

# 论文思路

## 摘要

## 第一章 绪论

### 1. 研究的背景和意义

### 2. 视觉语言模型研究现状

当前多模态agent模型能力相比LLM还有差距，达不到军用标准，需要进一步微调，构建自动化训练数据生成工具能够大幅提高效率。

### 3. 基于大模型的Agentic AI 研究现状

### 4. 目标检测识别领域智能问答系统研究现状

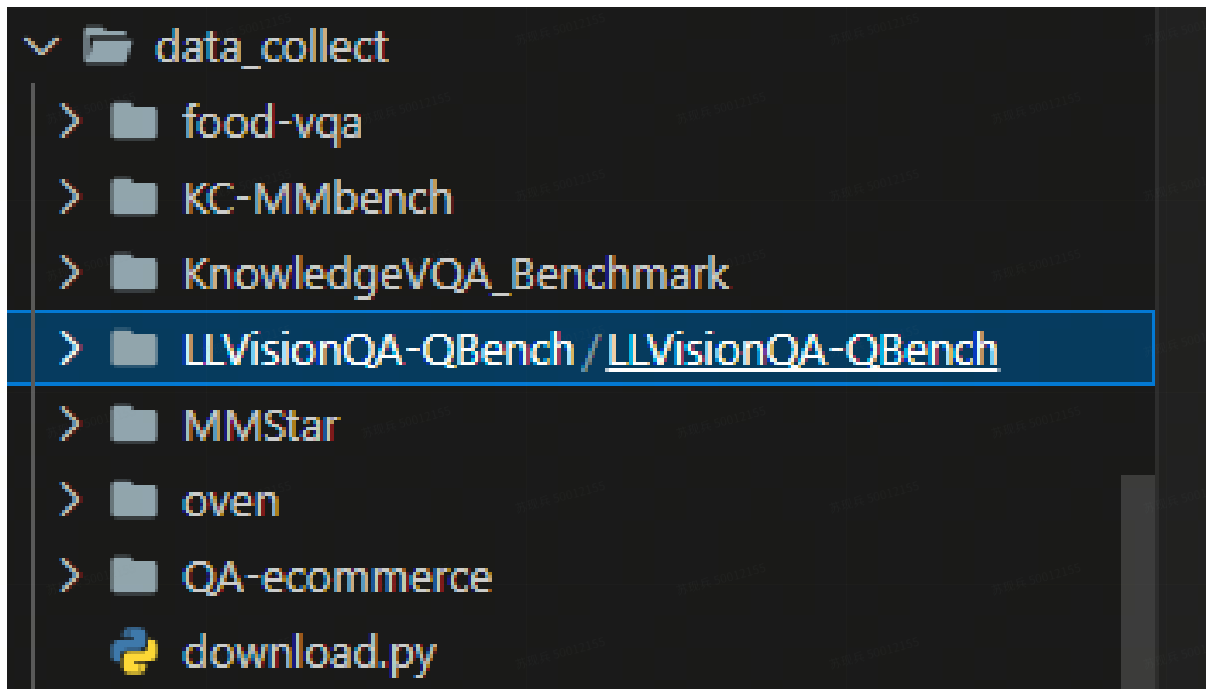
### 5. 研究内容

### 6. 论文结构

## 第二章 模型领域化微调


### 7. 预训练数据准备

原始：



清洗完成：

轮次
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1

 AOKVQA.xlsx

## 8. 训练方法

预训练：

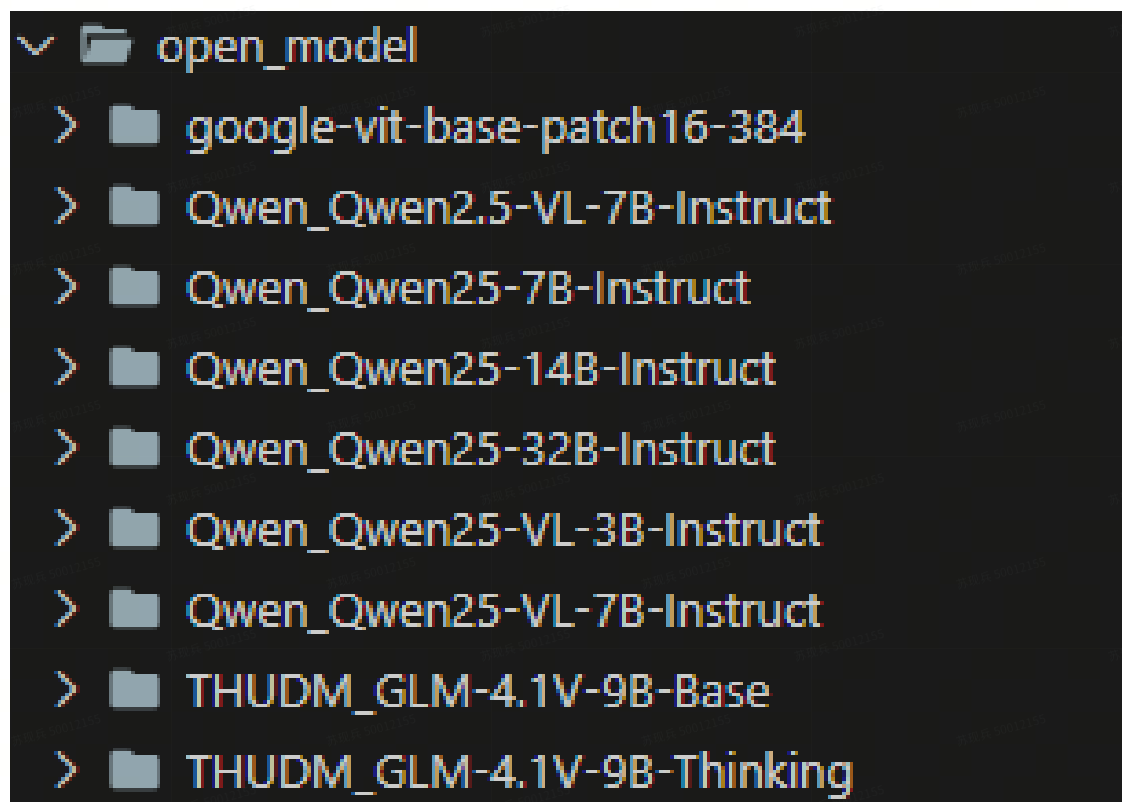
- 自回归预训练
- 监督微调
- Reinforcement Learning with Curriculum Sampling (RLCS)

微调方法：

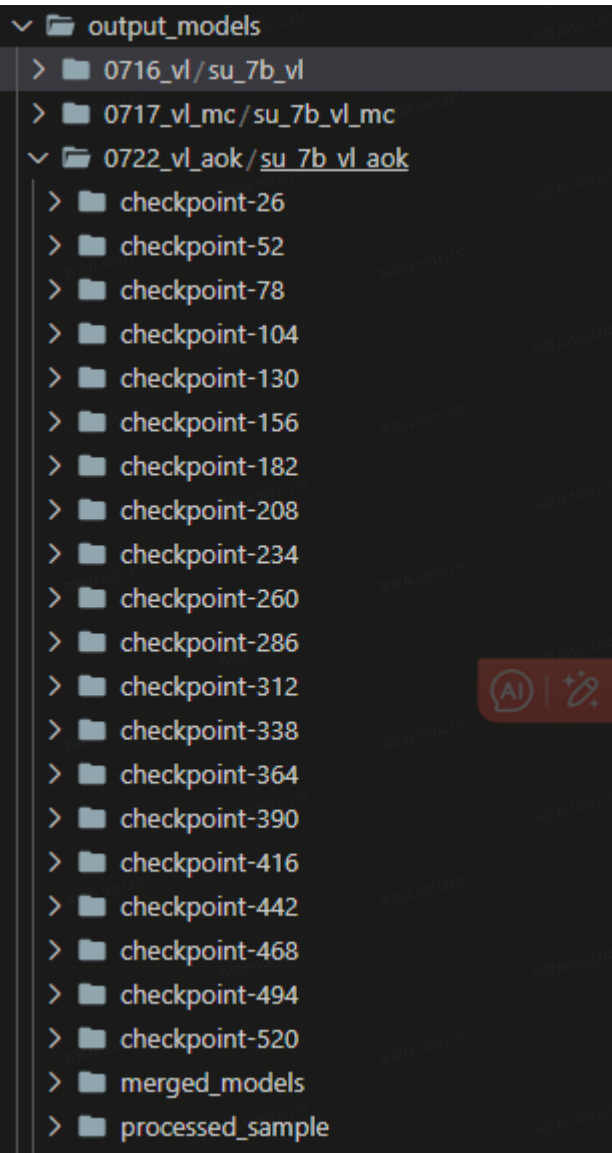
Qlora微调

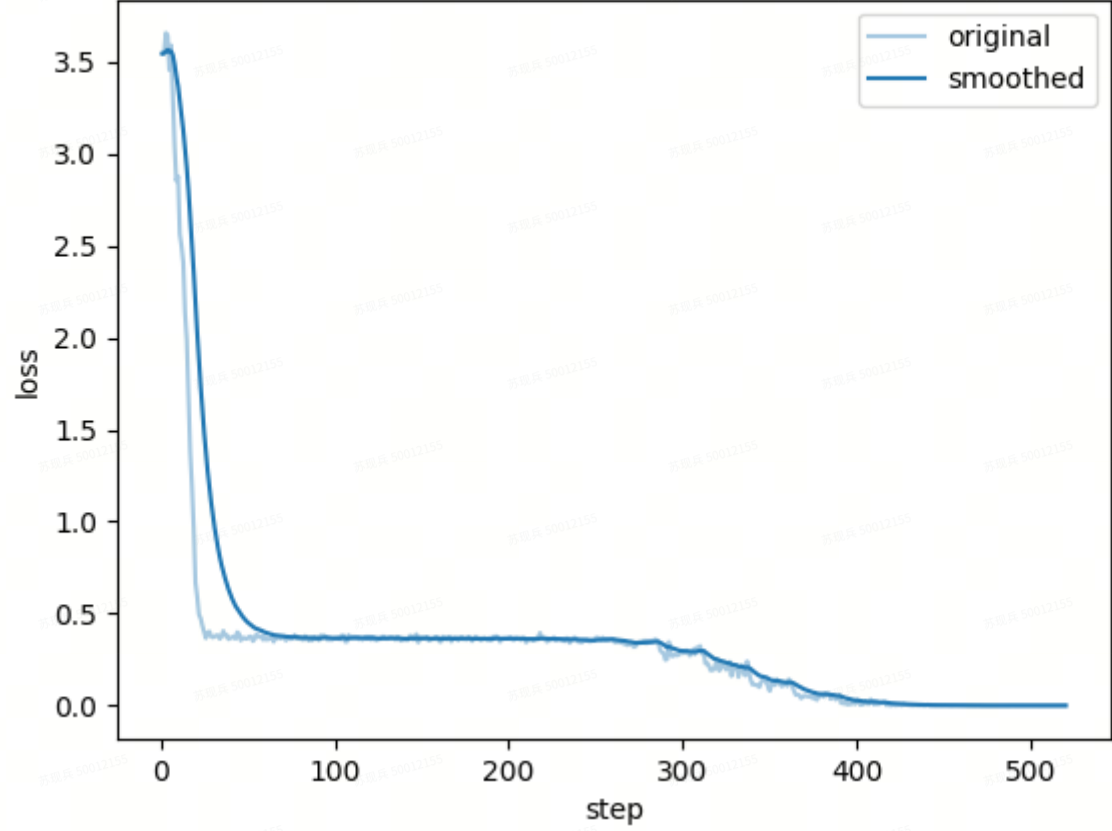
## 9. 对比实验

基座模型：



微调模型：





<think>用户现在需要描述图片内容，先看图片里的元素。首先是一个宣传海报，木质框架，背景有食物（可能是煎蛋类？），然后文字信息。

首先看整体结构：海报上有人字“好评 + 打卡”，下面是“送煎蛋或矿泉水瓶1瓶”。然后操作步骤分三步，第一步打开抖音、美团、大众点评；第二步搜索“将二两碗杂面”；第三步五星好评送煎蛋或矿泉水瓶。还有图片元素，比如煎蛋的图，矿泉水的图，底部有“送！送！送！”的黄色条带，右下角有小红书水印，背景有植物和地面。

现在组织语言，按顺序描述：海报是宣传类，木质支架，内容有文字、图片、步骤。文字部分有人标题“好评 + 打卡”，下面说明送煎蛋或矿泉水瓶。操作步骤分三步，每步用黑底黄字或类似排版。还有煎蛋和矿泉水的图片，底部黄色条带写“送！送！送！”，右下角有“小红书”水印，背景是户外，有植物和地面。

现在详细描述：

这是一张放置在户外的宣传海报，海报由木质支架支撑，背景可见绿色植物和地面。海报主体为宣传“好评打卡送礼品”的活动，包含文字、图片和操作步骤等信息。

海报上方有大号黑色手写字体“好评 + 打卡”，右侧有红色圆形印章样式文字。中间位置用黑色字体写着“送煎蛋或矿泉水瓶1瓶”，下方是“操作步骤”板块，以黑底黄字列出三步：

- 第一步：打开抖音、美团、大众点评
- 第二步：搜索“将二两碗杂面”
- 第三步：五星好评送煎蛋或矿泉水瓶1瓶

海报中还有煎蛋和矿泉水的图片，煎蛋图旁有“送”字，矿泉水瓶图也有“送”字，突出赠送福利。底部有一条黄色条带，用红色字体写着“送！送！送！”以强调优惠。海报右下角有“小红书”水印，表明图片来源。整体设计色彩丰富，以黄色、黑色、红色为主，信息明确，引导顾客通过点评和打卡获得赠品。

</think><answer>这是一张放置在户外的宣传海报，由木质支架支撑，背景可见绿色植物与地面。海报围绕“好评打卡送礼品”活动展开，包含文字、图片与操作步骤等信息：

- \*\*文字信息\*\*：海报上方有大号黑色手写字体“好评 + 打卡”，右侧有红色圆形印章样式文字；中间位置用黑色字体标注“送煎蛋或矿泉水瓶1瓶”；下方“操作步骤”板块以黑底黄字列出三步：第一步“打开抖音、美团、大众点评”，第二步“搜索‘将二两碗杂面’”，第三步“五星好评送煎蛋或矿泉水瓶1瓶”；底部黄色条带用红色字体写着“送！送！送！”。
- \*\*图片元素\*\*：海报中有煎蛋和矿泉水的实物图，煎蛋图旁、矿泉水瓶图旁均标注“送”字，直观展示赠送福利。
- \*\*其他细节\*\*：海报右下角有“小红书”水印，表明图片来源；整体设计以黄色、黑色、红色为主色调，信息清晰，引导顾客通过点评、打卡获得赠品。

</answer>

## 10. 本章小结

领域化微调效果和原始效果对比：



微调：prompt = "帮我分析途中飞机在目标检测过程中可能出现的影响因素并且分析影响因素的影响能力"

<think>用户现在需要分析图片中飞机在目标检测过程中可能出现的影响因素以及影响能力，然后根据给定的工具来编排任务。首先，得先理解目标检测的影响因素，比如图像分辨率、遮挡、背景复杂度、颜色对比度、目标大小、纹理特征这些方面。然后看图片里的情况：图片是飞机的俯视图，背景有涂鸦、建筑、道路，颜色对比可能有的地方有，但涂鸦和飞机颜色可能有些地方相似？不过先分析影响因素：

首先，影响因素1：背景复杂度。图片中飞机周围有涂鸦（蓝色和黄色图案）、建筑物、道路等，背景元素多且形状多样，可能干扰目标检测算法，因为算法需要区分飞机和背景。影响能力：较高，因为背景元素多，容易混淆。

影响因素2：目标大小与图像分辨率。飞机在图片中是较大的目标，但图像分辨率如果不算特别高，可能影响检测精度？不过这张图里飞机占比较大，但背景细节多，可能影响检测时的特征提取。

影响因素3：颜色对比度。飞机主体是白色，背景有蓝色和黄色涂鸦，颜色对比明显，这可能有助于检测，但如果涂鸦颜色和飞机部分区域颜色接近（比如某些涂鸦的白色？不过看图片里飞机是白色，涂鸦有黄色和蓝色，可能对比度还是有的，但需要看具体算法）。影响能力：中等，因为颜色对比有，但背景涂鸦的图案可能分散注意力。

原始 prompt = "帮我分析途中飞机在目标检测过程中可能出现的影响因素并且分析影响因素的影响能力"

影响因素4：遮挡情况。图片中飞机是否有被遮挡？看图片里飞机是完整的，没有被其他物体遮挡，所以遮挡因素影响小，影响能力低。

影响因素5：纹理特征。飞机的机身纹理和涂鸦的纹理不同，涂鸦有图案，飞机是光滑的金属质感（虽然图片是俯视图，可能纹理不明显），但纹理差异可能帮助检测，不过如果算法依赖纹理，这可能是个因素。影响能力：中等，因为纹理差异存在，但可能不是主要因素。

影响因素6：目标姿态。飞机是正上方俯视图，姿态统一，姿态因素影响小，影响能力低。

然后，现在需要把这些分析转化为任务，但用户需要的是用给定的工具编排DSL。不过看提供的工具，工具里没有直接的目标检测分析工具，可能需要用general\_qa来提问，或者看是否有其他工具？但候选工具里没有目标检测相关的工具，所以可能需要用general\_qa来询问目标检测的影响因素，或者分析图片中的情况。

不过用户的问题是“帮我分析途中飞机在目标检测过程中可能出现的影响因素并且分析影响因素的影响能力”，然后需要输出DSL格式的任务列表。但工具中没有直接的目标检测分析工具，所以可能需要用general\_qa来提问，比如询问目标检测的影响因素，或者分析图片中的情况。

现在看候选工具，general\_qa是图片问答，可以用来询问关于目标检测影响因素的问题。所以任务可能需要调用general\_qa，输入关于影响因素的问题。

那现在编排任务：

首先，确定需要分析的因素，然后构建问题。比如，询问“在目标检测过程中，图片中飞机可能受到哪些影响因素？各因素的影响能力如何？”然后用general\_qa工具，参数query是这个问题。

然后，看DSL格式，任务列表，每个任务ID从1开始。

现在检查工具列表：

工具名称有wlan、food、buy、check\_in、shop\_comments、general\_qa。这里需要用general\_qa，因为其他工具都是关于wifi、美食、团购、收藏、评论，和目标检测无关，所以只能用general\_qa来询问。



所以任务列表应该是：

```
[[ '1', 'general_qa(query="在目标检测过程中，图片中飞机可能受到哪些影响因素？各因素的影响能力如何？")' ]]
```

不过需要确认是否还有其他工具可能适用，但其他工具都不相关，所以只能用general\_qa。

现在再仔细分析影响因素：

1. 背景复杂度：周围有涂鸦、建筑、道路等复杂背景，干扰检测，影响能力高。
2. 颜色对比度：飞机白色与背景蓝、黄色涂鸦对比，有助于检测，但涂鸦图案分散注意力，影响能力中等。
3. 目标大小与分辨率：飞机在图中占比较大，但背景细节多可能影响特征提取，影响能力中等。
4. 遮挡：无遮挡，影响能力低。
5. 目标姿态：俯视图，姿态统一，影响能力低。
6. 纹理特征：飞机机身纹理与涂鸦纹理差异，有助于检测，影响能力中等。

然后，用general\_qa询问这些问题，所以DSL格式就是：

```
[[ '1', 'general_qa(query="在目标检测过程中，图片中飞机可能受到哪些影响因素？各因素的影响能力如何？")' ]]
```

不过可能需要更具体的提问，比如分点询问，但DSL中query是一个问题，所以综合起来提问。

总结，因为候选工具中没有目标检测相关的工具，只能用general\_qa来询问，所以任务列表包含调用general\_qa来分析影响因素和影响能力。</think>

<answer>以下是针对该问题的工具编排DSL结果，由于候选工具中无直接针对目标检测分析的工具，故使用 `general_qa` 工具来询问影响因素及影响能力：

```
[[ '1', 'general_qa(query="在目标检测过程中，图片中飞机可能受到哪些影响因素？各因素的影响能力如何？")' ]]
```

（注：若需更细分的任务拆分，可针对每个影响因素单独用 `general_qa` 提问，但根据当前工具限制，以上是综合性的任务编排。若未来有更针对性的目标检测分析工具，可进一步优化任务结构。）

</answer>



prompt = "帮我分析以下图片中的飞机哪个更难检测，并指出编号，并且分析受到哪些影响因素的影响导致难以检测"

<think>用户现在需要解决的问题是分析图片中哪个飞机更难检测，并指出编号和影响因素。首先，得先理解每个工具，但这里可能需要用general\_qa来提问，或者看候选工具里有没有适合的。不过看候选工具，wlan这些和图片检测无关，所以应该用general\_qa来提问，比如“分析图片中哪个编号的飞机更难检测，并指出编号和影响因素”。然后构建任务列表。

首先，确定需要使用的工具是general\_qa，因为需要分析图片内容，属于图片问答类。然后参数query就是用户的问题，比如“分析图片中的飞机，哪个更难检测，指出其编号及影响因素”。

然后构建DSL格式，任务列表，第一个任务用general\_qa，参数query是用户的问题表述。

现在需要确定DSL格式，比如：

```
[[1, 'general_qa(query="分析图片中的飞机，哪个更难检测，指出其编号及影响因素")']]
```

不过要检查工具参数，general\_qa的参数query是必选，所以需要把问题写进去。然后任务列表的格式符合要求，ID从1开始，每个任务第一项是ID，第二项是工具名称(参数=值)。

现在分析图片，可能需要看飞机的显示情况，比如是否有遮挡、颜色、大小、背景等。但因为是分析，需要用general\_qa来处理。所以最终的DSL应该是调用general\_qa工具，参数query包含分析问题。

现在整理步骤：

1. 确定需要使用的工具是general\_qa，因为需要分析图片内容，属于图片问答，有常识通识能力。
2. 构建问题，让工具分析哪个飞机更难检测，指出编号和影响因素。
3. 按照DSL格式输出，任务列表。

所以最终的DSL格式应该是：

```
[[1, 'general_qa(query="分析图片中的飞机，哪个更难检测，指出其编号及影响因素")']]
```

不过需要确认候选工具中是否有合适的，这里候选工具里general\_qa是图片问答，适合分析图片内容，所以选择这个工具。然后参数query包含用户的问题，所以构建问题语句。</think><answer>以下是符合要求的DSL格式任务列表，用于分析图片中哪个飞机更难检测并指出编号及影响因素：

```
[[{'1', 'general_qa(query="分析图片中的飞机，哪个更难检测，指出其编号及影响因素")'}]]
```

说明：

1. 选择的工具为 `general_qa`（图片问答工具），因其具备图片分析与常识通识能力，可处理对图片内容的分析类问题。
2. 工具参数 `query` 设为“分析图片中的飞机，哪个更难检测，指出其编号及影响因素”，符合工具“对图片相关信息提问并回答”的功能描述。
3. 任务列表格式遵循要求（任务ID从1开始、工具名称与参数结合），仅使用给定候选工具，且结合图片内容综合理解后确定调用 `general_qa` 工具分析问题。</answer>

## 第三章 智能问答系统知识库构建

### 11. 知识库架构设计

#### 11.1 使用langchain架构搭建RAG

#### 11.2 原始RAG与优化RAG对比

#### 11.3 结合向量数据库构建知识图谱

<https://www.53ai.com/news/MultimodalLargeModel/2025060418603.html>

### 12. 数据预处理

#### 12.1 数据爬取

#### 12.2 数据清洗

#### 12.3 数据整理

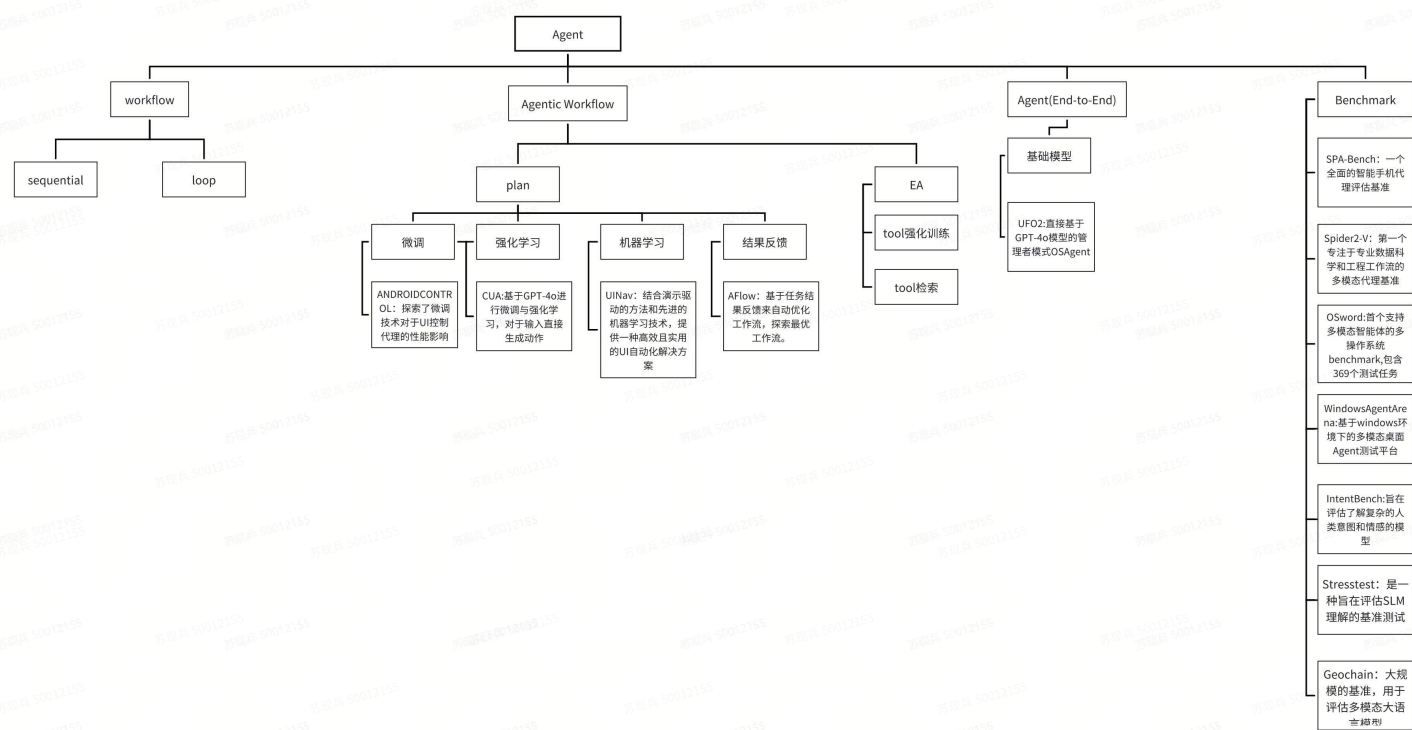
# 13. indexing、Retrieval、Generation优化

- 1. indexing: 5\_Levels\_Of\_Text\_Splitting、embedding模型
- 2. Retrieval: 使用混合检索

# 14. 对比试验

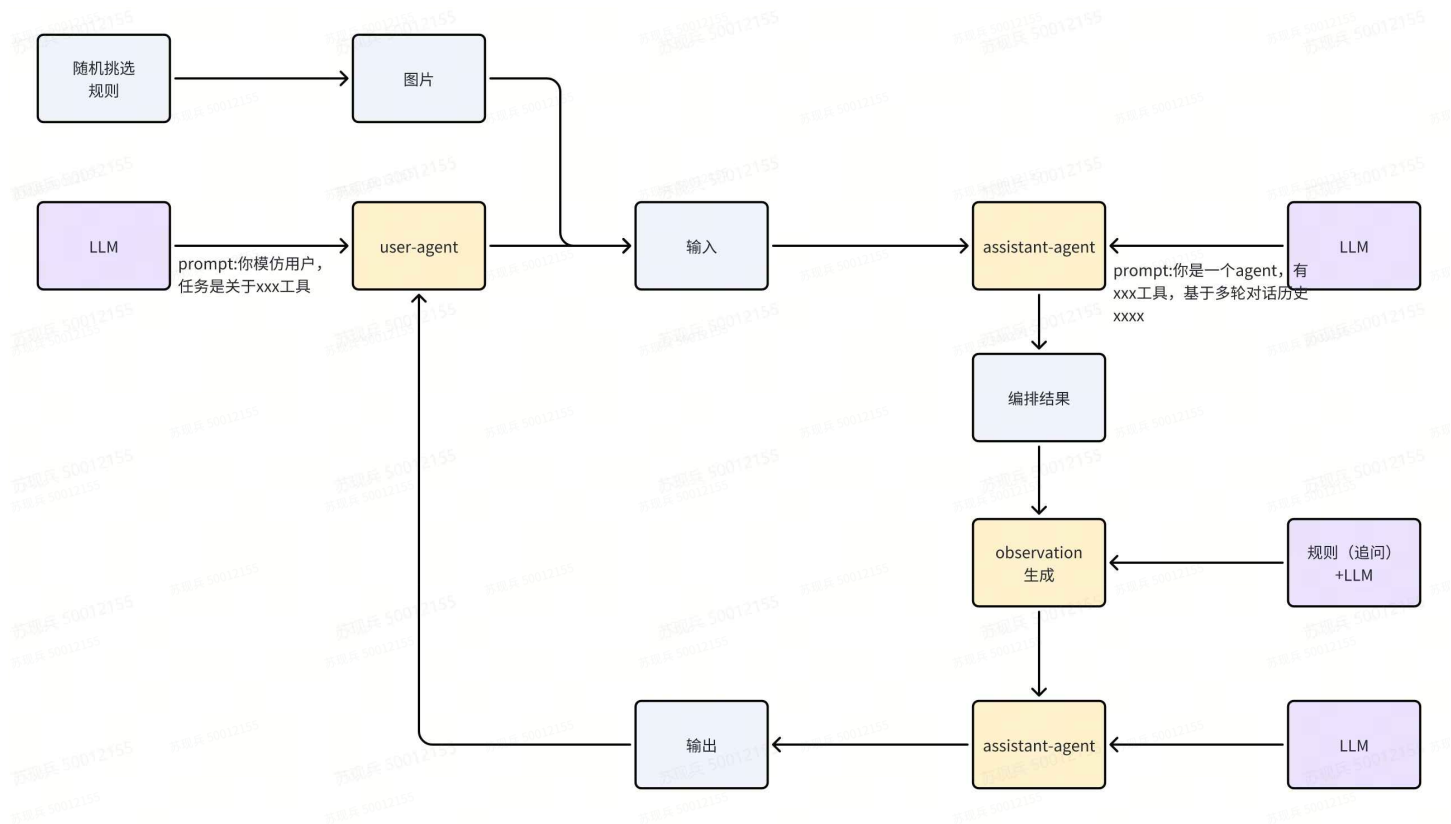
# 15. 本章小结

## 第四章 智能Agent系统设计



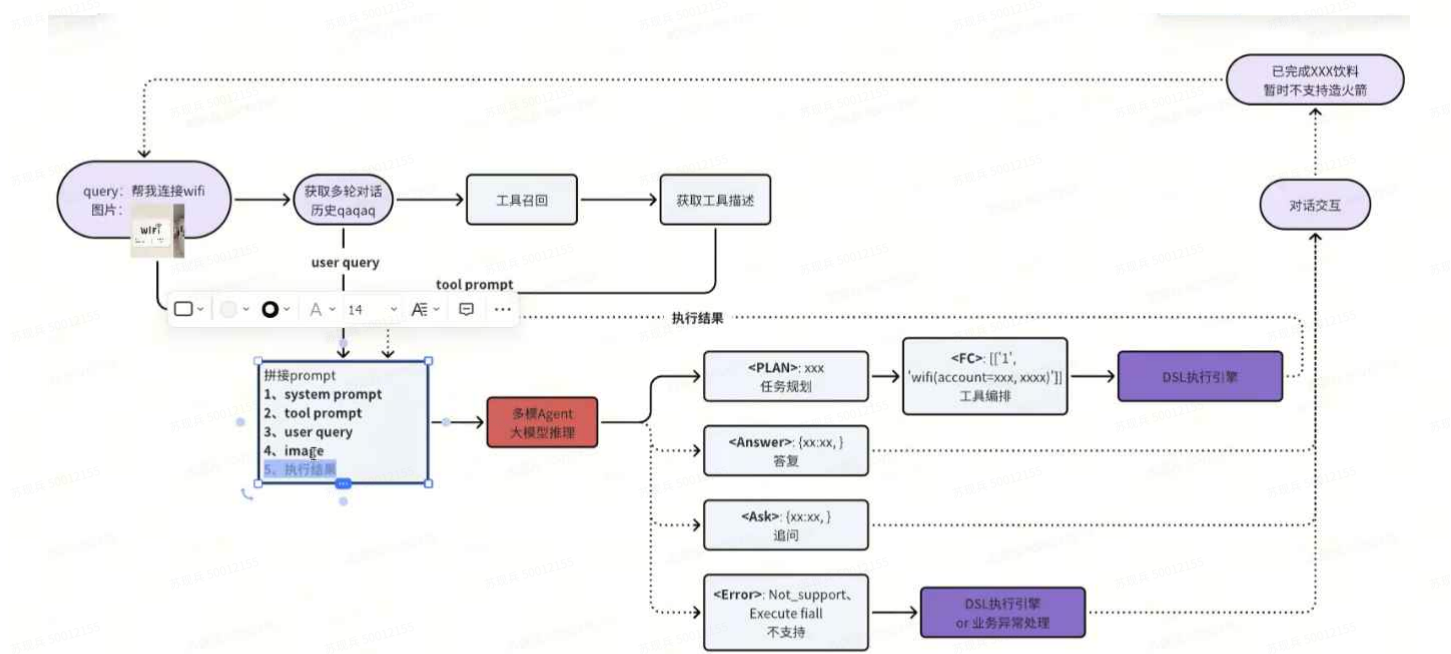
## 系统方案

架构设计：



- 执行结果格式: `[{"taskid": "xxx", "state": "error", "log": "网络异常"}]`
- 以下这几个是必选参数, 如果出现缺失, 则执行结果可以设置为 `{"taskid": "xxx", "state": "not finish", "log": "xx参数缺少"}`
- 每一次话术+图片至少覆盖如下类型: 继承、切域、图文相关、图文相关但是图片无信息、图文不相关

## Host-agent



## Work-agent

对比试验

## 第五章 系统部署

服务部署

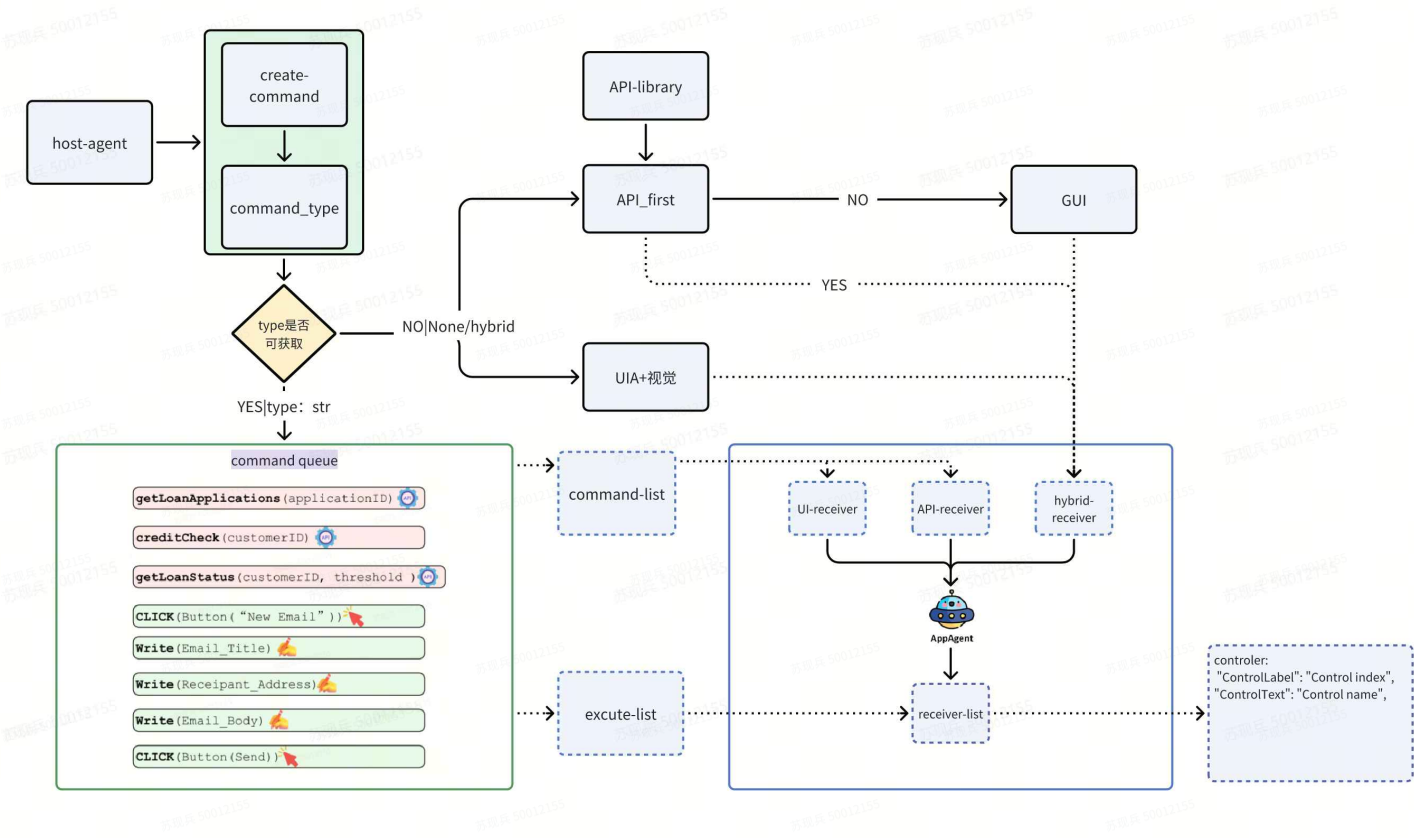
效果实验

本章小结

## 第六章 总结与展望

论文总结

工作展望



[https://llamafactory.readthedocs.io/zh-cn/latest/getting\\_started/installation.html](https://llamafactory.readthedocs.io/zh-cn/latest/getting_started/installation.html) llama-factory 地址