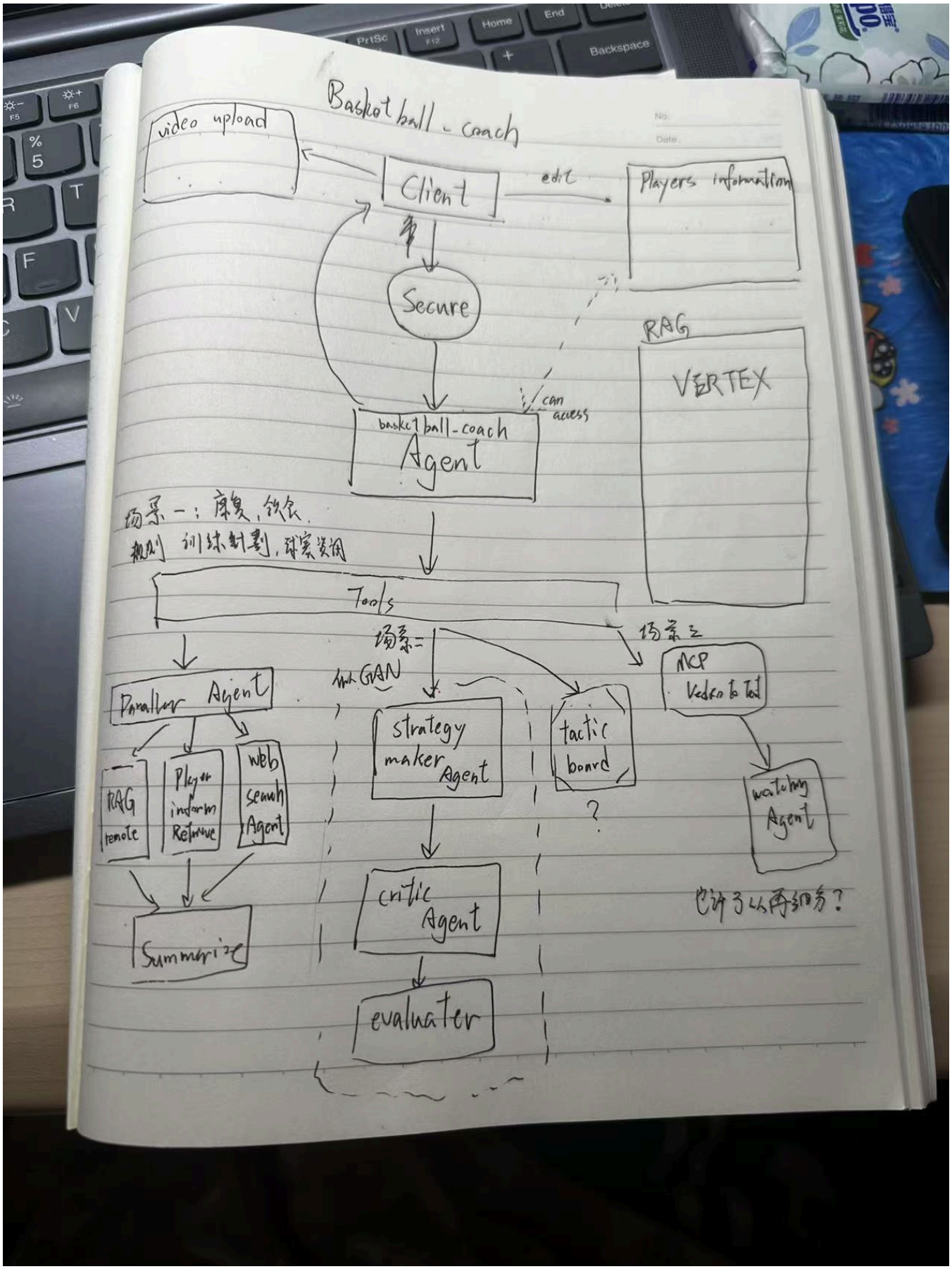
V3 開發手冊☺

潦草的架构草稿。（等待被figma）


Basket ball - coach
video upload
Client
edit
Players information
Secure
RAG
VERTEX
can access
basket ball - coach
Agent
场景一：康复，饮食，规划 训练计划，球赛资讯
Tools
场景二
场景三
MCP
Video to text
Parallar Agent
with GAN
strategy maker Agent
tactic board
RAG remote
Player inform Retrieve
web search Agent
watching Agent
critic Agent
Summarize
evaluater
?
也许可以再细分？

Developer Log

2025.6.25更新 -dyt

完成了prompt的基本要求，还需进行temperature的调参等。
同步在报告中进行了更新（报告就是main.tex,编译命令是该文件开头的四行注释，如果有参考文献直接在example_paper.bib中修改。在basketball_coach的文件夹中加上aibasis_report文件夹只是为了方便编译（在同一个根目录而已qwq

2025.6.25更新 -xsy

- 调整了文件结构
- 进行了文件重命名
  - `prompt.py` -> `prompts.py` （已集成dyt的更新）
  - `strategy_maker.py` -> `strategy.py`
  - `toolkit.py` -> `guardrail.py`
  - `knowledge_collect.py` -> `search.py` （RAG search Agent新增在这里）
  - `game_video_analysis.py` -> `video.py`
- 新增文件
  - `config.py`主要针对.env和主程序之间的配置链接
  - `service_key.json`与RAG系统的配置有关
- 功能完成
  - 成功部署Vertex RAG Engine，Agent可以正常连通
  - 调整了root_agent的prompt来适应变化和improve Guardrail

任务追踪

- hbc
  - 维护Vertex RAG Engine内容
    - [ ] 完善和更新RAG的参考资料（小红书/公众号等搬运、专业相关书籍等、主要是一些浏览器无法直接获取的内容）（支持文件类型）
  - 用户的本地球员资料库设计
    - [ ] 设计相关的文件结构
    - [ ] 设计资料库具体字段（例如球员姓名、球员位置、打球风格etc. ）（字段的内容可以比较灵活，毕竟语言模型能看懂文字）

        - [ ] 资料库实现（方式自选）
        - [ ] 写一个tool 和/或 agent可以获取这个资料库的资源（保证健壮性:找不到或者出现异常需要如实告知自己找不到）
        - [ ] 撰写说明文档：指导用户维护自己的资料库
        - [ ] （可选）（锦上添花）写一个tool 和/或 agent可以更新资料库，令用户可以直接与agent对话来维护这个数据库
    - Gloudglue用户影片上传与Agent影片获取
        - [ ] 写一个tool 和/或 agent可以成功调用这个MCP并获取结果
        - [ ] 写一个Agent（如果上一步已经完成的话可以略过）可以完整地：
            a. 和用户对话
            b. 调用上一点的工具
            c. 解读工具返回的内容，并合适地回应用户
        - [ ] 调试和优化prompt
    - [ ] 完成 技术报告：实现细节 中与自己的工作相关的部分
    - 顺带一提Vertex上部署貌似素要钱滴。。。所以不部署了。。
- xsy
    - [x] 在Vertex上搭建RAG
    - [ ] 把架构图里的结构用代码写出来
      确定具体应用的工具和技术（什么tool什么MCP 操作战术板是否可行）
    - 尽可能满足作业列出的技术深度和深度要求
    - 确保用户交互不出问题（至少adk web没问题）
    - session的恢复（目前都没有在保存对话的memory之类的）
    - 研究一下litellm集成如何支持vlm
    - 集成TTS功能
    - CLI用户助手（下载所有支持、检查和完成配置、切换使用模型 etc）
    - [ ] 完成 技术报告：实现细节 中与自己的工作相关的部分
    - [ ] 完成 技术报告：需求分析、技术选型 和 系统架构设计图
    - [ ] 完成 技术报告：需求分析、技术选型
- dyt
    - 技术报告
        - [x] 研究提供的latex模板
        - [ ] 技术报告内容整合与优化
    - Agent系统架构提示词工程`prompts.py`
        - 填充提示词（与系统架构implement同步）
        - 提示词调优（包括高级参数temperature/top_p的调优）（作业要求展示调优过

程....虽然不知道具体是要怎么展示但可以适当记录一下自己用的输入、prompt、参数和输出）
    - improve、补充系统架构
      - [ ] 提示词安全补充 `guardril.py`
        可以是自己写prompt 也可以用一个Agent来负责检测、目前只完成了Client输入的过滤（可以补充对"函数调用参数合法性检测"、“工具返回值检测”等一系列的应用健壮性的保护）
      - [ ] 结构化输入输出设计
        优化整个Agent系统对结构化输入输出的应用（例如某些AgentTool也许能用schema优化其调用）
  - [ ] 测试整体应用（保证功能性和健壮性：cover所有的功能，试试看不同的use case，试试看不安全/不相关的user inquiry etc.）
  - [ ] 完成 技术报告：评估对比 （可以和主流产品对比？）
  - [ ] 完成 技术报告：实现细节 中与自己的工作相关的部分
- 收尾
    - 部署方案README
    - 完整技术文档（要求：用户手册+API?文档）
    - 演示视频（要求：功能演示及技术解说）
    - 真实用户使用收集(以及star)
    - 技术报告：反思

ddl：6.27，啊啊啊尽量26号能把"收尾"之外的东西搞定吧.....

注意事项

更新requiements.txt

你在开发过程中可能新增了对一些第三方库的使用，导致需求的更新。如果发生这种情况，可以在工作路径下面执行:

```
pipreqs . --encoding=utf8 --force
```

来生成新的`requirements.txt`

你可能需要先在当前环境中安装pipreqs这个库:

```
pip install pipreqs
```

代码即文档

注意类、方法、函数、变量命名的可读性。

变量命名规范：全小写（非常量）/全大写（常量），可以使用下划线。

不要忘记留下必要的docstring和comment。（时间充裕的话）

资源整理

Document Link

- （需要VPN）Gemini API doc
- （需要VPN）Google ADK doc
- Siliconflow API doc
- Gloudglue doc

Prompt调试工具

- (LLM app)Gemini
- (Playground 单个prompt)Gemini
- (Playground 单个prompt)Deepseek & Qwen
- (整个Agent系统的调试)ADK WEB Evaluation
- (AI助手)Prompt 优解
- (AI助手)Prompt Pilot

Prompt准则和示范样例

- (其他篮球Agent)豆包
- (其他篮球Agent)GPT
- (文档)各类型prompt规范参考LLM Agent
- (文档)各类型prompt规范参考LLM Agent 和 tool设置
- (文档)各类型prompt规范参考Agent 与 sub Agent
- (文档)各类型prompt规范参考Adding Safety

- (文档)各类型prompt规范参考用LLm来做SafetyGuardrail
- (github贡献)Agent Example1(相对简单)(https://github.com/google/adk-python/tree/main/contributing/samples)
- (github贡献)Agent Example2(https://github.com/google/adk-samples/tree/main/python)

Vertex AI RAG

- RAG simple example
- RAG example
- Youtube: Build Your First RAG Agent with Agent Development Kit
- (文档)测测你的Agent
- (文档)部署你的Agent
- (文档)VertexAiRag记忆（示范为对话记忆）
- (文档)VertexAiRAG搜寻
- 推荐的篮球公众号:撩篮球

影片交互

- Cloudglue MCP Server Document

之前note.md遗留的知识视频

How to Protect your LLM

https://www.promptingguide.ai/zh/risks/adversarial#参数化提示组件
https://www.youtube.com/watch?v=6bYGhY9HB8k
https://www.youtube.com/watch?v=jrHRe9lSqqA
https://zhuanlan.zhihu.com/p/30480330292

What is structured output:

https://www.youtube.com/watch?v=xpvFinvqRCA

What is an Agent:

https://openai.github.io/openai-agents-python/
https://zhuanlan.zhihu.com/p/24432308656
https://zhuanlan.zhihu.com/p/657937696

https://www.zhihu.com/question/1894891236617332066/answer/1900585340592424543
https://zhuanlan.zhihu.com/p/32230066307

What is agentic workflow:

https://www.anthropic.com/engineering/building-effective-agents