

需求分析-数据分析智能体多端开发

1. 引言

1.1 编写目的

本软件需求规格说明书用于系统化描述《数据分析智能体多端开发》的业务需求和功能需求，明确系统要实现的功能、性能以及相关约束。本文档面向的主要读者包括：项目开发人员、测试人员、课程指导教师以及项目相关干系人。本说明书将作为后续概要设计、详细设计、编码实现和测试工作的依据与评审标准。

1.2 项目背景

当前高校学生在课程作业、学科竞赛及学生工作中频繁面临数据分析需求，但现有工具存在显著痛点：专业工具（Python、SPSS、R）学习成本高，操作流程复杂，需掌握编程或专业操作技能；轻量化工具（如 Excel）功能有限，难以满足多样化分析与可视化需求；多数工具局限于单一终端，跨设备数据同步不便，实验室电脑完成的分析需手动传输至手机或桌面端；商业工具存在付费门槛，学生团队难以承担长期使用成本。为解决上述问题，结合课程设计要求，项目组拟开发一款支持网页端、移动端 APP、桌面软件的数据分析智能体，集成开源大模型能力，实现“自然语言交互→代码生成→直接分析结果输出”的全流程，降低学生数据分析门槛，同时锻炼团队跨端开发与 AI 集成的工程实践能力。

1.3 术语与缩略语

表 1 术语

术语 / 缩写	含义说明
数据分析智能体	支持自然语言交互、多工具代码生成、数据可视化的跨端分析工具，仅履行数据分析师职责
多端	包含网页端（PC/平板浏览器）、Android/iOS 移动端 APP、Windows/macOS 桌面软件
LLM	Large Language Model（大语言模型），本项目采用 Phi-3-mini 轻量化开源模型
双模式输出	同时支持“生成可复用代码”和“直接输出分析结果（结论+可视化）”
SRS	Software Requirements Specification（软件需求规格说明书）
同步冲突	同一分析记录在多端同时修改导致的数据不一致问题
基础爬虫	仅抓取静态网页中表格数据的功能，严格遵守 robots 协议

1.4 参考资料

1.课程实践教学任务书

- 2.《软件工程（第 5 版）》张海藩著
- 3.同类开源数据分析工具公开文档（LangChain DataAnalysis 示例、DataChat 开源项目）
- 4.Phi-3-mini 官方技术文档、vLLM 量化优化指南
- 5.Flutter 跨端开发官方手册、Electron 快速入门教程
- 6.国家统计局公开数据接口使用规范、robots 协议行业标准
- 7.MySQL 8.0 官方文档、ECharts 可视化开发指南

2. 项目概述

2.1 系统目标

- 1.为学生提供“零成本、易上手、多端可用”的数据分析工具，覆盖基础数据统计、可视化与代码生成需求；
- 2.实现自然语言到数据分析指令的转化，输出 Python/SPSS/R 可复用代码（带注释）+ 直接输出结构化分析结论与可视化图表；
- 3.支持本地文件上传、联网检索公开数据、基础静态网页爬虫三种数据输入方式，丰富数据来源；
- 4.保障多端数据实时同步，实现分析记录跨设备无缝访问与管理；
- 5.8 周内完成最小可行产品，满足学生课程作业、竞赛、学生工作的核心数据分析需求；
- 6.架构设计兼顾可扩展性，便于后续新增分析工具或优化模型。

2.2 系统范围与定位

包含功能范围	暂不支持功能
CSV/Excel 文件上传（≤10MB）	复杂机器学习模型训练与部署（如回归、聚类算法）
自然语言交互，双模式输出（代码+分析结果）	动态网页爬虫与大规模数据抓取
折线图/柱状图/饼图/散点图生成与 PNG 导出	在线支付、多用户协作与数据共享功能
折线图/柱状图/饼图/散点图生成与 PNG 导出	跨学校/跨区域数据访问与同步
基础静态网页表格爬虫（遵守 robots 协议）	高并发访问支持（仅适配单用户/小规模测试场景）
多端数据同步（分析记录、用户配置）	付费第三方服务集成（短信验证、商业 API 调用）
分析记录管理（查看、检索、删除、导出 Word 报告）	自定义图表样式与复杂数据清洗功能
代码复制/下载、数据本地缓存、断网查看历史记录	...

表 2 功能介绍

2.3 用户角色与典型用户

角色	描述
学生用户	核心用户，涵盖计算机、数据科学、经济学、管理学等专业大二至大四学生；每周有 1-3 次数据分析需求（课程作业占 65%、竞赛占 20%、学生工作占 15%）；具备基础电脑操作能力，熟悉 Excel 基础功能，Python/SPSS/R 使用熟练度较低；偏好“操作步骤少、反馈及时、零成本”的工具，需多端使用场景支持
课程指导教师	次要用户，用于审核项目功能完整性、评估技术实现合理性与需求落地性；关注功能是否符合课程要求、技术选型是否适配学生能力、文档是否规范
游客	未登录用户，仅可浏览系统功能介绍、查看操作演示视频，无法使用核心分析功能

表 3 用户说明

典型用户描述

- 李同学（大三经管专业）：需完成课程作业中的成绩统计与可视化，熟悉 Excel 但 Python 基础薄弱，希望通过简单交互快速获取分析结果、图表及可直接复用的代码，需在实验室电脑与手机间同步数据；
- 王同学（计算机专业）：参与学科竞赛需处理公开数据，需要工具支持联网检索与基础爬虫功能，希望导出代码用于竞赛报告，需在 Windows 桌面端与 Android 手机间切换使用；
- 张同学（社团负责人）：需统计社团活动报名数据，通过学校官网静态网页获取报名信息，需生成饼图展示不同学院报名占比，同时需要 SPSS 操作步骤用于后续数据存档。

2.4 业务场景示例（用户故事）

场景 1：课程作业成绩分析

李同学打开网页端登录账号，上传“班级概率论成绩.csv”文件，在需求输入框中输入“统计 80-90 分人数、计算班级平均分并生成柱状图，输出 Python 代码及操作步骤”，选择 Python 工具后点击“开始分析”。系统解析文件后，生成完整 Python 代码、分步操作指南与柱状图，李同学复制代码提交作业，导出图表插入报告，并通过同步功能在手机端查看结果。

场景 2：竞赛公开数据补充

王同学在移动端 APP 输入需求“获取 2023 年全国高校毕业生人数与 2022 年对比数据，生成折线图”，点击“联网检索”。系统调用国家统计局公开接口获取数据，完成分析后生成对比折线图与 SPSS 操作流程，王同学将结果同步至桌面端，用于竞赛报告撰写。

场景 3：社团活动数据统计

张同学通过桌面端“爬虫工具”输入学校官网活动报名页面 URL，抓取报名数据后，输入需求“统计不同学院报名人数占比，生成饼图并输出 R 代码”。系统完成分析后，张同学导出饼图用于社团工作总结，同时将 R 代码分享给团队成员用于数据备份。

3. 需求总体说明

3.1 功能需求概述

系统功能需求划分为六大核心模块，模块间通过明确接口通信，确保低耦合：

- 1. 数据输入模块：**负责文件上传解析、联网数据检索、基础爬虫数据抓取；
- 2. 智能分析模块：**核心模块，含自然语言需求解析、LLM 调用、多工具代码生成、职责边界约束（仅响应数据分析需求）；
- 3. 可视化模块：**负责图表生成、格式转换与导出处理；
- 4. 多端同步模块：**处理同步触发、冲突检测与解决、跨端数据传输；
- 5. 用户管理模块：**涵盖注册（校园邮箱验证）、登录、个人信息维护、密码修改；
- 6. 辅助功能模块：**分析记录管理、代码复制/下载、Word 报告导出、日志记录。

3.2 非功能需求概述

除功能需求外，系统需满足以下非功能性要求：

- 1. 性能需求：**实验室环境下，本地数据分析响应时间 ≤ 10 秒，联网检索 ≤ 15 秒，多端同步延迟 ≤ 30 秒；
- 2. 安全性需求：**密码加密存储、权限分级控制、数据隔离、爬虫合规性约束；
- 3. 易用性需求：**核心操作步骤 ≤ 3 步，界面风格统一，错误提示明确，初次使用 5 分钟内可掌握基础功能；
- 4. 可靠性需求：**网络中断后支持断点续传，关键操作原子性，异常场景友好提示；
- 5. 可维护性需求：**模块解耦、配置集中管理、代码注释规范，便于后期功能扩展；
- 6. 兼容性需求：**适配主流浏览器、移动操作系统与桌面系统，安装包体积可控。

3.3 约束与假设

- 开发约束：团队 5 人，开发周期 8 周，技术栈限定为 Flutter+Electron+Flask+MySQL+Phi-3-mini；
- 硬件约束：运行环境为学生个人电脑或校园免费云服务器（ ≥ 2 核 8GB 内存），无高端 GPU 资源；
- 成本约束：全流程使用开源工具与免费服务，无商业软件授权或付费 API 调用；
- 合规约束：爬虫功能严格遵守 robots 协议，访问频率 ≤ 1 次/秒，仅抓取非涉密、非付费内容；
- 用户假设：所有用户均为本校学生，通过校园邮箱注册，数据仅用于个人学习用途，不用于商业传播；
- 网络假设：默认用户使用场景具备网络连接，断网时仅支持查看本地缓存的历史记录。

4. 功能需求详细说明

4.1 业务用例

角色	核心用例
学生	注册账号、登录系统、修改个人信息、修改密码、上传文件、输入分析需求、选

用户	择分析工具、联网检索数据、启动基础爬虫、查看分析结果、复制代码、导出图表、导出报告、同步分析记录、查看历史记录、检索历史记录、删除记录、处理同步冲突、断网查看缓存
课程指导教师	浏览系统功能、查看演示版本、审核功能完整性、评估需求落地性
游客	浏览功能介绍、查看操作指南、观看演示视频

表 4 业务用例表

4.2 用例列表

用例编号	用例名称	参与者	简要说明
UC001	用户注册	学生用户	通过校园邮箱注册账号，设置密码并完成基础信息填写
UC002	用户登录	学生用户	输入校园邮箱与密码登录系统，获取核心功能使用权限
UC003	个人信息维护	学生用户	修改用户名、绑定设备、更新个人偏好设置
UC004	密码修改	学生用户	验证原密码后修改新密码，支持密码找回（邮箱验证）
UC005	文件上传解析	学生用户	上传 CSV/Excel 文件，系统解析并展示数据预览（前 10 行）
UC006	自然语言分析	学生用户	输入分析需求与目标工具，系统生成代码、操作步骤与结果
UC007	联网检索数据	学生用户	输入检索关键词，系统调用公开接口获取数据并支持分析
UC008	基础爬虫	学生用户	输入静态网页 URL，抓取表格数据并解析为标准化格式
UC009	图表生成与导出	学生用户	系统根据分析结果生成图表，支持 PNG 格式导出至本地
UC010	代码复制	学生用户	一键复制生成的 Python/SPSS/R 代码至剪贴板
UC011	报告导出	学生用户	将分析结果、代码、图表导出为 Word 格式报告
UC012	多端同步	学生用户	手动触发或自动同步分析记录与用户配置至多端
UC013	同步冲突处理	学生用户	当多端修改同一记录时，选择保留版本或合并数据
UC014	历史记录管理	学生用户	查看、检索、删除历史分析记录，支持按时间/工具筛选
UC015	断网查看缓存	学生用户	断网状态下查看已缓存的历史分析记录与本地数据

表 5 用例列表

4.3 关键用例详细说明（示例）

用例编号：UC006

用例名称：自然语言分析

参与者：学生用户

触发条件：用户登录后，完成数据输入（文件上传/联网检索/爬虫），点击“开始分析”按钮

前置条件：

- 用户已注册并成功登录系统；
- 已完成数据输入，系统获取标准化原始数据；
- 用户输入的需求为数据分析相关（非其他领域需求）；
- 选择的分析工具为支持类型（Python/SPSS/R）。

后置条件：

- 系统生成对应工具的分析代码、详细操作步骤；
- 生成分析结果与可视化图表；
- 分析记录（含原始数据、代码、结果、图表）保存至用户账号；
- 同步模块标记该记录为“待同步”状态。

主成功场景（基本流程）：

- 用户在数据预览页面输入自然语言分析需求（如“统计各分数段人数并生成柱状图”）；
- 用户从下拉框选择目标分析工具（如 Python）；
- 用户选择输出模式（“仅生成代码”或“代码+直接分析”）；
- 用户点击“开始分析”按钮；
- 系统校验需求合法性（是否为数据分析相关内容）；
- 系统调用智能分析模块，传入原始数据信息、分析需求、工具类型与输出模式；
- 智能分析模块通过 Prompt 模板调用 Phi-3-mini 模型，生成对应工具的分析代码与分步操作步骤；
- 若用户选择“代码+直接分析”模式，系统在本地沙箱环境执行代码，生成结构化分析结果；
- 系统调用可视化模块，根据分析结果生成指定类型图表（如柱状图）；
- 系统将分析记录（原始数据 ID、需求文本、代码、操作步骤、结果、图表数据）写入数据库；
- 系统在客户端展示代码预览、操作步骤、分析结果与图表；
- 客户端提供“复制代码”“导出图表”“导出报告”“同步记录”按钮。

扩展场景（异常情况）：

4a. 需求超出数据分析范畴：

1. 系统提示“仅支持数据分析相关操作，请输入合法需求”；
2. 用户修改需求后重新提交（返回步骤 1）。

工具类型选择错误（如选择未支持工具）：

- 系统提示“当前仅支持 Python/SPSS/R 工具，请重新选择”；
- 用户更换工具后重新提交（返回步骤 2）。

代码执行失败（如数据格式错误、缺失值问题）：

- 系统检测错误类型，提示具体原因（如“数据中存在缺失值，请清理后重试”）；
- 用户选择“自动清理数据”（系统自动处理缺失值）或“手动修改数据”（跳转数据编辑页面）；
- 处理完成后重新执行分析（返回步骤 6）。

LLM 调用超时：

- 系统提示“分析服务暂时繁忙，请重试”；
- 系统提示“分析服务暂时繁忙，请重试”；
- 用户点击“重新分析”按钮，系统重新发起 LLM 调用（返回步骤 6）。

5. 非功能需求详细说明

5.1 性能需求

- 单用户操作场景下，本地数据分析响应时间 ≤ 10 秒（处理 ≤ 1000 行数据）；
- 联网检索数据响应时间 ≤ 15 秒（含数据获取、分析、图表生成）；
- 基础爬虫单页面抓取时间 ≤ 30 秒，支持同时抓取 1 个页面（避免资源占用过高）；
- 多端同步延迟 ≤ 30 秒，同步数据量 $\leq 10\text{MB}$ 时无明显卡顿；
- 前端页面加载时间 ≤ 3 秒，图表渲染时间 ≤ 2 秒；
- 支持单用户存储 ≤ 50 条分析记录，单条记录数据量 $\leq 5\text{MB}$ ；
- 桌面端安装包大小 $\leq 200\text{MB}$ ，移动端 APP 占用内存 $\leq 100\text{MB}$ ；
- 数据库查询响应时间 ≤ 1 秒（单表数据量 ≤ 1 万条）。

5.2 安全性需求

- 用户密码采用 MD5 加密存储，禁止明文保存或传输；
- 未登录用户仅可浏览功能介绍，无法访问核心分析、数据存储与同步功能；
- 分析数据按用户隔离存储，仅当前用户可查看、修改或删除自己的记录；
- 爬虫功能添加访问频率限制（ ≤ 1 次/秒），校验目标网站 robots 协议，禁止抓取涉密、付费或违规网站；
- 对用户输入的需求文本、URL、文件内容进行过滤，避免 SQL 注入、恶意代码执行、文件上传漏洞等安全风险；
- 敏感数据（如用户邮箱、分析记录 ID）传输过程中采用 HTTPS 加密；
- 支持账号异常登录检测（如异地登录），并向用户邮箱发送提醒。

5.3 可用性与易用性需求

- 系统界面风格统一，导航栏位置固定，关键操作按钮（上传、分析、导出）突出显示（颜色区分、尺寸放大）；
- 核心分析流程操作步骤 ≤ 3 步（数据输入 \rightarrow 需求提交 \rightarrow 查看结果），避免复杂层级跳转；
- 初次使用用户可通过内置操作指南（图文结合）快速上手，5 分钟内掌握基础功能；
- 错误输入或操作失败时，提供明确、可操作的提示信息（如“文件格式错误，请上传 CSV/Excel 文件（.csv/.xlsx/.xls）”），而非简单“操作失败”字样；
- 支持操作撤销（如删除记录后 30 秒内可恢复）与历史记录回溯；
- 关键操作（如删除记录、同步冲突处理）需用户二次确认，避免误操作；
- 多端界面适配不同设备尺寸：移动端支持手势缩放图表、下拉刷新同步；

- 桌面端支持文件拖拽上传、快捷键操作（如 Ctrl+C 复制代码）；
- 网页端适配屏幕分辨率 $\geq 1366 \times 768$ px，支持响应式布局。

5.4 可靠性与容错需求

- 网络短暂中断后，系统可在恢复连接后自动续传未完成的操作（如同步、导出），不导致数据丢失或混乱；
- 关键数据操作（如分析记录保存、同步、代码执行）保证原子性，要么全部成功，要么全部回滚；
- 系统异常（如 LLM 推理失败、爬虫访问受限、数据库连接异常）时，返回友好提示页面，隐藏服务器错误详情（如堆栈信息）；
- 支持数据本地缓存，断网状态下可查看已同步的历史分析记录与本地上传的文件；
- 文件上传过程中支持断点续传，避免网络中断后重新上传；
- 系统运行过程中自动记录关键操作日志（登录、分析、同步、错误），便于问题排查与故障恢复。

5.5 可维护性与可扩展性需求

系统采用分层架构（表示层、业务逻辑层、数据访问层），模块间解耦，支持单独替换某一层技术栈（如前端框架从 Flutter 替换为 React Native）；

新增分析工具（如 MATLAB、Stata 代码生成）时，无需大范围修改核心逻辑，仅需扩展智能分析模块的代码生成模板；

配置参数（如文件大小限制、图表类型、同步频率、LLM 模型参数）集中存储在配置文件，支持动态调整，无需修改代码；

代码添加必要注释（模块功能、核心函数、关键变量），注释覆盖率 $\geq 60\%$ ，便于后续维护与迭代；

数据库设计预留扩展字段（如用户表预留“ext_info”字段、分析记录表预留“ext_params”字段），支持新增需求时无需修改表结构；

预留第三方接口接入点（如新增公开数据接口、替换 LLM 模型），便于后期功能扩展。

5.6 兼容性需求

- 网页端兼容主流浏览器：Chrome 90+、Firefox 88+、Edge 90+、Safari 14+；
- 移动端支持操作系统：Android 10.0+、iOS 14.0+；
- 桌面端支持操作系统：Windows 10+（32/64 位）、macOS 11+；
- 支持的文件格式：CSV（.csv）、Excel（.xlsx/.xls），导出格式：PNG（图表）、Word（.docx）；
- 兼容不同屏幕尺寸：移动端屏幕 $\geq 375 \times 667$ px，桌面端屏幕 $\geq 800 \times 600$ px；
- 支持在校园网、家庭网络等不同网络环境下正常使用（带宽 ≥ 1 Mbps）。

6. 数据需求与业务规则

6.1 关键业务数据项说明（概念层面）

数据对象	业务含义	核心属性
用户（User）	系统注册账号，用于身份认证与权限控制	账号 ID、用户名、校园邮箱、加密密码、注册时间、账号状态、个人偏好
分析记录（AnalysisRecord）	用户的完整分析流程数据，用于存储与同步	记录 ID、用户 ID、数据名称、原始数据 ID、分析需求、工具类型、生成代码 ID、分析结果、图表数据 ID、创建时间、更新时间、同步状态
原始数据（RawData）	用户上传、检索或抓取的原始数据	数据 ID、数据内容（JSON 格式）、数据来源（文件/检索/爬虫）、文件名称（如有）、数据大小、创建时间
生成代码（GeneratedCode）	系统输出的分析代码及操作步骤	代码 ID、Python 代码、SPSS 代码、R 代码、操作步骤描述、代码长度
可视化结果（VisualResult）	分析生成的图表数据与导出信息	结果 ID、图表类型（折线图/柱状图/饼图）、图表数据（JSON 格式）、导出文件路径、生成时间
同步记录（SyncRecord）	多端同步状态信息，用于冲突检测	同步 ID、用户 ID、设备类型、设备 ID、最后同步时间、同步状态（同步中/成功/失败）、未同步记录 ID 列表
设备信息（Device）	用户绑定的多端设备信息	设备 ID、用户 ID、设备类型、设备名称、绑定时间、最后登录时间

6.2 业务规则与约束

- 1.用户注册规则：仅支持校园邮箱注册（邮箱后缀为学校域名，如“@xxx.edu.cn”）；
- 2.密码长度 ≥ 6 位，包含字母与数字（增强安全性）；
- 3.用户名长度 ≤ 20 字符，不支持特殊符号（除下划线）。
- 4.数据输入规则：上传文件大小 $\leq 10\text{MB}$ ，仅支持 CSV/Excel 格式，超出限制则提示“文件过大，请上传 $\leq 10\text{MB}$ 的 CSV/Excel 文件”；
- 5.联网检索仅支持公开的非涉密数据（如政府统计、高校公开数据），不支持检索付费或违规数据；
- 6.爬虫仅支持静态网页表格数据抓取，URL 需以“http://”或“https://”开头，且目标网站 robots 协议允许抓取；
- 7.原始数据行数 ≤ 1000 行，超出则提示“数据量过大，请精简数据后重试”（适配学生场景与性能需求）。
- 8.分析与代码生成规则：仅响应数据分析相关需求，拒绝聊天、解题、娱乐等非数据分析请求；
- 9.生成的代码长度 ≤ 500 行，操作步骤描述 ≤ 10 条，确保易用性与可读性；
- 10.支持同时生成 1 种工具代码（用户选择），暂不支持多工具代码同时输出；
- 11.代码需包含必要注释，操作步骤需明确、可落地（如“打开 SPSS→导入数据→点击分析→选择描述统计→设置变量→点击确定”）。
- 12.存储与同步规则：单用户最多存储 50 条分析记录，超出则提示“记录数量已达上限，请删除不常用记录后重试”；
- 13.同步优先级：按更新时间排序，最新修改的记录优先同步；
- 14.同步冲突处理规则：当同一记录在多端均被修改时，以用户手动选择的版本为准，或仅保留最新修改的版本（用户可选择）；
- 15.分析记录删除后不可恢复，删除前需用户二次确认。
- 16.导出规则：图表导出格式仅支持 PNG，分辨率 $\geq 1080 \times 720\text{px}$ ，保证清晰度；

17. Word 报告包含固定结构：报告标题、分析需求、原始数据预览、生成代码、操作步骤、分析结果、图表，便于作业或报告排版；
18. 导出文件命名规则：“分析记录_数据名称_工具类型_时间戳”（如“分析记录_概率论成绩_Python_202505201430”）。

7. 外部接口需求

7.1 用户接口需求（UI 层面）

1. 系统首页需包含：顶部：系统 Logo、导航菜单（首页、我的记录、帮助中心）、登录/个人中心入口；
2. 左侧：功能入口区（文件上传按钮、联网检索输入框、爬虫 URL 输入框）；
3. 中间：需求输入框（支持多行输入、占位提示文本）、工具选择下拉框、输出模式选择按钮、“开始分析”按钮；
4. 右侧：结果展示区（分页标签：代码、操作步骤、结果、图表）；
5. 底部：导出按钮组（复制代码、导出图表、导出报告）、同步按钮、清空按钮。
6. 个人中心页面需包含：基本信息：用户名、校园邮箱、注册时间；
7. 功能入口：账号设置（修改密码、个人信息）、设备管理（绑定设备列表、解除绑定）、同步管理（手动同步、冲突处理）；
8. 偏好设置：默认分析工具、图表默认类型、同步频率（自动/手动）。
9. 我的记录页面需包含：顶部：搜索框（按数据名称检索）、筛选框（按数据来源/工具类型/时间筛选）；
10. 中间：记录列表（每条记录含数据名称、分析时间、工具类型、图表缩略图、操作按钮）；
11. 操作按钮：查看详情、再次分析、导出、删除。
12. 通用 UI 规则：颜色方案：主色调采用蓝色（专业、稳重），辅助色采用灰色（中性、不干扰），按钮高亮色采用橙色（突出操作）；
13. 字体：统一使用无衬线字体（如微软雅黑、Roboto），标题字号 $\geq 16\text{px}$ ，正文字号 $\geq 14\text{px}$ ，确保可读性；
14. 加载状态：所有耗时操作（分析、同步、导出）需显示加载动画与进度提示（如“分析中...30%”）；
15. 错误提示：采用红色图标+文字说明的形式，位置靠近操作区域，便于用户关联。

7.2 外部系统接口

1. 公开数据接口：调用国家统计局、教育部等公开数据接口，获取宏观统计数据；
2. 接口格式：RESTful API，数据返回格式为 JSON；
3. 调用约束：遵守接口限流规则，不频繁调用（ ≤ 1 次/5 秒）。
4. 校园邮箱接口（可选）：用于用户注册验证、密码找回，调用学校邮件服务器接口发送验证邮件；
5. 接口要求：支持 SMTP 协议，需配置校园邮箱账号与授权码。
6. 无其他外部系统接口依赖（如支付、短信验证等付费服务），降低开发与维护成本。

7.3 硬件与其他接口

1. 客户端硬件接口需求：支持文件读取（本地文件上传）、存储（导出文件、本地缓存）；
2. 移动端支持相机（可选，用于拍摄表格图片识别数据，暂不实现）、网络连接（Wi-Fi/移动网络）；
3. 桌面端支持文件拖拽、快捷键操作、本地数据库存储（SQLite 缓存）。

4.无特殊硬件接口需求（如扫码枪、打印机等），适配学生常规设备。

8. 其他需求

8.1 法规与政策相关需求

- 1.遵守《网络安全法》《个人信息保护法》，不收集、存储、传输用户敏感信息（如身份证号、完整学号、手机号）；
- 2.用户数据仅用于个人学习用途，不进行商业变现、共享或泄露；
- 3.爬虫功能严格遵守《互联网爬虫自律公约》与目标网站 robots 协议，禁止抓取违规内容，保留爬虫操作日志（≥30 天）以备查询；
- 4.导出的报告、图表、代码需注明数据来源（如“数据来源：国家统计局公开数据”“代码生成：数据分析智能体”），尊重数据版权与知识产权。

8.2 数据备份与日志需求

- 1.数据备份：服务器端数据库每周自动备份 1 次，备份文件保留 1 个月；
- 2.客户端支持用户手动导出备份分析记录（JSON 格式），便于数据迁移或恢复。
- 3.日志记录：记录关键操作日志：用户登录/退出、文件上传、分析任务、同步操作、导出操作、错误信息；
- 4.日志内容包含：操作时间、用户 ID、设备类型、操作类型、操作结果、错误码（如有）；
- 5.日志存储期限：服务器端日志保留 3 个月，客户端日志保留 1 个月，超出自动清理。

8.3 项目交付与验收标准

- 1.功能验收：所有“必须实现”的用例（UC001-UC015 中的核心用例）需正常可用，无严重功能缺陷；
- 2.性能验收：满足 5.1 节性能指标要求，在实验室环境下连续测试 1 小时无明显性能下降；
- 3.安全性验收：通过基础安全测试（SQL 注入、XSS 攻击、文件上传漏洞测试），无高危安全隐患；
- 4.兼容性验收：在指定的浏览器、操作系统版本上测试核心功能，无兼容性问题；
- 5.文档验收：交付完整的需求规格说明书、概要设计文档、测试报告、用户手册；
- 6.代码验收：代码注释覆盖率≥60%，无明显冗余代码，核心模块单元测试覆盖率≥50%。