

# 大语言模型与数字孪生水利融合 应用场景分析探讨

李 意<sup>1</sup> 顾小林<sup>2</sup> 冯贤林<sup>1</sup> 刘心怡<sup>1</sup> 黄 瑞<sup>1</sup>

(1.中国电建集团贵阳勘测设计研究院有限公司,贵州 贵阳 550000;

2.贵州省水文水资源局,贵州 贵阳 550000)

**[摘要]** 人工智能大语言模型 (LLM) 正在赋能千行百业实现智慧化提升,数字孪生水利作为新阶段智慧水利的 implements 措施,与大模型深度融合发展是应有之义。然而,如何将数字孪生与 LLM 深度融合,推动水利高质量发展,亟需明晰 LLM 与数字孪生水利相结合的应用场景,进一步推进数字孪生水利建设。本文重点选取了数据底板构建、模型平台建设、知识平台建设等数字孪生水利的重要部分进行探讨,分析了 LLM 如何赋能满足不同水利业务场景的数据需求,如何解决模型平台通用性不足、多种类模型配置技术缺乏、复杂场景模拟能力不足、多业务场景动态支撑能力缺失等问题,以及如何建立水专家 LLM 实现知识平台的动态更新和实时响应。

**[关键词]** 大语言模型;数字孪生水利;数据底板;模型平台;知识平台

**[中图分类号]** TP391.9; TV213.4

**[文献标识码]** B

## Analysis and discussion of application scenarios of integration of large language model and digital twin water conservancy

Li Yi<sup>1</sup>, Gu Xiaolin<sup>2</sup>, Feng Xianlin<sup>1</sup>, Liu Xinyi<sup>1</sup>, Huang Rui<sup>1</sup>

(1.China Power Construction Corporation Guiyang Survey, Design & Research Institute Co., Ltd.,

Guiyang 550000, China; 2.Guizhou Provincial Bureau of Hydrology and Water Resources,

Guiyang 550000, China)

**Abstract:** Large language models (LLM) in artificial intelligence are empowering various industries to achieve intelligent upgrades. As an implementation measure for the new stage of smart water conservancy, the deep integration of digital twin system for water resources management with LLM is an inevitable trend. However, for how to deeply integrate digital twin system for water resources management with LLM to promote high-quality development in water conservancy, it is urgent to clarify the application scenarios of the combination of LLM and digital twin system for water resources management, so as to guide the further advancement of digital twin system for water resources management construction. This article focuses on discussing important parts of digital twin water conservancy such as the construction of data baseplates, model platform construction, and knowledge platform construction, analyzing how LLM can empower to meet the data demands of different water conservancy business scenarios, how to solve problems such as insufficient universality of model platforms, lack of configuration technologies for multiple types of models, insufficient complex scenario simulation capabilities, and the absence of dynamic support capabilities for multiple business scenarios, as well as how to establish a water expert large model to achieve dynamic updates and real-time responses of the knowledge platform.

**Keywords:** Large language model; digital twin system for water resources management; data baseplates; model platform; knowledge platform

**[收稿日期]** 2025-05-23

**[基金项目]** 基于分级中心化的大型能源基础设施建设工程时空数据服务研究(黔科合支撑[2024]一般 138)。

**[作者简介]** 李 意(1993-),男,贵州修文人,硕士,工程师,主要从事数字孪生水利研究。

## 0 引言

近年来,水利部大力推进数字孪生水利建设,发布了《数字孪生流域建设技术大纲(试行)》《数字孪生水网建设技术导则(试行)》等十余项导则、技术指南,在全国范围内先后组织了数字孪生流域、数字孪生水利工程、数字孪生灌区等先行先试工作。总体上,数字孪生水利建设取得明显成效,数字孪生水利框架体系构建基本完成。尽管在数字孪生水利建设深入推进中,成果不断涌现,其在实践中距离水利部提出的“需求牵引、应用至上、数字赋能、提升能力”总要求依然存在较大差距。推动各类应用赋能提升水利行业数字化、网络化、智能化水平是建设数字孪生水利的根本目标,然而孪生应用场景广度不够、深度不足<sup>[1]</sup>。

目前,应用在数字孪生水利中的新一代信息技术主要有机器学习、物联网、大数据等,然而机器学习时常表现出过拟合、欠拟合等泛化能力差的现象,物联网设备对于所生成的非结构化数据难以获取有效语义信息,大数据存储与分析技术门槛高,易用性和普适性不强。未来以 LLM 为水利大脑,利用 LLM 的数据分析能力、模拟预测能力、逻辑推理能力和智能决策能力,有效解决机器学习、物联网、大数据在数字孪生水利中的应用短板,整体实现并放大数字孪生水利的应用价值具有巨大潜力。钱峰等<sup>[2]</sup>从水利大模型的角度,研究了水利大模型的构建思路与总体框架,分析了需突破的关键技术,介绍了行业应用场景。李巍等<sup>[3]</sup>基于已于知识库训练大语言模型,实现了水利知识检索和智能问答。杨柳等<sup>[4]</sup>针对水利政府网站中公共服务机械、低效问题,引入 LLM 技术和理念,设计构建了一套以封装后的智能框架平台为服务支撑,以多元化的应用服务实现各业务流程,以多维度数据为资源存储的水利智能公共服务体系。梁思涵等<sup>[5]</sup>从数字孪生平台使用、构建和公众科普三方面对 GPT 预训练模型在水利行业的应用进行了探讨,简述了 GPT 在水利行业应用的风险和局限性,包括信息安全、信息误导和成本高昂方面。Ren 等<sup>[6]</sup>利用水文知识,构建水文知识库,结合微调等手段,训练了水文专家大模型,取得了良好的应用成效。综合来看,LLM 在水利行业的场景应用尚处于起步阶段。因此,为充分利用 LLM 能力,赋能

提升数字孪生水利发挥效益,急需对 LLM 与数字孪生水利深度融合应用场景进行分析研究,为数字孪生水利落地提供辅助。

## 1 大语言模型概述

### 1.1 大语言模型简介

LLM 的发展源于深度学习技术的进步,特别是 Transformer 架构的提出<sup>[7]</sup>。这种架构采用自注意力机制,能够高效处理长距离依赖关系,为模型学习复杂语言模式奠定了基础。随着模型规模的不断增加,从最初的数百万参数到当今的上万亿参数,LLM 展现出“涌现能力”,即模型在训练过程中自发产生了超越训练目标的智能特质,主要包括少样本或零样本学习、上下文学习、多模态学习、复杂推理、创造性生成等能力。国内主流 LLM 有文心一言、通义千问、豆包、智谱、盘古大模型、DeepSeek 等系列,国外主流 LLM 有 GPT、Gemini、Claude、Grok、LLaMA 等系列。截止 2024 年第二季度,全球的 LLM 企业超 3 万家,中国占全球的 15%,位居全球第二位<sup>[8]</sup>。

### 1.2 大语言模型应用方式

大语言模型通常有 RAG、Agent、工作流、微调、预训练等应用方式<sup>[9]</sup>。多数情况下选择其中一种方式就可以满足要求,部分场景需要多种方式组合,才能解决实际问题<sup>[10]</sup>。LLM 通过整合这些方式,构建了一个完整的 AI 系统,能够处理复杂的任务并适应不同的应用场景。LLM 的不同应用方式如表 1 所示。

大模型结合数字孪生水利领域内的新闻报道已经逐步显现,但截止目前,明确文献探讨如何结合的文献较少。RAG 的应用有水利知识检索和智能问答<sup>[3]</sup>;Agent 的应用有基于 Agent 实现智能公共服务体系<sup>[4]</sup>,以及利用 agent 理解用户的自然语言提问,驱动大模型执行水利一张图任务<sup>[11]</sup>;工作流与微调与数字孪生水利的结合暂时未见文献总结,但在电力、媒体、法律等领域已有应用<sup>[12-14]</sup>;预训练方式主要是大模型企业训练基座模型较多,目前在垂直行业领域应用较鲜见。

## 2 大模型与数字孪生水利融合

数字孪生平台由数据底板、模型平台、知识平

表 1		LLM 的不同应用方式	
应用方式	定义	作用	
RAG	RAG 是一种结合检索(Retrieval)和生成(Generation)的技术, 通过从外部知识库中检索相关信息来增强生成模型的能力。	在问答系统、内容生成等任务中,RAG 能够提供更准确、更丰富的信息输出。	
Agent	Agent 是一个能够自主决策、执行任务并与环境交互的智能体。它通常基于强化学习(RL)或规则引擎驱动。	在多步骤任务中,Agent 可以规划、执行和优化任务流程,如自动化客服、机器人控制或多轮对话系统。	
工作流	工作流是一种任务管理和流程自动化技术,用于定义、执行和监控复杂任务的步骤和逻辑。	在 AI 系统中,工作流用于协调多个模块,确保任务的高效执行。	
微调	微调是指在预训练模型的基础上,使用特定领域的数据对模型进行进一步训练,以提高其在特定任务上的性能。	使通用模型适应特定场景(如医疗、法律、金融等领域),提升模型的准确性和实用性。	
预训练	预训练是指在大规模通用数据集上训练模型,使其学习到广泛的语言、视觉或其他领域的知识。	为模型提供强大的泛化能力,作为后续微调和任务适配的基础。	

台等构成<sup>[15]</sup>,是为数字孪生水利提供算据、算法的核心,也是数字孪生流域、数字孪生水利工程、数字孪生水网等的核心组成部分。因此,研究探讨 LLM 与数字孪生平台的深度融合,对在数字孪生水利领域充分发挥 LLM 能力具有重要意义。

2.1 数据底板构建

数字孪生水利建设的首要任务是构建支撑有力的数据底板,然而数据底板具备多源异构、空间分散、部分重要数据实时更新性弱等特点,面对不同水利业务场景的数据需求,数据来源不统一、数据属性分散不全面、数据实时性不足等问题频发<sup>[16]</sup>。

针对数据来源不统一,通过利用 LLM 的自然语言处理(NLP)能力,开发智能数据解析器,自动识别和解析不同来源、不同格式的数据(如文本、表格、图像、传感器数据等),并将其转换为统一的标准化格式。针对数据属性不全面,利用 LLM 构建水利数据的关联网路;再通过 LLM 的语义理解能力,自动识别不同数据源之间的关系,整合数据所有属性,实现跨系统的数据融合,提升数据的整体可用性。针对数据实时性不足,通过 LLM 的时序预测能力,结合历史数据和实时数据,预测缺失或滞后的数据。同时,LLM 可以结合外部公开卫星遥感,动态连接下载,利用 LLM 的数据处理分析能力,实时更新数据底板。

2.2 模型平台研发

模型平台是数字孪生水利的重要组成部分,为上层“2+N”业务应用提供算法支撑,是决定数字孪生水利“四预”(预报、预警、预演、预案)功能是否实用、管用的核心。尽管国内已经具备种类繁多的水利专业模型,但是依然存在模型通用性不足、

多种类模型配置技术缺乏、复杂场景模拟能力不足、多业务场景动态支撑能力缺失等问题<sup>[17]</sup>。

对于模型通用性不足,传统研究思路是开发接口标准统一的水利专业模型,基于 LLM 可考虑在不改造已有水利专业模型的情况下,构造“统一语义接口”,用户通过自然语言描述业务需求,LLM 自动解析需求,自动将用户输入参数转换为各子模型所需格式,解决模型接口异构问题,进一步匹配并调用底层适配的水利专业模型组合,同时也解决了多种类模型配置的难题。对于复杂场景模拟能力不足,可以利用 LLM 的生成能力快速构建各类复杂场景,如极端降雨情况、水库群联合调度时各类自然地理约束条件、地质灾害等,而后以 LLM 生成的场景为基础数据,训练水利专业模型,增强水利专业模型的模拟能力。对于多业务场景动态支撑能力缺失,利用 LLM 的多任务实时学习和动态调整能力,使 LLM 能够同时处理洪水预警、水资源调度、生态保护等多种业务需求,例如在预报未来水库洪水来临之际,综合考虑水库水资源的充分利用、水库及下游防洪对象安全、沿岸生态保护等条件,动态组合多类模型协同模拟,提供动态决策支持。

2.3 知识平台建设

知识平台将水利领域的专业知识、历史经验、调度预案等数据系统化存储,形成可共享的知识库,推进水利知识的传承和复用,是积累、挖掘、利用水利知识的重要承载。当前构建知识平台的主要技术路线是知识图谱,但由于水利实体的多样性、命名不规范等现状造成水利实体自动识别难,实体属性和实体之间的关系复杂多样、动态变化,导致水利关系和实体属性难以快速全面建立且实



时响应变化。

利用 LLM 的预训练语言理解能力,结合水利领域的专业语料进行微调,建立理解水利行业的“水专家”大模型。通过水专家大模型的命名实体识别(NER)能力,可以很准确地从非结构化文本中提取水利实体,进一步利用 LLM 语义理解能力,对识别出来的实体进行归一化处理,如将“长江”和“扬子江”映射到同一实体。通过水专家 LLM 的关系和属性抽取技术,实现从文本中提取水利实体间的关系和水利实体的属性,结合 LLM 的增量学习和实时推理能力,实现知识平台的动态更新和实时响应。

### 3 结论

“千模大战”的背景下,新的 LLM 依旧不断发布,LLM 的技术发展与应用仍然在更新迭代,尚未明确出现一种可以占据主流地位且持续稳定的业务融合模式,但是 LLM 赋能千行百业的趋势已经不可阻挡。数字孪生水利作为新阶段智慧水利的 implements 措施,与 LLM 深度融合发展是大势所趋。本文重点选取了数据底板构建、模型平台建设、知识平台建设等数字孪生水利的重要部分进行探讨,分析了 LLM 如何赋能满足不同水利业务场景的数据需求,如何解决模型平台通用性不足、多类型模型配置技术缺乏、复杂场景模拟能力不足、多业务场景动态支撑能力缺失等问题,以及如何建立水专家大模型实现知识平台的动态更新和实时响应。

然而,LLM 的应用也面临着一定的风险,如 LLM 幻觉问题,LLM 在自己并不精通的问题上,偶尔表现出答非所问、胡编乱造,甚至是自圆其说的“伪专家”,该问题的解决依赖于模型本身的推理纠错能力,更依赖于模型的专业训练,这对训练数据和模型训练人员要求较高。其次,LLM 存在数据泄露问题,尽管现有诸多 LLM 私有化部署方案,但模型本身并不具备识别用户权限的能力,可能会按照被接收到的系统指令和用户提示,泄露本不该输出的数据。再如 LLM 的可解释性,尽管当前如

DeepSeek-R1 模型推理能力十分强大,但是在水利专业领域,LLM 依然类似黑箱模型,对专业词汇、专业场景的理解还有很大提升空间。□

#### 参考文献:

- [1] 谢明霞. 数字孪生水利内涵及应用场景研究 [J]. 人民长江, 2024, 55(2): 245-251+264.
- [2] 钱峰, 成建国, 夏润亮, 丁昱凯, 谢文君, 陆佳民, 李冰. 水利大模型的建设思路、构建框架与应用场景初探[J]. 中国水利, 2024(9): 9-19.
- [3] 李巍, 李树元. 基于大语言模型和知识图谱的数字孪生流域知识平台设计[J]. 海河水利, 2023(12): 75-78.
- [4] 杨柳, 姚薇, 马辉, 李珊珊, 翁春元, 曹蒙. LLM 在水利政府网站公共服务中的应用研究[J]. 水利信息化, 2024(2): 58-62.
- [5] 梁思涵, 李涛, 宾身堂, 等. GPT 预训练模型在水利行业应用进展 [C]//水利部防洪抗旱减灾工程技术研究中心,《中国防汛抗旱》杂志社,中国水利学会减灾专业委员会.第十四届防汛抗旱信息化论坛论文集.黄河水利委员会黄河水利科学研究院,河南省智慧水利工程技术研究中心, 2024: 62-70. DOI:10.26914/c.cnkihy. 2024.003860.
- [6] Ren Y, Zhang T, Dong X, Li W, Wang Z, He J, Zhang H, Jiao L. WaterGPT: Training a Large Language Model to Become a Hydrology Expert. Water, 2024, 16: 3075.
- [7] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]// The 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017: 6000-6010.
- [8] 中国信息通信研究院, 全球数字经济发展研究报告 (2024 年) [R]. (2025-01) [2025-03-31]
- [9] 韩炳涛, 刘涛. 大模型关键技术与应用 [J]. 中兴通讯技术, 2024, 30(2): 76-88.
- [10] 刘威辰, 杨华锋, 江民氏, 刘敬谦. 大模型应用服务平台建设研究 [J]. 信息技术与政策, 2024, 50(12): 21-30.
- [11] 明晨曦, 杨鹏, 张志鑫, 等. 基于水利一张图的地理空间信息问答智能体技术 [J]. 人民长江, 2025(4): 1-13.
- [12] 牛泽原, 李嘉媚, 艾芊. 大语言模型在电力系统中的应用初探 [J]. 电网技术, 2025(4): 1-12.
- [13] 史永恒. 基于开源大模型的短视频与直播 AI 实训系统的设计与实现 [J]. 信息技术与信息化, 2024(8): 195-198.
- [14] 王劲滔, 孟琪翔, 高志霖, 等. 基于大语言模型指令微调的案件信息要素抽取方法研究 [J]. 计算机科学与探索, 2025(4): 1-17.
- [15] 谢文君, 李家欢, 李鑫雨, 等.《数字孪生流域建设技术大纲(试行)》解析 [J]. 水利信息化, 2022(4): 6-12.
- [16] 崔培, 张涛, 曾斌, 等. 数字孪生水利动态时空数据底板构建研究 [J]. 中国水利, 2025(2): 52-64.
- [17] 罗斌, 周超, 张振东. 数字孪生水利专业模型平台构建关键技术及应用 [J]. 人民长江, 2024, 55(6): 227-233.