable Cities and Society, 80*, Article 103706. https://doi.org/10.1016/j.scs.2022.103706

Bendig, D., & Bräunche, A. (2024). The role of artificial intelligence algorithms in information systems research: A conceptual overview and avenues for research. *Management Review Quarterly*. https://doi.org/10.1007/s11301-024-00451-y

Bentley, P. J., Lim, S. L., Mathur, R., & Narang, S. (2024). Automated real-world sustainability data generation from images of buildings. In *arXiv*. https://doi.org/10.48550/arXiv.2405.18064

Bettencourt, L., & West, G. (2010). A unified theory of urban living. *Nature, 467*(7318), 912–913. https://doi.org/10.1038/467912a

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies, 1*(1), 363–376. https://doi.org/10.1162/qss_a_00018

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). *On the opportunities and risks of foundation models*. arXiv.Org. https://arxiv.org/abs/2108.07258v3.

Bowen, W. M., Dunn, R. A., & Kasdan, D. O. (2010). What is "urban studies"? Context, internal structure, and content. *Journal of Urban Affairs, 32*(2), 199–227. https://doi.org/10.1111/j.1467-9906.2009.00474.x

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. In *arXiv*. https://doi.org/10.48550/arXiv.2005.14165

Çalışkan, E. B. (2025). Land cover analysis of two university campuses: Examination over satellite images by chat GPT. *International Journal of Engineering and Geosciences, 10*(1), 124–136. https://doi.org/10.26833/ijeg.1504721

Campo-Ruiz, I. (2025). Artificial intelligence may affect diversity: Architecture and cultural context reflected through ChatGPT, midjourney, and google maps. *Humanities and Social Sciences Communications, 12*(1), 1–13. https://doi.org/10.1057/s41599-024-03968-5

Cao, D., Zeng, K., Wang, J., Sharma, P. K., Ma, X., Liu, Y., & Zhou, S. (2022). BERT-based deep spatial-temporal network for taxi demand prediction. *IEEE Transactions on Intelligent Transportation Systems, 23*(7), 9442–9454. https://doi.org/10.1109/TITS.2021.3122114

Cao, S., Zhao, Xuesheng, & Du, S. (2024). Multi-type and fine-grained urban green space function mapping based on BERT model and multi-source data fusion. *International Journal of Digital Earth, 17*(1), Article 2308723. https://doi.org/10.1080/17538947.2024.2308723

Castells, M. (1976). *Theory and ideology in urban sociology*. Routledge.

Chakrabarty, B. K. (2001). Urban management: Concepts, principles, techniques and education. *Cities, 18*(5), 331–345. https://doi.org/10.1016/S0264-2751(01)00026-9

Chang, C., Wang, W.-Y., Peng, W.-C., & Chen, T.-F. (2024). LLM4TS: Aligning pre-trained LLMs as data-efficient time-series forecasters. In *arXiv*. http://arxiv.org/abs/2308.08469.

Charfaoui, K., & Mussard, S. (2024). Sentiment analysis for tourism insights: A machine learning approach. *Stats, 7*(4). https://doi.org/10.3390/stats7040090. Article 4.

Chauncey, S. A., & McKenna, H. P. (2024). Creativity and innovation in civic spaces supported by cognitive flexibility when learning with AI chatbots in smart cities. *Urban Science, 8*(1), 16. https://doi.org/10.3390/urbansci8010016

Chen, C. (2024). Perception and reality: Urban green space analysis using language model-based social media insights. A case study within Shanghai's inner ring. In *Accelerated design - Proceedings of the 29th CAADRIA conference, Singapore*. https://papers.cumincad.org/cgi-bin/works/paper/caadria2024_68.

Chen, H., & Ding, Y. (2025). Implementing traffic agent based on LangGraph. In *Fourth international conference on intelligent traffic systems and smart city (ITSSC 2024), 13422* (pp. 582–587). https://doi.org/10.1117/12.3050639

Chen, J., Chen, L., Li, Y., Zhang, W., & Long, Y. (2023). Measuring physical disorder in urban street spaces: A large-scale analysis using street view images and deep learning. *Annals of the American Association of Geographers, 113*(2), 469–487. https://doi.org/10.1080/24694452.2022.2114417

Chen, W., Liang, Y., Zhu, Y., Chang, Y., Luo, K., Wen, H., … Zheng, Y. (2024). Deep learning for trajectory data management and mining: A survey and beyond. In *arXiv*. https://doi.org/10.48550/arXiv.2403.14151

Chen, Y., Hu, M., Chen, X., Wang, F., Liu, B., & Huo, Z. (2023). An approach of using social media data to detect the real time spatio-temporal variations of urban waterlogging. *Journal of Hydrology, 625*, Article 130128. https://doi.org/10.1016/j.jhydrol.2023.130128

Chen, Y., Liu, B., Zhang, Z., & Kim, H.-S. (2022). An end-to-end deep learning framework for multiple audio source separation and localization. In *ICASSP 2022 - 2022 IEEE International conference on acoustics, speech and signal processing (ICASSP)* (pp. 736–740). https://doi.org/10.1109/ICASSP43922.2022.9746950

Chen, Y., Wang, X., & Xu, G. (2023). GATGPT: A pre-trained large language model with graph attention network for spatiotemporal imputation. *arXiv*. http://arxiv.org/abs/2311.14332.

Chen, Y., Zhang, S., Han, T., Du, Y., Zhang, W., & Li, J. (2024). Chat3D: Interactive understanding 3D scene-level point clouds by chatting with foundation model for urban ecological construction. *ISPRS Journal of Photogrammetry and Remote Sensing, 212*, 181–192. https://doi.org/10.1016/j.isprsjprs.2024.04.024

Choi, S., Liu, K., & Hanaoka, S. (2024). Perceived responses of international tourists to transportation and tourism services during Typhoons Faxai and Hagibis in Japan. *Sustainability, 16(20)*. https://doi.org/10.3390/su16209114. Article 20.

Choi, S., & Yoon, S. (2025). AI agent-based intelligent urban digital twin (I-UDT): Concept, methodology, and case studies. *Smart Cities, 8(1), Article 1*. https://doi.org/10.3390/smartcities8010028

Cusumano, L., Olsson, N., Granath, M., Jockwer, R., & Rempling, R. (2025). Clustering techniques and keyword extraction with large language models for knowledge discovery in building defects data. *Construction Innovation, 25*(7), 76–97. https://doi.org/10.1108/CI-04-2024-0123

De Zarzà, I., De Curtò, J., Roig, G., & Calafate, C. T. (2023). LLM multimodal traffic accident forecasting. *Sensors, 23*(22), 9225. https://doi.org/10.3390/s23229225

Deng, J., Chai, W., Huang, J., Zhao, Z., Huang, Q., Gao, M., Guo, J., Hao, S., Hu, W., Hwang, J.-N., Li, X., & Wang, G. (2024). CityCraft: A real crafter for 3D city generation. *arXiv*. http://arxiv.org/abs/2406.04983.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv*. https://doi.org/10.48550/arXiv.1810.04805

Dortheimer, J., Martelaro, N., Sprecher, A., & Schubert, G. (2024). Evaluating large-language-model chatbots to engage communities in large-scale design projects. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 38*, Article e4. https://doi.org/10.1017/S0890060424000027

Elgendy, H., Sharshar, A., Aboeitta, A., Ashraf, Y., & Guizani, M. (2024). *GeoLLaVA: Efficient fine-tuned vision-language models for temporal change detection in remote sensing. arXiv*. https://doi.org/10.48550/arXiv.2410.19552

Fernandez, A., & Dube, S. (2023). Core building blocks: Next gen geo spatial GPT application. *arXiv*. http://arxiv.org/abs/2310.11029.

Fink, A. (2019). *Conducting research literature reviews: From the internet to paper*. Sage publications.

Fontes, T., Murçós, F., Carneiro, E., Ribeiro, J., & Rossetti, R. J. F. (2023). Leveraging social media as a source of mobility intelligence: An NLP-based approach. *IEEE Open Journal of Intelligent Transportation Systems, 4*, 663–681. https://doi.org/10.1109/OJITS.2023.3308210

Fu, S., Schultz, D. M., Lyu, H., Zheng, Z., & Zhang, C. (2025). Creating a national urban flood dataset for China from news texts (2000−2022) at the county level. *Hydrology and Earth System Sciences, 29*(3), 767–783. https://doi.org/10.5194/hess-29-767-2025

Fu, X., Sanchez, T. W., Li, C., & Reu Junqueira, J. (2024). Deciphering public voices in the digital era: Benchmarking ChatGPT for analyzing citizen feedback in Hamilton, New Zealand. *Journal of the American Planning Association, 1–14*. https://doi.org/10.1080/01944363.2024.2309259

Fu, X., Wang, R., & Li, C. (2024). Can ChatGPT evaluate plans? *Journal of the American Planning Association, 90*(3), 525–536. https://doi.org/10.1080/01944363.2023.2271893

Fu, Y., Peng, H., Ou, L., Sabharwal, A., & Khot, T. (2023). Specializing smaller language models towards multi-step reasoning. In *Proceedings of the 40th international conference on machine learning* (pp. 10421–10430). https://proceedings.mlr.press/v202/fu23d.html.

Gao, Y., Xiong, Y., Wang, S., & Wang, H. (2022). GeoBERT: Pre-training geospatial representation learning on point-of-interest. *Applied Sciences, 12*(24), 12942. https://doi.org/10.3390/app122412942

Giménez Manuel, J. G., Giner Pérez De Lucia, J., Celdrán Bernabeu, M. A., Mazón López, J. N., Cano Escribá, J. C., & Cecilia Canales, J. M. (2024). Advancing smart tourism destinations: A case study using bidirectional encoder representations from transformers-based occupancy predictions in Torrevieja (Spain). *IET Smart Cities, smc2.12085*. https://doi.org/10.1049/smc2.12085

Graham, S., & Marvin, S. (2002). *Splintering urbanism: Networked infrastructures, technological mobilities and the urban condition*. Routledge.

Guan, Q., Ouyang, J., Wu, D., & Yu, W. (2024). CityGPT: Towards urban IoT learning, analysis and interaction with multi-agent system. *arXiv*. http://arxiv.org/abs/2405.14691.

Guo, Y., Tang, J., Liu, H., Yang, X., & Deng, M. (2024). Identifying up-to-date urban land-use patterns with visual and semantic features based on multisource geospatial data. *Sustainable Cities and Society, 101*, Article 105184. https://doi.org/10.1016/j.scs.2024.105184

Hadi, M. U., Tashi, Q. A., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., & Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. TechRxiv. https://www.authorea.com/doi/full/10.36227/techrxiv.23589741.v1?commit=b1cb46f5b0f749cf5f2f33806f7c124904c14967.

Hall, P. (2014). *Cities of tomorrow: An intellectual history of urban planning and design since 1880*. John Wiley & Sons.

Han, B., Yang, Y., Caspi, A., & Howe, B. (2024). *Towards zero-shot annotation of the built environment with vision-language models (vision paper). arXiv*. https://doi.org/10.48550/arXiv.2408.00932

Han, X., Zhang, Z., Zhao, X., Shen, G., Kong, X., Wei, X., Nie, L., Ye, J., & Zhu, Y. (2024). *GARLIC: GPT-augmented reinforcement learning with intelligent control for vehicle dispatching. arXiv*. https://doi.org/10.48550/arXiv.2408.10286

Han, Y., Liu, J., Luo, A., Wang, Y., & Bao, S. (2025). Fine-tuning LLM-assisted Chinese disaster geospatial intelligence extraction and case studies. *ISPRS International Journal of Geo-Information, 14(2), Article 2*. https://doi.org/10.3390/ijgi14020079

Hao, X., Chen, W., Yan, Y., Zhong, S., Wang, K., Wen, Q., & Liang, Y. (2025). *UrbanVLP: Multi-granularity vision-language pretraining for urban socioeconomic indicator prediction. arXiv*. https://doi.org/10.48550/arXiv.2403.16831

Haydari, A., Chen, D., Lai, Z., Zhang, M., & Chuah, C.-N. (2024). MobilityGPT: Enhanced human mobility modeling with a GPT model. *arXiv*. http://arxiv.org/abs/2402.03264.

Hayden, D. J. (2023). Generative AI: Mining housing data with a higher powered shovel. *Cityscape, 25*(3), 199–206.

He, H., Luo, H., & Wang, Q. R. (2024). *ST-MoE-BERT: A spatial-temporal mixture-of-experts framework for long-term cross-city mobility prediction. arXiv*. https://doi.org/10.48550/arXiv.2410.14099

Howard, J., & Ruder, S. (2018). *Universal language model fine-tuning for text classification. arXiv*. https://doi.org/10.48550/arXiv.1801.06146

Hu, K., Li, L., Tao, X., & Zhang, J. (2024). Semantics and geography aware hierarchical learning for sequential crime prediction. *IEEE Signal Processing Letters, 31*, 1234–1238. https://doi.org/10.1109/LSP.2024.3393863

Hu, Y., Mai, G., Cundy, C., Choi, K., Lao, N., Liu, W., … Joseph, K. (2023). Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science, 37*(11), 2289–2318. https://doi.org/10.1080/13658816.2023.2266495

Huang, Y., Mao, X., Guo, S., Chen, Y., Shen, J., Li, T., Lin, Y., & Wan, H. (2024). STD-PLM: Understanding both spatial and temporal properties of spatial-temporal data with PLM. *arXiv*. http://arxiv.org/abs/2407.09096.

Jang, K. M., & Kim, J. (2025). Multimodal large language models as built environment auditing tools. *The Professional Geographer, 77*(1), 84–90. https://doi.org/10.1080/00330124.2024.2404894

Jannani, A., Bouhsissin, S., Sael, N., & Benabbou, F. (2025). Topic modeling and sentiment analysis of arabic news headlines for a societal well-being scoring and monitoring system: Moroccan use case. *IEEE Access, 13*, 26345–26364. https://doi.org/10.1109/ACCESS.2025.3538888

Javaid, S., Khan, M. A., Fahim, H., He, B., & Saeed, N. (2025). Explainable AI and monocular vision for enhanced UAV navigation in smart cities: Prospects and challenges. *Frontiers in Sustainable Cities, 7*, Article 1561404. https://doi.org/10.3389/frsc.2025.1561404

Jiang, H., Li, M., Witte, P., Geertman, S., & Pan, H. (2025). Urban chatter: Exploring the potential of ChatGPT-like and generative AI in enhancing planning support. *Cities, 158*, Article 105701. https://doi.org/10.1016/j.cities.2025.105701

Jiang, Y., Chao, Q., Chen, Y., Li, X., Liu, S., & Cong, G. (2024). UrbanLLM: Autonomous urban activity planning and management with large language models. *arXiv*. http://arxiv.org/abs/2406.12360.

Jin, K., Wi, J., Lee, E., Kang, S., Kim, S., & Kim, Y. (2021). TrafficBERT: Pre-trained model with large-scale data for long-range traffic flow forecasting. *Expert Systems with Applications, 186*, Article 115738. https://doi.org/10.1016/j.eswa.2021.115738

Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J. Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S., & Wen, Q. (2024). Time-LLM: Time series forecasting by reprogramming large language models. *arXiv*. http://arxiv.org/abs/2310.01728.

Jonnala, R., Liang, G., Yang, J., & Alsmadi, I. (2024). Using large language models in public transit systems, San Antonio as a case study. *arXiv*. http://arxiv.org/abs/2407.11003.

Jonveaux, L. (2024). *Using large language models for a standard assessment mapping for sustainable communities. arXiv*. https://doi.org/10.48550/arXiv.2411.00208

Ju, C., Liu, J., Sinha, S., Xue, H., & Salim, F. (2025). *TrajLLM: A modular LLM-enhanced agent-based framework for realistic human trajectory simulation. arXiv*. https://doi.org/10.48550/arXiv.2502.18712

Kalyuzhnaya, A., Mityagin, S., Lutsenko, E., Getmanov, A., Aksenkin, Y., Fatkhiev, K., … Boukhanovsky, A. (2025). LLM agents for smart city management: Enhancing decision support through multi-agent AI systems. *Smart Cities, 8(1), Article 1*. https://doi.org/10.3390/smartcities8010019

Kanth, A. K., Chitra, P., & Sowmya, G. G. (2022). Deep learning-based assessment of flood severity using social media streams. *Stochastic Environmental Research and Risk Assessment, 36*(2), 473–493. https://doi.org/10.1007/s00477-021-02161-3

Karren, K., Schmitz, M., & Schaffer, S. (2024). Improving conversational user interfaces for citizen complaint management through enhanced contextual feedback. *ACM Conversational User Interfaces, 2024*, 1–11. https://doi.org/10.1145/3640794.3665562

Kuckreja, K., Danish, M. S., Naseer, M., Das, A., Khan, S., & Khan, F. S. (2023). *GeoChat: Grounded large vision-language model for remote sensing. arXiv*. https://doi.org/10.48550/arXiv.2311.15826

Kuganesan, T., & Kuganesan, S. (2024). A GPT approach for transcending decision-making in urban environment through the integration of multiple earth observation and in-situ data sources. In *2024 international conference on machine intelligence for geoanalytics and remote sensing (MIGARS)* (pp. 1–3). https://doi.org/10.1109/MIGARS61408.2024.10544716

Lai, S., Xu, Z., Zhang, W., Liu, H., & Xiong, H. (2024). LLMLight: Large language models as traffic signal control agents. *arXiv*. http://arxiv.org/abs/2312.16044.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, C., Paramanayakam, V., Karatzas, A., Jian, Y., Fore, M., Liao, H., Yu, F., Li, R., Anagnostopoulos, I., & Stamoulis, D. (2025). *Multi-agent geospatial copilots for remote sensing workflows. arXiv*. https://doi.org/10.48550/arXiv.2501.16254

Li, J., Shi, Y., & Li, S. (2024). Analysis of Beijing traffic violations based on the BERT-CRF model. *Promet - Traffic&Transportation, 36*(2), 279–293. https://doi.org/10.7307/ptt.v36i2.366

Li, L., Hu, S., Shaw, A., & Hemphill, L. (2024). Crowdsourcing public attitudes toward local services through the lens of google maps reviews: An urban density-based perspective. *arXiv*. http://arxiv.org/abs/2404.13156.

Li, P., Yang, J., Islam, M. A., & Ren, S. (2025). *Making AI less 'thirsty': Uncovering and addressing the secret water footprint of AI models. arXiv*. https://doi.org/10.48550/arXiv.2304.03271

Li, R., Zhao, L., Xie, Z., Ji, C., Mo, J., Yang, Z., & Feng, Y. (2025). Mining and analyzing the evolution of public opinion in extreme disaster events from social media: A case study of the 2022 Yingde flood in China. *Natural Hazards Review, 26*(1), Article 05024015. https://doi.org/10.1061/NHREFO.NHENG-2107

Li, S., Feng, J., Chi, J., Hu, X., Zhao, X., & Xu, F. (2024). *LIMP: Large language model enhanced intent-aware mobility prediction. arXiv*. https://doi.org/10.48550/arXiv.2408.12832

Li, S., Tran, T., Lin, H., Krumm, J., Shahabi, C., & Xiong, L. (2024). *Geo-llama: Leveraging LLMs for human mobility trajectory generation with spatiotemporal constraints. arXiv*. https://doi.org/10.48550/arXiv.2408.13918

Li, X., Huang, F., Lv, J., Xiao, Z., Li, G., & Yue, Y. (2024). Be more real: Travel diary generation using LLM agents and individual profiles. *arXiv*. http://arxiv.org/abs/2407.18932.

Li, Y., Chen, S., Hwang, K., Ji, X., Lei, Z., Zhu, Y., … Liu, M. (2024). Spatio-temporal data fusion techniques for modeling digital twin city. *Geo-spatial Information Science, 1–24*. https://doi.org/10.1080/10095020.2024.2350175

Li, Z., Su, Y., Wang, H., & Zhao, W. (2025). *BuildingView: Constructing urban building exteriors databases with street view imagery and multimodal large language mode. arXiv*. https://doi.org/10.48550/arXiv.2409.19527

Liang, Q., Hu, C., & Chen, S. (2021). Evaluation of the optimal topic classification for social media data combined with text semantics: A case study of public opinion analysis related to COVID-19 with microblogs. *ISPRS International Journal of Geo-Information, 10*(12), 811. https://doi.org/10.3390/ijgi10120811

Lim, J., & Hwang, J. (2024). Exploring diverse interests of collaborators in smart cities: A topic analysis using LDA and BERT. *Heliyon, 10*(9), Article e30367. https://doi.org/10.1016/j.heliyon.2024.e30367

Liu, C., Yang, S., Xu, Q., Li, Z., Long, C., Li, Z., & Zhao, R. (2024). Spatial-temporal large language model for traffic prediction. In *2024 25th IEEE international conference on mobile data management (MDM)* (pp. 31–40). https://doi.org/10.1109/MDM61037.2024.00025

Liu, J., Meng, B., Wang, J., Chen, S., Tian, B., & Zhi, G. (2021). Exploring the spatiotemporal patterns of residents' daily activities using text-based social media data: A case study of Beijing, China. *ISPRS International Journal of Geo-Information, 10*(6), 389. https://doi.org/10.3390/ijgi10060389

Liu, M., Zhang, L., Chen, J., Chen, W.-A., Yang, Z., Lo, L. J., … O'Neill, Z. (2025). Large language models for building energy applications: Opportunities and challenges. *Building Simulation, 18*(2), 225–234. https://doi.org/10.1007/s12273-025-1235-9

Liu, X., Hu, J., Li, Y., Diao, S., Liang, Y., Hooi, B., & Zimmermann, R. (2024). UniTime: A language-empowered unified model for cross-domain time series forecasting. *arXiv*. http://arxiv.org/abs/2310.09751.

Long, Y., & Zhang, E. (2021). *Data augmented design: Embracing new data for sustainable urban planning and design*. Springer International Publishing. https://doi.org/10.1007/978-3-030-49618-0

Luo, Y., Cao, Z., Jin, X., Liu, K., & Yin, L. (2024). Deciphering human mobility: Inferring semantics of trajectories with large language models. In *2024 25th IEEE international conference on mobile data management (MDM)* (pp. 289–294). https://doi.org/10.1109/MDM61037.2024.00060

Ma, X., Karimpour, A., & Wu, Y.-J. (2023). Eliminating the impacts of traffic volume variation on before and after studies: A causal inference approach. *Journal of Intelligent Transportation Systems, 1–15*. https://doi.org/10.1080/15472450.2023.2245327

Madani, N., Bagalkotkar, A., Anand, S., Arnson, G., Srihari, R., & Joseph, K. *A recipe for building a compliant real estate chatbot. arXiv.Org*. (2024, October 7). https://arxiv.org/abs/2410.10860v1.

Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R., & Lao, N. (2023). On the opportunities and challenges of foundation models for geospatial artificial intelligence. *arXiv*. http://arxiv.org/abs/2304.06798.

Makridakis, S., Petropoulos, F., & Kang, Y. (2023). Large language models: Their success and impact. *Forecasting, 5(3), Article 3*. https://doi.org/10.3390/forecast5030030

Malekzadeh, M., Willberg, E., Torkko, J., & Toivonen, T. (2024). Urban visual appeal according to ChatGPT: Contrasting AI and human insights. *arXiv*. http://arxiv.org/abs/2407.14268.

Manvi, R., Khanna, S., Burke, M., Lobell, D., & Ermon, S. (2024). *Large language models are geographically biased. arXiv*. https://doi.org/10.48550/arXiv.2402.02680

Memduhoğlu, A., Fulman, N., & Zipf, A. (2024). Enriching building function classification using large language model embeddings of OpenStreetMap tags. *Earth Science Informatics, 17*(6), 5403–5418. https://doi.org/10.1007/s12145-024-01463-8

Mo, G., Xing, Z., Zhu, Y., Huang, L., Chen, W., & Li, J. (2025). Social media data-based rapid hazard assessment of urban waterlogging event: A case study of Guilin 6.19 waterlogging. *Water, 17(3), Article 3*. https://doi.org/10.3390/w17030354

Muhtar, D., Li, Z., Gu, F., Zhang, X., & Xiao, P. (2024). LHRS-bot: Empowering remote sensing with VGI-enhanced large multimodal language model. *arXiv*. http://arxiv.org/abs/2402.02544.

Musleh, M., & Mokbel, M. F. (2023). Kamel: A scalable BERT-based system for trajectory imputation. *Proceedings of the VLDB Endowment, 17*(3), 525–538. https://doi.org/10.14778/3632093.3632113

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., … Mian, A. (2024). *A comprehensive overview of large language models. arXiv*. https://doi.org/10.48550/arXiv.2307.06435

Nwafor, E., Vaughan, R., & Kolimago, C. (2021). Covid vaccine sentiment analysis by geographic region. In *2021 IEEE international conference on big data (big data)* (pp. 4401–4404). https://doi.org/10.1109/BigData52589.2021.9671854

Onsu, M. A., Lohan, P., Kantarci, B., Syed, A., Andrews, M., & Kennedy, S. (2025). *Leveraging multimodal-LLMs assisted by instance segmentation for intelligent traffic monitoring. arXiv*. https://doi.org/10.48550/arXiv.2502.11304

OpenAI. (n.d.). Create GPTs. Retrieved 19 April 2025, from https://chatgpt.com.

O'Sullivan, A. (2012). *Urban economics* (8. ed.). Chicago: Irwin.

Ouyang, T., Zhang, X., Han, Z., Shang, Y., & Li, Y. (2024). Health CLIP: Depression rate prediction using health related features in satellite and street view images. In *Companion proceedings of the ACM web conference 2024* (pp. 1142–1145). https://doi.org/10.1145/3589335.3651451

Paganelli, F. (2019). Urban mobility and transportation. In W. Leal Filho, U. Azeiteiro, A. M. Azul, L. Brandli, P. G. Özuyar, & T. Wall (Eds.), *Sustainable cities and communities* (pp. 1–13). Springer International Publishing. https://doi.org/10.1007/978-3-319-71061-7_67-1.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135. https://doi.org/10.1561/1500000011

Papadimitriou, C. H., Raghavan, P., Tamaki, H., & Vempala, S. (2000). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences, 61*(2), 217–235. https://doi.org/10.1006/jcss.2000.1711

Patel, T. A., & Nyirenda, C. N. (2022). A twitter-driven deep learning mechanism for the determination of vehicle hijacking spots in cities. *arXiv*. http://arxiv.org/abs/2208.10280.

Penumadu, V. K., Methani, N., & Sohoney, S. (2022). Learning geospatially aware place embeddings via weak-supervision. In *Proceedings of the 30th international conference on advances in geographic information systems* (pp. 1–10). https://doi.org/10.1145/3557915.3561016

Prabhune, A., Sri Hari, V., Sethiya, N. K., & Gauniyal, M. (2024). Developing causality and severity assessment frameworks for food safety signals using social media reviews: A technical report based on data from an urban Indian suburb. *Cureus*. https://doi.org/10.7759/cureus.64426

Pratama, D., & Akbar, S. (2023). Analysis of public opinion on public transportation in Bandung and Jakarta in Twitter using Indonesian bidirectional encoder representations from transformer. In *2023 IEEE international conference on industry 4.0, artificial intelligence, and communications technology (IAICT)* (pp. 179–183). https://doi.org/10.1109/IAICT59002.2023.10205608

Pullanikkat, R., Poddar, S., Das, A., Jaiswal, T., Singh, V. K., Basu, M., & Ghosh, S. (2024). Utilizing the twitter social media to identify transportation-related grievances in Indian cities. *Social Network Analysis and Mining, 14*(1), 118. https://doi.org/10.1007/s13278-024-01278-x

Qin, Z., He, D., Li, Q., Meng, B., Zhou, J., Chen, S., Chen, Z., & Chen, J. (2025). Research on the Chinese capital culture perception based on weibo big data. *Heliyon, 11*(4). https://doi.org/10.1016/j.heliyon.2025.e42758

Qu, Y., & Wang, J. (2024). Performance and biases of large language models in public opinion simulation. *Humanities and Social Sciences Communications, 11*(1), 1–13. https://doi.org/10.1057/s41599-024-03609-x

Quan, S. J., & Lee, S. (2025). Enhancing participatory planning with ChatGPT-assisted planning support systems: A hypothetical case study in Seoul. *International Journal of Urban Sciences, 29*(1), 89–122. https://doi.org/10.1080/12265934.2025.2462823

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog, 1*(8), 9.

Ramalingam, S. P., & Kumar, V. (2025). Building usage prediction in complex urban scenes by fusing text and facade features from street view images using deep learning. *Building and Environment, 267*, Article 112174. https://doi.org/10.1016/j.buildenv.2024.112174

Ramos-Henriquez, J. M., & Morini-Marrero, S. (2025). Airbnb customer experience in long-term stays: A structural topic model and ChatGPT-driven analysis of the reviews of remote workers. *International Journal of Contemporary Hospitality Management, 37* (1), 161–179. https://doi.org/10.1108/IJCHM-01-2024-0034

Ran, C., Xueqi, Y., Xuhui, J., Zhengqi, H., Jingze, G., Xianyue, Z., Chunyu, L., Chumin, L., Jing, Z., Zeke, L., Jingjing, Z., & Keke, L. (2023). HSC-GPT: A large language model for human settlements construction. *arXiv*. http://arxiv.org/abs/2401.00504.

Ren, Y., Chen, Y., Liu, S., Wang, B., Yu, H., & Cui, Z. (2024). TPLLM: A traffic prediction framework based on pretrained large language models. *arXiv*. http://arxiv.org/abs/2403.02221.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). *A systematic survey of prompt engineering in large language models: Techniques and applications. arXiv*. https://doi.org/10.48550/arXiv.2402.07927

Saito, R., & Haruyama, S. (2023). Estimating time-series changes in social sentiment @twitter in U.S. metropolises during the COVID-19 pandemic. *Journal of Computational Social Science, 6*(1), 359–388. https://doi.org/10.1007/s42001-022-00186-4

Saito, R., & Tsugawa, S. (2025). Understanding citizens' response to social activities on twitter in US metropolises during the COVID-19 recovery phase using a fine-tuned large language model: Application of AI. *Journal of Medical Internet Research, 27*(1), Article e63824. https://doi.org/10.2196/63824

Salam, S., Islam, M. S., Ahmed, F., Khan, L., Kim, D., Allo, N., & Nwariaku, O. (2021). Exploring the roles of social media data to identify the locations and severity of road traffic accidents. In *2021 IEEE fourth international conference on artificial intelligence and knowledge engineering (AIKE)* (pp. 62–71). https://doi.org/10.1109/AIKE52691.2021.00016

Santamouris, M. (Ed.). (2013). *Energy and climate in the urban built environment*. Routledge. https://doi.org/10.4324/9781315073774.

Scribano, C., Sapienza, D., Franchini, G., Verucchi, M., & Bertogna, M. (2021). All you can embed: Natural language based vehicle retrieval with spatio-temporal transformers. *arXiv*. http://arxiv.org/abs/2106.10153.

Sevim, N., Ibrahim, M., & Ekin, S. (2024). *Large language models (LLMs) assisted wireless network deployment in urban settings. arXiv*. https://doi.org/10.48550/arXiv.2405.13356

Shahrabani, M. M. N., & Apanaviciene, R. (2024). An AI-based evaluation framework for smart building integration into smart city. *Sustainability, 16(18), Article 18*. https://doi.org/10.3390/su16188032

Shahriar, S., Lund, B., Mannuru, N. R., Arshad, M. A., Hayawi, K., Bevara, R. V. K., Mannuru, A., & Batool, L. (2024). Putting GPT-4o to the sword: A comprehensive evaluation of language, vision, speech, and multimodal proficiency. *arXiv*. http://arxiv.org/abs/2407.09519.

Shalf, J. (2020). The future of computing beyond Moore's law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 378*(2166), Article 20190061. https://doi.org/10.1098/rsta.2019.0061

Shang, Y., Chen, J., Fan, H., Ding, J., Feng, J., & Li, Y. (2024). UrbanWorld: An urban world model for 3D city generation. *arXiv*. http://arxiv.org/abs/2407.11965.

Shi, J., Yang, Z., & Xu, W. (2023). Road-constrained trajectory recovery via BERT-based multi-task learning. In I. L. Wang, & X. Liu (Eds.), *International conference on intelligent systems, communications, and computer networks (ISCCN 2023), 12702* (pp. 606–611). https://doi.org/10.1117/12.2679598

Siering, M., Jan, M., Grčar, M., Jožef Stefan Institute, & Slovenia.. (2021). Design principles for robust fraud detection: The case of stock market manipulations. *Journal of the Association for Information Systems, 22*(1), 156–178. https://doi.org/10.17705/1jais.00657

Singh, S., Fore, M., & Stamoulis, D. (2024). GeoLLM-engine: A realistic environment for building geospatial copilots. *arXiv*. http://arxiv.org/abs/2404.15500.

Solatorio, A. V. (2023). GeoFormer: Predicting human mobility using generative pre-trained transformer (GPT). In *Proceedings of the 1st international workshop on the human mobility prediction challenge, 11–15*. https://doi.org/10.1145/3615894.3628499

Song, S., Li, X., Li, S., Zhao, S., Yu, J., Ma, J., Mao, X., & Zhang, W. (2023). How to bridge the gap between modalities: A comprehensive survey on multimodal large language model. *arXiv*. http://arxiv.org/abs/2311.07594.

Spennemann, D. H. R. (2024). Will artificial intelligence affect how cultural heritage will be managed in the future? Responses generated by four genAI models. *Heritage, 7*(3), 1453–1471. https://doi.org/10.3390/heritage7030070

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. *Proceedings of the AAAI Conference on Artificial Intelligence, 34(09), Article 09*. https://doi.org/10.1609/aaai.v34i09.7123

Sun, M., Li, Q., Pan, J., Wei, H., Liu, C., Xu, X., & Li, Y. (2025). POI extraction from digital city: An engineering exploration with large-language models. *Expert Systems, 42*(4), Article e70001. https://doi.org/10.1111/exsy.70001

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). *Understanding the capabilities, limitations, and societal impact of large language models. arXiv*. https://doi.org/10.48550/arXiv.2102.02503

Tan, C., Cao, Q., Li, Y., Zhang, J., Yang, X., Zhao, H., Wu, Z., Liu, Z., Yang, H., Wu, N., Tang, T., Ye, X., Chai, L., Liu, N., Li, C., Mu, L., Liu, T., & Mai, G. (2023). On the promises and challenges of multimodal foundation models for geographical, environmental, agricultural, and urban planning applications. *arXiv*. http://arxiv.org/abs/2312.17016.

Tang, J., & Long, Y. (2019). Measuring visual quality of street space and its temporal variation: Methodology and its application in the hutong area in Beijing. *Landscape and Urban Planning, 191*, Article 103436. https://doi.org/10.1016/j.landurbplan.2018.09.015

Tang, P., Yang, C., Xing, T., Xu, X., Jiang, R., & Sezaki, K. (2024). *Instruction-tuning llama-3-8B excels in city-scale mobility prediction. arXiv*. https://doi.org/10.48550/arXiv.2410.23692

Terashima, H., Tamura, N., Shoji, K., Katayama, S., Urano, K., Yonezawa, T., & Kawaguchi, N. (2023). Human mobility prediction challenge: Next location prediction using spatiotemporal BERT. In *Proceedings of the 1st international workshop on the human mobility prediction challenge, 1–6*. https://doi.org/10.1145/3615894.3628498

Topsakal, O., & Akinci, T. C. (2023). Creating large language model applications utilizing LangChain: A primer on developing LLM apps fast. *International Conference on Applied Engineering and Natural Sciences, 1*(1), 1050–1056. https://doi.org/10.59287/icaens.1127

Ullah, A., Qi, G., Hussain, S., Ullah, I., & Ali, Z. (2024). *The role of LLMs in sustainable smart cities: Applications, challenges, and future directions. arXiv*. https://doi.org/10.48550/arXiv.2402.14596

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). *Attention is all you need. arXiv*. https://doi.org/10.48550/arXiv.1706.03762

Veltmeijer, E., & Gerritsen, C. (2023). SentiMap: Domain-adaptive geo-spatial sentiment analysis. In *2023 IEEE 17th international conference on semantic computing (ICSC)* (pp. 17–24). https://doi.org/10.1109/ICSC56153.2023.00010

Verma, D., Mumm, O., & Carlow, V. M. (2023). Generative agents in the streets: Exploring the use of large language models (LLMs) in collecting urban perceptions. *arXiv*. http://arxiv.org/abs/2312.13126.

Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. (2019). SuperGLUE: A stickier benchmark for general-purpose language understanding systems. *Advances in Neural Information Processing Systems, 32*. https://proceedings.neurips.cc/paper/2019/hash/4496bf24afe7fab6f046bf4923da8de6-Abstract.html.

Wang, D., Wang, Y., Fu, X., Dou, M., Dong, S., & Zhang, D. (2023). Revealing the spatial co-occurrence patterns of multi-emotions from social media data. *Telematics and Informatics, 84*, Article 102025. https://doi.org/10.1016/j.tele.2023.102025

Wang, J. (2018). Urban policy. In A. Farazmand (Ed.), *Global encyclopedia of public administration, public policy, and governance*. Springer International Publishing. https://doi.org/10.1007/978-3-319-20928-9_226.

Wang, M., Pang, A., Kan, Y., Pun, M.-O., Chen, C. S., & Huang, B. (2024). LLM-assisted light: Leveraging large language model capabilities for human-mimetic traffic signal control in complex urban environments. *arXiv*. http://arxiv.org/abs/2403.08337.

Wang, P., Zhang, Y., Hu, T., & Zhang, T. (2023). Urban traffic flow prediction: A dynamic temporal graph network considering missing values. *International Journal of Geographical Information Science, 37*(4), 885–912. https://doi.org/10.1080/13658816.2022.2146120

Wang, S., Liu, C., & Zhou, Z. (2024). Government-enterprise green collaborative governance and urban carbon emission reduction: Empirical evidence from green PPP programs. *Environmental Research, 257*, Article 119335. https://doi.org/10.1016/j.envres.2024.119335

Wang, S., Zhao, Z., Xie, Y., Ma, M., Chen, Z., Wang, Z., Su, B., Xu, W., & Li, T. (2024). *Recent surge in public interest in transportation: Sentiment analysis of baidu apollo go using weibo data. arXiv*. https://doi.org/10.48550/arXiv.2408.10088

Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: Negative sentiment analysis on social media in China via BERT model. *IEEE Access, 8*, 138162–138169. https://doi.org/10.1109/ACCESS.2020.3012595

Wang, X., Ling, X., Zhang, T., Li, X., Wang, S., Li, Z., Zhang, L., & Gong, P. (2023). Optimizing and fine-tuning large language model for urban renewal. *arXiv*. http://arxiv.org/abs/2311.15490.

Wang, X., & Long, Y. (2023). Future shrinking cities on the globe: A projection map for 2020–2100 based on global gridded population dataset. *Environment and Planning B: Urban Analytics and City Science, 50*(7), 1994–1997. https://doi.org/10.1177/23998083231186153

Wang, Y., Nagireddy, S. R., Thota, C. T., Ho, D. H., & Lee, Y. (2022). Community-in-the-loop: Creating artificial process intelligence for co-production of city service. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW2), 1–21. https://doi.org/10.1145/3555176

Wu, J., Gan, W., Chao, H.-C., & Yu, P. S. (2024). *Geospatial big data: Survey and challenges. arXiv*. https://doi.org/10.48550/arXiv.2404.18428

Wu, Y., Xu, J., & Lin, Y. (2022). Social media-based analysis of bus service satisfaction. In Z. Zhang (Ed.)*, 901. 2021 6th international conference on intelligent transportation engineering (ICITE 2021)* (pp. 540–547). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-2259-6_47.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., … Gui, T. (2023). The rise and potential of large language model based agents: A survey. *arXiv*. http://arxiv.org/abs/2309.07864.

Xiao, B., Yin, Z., & Shan, Z. (2023). Simulating public administration crisis: A novel generative agent-based simulation system to lower technology barriers in social science research. *arXiv*. http://arxiv.org/abs/2311.06957.

Xiao, X., Fang, C., Lin, H., Liu, L., Tian, Y., & He, Q. (2022). Exploring spatiotemporal changes in the multi-granularity emotions of people in the city: A case study of Nanchang, China. *Computational Urban Science, 2*(1), 1. https://doi.org/10.1007/s43762-021-00030-x

Xu, F., Zhang, J., Gao, C., Feng, J., & Li, Y. (2023). *Urban generative intelligence (UGI): A foundational platform for agents in embodied city environment. arXiv*. https://doi.org/10.48550/arXiv.2312.11813

Xu, L., Du, Z., Mao, R., Zhang, F., & Liu, R. (2020). GSAM: A deep neural network model for extracting computational representations of Chinese addresses fused with geospatial feature. *Computers, Environment and Urban Systems, 81*, Article 101473. https://doi.org/10.1016/j.compenvurbsys.2020.101473

Xue, H., Voutharoja, B. P., & Salim, F. D. (2022). Leveraging language foundation models for human mobility forecasting. In *Proceedings of the 30th international conference on advances in geographic information systems* (pp. 1–9). https://doi.org/10.1145/3557915.3561026

Yan, B., Mai, F., Wu, C., Chen, R., & Li, X. (2024). A computational framework for understanding firm communication during disasters. *Information Systems Research, 35*(2), 590–608. https://doi.org/10.1287/isre.2022.0128

Yan, Y., Wen, H., Zhong, S., Chen, W., Chen, H., Wen, Q., … Liang, Y. (2024). *UrbanCLIP: Learning text-enhanced urban region profiling with contrastive language-image pretraining from the web. arXiv*. https://doi.org/10.48550/arXiv.2310.18340

Yang, Z., Lin, X., He, Q., Huang, Z., Liu, Z., Jiang, H., Shu, P., Wu, Z., Li, Y., Law, S., Mai, G., Liu, T., & Yang, T. (2024). *Examining the commitments and difficulties inherent in multimodal foundation models for street view imagery. arXiv*. https://doi.org/10.48550/arXiv.2408.12821

Yao, H., Da, L., Nandam, V., Turnau, J., Liu, Z., Pang, L., & Wei, H. (2025). *CoMAL: Collaborative multi-agent large language models for mixed-autonomy traffic. arXiv*. https://doi.org/10.48550/arXiv.2410.14368

Yao, J. (2023). Elevating urban tourism: Data-driven insights and AI-powered personalization with large language models brilliance. In *2023 IEEE 3rd international conference on social sciences and intelligence management (SSIM)* (pp. 138–143). https://doi.org/10.1109/SSIM59263.2023.10468915

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). *A survey on multimodal large language models. arXiv*. https://doi.org/10.48550/arXiv.2306.13549

Yu, B., Xing, H., Ge, W., Zhou, L., Yan, J., & Li, Y. (2024). Advanced susceptibility analysis of ground deformation disasters using large language models and machine learning: A Hangzhou city case study. *PLoS ONE, 19*(12), Article e0310724. https://doi.org/10.1371/journal.pone.0310724

Yu, J., & McKinley, G. (2024). Synthetic participatory planning of shared automated electric mobility systems. *Sustainability, 16*(13), 5618. https://doi.org/10.3390/su16135618

Yu, S., Yang, Y., Li, J., Guo, K., Wang, Z., & Liu, Y. (2024). Exploring low-carbon and sustainable urban transformation design using ChatGPT and artificial bee colony algorithm. *Humanities and Social Sciences Communications, 11*(1), 252. https://doi.org/10.1057/s41599-024-02765-4

Yuan, Y., Ding, J., Feng, J., Jin, D., & Li, Y. (2024). UniST: A prompt-empowered universal model for urban spatio-temporal prediction. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining, 4095–4106*. https://doi.org/10.1145/3637528.3671662

Yuan, Z., Lai, S., & Liu, H. (2025). *CoLLMLight: Cooperative large language model agents for network-wide traffic signal control. arXiv*. https://doi.org/10.48550/arXiv.2503.11739

Zamiralov, A., Khodorchenko, M., & Nasonov, D. (2020). Detection of housing and utility problems in districts through social media texts. *Procedia Computer Science, 178*, 213–223. https://doi.org/10.1016/j.procs.2020.11.023

Zeng, Q., Yang, Q., Dong, S., Du, H., Zheng, L., Xu, F., & Li, Y. (2024). *Perceive, reflect, and plan: Designing LLM agent for goal-directed city navigation without instructions. arXiv*. https://doi.org/10.48550/arXiv.2408.04168

Zeng, Z., Goo, J. M., Wang, X., Chi, B., Wang, M., & Boehm, J. (2024). Zero-shot building age classification from facade image using GPT-4. *arXiv*. http://arxiv.org/abs/2404.09921.

Zengin, M. S., Arslan, R., & Akgun, M. B. (2022). *Distributed sentiment analysis for geo-tagged twitter data. 2022 30th signal processing and communications applications conference (SIU), 1–4*. https://doi.org/10.1109/SIU55565.2022.9864702

Zenkert, J., & Fathi, M. (2024). *Taxonomy mining from a smart city CMS using the multidimensional knowledge representation approach. 2024 IEEE 14th annual computing and communication workshop and conference (CCWC), 0257–0261*. https://doi.org/10.1109/CCWC60891.2024.10427816

Zhang, B., Sun, X., Zhou, J., Li, X., Liu, D., & Wang, S. (2023). Analysis of public opinion evolution in public health emergencies based on multi-fusion model. In S. Yang, & S. Islam (Eds.), *Web and big data. APWeb-WAIM 2022 international workshops, 1784, 110–124*. Springer Nature Singapore. https://doi.org/10.1007/978-981-99-1354-1_11.

Zhang, J., Li, Y., Fukuda, T., & Wang, B. *Revolutionizing urban safety perception assessments: Integrating multimodal large language models with street view images. arXiv. Org.* (2024, July 29). https://arxiv.org/abs/2407.19719v2.

Zhang, R., Han, L., Sun, L., Liu, Y., Wang, J., & Lv, W. (2023). Regions are who walk them: A large pre-trained spatiotemporal model based on human mobility for ubiquitous urban sensing. *arXiv*. http://arxiv.org/abs/2311.10471.

Zhang, W., Liu, Y., Wang, X., Chen, X., Gao, C., & Chen, X. (2024). Demo abstract: Embodied aerial agent for city-level visual language navigation using large language model. In *2024 23rd ACM/IEEE international conference on information processing in sensor networks (IPSN)* (pp. 265–266). https://doi.org/10.1109/IPSN61024.2024.00033

Zhang, X., Shah, J., & Han, M. (2023). ChatGPT for fast learning of positive energy district (PED): A trial testing and comparison with expert discussion results. *Buildings, 13*(6), 1392. https://doi.org/10.3390/buildings13061392

Zhang, Y., Chen, Z., Zheng, X., Chen, N., & Wang, Y. (2021). Extracting the location of flooding events in urban systems and analyzing the semantic risk using social sensing data. *Journal of Hydrology, 603*, Article 127053. https://doi.org/10.1016/j.jhydrol.2021.127053

Zhang, Y., Zhang, F., & Chen, N. (2022). Migratable urban street scene sensing method based on vision language pre-trained model. *International Journal of Applied Earth Observation and Geoinformation, 113*, Article 102989. https://doi.org/10.1016/j.jag.2022.102989

Zhang, Z., Sun, Y., Wang, Z., Nie, Y., Ma, X., Sun, P., & Li, R. (2024). *Large language models for mobility in transportation systems: A survey on forecasting tasks. arXiv*. https://doi.org/10.48550/arXiv.2405.02357

Zhao, Y., Xu, K., Zhu, Z., Hu, Y., Zheng, Z., Chen, Y., Ji, Y., Gao, C., Li, Y., & Huang, J. (2025). *CityEQA: A hierarchical LLM agent on embodied question answering benchmark in city space. arXiv*. https://doi.org/10.48550/arXiv.2502.12532

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology, 5(3)*. https://doi.org/10.1145/2629592, 38:1-38:55.

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S., & Sun, L. (2023). A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT. *arXiv*. http://arxiv.org/abs/2302.09419.

Zhou, J., & Hu, H. (2025). Citizens' appeals classification using a mixed method: A case study of Yichang government public service hotline data. *Information Development, 02666669241312899*. https://doi.org/10.1177/02666669241312899

Zhou, M., Wang, Y., Hou, J., Luo, C., Zhang, Z., & Peng, J. (2024). *SceneX:Procedural controllable large-scale scene generation via large-language models. arXiv*. https://doi.org/10.48550/arXiv.2403.15698

Zhou, Z., Lin, Y., & Li, Y. (2024). Large language model empowered participatory urban planning. *arXiv*. http://arxiv.org/abs/2402.01698.

Zhu, H., Zhang, W., Huang, N., Li, B., Niu, L., Fan, Z., … Liu, X. (2024). *PlanGPT: Enhancing urban planning with tailored language model and efficient retrieval. arXiv*. https://doi.org/10.48550/arXiv.2402.19273

Zou, X., Yan, Y., Hao, X., Hu, Y., Wen, H., Liu, E., Zhang, J., Li, Y., Li, T., Zheng, Y., & Liang, Y. (2025). Deep learning for cross-domain data fusion in urban computing: Taxonomy, advances, and outlook. *Information Fusion, 113*, Article 102606. https://doi.org/10.1016/j.inffus.2024.102606

Zuo, J., & Li, K. (2021). Research on named entity recognition based on text data of Beijing subway accident. In *2021 international conference on internet, education and information technology (IEIT)* (pp. 114–118). https://doi.org/10.1109/IEIT53597.2021.00032