党史问答系统：LoRA 微调 + RAG 检索增强生成实验报告

一、项目概述

1.1 项目目标

- 针对党史领域问答场景，构建一套基于 LoRA 微调大模型 + 检索增强生成（RAG） 的端到端问答系统，满足以下核心要求：
- 使用 5041 领域 QA 数据，构建本地向量数据库；
- 支持 32k 长上下文、多轮对话、引用来源溯源、低置信度问题拒绝回答；
- 实现系统性能量化评估（准确率、引用 F1 值、幻觉率），并提供交互式 Web Demo 部署。

1.2 核心功能

- 向量检索：基于本地 bge-large-zh-v1.5 模型构建 FAISS 向量库，支持 Top-K 检索 + 重排序
- RAG 生成：融合 LoRA 微调 Qwen3-0-6B 模型，实现置信度判断、32k 长上下文、引用溯源

二、数据来源与处理

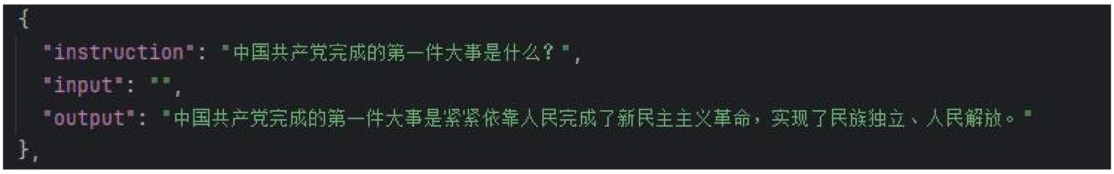
2.1 数据来源

实验采用《毛泽东邓小平江泽民胡锦涛关于中国共产党历史论述摘编 (中共中央党史和文献研究院)》以及《中国共产党简史》txt 文件。

2.2 QA 数据生成

党史 QA 数据集，数据格式为 JSON，核心字段：

- instruction：党史问题（如“中国共产党成立的标志是什么？”）；
- output：标准答案（如“1921 年 7 月召开的中共一大”）；
- input：空字段（无补充输入）。共 5321 对。



```
{
  "instruction": "中国共产党完成的第一件大事是什么？",
  "input": "",
  "output": "中国共产党完成的第一件大事是紧紧依靠人民完成了新民主主义革命，实现了民族独立、人民解放。"
},
```

2.3 向量数据库构建

- Embedding 模型：本地/models/bge-large-zh-v1.5，生成归一化向量；

三、基础模型

选择 Qwen-3-0.6B 为基础模型。使用 model_download.py 文件进行下载

四、方法

4.1 RAG 基础框架

- 检索器：基于本地 bge-large-zh-v1.5 模型构建 FAISS 向量库，支持 Top-K 检索 + 重排序

4.2 LoRA 微调

- 基座模型：Qwen-3-0.6B
- 训练参数：r=32, lora_alpha=128, max_seq_length=1024, batch_size=8, epochs=3
- 训练：rtx 5090 32G

4.3 评估方法

- 准确率：回答包含标准答案的比例；
- 引用 F1 值：来源标注的精准率 × 召回率 ×2/(精准率 + 召回率)；
- 幻觉率：无来源标注且非拒绝回答的比例。

五、实验结果

5.1 微调前后对比表

版本	Precision	Recall	F1	幻觉率
基座	0.497	0.654	0.524	0.200
Lora 微调	0.686	0.732	0.708	0.215
Lora+增强检索	0.741	0.818	0.807	0.114

六、demo 截图

系统状态

发现 2 个党史文档文件，正在加载...

成功加载 2 个文档片段

切分为 776 个文本块

检测到已有向量库，直接加载...

✅ RAG 系统已就绪

系统就绪！知识库包含 2 个党史文档

模型: 本地 Qwen3-0.6B

嵌入模型: BAAI/bge-large-zh-v1.5

使用提示

您可以询问关于党史的以下内容：

- 重要历史事件
- 党的历次代表大会
- 重要历史人物
- 党的理论发展
- 历史经验和教训

党史知识 RAG 问答系统

基于本地大模型的党史知识智能问答系统

您好！我是党史知识问答助手，可以为您解答关于中国共产党历史的各种问题。

中国共产党成立的标志是什么

<think> 好的，我需要回答用户的问题：“中国共产党成立的标志是什么？”首先，我要仔细查看提供的上下文内容，找到相关的信息点。
首先看第一个文档，里面提到“党的一大确定党的名称为‘中国共产党’。”这里明确提到了党的一大确立了党的名称，这是成立的重要标志之一。接下来，第二个文档提到“党的一大宣告中国共产党正式成立。”这里的表述更直接说明了成立的标志，即正式成立。

然后，第三个文档提到“中国共产党作为中国最先进的阶级——工人阶级的政党，不仅代表……”这部分虽然强调了党的性质，但并没有具体说明成立的标志是什么。不过，结合前面的内容，可以确认党的一大确实标志着中国共产党的成立。

需要注意的是，用户可能希望得到具体的标志，而不仅仅是描述成立的过程。因此，应该重点突出党一大确立名称和正式成立这两个关键点。同时，要确保时间、地点和人物的正确性，比如党一大召开的具体时间和地点，以及主要参与人员。

另外，检查是否有其他相关信息可能补充，但看起来上下文中只有这两个关键点。因此，综合这些信息，回答应包括党一大确立名称和正式成立这两个标志。 </think>

中国共产党成立的标志是**党的一大**。

具体说明如下：

1. **党的一大**于1921年7月23日在**上海法租界望志路**106号召开，标志着中国共产党正式成立。大会通过了党的纲领和章程，并确定党的名称为“中国共产党”。
2. 党的成立不仅是政治上的里程碑，更是思想理论上的奠基，体现了中国共产党作为工人阶级先锋队的性质和宗旨。

这一历史节点明确了中国共产党的核心特征和使命，奠定了其在民族复兴中的基础地位。

请输入您想了解的党史相关问题，例如：'中国共产党成立的历史背景是什么？'

系统状态

发现 2 个党史文档文件，正在加载...

成功加载 2 个文档片段

切分为 776 个文本块

检测到已有向量库，直接加载...

✅ RAG 系统已就绪

系统就绪！知识库包含 2 个党史文档

模型: 本地 Qwen3-0.6B

嵌入模型: BAAI/bge-large-zh-v1.5

使用提示

您可以询问关于党史的以下内容：

- 重要历史事件
- 党的历次代表大会
- 重要历史人物
- 党的理论发展
- 历史经验和教训

重要历史人物

<think> 好的，我现在需要处理用户的查询：“重要历史人物”。首先，我要仔细阅读提供的上下文，找到所有提到的重要历史人物及其相关信息。
首先看第一个文档，里面提到了孙中山、毛泽东、邓小平。孙中山是辛亥革命的主要领导人，推动了中国近代化的革命，但未能改变旧中国的性质。接下来是毛泽东，他在革命中发挥了关键作用，尤其是他的理论贡献，比如《论持久战》等。然后是邓小平，他是改革开放的总设计师，带领中国走向现代化。

第二个文档主要讨论了邓小平的《文选》内容，涉及他对历史的评价和对党的领导的重要性，但这里可能更多是关于他的思想，而不是直接列出人物。

第三个文档提到毛泽东的《改造我们的学习》，但同样没有直接列出具体的人物名字。第四个文档里，胡锦涛提到了改革开放和邓小平，但同样没有具体人物名称。

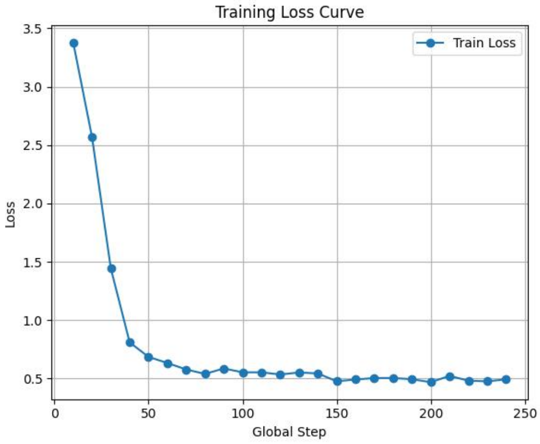
需要注意的是，每个文档的信息都是关于不同历史时期的人物，所以需要综合这些内容，确保没有遗漏重要人物。同时，要确认时间、地点和事件是否正确，例如孙中山的革命活动发生在1911年，毛泽东和邓小平分别在不同的时间段做出贡献。

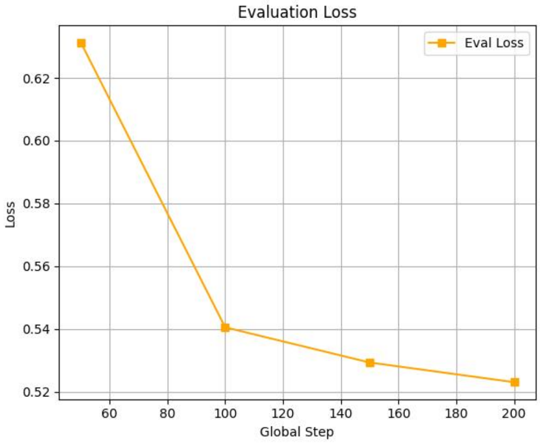
另外，用户要求回答要专业、准确，引用历史事件的时间、地点和人物，还要注意重要历史事件和决策的意义。因此，我需要将这些信息整理成结构清晰的回答，分点列出重要人物及其贡献，并注明相关的历史背景和意义。 </think>

根据提供的上下文，以下是重要历史人物及其相关分析：

1. **孙中山**
   - **贡献**：领导辛亥革命（1911年），推翻清朝统治，开启中国近代民主革命。尽管未能改变旧中国半殖民地半封建的社会性质，但为中国革命开辟了新方向。
   - **意义**：标志着中国现代史上“民族独立、民主共和”的开端，成为近代中国变革的关键人物。
2. **毛泽东**
   - **贡献**：
     - 领导中国共产党成立，推动新民主主义革命胜利（1949年后）。
     - 提出“实事求是”的思想路线，奠定中国特色社会主义理论的基础。
   - **意义**：作为马克思主义中国化的奠基者，其理论对中国革命和建设具有深远影响。
3. **邓小平**
   - **贡献**：
     - 开始改革开放，建立社会主义市场经济体制，推动中国经济发展进入新时代。
     - 主持“一国两制”进程，促进国际地位提升。
   - **意义**：中国历史上最具影响力的改革开放领袖之一，为全球发展提供了中国方案。
4. **江泽民**
   - **贡献**：
     - 在“十五大”上提出“科学发展观”，推动中国特色社会主义进入新阶段。
     - 组织“西双版纳”考察团，践行“中国梦”理念。

请输入您想了解的党史相关问题，例如：'中国共产党成立的历史背景是什么？'





七、未来改进方向

7.1 数据层优化

- 补充党史领域公开 QA 数据（如党史学习教育题库、权威文献），扩充有效数据量

- 构建数据质量评估体系（问题多样性、答案准确性、文本长度分布），提升数据基础质量。

7.2 检索层优化

- 替换单一向量检索为 “BM25 + 向量” 混合检索，提升低语义相似度问题的召回率；
- 引入重排序模型（如 bge-reranker-large），优化检索结果精准度，间接提升引用 F1 值。

7.3 模型层优化

- 微调 LoRA 模型时加入 “引用来源” 监督信号，让模型主动生成规范来源标注；
- 尝试 8bit 量化 / 模型蒸馏，进一步降低部署资源消耗。

7.4 功能层扩展

- 增加问答结果导出（JSON/Excel）、历史对话保存功能；
- 优化 Web Demo 界面，添加 “检索结果展示” 模块，提升透明度。

7.5 评估层完善

- 扩充评估样本量，增加 “响应速度”“用户满意度” 等主观 + 客观指标；
- 对比不同检索策略（向量 / BM25 / 混合）、不同 Embedding 模型的性能差异，形成优化对比报告。