# 赛题14 对话式数据报表智能助手

（本赛题仅限于本科生选择，命题企业为一等奖获得者提供实习岗位）

**命题企业介绍**

浙江鸿程计算机系统有限公司（简称：鸿程系统）成立于1996年，目前注册资本14196万元，是由中国电信与浙江省能源集团主要战略投资的信息化服务提供商，通过研究大数据、人工智能等创新技术，专注于面向数据大脑、数据运营等数字智能化场景应用，构建数据应用解决方案和复杂信息系统集成解决方案，致力于为通讯、能源、交通等重点企业提供高效的数字化解决方案和技术服务，为智慧城市建设提供城市数字化转型和城市数据智能运营服务。

鸿程系统以“创新、沟通、追求卓越”为企业文化的核心理念，倡导“鹰一样的个人，雁一样的团队”企业精神，聚集一流人才，紧跟世界信息技术发展潮流，创建卓越的信息化服务提供商。

**1、背景说明**

**【项目的行业背景】**

人工智能（AI）技术自20世纪50年代起便已起步，初期旨在创建能模仿人类智能与思维过程的机器，以解决各类问题。尽管如此，AI的进展在随后数十年内因计算资源有限以及数据量不足而严重受阻。直至近年，随着计算机性能的显著提升和数据存储能力的增强，人工智能才迎来了突破性发展，其潜力被全面激发，并已在语音识别、自然语言处理、图像识别、以及机器人控制等众多领域得到实际应用。

随着人工智能领域，尤其是自然语言处理（NLP）和机器学习技术的迅猛进展，机器对人类语言的理解和处理能力得以显著增强。这一进步构筑了对话式智能报表系统发展的技术根基，使这些系统在理解用户查询意图方面更为精准，并能够提供相应的数据分析与报表产出。

而在数字化转型的大背景下，企业对于数据处理和分析的需求日益增长。对话式智能报表系统能够帮助用户通过日常对话的方式轻松获取和分析数据，从而降低了数据分析的门槛、提高数据分析和报表生成的效率，使得非专业人员也能够进行数据分析任务，快速便捷地生成所需报表。

**【项目的业务背景】**

在传统商业智能方案中，数据需要中心化处理与分析，这样能够满足部分通用需求，但无法实现每位数据阅览角色的个性化分析需求。而在实际应用过程中，企业中的每位角色都可能存在个性化的数据需求。在传统方案中，需要专门的数据分析人员理解、分析不同人的需求，通过数据查询和分析，获得相关数据分析结果并制作报表。

而通过大语言模型和BI工具的应用，每个人都可以根据自己的需求，自助进行数据分析，并且方便地将分析结果在团队中共享沟通，通用的报表制作时间也只有使用传统工具的几十分之一甚至几百分之一，让项目实施效率显著增加。

**2、项目说明**

**【问题说明】**

企业用户对报表的需求频率高、需求报表形式多样、数据来源丰富，针对企业各项业务和数据库，需要从不同的数据库中调取、筛选数据，并绘制成适宜的报表，是一项相当耗费精力的工作。应用多轮对话、调取不同的数据库表，实现智能报表。需解决如下问题：

用户对数据存储不了解，仅对报表有需求，需要借助AI实现报表配置；

用户使用语言模式可能与业务数据库的用词不一致，需要借助AI工具明确专业术语；

业务数据来源丰富，结构化数据存储于不同的数据库表中，而智能报表需要实现数据的综合分析。

**【用户期望】**

搭建一个以智能对话为基础的数据智能助手，其功能包括但不限于以下：

应用多轮对话，精准识别用户数据报表需求，包括数据源、实现报表类型；

调用不同的数据库表、智能筛选数据；

应用大语言模型，实现多轮会话，并输出多模态的结果，包括文字、表格、图片等；

针对业务属性的专业术语，实现大语言模型的小范围微调；

较好地实现报表的制作和展示。

**3、任务要求**

1.请项目团队设计开发一个对话式数据报表智能助手工具，并实现本地化部署；

2.交付结果：设计文档、源码、数据标注方法与结果、数据训练方

法与结果；

3.效果评价：可通过自然语言对话，实现多项数据和报表输出；可通过多次对话，逐步明确报表需求。

4.在生态环境、交通、能源等领域自行收集相关数据并实现上述目标。

a)数据库包含至少五个结构化数据集和至少一种非结构化数据；

b)可根据非结构化数据中的信息，检索相关数据，回复自然语言对话中的问题；

例如，根据对话“2022年全国铁路营业里程是多少？”可调取非结构化数据“统计公报”，并给出准确回复“15.5公里”。



c)可根据结构化数据中的字段实现数据汇总、均值、最大值、最小值、标准差等统计量的计算；

d)可根据结构化数据中的字段，实现数据图表的绘制；

例如：根据对话“2024年4月杭州市PM2.5的平均浓度”，可调取

“设区城市环境空气质量自动监测信息（实时）”，并计算字段“城市名称”为杭州市内区/县、“监测时间”在2024年4月范围内，“PM2.5实时浓度”字段数据的平均值；



根据对话“绘制2024年1~4月杭州市各区/县月均PM2.5浓度的变化曲线”，绘制以PM2.5浓度月均值为纵坐标、时间（月份）为横坐标、各区/县为不同图例的折线图。