



DOI:10.12404/j.issn.1671-1815.2405919

引用格式:许娜,陈曦,杨建平,等.大语言模型在土木工程领域中的研究现状、挑战及趋势[J].科学技术与工程,2025,25(21):8773-8783.  
Xu Na, Chen Xi, Yang Jianping, et al. Research status, challenges, and trends of large language models in the field of civil engineering[J].  
Science Technology and Engineering, 2025, 25(21): 8773-8783.

建筑科学

# 大语言模型在土木工程领域中的研究现状、挑战及趋势

许娜<sup>1,2</sup>, 陈曦<sup>1</sup>, 杨建平<sup>3</sup>, 张博<sup>4</sup>, 陈伟<sup>4</sup>

(1. 中国矿业大学力学与土木工程学院, 徐州 221116; 2. 中国矿业大学人工智能研究院, 徐州 221116;  
3. 徐州工程学院土木工程学院, 徐州 221018; 4. 中国矿业大学计算机科学与技术学院, 徐州 221116)

**摘要** 土木工程行业在信息化转型中面临着大量的非结构化的文本信息,大语言模型(Large Language Models, LLMs)由于其强大的自然语言处理能力,为行业领域的智能化变革提供了新的机遇。采用系统性文献回顾的方法,在梳理LLMs的技术架构及在垂直领域研究现状的基础上,提出了LLMs在土木工程领域的四大应用场景及技术路线、面临的挑战及研究趋势。研究发现,LLMs已在土木工程领域有探索性的研究与应用,目前主要集中在内容生成类、智能问答类、文本摘要类及分析推理类四大应用场景,覆盖土木工程项目全生命周期阶段,并具有跨学科、跨模态融合的特性。然而,LLMs的应用仍面临知识专业性低、信息时效性差、数据质量及交互性低等挑战。基于此,提出了一系列未来研究机遇,在模型优化方面,利用参数高效微调技术注入专业知识,增强LLMs在土木工程领域应用的广度和深度;与知识图谱结合,提升LLMs在回答中的精准性、可解释性与时效性;融合多模态的数据类型,扩展LLMs在土木工程领域的应用场景;开发适用的模型评估方法,量化LLMs在土木工程领域应用的价值及性能表现。在应用场景方面,结合LLMs和土木工程领域特点,可以拓展LLMs在文档生成、问答系统、信息抽取、合规性审查等复杂任务中的应用,提高从业者与数据间的交互效率。研究旨在为学术界和企业界进一步将LLMs应用于土木工程领域提供借鉴与参考。

**关键词** 土木工程; 大语言模型; 自然语言生成; 生成式人工智能

中图分类号 TU17; 文献标志码 A

## Research Status, Challenges, and Trends of Large Language Models in the Field of Civil Engineering

XU Na<sup>1,2</sup>, CHEN Xi<sup>1</sup>, YANG Jian-ping<sup>3</sup>, ZHANG Bo<sup>4</sup>, CHEN Wei<sup>4</sup>

(1. School of Mechanics and Civil Engineering, China University of Mining and Technology, Xuzhou 221116, China;  
2. Artificial Intelligence Research Institute, China University of Mining and Technology, Xuzhou 221116, China;  
3. School of Civil Engineering, Xuzhou University of Technology, Xuzhou 221018, China;  
4. School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116, China)

**[Abstract]** The civil engineering industry faces with a vast array of unstructured textual information during its digital transformation. Large language models (LLMs) provide a new opportunity for the intelligent transformation of the industry because of its powerful natural language processing capability. A systematic literature review approach was employed, and based on the technical framework of LLMs and the current state of research in vertical domains, four major application scenarios for LLMs in civil engineering were suggested, along with corresponding technological approaches, challenges faced, and research trends. It is found that exploratory research and application of LLMs in civil engineering have been conducted, primarily focusing on content creation, intelligent Q & A, text summarization, and analytical reasoning, covering the entire lifecycle of civil engineering projects and featuring interdisciplinary and multi-modal integration. However, the utilization of LLMs struggles with low specificity of knowledge, poor timeliness of information, and inferior data quality and interactivity. Based on this, a series of future research opportunities were proposed to enhance the breadth and depth of LLMs application in the field of civil engineering by using parametric efficient fine-tuning technology to inject expertise in model optimization. Combined with knowledge graph, LLMs can improve the accuracy, interpretability and timeliness of answers. Multi-

收稿日期: 2024-08-06 修订日期: 2025-03-13

基金项目: 国家社会科学一般项目(23BGL277)

第一作者: 许娜(1982—),女,汉族,江苏徐州人,博士,副教授。研究方向:基于人工智能的土木工程管理。E-mail:xuna@cumt.edu.cn。

投稿网址: www.stae.com.cn

modal data types were integrated to expand the application scenarios of LLMs in civil engineering. Appropriate model evaluation methods were developed to quantify the value and performance of LLMs applications in civil engineering. In terms of application scenarios, combined with the characteristics of LLMs and civil engineering fields, the application of LLMs in complex tasks such as document generation, question and answer system, information extraction and compliance review can be expanded, and the interaction efficiency between practitioners and data can be improved. The purpose of the study is to provide reference for the academic and business circles to further apply LLMs in the field of civil engineering.

[**Keywords**] civil engineering; large language models (LLMs); natural language generation (NLG); generative artificial intelligence (Gen-AI)

土木工程行业蕴含着市场动态、合同法规、设计文件、施工报告、变更索赔等诸多具有价值的信息<sup>[1]</sup>,其中80%的信息以文本形式分散存储于项目各参与方或主管部门<sup>[2]</sup>,难以实现共享和再利用,通过自然语言处理(natural language processing, NLP)技术挖掘储存在这些文本数据中有价值的信息,将有助于实践者和决策者提高管理成效<sup>[3]</sup>。

早期的NLP技术研究主要基于正则表达式或匹配词的规则匹配方法<sup>[4]</sup>,但专业领域知识的复杂性和语言表达的多样性使得其难以涵盖所有语言现象。因此,基于机器学习的方法开始出现,通过建立土木工程领域内文本资料的语料库,利用统计模型,如隐马尔可夫模型(hidden Markov model, HMM)<sup>[5]</sup>、K最近邻(K-nearest neighbor, KNN)<sup>[6]</sup>和条件随机场(conditional random field, CRF)等<sup>[7]</sup>,可从大量文本信息中学习和发现数据规律,但基于机器学习的统计模型仍难以处理语句中的深层语义信息。深度学习的飞速发展推动NLP研究进入了一个新阶段。在土木工程领域,卷积神经网络(convolutional neural networks, CNN)、循环神经网络(recurrent neural network, RNN)<sup>[8]</sup>、注意力机制和变换器(Transformer)架构<sup>[9]</sup>等模型的引入都大幅提升了领域内自然语言处理任务的性能和准确性<sup>[10]</sup>。

2022年底,基于大语言模型(LLMs)的ChatGPT应用展示了LLMs在通用领域强大的自然语言理解及推理能力<sup>[11]</sup>。在土木工程领域中,虽然大语言模型的研究与应用尚处于起步阶段,但已在文本生成、语义理解和审查分类等诸多NLP任务中展示出了巨大的应用潜力。

为充分把握当下AI热潮的机遇与挑战,应更多关注LLMs研究与发展,现采取系统性的文献回顾方式,通过梳理现有基于LLMs在土木工程行业内的应用研究,按照其应用场景分为内容生成、智能问答、文本摘要和分析推理四类,对其进行综述与讨论,并提出LLMs在土木行业应用面临的挑战,最后从模型性能优化和实际应用场景的拓宽两个维度,展望了未来研究的趋势方向,旨在为该交叉领域的深入发展提供参考与指导。

## 1 大语言模型的技术架构及特性

大语言模型(large language models, LLMs)是一种基于大量文本数据训练的NLP模型<sup>[12]</sup>,其核心主要基于变换器(Transformer)架构(图1)<sup>[9]</sup>,区别于传统NLP模型,大多数研究者认为,LLMs具有数十亿以上的参数<sup>[13-14]</sup>。

在不同的NLP任务场景中,LLMs在训练方法、结构设计及应用领域展现出明显的多样性。根据技术架构的不同,LLMs可以分为三类:仅编码器大语言模型(encoder-only LLMs)、编解码器大语言模型(encoder-decoder LLMs)和仅解码器大语言模型(decoder-only LLMs)。各架构中具有代表性的目前主流的LLMs发展时间线如图2所示,基本信息如表1所示。

表1 具有代表性的大语言模型基本信息一览

Table 1 Basic information of existing mainstream LLMs

模型名称	发布机构	技术架构	发布时间	参数量/ $10^9$
BERT	Google	仅编码器	2018-10	0.11
ERNIE	Baidu	仅编码器	2019-03	0.114
ALBERT	Google	仅编码器	2019-09	0.235
ELECTRA	Stanford University	仅编码器	2020-03	0.025 ~ 0.3
DeBERTa	Microsoft	仅编码器	2020-12	0.9 ~ 1.5
GPT-2	OpenAI	仅解码器	2019-02	1.17 ~ 15.77
XLNet	Google/CMU	仅解码器	2019-06	0.11 ~ 0.34
Galactica	Meta	仅解码器	2019-08	6.7
GPT-3	OpenAI	仅解码器	2020-06	175
InstructGPT	OpenAI	仅解码器	2021-11	175
PaLM	Google	仅解码器	2022-04	540
FLAN-PaLM	Google	仅解码器	2022-10	540
Llama2	Meta	仅解码器	2023-02	7
Bard	Google	仅解码器	2023-02	1.3
GPT-4	OpenAI	仅解码器	2023-04	175
GPT-4-Turbo	OpenAI	仅解码器	2023-11	175
BART	FAIR	编解码器	2019-10	0.14
T5	Google	编解码器	2019-10	0.77 ~ 1.1
mT5	Google	编解码器	2020-10	13
UL2	Google	编解码器	2022-05	20
Flan-T5	Google	编解码器	2022-11	3
ChatGLM-130B	Tsinghua University	编解码器	2023-03	130
ChatGLM-6B	Tsinghua University	编解码器	2023-06	6
BaiChuan	Alibaba	编解码器	2023-07	13

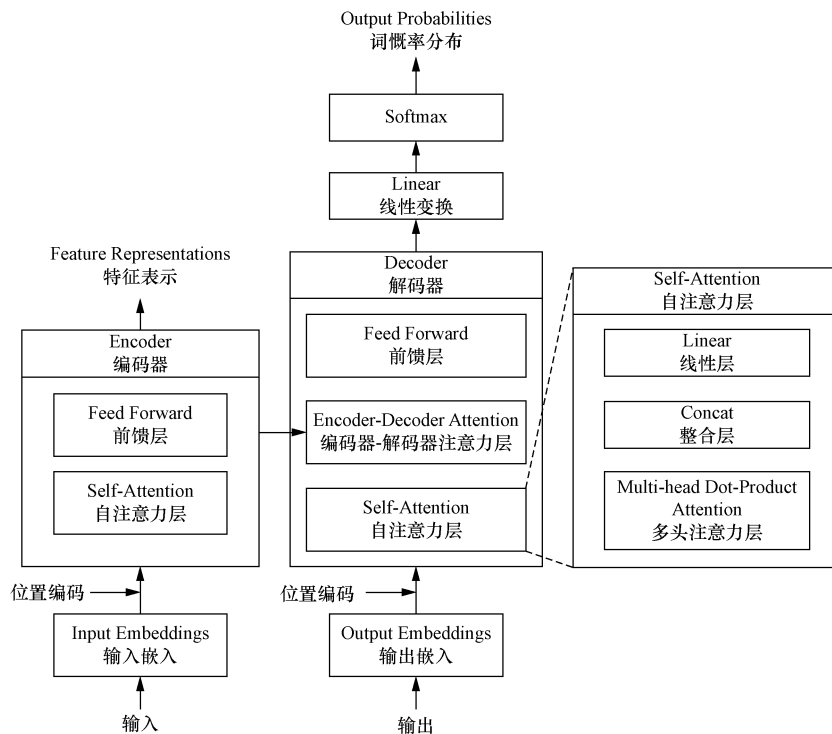


图 1 基于 Transformer 模型的 LLMs 核心技术架构

Fig. 1 The core architecture of LLMs based on the Transformer model

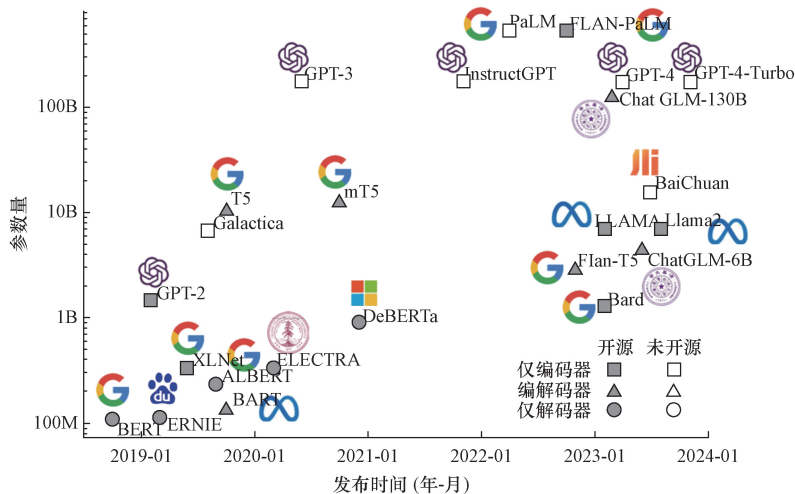


图 2 具有代表性的大语言模型发展时间线

Fig. 2 Development timeline of representative LLMs with different architectures

随着时代算力的显著增强,大语言模型的最大参数量呈指数级上涨,ChatGPT 的成功证明了在庞大数据及强大算力的加持下,仅解码器架构 LLMs 在广泛的场景应用的可能,这使得基于编解码器或纯解码器架构的 LLMs 逐渐受到公众关注。

(1) 仅编码器大语言模型。仅编码器大语言模型通过编码器对输入的句子进行向量化,并根据上下文权重参数分配注意力来理解单词间的关系。这些模型适用于需要理解整个文本的任务,例如文本分类<sup>[15]</sup>和命名实体识别<sup>[16]</sup>,但并不能直接用于

内容生成类型的任务,且泛化性较差,通常只能用于执行特定的少量任务。

(2) 编解码器大语言模型。编解码器大语言模型相较于仅基于编码器的模型,额外集成了一个解码器组件,这使得模型不仅能够对输入序列进行高效编码,还能利用解码器将编码后的向量映射回词汇空间,从而具备了生成连续文本的能力<sup>[17]</sup>。这种双向的信息流设计使 LLMs 特别适用于需要深入理解输入并生成相关输出的任务,如文本摘要生成、机器翻译以及问答系统<sup>[18]</sup>,编解码器架构因其强大

的上下文建模能力而表现出色。

(3) 仅解码器大语言模型。仅解码器大语言模型亦称为自回归模型,其完全依赖于解码器模块来生成连贯的输出文本。仅解码器大语言模型的一个显著特点是它们能够在没有针对特定任务进行微调的情况下,仅通过少量示例或提示(prompts)便执行复杂的自然语言处理任务,这一现象被称为(少或零)样本学习(few/zero-shot learning)<sup>[19]</sup>。这意味着模型可以利用其在预训练阶段学到的广泛知识,在面对新任务时无需额外的数据即可推断出解决方案。因此,其在生成特定文本、问答等任务方面表现强大,而预训练-微调范式(pretrain-finetune paradigm)的灵活性<sup>[20]</sup>允许模型快速适应多种下游任务,降低了针对每项任务重新训练模型的需求,极大地提高了效率和实用性。

综上所述,3种架构的大语言模型具有其各自的特点和适用场景。仅编码器模型专注于解释文本,适用于文本分类与实体识别;编解码器模型强理解与生成文本的能力,适用于文本摘要、翻译与问答系统;仅解码器模型则偏重生成文本,适用于文本创作和问答任务。值得一提的是,尽管在理论上编码器-解码器架构更能结合上下文表现出更强大的性能,但同时也会消耗更多的计算资源,增加处理任务的时间<sup>[21]</sup>。仅解码器架构只涉及生成输出序列的过程,其计算复杂度相对较低,使训练和生成过程更高效,这也是GPT系列等主流大语言模型均采用了仅解码器架构的原因之一。

## 2 大语言模型在土木工程领域的研究现状

### 2.1 内容生成类应用场景及其技术路线

土木工程项目从规划到运营的全生命周期中,存在大量文档的撰写与管理,包括但不限于技术报告、合同条款、施工方案和安全指导书。这些文档不仅要求专业严谨,而且涉及复杂的行业知识和标准。传统上,这些工作由工程师和专家团队耗时完成,而仅编码器与编解码器架构的LLMs能够捕捉用户对话中的语义,从而依靠预训练获得的知识中理解和生成逻辑连贯的自然语言文本,这使得引入LLMs使其自动化成为了可能。例如,Chung等<sup>[22]</sup>的研究显示,类似ChatGPT和BARD这样的LLMs可以基于项目需求和行业规范,自动生成施工文档,甚至解答相关问题。Prieto等<sup>[23]</sup>进一步证明,LLMs能根据项目进度生成连贯的施工计划,对任务进行逻辑排序,满足项目约束条件,极大地提高了工作效率。

Pu等<sup>[24]</sup>则引入了多模态信息处理的概念,其提出的AutoRepo框架结合了无人机图像采集与Vision Transformer编码器,将现场图像转化为向量表示,再由LLMs生成标准化的施工检查报告,实现了从图像到文本的有效转换。Prabhu等<sup>[25]</sup>探索了LLMs与叙事结构框架的结合,自动创建建筑项目状态更新报告,不仅涵盖了从文本到文本的转换,还涉及了文本到语音的过程,引入情境感知框架以提升信息流通性和项目管理的透明度与协作性。

在该类任务的研究中,技术路线主要为多模态的信息获取、表示学习、成果生成(图3)。多模态信息获取指从不同的传感器或数据源中获取多种类型的信息,包括文本、图像、语音、视频等。在表示学习阶段,多模态数据经过编码器输出为向量表征,以便模型能够更好地处理。LLMs作为该框架的最终环节,负责解析获取的数据,理解任务需求,并输出连贯、逻辑性强的自然语言文本。在此过程中,幻觉现象(hallucination)是亟待解决的问题,即模型可能生成看似合理但实际错误的信息。为克服这一障碍,研究者通常采取特定的提示策略或模型微调方案来指导LLMs在专业领域内更准确地生成文本。此外,高质量内容生成往往对模型参数量有较高要求,因此,利用商业模型提供的API成为一种高效的实践策略。

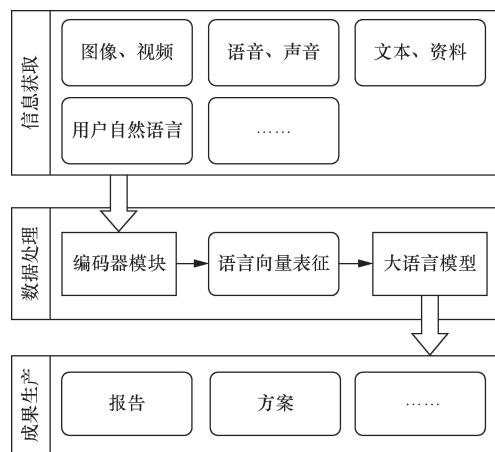


图3 内容生成类应用场景的技术路线

Fig. 3 Technology roadmap for generating application

### 2.2 智能问答类应用场景及其技术路线

LLMs的特性展现出了理解自然复杂语言查询的强大能力,在Prompt的加持下能结合知识库提供精准的答案,这在智能问答类任务中尤为重要。对于土木工程行业,工程师和管理人员常常面临查找规范、标准以及最佳工程实践信息的紧迫需求,以应对日常工作中遇到的具体问题。LLMs在此背景

下扮演着专家顾问的角色,能够解答关于材料选择、施工方法以及规范解读等多方面疑问,极大地提升了工作效率和决策质量。

在智能问答的实践中,Zheng 等<sup>[26]</sup>提出的 BIM-GPT 集成框架展示了 LLMs 在建筑规范解读方面的潜力。该框架不仅能够处理自然语言查询,还能通过动态提示生成过程,解释并总结数据库信息,最终生成符合用户查询的响应,并通过云端的建筑信息模型 (building information modeling, BIM) 平台获得 3D 可视化。在零样本和少样本场景下,对意图分类、建筑对象类别分类、过滤参数和投影参数识别、值提示的任务均取得了良好的表现。覃思中等<sup>[27]</sup>基于检索增强生成 (retrieval-augmented generation, RAG) 框架,提出了一种融合提示词策略与本地知识库的方法来优化问题回答的准确性,通过 sentence-transformers 库将用户问题转化为富有语义的向量,并采用 FAISS 向量索引库进行相似度检索,从本地知识库中识别和挑选出与用户问题语义最接近的知识片段,这些高度相关的知识随后结合用户问题构成提示词,被送入答案生成模型以产生精准的回答。将此方案应用于 ChatGPT、ChatGLM、ERNIE-bot 和 Llama 这 4 种模型,并对其在土木工程领域内专业问题回答性能进行评估对比,结果证实了该方法能够提高模型在问答系统中的准确率。类似地,该方法也应用于施工档案管理<sup>[28]</sup>、合同文档查询<sup>[29]</sup>、企业内部知识管理<sup>[30]</sup>等知识数据交互场景,以提升信息的获取效率。

在土木工程这样专业知识密集领域,直接应用 LLMs 预训练获得的知识难以覆盖所有可能的问题,这将导致模型偶尔会提供不准确或误导性的回答。在此挑战之下,RAG 技术的引入成为关键<sup>[31]</sup>,其通常涉及 3 个核心组件:知识库、大语言模型和查询模块(图 4)。当用户提出问题,查询模块利用这些关键信息在结构化的知识库中进行高效检索,找到与问题相关的知识片段,大语言模型结合检索结果生成精确且详尽的回答。在 Guu 等<sup>[12]</sup>与 Logan 等<sup>[32]</sup>的研究中,证明了知识图谱可以作为结构化的知识库,以辅助 LLMs 使用训练集外的知识来描述事实。

在 RAG 技术之外,通过 Prompt 工程可以引导大语言模型自我思考,提高对复杂问题回答的完整性和准确性<sup>[33-34]</sup>。例如,思维链(chain of thought, CoT)策略的提示可以使模型逐步给出推理逻辑,以更多的生成时间换取更高的准确性<sup>[35]</sup>;自我评估(deliberate then generate, DTG)策略的提示还会包含错误信息以引导模型进行自我思考判断,从而

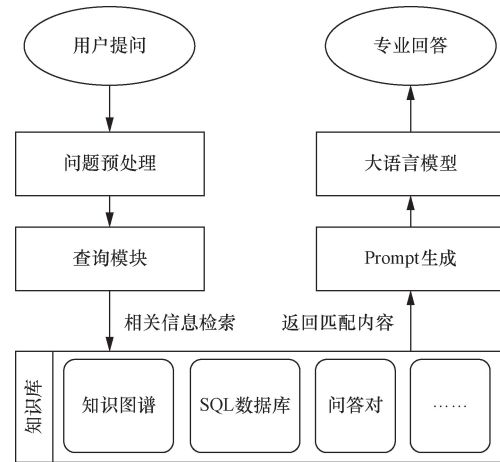


图 4 智能问答类应用场景的技术路线

Fig. 4 Technology roadmap for question-answering application

在文本生成时进行检测和错误更正<sup>[36]</sup>。然而,即便如此,保持知识库的更新和维护,以及避免模型的过拟合,仍然是需要解决的难题。未来的研究应继续探索如何优化模型的训练方式,以更好地适应不断变化的知识环境,同时确保模型的泛化能力和稳定性。

### 2.3 文本摘要类应用场景及其技术路线

在文本摘要类任务中,LLMs 凭借注意力机制的特性,展现了从冗长文档中筛选出核心信息的能力,从而生成精炼的摘要,极大地提升了信息获取的效率与质量。这一特性在教育、医学、法律等多个领域已得到广泛应用。在土木工程场景下,LLMs 能够得到工程合同、规范文档和研究报告的快速解读,迅速提炼出关键信息,使工程师能够在多变的项目环境中迅速掌握要点,洞察技术进步,预测潜在风险。

在具体的应用场景中,LLMs 主要用于理解和解释规范文本,并生成用户需要的决策建议。Zheng 等<sup>[37]</sup>提出基于大语言模型的函数识别方法(LLM-FuncMapper),提取建筑规范条文及其自动化规则解释,即通过构建一系列原子函数来捕捉建筑规范中隐含属性和复杂约束的共享计算逻辑生成函数库,使用链式思考等一系列提示策略,使大型语言模型能有效识别函数,同时,为了提高大语言模型识别功能的可表达性,设计了一个基于分类器的调谐策略,提高了识别相关预定义函数的召回率。Wong 等<sup>[38]</sup>提出,基于知识增强的语言模型可用于识别建筑合同中的风险,其研究构建了建筑合同条款库与专家知识库,通过计算项目条款与案例条款之间的向量相似性,根据相似度得分提取最相关的案例条款及其风险,并采用两阶段的提示过程让大语言模型生成可靠的结果。

在文本摘要类应用场景的研究中,其技术路线主要包括数据收集、清洗分割、标记、微调训练4个步骤(图5)。对摘要任务进行定义后,需要收集与任务相关的用于训练和评估模型的高质量数据集,对数据集进行相应的清洗,以剔除冗余、错误信息,并统一文本格式,而后将文本分解成句子和单词并进行相应的标注,微调过程中根据监督信号不断调整模型权重,使其达到最优化状态。由于大语言模型需要处理和理解超长文档,这要求模型具备足够多的编码器层以进行细致的上下文关联分析,因此,采用基于编码器架构的LLMs或编解码器架构的LLMs将有助于提升模型的性能。对于多模态的文本输入,还需要文本解析器(例如OCR等)进行相应的转换处理;在涉及专业的报告生成时,可结合外部知识库,以确保提高模型输出结果的准确性。

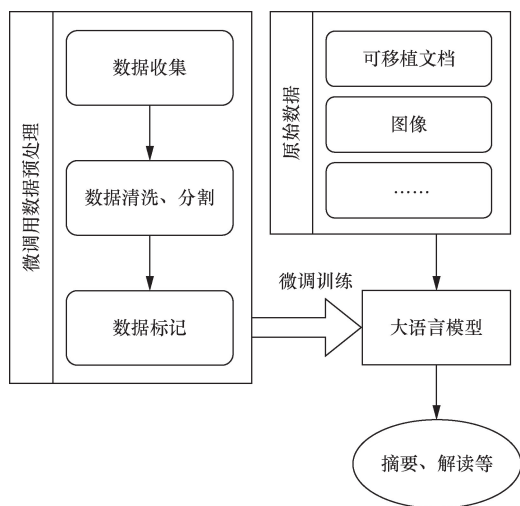


图5 文本摘要类应用场景的技术路线

Fig. 5 Technology roadmap for summary application

## 2.4 分析推理类应用场景及其技术路线

在土木工程项目管理与决策过程中,LLMs展现出了处理复杂数据的强大潜力,包括结构化数据(如数据库表格)和非结构化数据(如文本报告和现场图像)。LLMs能够从这些数据中提炼关键信息,识别潜在的规律与趋势,从而辅助进行逻辑推理和数据分析。这种能力对于预测成本、优化进度安排以及评估项目风险至关重要,有助于提升决策效率与准确性。

LLMs的分析类应用在土木工程领域主要发挥其分析及逻辑推理能力,完成对数据进行分类、标注、排序、计算等一系列任务。例如,Chen等<sup>[39]</sup>构建了一种预训练的多模态混合编解码器(multimodal mixture of Encoder-Decoder, MED)视觉-语言模型,其框架采用一个编码器负责处理视觉数据,一个解码

器负责生成相应的文本描述,模型集成了描述图像中场景的图像标注(image annotation, IC)功能,回答与图像内容相关的问题的视觉问答(visual question answering, VQA)功能;以及将图像内容与相关的文本信息精确匹配的图像-文本检索(image-text retrieval, ITR)功能。该研究还将该模型与增强现实(augmented reality, AR)技术相结合,开发了面向建筑工人的安全辅助系统VCSQ,该系统能实时处理和分析施工现场的图像,提供相应的安全警示和操作指导,从而为识别和检验潜在的安全隐患提供支持。Yoo等<sup>[40]</sup>的研究利用GPT通过迁移学习来预测六种类型的建筑事故,为了增强微调后的GPT模型的可解释性,研究提出了一种显著性可视化方法,能够从非结构化的自由文本数据中识别事故前兆,有助于主动预防建筑事故。在检测六种事故类型方面达到了82%的准确率。

在施工自动化方面,You等<sup>[41]</sup>提出了RoboGPT的框架,该框架利用ChatGPT的推理能力在基于机器人的施工装配中进行自动序列规划,以控制机器人的当前状态、目标、运动及位置等参数,解决了当前机器学习的方法在适应动态施工环境方面存在的局限性。Wang等<sup>[42]</sup>提出了一种基于LLMs的框架,使机器人能够根据现场工人的指令执行装配任务,展示了LLMs在人机协作执行建筑装配任务方面的潜力。

在分析推理类应用场景的研究中,技术路线通常遵循3个核心步骤:预处理问题、确定需求,以及生成结论(图6)。预处理阶段主要从文本中精确提取关键词、实体和属性等关键性信息,为后续的分析步骤提供必要的基础数据和语境环境。而后,大型语言模型针对问题内容的具体特点,识别并设定相应的任务类型,识别并整合必要的辅助信息或知识源,以支持有效和深入的逻辑推理。对于需求领域专业知识的情境,模型将构建一个基于知识库的推理链;对于依赖通用常识或普适逻辑的情境,则会应用如通用的推理分析原则。大语言模型作为分析推理的工具,能对输入的非结构化数据进行标注、提取,按指定的规则使其转化为结构化的数据,并在后续的框架中结合其他技术完成特定的任务。此类研究的大语言模型通常经过专门的微调,缺乏较好的泛化性,只能执行特定的任务。

## 2.5 综合评述

大语言模型在土木工程领域已展现出强大的生命力,且正在深度变革领域的传统技术。

从应用阶段看,基本覆盖全生命期的设计阶段、

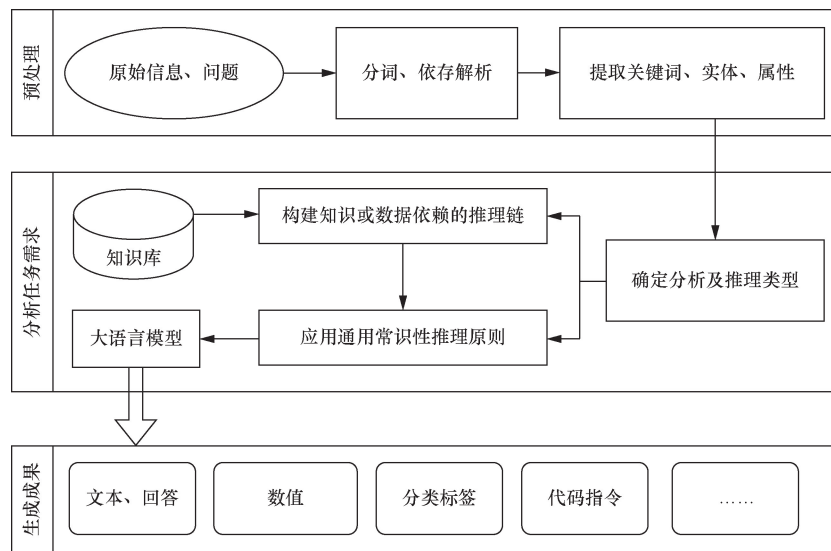


图6 分析推理类应用场景的技术路线

Fig. 6 Technology roadmap for data processing application

招投标阶段及施工阶段。在设计阶段,LLMs 被用于解读复杂的建筑规范条款,回答专业问题,提升设计人员的工作效率;在招投标阶段,LLMs 为合同风险评估提供支持;而在施工阶段,LLMs 根据现场实际生成进度计划、检查报告,并为工人提供安全建议。从中可以看出,LLMs 的应用与实际工作流程的耦合更多处于项目的管理、执行环节等中下游层面。

从结合策略看,越来越多的 LLMs 研究开始融合计算机视觉、机器人技术、多模态数据处理以及数据可视化等技术,这些技术结合 LLMs,实现了更好的环境感知、知识表示和推理,推进专业任务的智能化实现。

从发展趋势看,LLMs 应用的重心也在发生变化:从被动地生成文本,转向主动地提供专业建议和决策支持。这方面的探索主要集中在领域自适应和知识增强上,以提高模型的专业性,这使得大语言模型在建筑信息化中发挥着日益重要的作用。

### 3 大语言模型在土木工程领域面临的挑战

大语言模型在垂直领域的应用面临多种挑战,在使用者方面,使用商业化的大语言模型上传敏感信息将导致潜在的隐私泄露问题或版权问题。在大语言模型方面,除大语言模型在通用领域所面临的幻觉现象<sup>[43]</sup>、泛化性差<sup>[44]</sup>、可解释性差<sup>[45]</sup>、伦理问题<sup>[46]</sup>的不足外,土木工程领域应用大语言模型还存在知识准确性要求高、数据质量及交互性低、信息时效性要求高的挑战(图7)。

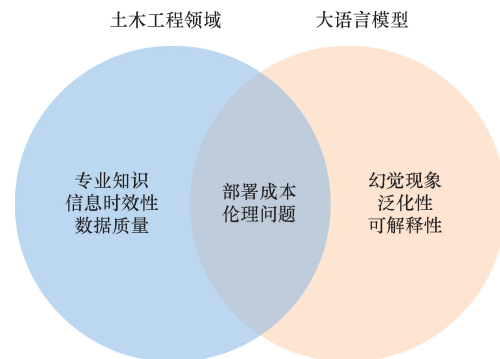


图7 大语言模型在土木工程领域面临的挑战

Fig. 7 Challenges faced by LLMs in the field of civil engineering

#### 3.1 知识准确性要求高

土木工程学科的知识体系跨越多阶段与专精领域,每一分支领域均承载着独特的术语与实践规范,这使得利用 LLM 生成领域内专业内容成为一项复杂且极具探索价值的研究课题。目前有诸多研究聚焦解决知识的获取问题,例如通过本地知识库对大型语言模型实施微调,以及优化提示工程,这些策略能有效遏制模型输出中的虚构信息,生成更有效的信息文本<sup>[47]</sup>,无论采用何种方式,专业知识的整合与嵌入是实现大语言模型在土木工程领域内应用的基础。

为了减轻这些幻觉带来的影响,除了使用专业的领域知识进行增强外,验证预测的模拟测试、对不确定性的持续监控以及引入人工监督 LLMs 的决策<sup>[48]</sup>都是可行的方案。

#### 3.2 信息时效性要求高

由于材料、施工工艺和法律规范在现实中是不

断变化的,存储在模型中的数据在一定时间后若不进行定期更新将会过时,从而提供不可靠的信息。定期根据新数据重新训练模型至关重要,但大模型的规模较为复杂,训练成本高昂。潜在的解决方案包括:简化更新的模块化模型架构、利用知识图谱的可更新性质<sup>[49]</sup>以及迁移学习等轻量级模型适应技术<sup>[50]</sup>。

### 3.3 数据质量及交互性低

数据的可用性和质量一直是人工智能在行业中应用的主要挑战<sup>[51]</sup>。在训练和微调模型时需要标准化、结构化的数据,而土木工程领域中包含信息的数据集有不同的格式,例如 PDF、DOC、HTML、DWG、IFC 等,大多缺乏统一的标准,但大语言模型的训练与微调目前只接受结构化如 JSON、CSV 等格式的数据集。尽管目前有 IFC 等标准化形式提高了数据格式之间的交互性,但由于版本兼容等原因,转换过程中会使数据的丰富性降低,造成一定程度的信息丢失现象<sup>[52]</sup>。

因此,将大语言模型应用于土木工程中的实际场景时,必须确保其生成内容的准确性与专业度,并采取有效措施解决模型在技术伦理上的问题,以更好地推进大语言模型的应用。

## 4 大语言模型在土木工程领域的研究趋势

### 4.1 模型优化层面

基于以上提及的挑战,大语言模型在土木工程行业进行应用研究时,可以考虑从参数高效微调、知识图谱增强、引入人工监督并进行评估 3 个方面进行模型优化研究。

#### 4.1.1 参数高效微调

使用土木工程领域中的设计手册、规范、合同、技术规程和 BIM 等数据对预训练语言模型进行微调是模型应用的基础,然而,在微调可能会带来过度拟合、灾难性遗忘等不利后果<sup>[53]</sup>。通过诸如 QLoRA<sup>[54]</sup>、P-Tuning<sup>[55]</sup>、Adapter Tuning<sup>[56]</sup> 等参数高效微调(parameter-efficient fine-tuning, PEFT)技术,可以提高预训练模型在新任务上的性能,同时减少计算成本和训练时间。

#### 4.1.2 知识图谱增强

通过将大语言模型与知识图谱相互结合,可以结合它们的优势,共同驱动推理过程。现有研究可分为大语言模型构建知识图谱与知识图谱增强大语言模型两部分。

大语言模型增强方面,已有研究利用大语言模型辅助知识图谱的构建<sup>[57]</sup>、补全<sup>[58]</sup>、推理<sup>[59]</sup>流程。

利用大语言模型的文本理解能力,能够从非结构化文本中自动化地构建专业领域知识图谱,相较于传统的人工构建能够大大提升效率。对于不完整的知识图谱,大语言模型可以通过预测缺失的关系或属性的方式补全这些信息。基于现有的知识图谱,大语言模型亦可推断两个实体之间是否存在潜在联系,或是探索不同领域知识之间的隐含关联。这种推理能力可以帮助发现新的知识,为研究、决策等情景提供有价值的信息。

知识图谱增强大语言模型主要有查询-生成<sup>[60]</sup>与生成-查询<sup>[61]</sup>两类策略。查询-生成类策略将知识图谱中的相关结构化知识转化为文本提示输入至大语言模型中,由大语言模型进行推理生成答案。生成-查询类策略由大语言模型解析问题的语义,生成查询检索命令以供知识图谱检索和推理得到答案。通过这种方式,在智能问答类任务中能有效提高模型的性能与可解释性,知识图谱作为独立的模块,易于更新维护的特性也解决了信息时效性的问题。

#### 4.1.3 多模态交互

多模态交互涉及从文本、图像、声音等多种输入模式中理解、推理和生成多种输出模式的过程。在土木工程领域的应用中,根据各阶段的任务不同,可以基于大语言模型,选取如 DALL-E3、Parti、IMAGEN、StyleGAN、Phenaki、Magic3D 等模态转换的模型或模块搭建多模态的实现框架。其实现框架如图 8 所示,各个模态的原始信息经过相应的编码器转化为对应的语义向量,进入大语言模型中完成任务,并经由文本解码器或其他模态的解码器生成相应的内容。

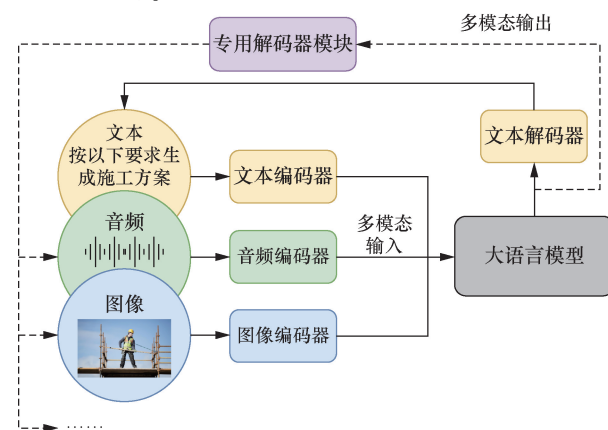


图 8 多模态大语言模型交互框架

Fig. 8 Multimodal LLMs interaction framework

#### 4.1.4 引入人工监督并进行评估

大语言模型并不是一个完美的黑盒,仍然需要人工监督来验证输出的质量和准确性,将 LLMs 生成与人类判断相结合的人机交互方法可以兼顾两

者的优势。同时仍应设计实验,使用可以量化的关键指标来评估 LLMs 产生的影响,诸如生产力、成本、时间、风险等,这有助于量化 LLMs 投资为企业带来的收益。

## 4.2 应用场景层面

大语言模型与土木工程领域的融合应用前景广阔,未来土木工程领域基于大语言模型的研究主要可以基于前文提到的四类任务应用场景进行扩展。

### 4.2.1 内容生成类

在土木工程领域大量基于模板的工作任务中,大语言模型可以发挥其能力,大大提升从业人员的生产效率。例如,根据模板快速生成可行性报告、工程合同、概预算书、施工方案及交底资料等文档,极大提高文件制作的效率;也可以通过模态转换将各类数据信息以直观的图表、3D 模型等形式呈现,提高从业者决策的效率;还可以按设计师的要求自动生成多种设计理念的文字与简图,提高业主单位与设计单位的沟通效率。

### 4.2.2 智能问答类

在项目事实的各个阶段,信息的交互至关重要,未来研究可以专注于如何结合问答系统和可视化工具改善从业者与数据、从业者之间的交互效率,实现知识积累与流动。例如,工程咨询问答系统中,语言模型可以对用户提出的有关工程设计、施工、计量、合同、招标等专业问题给出解答,极大地拓展了工程技术人员获取咨询意见的渠道;行业中现有的规范标准众多且冗杂,通过规范标准问答系统,即可快速定位相关内容并做出解释,展示相关条文的联系;在运维阶段,可创建具备知识管理的运维知识问答系统,提供设备运维所需的操作技能、维修指导、故障处理等知识,为维修人员提供一定的参考。此外,可以探讨集成语音识别和计算机视觉技术,以支持更自然、直观的交互方式,如通过语音指令操作建模软件或使用增强现实(AR)技术辅助现场工作。

### 4.2.3 文本摘要类

在土木工程领域,信息提取关键在于从海量的非结构化数据中(如工程日志、现场照片、设计文档等)提取出关键的信息,例如工期进度、材料消耗量、成本开支等。未来研究可以聚焦于开发和训练大语言模型,使其能够理解和处理行业特有的术语和数据结构范式,自动化地提取和归类这些关键信息,以便直接用于进一步的数据分析和决策。

### 4.2.4 推理分析类

大语言模型可以通过其强大的推理分析能力,自动化或半自动化地协助从业者完成一些复杂任

务,例如,合规性审查存在于各生命周期阶段,复杂而烦琐,通过整理和预处理土木工程领域的相关法规、条例文档,大语言模型可被用于自动地审查设计、施工文件,并发现其中不符合审查条例的问题。此外,在该研究方向上也可进行多模态的扩展:结合 BIM 数据和规范数据库,大语言模型可以分析设计的具体参数,并与相应的条例要求进行比较;利用无人机对施工现场进行检查,将图像数据与检查条例进行比对,生成检查结果报告,以提前避免潜在的问题。

综合以上应用场景来看,未来的研究方向将主要围绕着信息化、智能化方向开展,利用大语言模型相关的技术处理各种与自然语言相关的任务可以使项目提高效率,降低成本。理解和应用大语言模型在土木工程领域的能力将是未来技术的重要发展方向之一。

## 5 结论

从大语言模型爆发式增长的背景下出发,通过文献回顾的方式,探讨了大语言模型在土木工程行业内应用的现状、挑战及未来可能研究的方向。结论如下。

(1)大语言模型在垂直领域的主要应用任务可分为内容生成类、智能问答类、文本摘要类和分析推理类四大方向。

(2)大语言模型在土木工程领域中设计、招投标、施工和运营阶段均存在应用的场景及潜力,在需要处理自然语言文本的任务中表现出色,并结合其他技术构建多模态的语言模型,实现更广泛的应用。

(3)大语言模型在土木工程领域应用面临着诸多挑战,主要包括知识准确性要求高、数据质量及交互性低、信息时效性要求高,以及大语言模型本身存在的幻觉和准确性、泛化性和可解释性等问题。

(4)在模型优化层面,未来研究趋势包括:参数高效微调、与知识图谱结合、融合多模态数据、开发适用于模型的评估方法。在应用场景层面,未来研究可以拓展 LLMs 在文档生成、问答系统、信息抽取、合规性审查等复杂任务中的应用,提高从业者与数据间的交互效率。

全面信息化将是土木工程行业未来发展的大方向,大语言模型作为一种新兴的通用型人工智能技术,已经展示了其巨大的潜力,该综述为认识大语言模型在该行业的能力和提供了一定的文献基础,未来如何开发大语言模型具体在土木工程行业应用的研究将是重点所在。

参 考 文 献

- [1] Soibelman L, Wu J, Caldas C, et al. Management and analysis of unstructured construction data types[J]. *Advanced Engineering Informatics*, 2008, 22(1): 15-27.
- [2] Inmon W H. *Data architecture: a primer for the data scientist*[M]. Boston: Morgan Kaufmann, 2014.
- [3] Ur-Rahman N, Harding J A. Textual data mining for industrial knowledge management and text classification: a business oriented approach[J]. *Expert Systems with Applications*, 2012, 39(5): 4729-4739.
- [4] Moens M F. *Information extraction: algorithms and prospects in a retrieval context* [M]. Heidelberg: Springer Science & Business Media, 2006.
- [5] 刘晓波, 孔屹刚, 李涛, 等. 采煤机调高泵隐半马尔可夫模型磨损故障预测[J]. *科学技术与工程*, 2020, 20(29): 11980-11986.  
Liu Xiaobo, Kong Yigang, Li Tao, et al. Research on wear fault prognostics of hidden semi-Markov model of shearer pump[J]. *Science Technology and Engineering*, 2020, 20(29): 11980-11986.
- [6] 杨程, 颜海泉, 董正方. 基于 K 近邻算法的钢筋混凝土柱地震破坏模式判别方法[J]. *科学技术与工程*, 2023, 23(25): 10910-10917.  
Yang Cheng, Yan Haiquan, Dong Zhengfang. Seismic failure mode identification method of reinforced concrete columns based on KNN algorithm[J]. *Science Technology and Engineering*, 2023, 23(25): 10910-10917.
- [7] 董国鹏, 徐旭升. 建筑安全事故通告关键信息自动提取方法[J]. *科学技术与工程*, 2022, 22(10): 4026-4032.  
Dong Guopeng, Xu Xusheng. Automatic extraction method of key information in construction safety accident notification[J]. *Science Technology and Engineering*, 2022, 22(10): 4026-4032.
- [8] 李舟军, 范宇, 吴贤杰. 面向自然语言处理的预训练技术研究综述[J]. *计算机科学*, 2020, 47(3): 162-173.  
Li Zhoujun, Fan Yu, Wu Xianjie. Survey of natural language processing pre-training techniques[J]. *Computer Science*, 2020, 47(3): 162-173.
- [9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. *ArXiv*, 2017; 1706.03762.
- [10] Xu Y, Zhou Y, Sekula P, et al. Machine learning in construction: from shallow to deep learning[J]. *Developments in the Built Environment*, 2021, 6: 100045.
- [11] Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models[J]. *ArXiv Preprint ArXiv*, 2022; 2206.07682.
- [12] Guu K, Lee K, Tung Z, et al. REALM: retrieval augmented language model pre-training[J]. *ArXiv*, 2020; 2002.08909.
- [13] Carlini N, Tramer F, Wallace E, et al. Extracting training data from large language models[J]. *ArXiv*, 2021; 2012.07805.
- [14] Zhao W X, Zhou K, Li J, et al. A survey of large language models[J/OL]. *ArXiv Preprint ArXiv*, 2023; 2303.18223.
- [15] Yao L, Mao C, Luo Y. KG-BERT: BERT for knowledge graph completion[J]. *ArXiv*, 2019; 1909.03193.
- [16] Hakala K, Pyysalo S. Biomedical named entity recognition with multilingual BERT [C]//*Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*. Hong Kong: IEEE, 2019; 56-61.
- [17] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. *Journal of Machine Learning Research*, 2020, 21(140): 1-67.
- [18] Xue L, Constant N, Roberts A, et al. mT5: a massively multilingual pre-trained text-to-text transformer[J]. *ArXiv*, 2021; 2010.11934.
- [19] Brown TomB, Mann B, Ryder N, et al. Language models are few-shot learners[J]. *ArXiv*, 2020; 2005.14165.
- [20] Robinson J, Jegelka S, Sra S. Strength from weakness: fast learning using weak supervision [C]//*International Conference on Machine Learning*. Online: PMLR, 2020; 7780-8550.
- [21] Sanh V, Webson A, Raffel C, et al. Multitask prompted training enables zero-shot task generalization [C]//*International Conference on Learning Representations*. Online: ICLR, 2021.
- [22] Chung S, Moon S, Kim J, et al. Comparing natural language processing (NLP) applications in construction and computer science using preferred reporting items for systematic reviews (PRISMA) [J]. *Automation in Construction*, 2023, 154: 105020.
- [23] Prieto S A, Mengiste E T, García de Soto B. Investigating the use of ChatGPT for the scheduling of construction projects[J]. *Buildings*, 2023, 13(4): 857.
- [24] Pu H, Yang X, Li J, et al. AutoRepo: a general framework for multi-modal LLM-based automated construction reporting [J]. *ArXiv Preprint ArXiv*, 2023; 2310.07944.
- [25] Prabhu P, Athavale A A, Singh V. Development of an automated report generator using LLMs and storytelling frameworks to support broadcasting in construction projects[J]. *Automation in Construction*, 2022, 145: 104-115.
- [26] Zheng J, Fischer M. Dynamic prompt-based virtual assistant framework for BIM information search [J]. *Automation in Construction*, 2023, 155: 105067.
- [27] 覃思中, 郑哲, 顾焱, 等. 大语言模型在建筑工程中的应用测试与讨论[J]. *工业建筑*, 2023, 53(9): 162-169.  
Qin Sizhong, Zheng Zhe, Gu Yi, et al. Exploring and discussion on the application of large language models in construction engineering[J]. *Industrial Architecture*, 2023, 53(9): 162-169.
- [28] 李培源. 智能化技术在建筑工程档案管理中的应用[J]. *智能城市*, 2024, 10(4): 87-89.  
Li Peiyuan. The application of Intelligent technology in the management of construction engineering archives [J]. *Smart City*, 2024, 10(4): 87-89.
- [29] Wong S, Zheng C, Su X, et al. Construction contract risk identification based on knowledge-augmented language models [J]. *Computers in Industry*, 2024, 157: 104082.
- [30] Lee J, Jung W, Baek S. In-house knowledge management using a large language model: focusing on technical specification documents review[J]. *Applied Sciences*, 2024, 14(5): 2096.
- [31] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 9459-9474.
- [32] Logan R, Liu N F, Peters M E, et al. Barack's wife hillary: using knowledge graphs for fact-aware language modeling [J]. *ArXiv*, 2019; 1906.07241.
- [33] Tan K, Pang T, Fan C. Towards applying powerful large AI models in classroom teaching: opportunities, challenges and prospects [J]. *ArXiv Preprint ArXiv*, 2023; 2305.03433.
- [34] Liu B, Jiang Y, Zhang X, et al. LLM + P: empowering large

- language models with optimal planning proficiency[J]. ArXiv Preprint ArXiv, 2023: 2304.11477.
- [35] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837.
- [36] Li B, Wang R, Guo J, et al. Deliberate then generate: enhanced prompting framework for text generation[J]. ArXiv Preprint ArXiv, 2023: 2305.19835.
- [37] Zheng Z, Chen K Y, Cao X Y, et al. LLM-FuncMapper: function identification for interpreting complex clauses in building codes via LLM[J]. ArXiv Preprint ArXiv, 2023: 2308.08728.
- [38] Wong S, Zheng C, Su X, et al. Construction contract risk identification based on knowledge-augmented language model[J]. ArXiv Preprint ArXiv, 2023: 2309.12626.
- [39] Chen H, Hou L, Wu S, et al. Augmented reality, deep learning and vision-language query system for construction worker safety[J]. Automation in Construction, 2024, 157: 105158.
- [40] Yoo B, Kim J, Park S, et al. Harnessing generative pre-trained transformers for construction accident prediction with saliency visualization[J]. Applied Sciences, 2024, 14(2): 664.
- [41] You H, Ye Y, Zhou T, et al. Robot-enabled construction assembly with automated sequence planning based on ChatGPT: RoboGPT[J]. ArXiv Preprint ArXiv, 2023: 2304.11018.
- [42] Wang M, Li Y, Li S. Robotic assembly of interlocking blocks for construction based on large language models[J]. Construction Research Congress, 2024(3): 777-786.
- [43] Bang Y, Cahyawijaya S, Lee N, et al. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity[J]. ArXiv Preprint ArXiv, 2023: 2302.04023.
- [44] Fui-Hoon Nah F, Zheng R, Cai J, et al. Generative AI and ChatGPT: applications, challenges, and AI-human collaboration[J]. Journal of Information Technology Case and Application Research, 2023, 25(3): 277-304.
- [45] Zini J E, Awad M. On the explainability of natural language processing deep models[J]. ACM Computing Surveys, 2022, 55(5): 1-31.
- [46] Patton D U, Landau A Y, Mathiyazhagan S. ChatGPT for social work science: ethical challenges and opportunities[J]. Journal of the Society for Social Work and Research, 2023, 14(3): 553-562.
- [47] Wang X, Kapanipathi P, Musa R, et al. Improving natural language inference using external knowledge in the science questions domain[J]. ArXiv, 2019: 1809.05724.
- [48] Meskó B, Topol E J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare[J]. npj Digital Medicine, 2023, 6(1): 120.
- [49] Saka A, Taiwo R, Saka N, et al. GPT models in construction industry: opportunities, limitations, and a use case validation[J]. ArXiv Preprint ArXiv, 2023: 2305.18997.
- [50] Liu Y, Yang Z, Yu Z, et al. Generative artificial intelligence and its applications in materials science: current situation and future perspectives[J]. Journal of Materiomics, 2023, 9(4): 798-816.
- [51] Abioye S O, Oyedele L O, Akanbi L, et al. Artificial intelligence in the construction industry: a review of present status, opportunities and future challenges[J]. Journal of Building Engineering, 2021, 44: 103299.
- [52] Saka A B, Oyedele L O, Akanbi L A, et al. Conversational artificial intelligence in the AEC industry: a review of present status, challenges and opportunities[J]. Advanced Engineering Informatics, 2023, 55: 101869.
- [53] Chen S, Hou Y, Cui Y, et al. Recall and learn: fine-tuning deep pretrained language models with less forgetting[J]. ArXiv Preprint ArXiv, 2020: 2004.12651.
- [54] Dettmers T, Pagnoni A, Holtzman A, et al. QLoRA: efficient finetuning of quantized LLMs[J]. ArXiv, 2023: 2305.14314.
- [55] Liu X, Ji K, Fu Y, et al. P-tuning v2: prompt tuning can be comparable to fine-tuning universally across scales and tasks[J]. ArXiv Preprint ArXiv, 2021: 2110.07602.
- [56] Houlisby N, Giurugi A, Jastrzebski S, et al. Parameter-efficient transfer learning for NLP[J]. ArXiv, 2019: 1902.00751.
- [57] 徐春, 苏明钰, 孙彬. 基于 ChatGLM 和提示微调的旅游知识图谱构建[J]. 科学技术与工程, 2024, 24(31): 13484-13492.
- Xu Chun, Su Mingyu, Sun Bin. Tourism knowledge graph construction based on ChatGLM and prompt-tuning[J]. Science Technology and Engineering, 2024, 24(31): 13484-13492.
- [58] Zhang Y, Chen Z, Guo L, et al. Making large language models perform better in knowledge graph completion[C]//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne: ACM, 2024: 233-242.
- [59] Luo L, Ju J, Xiong B, et al. Chatrule: mining logical rules with large language models for knowledge graph reasoning[J]. ArXiv Preprint ArXiv, 2023: 2309.01538.
- [60] Sun J, Xu C, Tang L, et al. Think-on-graph: deep and responsible reasoning of large language model with knowledge graph[J]. ArXiv Preprint ArXiv, 2023: 2307.07697.
- [61] Luo H, Tang Z, Peng S, et al. ChatKBQA: a generate-then-retrieve framework for knowledge base question answering with fine-tuned large language models[J]. ArXiv Preprint ArXiv, 2023: 2310.08975.