**专创融合课程项目要求**

**作业方向：**

**面向图形用户界面自动化操作的AI智能体**

1. **任务描述：**

在移动应用的日常使用中，用户需要频繁完成各种图形用户界面(Graphical User Interface, GUI)任务，例如在微信中查收和回复消息、在日历中添加提醒、在文件管理中查找并发送文档等。这些操作虽然直观，但往往需要多步点击和应用切换，效率较低，也容易造成时间浪费。

本项目的目标是设计并实现一个能够在移动端环境中自动执行GUI任务的智能体系统。该系统以自然语言作为输入，帮助用户高效完成跨应用的复杂操作，从而减少重复劳动，提升使用体验。

主要任务目标包括：

1. 消息处理与回复：能够读取并理解即时通讯应用中的消息，根据用户需求自动生成和发送合适的回复。

2. 文件操作与分享：支持在移动设备中快速检索文件，并在不同应用间完成分享或发送。

3. 日程管理与提醒：根据用户的自然语言指令创建、修改或取消日程提醒，实现智能化的时间管理。

4. 多步任务规划：将用户的复杂需求分解为一系列可执行的GUI操作，并在执行过程中动态调整策略。

5. 上下文感知与个性化：在多轮交互中保持上下文信息，结合用户偏好提供更符合个人需求的执行方式。

通过以上功能，该系统能够让用户只需通过自然语言指令，即可在移动端完成多应用协同的复杂任务，真正实现高效、便捷和安全的智能化交互体验。

相关框架、模型、算法推荐：

* 成熟框架： LangChain、LlamaIndex、Auto-GUI、OpenAI Function Calling/Tool-Use
* 模型：GPT系列、Claude系列、Gemini系列、Vision-Language Models等
* 算法/方法：MobileAgent、MobileAgentV2、ShowUI、CogAgent等

1. **组队：**

由修课同学组队（每组人员≤5人）

1. **输出物：**

中期：中期方案（技术创新报告）+系统演示（可选项）

期末：一篇总体方案（综述报告+技术创新报告）+系统演示

1. **附加分：**

专利申请或学术论文

1. **重要时间节点：**

中期：每组汇报GUI自动化操作智能体技术方案；

第十七周：检查总体实现和各个报告文档；

**\*\*报告与实现要求\*\*：**

**1、综述报告包括：**题目、摘要、国内外研究现状+国内外相关工作陈述、结论和参考文献；

**2、技术创新报告包括：**

内容包括但不限于：引言、对任务的设计和分析，实现方案（方法），实验结果，结论和参考文献。

**3、系统代码实现要求：**

情况一、如无参考开源项目，则在报告中详细阐述技术路线和实现方案；情况二、如参考开源项目，则在报告中注明项目来源，写清项目网址，在报告中写清楚项目创新点，并最终需要在代码中注明每一个创新点。

**注意：如使用开源代码不经创新修改，并且不在报告和代码中注明的小组，一经查实从严处理！！**

**\*\*报告与实现要求\*\*：**

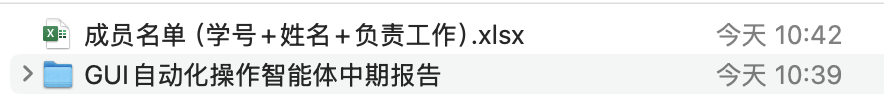
**关于抄袭：**

每个小组选题不同，组内成员互相帮助，组与组之间禁止互相抄袭，抄袭小组和被抄袭小组均得0分：

**实现语言：**

不限

**1、中期报告提交**



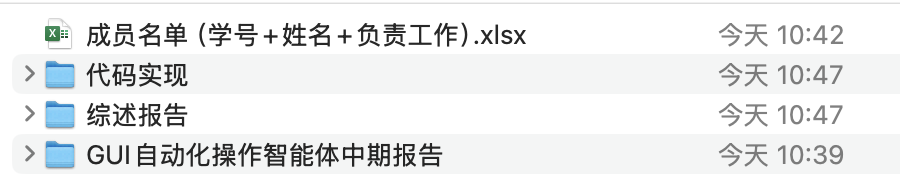
**将提交文件按照以上格式进行整理打包。**

**特别说明：**

**文件夹命名：**按小组号命名，如上图1号小组就是命名为**0.zip**；

**成员名单：**小组成员学号+姓名+每人负责的部分；

**2、期末作业提交：**



**将提交文件按照以上格式进行整理打包。**

**特别说明：**

**文件夹命名：**按小组号命名，如上图1号小组就是命名为**1.zip**；

**代码实现文件夹：**需要注明依赖的库以及其版本号，放到一个独立txt或markdown文件中；

**成员名单：**小组成员学号+姓名+每人负责的部分；

**期末作业提交：**

报告word文档或pdf文档：提交期末技术综述以及技术方案的报告。

代码以及系统演示压缩包：提交系统代码，附演示视频或演示截图等。

**成员名单excel文件：**小组成员学号+姓名+每人负责的部分。

**文件夹命名：**按小组号命名，如上图1号小组命名为**1.zip**。

**\*\*报告与实现要求\*\*：**

**1、综述部分包括：**题目、摘要、国内外研究现状+国内外相关工作陈述、结论和参考文献；

**2、技术方案与创新部分包括：**

内容包括但不限于：引言、对任务的设计和分析，实现方案（方法），实验结果，结论和参考文献。

**3、系统代码实现要求：**

情况一、如无参考开源项目，则在报告中详细阐述技术路线和实现方案；情况二、如参考开源项目，则在报告中注明项目来源，写清项目网址，在报告中写清楚项目创新点，并最终需要在代码中注明每一个创新点。

**注意：如使用开源代码不经创新修改，并且不在报告和代码中注明的小组，一经查实从严处理！！**

**\*\*参考文献\*\***

[1] Wang J, Xu H, Ye J, et al. Mobile-agent: Autonomous multi-modal mobile device agent with visual perception[J]. arxiv preprint arxiv:2401.16158, 2024.

[2] Wang J, Xu H, Jia H, et al. Mobile-agent-v2: Mobile device operation assistant with effective navigation via multi-agent collaboration[J]. Advances in Neural Information Processing Systems, 2024, 37: 2686-2710.

**界面解析/视觉理解相关论文：**

[3] Wan J, Song S, Yu W, et al. Omniparser: A unified framework for text spotting key information extraction and table recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 15641-15653.

[4] Zeng Z, Huang J, Zheng L, et al. UItron: Foundational GUI Agent with Advanced Perception and Planning[J]. arxiv preprint arxiv:2508.21767, 2025.

**GUI Grounding论文：**

[5] Hong W, Wang W, Lv Q, et al. Cogagent: A visual language model for gui agents[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 14281-14290.

[6] Cheng K, Sun Q, Chu Y, et al. Seeclick: Harnessing gui grounding for advanced visual gui agents[J]. arxiv preprint arxiv:2401.10935, 2024.