Review | Received 4 August 2026; Revised 5 October 2026; Accepted 8 November 2026; Published XX November 2026
https://doi.org/10.55092/sc20260023

Large foundation models in smart construction: a triple perspective review of data foundations, algorithmic adaptation, and workflow integration

Yi Li[1], Xinyue Guo[1], Yujie Han[2,3], Svetlana Besklubova[4], Ioannis Brilakis[4], Miroslaw J. Skibniewski[5], Zhiheng Zhao[2,3] and Ray Y. Zhong[1,*]

[1] Department of Data and Systems Engineering, The University of Hong Kong, Hong Kong, China

[2] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China

[3] Research Institute for Advanced Manufacturing, The Hong Kong Polytechnic University, Hong Kong, China

[4] Department of Engineering, University of Cambridge, Cambridge, UK

[5] Department of Civil and Environmental Engineering, University of Maryland, College Park, USA

* Correspondence author; e-mail: zhongzry@hku.hk.

Highlights:

- A data-algorithm-application framework synthesizes 136 construction LFM studies.
- 86 datasets are assessed for accessibility, semantic granularity, and benchmark readiness.
- 4 construction processes with 14 sub-tasks are summarized from the construction workflow.
- LFM algorithms evolve from semantic alignment and RAG to multimodal tool use.
- 6 future directions support traceable, executable, and workflow-integrated construction AI.

Abstract: Large Foundation Models (LFMs) have been applied to substantial construction tasks, including document analysis, multimodal understanding, and engineering software interaction. However, most existing studies focus on individual tasks or technical components, with limited attention to how data preparation, model adaptation, and lifecycle applications are integrated in an end-to-end intelligent workflow. To address this gap, this paper conducts a systematic review from three complementary perspectives: data, algorithms, and applications. It reviews how construction data support LFM development, how general-purpose LFMs are adapted to specific construction domains, and how these capabilities are used across lifecycle workflows. This review collects 136 studies published between November 2022 and August 2026 through an iterative search and screening process. A scientometric analysis is conducted to examine publication trends, major research topics, and the distribution of studies across publication venues. From the data perspective, 86 construction datasets reported in the 136 studies

Copyright©2026 by the authors. Published by ELSP. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

are identified and analyzed in data modality, accessibility, semantic granularity, data processing, and benchmark readiness. From the algorithmic perspective, the review examines the LFM-based methods for construction-specific knowledge and tasks. From the application perspective, the review classifies LFM-based construction applications into four main lifecycle processes: pre-construction planning, design and engineering, on-site construction execution and management, and post-construction operation and maintenance. The practical application of LFMs in construction is still constrained by limited access to construction data and insufficient validation in real-world projects. This review proposes future directions toward traceable, multimodal, executable, and workflow-integrated construction intelligence.

Keywords: smart construction; construction management; artificial intelligence; large foundation model; systematic review

1. Introduction

Construction projects consist of multiple interdependent activities and continuously generate large amounts of heterogeneous information across different phases [1]. To increase the level of automation in information processing, Artificial Intelligence (AI) is gradually being applied to construction data analysis and related tasks [2]. However, existing construction AI still relies heavily on task-specific Deep Learning (DL) models, which typically require specialized labeled data and are trained around fixed input-output relationships [3]. When project conditions, data distributions, or task requirements change, the reusability and transferability of these models are limited [4,5]. Meanwhile, they struggle to integrate information from different activities, incorporate domain expertise, and support a broader range of engineering reasoning. Consequently, the development of construction AI is shifting from single-task models toward workflow-level AI systems that can integrate heterogeneous information and support reasoning across the entire construction process.

Large Foundation Models (LFMs) are large-scale pretrained models with general-purpose language, vision-language, or multimodal capabilities [6]. In construction, LFMs have been applied to a wide range of construction tasks, such as document analysis, Building Information Modeling (BIM)-related reasoning, safety inspection, and building performance analysis [7–11]. Unlike task-specific DL models, LFMs obtain reusable representations through large-scale pre-training and are capable of integrating broader context and external domain knowledge [12–14]. As a result, they can support different construction tasks based on the same foundation models [15,16]. However, existing studies primarily come from single-task studies and prototype systems. These successes do not directly indicate that LFMs can be reliably integrated into real-world construction workflows. The fundamental reason lies in the fact that current research on data, algorithms, and deployment remains relatively fragmented, and there is a lack of systematic analysis of the connections among these three areas. A systematic review is therefore needed to examine these three components together and clarify the path from construction data foundations, through algorithmic adaptation, to workflow deployment.

Current reviews on construction Large Language Models (LLMs) have covered a broad range of topics. One category discusses the role of generative AI in Industry 4.0/5.0 and summarizes opportunities and challenges for LLM adoption in construction [13,17,18]. Another focuses on specific application

areas such as construction contract management and site safety [19]. However, most of these reviews are still primarily organized by technology types or application scenarios. They rarely provide further analysis of the interrelationships among the construction data foundation, model adaptation methods, and practical applications. It remains unclear which construction data are suitable for LFM development, how general-purpose models are adapted to construction tasks, and whether the capabilities can be reliably transferred to real-world applications. Therefore, the remaining review gap is the absence of a systematic analysis from a cross-layer perspective in data readiness, algorithmic adaptation, and workflow deployment.

To address this gap, this paper proposes a systematic review of LFM-enabled construction from the three connected perspectives: data foundations, algorithmic adaptation, and workflow application. At the data level, the review examines dataset modality, source, accessibility, semantic granularity, processing, and benchmarks in LFM-enabled studies. At the algorithm level, the review analyzes how general-purpose LFMs are adapted to construction knowledge and tasks through semantic adaptation, retrieval-augmented reasoning, multimodal and spatial understanding, and tool integration. At the application level, existing studies are classified into four construction lifecycle processes: pre-construction planning, design and engineering, on-site construction execution and management, and post-construction operation and maintenance. The review then connects these three levels to examine how data conditions influence model adaptation and how adapted model capabilities are used in construction tasks and workflows. The review further identifies cross-layer challenges and proposes research directions.

The remainder of this paper is organized as follows. Section 2 describes the review methodology. Section 3 presents the scientometric analysis. Section 4, Section 5, and Section 6 provide analyses from the data, algorithm, and application perspectives. Section 7 presents the cross-perspective analysis. Section 8 summarizes the key challenges and proposes future directions. Section 9 concludes this review.

2. Methodology

This section presents the systematic review method, as shown in Figure 1. It consists of the research questions, the literature collection, the inclusion and exclusion criteria, and the literature information extraction.

2.1. Research questions

This review is organized based on one general research question: “How well do current construction data and LFM adaptation methods support the reliable deployment in construction workflows?” Three sub-research questions (RQs) are proposed to examine data readiness, algorithmic adaptation, and workflow deployment in LFM-enabled construction.

- RQ1: Data readiness. What construction datasets are currently available for LFM development and evaluation?
- RQ2: Algorithmic adaptation. How are general-purpose LFMs adapted to construction-specific knowledge and tasks?
- RQ3: Workflow deployment. Where have LFMs been applied across the construction lifecycle?

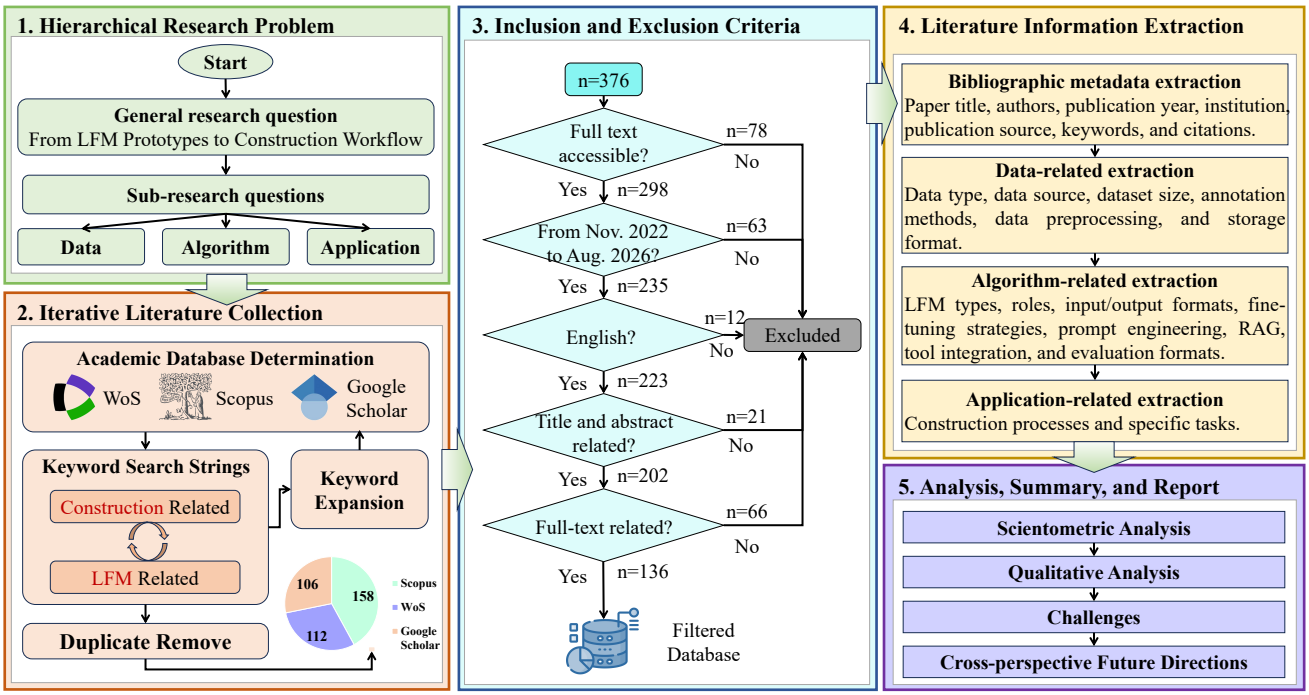

Figure 1. Overview of the review methodology.

2.2. Literature collection

The papers were collected from three databases: Web of Science (WoS), Scopus, and Google Scholar. The search was conducted using two categories of keyword strings, which included both construction-related and LFM-related keywords. The two blocks were combined using the Boolean operator AND, while terms within each block were connected by OR. Because the three databases use different search fields and syntax, the same keyword concepts were retained while the query format was adapted to each database. In WoS, the search was conducted in the Topic field using the structure TS=(construction-related terms) AND TS=(LFM-related terms). In Scopus, the corresponding search was performed in the title, abstract, and keyword fields using TITLE-ABS-KEY(construction-related terms) AND TITLE-ABS-KEY(LFM-related terms). Google Scholar was used as a supplementary search source. It is noted that the keyword list was developed iteratively. After the initial search, the titles, abstracts, and highlights of relevant papers were examined to identify additional terminology used in construction LFM research. Newly identified terms were incorporated into the corresponding keyword blocks, and the search was repeated until no additional relevant search terms were identified. The final keyword set is reported in Table 1. Because recently published articles may not yet be fully indexed in WoS or Scopus, an additional search was conducted for the latest months of the review period. Early Access records in WoS and Articles in Press in Scopus were checked, and Google Scholar results were restricted to recent publications and sorted by date. In addition, the online-first and latest-article pages of the main journals were directly checked for online papers. The publication status and date of these recent records were verified against the publisher page and DOI record before inclusion. The searches produced 621 records in total. 245 duplicate records were removed. 376 literature records remained for subsequent screening: 158 from Scopus, 112 from WoS, and 106 from Google Scholar.

Table 1. Search strings used for literature retrieval.

Topic	Keywords
Construction	"construction" OR "building" OR "infrastructure" OR "civil engineering" OR "built environment" OR "architecture" OR "architectural engineering" OR "structural engineering" OR "construction engineering" OR "construction management" OR "project management" OR "facility management" OR "asset management" OR "AEC" OR "AECO" OR "Architecture, Engineering, and Construction" OR "Architecture, Engineering, Construction, and Operations" OR "BIM" OR "building information modeling" OR "construction industry" OR "construction site" OR "construction process" OR "construction phase" OR "construction project" OR "construction lifecycle"
Large Foundation Models	"foundation model" OR "large foundation model" OR "large-scale foundation model" OR "large language model" OR "LLM" OR "vision-language model" OR "VLM" OR "large vision-language model" OR "multimodal foundation model" OR "multimodal large language model" OR "MLLM" OR "large multimodal model" OR "pretrained model" OR "large pretrained model" OR "pretrained language model" OR "generative pre-trained transformer" OR "generative artificial intelligence" OR "generative AI" OR "GenAI" OR "transformer-based model" OR "ChatGPT" OR "GPT" OR "BERT" OR "LLaMA" OR "PaLM" OR "Claude" OR "Gemini" OR "Qwen" OR "DeepSeek" OR "retrieval-augmented generation" OR "RAG" OR "GraphRAG" OR "prompt engineering" OR "fine-tuning" OR "instruction tuning" OR "knowledge graph" OR "tool use" OR "LLM agent" OR "multi-agent" OR "BIM-LLM"

Note: AEC, architecture, engineering, and construction; AECO, architecture, engineering, construction, and operations; MLLM, multimodal large language model; BERT, bidirectional encoder representations from transformers.

2.3. Inclusion and exclusion criteria

The inclusion and exclusion process consisted of five steps. First, 78 records were removed because their full texts were unavailable. Second, the publication period was restricted to November 2022-August 2026, corresponding to the rapid development and growing research attention surrounding Generative Pre-trained Transformer (GPT) models. 63 papers were excluded. Third, 12 non-English studies were excluded. Fourth, the titles and abstracts of the remaining studies were initially screened for topic relevance. 21 clearly unrelated papers were excluded. Finally, detailed full-text assessment was conducted to examine both topic relevance and methodological adequacy. For research papers, a study was retained only when (1) construction or AEC formed a substantive research context; (2) an eligible foundation model formed a substantive component of the proposed method or application; (3) the research objective, model role, data or input source, and methodological procedure were reported sufficiently for reliable extraction; and (4) empirical, benchmark, case-based, expert-based, functional, or other evaluation evidence was provided. For review papers, the literature search, study-selection, and synthesis procedures were required to be explicitly reported. The same criteria were used throughout the screening, and no study was treated differently. During the full-text review, 13 papers were removed because they were not related to the review topic. Another 53 were excluded because the reported methods or evaluation details were not adequate for the subsequent analysis. This resulted in a final set of 136 publications. The screening and coding processes were conducted by three reviewers. They first independently reviewed the studies and then compared their decisions. Cases with different judgments on eligibility, methodological adequacy, or coding were discussed among the reviewers before a final decision was made.

2.4. Literature information extraction

A predefined information extraction scheme was applied to the 136 included studies. At the studylevel, bibliographic metadata covered the paper title, authors, publication year, institution, keywords, and citations to present the publication characteristics. At the data level, the construction datasets in reviewed papers were collected, including data modality, source, accessibility, semantic granularity, and preprocessing procedure. At the algorithm level, LFM type and functional role, as well as fine-tuning, prompt engineering, Retrieval-Augmented Generation (RAG), and external tools were recorded. Evaluation information was also extracted to examine how the reported LFM outputs were validated. At the application level, each study was classified to the corresponding construction lifecycle stage and specific subtask.

3. Scientometric analysis

3.1. Publication overview and research distribution

Figure 2 shows how the research focus of LFM-related construction studies has changed over time. That reveals a clear transition in research focus from initial model adaptation to task-level exploration, then to workflow expansion and system integration. Since the literature review was completed in August 2026, the data for 2026 cover only a portion of the published literature. These data were included to reflect recent research trends, rather than for direct comparison with previous years. In the first phase (late 2022-early 2023), studies mainly examined whether pretrained models could be used for construction-specific knowledge and tasks [20]. From mid-2023 to early 2024, studies increasingly applied LFMs to specific construction tasks rather than at a conceptual level [21]. Since 2024, the field has entered a workflow expansion phase. More studies have begun to combine LFMs with external knowledge sources and engineering tools to support more structured construction workflows. Some studies further connected several of these components within a single system in late 2025, including multimodal RAG, knowledge-based reasoning, tree-indexed BIM retrieval, and multi-agent LLM systems [22–24]. The four development phases are used to describe how the focus of the field changed over time. They were defined from the timing and research focus of representative studies. These phases describe changes in the research focus of the reviewed literature rather than changes in industry adoption or deployment maturity. Although workflow-oriented and integrated systems receive increasing research attention, most reported applications remain at the prototype, case-study, or controlled-evaluation stage.

Figure 3a summarizes the publication venues of the 136 included studies. *Automation in Construction* published 37 papers (27.2%), followed by *Advanced Engineering Informatics* with 19 papers (14.0%). *Journal of Computing in Civil Engineering*, *Journal of Building Engineering*, and *Applied Energy* contributed 12 (8.8%), 9 (6.6%), and 7 (5.1%) papers, respectively. The remaining 52 papers (38.2%) were distributed across other journals. Figure 3b gives the annual distribution of these publications. In *Automation in Construction*, the number of papers increased from 3 in 2023 to 11 in 2024 and 15 in 2025. *Advanced Engineering Informatics* followed the same general pattern, with 1, 6, and 8 papers in these three years.

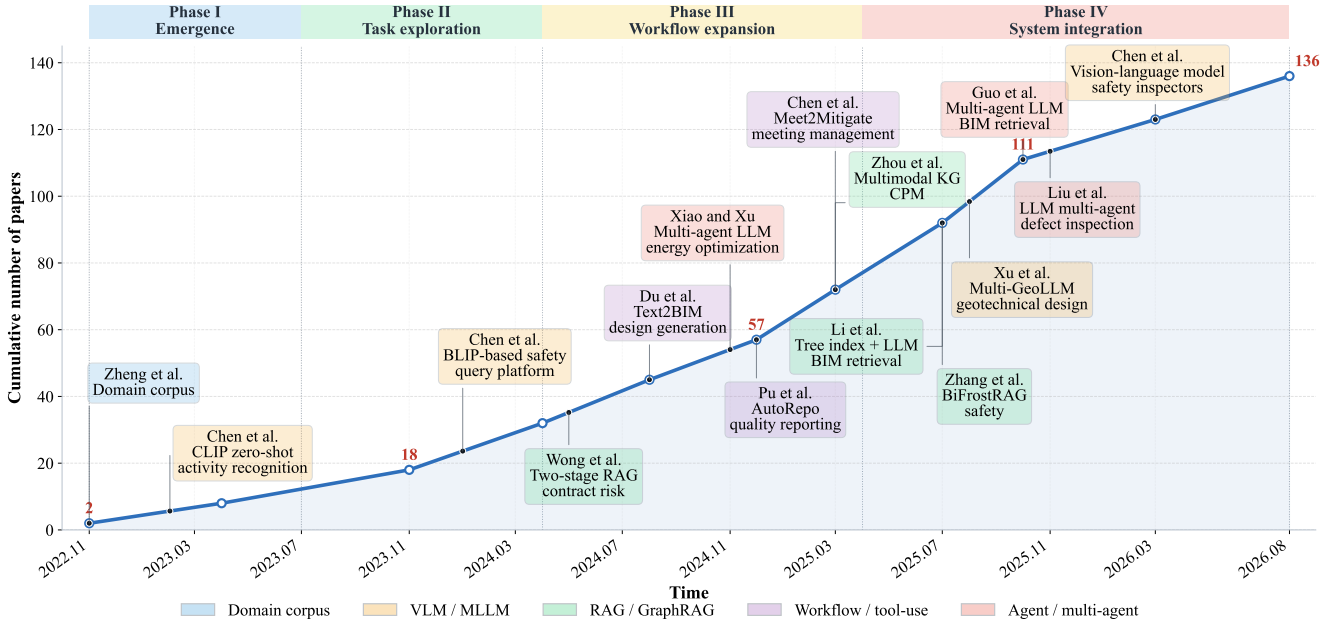

Figure 2. Publication trend and research focus of the 136 studies.

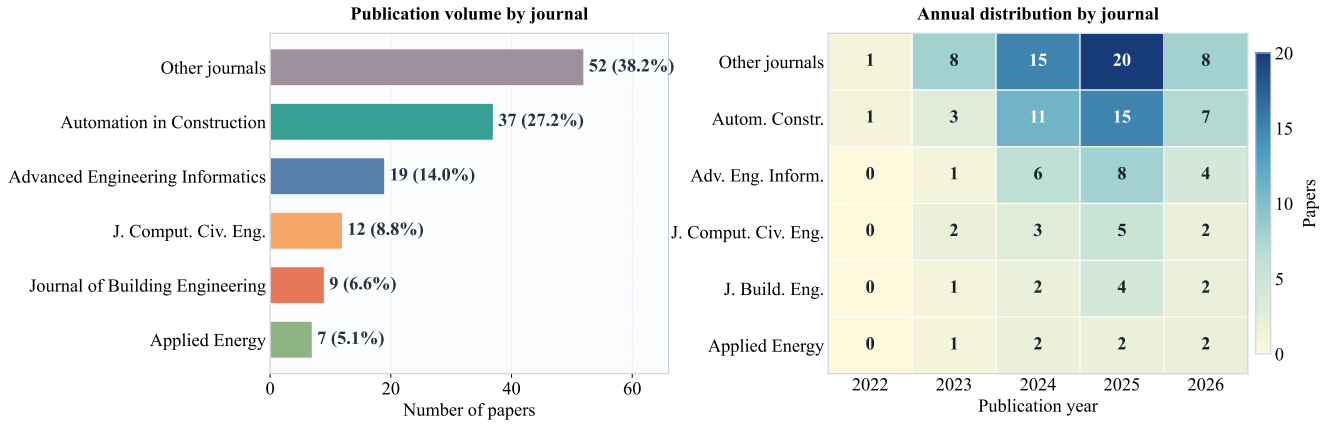

Figure 3. Journal distribution of the 136 included LFM-related construction studies.

Figure 4 shows the regional origins of the 136 reviewed studies and the institutions with relatively high publication counts. Mainland China contributed 50 studies, followed by the United States with 24, Hong Kong SAR (China) with 20, and Singapore with 8. At the institutional level, Tsinghua University contributed 19 studies and Zhejiang University 10. Southeast University, the National University of Singapore, and The Hong Kong Polytechnic University each contributed 8 studies. These results show that a substantial share of the reviewed work was produced by a relatively small group of institutions in East Asia, Singapore, and North America. (Note: THU = Tsinghua University; ZJU = Zhejiang University; SEU = Southeast University; NUS = National University of Singapore; PolyU = The Hong Kong Polytechnic University; HKUST = The Hong Kong University of Science and Technology; HKU = The University of Hong Kong; UIUC = University of Illinois Urbana-Champaign; CMU = Carnegie Mellon University; TAMU = Texas A&M University; NCSU = North Carolina State University; UofT = University of Toronto.)

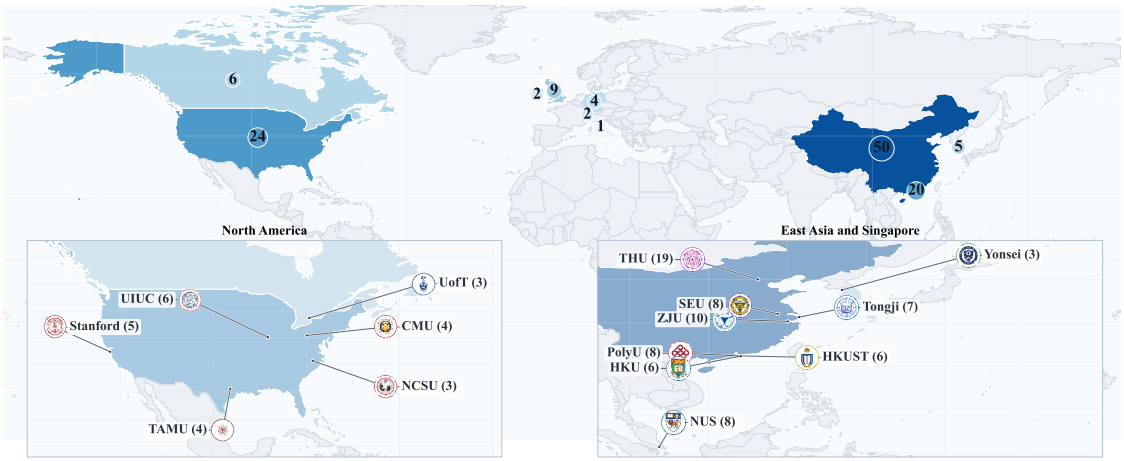

Figure 4. Regional and institutional distribution of the reviewed studies.

Figure 5 presents research flow across journals, regions, institutions, and application processes. The post-2024 studies show two relatively visible patterns in the Sankey diagram. The first route runs from construction informatics journals, particularly *Automation in Construction*, to mainland China and Hong Kong (China), then to leading institutions such as THU, ZJU, SEU, Tongji, PolyU, HKUST, and HKU, and finally to BIM/design, safety/quality, and execution/project management applications. In contrast, papers published in *Applied Energy* and other journals show a broader distribution, covering regions such as Singapore and the United States. The application areas of these papers are more closely related to Operations and Maintenance (O&M), energy, and end-of-life management. Several recurring combinations of publication venues, regions, institutions, and application areas can be observed in the Sankey diagram. This suggests that LFM research in construction is developing through multiple regional and application-specific clusters. In addition, there are differences in research focus among different regions and institutional groups, rather than a single, common development model.

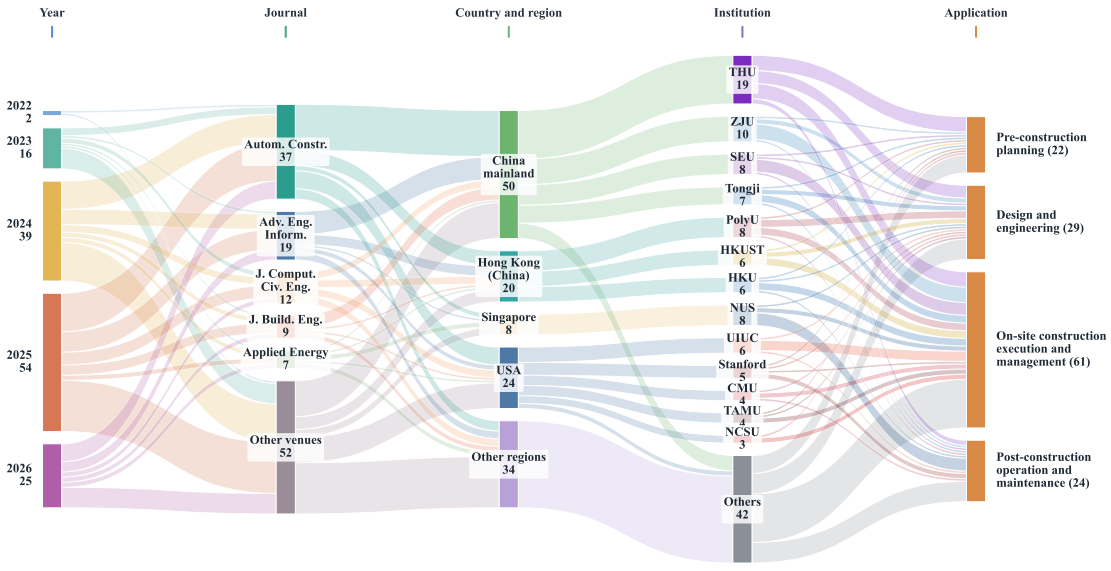

Figure 5. Research flows of the 136 included studies across publication year, journal, country/region, institution, and application domain.

3.2. *Keyword co-occurrence and author cluster analysis*

VOSviewer was used to construct the keyword co-occurrence and co-authorship networks. Figure 6 contains 49 keywords with at least five occurrences, while Table 2 groups them into six thematic clusters. The first cluster relates to LFM technologies. The term "large language model" occupies the central position, with close connections to ChatGPT, GPT, and generative pre-trained transformer. The second cluster links computer vision with construction safety, image captioning, Contrastive Language-Image Pre-training (CLIP), and safety inspection. BIM, artificial intelligence, project controls, and prompt engineering make up the third cluster, mainly around information retrieval and project-level decision support. The fourth cluster is associated with document processing, including BERT, named entity recognition, transfer learning, and construction management. The fifth and sixth clusters capture knowledge retrieval, construction dispute analysis, and model fine-tuning, respectively. Terms include RAG, knowledge graphs, construction contracts, risk identification, datasets, natural language processing, construction disputes, and fine-tuning. These terms reflect the use of external knowledge, language processing, and model adaptation in construction LFM studies.

Table 2. Thematic clusters identified from the keyword co-occurrence network.

Cluster	Theme	Representative keywords
C1 (blue)	LFM technologies	large language model, ChatGPT, GPT, generative pre-trained transformer, artificial general intelligence
C2 (red)	Vision-based construction perception	computer vision, construction safety, deep learning, image captioning, CLIP, safety inspection, personal protective equipment (PPE), convolutional neural network, natural language generation, intelligent construction
C3 (green)	BIM, AI, and construction workflow applications	building information modeling, artificial intelligence, machine learning, virtual assistant, prompt engineering, project controls, building energy modeling, construction site safety monitoring
C4 (purple)	Semantic extraction and document understanding	natural language processing (NLP), BERT, named entity recognition (NER), transfer learning, construction management, bridge inspection
C5 (yellow)	Knowledge retrieval, contract reasoning, and data resources	retrieval-augmented generation (RAG), knowledge graph, construction contract, risk identification, dataset, deep neural networks, construction site
C6 (cyan)	Natural language processing and construction dispute analysis	natural language processing, construction dispute, fine-tuning

Figure 7 presents the main co-authorship links in the reviewed literature. Several relatively distinct author groups can be identified. Bo Xiao and his coauthors have worked on construction-site applications involving visual reporting, activity recognition, question answering, and safety hazard identification [21,25–27]. Jack C. P. Cheng and his coauthors have mainly studied BIM information retrieval, multimodal RAG, semantic query processing, and construction safety monitoring [24,28–30]. Most collaborations remain concentrated within individual research teams, with fewer links between different groups.

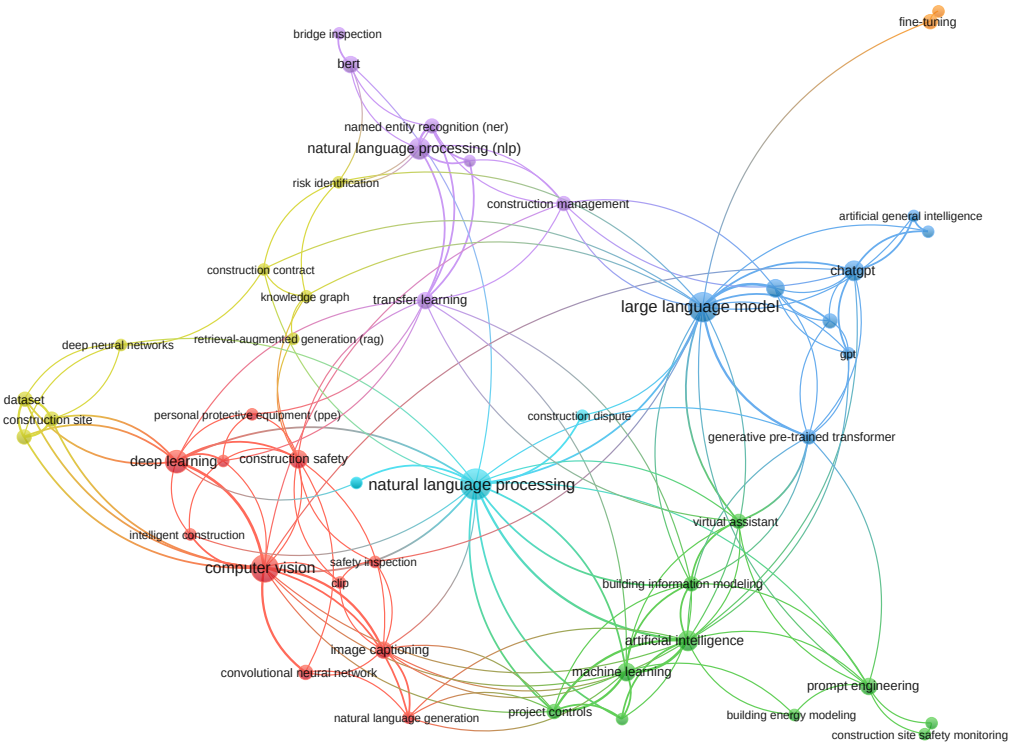

Figure 6. Keyword co-occurrence network.



Figure 7. Co-authorship network.

4. Qualitative analysis from the data perspective

From the 136 included studies, 86 unique construction datasets were identified for dataset-level analysis. This section reviews construction data from five aspects: modality distribution, source accessibility, semantic granularity, data processing, and benchmark readiness. Details of the 86 datasets are provided in Appendix A.

4.1. Data type and modality distribution

Construction data used in the reviewed LFM studies span text, visual records, three-dimensional models, and time-series records. Figure 8 summarizes the modality and subtype distribution of 86 construction datasets used in LFM-related studies. Text data dominate the corpus, with 49 datasets (57.0%), followed by image-text data (17, 19.8%), sensor data (13, 15.1%), 3D/BIM data (4, 4.7%), and image data (3, 3.5%). Within the text category, codes, standards, and guidance documents continue to support regulatory interpretation, compliance analysis, and domain-specific question answering [16,29,31–33], whereas contracts, procurement documents, and litigation records provide the basis for contractual reasoning and legal analysis [34–37]. Recent studies further move from document collections toward purpose-built evaluation and adaptation resources. AECBench and Arch-Eval provide structured benchmarks for assessing construction- and architecture-specific knowledge [38,39], while contract benchmarks, fine-tuning corpora, and large collections of construction-law cases support more systematic evaluation of contractual and regulatory reasoning [40–44]. Moreover, project-level and operational records cover planning meeting transcripts [45], construction management best-practice cases [46], defect inspection reports [47], energy modeling data [48,49], building energy operation records [50], design software documents [51], and construction news [52]. Visual and multimodal datasets remain less prevalent than text, but their scope has broadened considerably. Earlier image datasets mainly represent construction-site objects and equipment [53], workers and site activities [25,54], architectural and structural drawings [55], and defects such as cracks [56]. Image-text datasets further support safety reasoning and visual question answering [11,26,57–59], project-management and geotechnical reasoning [10,60], worker ergonomics [61], and automated inspection reporting [62]. More recent datasets extend these uses beyond recognition and captioning. Adapted vision-language models are trained and evaluated on construction fall-hazard image-text data [63], while UndergrBench organizes multimodal evidence for construction reasoning in underground engineering [64]. Sensor data remain relatively limited and are mainly used in building energy and operation studies, where long-term environmental, equipment, and energy-consumption measurements support operational monitoring and temporal analysis [50,65]. 3D data are primarily represented by BIM/Industry Foundation Classes (IFC) models, point clouds, and scan-to-BIM records, providing geometric and semantic information for model interpretation, spatial reasoning, and site assessment [49,66,67]. Although still less common than text and image-based resources, these two data types extend construction LFMs from document- and image-based understanding toward temporal and geometry-aware reasoning.

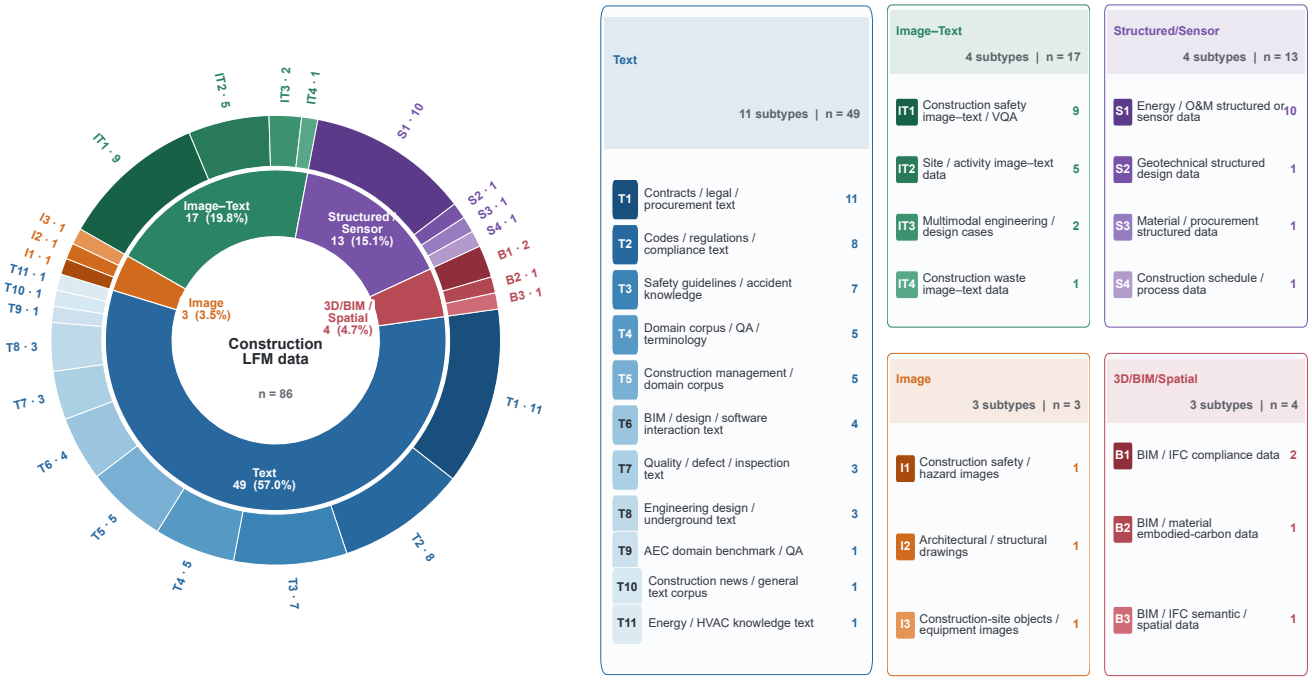

Figure 8. Type and modality distribution of the 86 datasets.

4.2. Data source and accessibility distribution

Figure 9 shows the distribution of the 86 identified construction datasets by primary data source (a) and accessibility status (b). Current construction LFM data primarily come from four types of sources. The first category comprises data collected from real-world construction projects and operational activities, often supplied by industry partners. This includes bid documents, project records, and construction monitoring data [65,68]. The second category comes from government agencies, professional organizations, and other industry institutions. This category primarily comprises standards, specifications, safety guidelines, and technical documents, which typically provide expert knowledge or a rule-based foundation [32,51]. The third category consists of existing datasets or publicly available online resources directly reused by researchers, such as construction site photos, news reports, and other published materials [52,69]. The fourth category consists of research-generated data, such as design case studies or simulation data [70,71].

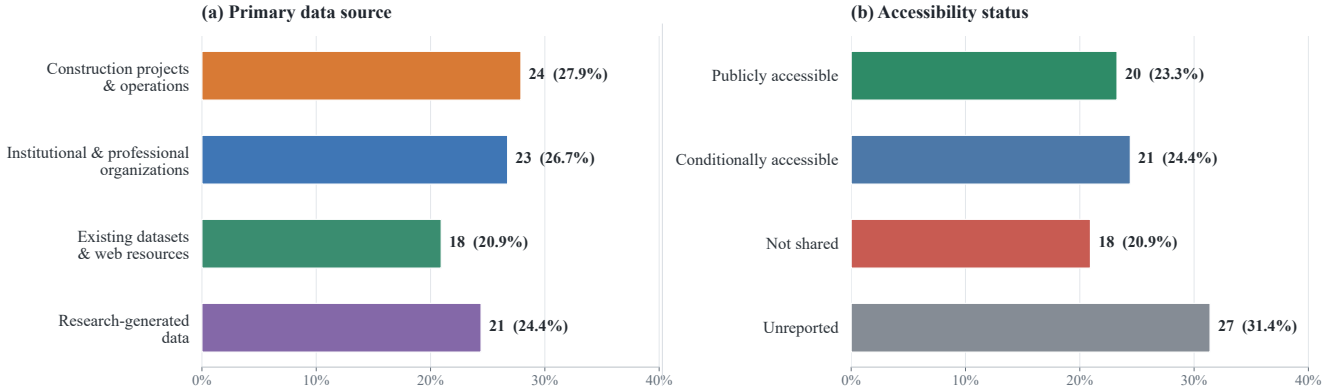

Figure 9. Data sources and accessibility of the 86 datasets.

Accessibility varies across these source groups and can be separately categorized as publicly accessible, conditionally accessible, or not shared. Unreported access conditions are recorded separately. Confidentiality and ownership restrictions on project and operational data lead to most of the data being available via request-based access or partial releases [68,72]. Institutional and professional materials, by contrast, are distributed through official websites and publishers. For example, OSHA regulations provide publicly accessible reference materials [22]. However, established distribution channels do not necessarily provide free or unrestricted access, as commercial materials such as FIDIC contract forms require purchase and remain subject to licensing conditions. Existing datasets offer another route to obtaining research inputs without collecting new project records, but their reuse depends on the completeness of the released materials. VisualSiteDiary, for example, provides captions while directing users to the original authors for the corresponding images [69]. Web collections additionally require a record of the selected items and collection period, since access to a website alone does not identify the exact material used in a study. Research-generated data reduce reliance on existing records, but shift the reproducibility requirement toward the release of generated samples. Thus, each source requires a different form of access documentation: provider authorization for project records, publication and licensing conditions for professional materials, original files and collection details for reused data, and sample or generation-procedure releases for generated data.

4.3. Data-task alignment and semantic granularity

The reviewed datasets remain better suited to perception, extraction, retrieval, and question answering than to higher-level construction reasoning. Some datasets still inherit conventional DL, Computer Vision (CV), or NLP dataset formulations, where construction information is represented as object categories, bounding boxes, defect classes, named entities, or document-level labels [25,53–56]. However, they provide limited support for LFM-oriented tasks, which require more explicit connections among document chunks, source metadata, cited evidence, visual regions, textual descriptions, and domain concepts. This difference is also reflected in data granularity. Textual data are often organized at the document, paragraph, clause, or QA-pair level, which supports retrieval and answer generation, but entity-relation structures and rule-level representations are less consistently encoded. Image and image-text datasets extend the data unit from isolated visual labels to paired visual-linguistic samples. However, these image-text datasets mainly provide image-caption, image-question, or inspection image-report correspondences [11,26,57–59]. Finer-grained semantic links such as region-level grounding, hazard-cause relations, defect-consequence relations, and corrective-action rationales are less commonly represented. Similarly, video datasets record temporal construction activities rather than event-process-outcome chains [21,25,46]. Current construction datasets mainly support LFMs in retrieval, extraction, perception, and answer generation. Fewer datasets explicitly represent rule constraints, process dependencies, state transitions, resource decisions, role-specific interactions, or human feedback. This indicates that data structures for planning-oriented, process-aware, and decision-centered construction LFMs are still at an early stage.

4.4. *Data transformation and processing*

A critical characteristic of construction LFM datasets is the transformation of heterogeneous raw data into model-ready representations. As shown in Figure 10, construction data are collected from heterogeneous sources and stored in diverse formats such as PDF, CSV, JSON, XML, JPG, and RVT. Most datasets require additional processing before they are used for LFM training, retrieval, or evaluation. The processing pipeline for textual datasets includes document parsing, Optical Character Recognition (OCR), semantic chunking, and vectorization or embedding [34,35,73]. Image and image-text datasets usually require object annotation, captioning, visual question construction, or region-phrase alignment [57,58,62,74]. For structured engineering data, BIM/IFC models, energy models, and building operation records need to be converted into textual descriptions, parameter-scenario pairs, component attributes, or machine-readable representations [48–51]. These transformations show that construction LFM datasets are increasingly moving from raw data collection toward task-oriented data reconstruction. However, the processing strategies remain heterogeneous across studies. Therefore, the readiness of a construction dataset depends not only on its modality or size, but also on whether its processing pipeline retains the evidence, context, and task structure required by LFMs.

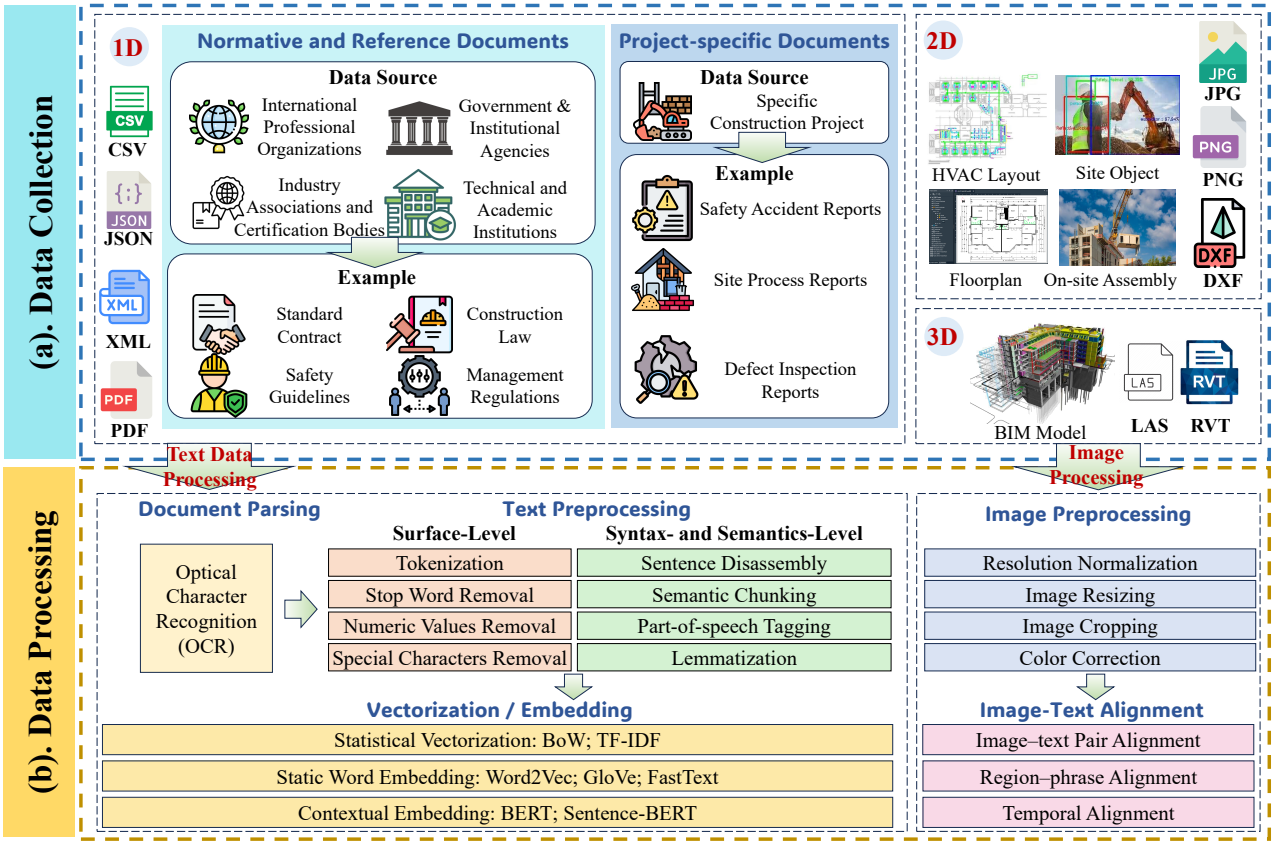

Figure 10. Transformation pipeline from raw construction data to LFM-ready construction datasets.

4.5. *Benchmark readiness and evaluation use*

Current construction LFM datasets serve four main functions: domain pre-training, external knowledge support, task fine-tuning, and model evaluation. Domain corpora are used to learn the language and domain-specific knowledge of construction, such as Chinese AEC regulatory corpora [20], structural

engineering corpora [73], and Construction Management Systems (CMS) academic corpora [75]. The knowledge base provides models with external information during the reasoning phase via retrieval or RAG. The external information may include construction specifications, safety regulations, best practices, and energy audit knowledge [16,29,31,57,76]. Fine-tuning datasets further organize domain data into input and output samples suitable for specific tasks [48,49]. In contrast, data resources used for model evaluation are relatively limited. Although many studies have constructed test datasets to validate specific models or use cases, such evaluation data do not qualify as shared benchmarks [26,32,77–79].

5. Qualitative analysis from the algorithm perspective

5.1. Foundation model evolution and construction-specific adaptation

Since the public release of ChatGPT in late 2022, LFM capabilities have expanded from conversational language processing to multimodal information, complex reasoning, and external tool execution. First, instruction tuning, dialogue alignment, and human-preferred optimization have improved the model's ability to follow user instructions [80,81]. Then, longer context windows enabled the model to process long documents, project logs, and continuous task histories [82–84]. At the same time, multimodal models expanded the model inputs from text to images, audio, and video [85,86]. Recently, reasoning models have improved performance on multi-step problems, particularly in mathematics, programming, and scientific tasks [87–89]. Tool-use capabilities enable LFMs to call functions, execute code, and interact with external software [90,91].

LFM-enabled construction research, however, has not adopted these capabilities at the same pace. Language-based adaptation and retrieval are now relatively common, whereas multimodal reasoning, tool use, and workflow-level integration remain much less established. Early work in construction was largely text-based. BERT-style encoders were used for text classification, entity recognition, semantic similarity, and extractive question answering in contracts, regulations, and safety documents [35,36,74,75,77]. Generative LLMs extended these applications to document analysis, knowledge retrieval, report generation, and construction-specific question answering [16,58,76,92,93]. Open-source LLMs also made domain-specific fine-tuning and private deployment more practical, and were subsequently explored for construction management, schedule reasoning, and policy interpretation [46,73,76]. Long-context models have also been examined for interpreting building codes with dependencies across provisions [94]. Multimodal construction studies have also begun to use CLIP, BLIP, GPT-4V, and Qwen2-VL for visual inspection, safety checking, image-based reasoning, and multimodal retrieval [21,54,58,95,96]. Recent tool-supported and multi-agent studies connect LFMs with engineering software for BIM authoring, automated compliance checking, and building energy modeling [9,97,98]. Construction applications thus extend from individual document and image analysis tasks to assistance with multi-step engineering workflows.

5.2. Domain adaptation for construction tasks

Construction information is domain-specific, heterogeneous, and project-dependent. Therefore, construction LFMs require domain adaptation to achieve reliable interpretation, reasoning, and decision support. Existing studies mainly follow five adaptation paths.

5.2.1. Model-level domain and task adaptation

Model-level adaptation modifies pretrained models using construction data so that their internal representations and output behavior better reflect domain terminology, task requirements, and engineering constraints. Existing studies mainly involve three learning objectives: domain-adaptive pretraining, supervised fine-tuning, and reward-based post-training. Domain-adaptive pretraining continues training on large construction corpora and is therefore suited to learning broad construction domain language before a specific downstream construction task is defined. Its main advantage is broader domain coverage, but it requires sufficiently large and representative corpora and introduces additional training cost [75]. Supervised fine-tuning (SFT), by contrast, directly learns mappings from construction inputs to task-specific outputs and is therefore more targeted. This approach has been applied to building-code classification and named entity recognition in Qwen-AEC, as well as construction-law identification and building-code question answering [3,43,44]. SFT can adapt a model to a specific construction task with a relatively limited amount of task-specific data. However, it remains sensitive to the coverage and quality of task-specific training examples. Parameter-efficient fine-tuning (PEFT) reduces the number of trainable parameters relative to full-model fine-tuning. In construction, LoRA has been used in conjunction with ChatGLM3-6B for named entity recognition [73]. UndergrGPT adopts parameter-efficient adaptation for underground-engineering language tasks, and BuildThemis uses QLoRA for building-code compliance [1,99]. More recent works introduce reward-based post-training to optimize outputs by task-specific rewards [40,71]. The effectiveness of these approaches depends on the design of reward signals and the scope of training cases.

5.2.2. Prompt-level adaptation strategies

Prompt-level adaptation guides LFM behavior at inference without updating model parameters. Existing construction studies mainly use four complementary strategies: instruction and constraint prompting, contextual and example-based prompting, structured reasoning prompting, and role-based prompting, as shown in Figure 11. Instruction prompts define the task, required inputs, engineering constraints, and output format. For example, NADIA uses explicit instructions to guide exterior-wall detailing, while Text-to-Layout specifies spatial and formatting requirements for generating structured floor-plan representations [100,101]. Contextual and example-based prompts provide project information or demonstrations that are relevant to the current task. Few-shot examples and system-level context have been used for translating building regulations into structured rules, while Multi-GeoLLM retrieves similar design cases to construct task-specific geo-prompts [10,102]. Structured reasoning prompts further organize how a model processes complex tasks. Construction studies have explored chain-of-thought and self-consistency for regulatory interpretation, TriPhase and decision-tree prompts for visual safety

compliance, and structured prompts for accident-cause analysis [29,102,103]. More task-specific designs also use rule-based adaptive prompts to map regulatory clauses to executable functions [104]. Role-based prompting assigns professional responsibilities to the model or individual agents. For example, recent building-layout systems separate architectural, accessibility, and user-experience perspectives during design reasoning [105]. Compared with model-level adaptation, these strategies require no additional parameter training and can be modified for different projects or tasks. However, their effectiveness remains sensitive to the information and constraints provided in the prompt. Prompting alone cannot reliably supply missing or updated domain knowledge. This limitation motivates the use of external knowledge representation and retrieval mechanisms.

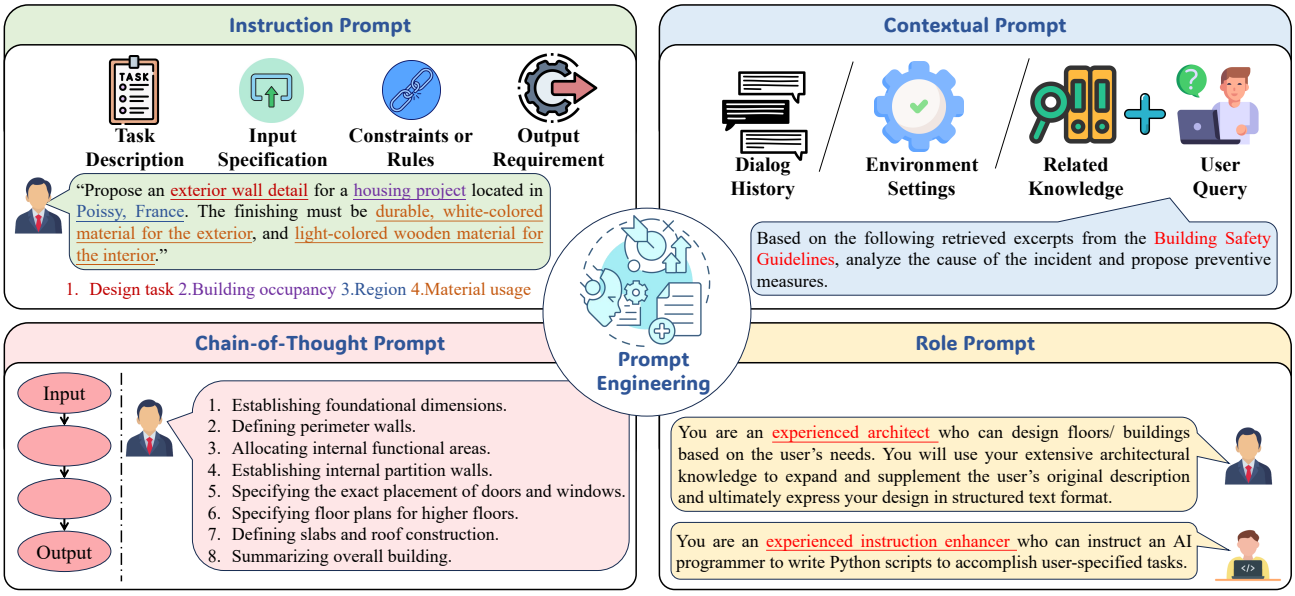

Figure 11. Prompt-level strategies for construction semantic adaptation.

5.2.3. Knowledge representation and retrieval-augmented reasoning

Knowledge representation and retrieval-augmented reasoning connect LFMs with external construction knowledge. This external knowledge is difficult to encode completely in model parameters, particularly project records, standards, contracts, safety cases, and other frequently updated domain documents. As shown in Figure 12, existing approaches can be organized into vector-based, graph-based, and hybrid RAG according to their knowledge index and retrieval mechanism. Vector RAG retrieves semantically related document segments through embeddings. Earlier studies used vectorized domain knowledge for contract and construction-management question answering [7,11]. More recent work improves this process through domain-adapted embeddings, semantic chunking, hierarchical knowledge units, and knowledge-guided retrieval [41,106,107]. These methods are well suited to large collections of unstructured documents, but their performance is sensitive to how the source material is segmented and whether important cross-sentence or cross-section dependencies are retained. Graph-based RAG represents domain knowledge through entities, relations, clauses, or document structures and retrieves information through graph relations or traversal. For example, Zheng *et al.* [8] organized contractual clauses and their relations in a nested contract knowledge graph to support automated contract review. Zhou *et al.* [26]

used a multimodal knowledge graph to connect project-management information for question answering. In construction safety, Lee *et al.* [32] represented safety-guideline knowledge as a graph to support knowledge retrieval and reasoning. Graph structures preserve explicit relations more effectively than isolated text chunks. Hybrid RAG combines complementary retrieval channels from Vector RAG and Graph RAG. Building-engineering knowledge management has combined vector, property-graph, and keyword retrieval; BifrostRAG integrates semantic vector search with traversal over entity and document graphs; and recent off-site construction quality-control research aligns dense retrieval, sparse retrieval, and subgraph traversal [22,31,108].

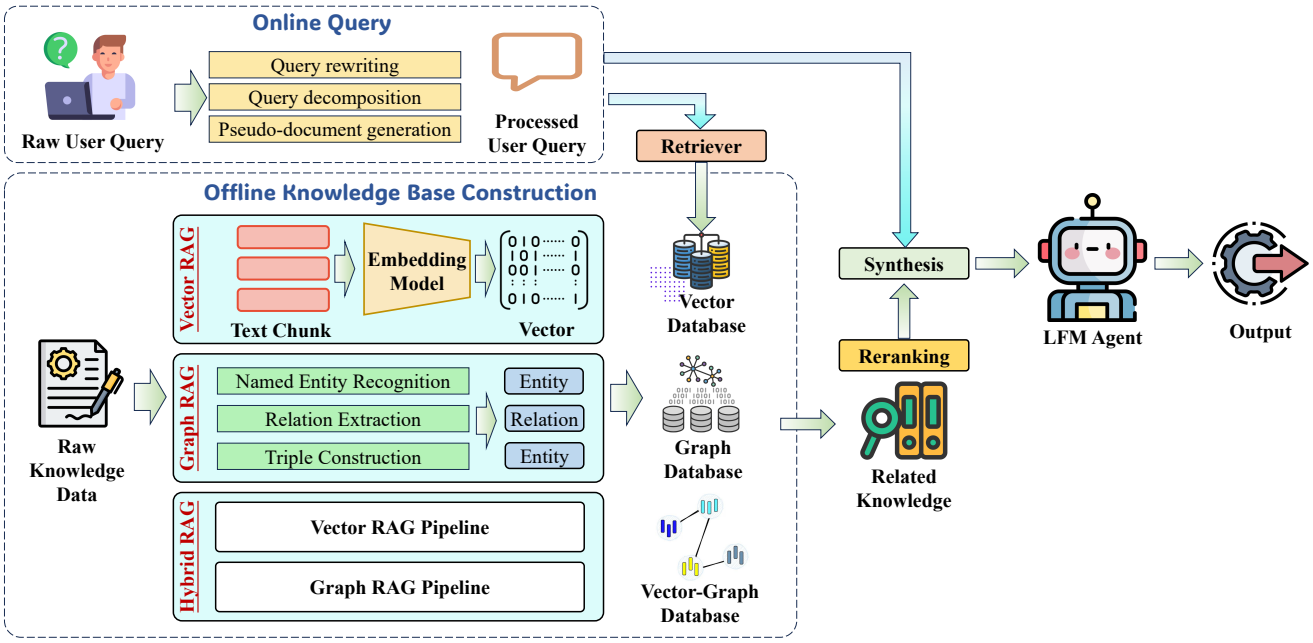

Figure 12. Knowledge representation and retrieval-augmented reasoning for construction LFMs.

5.2.4. Multimodal adaptation and spatial understanding

Multimodal and spatial understanding is essential for construction LFMs. Substantial construction knowledge is distributed across heterogeneous textual, visual, geometric, and temporal sources. These sources jointly describe the physical state, spatial configuration, safety condition, and management process of a project. Multimodal construction tasks require LFMs not only to understand "what information is contained" in images, text, or BIM data, but also to determine "which objects this information corresponds to, where they are located, and what spatial relationships they have" in the real-world engineering environment. The encoding and fusion components in Figure 13 illustrate how visual information can be associated with language representations. Existing research has begun to utilize VLMs to process construction site images, engineering drawings, and point cloud data, and has enhanced the understanding of engineering scenarios through domain knowledge, region segmentation, geometric matching, and BIM element mapping. Liang *et al.* [54] use CLIP-based multimodal prototypes for few-shot recognition of temporary site objects without additional model training. For image-to-text tasks, ChatCH adapts Qwen2-VL-7B for hazard identification and structured report generation, while CS-VLM uses standardized image-caption pairs and parameter-efficient fine-tuning to describe fall hazards [59,63].

These studies use paired visual and textual data to associate site conditions with task-specific descriptions. Construction knowledge can also be introduced during image or multimodal reasoning. Chan *et al.* [28] use ontology-guided prompts to incorporate safety concepts, whereas Guo *et al.* [29] retrieve relevant regulations when assessing conditions observed in site images. Tasks involving spatial information require an additional link between model outputs and identifiable regions or building elements. Ali *et al.* [109] divide Mechanical, Electrical, and Plumbing (MEP) drawings into informative regions before using a VLM to classify and locate equipment symbols. Gong *et al.* [67] combine MLLM-assisted point-cloud segmentation with scan-vs-BIM matching and geometric verification to associate site observations with BIM elements. Video-based reporting provides a related multimodal workflow, where computer vision first extracts site activities and productivity measures and ChatGPT then converts these results into a textual report [25]. However, a fluent or semantically correct description does not guarantee accurate spatial interpretation. A model may correctly identify a safety hazard or component type while associating it with the wrong image region or BIM element, or misinterpreting its dimensions and spatial relationships. Therefore, for construction tasks involving spatial information, model outputs should be further verified against corresponding image regions, geometric relationships, or BIM elements. Precise geometric reasoning, understanding of scale and dimensions, interpretation of engineering drawing symbols, BIM semantic mapping, and continuous tracking of construction status over time remain unresolved challenges in current VLM/MLLM research.

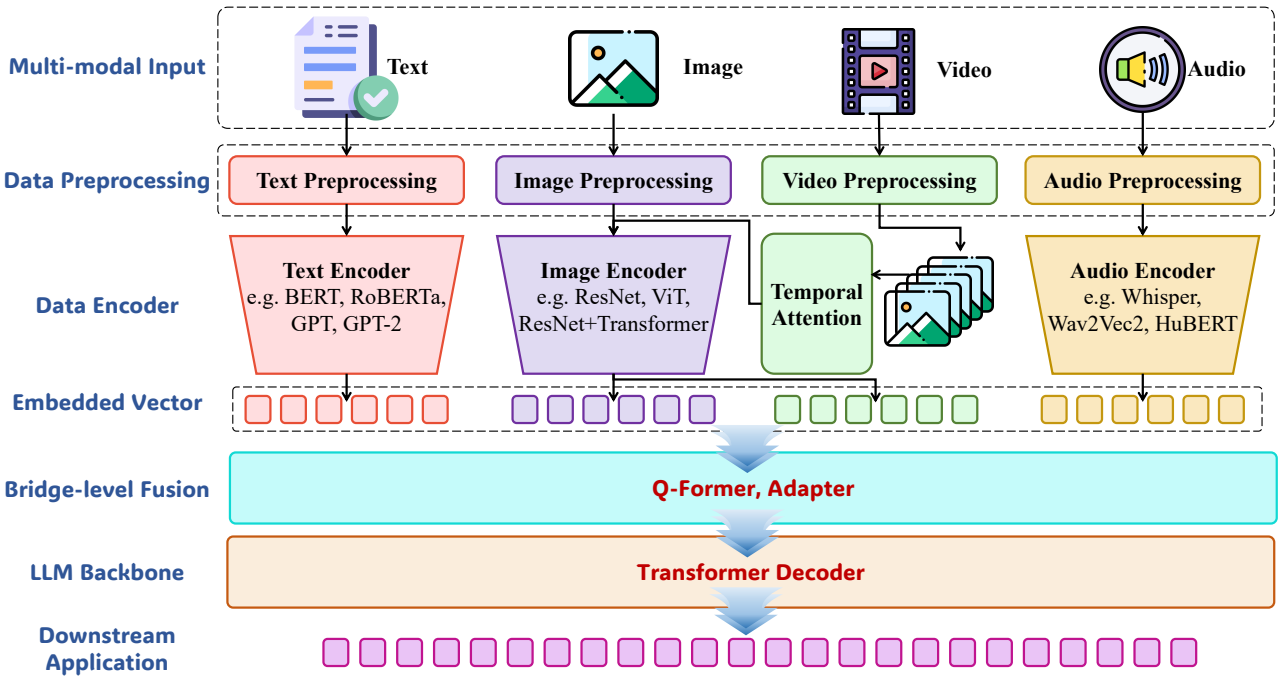

Figure 13. An illustrative architecture of LLM-based multimodal models.

5.2.5. Tool-augmented generation and executable workflows

Construction workflows often require tool-based interaction, as many tasks involve database queries, digital model modification, or software execution. Consequently, recent research has begun to integrate LFMs with Application Programming Interfaces (APIs), code interfaces, databases, and simulation

tools, enabling natural language instructions to be translated into specific engineering operations. Early methods primarily focused on a single function or API call. For example, Guo *et al.* [24] map natural-language BIM queries to corresponding Revit API functions. LLM-FuncMapper links regulatory clauses to predefined computational functions before generating the required checking code in compliance checking [104]. Text2BIM further uses callable Vectorworks functions to execute generated scripts and create editable building models [9]. Research has gradually expanded from single-step code generation to multi-step workflows, in which multiple agents or tools sequentially perform information parsing, model querying, code generation, verification, and execution. Intermediate results must therefore be passed reliably between successive steps. AutoGen-Revit separates code generation, checking, and execution among different agents [110]. IFC-Agent chains schema-aligned IfcOpenShell operations and shares entity information across agents for IFC querying and modification [111]. Automated compliance-checking systems also connect regulatory interpretation with BIM data extraction, check execution, and reporting [97]. Furthermore, some systems have begun to use actual software execution results to form a closed-loop feedback system. In Text2BIM, interpreter errors and model-checker results are returned to the agents so that generated scripts can be corrected before another execution attempt [9]. Building-energy workflows assign debugging agents to correct EnergyPlus input files [112]. Data2BEM connects the agents with OpenStudio and EnergyPlus for model generation, simulation, and calibration against measured energy consumption [98]. A similar feedback loop is used in building O&M, where agents invoke virtual in-situ modeling and fault-diagnosis applications as external tools [113]. In building disassembly, agent-generated plans are passed to a MoveIt action library and checked in MuJoCo before execution [114]. This suggests that, in such systems, the primary role of LFMs is not to replace BIM, simulation, or robotics software, but rather to interpret tasks, select and organize tools, and coordinate subsequent operations based on execution feedback. Engineering calculations and physical verification are still performed by specialized software. At the same time, existing evaluations show that successful software interaction does not necessarily imply reliable engineering execution. Reliability decreases as the task and software interaction become more complex. Alwashah *et al.* [115] evaluated 31 BIM automation tasks and recorded 238 execution and modeling errors across ten categories, of which API misuse accounted for 48.3%. In the EnergyPlus model generation, although the syntax pass rate reached 98.33% after SFT processing, the end-to-end validity was only 24.67% [71]. Errors in LFM output may also directly affect engineering models and workflows. Tool-augmented construction workflows require the establishment of stricter control mechanisms in areas such as interface constraints, access control, execution verification, process logging, rollback, and manual approval.

5.3. Evaluation and validation of LFM outputs

In construction, the evaluation methods for LFMs are closely related to the format of model outputs. For tasks with clear reference answers, such as classification, named entity recognition, and closed-ended question-answering, existing research still primarily uses traditional metrics to compare results against annotated ground truth. For example, Qwen-AEC reports weighted F1 for building-code classification and macro F1 for entity recognition [44]. As LFMs are increasingly used in RAG and open-ended question-answering, evaluating only the final answer is no longer sufficient to reflect system performance.

RAG systems require separate validation of retrieved evidence and generated answers. For example, BifrostRAG evaluates whether the correct regulatory sections are retrieved using precision, recall, and F1, while CC-RAG focuses on the quality of answers to construction-contract examination questions [22,41]. For open-ended suggestions and long-form text responses, due to the lack of a single reference answer, experts or users are typically required to evaluate their technical accuracy, relevance, and completeness. For instance, Uhm *et al.* [92] invite researchers, safety experts, and workers to review generated safety advice. AECBench also uses expert-defined rubrics to evaluate long-form responses and reports differences between model-based and expert scores [38]. Multimodal tasks introduce an additional need for spatial verification. ConstructionSite 10k evaluates caption quality, safety-rule detection, and visual grounding separately [116]. Its results indicate that detailed textual descriptions do not necessarily correspond to accurate hazard localization. When models further generate code or engineering models, successful parsing and execution do not guarantee that the final results meet engineering constraints and code requirements. Evaluation must also verify geometric constraints, compliance with regulations, and whether the final task has been completed [97,115]. For numerical predictions and operational decisions, it is also necessary to confirm their feasibility under actual engineering conditions [98,117]. Therefore, the evaluation of construction LFMs is gradually expanding from traditional output matching to multiple levels, including professional judgment, spatial verification, execution validity, and final engineering outcomes.

6. Qualitative analysis from the application perspective

From the application perspective, the 136 reviewed studies are reorganized into four primary research processes: pre-construction planning, design and engineering, on-site construction execution and management, and post-construction operation and maintenance, as shown in Figure 14.

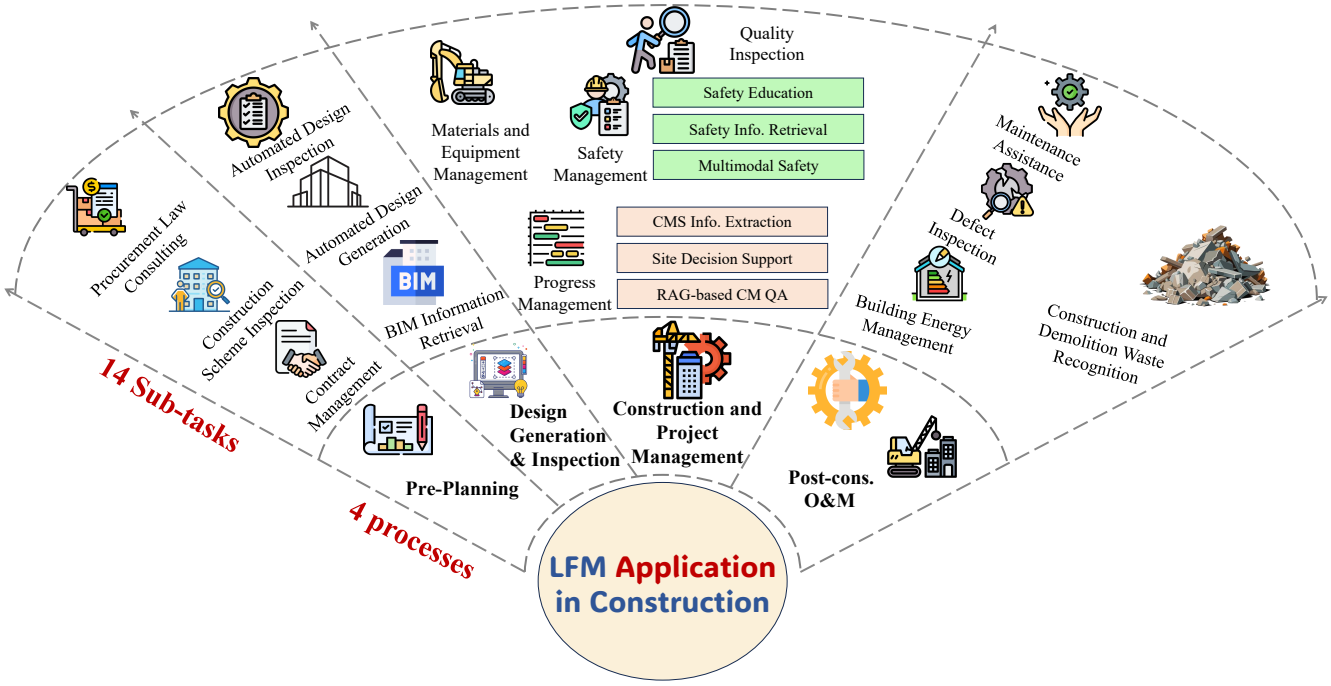

Figure 14. LFM applications across four construction lifecycle processes and 14 sub-tasks.

6.1. Pre-construction planning, contracting, and procurement

In the pre-construction stage, LFM studies focus on contract review, planning document analysis, construction scheme inspection, and procurement support. Contract-related research forms a particularly large part because contracts contain complex quality standards, acceptance criteria, risk allocation, and responsibility distribution. Early studies treated contract analysis as a set of text-processing tasks, and mainly applied BERT-based models for clause classification, entity extraction, and risk identification [34,35,118]. More recent LLM-based approaches move beyond clause-level analysis toward the review of longer documents. Long-context chunking and RAG enable LLMs to process lengthy contract documents that exceed the native context window [119]. Domain knowledge bases, two-stage prompting, and graph-enhanced retrieval have further been used to improve the recognition of implicit risks, nested clauses, and multi-layer contractual relationships [7,8]. LFMs have also been used for construction scheme inspection and procurement consulting. Early attempts employed ChatGPT to generate preliminary construction schedules directly [120]. BIM-NLP integration has also been used to transform design information into structured construction schedules [121]. Moreover, knowledge graphs and LLMs have been combined to parse construction schemes and check compliance with standards [122]. LFMs have also been applied to government construction procurement consulting and bid preparation, including requirement extraction, bid/no-bid support, and proposal drafting [37,68].

6.2. Design generation and compliance checking

During the design process, LFMs are mainly used for BIM information retrieval, design generation, and automated compliance checking. BIM integrates lifecycle information on building elements, spatial layouts, MEP components, and construction processes. However, effective information retrieval from BIM models requires familiarity with complex interfaces, object hierarchies, and domain terminology [123,124]. Early language-model applications used natural language as an interface to BIM information. RoBERTa-based models classified BIM queries into predefined categories, while GPT-based systems translated user questions into structured BIM queries and returned both textual and visualized results [125]. Recent studies extend this direction to spatial reasoning in IFC/BIM models, where multi-agent LLM frameworks coordinate natural-language parsing, spatial indexing, and relationship computation to answer queries about adjacency, distance, and component relationships [24,30]. Another major direction uses LLMs as design-generation or design-authoring agents. Early BIM-LLM workflows transformed BIM into XML or other structured representations, used LLMs to modify design details, and then transformed the results back to BIM [100,126]. Other studies embedded LLM-based copilots within BIM software to interpret natural-language instructions and invoke predefined modeling tools [51]. More recent work generates imperative code that directly calls BIM authoring APIs, which support editable BIM models with internal layouts and semantic attributes [9]. Recent studies increasingly use LLM-generated code to directly create or modify BIM models rather than limiting the model to the selection of predefined tools. This gives the workflow more direct control over geometry and object properties. LLMs have also been examined for Automated Compliance Checking (ACC) [127]. In these studies, regulatory provisions are interpreted and converted into computable rules or checking logic, which can then be

applied to BIM-based design review [102,128]. Related design-review applications also use ChatGPT and computer vision to retrieve historical RFI records and predict design or construction issues [74].

6.3. On-site construction execution and management

6.3.1. Execution management and site coordination

LFM applications in construction management cover progress reporting, planning support, meeting analysis, knowledge retrieval, site-object and activity recognition, and robotic assembly. For construction records, BERT and RoBERTa pretrained on construction-management corpora have been fine-tuned for text classification and named entity recognition [75]. Site videos are also used to prepare management reports. Xiao *et al.* [25] proposed a framework that extracts activity and productivity information from videos and then uses ChatGPT to generate daily construction reports. For planning meetings, BERT classifies constraints in meeting transcripts, while GPT-based tools help analyze their causes, impacts, and possible solutions [45]. Meet2Mitigate combines speaker diarization, speech recognition, and LLM summarization to identify project issues in meeting recordings. It then retrieves relevant best practices to propose action items for the issues discussed [46]. Knowledge retrieval and question answering help project teams access information in project documents and domain knowledge bases. RAG4CM organizes project documents into a hierarchical knowledge pool and incorporates user preferences to improve retrieval and answer generation [11]. Other work combines vector, property-graph, and keyword indexing to retrieve relevant engineering knowledge [31]. Multimodal knowledge graphs have also been used to provide construction-management knowledge for LLM-based question answering [26]. A related approach combines an external construction engineering management knowledge base with fine-tuned LLMs to answer professional and case-based questions [78]. For site monitoring, zero-shot activity recognition has been used to identify excavator activities and estimate earthmoving productivity from videos [21]. CLIP-based few-shot learning addresses the recognition of temporary construction objects using a limited number of labeled images [54]. RoboGPT applies ChatGPT to construction assembly planning by generating robot assembly sequences [129]. Another study proposes a hierarchical code-generation framework that connects LLMs with customized APIs to produce executable control programs for construction robots [79].

6.3.2. Safety, quality, and on-site compliance assurance

Early studies on construction safety examined ChatGPT for hazard recognition and safety training [130]. Hussain *et al.* [131] incorporated a ChatGPT-based conversational instructor into a virtual reality environment for training migrant workers. For scenario-based training, Naderi and Shojaei used a database of verified safety information to generate hazardous situations and feedback for learners [76]. RAG-based systems also retrieve relevant safety guidance to suggest preventive measures and generate context-specific safety information [32,92]. BifrostRAG addresses safety questions that require information from several regulatory clauses. It combines a graph of entity relations with a graph of document structure to support multi-hop question answering [22]. Multimodal applications use site images and textual information to identify hazards and assess safety conditions. The Visual Construction Safety Query (VCSQ) system combines image captioning, visual question answering, and image-text retrieval

with an augmented reality interface [58]. CLIP-based image captioning has been used to describe safety violations and assist inspection reporting [96]. Related work uses vision-language models to identify ergonomic problems and suggest corresponding solutions from images of construction workers [61,132]. ChatCH uses a fine-tuned vision-language model for hazard identification and supports the generation of structured hazard reports [59]. Other approaches combine multimodal LLMs with domain knowledge and multi-step reasoning for safety inspection [57], or retrieve relevant regulations to assess worker behavior and safety compliance in site images [29]. Quality-assurance applications cover document review, compliance checking, and inspection reporting. For prefabricated buildings, BERT-based text parsing and knowledge-graph querying support quality diagnosis and inspection-related question answering [33]. Knowledge-guided retrieval has also been used to locate regulatory information for reviewing construction disclosure documents [107]. For site inspection, AutoRepo uses scene information collected by unmanned vehicles as input to a multimodal LLM to generate construction inspection reports [62].

6.4. Operation, maintenance, energy management, and end-of-life

The post-construction applications reviewed here cover building energy management, defect inspection, maintenance support, digital-twin information requirements, and waste recognition. For building energy modeling, EPlus-LLM uses a fine-tuned T5 model to convert natural-language building descriptions into EnergyPlus input files and runs the resulting models through the simulation engine [48]. Related work explores the use of ChatGPT for simulation input generation, output analysis, and error diagnosis [133]. Semantic similarity and fine-tuned language models have also been used to match BIM spaces and constructions with relevant database entries, supplying missing semantic and thermal information for energy simulation [49]. For building operation, GPT has been used to examine patterns extracted from chiller-plant data and identify energy waste [50]. A multi-agent framework combines building information processing, performance diagnosis, and retrofit recommendations using unstructured building data [16]. LLMs have also been evaluated on their knowledge and reasoning in heating, ventilation, and air conditioning (HVAC) through professional examination questions [134]. For maintenance, language models help extract and retrieve information about asset defects. BERT-based named entity recognition identifies damage-related entities in bridge inspection reports [135], while knowledge-graph question answering combines BERT representations with semantic matching to answer bridge inspection questions [136]. For housing defects, a hybrid approach uses a closed-source LLM to generate training data and a locally deployed open-source model to answer defect-related questions [47]. Multimodal LLMs have been tested for steel-bridge crack classification at both image and patch levels [56]. In robotic inspection, object detection and image captioning are combined with GPT-based interpretation of voice and text instructions to guide indoor defect inspections and prepare inspection reports [137]. Maintenance support also includes interactive task guidance and the preparation of digital-twin requirements. One system combines OCR, ChatGPT, and Augmented Reality (AR) to assist users with maintenance procedures [138]. DT-GPT helps owners define model requirements, required functions, and nongeometric information for operation and maintenance digital twins [93]. For end-of-life management, the work discussed here focuses on waste material recognition. ConCLIP combines images and text to classify construction waste materials for automated sorting [95].

7. Cross-perspective distribution of LFM-enabled construction studies

Figure 15 connects 136 studies across six categories: data type and modality, data source, data processing, model type, domain adaptation, and application sub-tasks. Studies could receive multiple labels within each dimension. It shows how different data foundations influence subsequent algorithm selection and construction applications. Two research pathways can be observed.

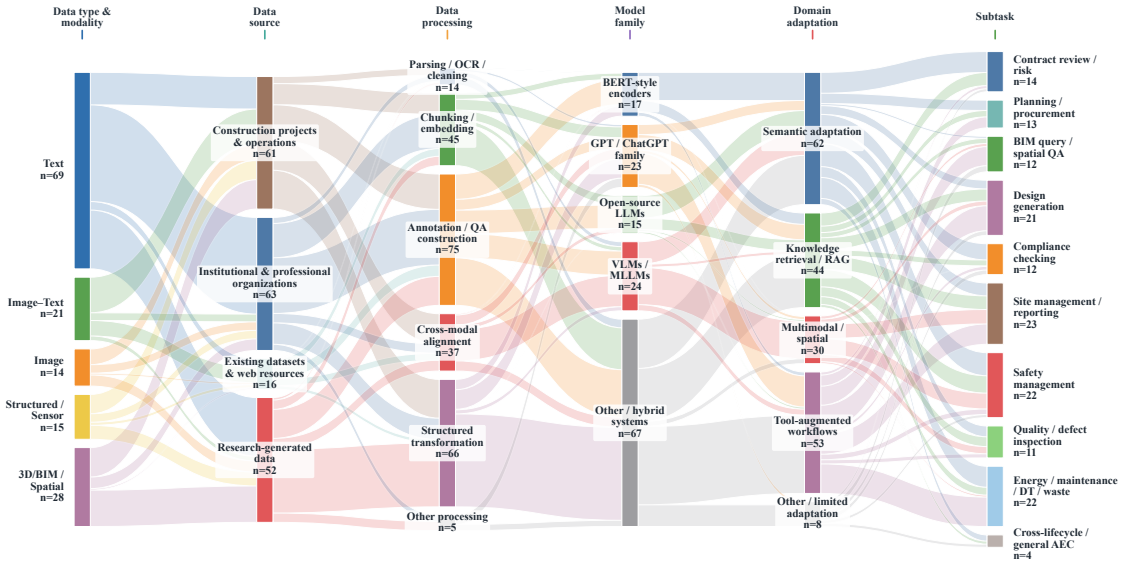

Figure 15. Sankey distribution of the 136 reviewed studies across data-algorithm-application.

The first pathway is based on text construction information. Regulations, contracts, technical documents, project reports, and other management materials constitute the main sources of construction information. These materials typically exist in unstructured or semi-structured formats, so the original documents must be converted into retrieval and processing information units through parsing, segmenting, or embedding. After this process, the text data can be integrated more directly with existing language models. BERT-class models primarily handle classification and information extraction, while GPT-class models further support question-answering, generation, and reasoning. Fine-tuning, prompts, and RAG are used to incorporate construction terminology, task requirements, and external domain expertise. Therefore, the downstream applications primarily focus on tasks such as contract review, construction management Q&A, safety standard retrieval, and quality document analysis. For classification and information-extraction tasks with annotated ground truth, results can be compared using accuracy and F1 scores. Its relative maturity stems not only from the large volume of text data but from the relatively direct technical alignment between text data, language models, and these application tasks.

The second pathway relies on multimodal data. These data include image-text data, construction site images, BIM/IFC models, point clouds, videos, and operational logs. These data directly describe real-world engineering objects and their spatial and dynamic states. This type of data requires establishing object correspondences across different modalities and systems while maintaining consistency in spatial location and temporal state. VLMs, MLLMs, and methods combined with structured queries and engineering tools play a more significant role in this pathway. Their applications have also expanded to tasks such as BIM queries, spatial question-answering, design generation, automated compliance checks,

construction progress assessment, and quality inspections. In these applications, a common feature is that model outputs must be further mapped to specific components, spatial locations, project statuses, and engineering rules. Their validation is often based on single cases, prototype testing, or functional experiments. Therefore, this pathway places higher requirements on data alignment, spatial semantics, state preservation, and result verification. It also remains relatively small in scale at present.

Research on LFMs in the construction field is beginning to shift from information-driven tasks to workflows with a greater emphasis on engineering. From a cross-layer perspective, developments in data, algorithms, and applications are not fully aligned. At the data level, current research still relies primarily on text and project-specific information, while multimodal, BIM/IFC, and sensor data remain relatively limited. Meanwhile, at the algorithm level, research has expanded from basic semantic alignment and knowledge retrieval to include multimodal reasoning, tool execution, and multi-agent collaboration. At the application level, existing research remains heavily concentrated on information extraction, retrieval, question-answering, report generation, and decision support, while the number of studies in engineering model modification, software execution, and continuous workflow collaboration remains relatively limited. The primary challenge currently lies in the interconnections among these three layers: whether data can provide sufficient engineering semantics; whether algorithms can transfer this information and generate reliable inferences; and whether applications can transform model outputs into verifiable engineering results. Consequently, the current bottleneck in construction LFM research is a lack of synchronized development among the data, algorithms, and practical applications.

8. Key challenges and future directions

8.1. Key challenges of LFM-enabled construction

Current limitations in construction LFMs can be grouped into three areas: data foundations, algorithmic adaptation, and practical deployment, as summarized in Figure 16. Table 3 summarizes challenges identified in the reviewed studies. For the application-limitation table, each study was assigned to one primary application subtask according to its main research objective and validation task.

8.1.1. Data-level challenges

First, reusable real-world engineering data remains insufficient [31,32,34,35,49]. A significant portion of this data comes from specific engineering projects and is restricted by commercial confidentiality, data ownership, and privacy requirements. Some studies do not release raw data, instead providing only samples, annotation results, or processed data [49,69]. This not only hinders the reproducibility of research but also limits independent validation across different projects and models.

Second, standardized representation and exchange methods between different data sources are still limited. Building data is fragmented across various systems and uses differing data formats, object structures, and encoding rules. Existing research typically focuses on task-specific design of data cleaning, transformation, and organization workflows. Consequently, data prepared for one task often cannot be directly transferred to another model or software environment [9,30,48,49,110].

Table 3. Distribution of the identified LFM challenges across 14 construction application sub-tasks.

Challenge	Pre-construction planning			Design and engineering			On-site construction execution and management				Post-construction operation and maintenance			
	Contract management	Planning / scheme inspection	Procurement consulting	BIM information retrieval	Design generation	Design inspection / compliance	Materials & equipment	Progress management	Safety management	Quality inspection	Building energy management	Defect inspection	Maintenance / DT assistance	CDW recognition
Data perspective														
Restricted accessibility and incomplete dataset releases	-	-	-	-	-	[128]	[139]	[26]	-	-	[16]	[47]	[93]	-
Data standardization and interoperability	-	-	-	[24]	-	-	-	-	[32]	-	[48,49]	-	[93]	-
Insufficient semantic granularity	[8]	[122]	-	[30]	-	[102]	-	[11,107]	[22]	-	[49]	-	-	-
Scarcity of multimodal, spatial, and temporal data	-	-	-	-	-	-	-	[25]	-	-	-	-	-	[95]
Limited benchmark readiness and evaluation consistency	-	-	-	-	[100]	[99]	-	[25,26,78, 107]	[22,32,59, 92]	[62]	-	[56]	-	[95]
Algorithm perspective														
Insufficient domain adaptation and limited transferability	[78]	[120,122]	[37]	-	[70,100]	[99,102]	[79]	[26,31,78]	[29,32,57,76, 92]	-	[49,134]	[47]	-	-
Long-context and hierarchical reasoning constraints	[8,119]	-	-	[24,30]	-	[102,128]	-	[11,46]	[22,57]	-	[134]	-	-	-
Multimodal, spatial, and temporal reasoning limitations	-	-	-	[30]	[9,10]	-	[140]	[25]	[29,57,59]	[62]	-	[23,56]	[138]	[95]
Limited tool execution and workflow reliability	-	-	-	[24]	[9,10,51,55, 110]	-	[79,129]	-	-	[62]	[48]	[23]	[138]	-
Incomplete verification of engineering outputs	-	-	-	-	[115]	[97]	-	-	[116]	-	[71]	-	-	-
Application perspective														
Limited real-world validation and daily workflow use	-	[120,122]	[37]	-	[95,110]	-	[79]	[46]	[59,76]	-	[134]	[23]	[93]	-
Dependence on professional review and unclear decision roles	-	[120]	-	-	[10]	-	[139,140]	-	[92,130, 131]	-	-	-	[138]	-
Confidentiality and access restrictions in deployment	[34,35]	-	[37]	-	-	-	-	[25,31,46]	[58]	-	-	[47]	[93]	-

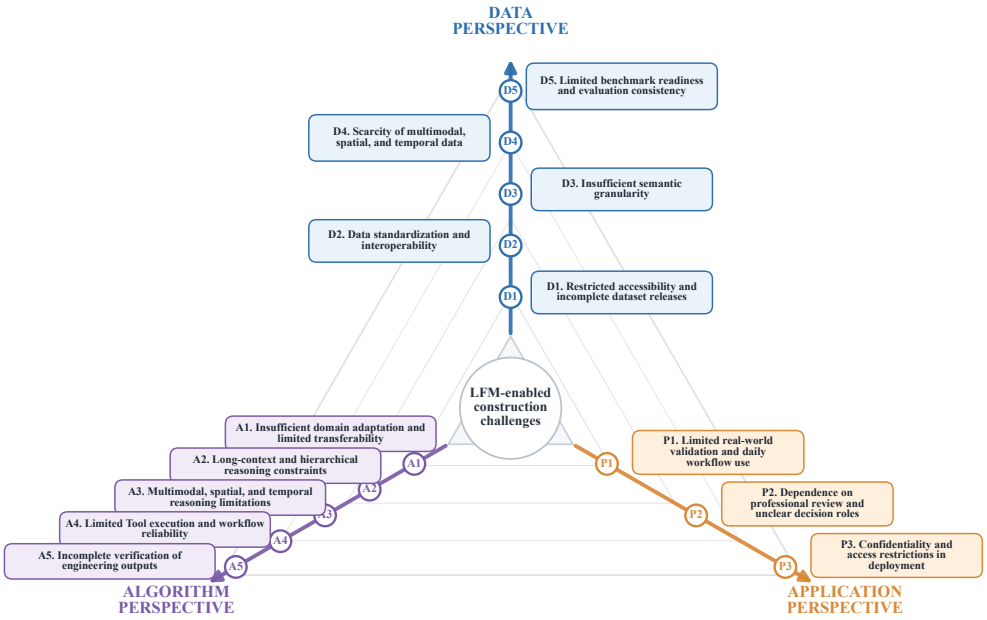

Figure 16. Three-perspective challenge map of LFM-enabled construction.

Third, available data still do not adequately capture engineering evidence and domain-specific relationships. Many datasets still use labels, text snippets, image descriptions, or question-answer pairs as their basic units, but often fail to fully preserve the connection between judgment outcomes and the engineering evidence [7,8,22,31]. For example, safety analyses should not only identify risks, but must also connect to risk locations, causes, relevant standards, and corrective measures. If these relationships are not explicitly preserved, it is difficult to verify whether model-generated answers are truly based on engineering evidence, even if they appear reasonable [93,102,119,136].

Fourth, multimodal and spatiotemporal data remains relatively limited. Current research makes relatively limited use of BIM/IFC, images, point clouds, videos, sensor data, and schedule data. Such data typically originate from specific projects, with high costs of collection and annotation. Furthermore, different data sources lack unified object identifiers, spatial relationships, and temporal correspondences. Consequently, it is difficult to use existing data to form a reusable and scalable multimodal engineering data foundation [9,26,59,62,116].

Fifth, benchmarks lack sufficient maturity. Although specialized benchmarks for building standards, safety management, and expert Q&A have emerged in recent years, most remain limited to specific tasks, projects, or regulatory environments. Differences among studies in sample selection, task definition, data splitting, and evaluation methods make direct comparison difficult. This issue is even more obvious for reasoning-based and tool-execution LFMs [26,48,49,59,62]. Evaluation should not be based solely on whether the final answer is correct; it must also assess whether the reasoning is well-founded, whether the spatial relationships are accurate, and whether the task has been fully completed.

8.1.2. Algorithm-level challenges

First, domain adaptation remains highly dependent on specific tasks and scenarios. Fine-tuning, prompt engineering, and RAG can enhance LFMs' ability to adapt to construction terminology, regulatory knowledge, and specific tasks [102,119,128,141]. These improvements are often based on specific data,

regulatory frameworks, or project environments [46,73,76,99]. The applicable conditions vary across different regions, projects, and tasks.

Second, hierarchical relationships and cross-document dependencies within complex engineering knowledge remain difficult to process reliably. Construction contracts, specifications, and technical documents typically contain hierarchical clauses, cross-references, and applicability conditions [7,8,11,22]. While long-context models and RAG can provide more relevant information, they may still overlook key conditions or incorrectly combine rules from different hierarchical levels.

, multimodal and spatiotemporal reasoning capabilities remain limited. Existing VLMs and MLLMs are still susceptible to errors in precise spatial relationships, BIM semantics, and changes in object states, particularly in scenarios across different data sources and phases [21,54,58,96]. Inconsistencies in object identity, location, and state directly impact progress assessments, safety inspections, and quality analyses.

Fourth, tool execution introduces new reliability challenges. When LFMs are used for code generation, API calls, BIM modifications, or other software operations, errors may arise from tool selection, parameter configuration, or object references [9,51,101]. In multi-step workflows, errors generated in a previous step may propagate to subsequent tools or agents [23,79,129,139,140]. Therefore, the successful execution of a program does not necessarily mean that the final engineering task has been completed correctly [23,79,110].

Fifth, existing evaluation methods are still insufficient to verify the correctness of LFM outputs. Retrieval accuracy, language quality, or success rates typically reflect only a single aspect of the workflow and cannot fully determine whether the final output meets engineering requirements [71,97,116]. For example, a correct textual description does not mean that it corresponds to the correct image region, BIM object, or specification conditions. Therefore, evaluation must also take into account evidence support, spatial correspondence, compliance with rules, and whether the task has been truly completed [38].

8.1.3. Application-level challenges

First, long-term stability in real-world projects has not yet been fully validated. Existing research remains primarily focused on prototypes, single-case studies, and controlled experiments [7,46,59,120,125]. It is difficult to demonstrate that LFMs can operate stably in dynamic, multi-stakeholder real-world construction environments [9,11,51,75]. Even though some systems have been integrated with BIM or engineering software, most remain at the functional validation stage and have not been fully incorporated into daily project workflows.

Second, the respective decision-making roles of humans and LFMs and the allocation of responsibility remain unclear. LFMs have already begun to participate in professional judgments such as design, safety, compliance, and contracts. However, there is still no clear definition of which tasks can be performed automatically and which results must be verified by professionals [7,37,102,119,122]. When model outputs conflict with professional judgment or lead to errors, the ultimate decision-making authority and assignment of responsibility also remain unclear [29,57,141].

Third, restrictions on the use of project data limit actual deployment. Real-world engineering applications typically involve sensitive data. However, being able to access data does not mean it can be directly submitted to LFMs for processing [34,35,37]. Actual deployment still requires clearly defining

data access permissions, the processing environment, and the scope of data that different users and agents are permitted to use [25,46,50,58,93].

8.2. Future directions

The challenges across the three dimensions reveal that the future development of construction LFMs does not depend on improvements in the capabilities of a single model. Instead, it requires the establishment of a comprehensive system that connects data foundations, intelligent reasoning, and engineering deployment. While current algorithmic research has advanced rapidly, high-quality construction data, engineering knowledge representation, and real-world workflow validation have not kept pace. As shown in Figure 17, future development of construction LFMs should address the coupled challenges identified from data, algorithm, and application perspectives.

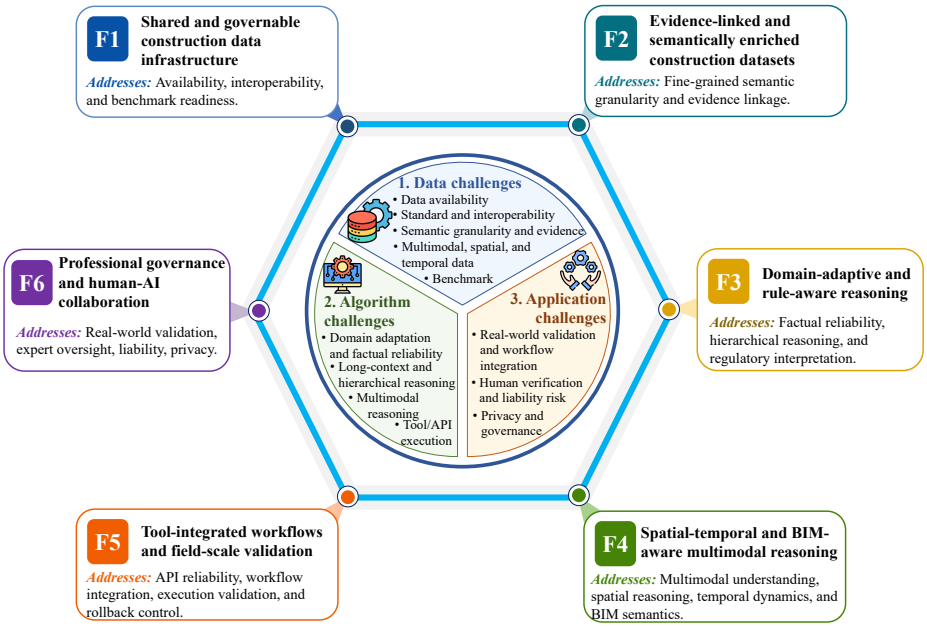

Figure 17. Challenge-oriented future directions for LFM-enabled construction.

To address the challenges of difficult data acquisition, inconsistent formats, and insufficient cross-system interoperability, it will be necessary in the future to establish unified standards for data description, metadata, and interfaces. A layered access and privacy protection mechanism should also be developed. The focus should not merely be on expanding the scale of data openness, but on ensuring that data can be continuously reused under clearly defined permissions. Since a large volume of engineering data is held by project owners, contractors, design firms, and public institutions, this effort requires both technical standards and data governance mechanisms.

To address the issues of insufficient semantic granularity in existing data and limited comparability of benchmarks, future data resources should be developed into evidence-linked and verifiable engineering benchmarks. In addition to inputs and labels, they should also preserve evidence such as specification clauses, spatial objects, process states, and decision rationale, and establish unified task definitions, data partitions, and evaluation rules.

To address the challenges of domain adaptation, specifically the difficulty in handling task-specific dependencies, rule hierarchies, and cross-document relationships, future research needs to shift further from simply incorporating domain knowledge to correctly applying engineering knowledge. Models should be capable of handling the scope of rule application, exception conditions, responsibility relationships, and cross-document dependencies. At the same time, domain knowledge should be continuously updated in accordance with changes in regulations, project requirements, and industry standards.

To address the shortcomings in the processing of multimodal, spatial, and temporal information, future research needs to further integrate images, BIM/IFC data, point clouds, video, sensor data, and schedule information. The focus should be on ensuring consistency in object identity, spatial location, and state changes across different data sources. This will enable models to perform cross-modal and spatiotemporal reasoning across project stages, rather than performing a single instance of recognition or description.

To address issues such as tool invocation errors and error propagation during multi-step execution, future LFM workflows will need to incorporate intermediate result checks, parameter validation, execution logs, error diagnosis, and rollback mechanisms. Evaluation should also be expanded from simply determining whether an API call succeeded or whether code executed to assessing whether the final engineering task was completed correctly. The validation environment should gradually expand from laboratory prototypes and sandbox testing to real software environments and supervised project applications.

To address issues related to professional review, allocation of responsibility, and data access permissions in real-world deployments, future approaches should determine the level of human involvement based on task risk. Low-risk information retrieval can be subject to lighter human review, while high-risk tasks should retain professional approval and audit mechanisms. At the same time, model inputs, evidence sources, execution processes, and final human decisions must be recorded. Industry associations, standards bodies, and regulatory authorities should jointly clarify the boundaries of data usage, model application, and accountability.

9. Conclusion

This paper systematically reviews 136 studies on LFM-enabled construction and 86 related construction datasets. It analyzes the current state of development from three interrelated perspectives: data, algorithms, and applications. The review finds that research on construction LFMs is expanding rapidly. However, the development of these three aspects is not synchronized. The central limitation is the lack of a stable connection between the data foundation, the reasoning ability of the model, and actual engineering deployment.

The future development of construction LFM must shift from optimizing individual models toward developing integrated engineering AI systems. First, it is necessary to establish a more open and governable data foundation so that project data can support model development and evaluation under secure and standardized conditions. Second, it is necessary to develop reasoning methods capable of understanding engineering rules, spatial states, and dynamic processes, and to integrate LFMs more reliably with BIM, simulation tools, and project management processes. Third, deploying these systems at scale requires the joint participation of professionals, industry organizations, software vendors, and regulatory agencies.

This study also has limitations. First, the scope of the review is limited by the databases, the search period, and the scope of English-language literature. Some relevant studies may not have been included.

Second, due to significant differences in existing studies, some classifications and comparisons inevitably involve a degree of subjective judgment. As research on construction LFMs continues to develop, future reviews will incorporate more project case studies and actual deployment results.

Data availability statement

The data or datasets generated or analyzed in this study (the coded literature database supporting the findings of this review) are available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used ChatGPT solely for grammar correction and language polishing. The authors reviewed and revised all AI-assisted outputs and take full responsibility for the integrity, originality, and accuracy of the manuscript.

Acknowledgments

This work is supported by the RGC Research Impact Fund (R7036-22), RGC Theme-based Research Scheme (T32-707-22-N), and RGC GRF project (17202124). This research is also supported by the Digital Roads Prosperity Partnership (DR). DR is supported by the Engineering and Physical Sciences Research Council (EPSRC) grant number [EP/V056441/1], Costain, National Highways, the University of Cambridge, Department for Transport and Didimi. The authors gratefully acknowledge the collaboration of all academic and industrial project partners. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the institutes mentioned above.

Authors' contribution

Conceptualization, Y.L., X.G., Y.H., S.B., I.B., M.J.S., Z.Z., and R.Y.Z.; methodology, Y.L., X.G., Y.H.; software, Y.L. and X.G.; validation, Y.H., S.B., I.B., M.J.S., Z.Z., and R.Y.Z.; formal analysis, Y.L., X.G., Y.H., and S.B.; investigation, Y.L., X.G., and Y.H.; resources, I.B., M.J.S., Z.Z., and R.Y.Z.; writing—original draft preparation, Y.L. and X.G.; writing—review and editing, Y.H., S.B., I.B., M.J.S., Z.Z., and R.Y.Z.; visualization, Y.L., X.G., and Y.H.; project administration, R.Y.Z.; funding acquisition, R.Y.Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Dataset-level summary of 86 construction datasets

Table A1 provides a dataset-level summary of the 86 identified construction datasets, covering their modality, data type, source, size, accessibility, and primary LFM use.

Table A1. 86 construction datasets used or reported in LFM-related studies.

No.	Dataset/resource	Modality	Data type	Reference	Source	Size	Accessibility	Primary LFM use
1	Alberta Construction Image Dataset (ACID)	Image–Text	Site/activity image–text data	[69]	Public construction image dataset enriched with two captions per image	4000 images + 8000 generated text captions	Public	Model adaptation/fine-tuning
2	ACTV construction activity image dataset	Image–Text	Site/activity image–text data	[69]	Public construction image dataset enriched with two captions per image	964 images + 1928 generated text captions	Public	Model adaptation/fine-tuning
3	AECBench	Text	Codes/regulations/compliance text	[38]	Building codes, proprietary internal documents, professional/licensing exam resources	4800 questions; 23 tasks	Mixed	Evaluation/benchmark
4	AEC close-domain Chinese corpus	Text	Domain corpus/QA/terminology	[20]	Close-domain Chinese text corpus	26,727 lines; 12,899,562 Chinese characters	Public	Domain pretraining
5	AEC in-domain Chinese corpus	Text	Codes/regulations/compliance text	[20]	Chinese AEC regulatory/domain websites	126,433 lines; 10,895,634 Chinese characters	Public	Domain pretraining
6	Arch-Eval	Text	AEC domain benchmark/QA	[39]	Official Chinese registered architect/planner exams, mock questions and university papers	10,440 single-choice questions	Mixed	Evaluation/benchmark
7	BBC News corpus	Text	Construction news/general text corpus	[52]	BBC News articles	954,749 articles	Public	Embedding/representation
8	GRPO reinforcement-learning BEM dataset	Structured/Sensor	Energy/O&M structured or sensor data	[71]	Additional natural-language/IDF building examples	300 GRPO examples	Synthetic	Model adaptation/fine-tuning
9	Natural-language–EnergyPlus IDF SFT dataset	Structured/Sensor	Energy/O&M structured or sensor data	[71]	Synthetic paired natural-language building descriptions and EnergyPlus IDF files	2100 SFT pairs	Synthetic	Model adaptation/fine-tuning
10	BEM generation held-out test set	Structured/Sensor	Energy/O&M structured or sensor data	[71]	Held-out natural-language/IDF building examples	300 test examples	Synthetic	Evaluation/benchmark
11	BEQA building-code exam benchmark	Text	Codes/regulations/compliance text	[94]	National Engineer Architecture certification exam questions	ND	Regulatory	Evaluation/benchmark
12	BifrostRAG multi-hop OSHA question benchmark	Text	Safety guidelines/accident knowledge	[22]	Multi-hop questions over OSHA 1926 Subparts L/M/P/X	93 questions: 30 L + 27 M + 17 P + 19 X; 10 from authoritative OSHA resources + 83 created with a CSP/OSHA-authorized trainer	Project-specific	Evaluation/benchmark
13	BINLQ BIM natural-language query dataset	Text	BIM/design/software interaction text	[125]	Public BIM natural-language query dataset	2065 natural-language queries	Public	Prompt/in-context input
14	Highway-bridge BIM compliance evaluation rules	3D/BIM/Spatial	BIM/IFC compliance data	[142]	Structured bridge rules applied to an IFC highway-bridge model	112 evaluation rules: 65 TP, 8 TN, 12 FP, 27 FN	Project-specific	Evaluation/benchmark

Table A1. *Cont.*

No.	Dataset/resource	Modality	Data type	Reference	Source	Size	Accessibility	Primary LFM use
15	Chinese bridge inspection KGQA dataset	Text	Quality/defect/inspection text	[136]	Bridge inspection knowledge graph + associated Chinese QA pairs	ND	Project-specific	Model adaptation/fine-tuning
16	Chinese bridge damage/defect inspection-report corpus	Text	Quality/defect/inspection text	[135]	Bridge inspection reports from multiple areas of China	>1300 reports; 330,414 characters; 13,116 sentences	Regulatory	Model adaptation/fine-tuning
17	Building regulatory text + BIM compliance case	3D/BIM/Spatial	BIM/IFC compliance data	[128]	Construction design regulations + BIM model represented with lightweight domain ontology	ND	Regulatory	Model adaptation/fine-tuning
18	Real-world rule-interpretation rule–script dataset	Text	Codes/regulations/compliance text	[99]	Real-world automated code-compliance rule interpretation project	ND	Regulatory	Model adaptation/fine-tuning
19	Hydraulic-construction accident instruction dataset	Text	Safety guidelines/accident knowledge	[103]	Hydraulic engineering construction accident investigation records transformed into specialized instructions	ND	Regulatory	Model adaptation/fine-tuning
20	CECA embodied-carbon training data	3D/BIM/Spatial	BIM/material embodied-carbon data	[143]	Typical building-project material/equipment information + embodied-carbon calculation knowledge	High-quality training dataset; exact size ND	Regulatory	Model adaptation/fine-tuning
21	CEM-QA case-based benchmark	Text	Construction management/domain corpus	[78]	Case-based construction-engineering-management questions	100 case-based questions	Project-specific	Evaluation/benchmark
22	CEM-QA multiple-choice benchmark	Text	Construction management/domain corpus	[78]	Professional examination/domain questions	5050 MCQs	Project-specific	Evaluation/benchmark
23	Semi-automatically labeled construction safety image-text data	Image–Text	Construction safety image–text/VQA	[28]	Construction-site video frames/images labeled through in-context learning	ND	Project-specific	Model adaptation/fine-tuning
24	Construction Hazard Dataset (CHD)	Image–Text	Construction safety image–text/VQA	[59]	Historical inspection records + public construction hazard datasets	1308 real hazard images	Mixed	Model adaptation/fine-tuning
25	CMS academic-publication corpora	Text	Construction management/domain corpus	[75]	Academic papers, conference papers, articles, whitepapers and selected books on construction management systems	~5.7 million words; four corpus variants	Public	Domain pretraining
26	CMS NER fine-tuning dataset	Text	Construction management/domain corpus	[75]	Automatic construction-control data	ND	Project-specific	Model adaptation/fine-tuning
27	CMS text-classification fine-tuning dataset	Text	Construction management/domain corpus	[75]	Infrastructure condition prediction data	ND	Project-specific	Model adaptation/fine-tuning
28	ConSTRAG safety QA evaluation dataset	Text	Safety guidelines/accident knowledge	[144]	Safety questions across the training domain	11,020 questions	Project-specific	Evaluation/benchmark
29	ConstructionSite 10 k	Image–Text	Site/activity image–text data	[116]	10,013 selected construction-site images	10,013 images; 70% train/30% test	Public	Model adaptation/fine-tuning

Table A1. *Cont.*

No.	Dataset/resource	Modality	Data type	Reference	Source	Size	Accessibility	Primary LFM use
30	Chinese construction case-law corpus	Text	Contracts/legal/procurement text	[43]	Construction judicial cases + relevant legal articles	81,472 construction cases; 303 acts	Public	Model adaptation/fine-tuning
31	Construction Visual Question Answering dataset	Image–Text	Site/activity image–text data	[60]	On-site images + SODA + ACID + other public/web sources	> 11,000 image-question pairs	Mixed	Model adaptation/fine-tuning
32	Contract risk-handling action dataset	Text	Contracts/legal/procurement text	[145]	Construction contract clauses/statements	ND	Regulatory	Model adaptation/fine-tuning
33	International construction contract benchmark	Text	Contracts/legal/procurement text	[42]	International construction contract knowledge questions	1131 questions	Project-specific	Evaluation/benchmark
34	Construction contract instruction-tuning QA dataset	Text	Contracts/legal/procurement text	[40]	FIDIC Yellow Book + interpretive materials	5000 instruction QA pairs	Project-specific	Model adaptation/fine-tuning
35	Construction specification extractive-QA set	Text	Domain corpus/QA/terminology	[146]	Construction specifications	855 extractive QA pairs	Regulatory	Model adaptation/fine-tuning
36	Construction safety-guideline open-ended QA set	Text	Safety guidelines/accident knowledge	[146]	Construction safety guidelines	694 open-ended QA pairs	Public	Model adaptation/fine-tuning
37	Work-at-height construction image corpus	Image	Construction safety/hazard images	[63]	Authentic construction work-at-height images including fall-hazard scenarios	Large-scale corpus; exact count ND	Project-specific	Model adaptation/fine-tuning
38	Standardized construction-safety terminology database	Text	Safety guidelines/accident knowledge	[63]	Standardized safety terminology/protocol vocabulary	ND	Regulatory	Model adaptation/fine-tuning
39	DfMA conversation-scenario evaluation set	Text	BIM/design/software interaction text	[147]	Conversation scenarios covering BIM-based DfMA interactions	692 conversation scenarios	Project-specific	Evaluation/benchmark
40	Natural-language description–IDF sentence pairs	Structured/Sensor	Energy/O&M structured or sensor data	[48]	Generated descriptions paired with EnergyPlus IDF sentences	70,000 description–IDF sentence pairs	Synthetic	Model adaptation/fine-tuning
41	EnergyPlus parameter–IDF scenario pairs	Structured/Sensor	Energy/O&M structured or sensor data	[48]	EnergyPlus IDF files generated from sampled parameters	70,000 parameter–IDF scenario pairs	Synthetic	Model adaptation/fine-tuning
42	Ergonomic problem–solution image-caption dataset	Image–Text	Construction safety image–text/VQA	[61]	YouTube and other web construction-work images/videos	1244 + 105 samples	Public	Model adaptation/fine-tuning
43	Ergonomic image-caption training/augmentation corpus	Image–Text	Construction safety image–text/VQA	[132]	Construction-site ergonomic images paired with ergonomic-problem and solution descriptions; data augmentation used	Total training-corpus size ND; image-caption supervision with data augmentation	Project-specific	Model adaptation/fine-tuning
44	Farsi construction contract risk statements	Text	Contracts/legal/procurement text	[118]	Construction contracts written in Farsi	ND	Regulatory	Model adaptation/fine-tuning
45	FIDIC 2017 Silver Book sentence corpus	Text	Contracts/legal/procurement text	[35]	2017 FIDIC Silver Book	1426 sentences	Regulatory	Model adaptation/fine-tuning

Table A1. *Cont.*

No.	Dataset/resource	Modality	Data type	Reference	Source	Size	Accessibility	Primary LFM use
46	Building-code clauses for LLM-FuncMapper	Text	Codes/regulations/compliance text	[104]	GB50016-2014 chapters 3–5 + atomic compliance-check functions	451 textual clauses -> 679 split clauses; 1,048 tabular clauses; 66 atomic functions	Project-specific	Other deployment data
47	Real construction project contracts	Text	Contracts/legal/procurement text	[34]	Real construction project contracts	35 project contracts; combined standard+real corpus = 20,612 sentences	Project-specific	Model adaptation/fine-tuning
48	Standard construction contracts	Text	Contracts/legal/procurement text	[34]	Standard contracts of global institutions	17 standard contracts; combined standard+real corpus = 20,612 sentences	Regulatory	Model adaptation/fine-tuning
49	GeoLLM single-pile design dataset	Structured/Sensor	Geotechnical structured design data	[70]	Synthetic/structured single-pile geotechnical design scenarios covering pile type, cross-section and design conditions	54 scenarios	Project-specific	Evaluation/benchmark
50	HVAC professional knowledge/skill evaluation question set	Text	Energy/HVAC knowledge text	[134]	HVAC-domain knowledge and skill questions/tasks	ND	Project-specific	Evaluation/benchmark
51	IFC semantic-interpretation benchmark models	3D/BIM/Spatial	BIM/IFC semantic/spatial data	[66]	IFC-based BIM models restructured into JSON + natural-language query/modification tasks	ND	Project-specific	Evaluation/benchmark
52	Construction defect case corpus	Text	Quality/defect/inspection text	[47]	Defect cases from a South Korean construction company	247 defect cases	Project-specific	Model adaptation/fine-tuning
53	KOSHA construction-safety corpus	Text	Safety guidelines/accident knowledge	[32]	86 Korean Occupational Safety and Health Agency documents	86 documents; 5502 data entries	Public	Model adaptation/fine-tuning
54	LIQA legal-interpretation benchmark	Text	Contracts/legal/procurement text	[94]	Building-code/regulation interpretation questions	171 open-ended questions	Project-specific	Evaluation/benchmark
55	Construction material-matching enterprise dataset	Structured/Sensor	Material/procurement structured data	[72]	Enterprise bill-of-material records	3773 raw records -> 2912 final records after removing 472 incomplete, 255 duplicates, 134 ambiguous	Project-specific	Model adaptation/fine-tuning
56	MILP-generated optimal building-energy schedules	Structured/Sensor	Energy/O&M structured or sensor data	[117]	Synthetic optimal schedules produced by mixed-integer linear programming	Total N ND	Synthetic	Model adaptation/fine-tuning
57	MOCS-based safety VQA set	Image–Text	Construction safety image–text/VQA	[58]	Multimodal Objects in Construction Sites (MOCS)	578 images with QA pairs	Public	Model adaptation/fine-tuning

Table A1. *Cont.*

No.	Dataset/resource	Modality	Data type	Reference	Source	Size	Accessibility	Primary LFM use
58	Unreinforced-footing multimodal design cases	Image–Text	Multimodal engineering/design cases	[10]	Foundation-engineering cases covering five footing styles and design/verification tasks	20 scenarios instantiated as 60 multimodal sets (text, image, text-image)	Project-specific	Evaluation/benchmark
59	Unreinforced-footing textual robustness cases	Text	Engineering design/underground text	[10]	Generated/curated textual footing-design cases	100 textual cases	Project-specific	Evaluation/benchmark
60	NBCC-derived fine-tuning QA dataset	Text	Codes/regulations/compliance text	[3]	Question-answer pairs derived from NBCC	ND	Project-specific	Model adaptation/fine-tuning
61	NREL standardized space-type database	Structured/Sensor	Energy/O&M structured or sensor data	[49]	NREL standardized space-type database	224 space types across 23 building types	Public	Model adaptation/fine-tuning
62	O&M hybrid fine-tuning/external knowledge base	Structured/Sensor	Energy/O&M structured or sensor data	[65]	Building O&M documents, knowledge and monitoring-derived information	ND	Project-specific	Model adaptation/fine-tuning
63	OpenStudio construction/material database	Structured/Sensor	Energy/O&M structured or sensor data	[49]	OpenStudio/EnergyPlus construction and material database	201 construction sets with material entries	Public	Model adaptation/fine-tuning
64	Construction procurement-law question bank	Text	Contracts/legal/procurement text	[37]	Government e-Procurement System procurement-law question bank	3717 True/False + 294 QA	Public	Model adaptation/fine-tuning
65	Architectural–structural design pairs	Image	Architectural/structural drawings	[55]	10 architectural design and research institutes	250 paired architectural–structural designs	Project-specific	Model adaptation/fine-tuning
66	Qwen-AEC ARC named-entity dataset	Text	Codes/regulations/compliance text	[44]	Building-code automated rule-compliance NER data	ND	Regulatory	Model adaptation/fine-tuning
67	Qwen-AEC building-code text-classification dataset	Text	Codes/regulations/compliance text	[44]	Building-code clauses	ND	Regulatory	Model adaptation/fine-tuning
68	ResStock stratified LLM evaluation sample	Structured/Sensor	Energy/O&M structured or sensor data	[148]	Random stratified sample across 49 states and 3 housing types	400 homes; 389 parameters each; 28 selected for GPT-4o narratives	Project-specific	Evaluation/benchmark
69	Revit C# API BIM-query benchmark	Text	BIM/design/software interaction text	[24]	Revit C# API query/function space	80 queries of varying complexity	Project-specific	Evaluation/benchmark
70	SAFE construction safety image dataset	Image–Text	Construction safety image–text/VQA	[69]	Public construction image dataset enriched with two captions per image	1762 images + 3524 generated text captions	Public	Model adaptation/fine-tuning
71	Site Object Detection Dataset (SODA)	Image–Text	Site/activity image–text data	[5469]	Public construction image dataset enriched with two captions per image	Exact cropped count ND | 1089 images + 2178 generated text captions	Public	Model adaptation/fine-tuning
72	Expert-validated contract-query evaluation set	Text	Contracts/legal/procurement text	[18]	Natural-language questions formulated from the hostel contract and validated by construction experts	30 draft questions -> 20 final questions validated by all 3 experts	Project-specific	Evaluation/benchmark

Table A1. *Cont.*

No.	Dataset/resource	Modality	Data type	Reference	Source	Size	Accessibility	Primary LFM use
73	Tell2Design layout data/case prompts	Text	BIM/design/software interaction text	[105]	Tell2Design natural-language spatial descriptions + office/multi-apartment layout cases	ND	Public	Prompt/in-context input
74	Temporary Objects on Construction Site (TOCS)	Image	Construction-site objects/equipment images	[54]	New construction-site temporary-object dataset	90 images in current manuscript; 22 object categories confirmed	Project-specific	Evaluation/benchmark
75	Tower-crane safety knowledge graph	Text	Safety guidelines/accident knowledge	[149]	Domain-specific tower-crane safety knowledge	100 nodes + 355 relations	Project-specific	Model adaptation/fine-tuning
76	Construction safety image-caption dataset	Image–Text	Construction safety image–text/VQA	[96]	Safety reports from multiple building construction sites; captions by three experienced safety inspectors	ND; covers 9 violation types	Project-specific	Model adaptation/fine-tuning
77	UndergrBench	Image–Text	Multimodal engineering/design cases	[64]	Engineering codes, geological reports, academic papers and underground disaster cases	10,888 Chinese multimodal samples; 9 tasks; 3 cognitive levels	Mixed	Evaluation/benchmark
78	UndergrData domain corpus	Text	Engineering design/underground text	[1]	Underground-engineering literature, textbooks, standards and geological reports	~500 million Chinese characters	Mixed	Domain pretraining
79	UndergrTest benchmark	Text	Engineering design/underground text	[1]	Expert-constructed underground-engineering questions	1000 questions: 600 multiple-choice + 400 open-ended	Project-specific	Evaluation/benchmark
80	Commercial-building master-schedule dataset	Structured/Sensor	Construction schedule/process data	[77]	10 commercial building project master schedules	10 projects; 35,998 activity sequence tuples	Project-specific	Model adaptation/fine-tuning
81	Vision–Language Construction and Solid Waste (VL-CSW) dataset	Image–Text	Construction waste image–text data	[95]	Construction and solid-waste material images paired with descriptions	ND	Project-specific	Model adaptation/fine-tuning
82	Construction-site dense-caption hazard image dataset	Image–Text	Construction safety image–text/VQA	[27]	Real-world construction-site images originally collected by a prior study	2000 site images	Public	Model adaptation/fine-tuning
83	Wuhan metro construction safety image-text pairs	Image–Text	Construction safety image–text/VQA	[57]	Large metro construction safety monitoring system in Wuhan	2619 image-text pairs	Project-specific	Model adaptation/fine-tuning
84	Construction structural-element NER dataset	Text	Domain corpus/QA/terminology	[73]	Semi-automatically generated from standards/textbooks/lexicons	33,500 structure-element named entities	Project-specific	Model adaptation/fine-tuning
85	Construction NER source corpus	Text	Domain corpus/QA/terminology	[73]	289 Chinese structural standards + 4 professional textbooks + 3 IME lexicons	296 source resources	Regulatory	Model adaptation/fine-tuning
86	Construction NER QA-style dataset	Text	Domain corpus/QA/terminology	[73]	Semi-automatically generated from construction sources	63,597 QA cases	Project-specific	Model adaptation/fine-tuning

Note: ND = not declared.

References

[1] Xu Z, Li T, Ren P, Zhang X, Li T. Framework and construction methodology of underground engineering domain knowledge large language model: UndergrGPT. *Smart Constr.* 2024, 1(2):0012.

[2] Baduge SK, Thilakarathna S, Perera JS, Arashpour M, Sharafi P, *et al.* Artificial intelligence and smart vision for building and construction 4.0: machine and deep learning methods and applications. *Autom. Constr.* 2022, 141:104440.

[3] Aqib M, Hamza M, Mei Q, Chui YH. Fine-tuning large language models and evaluating retrieval methods for improved question answering on building codes. *Smart Constr.* 2025, 2(3):0021.

[4] Omar T, Nehdi ML. Data acquisition technologies for construction progress tracking. *Autom. Constr.* 2016, 70:143–155.

[5] Paneru S, Jeelani I. Computer vision applications in construction: current state, opportunities & challenges. *Autom. Constr.* 2021, 132:103940.

[6] Zhao S, Liu S, Jiang Y, Zhao B, Lv Y, *et al.* Industrial foundation models (IFMs) for intelligent manufacturing: a systematic review. *J. Manuf. Syst.* 2025, 82:420–448.

[7] Wong S, Zheng C, Su X, Tang Y. Construction contract risk identification based on knowledge-augmented language models. *Comput. Ind.* 2024, 157–158:104082.

[8] Zheng C, Wong S, Su X, Tang Y, Nawaz A, *et al.* Automating construction contract review using knowledge graph-enhanced large language models. *Autom. Constr.* 2025, 175:106179.

[9] Du C, Esser S, Nousias S, Borrmann A. Text2BIM: generating building models using a large language model-based multiagent framework. *J. Comput. Civ. Eng.* 2026, 40(2):04025142.

[10] Xu H, Zhang N, Yin ZY, Njock PGA. Multimodal framework integrating multiple large language model agents for intelligent geotechnical design. *Autom. Constr.* 2025, 176:106257.

[11] Wu C, Ding W, Jin Q, Jiang J, Jiang R, *et al.* Retrieval augmented generation-driven information retrieval and question answering in construction management. *Adv. Eng. Inform.* 2025, 65:103158.

[12] Zhang L, Chen Z. Opportunities of applying large language models in building energy sector. *Renew. Sustain. Energy Rev.* 2025, 214:115558.

[13] Rane N. ChatGPT and similar generative artificial intelligence (AI) for building and construction industry: contribution, opportunities and challenges of large language models for Industry 4.0, Industry 5.0, and Society 5.0. *SSRN* 2023. SSRN:4603221.

[14] Yan X, Yang X, Jin N, Chen Y, Li J. A general AI agent framework for smart buildings based on large language models and ReAct strategy. *Smart Constr.* 2025, 2(1):0004.

[15] Kampelopoulos D, Tsanousa A, Vrochidis S, Kompatsiaris I. A review of LLMs and their applications in the architecture, engineering and construction industry. *Artif. Intell. Rev.* 2025, 58(8):250.

[16] Xiao T, Xu P. Exploring automated energy optimization with unstructured building data: a multi-agent based framework leveraging large language models. *Energy Build.* 2024, 322:114691.

[17] Ghimire P, Kim K, Acharya M. Opportunities and challenges of generative AI in construction industry: focusing on adoption of text-based models. *Buildings* 2024, 14(1):220.

[18] Taiwo R, Bello IT, Abdulai SF, Yussif AM, Salami BA, *et al.* Generative artificial intelligence in construction: a Delphi approach, framework, and case study. *Alex. Eng. J.* 2025, 116:672–698.

[19] Saka A, Taiwo R, Saka N, Salami BA, Ajayi S, *et al.* GPT models in construction industry: opportunities, limitations, and a use case validation. *Dev. Built Environ.* 2024, 17:100300.

[20] Zheng Z, Lu X, Chen K, Zhou Y, Lin J. Pretrained domain-specific language model for natural language processing tasks in the AEC domain. *Comput. Ind.* 2022, 142:103733.

[21] Chen C, Xiao B, Zhang Y, Zhu Z. Automatic vision-based calculation of excavator earthmoving productivity using zero-shot learning activity recognition. *Autom. Constr.* 2023, 146:104702.

[22] Zhang Y, Wang X, Hu M, Zhang Z. Bridging dual knowledge graphs for multi-hop question answering in construction safety. *Autom. Constr.* 2026, 183:106794.

[23] Liu J, Li H, Chai C, Chen K, Wang D. A LLM-informed multi-agent AI system for drone-based visual inspection for infrastructure. *Adv. Eng. Inform.* 2025, 68:103643.

[24] Guo P, Xue H, Ma J, Cheng J. Advancing BIM information retrieval with an LLM-based query-domain-specific language and library code function alignment system. *Autom. Constr.* 2025, 178:106374.

[25] Xiao B, Wang Y, Zhang Y, Chen C, Darko A. Automated daily report generation from construction videos using ChatGPT and computer vision. *Autom. Constr.* 2024, 168:105874.

[26] Zhou S, Liu K, Li D, Fu C, Ning Y, *et al.* Augmenting general-purpose large-language models with domain-specific multimodal knowledge graph for question-answering in construction project management. *Adv. Eng. Inform.* 2025, 65:103142.

[27] Wang Y, Xiao B, Bouferguene A, Al-Hussein M. Proactive safety hazard identification using visual–text semantic similarity for construction safety management. *Autom. Constr.* 2024, 166:105602.

[28] Chan C, Wong PK, Guo X, Cheng J, Chan J, *et al.* Context-aware vision-language model agent enriched with domain-specific ontology for construction site safety monitoring. *Autom. Constr.* 2025, 177:106305.

[29] Guo K, Wong PK, Cheng J, Chan CF, Leung PH, *et al.* Enhancing visual-LLM for construction site safety compliance via prompt engineering and bi-stage retrieval-augmented generation. *Autom. Constr.* 2025, 179:106490.

[30] Li A, Wong PKY, Tao X, Ma J, Cheng JC. An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent. *Adv. Eng. Inform.* 2025, 66:103375.

[31] Wang Z, Liu Z, Lu W, Jia L. Improving knowledge management in building engineering with hybrid retrieval-augmented generation framework. *J. Build. Eng.* 2025, 103:112189.

[32] Lee J, Ahn S, Kim D, Kim D. Performance comparison of retrieval-augmented generation and fine-tuned large language models for construction safety management knowledge retrieval. *Autom. Constr.* 2024, 168:105846.

[33] Chen G, Xiahou X, Li J, Chen L, Zou Y, *et al.* Knowledge graph-driven question answering for prefabricated building quality management through natural language processing and transfer learning. *Eng. Constr. Archit. Manag.* 2026, 33(9):6887–6913.

[34] Fu Y, Xu C, Zhang L, Chen Y. Control, coordination, and adaptation functions in construction contracts: a machine-coding model. *Autom. Constr.* 2023, 152:104890.

[35] Shuai B. A rationale-augmented NLP framework to identify unilateral contractual change risk for construction projects. *Comput. Ind.* 2023, 149:103940.

[36] Seo W, Kang Y. AI model for analyzing construction litigation precedents to support decision-making. *Autom. Constr.* 2024, 168:105824.

[37] Liu CY, Chou JS. Automated legal consulting in construction procurement using metaheuristically optimized large language models. *Autom. Constr.* 2025, 170:105891.

[38] Liang C, Huang Z, Wang H, Chai F, Yu C, *et al.* AECBench: a hierarchical benchmark for knowledge evaluation of large language models in the AEC field. *Adv. Eng. Inform.* 2026, 71:104314.

[39] Wu J, Jiang M, Fan J, Li S, Xu H, *et al.* Arch-Eval benchmark for assessing Chinese architectural domain knowledge in large language models. *Sci. Rep.* 2025, 15(1):13485.

[40] Zhang M, Xu C, Gan Y, Wang Y, Fu Y, *et al.* Automating construction contract question answering using large language model and fine-tuning. *Expert Syst. Appl.* 2026, 297:129493.

[41] Zhang L, Ning Y. Improving construction contract question answering through embedding optimization and semantic chunking in large language models. *Adv. Eng. Inform.* 2026, 69:104027.

[42] Xia H, Chen Y, Qi X, Wang Y. Benchmarking international construction contract knowledge of large language models. *Expert Syst. Appl.* 2026, 317:131754.

[43] Zhou S, Xie S, Yi W, Wang W. Construction case-relevant article-level law identification using fine-tuned large language models: a study in China's construction industry. *Adv. Eng. Inform.* 2026, 70:104144.

[44] Chen J, Tian J, Zheng T, Hui Y, Li Z. Qwen-AEC: instruction-tuning large language models for high-performance building code analysis. *Adv. Eng. Inform.* 2026, 71:104268.

[45] He C, Yu B, Liu M, Guo L, Tian L, *et al.* Utilizing large language models to illustrate constraints for construction planning. *Buildings* 2024, 14(8):2511.

[46] Chen G, Alsharef A, Ovid A, Albert A, Jaselskis E. Meet2Mitigate: an LLM-powered framework for real-time issue identification and mitigation from construction meeting discourse. *Adv. Eng. Inform.* 2025, 64:103068.

[47] Jeon K, Lee G. Hybrid large language model approach for prompt and sensitive defect management: a comparative analysis of hybrid, non-hybrid, and GraphRAG approaches. *Adv. Eng. Inform.* 2025, 64:103076.

[48] Jiang G, Ma Z, Zhang L, Chen J. EPlus-LLM: a large language model-based computing platform for automated building energy modeling. *Appl. Energy* 2024, 367:123431.

[49] Forth K, Borrmann A. Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM. *J. Build. Eng.* 2024, 95:110312.

[50] Zhang C, Zhang J, Zhao Y, Lu J. Automated data mining framework for building energy conservation aided by generative pre-trained transformers (GPT). *Energy Build.* 2024, 305:113877.

[51] Du C, Nousias S, Borrmann A. Towards a copilot in BIM authoring tool using a large language model-based agent for intelligent human-machine interaction. In *Proceedings of the 31st International Workshop on Intelligent Computing in Engineering, EG-ICE 2024*, Vigo, Spain, July 3–5, 2024, pp. 403–412.

[52] Chung S, Kim J, Baik J, Chi S, Kim DY. Identifying issues in international construction projects from news text using pre-trained models and clustering. *Autom. Constr.* 2024, 168:105875.

[53] An X, Zhou L, Liu Z, Wang C, Li P, *et al.* Dataset and benchmark for detecting moving objects in construction sites. *Autom. Constr.* 2021, 122:103482.

[54] Liang Y, Vadakkepat P, Chua DKH, Wang S, Li Z, *et al.* Recognizing temporary construction site objects using CLIP-based few-shot learning and multi-modal prototypes. *Autom. Constr.* 2024, 165:105542.

[55] Qin S, Guan H, Liao W, Gu Y, Zheng Z, *et al.* Intelligent design and optimization system for shear wall structures based on large language models and generative artificial intelligence. *J. Build. Eng.* 2024, 95:109996.

[56] Wang X, Yue Q, Liu X. Crack image classification and information extraction in steel bridges using multimodal large language models. *Autom. Constr.* 2025, 171:105995.

[57] Wang Y, Luo H, Fang W. An integrated approach for automatic safety inspection in construction: domain knowledge with multimodal large language model. *Adv. Eng. Inform.* 2025, 65:103246.

[58] Chen H, Hou L, Wu S, Zhang G, Zou Y, *et al.* Augmented reality, deep learning and vision-language query system for construction worker safety. *Autom. Constr.* 2024, 157:105158.

[59] Chen Q, Yin X. Tailored vision-language framework for automated hazard identification and report generation in construction sites. *Adv. Eng. Inform.* 2025, 66:103478.

[60] Yang B, Zhang B, Han Y, Liu B, Hu J, *et al.* Vision transformer-based visual language understanding of the construction process. *Alex. Eng. J.* 2024, 99:242–256.

[61] Yong G, Liu M, Lee S. Automated captioning for ergonomic problem and solution identification in construction using a vision-language model and caption augmentation. In *Proceedings of Construction Research Congress 2024*, Des Moines, USA, March 20–23, 2024, pp. 709–718.

[62] Pu H, Yang X, Li J, Guo R. AutoRepo: a general framework for multimodal LLM-based automated construction reporting. *Expert Syst. Appl.* 2024, 255:124601.

[63] Li Y, Xu F, Zhang Z, Mei X, Huang H. Construction site fall hazard identification and automated captioning using adapted vision-language models. *Autom. Constr.* 2026, 183:106790.

[64] Li T, Ge X, Wang Z, Ren P, Xu Z. Multimodal large language model (MLLM) benchmark for intelligent construction in underground engineering. *Autom. Constr.* 2026, 188:106997.

[65] Li Y, Ji M, Chen J, Wei X, Gu X, *et al.* A large language model-based building operation and maintenance information query. *Energy Build.* 2025, 334:115515.

[66] Jin S, Kim D, Lee J, Lee D. Evaluating large language models (LLMs) for semantic interpretation of IFC-based BIM data. *Autom. Constr.* 2026, 185:106879.

[67] Gong X, Wang B, Xu Y, Tao X, Chen W, *et al.* Automated environmental assessment of construction sites using multimodal large language model (MLLM)-based point cloud segmentation and instance-level scan-vs-BIM. *Autom. Constr.* 2026, 187:106949.

[68] Kwasafo OK, Saghatforoush E, Govender N. A generative AI model for automating construction bid preparation. *Int. J. Constr. Manag.* 2026, 26(14):3182–3202.

[69] Jung Y, Cho I, Hsu SH, Golparvar-Fard M. VisualSiteDiary: a detector-free vision-language transformer model for captioning photologs for daily construction reporting and image retrievals. *Autom. Constr.* 2024, 165:105483.

[70] Xu HR, Zhang N, Yin ZY, Guy Atangana Njock P. GeoLLM: a specialized large language model framework for intelligent geotechnical design. *Comput. Geotech.* 2025, 177:106849.

[71] Liu D, Zhou X, Li Y. Large language model–driven building energy model generation. *Energy Build.* 2026, 359:117255.

[72] Su Y, Li H, Wang M, Wen Y. Intelligent matching of construction materials: a multi-agent framework based on large language models and a multi-scale strategy. *Buildings* 2026, 16(17):3414.

[73] Zhou J, Ma Z. Named entity recognition for construction documents based on fine-tuning of large language models with low-quality datasets. *Autom. Constr.* 2025, 174:106151.

[74] Panahi R, Kivlin JP, Louis J. Request for information (RFI) recommender system for pre-construction design review application using natural language processing, Chat-GPT, and computer vision. In *Proceedings of Computing in Civil Engineering 2023*, Corvallis, USA, June 25–28, 2023, pp. 159–166.

[75] Zhong Y, Goodfellow SD. Domain-specific language models pre-trained on construction management systems corpora. *Autom. Constr.* 2024, 160:105316.

[76] Naderi H, Shojaei A. Large-language model (LLM)-powered system for situated and game-based construction safety training. *Expert Syst. Appl.* 2025, 283:127887.

[77] Jung Y, Hockenmaier J, Golparvar-Fard M. Transformer language model for mapping construction schedule activities to Uniformat categories. *Autom. Constr.* 2024, 157:105183.

[78] Zhou S, Liu X, Li D, Gu T, Liu K, *et al.* Integrating domain-specific knowledge and fine-tuned general-purpose large language models for question-answering in construction engineering management. *Autom. Constr.* 2025, 175:106206.

[79] Luo H, Wu J, Liu J, Antwi-Afari MF. Large language model-based code generation for the control of construction assembly robots: a hierarchical generation approach. *Dev. Built Environ.* 2024, 19:100488.

[80] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, *et al.* Training language models to follow instructions with human feedback. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, New Orleans, USA, November 28–December 9, 2022, pp. 27730–27744.

[81] OpenAI. Introducing ChatGPT. 2022. Available: https://openai.com/index/chatgpt/ (accessed on 29 June 2026).

[82] Gemini Team. Gemini 1.5: unlocking multimodal understanding across millions of tokens of context. 2024. Available: https://storage.googleapis.com/deepmind-media/gemini/gemini_v1_5_report.pdf (accessed on 29 June 2026).

[83] Grattafiori A, Dubey A, Jauhri A, Pandey A, Kadian A, *et al.* The Llama 3 herd of models. 2024. Available: https://ai.meta.com/research/publications/the-llama-3-herd-of-models/ (accessed on 29 June 2026).

[84] OpenAI. Introducing GPT-4.1 in the API. 2025. Available: https://openai.com/index/gpt-4-1/ (accessed on 29 June 2026).

[85] OpenAI. GPT-4o system card. 2024. Available: https://openai.com/index/gpt-4o-system-card/ (accessed on 29 June 2026).

[86] Meta AI. The Llama 4 herd: the beginning of a new era of natively multimodal AI innovation. 2025. Available: https://ai.meta.com/blog/llama-4-multimodal-intelligence/ (accessed on 29 June 2026).

[87] Anthropic. Claude 3.7 Sonnet and Claude Code. 2025. Available: https://www.anthropic.com/news/claude-3-7-sonnet (accessed on 29 June 2026).

[88] Guo D, Yang D, Zhang H, Song J, Zhang R, *et al.* DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning. 2025. Available: https://api-docs.deepseek.com/news/news250120 (accessed on 29 June 2026).

[89] Yang A, Li A, Yang B, Zhang B, Hui B, *et al.* Qwen3 technical report. 2025. Available: https://github.com/QwenLM/Qwen3 (accessed on 29 June 2026).

[90] Schick T, Dwivedi-Yu J, Dessì R, Raileanu R, Lomeli M, *et al.* Toolformer: language models can teach themselves to use tools. In *Proceedings of the 37th Conference on Neural Information Processing Systems*, New Orleans, USA, December 10–16, 2023.

[91] Yao S, Zhao J, Yu D, Du N, Shafran I, *et al.* ReAct: synergizing reasoning and acting in language models. In *Proceedings of the International Conference on Learning Representations*, Kigali, Rwanda, May 5–11, 2023.

[92] Uhm M, Kim J, Ahn S, Jeong H, Kim H. Effectiveness of retrieval augmented generation-based large language models for generating construction safety information. *Autom. Constr.* 2025, 170:105926.

[93] Bao S, Bu H. Defining and generating operation and maintenance management requirements in digital twin applications using the DT-GPT framework. *J. Build. Eng.* 2025, 104:112356.

[94] Lee J, Lee G. Long context window-based zero-shot legal interpretation of building codes and regulations. *Autom. Constr.* 2025, 179:106450.

[95] Sun Y, Gu Z, Yang SB. Probing vision and language models for construction waste material recognition. *Autom. Constr.* 2024, 166:105629.

[96] Tsai WL, Le PL, Ho WF, Chi NW, Lin JJ, *et al.* Construction safety inspection with contrastive language-image pre-training (CLIP) image captioning and attention. *Autom. Constr.* 2025, 169:105863.

[97] Iversen O, Huang L. Leveraging large language models for BIM-based automated compliance checking. *Autom. Constr.* 2026, 182:106707.

[98] Lu J, Zheng Z, Langtry M, Jackson M, Zhao Y, *et al.* Automated building energy modeling for energy retrofits using a large language model-based multi-agent framework. *iScience* 2025, 28(11):113867.

[99] Shi JWL, Solihin W, Yeoh JK. Fine-tuning a large language model for automated code compliance of building regulations. *Adv. Eng. Inform.* 2025, 68:103676.

[100] Jang S, Lee G, Oh J, Lee J, Koo B. Automated detailing of exterior walls using NADIA: natural-language-based architectural detailing through interaction with AI. *Adv. Eng. Inform.* 2024, 61:102532.

[101] Duggempudi J, Gao L, Senouci A, Han Z, Zhang Y. Text-to-Layout: a generative workflow for drafting architectural floor plans using LLMs. In *Proceedings of the International Conference on Transportation and Development 2026*, Detroit, USA, June 28–July 1, 2026, pp. 191–205.

[102] Fuchs S, Witbrock M, Dimyadi J, Amor R. Using large language models for the interpretation of building regulations. *J. Eng. Proj. Prod. Manag.* 2024, 14(4):0035.

[103] Wang F, Zhou J, Zheng X. Identification of hydraulic engineering construction accident causes using instruction tuning and structured prompting in large language models. *J. Constr. Eng. Manag.* 2026, 152(8):04026122.

[104] Zheng Z, Han J, Chen KY, Cao XY, Lu X, *et al.* Translating regulatory clauses into executable codes for building design checking via large language model driven function matching and composing. *Eng. Appl. Artif. Intell.* 2026, 163:112823.

[105] Zhang H, Zhang R. A multiagent large language model–based system for early-stage building layout planning. *J. Comput. Civ. Eng.* 2026, 40(6):04026095.

[106] Baek S, Park CY, Jung W. Automated safety risk management guidance enhanced by retrieval-augmented large language model. *Autom. Constr.* 2025, 176:106255.

[107] Xiao H, Zhuang J, Yang B, Han J, Yu Y, *et al.* Generative knowledge-guided review system for construction disclosure documents. *Adv. Eng. Inform.* 2025, 68:103618.

[108] Meng F, Pan M. Advancing quality control in off-site construction with large language models enhanced by hybrid retrieval-augmented generation. *Adv. Eng. Inform.* 2026, 71:104381.

[109] Ali RH, Park CY, Aryal A, Jeong HD. Automated MEP drawing analysis using computer vision preprocessing and vision language models. In *Proceedings of Construction Research Congress 2026*, San Antonio, USA, March 18–21, 2026, pp. 919–928.

[110] Dong Y, Zhan Z, Hu Y, Doe DM, Han Z. AI BIM coordinator for non-expert interaction in building design using LLM-driven multi-agent systems. *Autom. Constr.* 2025, 180:106563.

[111] Gao Y, Hu F, Chai C, Weng Y, Li H. Multi-agent framework for schema-guided reasoning and tool-augmented interaction with IFC models. *Autom. Constr.* 2026, 186:106888.

[112] Zhang L, Ford V, Chen Z, Chen J. Automatic building energy model development and debugging using large language models agentic workflow. *Energy Build.* 2025, 327:115116.

[113] Yoon S, Song J, Li J. Ontology-enabled AI agent-driven intelligent digital twins for building operations and maintenance. *J. Build. Eng.* 2025, 108:112802.

[114] Slezák S, Shevidi S, Shakeri Z, Ron G, Amtsberg F, *et al.* Human–robot collaboration in building disassembly: a multi-agent LLM architecture. *Constr. Robot.* 2026, 10:23.

[115] Alwashah Z, Xiao B, Liu H, Mueller ST, Shao X. Reliable LLM-driven BIM automation through capability-based multi-dimensional evaluation. *Autom. Constr.* 2026, 187:106910.

[116] Chen X, Zou Z. Are large pre-trained vision language models effective construction safety inspectors. *Data-Centric Eng.* 2026, 7:e11.

[117] Haghighi R, Bui VH, Su W. Fine-tuned large language models as surrogate schedulers for smart building energy management. *Energy Build.* 2026, 364:117588.

[118] Kazemi MH, Alvanchi A. Application of NLP-based models in automated detection of risky contract statements written in complex script system. *Expert Syst. Appl.* 2025, 259:125296.

[119] Gao Y, Gan Y, Chen Y, Chen Y. Application of large language models to intelligently analyze long construction contract texts. *Constr. Manag. Econ.* 2025, 43(3):226–242.

[120] Prieto SA, Mengiste ET, García de Soto B. Investigating the use of ChatGPT for the scheduling of construction projects. *Buildings* 2023, 13(4):857.

[121] Donia M, Elbeltagi E, Elhakeem A, Wefki H. An integrated BIM–NLP framework for design-informed automated construction schedule generation. *Designs* 2026, 10(2):43.

[122] Li H, Yang R, Xu S, Xiao Y, Zhao H. Intelligent checking method for construction schemes via fusion of knowledge graph and large language models. *Buildings* 2024, 14(8):2502.

[123] Borrmann A, König M, Koch C, Beetz J. *Building information modeling: why? What? How?* 1st ed. Cham: Springer International Publishing, 2018. pp. 1–24.

[124] Lin JR, Hu ZZ, Zhang JP, Yu FQ. A natural-language-based approach to intelligent data retrieval and representation for cloud BIM. *Comput.-Aided Civ. Infrastruct. Eng.* 2015, 31(1):18–33.

[125] Zheng J, Fischer M. Dynamic prompt-based virtual assistant framework for BIM information search. *Autom. Constr.* 2023, 155:105067.

[126] Jang S, Lee G. Interactive design by integrating a large pre-trained language model and building information modeling. In *Proceedings of Computing in Civil Engineering 2023: Visualization, Information Modeling, and Simulation*, Reston, USA, June 25–28, 2023, pp. 291–299.

[127] Zhang J. How can ChatGPT help in automated building code compliance checking? In *Proceedings of the 40th International Symposium on Automation and Robotics in Construction*, Chennai, India, July 5–7, 2023, pp. 63–70.

[128] Chen N, Lin X, Jiang H, An Y. Automated building information modeling compliance check through a large language model combined with deep learning and ontology. *Buildings* 2024, 14(7):1983.

[129] You H, Ye Y, Zhou T, Zhu Q, Du J. Robot-enabled construction assembly with automated sequence planning based on ChatGPT: RoboGPT. *Buildings* 2023, 13(7):1772.

[130] Uddin SMJ, Albert A, Ovid A, Alsharef A. Leveraging ChatGPT to aid construction hazard recognition and support safety education and training. *Sustainability* 2023, 15(9):7121.

[131] Hussain R, Sabir A, Lee DY, Zaidi SFA, Pedro A, *et al.* Conversational AI-based VR system to improve construction safety training of migrant workers. *Autom. Constr.* 2024, 160:105315.

[132] Yong G, Liu M, Lee S. Explainable image captioning to identify ergonomic problems and solutions for construction workers. *J. Comput. Civ. Eng.* 2024, 38(4):04024022.

[133] Zhang L, Chen Z, Ford V. Advancing building energy modeling with large language models: exploration and case studies. *Energy Build.* 2024, 323:114788.

[134] Lu J, Tian X, Zhang C, Zhao Y, Zhang J, *et al.* Evaluation of large language models (LLMs) on the mastery of knowledge and skills in the heating, ventilation and air conditioning (HVAC) industry. *Energy Built Environ.* 2025, 6(5):875–892.

[135] Liu J, Li W, Li F, Zhao X. Chinese named entity recognition for bridge damage and defects based on text mining and natural language pretraining models. *J. Constr. Eng. Manag.* 2025, 151(6):04025060.

[136] Yang J, Yang X, Li R, Luo M, Jiang S, *et al.* BERT and hierarchical cross attention-based question answering over bridge inspection knowledge graph. *Expert Syst. Appl.* 2023, 233:120896.

[137] Wen Y, Chen K. Autonomous detection and assessment of indoor building defects using multimodal learning and GPT. In *Proceedings of Construction Research Congress 2024*, Des Moines, USA, March 20–23, 2024, pp. 1001–1009.

[138] Xu F, Nguyen T, Du J. Augmented reality for maintenance tasks with ChatGPT for automated text-to-action. *J. Constr. Eng. Manag.* 2024, 150(4):04024015.

[139] Wang M, Li Y, Li S. Robotic assembly of interlocking blocks for construction based on large language models. In *Proceedings of Construction Research Congress 2024*, Des Moines, USA, March 20–23, 2024, pp. 777–786.

[140] Park S, Menassa CC, Kamat VR. Integrating large language models with multimodal virtual reality interfaces to support collaborative human–robot construction work. *J. Comput. Civ. Eng.* 2025, 39(1):04024053.

[141] Madireddy S, Gao L, Din ZU, Kim K, Senouci A, *et al.* Large language model-driven code compliance checking in building information modeling. *Electronics* 2025, 14(11):2146.

[142] Yang Y, Jing X, Liu YM. Automated checking of highway bridge BIM models based on large language models. *Buildings* 2025, 15(19):3465.

[143] Gu X, Chen C, Fang Y, Mahabir R, Fan L. CECA: an intelligent large-language-model-enabled method for accounting embodied carbon in buildings. *Build. Environ.* 2025, 272:112694.

[144] Chen Q, Yin X, Yuan B, Chen Q. Personalized safety training for construction workers: a large language model-driven multi-agent framework integrated with knowledge graph reasoning. *Comput. Ind.* 2026, 174:104399.

[145] Pham HTTL, Han S. Natural language processing with multitask classification for semantic prediction of risk-handling actions in construction contracts. *J. Comput. Civ. Eng.* 2023, 37(6):04023027.

[146] Wang S, Fu Y, Kim J. Toward construction-specialized, small language models: the interplay of domain adaptation, model scale and data volume. *Adv. Eng. Inform.* 2026, 69:104035.

[147] Han D, Zhao W, Yin H, Qu M, Zhu J, *et al.* Large language models driven BIM-based DfMA method for free-form prefabricated buildings: framework and a usefulness case study. *J. Asian Archit. Build. Eng.* 2025, 24(3):1500–1517.

[148] Shu L, Yeganeh A, Zhao D. Large language models for building energy retrofit decision-making: technical and sociotechnical evaluations. *Buildings* 2025, 15(22):4081.

[149] Wang L, Yuan J, Su S, Ding H, Bai Y, *et al.* Fusing enhanced YOLO and knowledge graph-based large language models for automatic risk perception in tower crane operations. *Autom. Constr.* 2026, 183:106823.