Engineering, Construction and Architectural Management

159

Received 4 November 2025
Revised 17 January 2026
Accepted 1 February 2026

Large language models in smart construction: a systematic review of implementation strategies, applications and future directions

Yu Gao and Tak Wing Yiu
School of Built Environment, University of New South Wales, Sydney, Australia
Xuesong Shen
School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia, and
Vivian W.Y. Tam
School of Engineering, Design and Built Environment, Western Sydney University, Sydney, Australia

Abstract

Purpose – With accelerating advances in artificial intelligence (AI) and digital transformation, the construction domain is rapidly adopting large language models (LLMs) to improve efficiency, safety and sustainability. However, a comprehensive understanding of their applications, adaptation strategies and evaluation practices in this domain remains limited. This paper provides a systematic landscape of LLM research and applications in construction, clarifying their roles, integration pathways and associated challenges.

Design/methodology/approach – A systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guideline, combining bibliometric and content analysis of publications from 2020 to 2025. The review examines the evolution, implementation strategies and performance characteristics of LLMs in construction domains.

Findings – LLMs have been increasingly implemented across diverse construction tasks. Most studies utilize pretrained models, adapted through fine-tuning, retrieval-augmented generation or prompt engineering. While LLMs demonstrate strong reasoning and semantic understanding capabilities, challenges persist in domain adaptation, hallucination control, data privacy and lifecycle integration.

Practical implications – The results provide insights for researchers and practitioners on how LLMs can be effectively implemented throughout the construction domain. The paper highlights the need for developing domain-specific evaluation standards, establishing governance frameworks and aligning LLM applications with existing digital ecosystems such as building information modeling and digital twins.

Originality/value – This paper presents a comprehensive systematic review focused on LLM applications in construction. By mapping the technological landscape and summarizing key implementation strategies, it establishes a theoretical and practical foundation for advancing intelligent, data-driven and human-centered construction in the era of generative AI.

Keywords Large language models (LLMs), Smart construction, Construction management, Systematic review

Paper type Literature review

1. Introduction

The construction industry is a critical pillar of the global economy, contributing approximately 13% to global gross domestic product (Abdelalim *et al.*, 2025). However, it has long been

Funding: This work was supported by Australian Research Council (award number: IC240100020 and IH200100010).

Engineering, Construction and Architectural Management
Vol. 33 No. 15, 2026
pp. 159-181
Emerald Publishing Limited
e-ISSN: 1365-232X
p-ISSN: 0969-9988
DOI 10.1108/ECAM-10-2025-1668

characterized by slow adoption of innovation and lagging productivity growth (Gambatese and and Hallowell, 2011). These limitations are largely attributed to its project-based nature, the need for multi-stakeholder collaboration, and a strong dependence on information integration and sharing. In recent years, technologies such as building information modeling (BIM), the Internet of Things, digital twins, big data analytics, robotics, and artificial intelligence (AI) have been increasingly adopted to address challenges including design complexity, low construction efficiency, and high operational costs (Zhang *et al.*, 2013). Smart construction refers to the digitally enabled, lifecycle-wide management of construction processes for efficiency and collaboration (Wu *et al.*, 2022). Driven by digital transformation, AI subfields such as machine learning, computer vision, and natural language processing (NLP) have been widely applied to construction-related tasks (Abioye *et al.*, 2021). Nevertheless, construction data are often fragmented, semantically heterogeneous, and poorly integrated across lifecycle stages. These characteristics pose significant challenges for traditional AI models in terms of generalization, contextual understanding, and task transferability (Akinosho *et al.*, 2020).

The emergence of generative AI, particularly Large Language Models (LLMs), marks a paradigm shift in AI capabilities. LLMs are a category of AI models with billions of parameters, trained on massive unlabeled text using self-supervised learning (Raiaan *et al.*, 2024). LLMs have already achieved notable success in data-intensive fields such as healthcare (Peng *et al.*, 2023), particularly in tasks involving information extraction, dialog systems, and generative outputs. In the construction domain, LLMs are emerging as a promising tool for enhancing design workflows, enabling intelligent management, and optimizing building operations (Saka *et al.*, 2024). Moreover, with fine-tuning and contextual prompting, LLMs can be adapted to diverse construction scenarios to support complex tasks such as automated safety inspection (Wang *et al.*, 2025a), building energy modeling (Jiang *et al.*, 2024), energy performance diagnostics (Liu *et al.*, 2025b), and design change management (Jang and Lee, 2024). These capabilities position LLMs as key enablers of smart construction.

Despite preliminary explorations of LLMs in construction, current studies remain fragmented and mostly tool-oriented, lacking a comprehensive analysis of their technological foundations, application pathways, and implementation challenges. Existing reviews have largely focused on traditional AI techniques, such as visual recognition (Akinosho *et al.*, 2020). However, LLM-specific discussions within these reviews remain limited. Saka *et al.* (2024) explored the potential of ChatGPT in construction management, yet did not establish a dedicated framework centered on LLMs. Taiwo *et al.* (2024) provided a comprehensive overview of generative AI applications in construction and proposed a data-driven approach for customized model development, but their focus remained on general generative models rather than LLMs. While their study offered valuable case summaries and identified implementation barriers, it primarily adopted a descriptive perspective, lacking a task-driven analytical framework and a deeper examination of model evolution or multi-phase integration strategies.

Accordingly, this paper aims to systematically review the research progress, application patterns, and key challenges of LLMs in the construction domain through a combination of bibliometric and content analysis, thereby addressing the existing lack of systematic understanding in this emerging field. The paper is guided by the following core research questions.

(1) What are the primary functional roles and technological pathways of LLMs in the construction sector?

(2) How are LLMs implemented and integrated across different construction stages and task types?

(3) What performance limitations, ethical concerns, and practical challenges do LLM applications face in construction?

This paper makes three main contributions. First, it develops an integrated framework to organize LLM research in the construction domain. Second, it provides a structured synthesis of architectures, implementation strategies, and tool ecosystems, and identifies representative task patterns and data-flow characteristics across the building lifecycle. Third, it outlines a research roadmap for smart construction that translates these findings into actionable directions. Beyond technical insights, the paper also highlights societal implications, particularly the potential effects on construction safety, workforce structure and reskilling, and the governance mechanisms needed to support responsible and accountable adoption.

2. Research methods and process

This study adopts a systematic literature review methodology guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Page *et al.*, 2021). PRISMA provides a well-established and empirically validated framework for documenting the identification, screening, eligibility, and inclusion of studies, and has been widely adopted in high-quality reviews within construction management (Gao *et al.*, 2025b). This approach is particularly suitable for emerging research domains where empirical evidence is fragmented and rapidly evolving, as it allows systematic consolidation of dispersed findings into a structured analytical framework. Accordingly, a top-down review strategy was employed, starting from broad search terms and progressively refining the dataset using predefined inclusion and exclusion criteria. The review process consists of five main steps, as illustrated in Figure 1, which summarizes the identification, screening, and final selection of records included in the analysis.

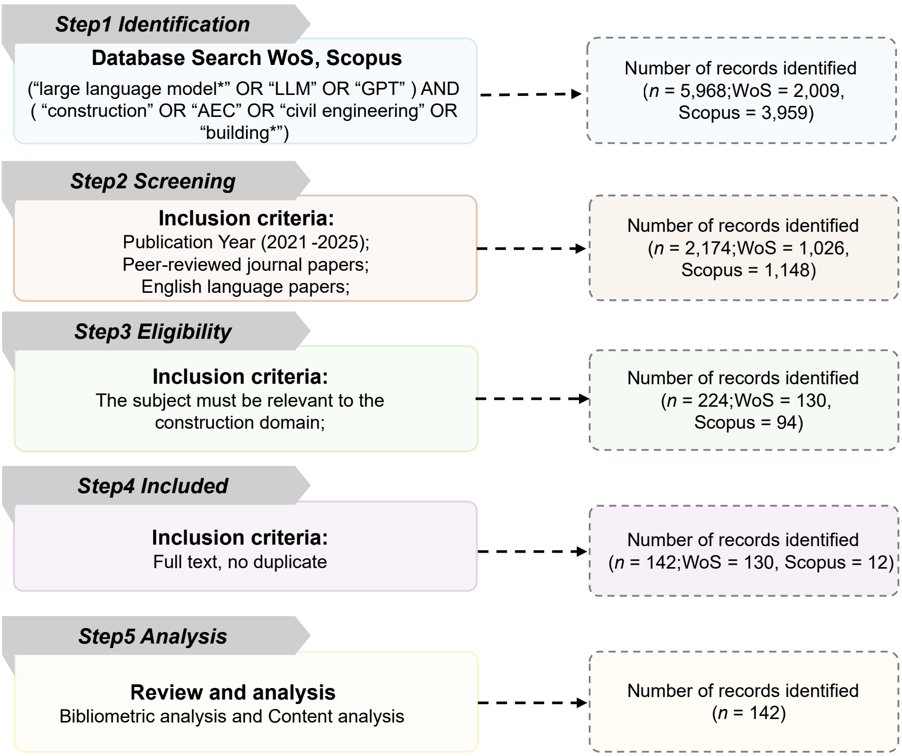

Figure 1. Flowchart of the article selection process. Source(s): Authors' own work

(1) Identification. Relevant articles were retrieved from two widely recognized academic databases Web of Science (WoS) Core Collection and Scopus due to their comprehensive coverage and high-quality indexing. The following search queries were applied:

- Scopus: (TITLE-ABS-KEY ("large language model*" OR "LLM" OR "GPT") AND TITLE-ABS-KEY ("construction" OR "architecture, engineering, and construction (AEC)" OR "civil engineering" OR "building*"))
- WoS Core Collection: "large language model*" OR "LLM" OR "GPT" (Topic) AND "construction" OR "AEC" OR "civil engineering" OR "building*" (Topic).

The "*" is used to include both singular and plural forms and related suffixes. A total of 5,968 records were initially retrieved.

(2) Screening. Initial screening was performed based on the following inclusion criteria: (1) publication year between 2021 and 2025; (2) peer-reviewed journal articles; and (3) English-language papers. After this step, 2,174 records remained.

(3) Eligibility. The remaining articles were assessed for relevance to the construction domain, based on the journal and title and abstract review, resulting in the retention of 224 records.

(4) Inclusion. Records were further screened to exclude duplicates and those without accessible full text. A total of 142 records were finally included for review.

(5) Analysis. The final dataset of 142 papers was subjected to bibliometric analysis and qualitative content analysis to identify key research trends, methodologies, and thematic distributions in the application of LLMs in the construction industry.

3. Overview of articles

3.1 Quantitative analysis

A total of 142 peer-reviewed papers were selected for analysis. As shown in Figure 2, the year-wise distribution of the selected papers reflects a significant surge in research interest in this field since 2023. Although relevant studies were rarely found before that time, the number of publications increased significantly in 2024 and continued to rise rapidly in 2025, suggesting that the application of LLMs in the construction domain has become a highly active and emerging research topic. Notably, about 95% of the collected papers were published within the past two years, highlighting the strong momentum and growing academic engagement in this area. Regarding publication venues, Figure 2 also summarizes the top ten journals that have

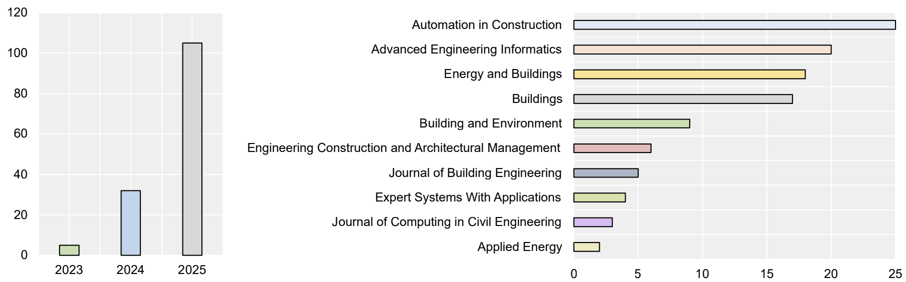

Figure 2. Yearly publication trends and journal distribution of selected papers. Source(s): Authors' own work

published the most relevant articles over the past five years. Most of these journals are widely recognized for their strong influence in construction disciplines.

To better capture the thematic structure and key research trajectories in this domain, a keyword co-occurrence analysis was conducted using VOSviewer, a widely adopted bibliometric tool known for its efficient clustering and visual output. Through this approach, the underlying relationships among frequently used keywords were mapped to reveal dominant themes and interconnections. From the 142 reviewed articles, a total of 609 unique keywords were extracted, including both author-defined and index terms. To ensure analytical clarity and reduce noise, the minimum occurrence threshold was set to three, and a manual normalization process was applied to eliminate duplicates and unify semantically similar terms. After refinement, 38 keywords met the inclusion criteria for visualization.

The resulting network, illustrated in Figure 3, presents each keyword as a node, where node size reflects its frequency of occurrence, and edge thickness indicates the strength of co-occurrence. Based on the association patterns, the keywords were grouped into five clusters, each represented by a distinct color. Among them, Cluster 1 (red) focuses on ChatGPT-driven applications in the AEC industry. Cluster 2 (green) is the largest group, centered on “large language model” and extended by “retrieval-augmented generation” and “knowledge graphs”, covering construction management, inspection, information extraction, and safety, highlighting LLMs’ role in knowledge-driven construction tasks. Cluster 3 (blue) emphasizes BIM and NLP based applications, with keywords such as “building information modeling”, “natural language processing”, “prompt engineering”, and “virtual assistant”; Cluster 4 (yellow) centers on methodological development, including terms like “framework”, “fine-tuning”, and “deep learning”; Cluster 5 (purple) relates to foundational AI and energy applications, involving “artificial intelligence”, “machine learning”, “building energy modeling”, and “architecture”.

3.2 Implementation strategies

The implementation architectures can be broadly categorized into standalone and integrated models. Standalone deployments treat the LLM as an independent application, typically built

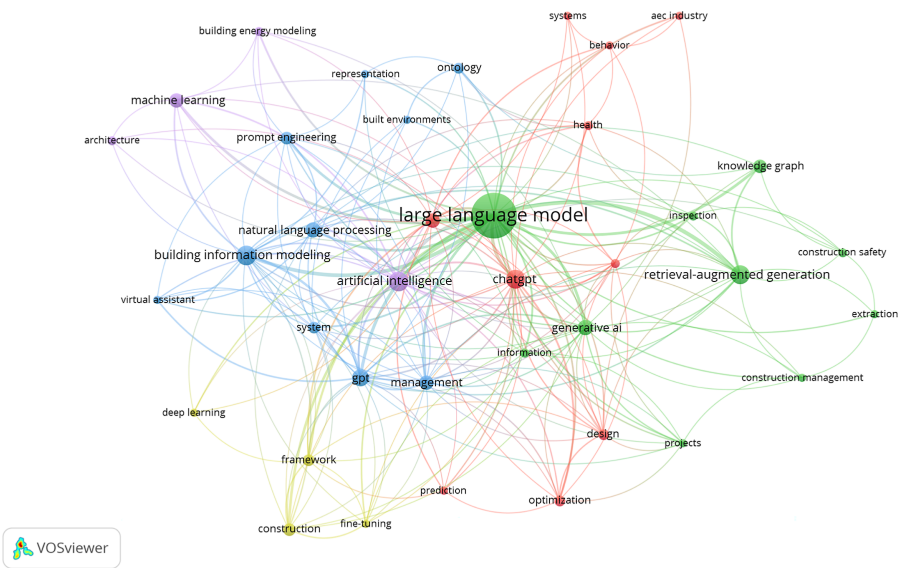

Figure 3. Keyword co-occurrence map. Source(s): Authors’ own work

for a specific task, such as an intelligent conversational assistant for building energy modeling (Li *et al.*, 2025b). This approach is flexible and quick to deploy but limited in multi-source data integration and cross-task reasoning. In contrast, integrated models embed the LLM module deeply into BIM, Digital Twin, or Internet of Things systems (Bao and Bu, 2025), connecting seamlessly to structured data (e.g. BIM models, sensor streams) via application programming interfaces (APIs) or knowledge graphs to enable context-aware reasoning and decision-making (Wang *et al.*, 2025b). As illustrated in Figure 4, three representative integration patterns can be identified: retrieval-augmented generation (RAG) pipelines, which combine information retrieval with generative reasoning to efficiently access and interpret domain knowledge (Lee *et al.*, 2024), integrating external knowledge bases to provide precise and up-to-date information (Zhang *et al.*, 2025c). In agent-based frameworks, LLMs serve as the coordination core among perception, reasoning, and control agents to achieve distributed task execution (Xiao and Xu, 2024). Cross-modal fusion systems link text, image, and sensor data for joint reasoning in tasks such as safety inspection (Wang *et al.*, 2025a). Beyond these integration architectures, some studies further enhance performance through adaptation strategies, including domain-specific fine-tuning with labeled datasets to improve task accuracy (Jiang and Chen, 2025), and prompt engineering using structured templates and dynamic context injection to align outputs with engineering-specific requirements (Zheng and Fischer, 2023).

Beyond architectural integration, implementation diversity extends across model families, data modalities, adaptation strategies, and application scenarios, revealing a complex ecosystem of LLM pathways in smart construction. As illustrated in Figure 5, these implementation routes demonstrate how LLMs are progressively aligned with construction-specific objectives, bridging unstructured language processing with structured engineering data and domain reasoning.

At the model level, five major categories dominate: GPT-based, BERT-based, LLaMA-based, Chinese-oriented models, and multimodal/vision-language models. GPT-based models play a leading role due to their strong text generation capacity, accounting for approximately 80.2% of all pretrained models adopted, particularly in compliance checking (Shi *et al.*, 2025b) and text analysis tasks (Xiao *et al.*, 2025). Bidirectional Encoder Representations from Transformers (BERT) and LLaMA models are more frequently applied in classification, reasoning, and document automation, while multimodal models are primarily directed toward

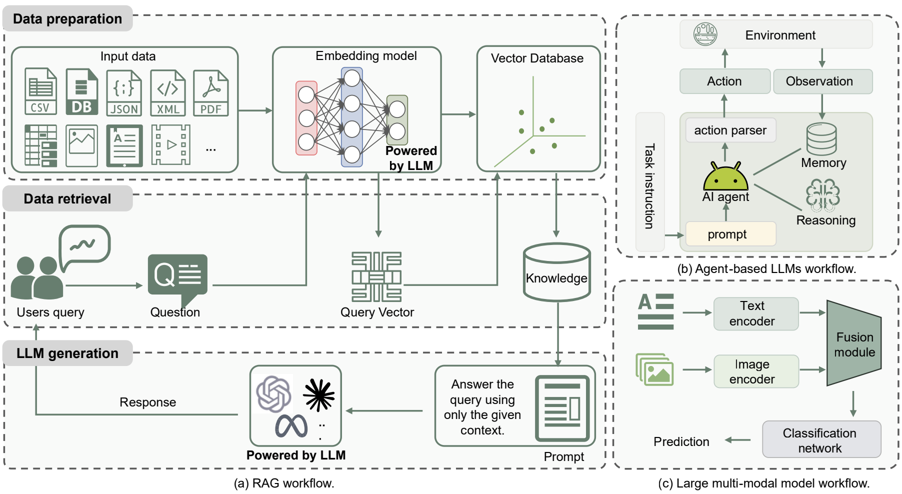

Figure 4. Typical workflows of LLMs-driven systems. Source(s): Authors' own work

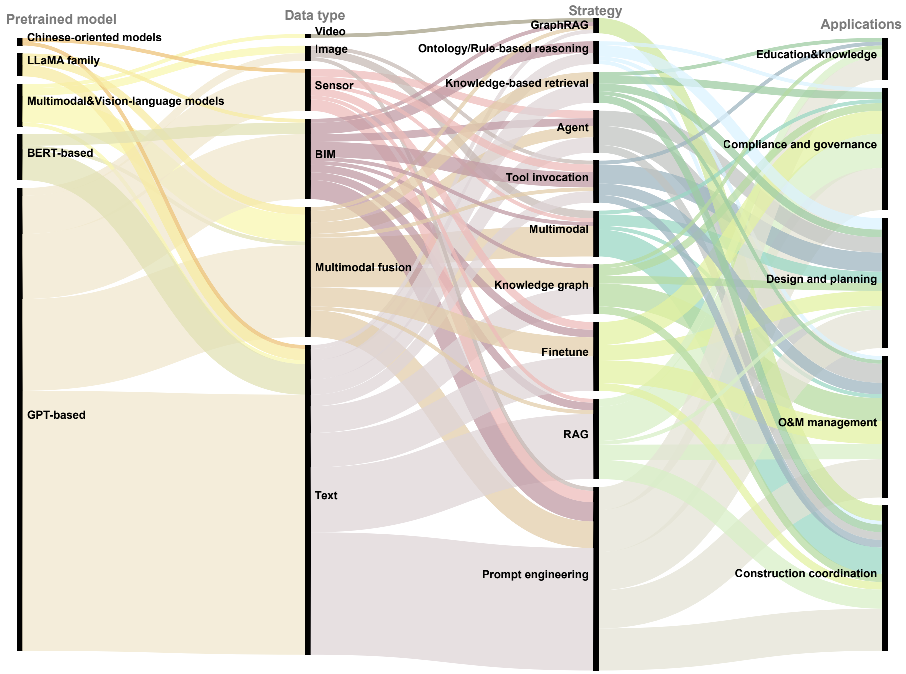

Figure 5. Implementation pathways of LLMs in smart construction. Source(s): Authors' own work

image- and video-driven tasks such as hazard detection (Luo *et al.*, 2025) and safety monitoring (Uhm *et al.*, 2025).

At the data type level, textual data remains the central medium, but substantial flows extend toward BIM data, sensor signals, images, videos, and multimodal fusion. This indicates the inherently multimodal nature of construction tasks: while language serves as the primary interface, practical applications increasingly require integration with visual (Wang *et al.*, 2025a) and sensor data (Xu *et al.*, 2025b). Multimodal pathways, in particular, emphasize the fusion of textual prompts with visual inputs, reflecting the high reliance on image and video information in construction environments (Tsai *et al.*, 2025).

At the strategy level, five dominant pathways can be identified: prompt engineering, RAG, finetune, knowledge graph integration, and multimodal fusion. Evidence indicates that LLM implementation in construction rarely relies on a single pathway; instead, hybrid combinations are increasingly common. For example, GPT-based models are often paired with RAG for compliance tasks (Joffe *et al.*, 2025), construction coordination frequently integrates agent-based (Zhang and Ge, 2025) and multimodal (Cai *et al.*, 2025) approaches, and process planning has leveraged LLM-enabled knowledge graph construction for knowledge extraction (Xu *et al.*, 2025a).

At the application level, distinct preferences are observed across tasks. In compliance and governance, GPT-based models combined with RAG and textual data are applied to regulatory interpretation and contract review (Jang and Lee, 2024). Construction coordination relies more on GraphRAG and multimodal fusion, integrating text with images and videos for defect analysis (Jeon and Lee, 2025). In design and planning, prompt engineering and BIM data serve as primary pathways for aligning natural language queries with complex structures and enabling constraint validation and code-driven retrieval (Guo *et al.*, 2025). O&M management is characterized by the use of tool calling and sensor data to interface with simulators and monitoring systems for optimization (Jiang *et al.*, 2024). In education and knowledge, prompt engineering (Sabir *et al.*, 2025) is frequently adopted, tailored to educational datasets to support Q&A, tutoring, and training.

4. Functional domains of LLMs applications

To provide a structured understanding of the functional roles of LLMs, a comprehensive content analysis was conducted. The analysis revealed five distinct functional domains (Figure 6), each comprising task clusters that highlight the specific contributions of LLMs. The detailed sub-sections below elaborate each domain.

4.1 Compliance and regulation

Recent studies have demonstrated a rapid advancement in applying LLMs to the interpretation, transformation, and compliance checking of building regulations. Traditional rule-based or supervised learning approaches often suffer from scalability issues or high data dependency. In contrast, LLMs offer a more flexible approach by leveraging contextual understanding, few-shot learning, prompt engineering, and RAG. As summarized in Table 1, a wide spectrum of tasks has been explored.

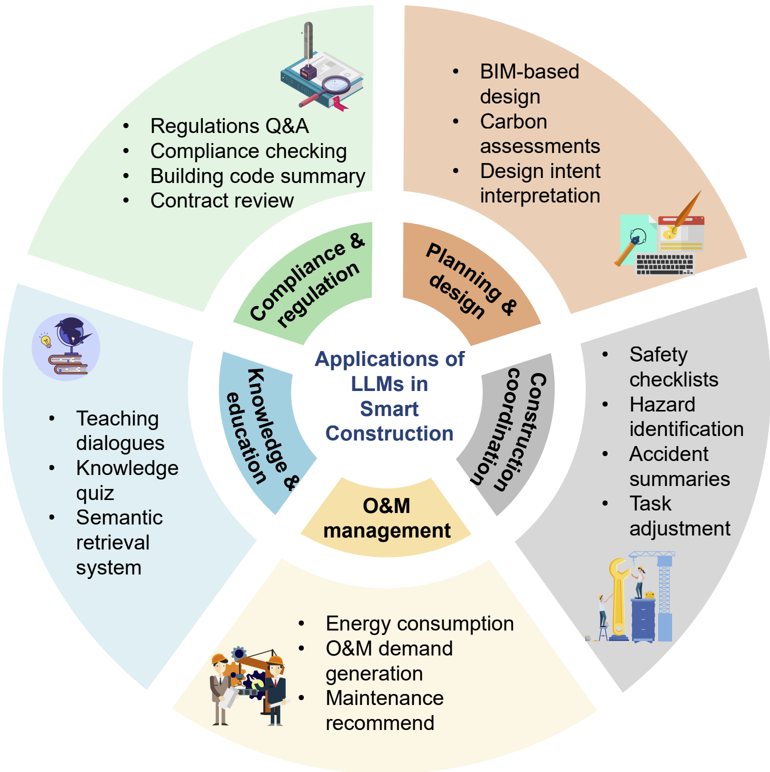

Figure 6. Categorization of LLMs in smart construction. Source(s): Authors' own work

Table 1. Summary of applications of LLMs in compliance and regulation

Applications	Data source	Data processing	Model	Advantages	Limitations	Reference
Summarize construction dispute data	Westlaw database (300 U.S. construction dispute cases)	Sentence-level segmentation, BERT-based summarization	OpenAI's models and BERT	Summary results are intuitive and geared towards legal practice scenarios	Inconsistent contextual coherence and abstract precision	Seo and Kang (2025)
Construction contract risk review	Contract	Graph-enhanced retrieval with nested contract knowledge graphs	LLM + Nested Contract Knowledge Graph (NCKG)	High semantic precision, hallucination mitigation	High cost of knowledge graph construction, lacks dynamic updates	Zheng *et al.* (2025)
Compliance checking	International Building Code 2015, 2022 Michigan state traffic crash data	Embedding	Prompt-engineered interaction framework with ChatGPT	For regulatory processing scenarios where data is scarce	Requires expert intervention for Prompt engineering	Yang and Zhang (2024)
Contract clause interpretation	Standard contract forms	Embedding	Langchain + GPT-3.5 + Chroma	Vector database supports semantic similarity search beyond traditional keyword matching	Does not involve complex reasoning or generation; suitable for closed-ended quizzes only	Saparamadu *et al.* (2025)
Named entity recognition for construction documents	Internal construction documents	Embedding	Fine-tuning LLMs, RAG	Effective on weakly supervised data; adaptable to diverse construction texts	Limited entity types	Zhou and Ma (2025)
Building codes and standards Q&A	National Building Code of Canada 2020	Modality alignment, Text segmentation, Embedding	RAG	Highly traceable: all Q&As come with original canonical citations	Plain text processing only	Joffe *et al.* (2025)
Legal consulting in construction procurement	Taiwan legal texts (in Traditional Chinese	Pilgrimage Walk Optimization-Differential Evolution (PWO-DE) Optimization hyperparameters	Llama, ChatGPT	Cultural adaptation, improved legal accuracy	Still need human judgment	Liu and Chou (2025)

(*continued*)

Table 1. Continued

Applications	Data source	Data processing	Model	Advantages	Limitations	Reference
Construction contract review	International Federation of Consulting Engineers (FIDIC)	Embedding	LLM + Domain Rule Templates	Strong semantic matching skills; rudimentary reasoning skills	Insufficient ability to generalize to new types of contracts	Kim *et al.* (2025)
Building code Q&A	Regulation texts	Embedding	RAG	Highly interpretable; combines symbolic and semantic reasoning	Reliance on high-quality regulatory corpora	He *et al.* (2025a)
BIM compliance check	BIM model, Residential Design Code GB-50096–2011	Term extraction, Semantic clustering	GPT-4	Improved extraction accuracy; reduced need for manual feature engineering	More complex processes, requiring deep integration of multiple modules	Chen *et al.* (2024)

Source(s): Authors' own work

The prevailing approaches reveal a fundamental trade-off between precision and generalizability. For regulation transformation, studies by Fuchs *et al*. (2024) and Yang and Zhang (2024) demonstrated that prompt engineering with large models like GPT can achieve high precision on specific code clauses. However, this performance is highly sensitive to prompt phrasing and the selection of few-shot examples, creating a reproducibility crisis. In response, RAG-based frameworks have emerged to improve transparency by explicitly grounding model outputs in source documents. These systems mitigate hallucination and provide citations, which is a significant step forward for trustworthiness. Saparamadu *et al*. (2025) benchmarked a vector-augmented chatbot against ChatGPT in contract interpretation tasks and showed superior factual accuracy and fewer hallucinations. However, their effectiveness remains closely dependent on the quality, coverage, and granularity of the underlying knowledge base.

Efforts to enhance explainability and domain alignment have increasingly turned to knowledge integration, yet these approaches often substitute one form of complexity for another. While the nested contract knowledge graph improves semantic search (Zheng *et al*., 2025), its value is contingent on the extensive, and often prohibitive, manual effort required for graph construction, creating a significant scalability bottleneck. Similarly, the integration of expert rules enhances coverage but at the cost of adaptability (Kim *et al*., 2025). Hybrid frameworks that combine LLMs, deep learning, and ontologies represent an architecturally ambitious approach, but they risk generating overly complex systems that are difficult to maintain, debug, and deploy in practice (Chen *et al*., 2024). The domain adaptation study successfully reduces hallucination for specific jurisdictions (Liu and Chou, 2025), but revealingly exposes a core weakness: the lack of a foundational, cross-jurisdictional understanding in these models, which are instead painstakingly retrofitted to a single legal context. The transition to real-time, interactive assistants is commendable but remains a proof-of-concept (Tran *et al*., 2024), its validation in the chaotic, multilingual, and high-stakes environment of a construction site, where reliability is paramount, remains the critical unsolved challenge. The heuristic segmentation strategies employed for lengthy contracts (Gao *et al*., 2025c) are a pragmatic workaround, but they inherently fracture the document's context, potentially creating misinterpretations at the seams between segments. Nevertheless, reliance on heuristic segmentation highlights persistent challenges in maintaining document-level context and ensuring reproducibility.

4.2 Planning and design

The application of LLMs in the planning and design phase has progressed rapidly, demonstrating enhanced semantic interpretation, functional reasoning, and task automation. Yet, most existing studies remain proof-of-concept prototypes, with limited evidence of scalability, interoperability, or seamless integration into established design workflows. An overview of representative applications, including their methodological components and technical advantages, is provided in Table 2.

Technical progress and limitations: The Natural-language-based Architectural Detailing through Interaction with AI (NADIA) system (Jang *et al*., 2024), which integrates GPT with Revit to automate exterior wall detailing, achieved 83.3% accuracy and reduced manual effort. However, its reliance on predefined JavaScript Object Notation (JSON) rules and rigid prompt chains constrain adaptability to novel design intents and complex, non-standard forms, limiting generalizability. Similarly, a multimodal framework that fuses façade imagery with textual signage for building use identification (Ramalingam and Kumar, 2025) illustrates novel vision–language integration but performs poorly in data-scarce contexts such as narrow streets or signage-deficient environments, reflecting the dependence of LLM-based design tools on high-quality multimodal data rarely available in early-stage design. A generative design framework for industrialized buildings (Gao *et al*., 2025a) further demonstrates the potential of combining LLMs with knowledge graph question answering and multi-objective optimization to support interdisciplinary integration and performance-driven decision-making.

Table 2. Summary of applications of LLMs in planning and design

Application focus	Data source	Data processing	Model	Advantages	Limitations	Reference
Automated detailing of exterior walls	Architectural design intentions (text) + Revit BIM environment	Prompt chaining, rule-based JSON mapping	GPT-3.5-turbo-0301	High automation, compliant with thermal rules, reduces manual modeling	Dependent on prompt quality and rule definitions, limited to wall detailing	Jang *et al.* (2024)
Building use identification	Google street view images	Building detection, text/image classification, ensemble fusion	Grounding DINO, DINOv2, CLIPSeg, GPT-3.5 Turbo, fine-tuned	Combines visual and textual cues for robust classification	Performance drop in narrow streets or buildings without signage	Ramalingam and Kumar (2025)
Accounting embodied carbon in buildings	Two buildings, CAD drawings, and construction plans, Chinese Life Cycle Database	Regular expressions and text parsing techniques	Fine-tuned Claude-3.5, GPT-4o	High precision: Claude-3.5 match accuracy of 84.12%	Models have regional adaptation limitations	Gu *et al.* (2025)
Generative design in industrialized construction	BIM data, Singapore Road Traffic Rules, interviews, Singapore Building Carbon Calculator	Data cleaning, normalization, embedding	LLM + Knowledge Graph Q&A	Achieves 15.8% cost reduction, 21.2% energy saving	High computational demands, partial realization of full framework in practice	Gao *et al.* (2025a)
Early-stage design decision support in additive construction	Domain ontologies, standards	Ontology formalization, knowledge graph construction, SPARQL querying	LLM + Knowledge Graph Q&A	bridges BIM and additive construction domains	High ontology modeling cost	Li and Petzold (2025)
BIM-based design for free-form prefabricated buildings	BIM model, rules	Extract features, prompt generation	Llama 2, Chat-Yuan, Chat-GPT, GPT-4	Bridges BIM design and manufacturing logic through semantic LLM integration	The initial costs associated with implementing can be substantial	Han *et al.* (2025)

Source(s): Authors' own work

Integration–performance trade-offs: Deeper integration with sustainability and manufacturability objectives has also been explored. For instance, linking BIM schedules with carbon emission databases (Gu *et al.*, 2025) illustrates the potential for automated sustainability accounting. Yet, the regional specificity of such databases severely restricts global applicability, underscoring the absence of standardized and interoperable environmental data schemas. Ontology-driven and BIM-based design-for-manufacture-and-assembly frameworks demonstrate academic rigor and formalize expert knowledge effectively (Li and Petzold, 2025), but high modeling and implementation costs render them feasible mainly for large-scale, repetitive projects, neglecting the long tail of bespoke constructions that dominate practice.

4.3 Construction and coordination management

In the construction phase, LLMs are increasingly integrated into key tasks such as safety management, site-level information coordination, equipment control, knowledge extraction, and human–machine interaction. These applications purport to enhance productivity and decision-making on site; however, a critical evaluation reveals a field in its infancy, often prioritizing technological novelty over practical robustness, and demonstrating a concerning gap between controlled research settings and the chaotic reality of the construction site.

Safety management: This remains the most active domain of adoption. LLMs are increasingly deployed in coordination, safety monitoring, and task scheduling, often through integration with sensor and image data. RAG and fine-tuning based models have demonstrated improved knowledge retrieval and structured risk identification compared with baseline GPT-4 outputs (Lee *et al.*, 2024). Graph-augmented retrieval has enhanced the parsing of fragmented planning constraints (He *et al.*, 2025b). In prefabricated construction, risk management agents have achieved 99.95% accuracy in risk definition (Zhang and Ge, 2025), while vision–language systems improve hazard recognition through Contrastive Language-Image Pre-training (CLIP)-based fusion (Tsai *et al.*, 2025). On a large-scale case basis, semantic alignment of job activities, equipment types, and mitigation advice has been supported by LLM-based retrieval and synthesis (Baek *et al.*, 2025). Multimodal reasoning capabilities, particularly from Vision language models (Chen *et al.*, 2025b), enable hazard detection by fusing on-site video or image streams with textual prompts and planning constraints. Despite these advances, the evaluation of safety systems often relies on synthetic datasets or narrow scenarios, raising concerns about generalizability.

Task coordination and multi-agent collaboration: LLMs have been used to transform meeting transcripts into actionable items, such as in the “Meet2Mitigate” framework (Chen *et al.*, 2025a), and to support multimodal reporting systems that combine UAV-collected imagery with textual inputs for field documentation (Pu *et al.*, 2024). Incident analysis has also been automated: causal chain reconstruction, accident-type classification (Zhang *et al.*, 2025d), equipment failure inference (Ray *et al.*, 2024), and scene graph-based visualization of accidents have all been demonstrated. In field inspections, LLMs orchestrate multi-agent workflows to interpret drone-captured visual data and automate reporting (Liu *et al.*, 2025a). However, many of these systems are limited by hallucination risks, dependency on structured reporting formats, and a lack of mechanisms for uncertainty communication. Most frameworks also assume cooperative, noise-free interactions, which may not hold in real-world construction meetings or when multiple stakeholders have conflicting priorities.

Process control and perception: Applications include zero-shot frameworks using CLIP and transformers for equipment classification without labels (Jeoung *et al.*, 2025), hierarchical code generation for robotic assembly (Luo *et al.*, 2024), and multimodal grouting parameter forecasting (Zhang *et al.*, 2025e). These methods illustrate the versatility of LLMs in integrating numeric, textual, and visual signals. Nonetheless, challenges remain in ensuring robustness under varying site conditions, computational feasibility for real-time operation, and the explainability of decisions in safety-critical tasks.

4.4 Operation and maintenance management

With the increasing digitalization of building operations, LLMs are being explored in the O&M phase due to their semantic understanding, contextual reasoning, and cross-modal integration capabilities. Current studies have concentrated on four main areas: energy modeling, system diagnosis, and information retrieval.

Energy modeling and prediction: LLMs have been used to automate the construction of building energy models, reducing reliance on expert input and improving prediction efficiency. Platforms such as EPlus-LLM integrate GPT models into energy simulation workflows (Jiang *et al.*, 2024), while Prompt-AM employs task decomposition and multi-agent orchestration for semantic coordination across modeling stages (Jiang *et al.*, 2025). UBEM-GPT extends prediction to the urban scale under data-scarce conditions (Choi and Yoon, 2024), and Empower leverages language-based representations of historical load and weather data for zero-shot energy demand inference (Zhou and Wang, 2025). Multilingual models have also been applied to enhance cross-language generalization (Forth and Borrmann, 2024). Despite these advances, most approaches remain experimental: they often rely on small datasets, lack standardized benchmarks for comparison, and have not been validated in large-scale operational settings.

System diagnosis and strategy optimization: LLMs have been combined with interpretable machine learning to express Heating, Ventilation and Air Conditioning (HVAC) causal chains in natural language, improving transparency (Zhang and Chen, 2024). Fine-tuned LLMs can detect anomalies from limited labeled samples (Zhang *et al.*, 2025a), while occupant feedback has been analyzed through NLP to support personalized comfort control (Sadick and Chinazzo, 2025). When integrated with digital twins, they enable contextual fault detection and autonomous control through ontology-based semantic reasoning (Yoon *et al.*, 2025). Knowledge graph–augmented prompting improves the generation of structured maintenance documents, enhancing clarity and accuracy (Shi *et al.*, 2025a). However, many studies rely on narrowly scoped training data, and there is limited evaluation of performance under real-time operational constraints or changing occupancy patterns.

Information retrieval and question answering: LLMs support semantic search of O&M manuals, maintenance records, and BIM-linked component data. Examples include ChatGPT with RAG for expert-level O&M queries (Li *et al.*, 2025b) and multi-agent architectures for real-time BIM data retrieval (Liu *et al.*, 2025c). Dynamic prompt generation has been used to improve context awareness (Zheng and Fischer, 2023), while generative clustering techniques classify unlabeled defect reports for structured maintenance knowledge (Cusumano *et al.*, 2025). Yet these systems face challenges in scalability, retrieval robustness, and ensuring the accuracy of outputs in safety-critical contexts.

4.5 Knowledge and education

At the level of knowledge management and education, LLMs have been applied not only to improve learning efficiency in construction-related education but also integrated into domain-specific Q&A systems, personalized training platforms, and structured knowledge retrieval workflows, helping to bridge the gap between academic training and practical application.

Domain-specific Q&A and knowledge management: Structured knowledge injection and RAG frameworks have been used to improve adaptability in project management and legal domains. Examples include knowledge graph for project management, which improved accuracy across multiple question types (Zhou *et al.*, 2025a), and knowledge integration approaches that boosted exam performance in specialized areas (Zhou *et al.*, 2025b). While these systems demonstrate measurable gains, they remain heavily dependent on curated datasets and lack robustness across diverse project contexts. Hybrid retrieval systems combining vectors, keywords, and graph structures have improved multi-source knowledge access throughout the building lifecycle (Wang *et al.*, 2025b). Spatial queries over BIM models have also been supported by natural language systems integrating geometric indexing and multi-agent coordination, achieving 92.1% accuracy (Li *et al.*, 2025a). Yet these approaches rely on domain-specific infrastructure and controlled settings, raising concerns about scalability and integration into existing workflows.

Education and training: In higher education, LLMs have been integrated with computer algebra systems to support inquiry-based learning, enhancing conceptual understanding and problem-solving in engineering education (Matzakos and Moundridou, 2025). Broader studies highlight potential benefits in curriculum design and vocational training, particularly for engagement and foundational knowledge delivery (Jelodar, 2025). However, most work remains limited to small-scale pilots, with little evidence of long-term improvement in task-specific or strategic skills, suggesting a gap between enhanced engagement and actual skill acquisition. VR-based personalized safety training (Sabir *et al.*, 2025) and game-based systems for risk awareness (Naderi and Shojaei, 2025) illustrate how LLMs can complement immersive technologies. While effective in controlled studies, questions remain about deployment costs, standardization of scenarios, and generalizability to different trades and cultural contexts.

5. Discussions

5.1 Limitations and future directions

Despite their promise, LLM adoption in construction remains constrained by data privacy and security concerns, high deployment costs, and hallucination-driven risks in safety-critical environments. Challenges around governance, liability, and workforce transition further complicate practical implementation. Moreover, limited domain-specific knowledge and the absence of standardized benchmarks hinder reliable evaluation and tool interoperability. Advancing secure architectures, efficient and uncertainty-aware models, clear regulatory frameworks, and construction-tailored datasets will be essential. Figure 7 summarizes these key limitations and future research directions.

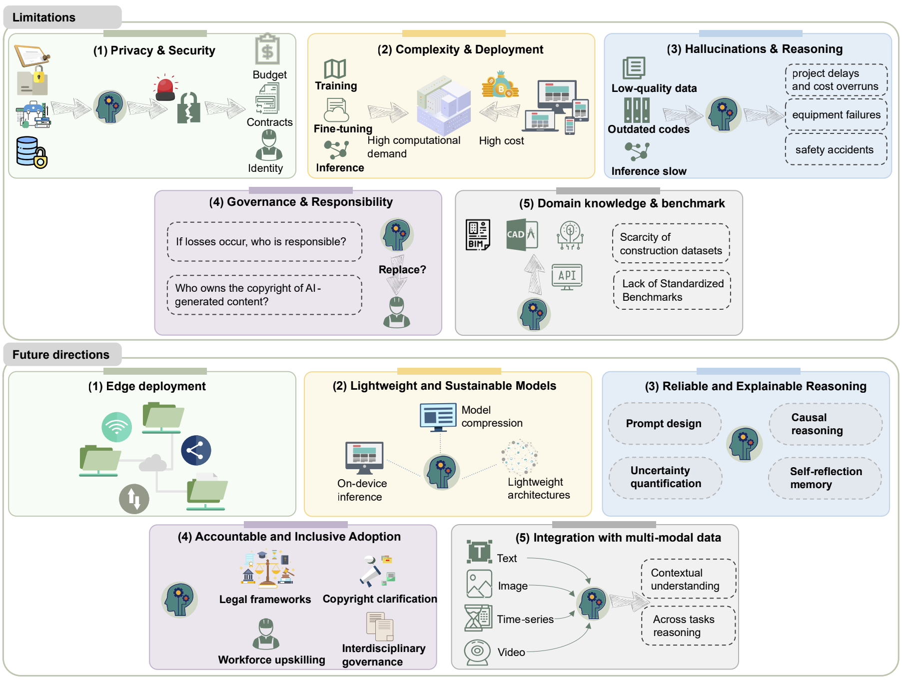

Figure 7. Current research limitations and future directions. Source(s): Authors' own work

5.1.1 Data privacy and security. Construction projects involve sensitive data, including design drawings, BIM models, contracts, and operational records. Reliance on cloud-hosted LLMs raises the risk of data leakage, unauthorized access, or cyber-attacks, which can compromise intellectual property and project confidentiality (Han *et al.*, 2025). Furthermore, LLMs trained on large-scale corpora may inadvertently expose fragments of sensitive information in their outputs, even when such content is not part of the input query (Kaddour *et al.*, 2023). Researchers have also identified multiple vulnerabilities in LLMs, including backdoor attacks, prompt injection, data poisoning, and jailbreak techniques, which can be exploited to extract sensitive data or manipulate outputs (Das *et al.*, 2025). These risks highlight the insufficiency of traditional anonymization and access control mechanisms in high-risk domains such as construction.

Future research should therefore prioritize the design of privacy-preserving mechanisms. Key directions include secure edge deployment, which performs inference locally to minimize transmission risks; federated learning, which enables collaborative training across multiple organizations without exposing raw data; and confidential computing, which uses hardware-based encryption to protect data during processing. Integrating these strategies into LLM workflows will enhance resilience against adversarial attacks, safeguard the confidentiality of construction data, and ensure that LLM-powered systems can be deployed responsibly in practice.

5.1.2 Complexity and deployment constraints. LLMs are inherently resource-intensive, requiring significant computational capacity for training, fine-tuning, and inference. In the construction domain, this translates into barriers for tasks such as large-scale BIM model parsing, real-time compliance checking, and continuous site monitoring, all of which demand substantial computing power. The resulting high energy consumption and carbon emissions are especially paradoxical given the sector's commitment to energy efficiency and decarbonization (Zhang and Chen, 2025). Financial costs add further constraints: API-based access can become prohibitively expensive when querying extensive project records or running frequent regulatory checks, while local deployment requires costly graphical processing unit clusters that must be regularly upgraded to support advancing model complexity (Zhang *et al.*, 2025b).

These limitations reduce the feasibility of LLM integration in site-level or resource-constrained project environments. Future research should therefore prioritize lightweight architectures, model compression, on-device inference, and edge deployment, alongside the adoption of smaller but capable models to ensure that LLM applications in construction remain both practical and sustainable.

5.1.3 Hallucination and reasoning instability. Hallucinations arise when LLMs generate outputs that appear coherent and credible but are factually incorrect or logically invalid (Ji *et al.*, 2023). In the construction domain, such errors can be particularly damaging in safety-critical contexts, such as compliance checking, schedule optimization, or equipment maintenance, where inaccurate recommendations may result in costly delays or safety incidents. Reasoning instability further exacerbates this challenge, as models may provide inconsistent or contradictory outputs under similar conditions. In addition, high computational demands can slow inference, undermining real-time monitoring and site-level decision making and increasing the likelihood of compounding errors.

Mitigation strategies are emerging. Carefully designed prompts that specify contextual information, including key project parameters such as energy peaks or schedule constraints, help reduce bias and improve relevance (Cao *et al.*, 2023). For example, in energy demand forecasting or schedule simulation, prompts should explicitly include task context and key data characteristics (such as peak load, average energy consumption, and schedule ranges) to reduce generation bias. It not only helps mitigate hallucination risks but also strengthens the alignment of LLMs with real-world construction and operation contexts.

Future research should emphasize uncertainty estimation, causal reasoning, and self-refinement mechanisms so that models can indicate confidence levels, identify potential errors, and revise outputs before use. Memory-enhanced architectures that learn from prior mistakes (Shinn *et al.*, 2023) offer additional promise for improving stability and supporting more transparent and verifiable decision-making across construction workflows.

5.1.4 Governance, liability and workforce transformation. LLM adoption in construction raises governance issues beyond technical feasibility, particularly around legal accountability, intellectual property, and workforce transformation. A key gap is the lack of clear liability frameworks: when LLM-generated outputs lead to design errors, cost overruns, or safety incidents, it is unclear whether responsibility lies with model developers, deploying organizations, or end-users. This ambiguity constrains practitioner trust and regulatory acceptance.

Closely linked to liability is the issue of copyright and intellectual property. Construction workflows rely on proprietary standards, codes, and technical documents, some of which are copyrighted (Zhang and Chen, 2025), and their use in training may risk unauthorized reproduction or rephrasing. Meanwhile, ownership of AI-generated content remains legally unresolved. LLMs also accelerate workforce transformation, raising concerns over job displacement and shifting skill demands (Mäkelä and Stephany, 2024).

Addressing these challenges requires an integrated governance agenda. Legal frameworks must establish accountability for AI-assisted decisions; copyright and licensing requirements must be clarified; and workforce adaptation strategies must be embedded into industry policies and organizational practices. Such efforts call for interdisciplinary collaboration among engineers, legal scholars, policymakers, and professional bodies to ensure that LLM adoption in construction is technically feasible, legally sound, ethically responsible, and socially sustainable.

5.1.5 Domain knowledge and benchmark. A major barrier to applying LLMs in construction is the lack of domain-specific expertise and the scarcity of high-quality, construction-oriented datasets (Zhou *et al.*, 2025b). General corpora fail to capture specialized terminology, regulations, safety standards, leading to biased or incomplete outputs. Curated datasets combining BIM documentation, project logs, safety records, and compliance guidelines are needed to support accurate reasoning in tasks such as code checking, energy modeling, and risk prediction. Interoperability is another challenge, as construction workflows depend on diverse systems such as computer-aided design, BIM, and project management platforms. Current models struggle to handle multi-modal and domain-specific file formats, limiting real-world applicability. Future studies should develop middleware solutions and standardized protocols that enable LLMs to communicate with existing tools via APIs or knowledge graphs.

The lack of unified evaluation standards further constrains progress (Banerjee *et al.*, 2024). Unlike NLP domains with the General Language Understanding Evaluation (GLUE) benchmark or the Massive Multitask Language Understanding (MMLU) benchmark, construction lacks shared datasets and metrics. Establishing open benchmarks for tasks such as compliance checking, BIM parsing, and safety management would support fair comparison, reproducibility, and adoption. Advancing domain datasets, interoperability frameworks, and benchmarking systems will enhance reliability and scalability, fostering the responsible integration of LLMs into digital construction ecosystems.

5.2 Implications and recommendations

The findings indicate that the strongest performance of current LLM applications lies in augmenting, rather than replacing, human decision-making. In the near term, LLMs are best positioned as decision-support tools rather than fully autonomous systems, particularly in safety and cost critical workflows; accordingly, human-in-the-loop verification, traceable retrieval, and audit logging should be treated as baseline requirements. From a governance perspective, organizations should establish clear protocols for data privacy, access control, and intellectual property before scaling deployment. The lack of standardized benchmarks remains a key bottleneck, underscoring the need for open, construction specific datasets, benchmarks, and evaluation protocols to enable reproducible research. At the societal and policy level, responsible adoption will require workforce reskilling and sector-level guidance to support transparent, auditable, and accountable use.

6. Conclusions

This paper presented a comprehensive review of the application of LLMs in the domain of smart construction, aiming to capture the evolving research landscape and emerging capabilities of these models. By systematically analyzing 142 peer-reviewed papers through bibliometric and content analysis, this paper highlights the expanding role of LLMs in supporting tasks. The findings revealed that LLMs, particularly those based on GPT and BERT architectures, are increasingly integrated into real-world construction workflows through prompt engineering, fine-tuning, and RAG strategies. These techniques have enabled the transformation of unstructured construction data into actionable knowledge, improved accessibility to domain expertise, and supported more adaptive and intelligent decision-making environments. This review can serve as a foundational reference for researchers and practitioners seeking to understand the potential, methods, and limitations of LLMs in smart construction. It not only summarizes current achievements but also outlines future directions for developing trustworthy, scalable, and domain-aligned language intelligence in smart construction.

Despite its contributions, this review has several limitations. The literature was mainly sourced from WoS and Scopus, potentially excluding relevant studies from other databases. In addition, detailed technical evaluations of model performance were beyond the review's scope. Future research should expand data sources and integrate systematic review with empirical benchmarking for a more comprehensive assessment.

References

Abdelalim, A.M., Salem, M., Salem, M., Al-Adwani, M. and Tantawy, M. (2025), "An analysis of factors contributing to cost overruns in the global construction industry", *Buildings*, Vol. 15 No. 1, p. 18, doi: 10.3390/buildings15010018.

Abioye, S.O., Oyedele, L.O., Akanbi, L., Ajayi, A., Davila Delgado, J.M., Bilal, M., Akinade, O.O. and Ahmed, A. (2021), "Artificial intelligence in the construction industry: a review of present status, opportunities and future challenges", *Journal of Building Engineering*, Vol. 44, 103299, doi: 10.1016/j.jobe.2021.103299.

Akinosho, T.D., Oyedele, L.O., Bilal, M., Ajayi, A.O., Delgado, M.D., Akinade, O.O. and Ahmed, A.A. (2020), "Deep learning in the construction industry: a review of present status and future innovations", *Journal of Building Engineering*, Vol. 32, 101827, doi: 10.1016/j.jobe.2020.101827.

Baek, S., Park, C.Y. and Jung, W. (2025), "Automated safety risk management guidance enhanced by retrieval-augmented large language model", *Automation in Construction*, Vol. 176, 106255, doi: 10.1016/j.autcon.2025.106255.

Banerjee, S., Agarwal, A. and Singh, E. (2024), "The vulnerability of language model benchmarks: do they accurately reflect true LLM performance?", arXiv preprint arXiv:.03597, doi: 10.48550/arXiv.2412.03597.

Bao, S. and Bu, H. (2025), "Defining and generating operation and maintenance management requirements in digital twin applications using the DT-GPT framework", *Journal of Building Engineering*, Vol. 104, 112356, doi: 10.1016/j.jobe.2025.112356.

Cai, R., Guo, Z., Chen, X., Li, J., Tan, Y. and Tang, J. (2025), "Automatic identification of integrated construction elements using open-set object detection based on image and text modality fusion", *Advanced Engineering Informatics*, Vol. 64, 103075, doi: 10.1016/j.aei.2024.103075.

Cao, D., Jia, F., Arik, S.O., Pfister, T., Zheng, Y., Ye, W. and Liu, Y. (2023), "Tempo: prompt-based generative pre-trained transformer for time series forecasting", arXiv preprint arXiv:.04948.

Chen, N., Lin, X., Jiang, H. and An, Y. (2024), "Automated building information modeling compliance check through a large language model combined with deep learning and ontology", *Buildings*, Vol. 14 No. 7, 1983, doi: 10.3390/buildings14071983.

Chen, G., Alsharef, A., Ovid, A., Albert, A. and Jaselskis, E. (2025a), "Meet2Mitigate: an LLM-powered framework for real-time issue identification and mitigation from construction meeting discourse", *Advanced Engineering Informatics*, Vol. 64, 103068, doi: 10.1016/j.aei.2024.103068.

Chen, Z., Chen, H., Imani, M., Chen, R. and Imani, F. (2025b), "Vision language model for interpretable and fine-grained detection of safety compliance in diverse workplaces", *Expert Systems with Applications*, Vol. 265, 125769, doi: 10.1016/j.eswa.2024.125769.

Choi, S. and Yoon, S. (2024), "GPT-based data-driven urban building energy modeling (GPT-UBEM): concept, methodology, and case studies", *Energy and Buildings*, Vol. 325, 115042, doi: 10.1016/j.enbuild.2024.115042.

Cusumano, L., Olsson, N., Granath, M., Jockwer, R. and Rempling, R. (2025), "Clustering techniques and keyword extraction with large language models for knowledge discovery in building defects data", *Construction Innovation*, Vol. 25 No. 7, pp. 76-97, doi: 10.1108/ci-04-2024-0123.

Das, B.C., Amini, M.H. and Wu, Y. (2025), "Security and privacy challenges of large language models: a survey", *ACM Computing Surveys*, Vol. 57 No. 6, pp. 1-39, doi: 10.1145/3712001.

Forth, K. and Borrmann, A. (2024), "Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM", *Journal of Building Engineering*, Vol. 95, 110312, doi: 10.1016/j.jobe.2024.110312.

Fuchs, S., Witbrock, M., Dimyadi, J. and Amor, R. (2024), "Using large language models for the interpretation of building regulations", arXiv preprint arXiv:.21060.

Gambatese, J.A. and Hallowell, M. (2011), "Enabling and measuring innovation in the construction industry", *Construction Management and Economics*, Vol. 29 No. 6, pp. 553-567, doi: 10.1080/01446193.2011.570357.

Gao, M.Y., Li, C., Petzold, F., Tiong, R.L.K. and Yang, Y. (2025a), "Lifecycle framework for AI-driven parametric generative design in industrialized construction", *Automation in Construction*, Vol. 174, 106146, doi: 10.1016/j.autcon.2025.106146.

Gao, Y., Xu, X., Yiu, T.W. and Wang, J. (2025b), "Transfer learning for smart construction: advances and future directions", *Automation in Construction*, Vol. 175, 106238, doi: 10.1016/j.autcon.2025.106238.

Gao, Y., Yihong, G., Yuting, C. and Chen, Y. (2025c), "Application of large language models to intelligently analyze long construction contract texts", *Construction Management and Economics*, Vol. 43 No. 3, pp. 226-242, doi: 10.1080/01446193.2024.2415676.

Gu, X., Chen, C., Fang, Y., Mahabir, R. and Fan, L. (2025), "CECA: an intelligent large-language-model-enabled method for accounting embodied carbon in buildings", *Building and Environment*, Vol. 272, 112694, doi: 10.1016/j.buildenv.2025.112694.

Guo, P., Xue, H., Ma, J. and Cheng, J.C.P. (2025), "Advancing BIM information retrieval with an LLM-based query-domain-specific language and library code function alignment system", *Automation in Construction*, Vol. 178, 106374, doi: 10.1016/j.autcon.2025.106374.

Han, D., Zhao, W., Yin, H., Qu, M., Zhu, J., Ma, F., Ying, Y. and Pan, A. (2025), "Large language models driven BIM-based DfMA method for free-form prefabricated buildings: framework and a usefulness case study", *Journal of Asian Architecture and Building Engineering*, Vol. 24 No. 3, pp. 1500-1517, doi: 10.1080/13467581.2024.2329351.

He, C., He, W., Liu, M., Leng, S. and Wei, S. (2025a), "Enriched construction regulation inquiry responses: a hybrid search approach for large language models", *Journal of Management in Engineering*, Vol. 41 No. 3, 04025001, doi: 10.1061/jmenea.meeng-6444.

He, C., Liu, M., He, W., Wang, Z. and Hsiang, S. (2025b), "Addressing the information silo of construction planning constraints – a graph retrieval augmented generation approach", *Engineering, Construction and Architectural Management*, Vol. ahead-of-print, doi: 10.1108/ecam-08-2024-1066.

Jang, S. and Lee, G. (2024), "Interactive design by integrating a large pre-trained language model and building information modeling", *Computing in Civil Engineering*, Vol. 2023, pp. 291-299, doi: 10.1061/9780784485231.035.

Jang, S., Lee, G., Oh, J., Lee, J. and Koo, B. (2024), "Automated detailing of exterior walls using NADIA: natural-language-based architectural detailing through interaction with AI", *Advanced Engineering Informatics*, Vol. 61, 102532, doi: 10.1016/j.aei.2024.102532.

Jelodar, M.B. (2025), "Generative AI, large language models, and ChatGPT in construction education, training, and practice", *Buildings*, Vol. 15 No. 6, p. 933, doi: 10.3390/buildings15060933.

Jeon, K. and Lee, G. (2025), "Hybrid large language model approach for prompt and sensitive defect management: a comparative analysis of hybrid, non-hybrid, and GraphRAG approaches", *Advanced Engineering Informatics*, Vol. 64, 103076, doi: 10.1016/j.aei.2024.103076.

Jeoung, J., Jung, S. and Hong, T. (2025), "Zero-shot framework for construction equipment task monitoring", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 40, pp. 3218-3238, doi: 10.1111/mice.13506.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A. and Fung, P. (2023), "Survey of hallucination in natural language generation", *ACM Computing Surveys*, Vol. 55 No. 12, pp. 248-38, doi: 10.1145/3571730.

Jiang, G. and Chen, J. (2025), "Efficient fine-tuning of large language models for automated building energy modeling in complex cases", *Automation in Construction*, Vol. 175, 106223, doi: 10.1016/j.autcon.2025.106223.

Jiang, G., Ma, Z., Zhang, L. and Chen, J. (2024), "EPlus-LLM: a large language model-based computing platform for automated building energy modeling", *Applied Energy*, Vol. 367, 123431, doi: 10.1016/j.apenergy.2024.123431.

Jiang, G., Ma, Z., Zhang, L. and Chen, J. (2025), "Prompt engineering to inform large language model in automated building energy modeling", *Energy*, Vol. 316, 134548, doi: 10.1016/j.energy.2025.134548.

Joffe, I., Felobes, G., Elgouhari, Y., Talebi Kalaleh, M., Mei, Q. and Chui Ying, H. (2025), "The framework and implementation of using large language models to answer questions about building codes and standards", *Journal of Computing in Civil Engineering*, Vol. 39 No. 4, 05025004, doi: 10.1061/jccee5.cpeng-6037.

Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R. and Mchardy, R. (2023), "Challenges and applications of large language models", arXiv preprint arXiv:.10169, doi: 10.48550/arXiv.2307.10169.

Kim, E.W., Shin, Y.J., Kim, K.J. and Kwon, S. (2025), "Development of an automated construction contract review framework using large language model and domain knowledge", *Buildings*, Vol. 15 No. 6, p. 923, doi: 10.3390/buildings15060923.

Lee, J., Ahn, S., Kim, D. and Kim, D. (2024), "Performance comparison of retrieval-augmented generation and fine-tuned large language models for construction safety management knowledge retrieval", *Automation in Construction*, Vol. 168, 105846, doi: 10.1016/j.autcon.2024.105846.

Li, C. and Petzold, F. (2025), "Ontology-driven mixture-of-domain documentation: a backbone approach enabling question answering for additive construction", *Buildings*, Vol. 15 No. 1, p. 133, doi: 10.3390/buildings15010133.

Li, A., Wong, P.K.-Y., Tao, X., Ma, J. and Cheng, J.C.P. (2025a), "An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent", *Advanced Engineering Informatics*, Vol. 66, 103375, doi: 10.1016/j.aei.2025.103375.

Li, Y., Ji, M., Chen, J., Wei, X., Gu, X. and Tang, J. (2025b), "A large language model-based building operation and maintenance information query", *Energy and Buildings*, Vol. 334, 115515, doi: 10.1016/j.enbuild.2025.115515.

Liu, C.-Y. and Chou, J.-S. (2025), "Automated legal consulting in construction procurement using metaheuristically optimized large language models", *Automation in Construction*, Vol. 170, 105891, doi: 10.1016/j.autcon.2024.105891.

Liu, J., Li, H., Chai, C., Chen, K. and Wang, D. (2025a), "A LLM-informed multi-agent AI system for drone-based visual inspection for infrastructure", *Advanced Engineering Informatics*, Vol. 68, 103643, doi: 10.1016/j.aei.2025.103643.

Liu, M., Zhang, L., Chen, J., Chen, W.-A., Yang, Z., Lo, L.J., Wen, J. and O'Neill, Z. (2025b), "Large language models for building energy applications: opportunities and challenges", *Building Simulation*, Vol. 18 No. 2, pp. 225-234, doi: 10.1007/s12273-025-1235-9.

Liu, D., Zhou, X. and Li, Y. (2025c), "Balancing performance and cost of LLMs in a multi-agent framework for BIM data retrieval", *Architectural Engineering and Design Management*, pp. 1-18, doi: 10.1080/17452007.2025.2456768.

Luo, H., Wu, J., Liu, J. and Antwi-Afari, M.F. (2024), "Large language model-based code generation for the control of construction assembly robots: a hierarchical generation approach", *Developments in the Built Environment*, Vol. 19, 100488, doi: 10.1016/j.dibe.2024.100488.

Luo, X., Qin, M., Gao, Z., Yan, H. and Yang, X. (2025), "Ground abstract structure concepts of scaffolding systems for automatic compliance checking based on reasoning segmentation", *Expert Systems with Applications*, Vol. 270, 126563, doi: 10.1016/j.eswa.2025.126563.

Mäkelä, E. and Stephany, F. (2024), "Complement or substitute? How AI increases the demand for human skills", arXiv preprint arXiv:.19754.

Matzakos, N. and Moundridou, M. (2025), "Exploring large Language models integration in higher education: a case study in a mathematics laboratory for civil engineering students", *Computer Applications in Engineering Education*, Vol. 33 No. 3, e70049, doi: 10.1002/cae.70049.

Naderi, H. and Shojaei, A. (2025), "Large-language model (LLM)-powered system for situated and game-based construction safety training", *Expert Systems with Applications*, Vol. 283, 127887, doi: 10.1016/j.eswa.2025.127887.

Page, M.J., Mckenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., Mcdonald, S., Mcguinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. and Moher, D. (2021), "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews", *BMJ*, Vol. 372, n71, doi: 10.1136/bmj.n71.

Peng, C., Yang, X., Chen, A., Smith, K.E., Pournejatian, N., Costa, A.B., Martin, C., Flores, M.G., Zhang, Y., Magoc, T., Lipori, G., Mitchell, D.A., Ospina, N.S., Ahmed, M.M., Hogan, W.R., Shenkman, E.A., Guo, Y., Bian, J. and Wu, Y. (2023), "A study of generative large language model for medical research and healthcare", *Npj Digital Medicine*, Vol. 6 No. 1, p. 210, doi: 10.1038/s41746-023-00958-w.

Pu, H., Yang, X., Li, J. and Guo, R. (2024), "AutoRepo: a general framework for multimodal LLM-based automated construction reporting", *Expert Systems with Applications*, Vol. 255, 124601, doi: 10.1016/j.eswa.2024.124601.

Raiaan, M.A.K., Mukta, M.S.H., Fatema, K., Fahad, N.M., Sakib, S., Mim, M.M.J., Ahmad, J., Ali, M.E. and Azam, S. (2024), "A review on large language models: architectures, applications, taxonomies, open issues and challenges", *IEEE Access*, Vol. 12, pp. 26839-26874, doi: 10.1109/access.2024.3365742.

Ramalingam, S.P. and 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*, Vol. 267, 112174, doi: 10.1016/j.buildenv.2024.112174.

Ray, U., Arteaga, C., Ahn, Y. and Park, J. (2024), "Enhanced identification of equipment failures from descriptive accident reports using language generative model", *Engineering, Construction and Architectural Management*, Vol. ahead-of-print, doi: 10.1108/ecam-09-2024-1259.

Sabir, A., Hussain, R., Pedro, A. and Park, C. (2025), "Personalized construction safety training system using conversational AI in virtual reality", *Automation in Construction*, Vol. 175, 106207, doi: 10.1016/j.autcon.2025.106207.

Sadick, A.-M. and Chinazzo, G. (2025), "What did the occupant say? Fine-tuning and evaluating a large language model for efficient analysis of multi-domain indoor environmental quality feedback", *Building and Environment*, Vol. 274, 112735, doi: 10.1016/j.buildenv.2025.112735.

Saka, A., Taiwo, R., Saka, N., Salami, B.A., Ajayi, S., Akande, K. and Kazemi, H. (2024), "GPT models in construction industry: opportunities, limitations, and a use case validation", *Developments in the Built Environment*, Vol. 17, 100300, doi: 10.1016/j.dibe.2023.100300.

Saparamadu, P.V.I.N., Sepasgozar, S., Guruge, R.N.D., Jayasena, H.S., Darejeh, A., Ebrahimzadeh, S.M. and Eranga, B.A.I. (2025), "Optimising contract interpretations with large language models: a comparative evaluation of a vector database-powered chatbot vs. ChatGPT", *Buildings*, Vol. 15 No. 7, 1144, doi: 10.3390/buildings15071144.

Seo, W. and Kang, Y. (2025), "Auto-summarization of the texts of construction dispute precedents", *Advanced Engineering Informatics*, Vol. 65, 103381, doi: 10.1016/j.aei.2025.103381.

Shi, F., Chen, L., Zhou, M., Zhao, Y. and Zheng, Y. (2025a), "A stepwise intelligence generative method for structured maintenance guidance documents based on knowledge graph augmented LLM", *Advanced Engineering Informatics*, Vol. 67, 103523, doi: 10.1016/j.aei.2025.103523.

Shi, J.W.L., Solihin, W. and Yeoh, J.K.W. (2025b), "Fine-tuning a large language model for automated code compliance of building regulations", *Advanced Engineering Informatics*, Vol. 68, 103676, doi: 10.1016/j.aei.2025.103676.

Shinn, N., Cassano, F., Gopinath, A., Narasimhan, K. and Yao, S. (2023), "Reflexion: language agents with verbal reinforcement learning", *Advances in Neural Information Processing Systems*, Vol. 36, pp. 8634-8652.

Taiwo, R., Bello, I.T., Abdulai, S.F., Yussif, A.-M., Salami, B.A., Saka, A. and Zayed, T. (2024), "Generative AI in the construction industry: a state-of-the-art analysis", arXiv preprint arXiv:.09939.

Tran, S.V.-T., Yang, J., Hussain, R., Khan, N., Kimito, E.C., Pedro, A., Sotani, M., Lee, U.-K. and Park, C. (2024), "Leveraging large language models for enhanced construction safety regulation extraction", *Journal of Information Technology in Construction*, Vol. 29, pp. 1026-1038, doi: 10.36680/j.itcon.2024.045.

Tsai, W.-L., Le, P.-L., Ho, W.-F., Chi, N.-W., Lin, J.J., Tang, S. and Hsieh, S.-H. (2025), "Construction safety inspection with contrastive language-image pre-training (CLIP) image captioning and attention", *Automation in Construction*, Vol. 169, 105863, doi: 10.1016/j.autcon.2024.105863.

Uhm, M., Kim, J. and Lee, G. (2025), "Automated analysis of construction safety accident videos using a large multimodal model and graph retrieval-augmented generation", *Automation in Construction*, Vol. 177, 106363, doi: 10.1016/j.autcon.2025.106363.

Wang, Y., Luo, H. and Fang, W. (2025a), "An integrated approach for automatic safety inspection in construction: domain knowledge with multimodal large language model", *Advanced Engineering Informatics*, Vol. 65, 103246, doi: 10.1016/j.aei.2025.103246.

Wang, Z., Liu, Z., Lu, W. and Jia, L. (2025b), "Improving knowledge management in building engineering with hybrid retrieval-augmented generation framework", *Journal of Building Engineering*, Vol. 103, 112189, doi: 10.1016/j.jobe.2025.112189.

Wu, C., Li, X., Guo, Y., Wang, J., Ren, Z., Wang, M. and Yang, Z. (2022), "Natural language processing for smart construction: current status and future directions", *Automation in Construction*, Vol. 134, 104059, doi: 10.1016/j.autcon.2021.104059.

Xiao, T. and Xu, P. (2024), "Exploring automated energy optimization with unstructured building data: a multi-agent based framework leveraging large language models", *Energy and Buildings*, Vol. 322, 114691, doi: 10.1016/j.enbuild.2024.114691.

Xiao, H., Zhuang, J., Yang, B., Han, J., Yu, Y. and Lai, S. (2025), "Generative knowledge-guided review system for construction disclosure documents", *Advanced Engineering Informatics*, Vol. 68, 103618, doi: 10.1016/j.aei.2025.103618.

Xu, Q., Qiu, F., Zhou, G., Zhang, C., Ding, K., Chang, F., Lu, F., Yu, Y., Ma, D. and Liu, J. (2025a), "A large language model-enabled machining process knowledge graph construction method for intelligent process planning", *Advanced Engineering Informatics*, Vol. 65, 103244, doi: 10.1016/j.aei.2025.103244.

Xu, Y., Zhu, S., Cai, J., Chen, J. and Li, S. (2025b), "A large language model-based platform for real-time building monitoring and occupant interaction", *Journal of Building Engineering*, Vol. 100, 111488, doi: 10.1016/j.jobe.2024.111488.

Yang, F. and Zhang, J. (2024), "Prompt-based automation of building code information transformation for compliance checking", *Automation in Construction*, Vol. 168, 105817, doi: 10.1016/j.autcon.2024.105817.

Yoon, S., Song, J. and Li, J. (2025), “Ontology-enabled AI agent-driven intelligent digital twins for building operations and maintenance”, *Journal of Building Engineering*, Vol. 108, 112802, doi: 10.1016/j.jobe.2025.112802.

Zhang, L. and Chen, Z. (2024), “Large language model-based interpretable machine learning control in building energy systems”, *Energy and Buildings*, Vol. 313, 114278, doi: 10.1016/j.enbuild.2024.114278.

Zhang, L. and Chen, Z. (2025), “Opportunities of applying large language models in building energy sector”, *Renewable and Sustainable Energy Reviews*, Vol. 214, 115558, doi: 10.1016/j.rser.2025.115558.

Zhang, L. and Ge, Q. (2025), “Offsite construction risk management agent – a GPTs based analysis model”, *Engineering, Construction and Architectural Management*, Vol. ahead-of-print, doi: 10.1108/ecam-05-2024-0624.

Zhang, S., Teizer, J., Lee, J.-K., Eastman, C.M. and Venugopal, M. (2013), “Building information modeling (BIM) and safety: automatic safety checking of construction models and schedules”, *Automation in Construction*, Vol. 29, pp. 183-195, doi: 10.1016/j.autcon.2012.05.006.

Zhang, J., Zhang, C., Lu, J. and Zhao, Y. (2025a), “Domain-specific large language models for fault diagnosis of heating, ventilation, and air conditioning systems by labeled-data-supervised fine-tuning”, *Applied Energy*, Vol. 377, 124378, doi: 10.1016/j.apenergy.2024.124378.

Zhang, L., Ford, V., Chen, Z. and Chen, J. (2025b), “Automatic building energy model development and debugging using large language models agentic workflow”, *Energy and Buildings*, Vol. 327, 115116, doi: 10.1016/j.enbuild.2024.115116.

Zhang, L., Fu, X., Li, Y. and Chen, J. (2025c), “Large language model-based agent schema and library for automated building energy analysis and modeling”, *Automation in Construction*, Vol. 176, 106244, doi: 10.1016/j.autcon.2025.106244.

Zhang, L., Hou, Y. and Ren, F. (2025d), “AIR agent—A GPT-based subway construction accident investigation report analysis chatbot”, *Buildings*, Vol. 15 No. 4, p. 527, doi: 10.3390/buildings15040527.

Zhang, T., Yu, H., Wang, X., Wang, J. and Ren, B. (2025e), “Knowledge-based cross-modal fusion for long-term forecasting of grouting construction parameters using large language model”, *Automation in Construction*, Vol. 172, 106036, doi: 10.1016/j.autcon.2025.106036.

Zheng, J. and Fischer, M. (2023), “Dynamic prompt-based virtual assistant framework for BIM information search”, *Automation in Construction*, Vol. 155, 105067, doi: 10.1016/j.autcon.2023.105067.

Zheng, C., Wong, S., Su, X., Tang, Y., Nawaz, A. and Kassem, M. (2025), “Automating construction contract review using knowledge graph-enhanced large language models”, *Automation in Construction*, Vol. 175, 106179, doi: 10.1016/j.autcon.2025.106179.

Zhou, J. and Ma, Z. (2025), “Named entity recognition for construction documents based on fine-tuning of large language models with low-quality datasets”, *Automation in Construction*, Vol. 174, 106151, doi: 10.1016/j.autcon.2025.106151.

Zhou, Y. and Wang, M. (2025), “Empower pre-trained large language models for building-level load forecasting”, *IEEE Transactions on Power Systems*, Vol. 40 No. 5, pp. 1-13, doi: 10.1109/tpwrs.2025.3548891.

Zhou, S., Liu, K., Li, D., Fu, C., Ning, Y., Ji, W., Liu, X., Xiao, B. and Wei, R. (2025a), “Augmenting general-purpose large-language models with domain-specific multimodal knowledge graph for question-answering in construction project management”, *Advanced Engineering Informatics*, Vol. 65, 103142, doi: 10.1016/j.aei.2025.103142.

Zhou, S., Liu, X., Li, D., Gu, T., Liu, K., Yang, Y. and Wong, M.O. (2025b), “Integrating domain-specific knowledge and fine-tuned general-purpose large language models for question-answering in construction engineering management”, *Automation in Construction*, Vol. 175, 106206, doi: 10.1016/j.autcon.2025.106206.

Corresponding author
Yu Gao can be contacted at: yu.gao4@unsw.edu.au