## 2024 英特尔-阿里天池LLM Hackathon项目说明文档

### 1. 项目名称

MindEaseAI

### 2. 团队信息(1-5人)

* 团队名称： MindEaseAI
* 队长名： 武星辰
* 各成员及分工：

|  |  |
| --- | --- |
| **成员** | **分工** |
| 武星辰 | 产品 |
| 赵淑婷 | 前后端开发 |
| 徐宁 | 前后端开发 |
| 王冰倩 | 算法开发 |
| 卢银鹏 | 算法开发 |

* 联系邮箱： shuting.zhao@outlook.com

### 3. 项目背景与目的

* 当前市场需求或技术趋势

近年来，随着人工智能和大数据技术的快速发展，各行各业对数据分析和机器学习的需求不断增加。企业和组织希望利用数据驱动决策，提高运营效率，发现新的商业机会。在这样的背景下，市场上对数据科学家和机器学习工程师的需求激增。同时，随着自然语言处理（NLP）技术以及大语言模型技术的进步，通过人工智能大语言模型实现知识库问答成为提高用户体验和自动化服务的重要工具。

* 项目动机

在心理健康领域，传统的心理咨询服务存在供需不平衡的问题：专业心理咨询师数量有限，而需要帮助的人群却在不断增加。加上心理健康问题的敏感性，很多人往往因为隐私或羞耻感而不愿主动寻求帮助。AI技术的快速发展，特别是在自然语言处理和情感分析领域，为心理健康领域引入AI技术提供了可能。通过AI，可以更高效、更广泛地提供心理健康支持，减轻心理咨询师的工作负担，同时也能为更多有需要的人提供及时和个性化的帮助。

* 项目希望解决的问题或改进的领域

1. **心理健康支持的便捷性**：提供一个随时随地可用的心理健康支持平台，打破时间和空间的限制。
2. **心理问题识别的效率和准确性**：通过AI分析用户的倾诉内容，迅速总结和归纳用户的心理疑难杂症，并提供有针对性的意见。
3. **心理咨询师的工作负担**：通过AI助手承担初步的咨询和分析工作，为专业心理咨询师提供更为准确和详细的用户心理状态报告，从而提高咨询效率。

* 项目目标

1. 总体目标

开发一款AI心理健康咨询助手，模拟心理咨询师与来访者的互动，对来访者的倾诉进行分析和总结，提供高效、准确和个性化的心理健康支持。

1. 具体目标

**语言模型框架：**使用英特尔部署工具（如 ipex-llm、openvino）在魔搭社区CPU上部署开源大语言模型

**知识库搭建**：建立一个涵盖广泛心理健康知识的结构化知识库，支持AI助手通过RAG技术进行回答和建议。

**对话系统开发**：开发一个智能对话系统，能够与用户进行自然流畅的互动，理解用户的问题并从知识库中检索相关答案。

**用户体验优化**：设计友好的用户界面，确保系统易于使用，并能够提供高质量的用户体验。

1. 期望成果

通过本项目的实施，我们期望能够：

* 提供一个高效、便捷的AI心理健康咨询助手，帮助用户获得及时和个性化的心理健康支持。
* 实现对用户心理状态的准确分析和总结，提高心理问题识别的效率和准确性。
* 减轻心理咨询师的工作负担，为专业心理咨询师提供更为准确和详细的用户心理状态报告。
* 鼓励更多用户主动寻求心理健康帮助，推动心理健康服务的普及和发展。

### 4. 方案介绍

**方案的核心思想和原理**

我们的解决方案旨在通过构建一个智能RAG系统来有效管理和利用心理咨询行业的专业知识。将先进的英特尔大模型推理系统与知识库相结合，实现针对性的简单心理咨询服务。

### 5. 技术特点

**使用的算法、模型及数据集**

大语言模型：

Qwen-2 (4B)：这是我们选择的核心大语言模型（LLM），用于处理自然语言理解和生成任务。Qwen-2 拥有强大的语言理解能力，能够处理复杂的自然语言问题并生成高质量的回答。

模型优化工具：

ipex-llm（Intel® Extension for PyTorch Large Language Models）：用于在 CPU 上优化和加速 Qwen-2 大语言模型的推理过程，确保在魔搭社区的 CPU 环境中高效运行。

数据集:

PsyQA 一个中文心理健康支持问答数据集，提供了丰富的援助策略标注。可用于生成富有援助策略的长咨询文本。

https://github.com/thu-coai/PsyQA

**数据处理和分析的方法**

数据收集与预处理：

开发工具接口处理json数据信息，结构化处理并便于存入vector database

内容总结和信息占比分析：通过大语言模型对收集到的内容进行总结，分析各个信息的占比。这有助于我们理解数据的分布情况，并为后续的 chunk 处理做好准备。

数据增强与存储：

RAG（Retrieval-Augmented Generation）：采用 RAG 技术，将检索到的相关知识与生成模型结合起来，提高回答的准确性和相关性。

Chunk 处理：

根据内容总结和信息占比分析的结果，将大段的文本内容分割成更小的 chunk，便于存储和检索。

### 6. 使用到的英特尔软硬件技术

项目硬件使用魔搭社区提供的免费云CPU资源 Intel(R) Xeon(R) Platinum 8369B CPU @ 2.70GHz，核心数8；

部署工具采用英特尔 IPEX-LLM ，加速常用大语言模型在英特尔CPU 上的高速推理。

### 7. 成果说明（初赛可展示阶段性成果）

1. 能够通过选择咨询类型，检索不同的专业心理问答数据库辅助LLM回答

2. 避免通用模型模版化的冰冷回答，回答更温和，引导来访者倾诉

3. 后续计划为来访者回答建立个人知识库，采用Neo4j等图形数据库进行GraphRAG，对来访者思维模式进行结构化理解

A screenshot of a chatbot

Description automatically generated

### 9. 支持材料（可选）

* Github 链接：<https://github.com/rachelxhoot/MindEaseAI>
* 魔搭Demo（有一定有效期）：https://dsw-gateway-cn-hangzhou.data.aliyun.com/dsw-608509/proxy/7860/?