

# 《数据分析与可视化实践》课程设计任务书

## 一、任务概述

### 1.1 任务名称

- 基于大模型的端到端视频生成。

### 1.2 任务要求

- 复现并改进 MM-StoryAgent 源码。
- 源码链接: [https://github.com/X-PLUG/MM\\_StoryAgent](https://github.com/X-PLUG/MM_StoryAgent)

### 1.3 教学目标

- 掌握端到端视频生成的基本架构;
- 掌握基于大模型的系统集成与开发能力;
- 掌握高水平代码复现能力;
- 培养科研素养与创新能力。

## 二、任务要求

### 2.1 复现原始代码, 实现端到端视频生成 (60 分)

原始代码包含如下几个组件:

- 主题输入: 支持用户输入故事主题或简要描述;
- 故事生成: 通过大语言模型 (LLM) 生成故事情节内容, 包括场景、角色与对话等分段信息;
- 图像生成: 调用文生图 API (如 Stable Diffusion、DALL·E 等) 或本地开源模型, 生成与故事内容相匹配的图像;
- 多媒体整合: 将文本与图像融合, 生成包含字幕解说与配图的视频展示。

## 2.2 代码改进，实现生成内容创新（40 分）

可选的改进包括但不限于如下几个方面（越多越好越难，分数越高）：

- 当前系统仅支持输入简短标题后由大模型自动生成故事。需改进架构，使其支持长文本（如大规模文档、长篇小说或新闻集合）输入并生成视频。
- 当前代码为端到端自动化流程，缺乏用户交互。可设计交互式系统，支持用户与大模型进行对话式生成，允许对不满意的内容进行局部编辑与优化，从而提升用户控制力与生成结果的可用性。
- 现有代码生成一个视频约需 40 分钟，效率较低。改进思路包括优化数据处理流程、提升并行计算效率、引入更高效的推理框架，从而缩短生成时间。
- 现有代码缺少分析环节。可利用现有 NLP/ML 工具（如话题建模、情感分析 API）辅助 LLM 对输入内容进行语义分析与要点提炼，再基于分析结果生成主题视频。
- 现有代码生成内容局限于简短动画，情节单一。可拓展为更丰富的叙事场景，例如：知名科学家的生平介绍、特定领域的科普故事、产品使用教程、近现代重大历史事件回顾等，以提升生成内容的多样性与应用价值。
- 现有视频仅依赖静态图像拼接，镜头语言较为单调。可优化图像到视频的平滑转换、转场效果与字幕排版；同时引入分镜头设计理念，合理安排每帧的顺序、画面元素布局，精心设计镜头的远近切换和局部特写等，增强叙事的沉浸感与表现力。
- 现有代码输出仅限于静态图像。可进一步引入数据可视化要素（如柱状图、折线图、散点图），以及动态图表、交互式叙事组件，以增强信息传递效果。
- 现有代码主要依赖图像生成 API。可进一步接入更先进的视频创作、编辑与剪辑 API（如 Runway、Pika 等），以获得更高质量的视频效果与更多创意表达形式。

### **三、交付要求**

#### **3.1 整体安排**

- 分小组（每组 4-5 人）共同完成代码复现与改进、项目报告撰写、PPT 制作与汇报；
- PPT 汇报时间和地点：2025-11-23 8:30-12:00 45 楼 A108（暂定）；
- 每组上传一份节课材料到智慧树，截止时间：2025-11-30。

#### **3.2 结课材料**

- 项目源码。含源码、注释、环境与运行说明。
- 项目报告。使用指定模板，要求内容详实、图文并茂、写作规范，不少于 12 页。
- 汇报 PPT。含背景与目标、技术路线，实现效果等。

附：报告格式

# 天津大学

## 《数据库应用实践》课程报告



小组编号：

姓名	学号	专业	年级

年 月 日

## **一、 报告摘要**

1、具体解决方案概述。

2、小组分工

## **二、 方案展示**

图文并茂展示全部实现内容。

## **三、 小组总结**

总结自己在上机实验过程中遇到的各类问题、困难以及解决过程中的收

获，对上机实验内容等方面的体会与建议。