2025/5/11

购物反诈小助手

Agent 作品策划

团队：SIGAI39

目录

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、产品介绍网页构建、工作流设计

邢梓涵：负责数据集构建、测试优化、一部分大模型提示词优化

刘丝语：负责市场分析、文档撰写

团队背景：

我们来自华中科技大学人工智能专业，具备一定的自然语言处理（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。

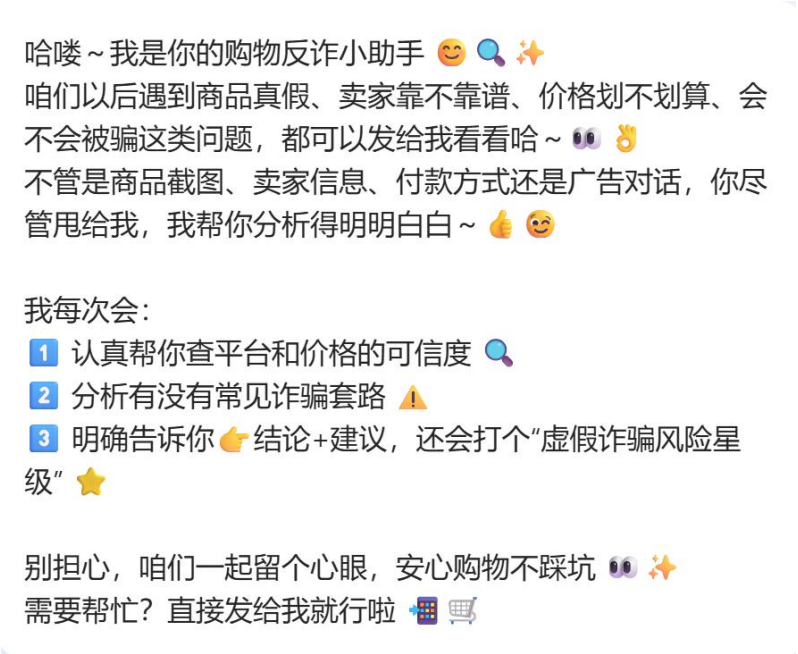
传统反诈手段依赖人工经验或简单关键词匹配，无法结合语义理解、动态数据验证和风险量化评估，导致消费者仍面临较高受骗风险。

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，例如，开场白如下：



每次回复带有 emoji 表情，采用温和的语气，能给用户身临其境的使用体验。并且在回复的末尾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。回复示例如下：

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。如下图中:周杰伦的演唱会门票通常非常抢手，价格远低于市场价往往诱骗消费者。

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。如下图中:闲鱼上个人卖家多，交易记录少，风险高。

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。如当出现“加微信转账”时，智能体会提醒您交易的风险性。

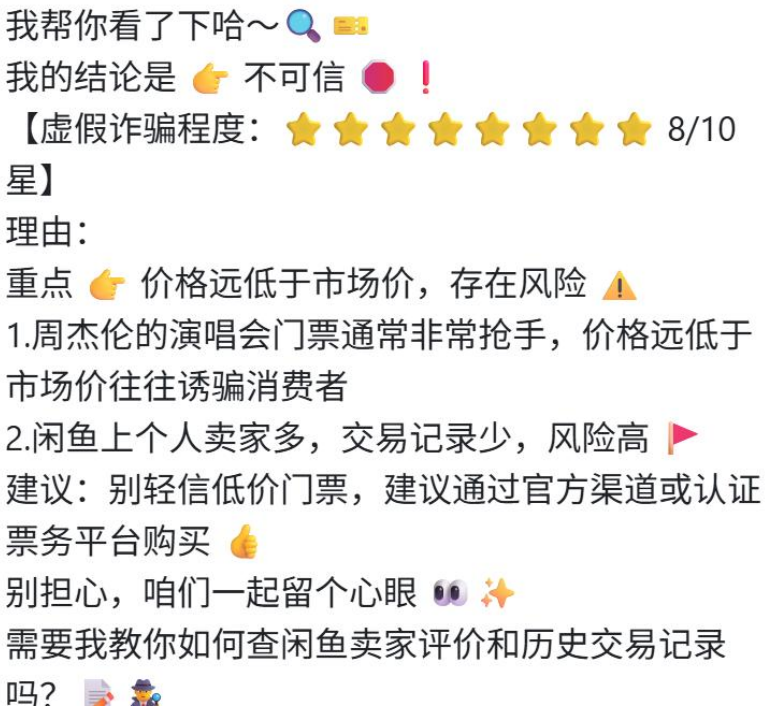
动态验证与风险量化

联网搜索：对存疑信息（如专利号、品牌备案、品牌功效.....）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

智能体能够基于不完备的信息进行推理，得出较为确定的结论，是帮助用户在复杂、不透明环境下做出果断决策的重要能力。在实际交易或信息判断场景中，用户往往无法获得全面详尽的数据，但通过合理分析现有线索，依然可以做出清晰的结论，比如：“较为可信”“不可信”“高风险，建议放弃”等，从而有效规避风险和损失。

为了帮助用户进一步落实安全操作，智能体能提供具体行动建议。

如同城验货、走平台担保支付、核实卖家身份与历史交易记录、保留聊天记录和付款凭证、要求卖家提供实时商品视频或特征验证、咨询有经验的朋友或专业人士、警惕明显低于市场价的商品主动放弃交易、使用货到付款或第三方中介服务、拒绝脱离平台私下交易等。

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：借助该工具，消费者能够快速验证商品信息与卖家的可信度，有效识别虚假宣传和潜在风险，降低受骗几率。同时，辅助用户理性判断购物需求，减少冲动消费，提升整体购物体验与满意度。

电商平台：可将该系统集成至平台客服与审核流程，实时辅助客服识别涉嫌违规的广告与可疑商品，提升处理效率与准确率。平台亦可借此加强对商家行为的监控，维护公平健康的交易环境，增强用户信任。

监管部门：支持监管机构实现大规模市场监测，批量识别与跟踪可疑商户和潜在欺诈行为，显著提升风险预警与执法效率。通过精准

定位高风险店铺与商品，助力精准执法与行业治理，保障消费者权益与市场秩序。

2.5 社会价值

本智能体通过有效降低在线购物中的欺诈风险，增强消费者对电商平台与在线交易环境的信任度，从而促进安全、透明、公平的数字消费生态建设。一方面，消费者在获得更高安全保障的同时，能够更放心地参与网络购物，激发市场活力；另一方面，平台与商家也因诚信经营与风险可控，获得更长远的用户黏性与市场声誉，助力整个电商产业链的良性循环与可持续发展。

未来，本智能体具备广阔的拓展潜力。可进一步集成多模态数据分析能力，涵盖商品图片、视频广告等丰富媒介内容，实现对视觉与文本信息的联合识别与风险预警。

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	每次对话调用

诈骗知识库	内置 8000+欺诈话术模板	自动匹配调用
识别诈骗工作流	识别价格、平台等是否不可信	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 →知识库检索→ 联网搜索→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	≤10 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（重点标注）

具体建议（分点列出）

后续问题建议（1-2 条）

4. 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：9.74 亿人（2024 年 12 月数据）

潜在用户比例：网民遭遇网络诈骗中遭遇网络购物诈骗的比例为 33.9%（《第 51 次中国互联网络发展状况统计报告》）

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

AI 工具替代潜力：预计可降低大量人工审核成本

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：4.76 万件

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 与人工审核对比

对比维度	传统人工审核	本项目 AI 智能体

效率	依赖人工逐条审核，处理速度慢，高峰期易积压	AI 自动化分析，响应速度快，支持高并发处理。
成本	人力成本高（需专职团队），培训周期长，且需持续更新经验库。	前期开发成本固定，后期边际成本趋近于零，支持自动化知识库更新。
一致性	受审核员主观经验影响，标准不统一（如同一商品不同人可能给出相反结论）。	基于统一算法模型，输出标准化风险评分（0-10 星）与建议，结果可复现。
动态适应能力	依赖人工发现新骗术后再更新规则，滞后性明显。	实时联网验证（如专利号、市场价）、自动学习新话术（欺诈知识库动态扩展）。

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：

8000+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

本智能体计划开源

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	拼多多第三方店→中等风险

		2. 平台可信度评估	
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