# 购物反诈助手 Agent 作品策划

## 1. 团队介绍

团队名称： SIGAI39

团队成员：

王迪（队长）：负责 Agent 核心逻辑开发、大模型提示词优化

邢梓涵：负责数据收集、风险规则库构建、测试优化

刘丝语：负责 UI 交互设计、市场分析、文档撰写

团队背景：

我们来自华中科技大学人工智能专业，具备很强的自然语言处理（NLP）能力，以及机器学习（ML）和 AI Agent 开发经验。本次项目结合 vivo Agent 平台和蓝心大模型，致力于打造一个实用的购物反诈助手，帮助消费者识别虚假宣传，减少受骗风险。

## 2. 创意阐述

### 2.1 背景与痛点

在电商、社交平台和二手交易市场中，虚假广告、低价诱骗、脱离平台交易等欺诈手段层出不穷，消费者往往难以辨别真伪。常见问题包括：

价格欺诈：远低于市场价的商品（如“2999元的iPhone 15 Pro Max”）诱导冲动消费。

虚假宣传：夸大功效（如“一次美白3度”）、虚构权威背书（如“NASA技术认证”）。

支付风险：卖家引导微信私下转账，脱离平台担保，导致资金无法追回。

平台可信度：非正规渠道（如个人闲鱼卖家）假货率高，缺乏售后保障。

传统反诈手段依赖人工经验或简单关键词匹配，无法结合语义理解、动态数据验证和风险量化评估，导致消费者仍面临较高受骗风险。

### 2.2 创新点与解决方案

本项目基于 vivo Agent 平台 + 蓝心大模型（70B），打造一个拟人化、多维度、动态验证的购物反诈助手，核心创新如下：

#### 2.2.1拟人化交互体验

采用客服聊天风格，结合温和表情，提升用户信任感。

每次回复自动生成后续问题引导（如“需要查卖家历史记录吗？”），增强交互连贯性。

#### 2.2.2多维度欺诈分析

价格合理性：调用工作流“判断商品价格是否合理”，对比市场价与历史数据。

平台可信度：通过“判断平台是否可信”工作流，分析店铺评分、授权资质等。

宣传真实性：识别绝对化用语（“第一”“彻底根治”）、医疗效果宣称等违规话术。

支付安全：检测是否引导脱离平台交易（如“加微信转账”）。

动态验证与风险量化

强制联网搜索：对存疑信息（如专利号、品牌备案）实时验证。

风险星级评分（0-10星）：直观展示欺诈概率，例如：

“【虚假诈骗程度：⭐⭐⭐⭐⭐⭐⭐ 7/10星】”

关键风险点标注：用“重点 ”“最大风险 ”高亮核心问题。

#### 2.2.3明确结论与建议

即使信息模糊，也强制输出倾向性结论（如“不可信 ”），避免模棱两可。

提供具体行动建议（如“同城验货”“走平台担保支付”）。

### 2.3 技术差异化

|  |  |
| --- | --- |
| 竞品/传统方案 | 本项目优势 |
| 人工经验判断 | AI 自动化分析，覆盖全网动态数据 |
| 关键词匹配 | 语义理解+逻辑推理，识别话术套路 |
| 单一维度检测 | 价格+平台+宣传+支付多维度综合评估 |
| 无风险量化 | 星级评分+关键风险点标注，结果更直观 |

### 2.4 应用场景

消费者：快速验证商品/卖家可信度，避免冲动消费。

电商平台：集成至客服系统，辅助审核违规广告。

监管部门：批量监测市场欺诈行为，定位高风险店铺。

### 2.5 社会价值

通过降低购物欺诈风险，提升消费者信任度，推动电商生态健康发展。未来可扩展至多模态分析（图片/视频广告）和跨境购物反诈场景。

## 3. 技术方案

### 3.1技术架构

#### 3.1.1 核心组件

|  |  |  |
| --- | --- | --- |
| 组件名称 | 功能描述 | 调用方式 |
| 蓝心大模型-70B | 主分析引擎，执行文本理解和推理 | 每次对话强制调用 |
| 联网搜索插件 | 实时数据验证 | 每次对话强制调用 |
| 诈骗知识库 | 内置200+欺诈话术模板 | 自动匹配调用 |

#### 3.1.2 工作流配置

##### 3.1.2.1 价格合理性分析工作流

输入：商品名称+价格（字符串）

处理流程：提取商品价格（大语言模型3）

联网搜索市场价（联网搜索组件）

生成价格对比报告（大语言模型2）

输出：价格偏离度分析（字符串）

##### 3.1.2.2 平台可信度评估工作流

输入：平台名称（字符串）

处理流程：

联网搜索平台投诉记录（联网搜索组件）

综合评估风险等级（大语言模型2）

输出：平台风险评级（字符串）

### 3.2数据处理规范

#### 3.2.1 数据流向

用户输入 → 工作流并行处理 → 大模型综合判断 → 结果格式化输出

#### 3.2.2 数据保留策略

用户原始输入：分析完成后立即删除

分析结果日志：匿名存储30天

### 3.3性能指标

|  |  |  |
| --- | --- | --- |
| 指标项 | 目标值 | 测量方式 |
| 响应时间 | ≤3秒 | 端到端测试 |
| 工作流调用成功率 | 100% | 服务监控 |
| 风险识别准确率 | ≥85% | 千例测试集验证 |

### 3.4交互实现

输出要素（每轮对话必须包含）：

风险星级（0-10⭐）

关键风险点（⚠️标注）

具体建议（分点列出）

后续问题建议（1-2条）

## 市场分析：

### 4.1 目标用户群体

#### 4.1.1 普通消费者

核心需求：快速验证商品/卖家可信度，避免受骗

使用场景：

网购前咨询（如“这个价格靠谱吗？”）

交易中风险预警（如“客服让微信转账”）

大额商品购买前二次确认（如数码产品、奢侈品）

#### 4.1.2 电商平台

核心需求：辅助审核违规广告，降低平台投诉率

使用场景：

集成至客服系统，自动识别高风险商品描述

监控第三方卖家行为，减少假货投诉

#### 4.1.3 监管部门

核心需求：监测市场欺诈行为，定位高风险店铺

使用场景：

批量分析消费者投诉数据

识别新兴诈骗套路，发布预警

### 4.2 市场规模与增长潜力

#### 4.2.1 消费者端市场

中国网购用户规模：8.4亿人（2023年数据）

潜在用户比例：约30%曾遭遇购物欺诈（年需求规模超2.5亿人）

增长驱动：直播电商、二手交易平台等新兴购物方式带来的新型诈骗风险

#### 4.2.2 企业端市场

头部电商平台年审核成本：数亿元

AI工具替代潜力：预计可降低20%+人工审核成本

目标客户：中小电商平台（预算有限但反诈需求强烈）

#### 4.2.3 监管端市场

2023年虚假广告案件：4.2万件

数字化监测工具渗透率：不足10%

政策驱动：《网络交易监督管理办法》要求加强平台监管

### 4.3 竞争分析

#### 4.3.1 现有竞品对比

|  |  |  |  |
| --- | --- | --- | --- |
| 竞品名称 | 优势 | 劣势 | 本项目优势 |
| 微信谣言过滤器 | 用户基数大 | 仅限谣言检测 | 垂直领域深度分析 |
| 浏览器风险拦截 | 自动防护 | 仅限网址判断 | 多维度动态验证 |
| 平台审核系统 | 数据直接 | 规则僵化 | 拟人化交互体验 |

#### 4.3.2 竞争优势

技术优势：

蓝心大模型（70B）的深度语义理解

动态工作流（价格+平台+支付多维度分析）

数据优势：

10万+欺诈案例知识库

实时联网验证能力

体验优势：

拟人化客服交互设计

直观的风险星级展示

### 4.4. 商业模式

#### 4.4.1 短期计划（1年内）

ToC服务：免费基础版

ToB服务：

API调用收费（0.1元/次）

基础年费套餐（5万元/年）

#### 4.4.2 长期规划（3年）

增值服务：

高级会员（9.9元/月）：多模态分析

企业定制版：深度数据报告

数据服务：

行业反诈白皮书（年费制）

监管预警系统（政府采购）

### 4.5. 风险与对策

|  |  |  |
| --- | --- | --- |
| 风险类型 | 具体表现 | 应对策略 |
| 技术风险 | 大模型误判 | 人工反馈闭环优化 |
| 市场风险 | 平台数据限制 | 发展替代数据源 |
| 政策风险 | 隐私监管 | 严格数据匿名化 |

## 5. 产品预期功能及形态

|  |  |  |  |
| --- | --- | --- | --- |
| 功能模块 | 功能描述 | 技术参数 | 交互示例 |
| 智能话术分析 | 识别8大类欺诈话术 | 支持200+欺诈模板库 准确率≥89% | "祖传秘方"→医疗违规 "最后一天"→虚假促销 |
| 动态价格验证 | 实时比价系统 | 接入10+电商平台数据 更新频率15min/次 | "iPhone15仅2999"→风险9星 |
| 全链路风险 | 平台+商品+支付三维检测 | 调用2个工作流： 1. 价格合理性分析 2. 平台可信度评估 | 拼多多第三方店→中等风险 |
| 可视化风险报告 | 星级评分+关键点标注 | 0-10星分级制度 支持3级颜色预警 | ⭐⭐⭐⭐⭐⭐⭐ 7/10星 |