

地整治的管理体系相衔接,因地制宜推行政府与村集体、农民用水合作组织、社会化服务主体等共同参与管护的机制,加强基层农田水利工程管理机构和队伍建设,鼓励各地探索成立农田水利管理中心、用水管理站等专门机构,将管护责任落实到田间地头。

在全域土地整治过程中,应完善农田水利工程产权管理制度,明晰所有权、使用权和收益权,建立工程维修养护基金,合理确定供水价格,将工程运行管护费用纳入水价,形成“谁用水、谁缴费、谁管理”的良性运行机制,制定农田水利工程设施维修养护技术规程,强化设施的经常性管护,及时修复损毁设施,确保工程完好率和利用率。

利用现代信息技术提升农田水利工程管理水平,推行信息化、精细化、智能化管理,加强灌溉管理信息系统建设,实现水量调度、用水监测、水费计收、设施运行等环节的信息化管控,提高农业用水效率,推广喷灌、微灌、管道输水等节水灌溉技术,发展精准灌溉、智能灌溉,实现农田水分的精准调控和资源优化配置<sup>[3]</sup>。

#### 4.4 推进农田水利与乡村产业融合发展,拓展综合效益

在全域土地整治背景下,农田水利工程建设应当与产业发展规划相衔接,为乡村产业布局提供支撑,统筹考虑灌溉、排水、蓄水、引水、供水等功能,优化乡村产业发展布局,促进农林牧渔等产业融合发展,依托农田水利工程建设农业产业园区,大力发展设施农业、林下经济、生态旅游等特色产业,打造现代农业产业集群,提升农业综合效益,支持龙头企业、农民合作社等主体围绕农田水利工程开展农产品加工、仓储物流、乡村旅游等二三产业,延长农业产业链,提升农产品附加值,拓宽农民就业增收渠道。

全域土地整治强调城乡融合发展,农田水利工程也应当发挥连接城乡的作用,要发挥农田水利工程在水土保持、水源涵养、生物多样性保护等方面的生态功能,统筹兼顾生态效益和经济效益,加强水生态环境评估和监测,严格用水定额管理,维护河湖生态流量,防止过度用水损害生态环境,在农田水利工程建设中融入海绵城市理念,采取“渗、滞、蓄、净、用、排”等措施,加强雨洪资源利用,维护水生态平衡。推进废弃农田水利设施生态化改造,因地制宜营造湿地、水源涵养林等生态系统,提升农田水利工程的景观、休闲、文化等综合价值。

#### 4.5 注重农田水利与生态环境协同治理,推进绿色发展

全域土地综合整治强调生态文明建设,要求统筹生产、

生活、生态三大空间,农田水利工程建设也应当秉持生态优先理念,实现与生态环境的协同治理,在全域土地整治中,应当将农田水利工程纳入整体的生态系统考量,使其成为维护区域生态安全的重要环节。

新时期农田水利工程建设要树立“绿水青山就是金山银山”理念,坚持生态优先、绿色发展,加强与水土保持、水环境治理、水生态修复等工作的协同推进,实现农田水利建设与生态文明建设相得益彰、互促共进。要立足流域和区域水资源禀赋,科学制定农田水利工程配置方案,严格水资源消耗总量和强度双控,优化水资源时空分布,提高水资源利用效率<sup>[4]</sup>。在工程建设中注重生态环保,优化施工工艺,减少对原有地貌和植被的扰动,加强弃渣管理和生态修复,最大限度降低工程建设的负面生态影响。

全域土地整治为农田水利工程的生态功能发挥提供了更大空间,大力发展生态农田水利,推广渠道防渗、管道输水等节水技术,建设集雨蓄水、水肥一体化等高效节水灌溉系统,提高农业用水效率。在农田水利工程建设中注重生物多样性保护,科学布设生态廊道和野生动物通道,为水生生物提供栖息环境,加强农业面源污染治理,在灌区建设植物缓冲带和人工湿地,削减农田退水中的氮磷等污染物,保护区域水环境,完善农田水利排水系统,促进农田退水循环利用,开展农村黑臭水体整治,改善农村人居环境。

## 5 结语

全域土地综合整治是国土空间治理的重要抓手,事关农业农村现代化和城乡融合发展全局,在全域土地综合整治中,农田水利工程建设发挥着基础性、先导性作用,是实现农业高质高效、乡村宜居宜业的关键举措,面对新形势新要求,必须坚持创新驱动,着力在体制机制、规划设计、建设管理、运行维护等方面积极探索、大胆实践,着力构建形成一套符合时代要求的现代农田水利工程建设管理新机制新模式。

### 参考文献

- [1] 武帅帅. 农田水利工程高效节水灌溉技术发展概述 [J]. 现代农业科技, 2025, (12): 149-152.
- [2] 张合生. 新农村中的小型农田水利工程管理模式探讨 [J]. 农机市场, 2025, (06): 65-67.
- [3] 刘桐宇. 智能化节水灌溉技术在农田水利工程中的应用 [J]. 科技与创新, 2025, (11): 88-91.
- [4] 陈景宏. 农田水利工程施工全流程质量控制策略研究 [J]. 农业开发与装备, 2025, (06): 146-148.

# Research on the Optimization of Water Conservancy Project Operation Management Question-Answering Robot Based on Large Language Model

Yiwei Yuan

Suzhou Administration of Taihu Basin Administration, Suzhou, Jiangsu, 215100, China

## Abstract

This article focuses on the optimization exploration of a question-answering robot for the operation and management of water conservancy projects based on large language models. In terms of the technical framework, a complete system has been constructed, which includes system architecture design, document sharding and vectorized optimization, as well as hybrid retrieval and reordering mechanisms. Empirical research was implemented on the Dify platform through experimental design and data preparation, and performance evaluation and comparison were carried out. At the same time, conduct in-depth analysis of key challenges, propose multi-knowledge point aggregation processing and long document processing optimization solutions, and take into account both ethical and security considerations. The research aims to enhance the accuracy and efficiency of question-answering in the operation and management of water conservancy projects, and provide theoretical and practical references for the optimization of intelligent question-answering systems in related fields.

## Keywords

Large Language Model Operation of water conservancy projects Optimization of Question-Answering Robots

## 基于大语言模型的水利工程运行管理问答机器人优化探究

袁一炜

太湖流域管理局苏州管理局, 中国·江苏 苏州 215100

## 摘要

文章聚焦基于大语言模型的水利工程运行管理问答机器人优化探究。在技术框架上, 构建了包含系统架构设计、文档分片与向量化优化、混合检索与重排序机制的完整体系。实证研究通过实验设计与数据准备, 在 Dify 平台实现落地, 并开展性能评估与对比。同时, 深入分析关键挑战, 提出多知识点聚合处理、长文档处理优化方案, 兼顾伦理与安全考量。研究旨在提升水利工程运行管理问答的准确性与效率, 为相关领域智能化问答系统的优化提供理论与实践参考。

## 关键词

大语言模型; 水利工程运行; 问答机器人优化

## 1 引言

随着数字孪生水利的深入推进, 基于大语言模型的私有知识库问答机器人被越来越多的水利工程运行管理单位部署在数字孪生系统上, 以提升工程管理的智能化水平。与此同时, 以大语言模型 (LLM) 为基础的知识问答系统构建方法不可避免的带来一些缺陷, 包括上层应用和模型基座深度绑定, 切换模型基座时, 上层逻辑不得不大量修改; 模型的处理环节不完善, 开发成本高, 比如向量存储和搜索, LLM 提示词生成, 数据链路 (导入、分片、加工) 等都需要投入成本; 缺少标准化流程, 难以结合场景灵活编排。本文针对上述问题, 尝试探索

## 2 大语言模型的水利工程运行管理问答机器人优化重要性

### 2.1 研究背景与意义

自 2022 年底 ChatGPT 发布以来, LLM 在自然语言处理领域展现出革命性突破。然而, 通用 LLM 在垂直领域存在知识滞后性 (如 GPT-3.5 训练数据截止至 2021 年) 和未公开数据覆盖不足等问题。构建基于私有知识库的问答机器人, 成为解决领域知识实时性与准确性的关键路径<sup>[1]</sup>。据 IDC 研究显示, 2025 年企业级 AI 问答系统市场规模将达 127 亿美元, 年复合增长率超 45%, 其中私有知识库集成方案占比超 68%。

水利领域作为典型垂直场景, 其工程手册、运行规范等专业文档具有强时效性和保密性。以太浦闸工程为例, 传统检索方式存在关键词匹配精度低、跨文档关联分析不足等问题, 导致技术人员查询效率低下。本研究的创新点在于:

**【作者简介】**袁一炜 (1996-), 男, 中国江苏南通人, 硕士, 工程师, 从事智慧水利工程研究。

提出“向量检索+语义聚合”的混合架构，解决多知识点关联查询难题；

结合水利领域语料特性，优化文档分片策略与向量化模型选择；

通过 Dify 平台实现低代码落地，验证了工程化方案的可行性。

## 2.2 国内外研究现状

国际方面，OpenAI 于 2023 年推出的 GPTs (GenerativePre-trainedTransformers) 允许用户自定义知识库，2025 年发布的 GPT-4Turbo 将上下文长度扩展至 128ktokens，大幅提升长文档处理能力。Google 的 Gemini 模型通过跨模态融合技术，在文档问答场景中实现了多源数据的语义关联。国内研究中，清华大学 ChatGLM-6B 通过量化技术实现消费级显卡部署，字节跳动 Doubao 模型在中文领域知识表示上取得显著进展。

在私有知识库构建方面，传统方法主要采用 Fine-Tuning 和 PromptEngineering，但前者存在参数更新成本高、后者受限于上下文长度的问题。2024 年提出的检索增强生成 (RAG) 框架成为主流方案，其核心在于通过向量数据库实现高效知识召回，结合 LLM 进行答案整合。据《Nature》子刊研究，RAG 架构在医疗、法律等领域的问答准确率较纯 LLM 方法提升 34% 以上。

## 3 技术框架与关键算法

### 3.1 系统架构设计

本研究提出的问答系统框架由三大模块组成：

知识预处理模块：完成文档分片、向量化与存储；

智能检索模块：结合向量相似度与语义理解实现混合检索；

答案生成模块：利用 LLM 进行结果整合与逻辑推理。

相较于传统 RAG 架构仅依赖单一向量索引的局限，该框架的核心创新在于提出“多级索引策略”。此策略突破了传统文档切分模式，采用语义块 (SemanticChunk) 切分技术，依据水利工程专业术语的语义关联性划分文档单元，每个语义块单独生成向量索引以保留上下文逻辑<sup>[2]</sup>。同时，通过关键词提取算法从语义块中提炼“坝体监测”“水位调控”等专业术语，构建倒排索引实现精准匹配。这种“语义+词法”的双重检索机制，既解决了传统向量检索对专业术语敏感性不足的问题，又弥补了单纯关键词检索难以理解语义关联的缺陷<sup>[3]</sup>。实验数据显示，在处理涉及多流程的水利调度方案、跨系统的设备故障诊断等复杂问题时，该策略使召回准确率较传统架构提升 19.2%，尤其在专业术语密集的场景中表现更为突出，显著增强了问答系统对水利工程领域复杂知识的处理能力。

### 3.2 文档分片与向量化优化

#### 3.2.1 语义块切分算法

针对水利文档的专业特性，设计了基于 Markdown 结构的分层切分器：

一级切分：按章节标题（如“3.2 闸门操作规

程”）进行初步分割；二级切分：利用递归字符分割器 (RecursiveCharacterTextSplitter) 按段落切分，控制每个分片在 100-150tokens 之间；三级切分：对表格、流程图等非文本内容进行语义解析，生成结构化元数据。

对比实验显示，该分层切分法较传统均匀切分法，在问题-答案匹配度上提升 23.5%（见表 1）。

表 1

切分方法	平均匹配度	耗时 (秒/100 页)
均匀切分	0.62	12.5
基于标题切分	0.71	8.3
分层语义切分	0.85	9.7

#### 3.2.2 向量化模型选型

测试了四种主流向量化模型在水利领域的表现：

text-embedding-ada-002 (OpenAI)：向量维度 1536，支持 8192tokens，中文相似度排序性较好但区分度不足（集中在 0.7 左右）；

text2vec-large-chinese：向量维度 768，支持 512tokens，在专业术语表示上更优；

paraphrase-multilingual-mpnet-base-v2：多语言支持，但中文语义表示较弱；

BAAI/bge-large-zh：2025 年发布的中文向量化模型，在水利领域的语义召回率较 ada-002 提升 11.3%。

最终选择 BAAI/bge-large-zh 作为基础向量化模型，配合 OpenAI 的 text-embedding-ada-002 进行跨模态融合，形成“基础向量+语义增强”的双重表示。

### 3.3 混合检索与重排序机制

设计了三层检索策略：

向量相似度检索：使用 HNSW 算法在向量数据库中召回 Top-K 结果；

关键词语义检索：通过 Elasticsearch 对问题进行分词，匹配文档中的专业术语；

LLM 辅助检索：利用 GPT-4Turbo 生成扩展查询词，补充潜在相关知识点。

在检索结果的精细化处理阶段，该框架进一步引入 Cohere 的 rerank-multilingual-v2.0 模型优化重排序机制，形成“检索-重排”的闭环增强逻辑。传统 RAG 架构的重排序多依赖简单相似度加权，难以处理水利工程领域“多步骤操作”“跨规范关联”等复杂问题的语义匹配<sup>[4]</sup>。而 rerank-multilingual-v2.0 模型通过深度理解专业术语的层级关系（如“船舶过闸”包含“闸室水位调节”“通航信号校验”等子环节），对初步检索得到的语义块进行二次打分，优先保留与问题意图高度契合的内容。实验数据表明，这一机制使最终答案的准确率提升 15.7%。在针对“船舶过闸操作流程与安全规范”这类涉及多文档交叉引用的复杂问题时，结合前文的多级索引策略，混合检索体系的准确率较单一向量检索提升 32.1%，尤其在处理包含专业缩略语、跨章节关联知识的查询时，展现出更强的上下文整合能力，有效解决了传统架构中“检索碎片化”“语义偏移”等痛点<sup>[5]</sup>。