###### 选题的目的和意义：

1. 选题意义：

理论意义：目前基于Transformer架构的大语言模型发展逐渐成熟，但应用领域较为单一，多用于简单的生成式问答和自然语言处理场景，目前，学者们已经探索了LLM应用于金融、医学等多个领域的知识问答，但是回答质量受领域知识缺乏和推理能力较弱的影响[3]。知识图谱通过结构化的方式表示和管理知识，在个性化推荐，搜索引擎优化等领域有着出色的表现，但传统的知识图谱构建方法依赖大量人工标注工作，耗时耗力且难以满足知识更新需求，本课题尝试通过结合大语言模型的强大语言理解和生成能力，降低知识图谱构建中的复杂性和成本问题。

现实意义：通过结合大语言模型与知识图谱技术，提升医疗信息处理的效率，降低传统知识图谱构建中的高成本与低效率问题。研究成果将推动智能医疗系统的发展，实现自动化的知识提取与管理，帮助普通人更快速、准确地获取医疗知识，优化诊疗流程。此外，本课题的研究还可推广至其他领域（如金融、法律等），推动跨领域知识管理的智能化，促进相关行业的技术进步和应用落地，为提升国家医疗资源共享和决策支持能力提供有力保障。

1. 选题目的：

旨在借助国内优秀的大模型训练平台和gStore国产图数据库构建优秀的医疗领域知识图谱，提升目前的医疗信息处理效率，并希望知识图谱反作用于大语言模型的提示词工程，实现促进智能医疗发展、解决实际应用中的痛点。

###### 国内外研究现状：

1. 国内研究现状：

对于将大模型的文本处理能力用于知识图谱的构建的研究，近一年内开始增多，涉及领域包含水利，核电控制，电子信息等，在医疗领域，也有用于肝病，鼻炎，中医等方向的研究。主要研究成果为建立知识库问答平台，依托的数据库大多为的Neo4j数据库[1]。具体的研究内容包括利用可商业化的开源大模型和基于思维链的Prompt工程，寻找文本内容中的实体,以及和现有的知识图谱构建方案的对比。

1. 国外研究现状：

相比于国内，国外的研究更倾向于利用构建完善的知识图谱来训练大模型，利用知识图谱这种外部知识源来解决小型LLM出现的知识不足和幻觉问题，此外，也有一些利用LLM来去生成新的三元组，用以完善不完整的知识图谱的研究[2]。

###### 理论依据和研究内容：

1. 理论依据：

经过详细调研得出：知识图谱技术和大规模预训练语言模型（LLM）各自存在缺陷，前者结构化表达领域知识但依赖人工标注，后者在无标注数据中泛化能力强但缺少领域知识和推理能力。此外，LLM能够弥补知识图谱中推理能力不足的问题，而知识图谱可以为LLM提供领域特定的知识，从而增强其回答领域问题的准确性。两者结合能够更好地满足从业人员从海量文档中提取相关知识的需求，提高知识利用的效率。

二、研究内容：

1.使用国内开源的大模型训练平台（如智谱AI，通义千问）对医疗文献、诊疗记录等非结构化数据中自动提取医疗实体和关系，使用国产图数据库gStore构建高质量的医疗领域知识图谱。

2.设计一套通用技术方案或平台，完成数据加工，实体识别，知识图谱构建的自动化，探究利用该平台实现低成本构建某一领域知识图谱的可能。

###### 研究方法：

1. 研究方法论

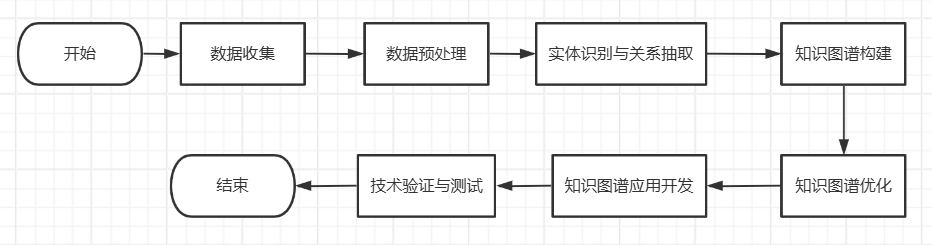
**1.文献综述法**：通过系统回顾国内外研究现状，了解相关领域的理论基础和实际需求。结合知识图谱与大语言模型的   理论，构建理论框架。

**2.数据驱动方法**：收集和分析大量的医疗数据，提取实体和关系，以此构建高质量的知识图谱。

**3.实验研究法**：设计实验来测试不同的数据预处理、实体识别和关系抽取的方法，以验证技术路径的有效性。

**4.敏捷开发法**：采用迭代式开发方法，不断改进知识图谱的构建流程，优化数据处理和模型输出。

1. 技术路线图



###### 步骤及进度安排

2024-10-30至2024-12-15:完成前期的数据收集工作；

2024-12-16至2025-1-15:对国内的开源大模型进行微调，利用python或shell等编写脚本对数据进行初步处理。

2025-1-16 至 2025-1-30：使用云服务器搭建gStore数据库环境，对初步处理的数据进行核验。

2025-2-01至 2025-2-15：将整理好的数据录如数据库中，完成毕业论文初稿。

2025-2-16 至202-3-5：搭建一套知识图谱可视化web平台，完善论文。

###### 重点和难点：

1. 前期数据集的收集和清洗：计划使用国内的专业数据集开源网站和利用爬虫等技术到网上搜集一些数据。
2. gStore数据库云环境的搭建：使用云服务器搭建一套线上开发和运行环境，给整个毕设提供支撑。
3. 设计优秀的提示词使生成的结果，符合知识图谱构建的要求。

###### 预期结果和成果形式：

结果：构建高质量的医疗领域知识图谱，计划针对某一垂直的领域进行，完成其持久化存储和可视化展示。

成果形式：基于gStore图数据库和Springboot技术栈，实现知识图谱的可视化，和问答平台的搭建。

###### 主要参考文献：

1. 张才科.基于大语言模型的知识图谱构建及应用研究[J].计算机科学与探索, 2024,18(7):3
2. Dong Shu.Knowledge Graph Large Language Model (KG-LLM) for Link Prediction[J]arxiv,2024,(4):3-6
3. 谢明华.基于大模型的电子信息领域知识图谱自动构建与检索技术[J]电讯技术,2024,64(08):4-8
4. 孔文强. 基于LLM Prompt技术的领域知识图谱构建框架研究[D]. 电子科技大学, 2024:3
5. 张子威 武志 学张薇. 大语言模型在医疗应急知识图谱问答服务中的智能化实践探索[J]. 软件导刊, 2002(10):4.