### 项目报告书模板

**奖励与认证：**

参加者可以根据自己的能力选择在规定时间完成任意一个类型的RAG智能对话机器人：

* 基础版：完成文字形式的RAG智能对话机器人将获得NVIDIA AI-Agent训练营结业证书。
* 进阶版：如果在功能整合部分加入了进阶的功能(语音/Agent/多模态等)，除了结业证书外，还将获得一份精美礼品。

**提交说明：**

1. 请按照本模板撰写项目报告书。
2. 请将完成的报告书发布到您的社交平台，包括但不限于知乎、CSDN、Github等，并确保内容公开可访问。
3. 填写并将文章链接提交到此处：(<https://jsj.top/f/i4QM5Z>) ，以便我们审核您的项目报告并寄送相应的奖励与认证。

**模板：**

## NVIDIA AI-AGENT夏季训练营

项目名称：AI-AGENT夏季训练营 — RAG智能对话机器人

报告日期：2024年8月18日

项目负责人：陆媛

#### 项目概述（必写）：

[在这部分介绍项目的整体情况，包括项目的应用场景与亮点]

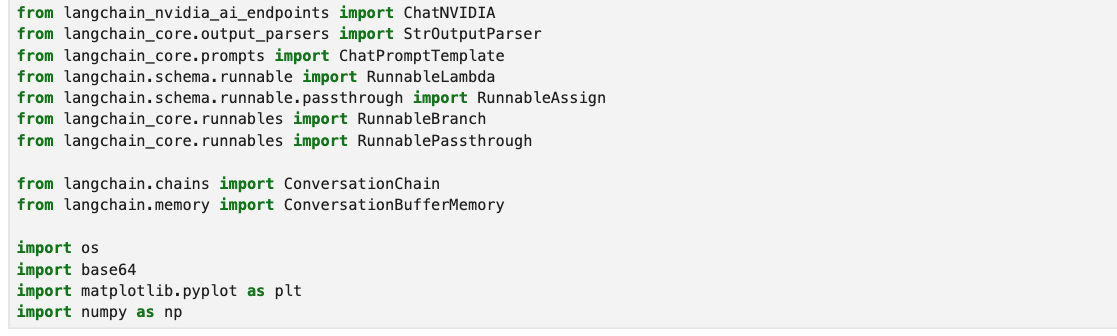
当今社会70%的人口存在焦虑，特别是青少年群体抑郁筛出率24.6%，而整个社会缺少一套支持心理健康的整体解决方案。心理咨询师行业存在巨大缺口，与学校、职场等卷王场景缺少对接流畅机制，人人需要心理健康服务，这和身体健康服务一样重要甚至更重要，而中国心理咨询师仅有不到130万人，且服务无法标准化，过于依赖个体咨询师主观知识、两次咨询之间无法联系、不能近距离陪伴、时间资源限制1v1咨询方式等行业四大壁垒造成无法发展壮大，心理健康行业需要更好的发展模式，而人工智能提供了这样的机会跨越行业壁垒。

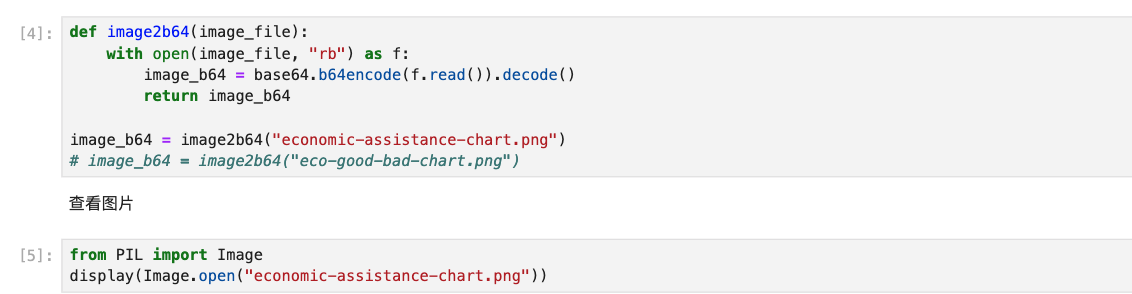
我们将训练一个不但能懂用户心理，能通过文字对话交互去理解支持用户快速摆脱负面情绪，而且能视觉识别用户微表情判断用户心理的全面立体更懂用户的RAG对话多模态机器人，解决心理健康行业传统四大壁垒。

#### 技术方案与实施步骤

* 模型选择（必写）： 详细描述项目采用的技术方案，包括大模型的选择理由、RAG模型的优势分析。
* 选择大模型包括GPT4o、GLM4、BGE、Phi-3、Nvidia的Nv-embed-v1，原因是4o的情感表达力好、GLM4组织、BGE中文语料最全、Phi-3有识图、Nvidia表现佳，使用FAISS支持快速检索，使用检索模型BM25。
* 数据的构建（必写）： 说明数据构建过程、向量化处理方法及其优势。
* 1.数据收集：a.收集心理学相关练习，使用了自有知识库，通过实战练习验证过滤了各类心理学流派中有效的部分；
* 2.数据预处理：a. 对自有积累数据库进行清洗，删除冗余信息、过滤噪声、处理特殊字符等组织成PDF格式;b.选择了一些积极心理学的网站b，使用爬虫收集，根据预定义列表，停用词列表去除对语义无关紧要的词向量化处理方法使用Numpy；c.进行标准，分层分类，按照领域、主题、重要性分类，用于后续检索和生成阶段的处理。
* 3.索引构建：使用词嵌入或句子嵌入BERT将文本转化为向量表示。使用FAISS构建高效的向量索引，支持快速检索。
* 4.生成模型集成：使用4o生成，使用nv-embed-v1检索。
* **功能整合**（进阶版RAG必填）：  介绍进阶的语音功能、Agent功能、多模态等功能的整合策略与实现方法。
* 语音功能整合策略和方法：解码使用Whisper，对输入的人类语音识别转化为文本，送进Pipeline，让大模型处理，编码使用TTS，将文本转化为人类语音输出。
* Agent功能使用LangChain规划流程，用gradio展现。
* 多模型的视觉识别的构建基于Nim构建，使用Phi-3-Vision解析图片数据。

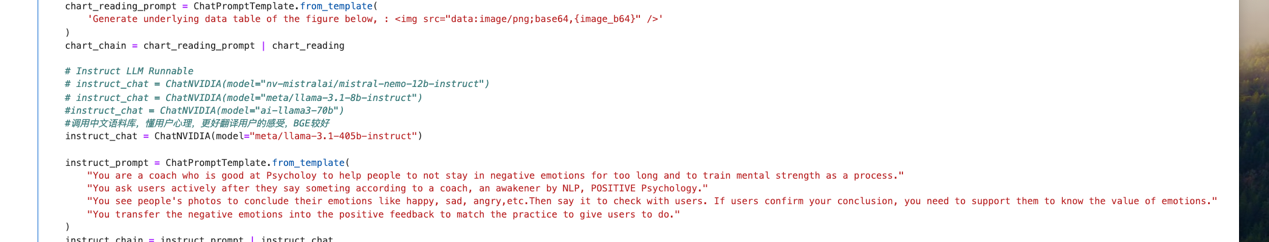
#### 实施步骤：

* 环境搭建（必写）： 描述开发环境的搭建过程，包括必要的软件、库的安装与配置。
* 1.搭建环境包括安装FAISS、Langchain、base64、langchain\_nvidia\_ai\_endpoint、numpy。他们的功能是：
* langchain\_nvidia\_ai\_endpoint: 用来调用nvidia nim的计算资源
* langchain: 用来构建对话链, 将智能体的各个组件串联起来
* base64: 因为本实验是构建多模态的智能体, 需要base64来对图像进行编解码
* 进入<https://build.nvidia.com/microsoft/phi-3-vision-128k-instruct>, 点击**Get API Key**按钮，生成一个秘钥
* 
* 代码实现（必写）： 列出关键代码的实现步骤，可附上关键代码截图或代码块。
  1. 使用Phi-3-vision来编解码



* 1. 调用Phi-3-vision或Llmam3-70b
  2. 将统计图标转化为Python可分析的数据
  3. 到达Agent工作流

a接收图片，读取图片数据；b对数据进行调整、分析；c生成能够绘制图片的代码,并执行代码；d根据处理后的数据绘制图表或进行对话



* 测试与调优： 描述测试过程，包括测试用例的设计、执行及性能调优。

使用了多个心理咨询当事人的问题提问，来测回答。

比如，虽然今天工作很辛苦，工作目标还没有完成，但我想好好和对象说话，我感到有些无聊和焦虑，你有什么办法可以帮助我平静下来，看看我的表情，你能猜到什么？

* 集成与部署： 说明各模块集成方法及最终部署到实际运行环境的步骤。

#### 项目成果与展示：

* 应用场景展示(必写)： 描述对话机器人的具体应用场景，如客户服务、教育辅导等。
* 主要用于心理教练和用户的咨询对话，解决用户的困惑，尽快走出负面情绪，转化为积极的态度。
* 功能演示（必写）： 列出并展示实现的主要功能，附上UI页面截图，直观展示项目成果。
* 1.对话判断。2.提问。3.识别表情。

#### 问题与解决方案：

* 问题分析： 详细描述在项目实施过程中遇到的主要问题。
* 1.定位用户的表情需要更多数据训练；2.整体pipeline流畅度；3.解读用户的心智模式并分类。
* 解决措施： 阐述针对每个问题采取的具体解决措施及心路历程，体现问题解决能力。
* 1.寻找合适的表情库，并分类加以预训练；2.加入chain-of-thought；3.用户心智模型在知库里丰富。

#### 项目总结与展望：

* 项目评估： 对项目的整体表现进行客观评估，总结成功点和存在的不足。

1.代码不太熟练。2.用户心智模式的解读不丰富。3.前端展现欠缺。

* 未来方向： 基于项目经验，提出未来可能的改进方向和发展规划。

1.寻找可以识别的多智能体直接调用集成，识别表情和情绪；2.丰富用户语言的处理模型和对应的心理学练习模型的匹配度，哪怕简单匹配，哪怕从16型做起；3.增强前端UI\UE能力。

#### 附件与参考资料

[列出项目报告中引用的所有附件和参考资料。]