# OLY\_agent 广告文案审核系统实验报告

## 1. 项目概述

OLY\_agent 是一个基于多智能体架构的广告文案审核系统，旨在通过人工智能技术对广告文案进行全方位的合规性、质量和效果评估。该系统采用了先进的大语言模型技术，结合多智能体协作架构，能够自动化地检测广告文案中的敏感词、错别字、合规性问题等多个维度的内容，并提供优化建议。

## 2. 技术架构

### 2.1 整体架构

OLY\_agent 系统采用了分层架构设计，主要包括以下几个核心组件：

1. **前端界面层**：基于 Gradio 构建的交互式 Web 界面
2. **核心控制层**：基于 Google ADK (Agent Development Kit) 的智能体管理系统
3. **智能体层**：包含根智能体和多个专业子智能体
4. **模型层**：基于 LiteLLM 的大语言模型接入层
5. **工具层**：辅助功能模块和验证工具

系统架构图如下：

+------------------+  
| Gradio UI |  
+--------+---------+  
 |  
+--------v---------+  
| Google ADK |  
| Runner System |  
+--------+---------+  
 |  
+--------v---------+  
| Root Agent |  
+--------+---------+  
 |  
+--------v---------+  
| Parallel Agents |<----+  
+------------------+ |  
 | |  
+--------v---------+ |  
| Merge Agent | |  
+--------+---------+ |  
 | |  
+--------v---------+ |  
| Sub-agents |-----+  
+------------------+

### 2.2 核心技术栈

* **前端框架**：Gradio 3.40.0
* **智能体框架**：Google ADK 1.0.0
* **大语言模型**：通义千问 Qwen3-235B-A22B
* **模型接口层**：LiteLLM 1.70.4
* **API 调用**：OpenAI API 兼容接口
* **异步处理**：asyncio 3.4.3
* **环境配置**：python-dotenv 1.0.1
* **日志系统**：loguru 0.7.2

## 3. 技术特点

### 3.1 多智能体协作架构

OLY\_agent 系统最显著的技术特点是采用了多智能体协作架构。系统包含一个根智能体（Root Agent）和多个专业子智能体，每个子智能体负责特定领域的检测任务：

1. **敏感词检测智能体**：检测政治、宗教、暴力、色情等敏感词
2. **错别字检测智能体**：检测文案中的错别字和语法错误
3. **长度检测智能体**：检测文案长度是否符合要求
4. **逻辑检测智能体**：检测文案逻辑是否连贯
5. **语法检测智能体**：检测文案语法是否正确
6. **合规检测智能体**：检测文案是否符合广告法等法规
7. **用户视角检测智能体**：从用户角度评估文案效果
8. **情感基调检测智能体**：检测文案情感基调是否适合
9. **营销效果检测智能体**：评估文案营销效果
10. **SEO优化检测智能体**：检测文案SEO优化程度
11. **格式检测智能体**：检测文案格式是否规范

这种多智能体架构的优势在于：

* **专业化分工**：每个智能体专注于特定领域，提高检测精度
* **并行处理**：通过 ParallelAgent 实现多个检测任务的并行执行，提高效率
* **可扩展性**：可以方便地添加新的智能体以扩展系统功能
* **结果整合**：通过 Merge Agent 整合各子智能体的检测结果，提供综合评估

### 3.2 基于 Google ADK 的智能体管理

系统使用 Google ADK（Agent Development Kit）作为智能体管理框架，具有以下特点：

1. **灵活的智能体组织**：
   * 使用 SequentialAgent 实现顺序执行
   * 使用 ParallelAgent 实现并行执行
   * 使用 LlmAgent 实现基于大语言模型的智能体

* **会话管理**：
  + 使用 InMemorySessionService 管理用户会话
  + 支持多轮对话和上下文保持
* **工具调用**：
  + 通过 AgentTool 实现智能体对外部工具的调用
  + 支持工具上下文（ToolContext）传递
* **事件流处理**：
  + 支持 SSE（Server-Sent Events）流式响应
  + 实现实时反馈和增量输出

### 3.3 高性能大语言模型应用

系统使用通义千问 Qwen3-235B-A22B 大语言模型作为核心推理引擎，通过 LiteLLM 进行统一接口调用：

1. **模型特点**：
   * 超大规模参数（235B）：具有强大的理解和生成能力
   * 对中文的深度优化：更好地理解中文语境和表达
   * 思考链（Chain-of-Thought）支持：通过 enable\_thinking 参数启用

* **性能优化**：
  + 流式输出（Streaming）：实现实时响应
  + 参数控制：通过 temperature、max\_tokens 等参数控制输出质量
  + 超时处理：设置 request\_timeout 防止请求阻塞
* **负载均衡**：
  + 为不同智能体配置不同的 API Key（OPENAI\_API\_KEY\_1、OPENAI\_API\_KEY\_3 等）
  + 避免单一 API Key 的请求限制（Rate Limit）

### 3.4 异步并行处理机制

系统采用异步编程模式，提高并发处理能力：

1. **异步执行**：
   * 使用 asyncio 实现异步操作
   * 通过 async/await 语法实现非阻塞调用

* **并行智能体**：
  + 使用 ParallelAgent 实现多个子智能体的并行执行
  + 提高系统吞吐量和响应速度
* **事件驱动**：
  + 通过事件流（event stream）实现实时反馈
  + 支持增量式输出和交互式体验

### 3.5 用户友好的交互界面

系统基于 Gradio 框架构建了美观且易用的 Web 交互界面：

1. **界面设计**：
   * 自定义 CSS 样式，实现现代化 UI
   * 响应式布局，适应不同设备屏幕

* **交互功能**：
  + 聊天机器人界面（Chatbot）：支持多轮对话
  + 实时反馈：通过流式输出实现打字效果
  + Markdown 渲染：支持富文本格式输出
* **用户体验优化**：
  + 清晰的使用说明
  + 一键清空历史记录
  + 复制按钮，方便用户复制内容

## 4. 系统工作流程

OLY\_agent 系统的工作流程如下：

1. **用户输入**：用户通过 Gradio 界面输入需要审核的广告文案
2. **内容验证**：系统首先通过 test\_message 函数验证输入内容是否合规
3. **会话创建**：创建或恢复用户会话，维护对话上下文
4. **并行检测**：
   * Root Agent 调用 ParallelContentDetector
   * ParallelContentDetector 并行调用多个子智能体进行检测
   * 每个子智能体独立完成各自的检测任务

* **结果合并**：Merge Agent 收集并整合所有子智能体的检测结果
* **结果呈现**：系统将整合后的结果以 Markdown 表格形式呈现给用户
* **交互继续**：支持用户进一步询问或提交新的文案

## 5. 技术创新点

### 5.1 多模型协同工作

系统设计允许不同子智能体使用不同的模型或模型配置，实现资源的最优分配：

* 基础检测（如敏感词、错别字）可以使用较小的模型
* 复杂检测（如合规性、营销效果）可以使用较大的模型
* 通过 API Key 的区分，实现负载均衡

### 5.2 可扩展的智能体架构

系统采用模块化设计，便于扩展和定制：

* 可以方便地添加新的子智能体以支持新的检测维度
* 可以调整子智能体的组合方式（并行或顺序）
* 可以根据需求修改各智能体的提示词（Prompt）

### 5.3 实时流式输出

系统支持流式输出（Streaming），提供更好的用户体验：

* 用户无需等待完整结果，可以看到实时生成的内容
* 支持思考链（Chain-of-Thought）的可视化
* 降低用户等待感知，提高交互体验

## 6. 系统局限性与未来改进方向

### 6.1 当前局限性

1. **模型依赖**：系统高度依赖于大语言模型的能力，模型性能直接影响系统效果
2. **API 成本**：使用大型模型（如 235B 参数规模）的 API 调用成本较高
3. **并发限制**：受限于 API 调用频率限制，高并发场景下可能面临挑战
4. **领域适应性**：当前系统主要针对广告文案，对其他类型内容的适应性需要验证

### 6.2 未来改进方向

1. **模型优化**：
   * 探索模型量化和本地部署方案，降低依赖和成本
   * 尝试更小参数量但针对特定任务优化的模型

* **功能扩展**：
  + 增加多模态支持，实现对图像、视频广告的审核
  + 添加行业特化的审核规则和智能体
* **性能提升**：
  + 实现智能体结果缓存机制，避免重复计算
  + 优化并行处理策略，提高系统吞吐量
* **用户体验**：
  + 增加可视化分析功能，展示审核结果的详细分析
  + 提供文案优化建议的自动应用功能

## 7. 结论

OLY\_agent 广告文案审核系统是一个基于先进大语言模型和多智能体架构的创新应用。通过将复杂的审核任务分解为多个专业子任务，并利用并行处理和异步机制，系统能够高效地对广告文案进行全方位审核。

系统的核心技术优势在于： 1. 基于 Google ADK 的多智能体协作架构 2. 通义千问大语言模型的强大理解和生成能力 3. 异步并行的高效处理机制 4. 友好直观的用户交互界面

这些技术特点使 OLY\_agent 系统能够为广告文案创作者提供及时、专业的审核反馈，帮助他们创作出既符合法规要求又具有良好营销效果的广告文案。