

杭州电子科技大学

毕业设计（论文）开题报告

基于 Node.js 的智能 AI 问答系统设
计与实现

学 院 计算机学院

专 业 软件工程

姓 名 龚德江

班 级 21052711

学 号 21052211

指导教师 王思轩

一、综述本课题国内外研究动态，说明选题的依据和意义

1.1 总体说明

随着人工智能技术的蓬勃发展，智能 AI 问答系统已成为信息技术领域的研究热点。这类系统能够模拟人类对话，理解和回答用户的问题，为用户提供便捷的信息获取途径。基于 Node.js^[1] 的智能 AI 问答系统，结合了 Node.js 的高效服务器端开发能力和 AI 技术的智能问答功能，具有广阔的应用前景和研究价值。

1.2 国内外现状

在国内，智能 AI 问答系统的发展势头强劲，众多科技企业和研究机构如百度、阿里巴巴、腾讯等都在积极探索这一领域的研发和应用。这些企业推出的智能问答产品，^[3]式。然而，国内 AI 问答系统的发展仍面临一些挑战^[2]，如提高系统的准确性和稳定性、处理复杂和模糊的问题等。

在国外，智能 AI 问答系统的发展同样如火如荼。OpenAI 推出的 GPT 系列大模型，如 GPT-3 和 GPT-4，以其卓越的自然语言处理能力和深度学习能力，为 AI 问答系统树立了新的标杆。这些大模型的应用，极大地提升了 AI 问答系统的智能化水平。

大模型技术日益成熟，企业大模型应用场景加速落地。知识库作为企业持续沉淀的数据资产，为大模型落地提供天然的数据基础，因此知识库成为企业大模型率先落地场景。

人工智能已经成为科技革命和产业变革的核心驱动力，在政府对人工智能产业发展的大力支持下，各行业开始进行试点，其中大模型+知识库因其成本低、周期短，成为大模型落地的优先场景。在各行业中，金融、政务、电信是大模型+知识库应用最领先的三个领域。以上三领域知识繁杂，内部员工使用频繁，且均需面向外部用户提供咨询服务支持，内外需求驱动业内机构快速落地大模型+知识库，如金融应用场景以智能客服、智能投顾、智能报告生成为代表，政务领域应用场景如 12345 热线、政策标准知识库搜索等。此外，教育、医疗、工业、能源等行业领先企业也在试点大模型+知识库，主要满足内部员工使用需求。如教育行业的智能教学、个性化学习推荐，医疗领域的药物研发、就医知识库等场景。

Node.js 以其高效、快速及丰富的生态系统成为 AI 项目的理想选择，能够解决 AI 问答系统中快速开发、跨平台部署及高效处理高并发请求等问题。

1.3 选题的依据

选择本题主要关注两个方面，一是基于 Node.js 技术，二是智能 AI 问答，所以选题依据主要从这两者展开：

(1) 技术趋势与需求增长：随着人工智能技术的快速发展，智能 AI 问答系统已成为

企业服务、在线教育、智能家居等多个领域^[4]的关键技术。Node.js 作为高效的服务器端开发平台，结合 AI 技术，能够构建高性能、可扩展的问答系统，满足市场对智能化服务的需求。

- (2) 技术融合与创新：Python 之所以是 AI 开发的主流语言，不仅因为其语法简洁、易于上手，更是因为它有丰富的 AI 库与框架（如 TensorFlow 和 PyTorch）支持底层的数学运算和硬件加速，这些框架为 AI 模型的训练与推理提供了高效的计算能力。但在许多实际的 AI 项目中，Node.js 仍然拥有其独特的地位。举个例子，许多 AI 应用需要前端与后端的交互，以提供实时响应的用户体验。Node.js 凭借其事件驱动与非阻塞 I/O^[5]的优势，可以高效处理用户请求，为用户带来顺畅的体验。例如，在数据可视化、API 接口和实时数据交互等方面，Node.js 依旧是开发者的重要选择。
- (3) 要解决的实际问题：使用 Node.js 完成后端接口，并对接已有大模型，并开发简介美观的前端页面，给用户提供优质的体验。同时使用 prompt^[6]和数据情绪不断优化回答效果。

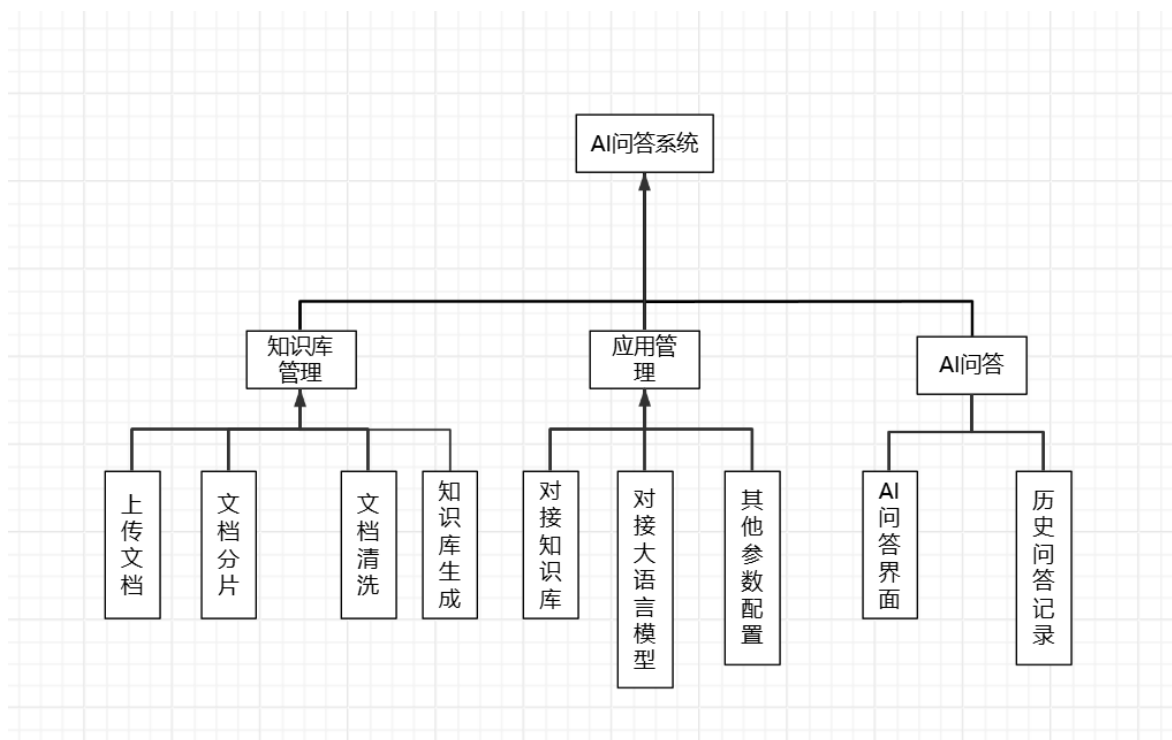
1.4 选题的意义

- (1) 技术前沿探索：本课题旨在探索 Node.js 与 AI 技术的结合，推动智能问答系统的技术创新，紧跟技术前沿。
- (2) 市场需求满足：随着智能化服务的普及，智能 AI 问答系统成为众多行业的标配。本课题的研究有助于满足市场需求，提升用户体验。
- (3) 性能优化实践：Node.js 的高效性能使其成为构建高性能问答系统^[7]的理想选择。本课题将实践 Node.js 在 AI 问答系统中的性能优化，为类似应用提供参考。
- (4) 知识图谱^[8]构建：AI 问答系统依赖于丰富的知识图谱。本课题将探索知识图谱的构建与更新方法，为系统提供持续的知识支持。
- (5) 自然语言处理深化：智能问答系统离不开自然语言处理技术的支持。本课题将深化对 NLP 技术的理解与应用，提升系统的问答准确性与流畅性。
- (6) 用户体验提升：通过智能 AI 问答系统，可以为用户提供更加便捷、个性化的服务，提升用户满意度与忠诚度。
- (7) 跨学科研究融合：本课题涉及计算机科学、人工智能、自然语言处理等多个领域，有助于促进跨学科研究的融合与发展。

二、研究的基本内容，拟解决的主要问题：

2.1 系统整体功能模块

系统整体模块如下，主要包括知识库管理、应用管理和 AI 问答三大部分。知识库管理主要包括上传文档、文档分片和清洗、知识库生成；应用管理包括对接知识库、对接大语言模型、其他参数配置；AI 问答包括 AI 问答交互界面、历史问答记录。



2.2 系统功能更详细的介绍

- (1) 上传文档：支持直接上传 docx、txt 等多种格式文档。
- (2) 文档分片：提供多种文档切片方式。
- (3) 文档清洗：提供多种文本清洗等 RAG 文档预处理方式；
- (4) 文档改写：将 RAG^[9]不能高效处理的结构化数据，通过文档改写修改为 RAG 友好的非结构化数据。
- (5) 知识库生成：提供多个预置场景 Prompt 库，支持生成高质量的 QA 问答对，支持基于 QA 的知识库生成功能，后续将提供更多的重写增强结构化处理等知识库管理能力。
- (6) 对接知识库：将应用和已有的知识库对接起来。
- (7) 对接大语言模型：支持对接 openAI 等大语言模型。
- (8) 其他参数配置：配置应用的名称、简介、开场白、找不到对应答案时的提示语等。
- (9) AI 问答界面：提供简介美观的 AI 问答界面，用户可以通过输入问题，或者点击已有问题向 AI 问答助手提问并得到解答。
- (10) 历史问答记录：用户过去的提问将被记录下来。

三、研究步骤、方法及措施：

3.1 研究步骤

3.1.1 总体工作流程如下

| 工作流程 | 日程安排 |
|----------------------|-------------|
| 系统分析和设计 | 1. 1-2. 1 |
| 完成系统核心功能，上线第一个最小可用版本 | 2. 1-2. 17 |
| 持续迭代，完善后续开发 | 2. 18-3. 18 |
| 优化系统架构，提高系统的并发处理能力 | 3. 19-3. 26 |

3.3.2 流程中的具体研究步骤

- (1) 需求分析与系统设计：
 - a) 调研市场现有 AI 问答系统，明确系统需实现的功能和性能指标。
 - b) 设计系统架构，包括前端交互界面、后端 Node.js 服务器、数据库及 AI 模型对接模块。
- (2) 技术选型与环境搭建：
 - a) 选择适合的 Node.js 框架（如 Express）和数据库（如 MongoDB）。
 - b) 搭建开发环境，配置必要的开发工具和库。
- (3) 模块开发与集成：
 - a) 开发文档上传、分片、清洗、改写等预处理模块。
 - b) 实现知识库生成与管理功能，包括 QA 对生成和结构化数据处理。
 - c) 对接大语言模型，开发模型调用和结果解析模块。
 - d) 设计并实现前端交互界面，确保用户友好。
- (4) 系统测试与优化：
 - a) 进行单元测试、集成测试和系统测试，确保各模块功能正常且系统稳定。
 - b) 根据测试结果优化代码和性能，提升系统响应速度和用户体验。
- (5) 部署与上线：
 - a) 在服务器上部署系统，配置必要的环境变量和参数。
 - b) 进行系统上线前的最终测试，确保系统稳定运行。
 - c) 正式上线并持续监控系统运行状态，及时响应和处理用户反馈。

3.2 研究方法

- (1) 文献调研：广泛查阅国内外关于 AI 问答系统、Node.js 开发、大语言模型对接等

方面的文献，了解研究背景和前沿技术，为系统设计提供理论依据。

- (2) 实践探索：通过动手实践，学习并掌握 Node.js 框架、数据库操作、前端交互设计等关键技术，积累实战经验，为系统实现打下坚实基础。
- (3) 案例分析：分析市场上已有的 AI 问答系统案例，总结其优点和不足，结合实际需求，提出改进方案，为系统设计提供参考。
- (4) 社区参与：积极参与相关技术社区和论坛，与同行交流心得，获取技术支持，解决开发过程中遇到的问题。
- (5) 实验验证：在系统开发过程中，通过单元测试、集成测试和系统测试，验证各模块功能的正确性和系统的稳定性，确保系统满足设计要求。

3.3 措施和实施

3.3.1 技术和工具选型

- 前端：Vue、Ant Design、Axios
- 后端：Node.js、LangChain、PostgreSQL/pgvector、Nginx、RabbitMQ、Redis
- 其他工具：Visual Studio Code、Node-RED、ML.js
- 大模型：openAI 等国内外大模型

3.3.2 采用基于原型的敏捷编程进行系统设计实现

在基于 Node.js 的智能 AI 问答系统的设计与实现中，我们采用基于原型的敏捷编程方法。该方法强调快速迭代和用户反馈，通过构建系统的原型来逐步明确需求和功能。首先，我们根据初步需求分析，快速搭建一个基本可行的系统原型，包括用户交互界面、AI 问答逻辑等核心功能。随后，通过用户测试和反馈，不断迭代优化原型，逐步添加和完善系统模块，如文档处理、知识库管理、大语言模型对接等。在开发过程中，我们注重团队协作和代码复用，采用敏捷开发流程^[10]，确保系统能够快速响应需求变化，同时保持代码的可维护性和可扩展性。通过这种方法，我们能够高效地完成系统设计与实现，为用户提供一个功能完善、性能稳定的智能 AI 问答系统。

四、研究工作进度：

| 序号 | 时间 | 内容 |
|----|---------------------------|-----------------|
| 1 | 2024. 10. 31-2024. 11. 5 | 选好毕业设计题目并准备相关资料 |
| 2 | 2024. 11. 6-2024. 11. 10 | 接受任务书 |
| 3 | 2024. 11. 11-2024. 12. 24 | 搜集资料, 准备开题报告 |
| 4 | 2024. 12. 25-2024. 12. 29 | 开题报告会 |
| 5 | 2024. 12. 30-2025. 1. 28 | 按照开题计划进行实验思路设计 |
| 6 | 2025. 1. 29-2025. 3. 26 | 处理数据、编写程序、对比试验 |
| 7 | 2025. 3. 27-2025. 4. 16 | 撰写毕业论文 |
| 8 | 2025. 4. 17-202. 4. 25 | 论文审核及查重 |
| 9 | 2025. 4. 26-2025. 5. 09 | 论文自查、专家抽查论文 |
| 10 | 2025. 5. 09-2025. 5. 11 | 答辩报告会 |

五、主要参考文献：（所列出的参考文献不得少于 12 篇，其中外文文献不得少于 2 篇，发表在期刊上的学术论文不得少于 4 篇。）

- [1] 周绍彬, 肖铮. 基于开源大语言模型的知识库问答系统实践[C]//教育部高等学校图书情报工作指导委员会, CALIS 管理中心, 上海交通大学图书馆, 武汉大学图书馆. 2024 年第十七届图书馆管理与服务创新论坛论文集. 厦门大学; , 2024:2. DOI:10. 26914/c. cnkihy. 2024. 041639.
- [2] 张君冬, 刘江峰, 王震宇, 等. 用户响应式场景下大模型驱动的 AI 问答研究: 以医疗分诊为例[J/OL]. 情报理论与实践, 1-13[2024-12-15]. <http://kns.cnki.net/kcms/detail/11.1762.G3.20241018.1002.002.html>.
- [3] 晋赞霞, 覃京燕, 殷绪成. AI 设计下的文本视觉问答技术[J]. 包装工程, 2021, 42(06):7-12. DOI:10.19554/j.cnki.1001-3563.2021.06.002.
- [4] 林德力, 范玉婷, 廖心妍, 等. 基于 GPT 模式的 LLM 模型构建高职学校人才培养工作状态数据采集与管理平台问答系统研究[J]. 创新创业理论与实践, 2023, 6(16):9-14.
- [5] Lei W ,Chen M ,Xueyang F , et al.A survey on large language model based autonomous agents[J].Frontiers of Computer Science,2024,18(06):
- [6] Hao C ,Liu Y ,Chao M , et al.Telecom user churn prediction scheme based on large language model[J/OL].The Journal of China Universities of Posts and Telecommunications,1-8[2024-12-15].<https://doi.org/10.19682/J.Cnki.1005-8885.2024.1013>.
- [7] 王晓宇, 李欣, 胡勉宁, 等. 基于大语言模型的 CIL-LLM 类别增量学习框架[J/OL]. 计算机科学与探索, 1-13[2024-12-

15]. <http://kns.cnki.net/kcms/detail/11.5602.tp.20241018.1053.002.html>.

- [8] 李晓庆, 马昊聪, 周军华, 等. 基于大语言模型的知识图谱构建与应用[C]//中国仿真学会. 第三十六届中国仿真大会论文集. 北京市复杂产品先进制造系统工程技术研究中心北京仿真中心; 复杂产品智能制造系统技术国家重点实验室北京电子工程总体研究所; 航天系统仿真重点实验室北京仿真中心; , 2024: 11. DOI:10.26914/c.cnkihy.2024.044781.
- [9] 陈晓皎, 束云峰, 汪睿涵, 等. 大语言模型驱动的 UI 评估系统[J/OL]. 图学学报, 1-11[2024-12-15]. <http://kns.cnki.net/kcms/detail/10.1034.T.20241016.1741.002.html>.
- [10] 张恒郡, 李飞, 刘鸣飞, 等. 基于大语言模型的风机装配工艺文档多模态图谱建模技术[J/OL]. 东华大学学报(自然科学版), 1-13[2024-12-15]. <https://doi.org/10.19886/j.cnki.dhdz.2024.0229>.
- [11] 肖洪云, 徐佳钰. 基于 LLM 的“混合式教学设计”智能体构建与应用研究[J]. 沧州师范学院学报, 2025, 41(01): 97-102. DOI:10.13834/j.cnki.czsfxxyb.2025.01.018.
- [12] 查英华, 郭朝霞, 鞠慧光. 基于大语言模型的智能学习助手设计与实现[J]. 现代信息科技, 2025, 9(03): 50-55. DOI:10.19850/j.cnki.2096-4706.2025.03.010.

六、开题答辩小组评审意见：

| 考核点 | 背景及 意义阐述情况 | 研究内容 与任务书的匹配程 度 | 研究方案 合理性 | 进度安排 情况 | 答辩情 况 | 总分 |
|-----|---------------|-----------------------|-------------|------------|----------|-----|
| 满分 | 20 | 30 | 30 | 10 | 10 | 100 |
| 评分 | | | | | | |

开题答辩小组负责人签字：_____

2024 年 月 日