## 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：戴欣

#### 项目概述：

本项目通过英伟达的模型生成一个太阳系的知识专家，应用场景主要是帮助学生解答太阳系相关问题，亮点是使用了百度百科中的很多知识进行了参考。

#### 技术方案与实施步骤

* **模型选择**：

Embedding模型  "nvidia/nv-embedqa-e5-v5"作为检索参考。而推理模型选择了"ai-mixtral-8x7b-instruct"语言模型，RAG的优势是使用相对小的模型plus知识库就可以训练出接近大模型的效果。

**数据的构建**：通过百度百科的词条进行数据索引(retriever)

* **功能整合**：  通过语音包speech\_recognization将语音识别功能整合到聊天机器人里。

#### 实施步骤：

* 环境搭建： 创建了一个环境，python版本3.9,安装了openai, nvidia 以及speech\_recognization相关库。
* 代码实现： 详见github代码库。
* 测试与调优： 代码进行了一些迭代。
* 集成与部署： 部署需要英伟达相关的库，专业的部署工具。

#### 项目成果与展示：

* 应用场景展示：

|  |
| --- |
| 可以向学生介绍太阳系相关的知识的聊天机器人。 |

* 功能演示：

|  |
| --- |
|  |

可见，该聊天机器人可以更准确的回答太阳系相关的问题

#### 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，api过期，包没有安装等。
* 解决措施： 重新安装api，以及安装对应的包。

#### 项目总结与展望：

* 项目评估：项目可以使聊天机器人理解太阳系相关知识。
* 未来方向：由于网络防火墙的存在，好像无法使用语音识别模块中google云的部分，国外的同学可以试一试。

#### 附件与参考资料

百度百科中关于九大行星和太阳系的介绍。