

GUI Agent 在公安业务流程自 动化中的应用

Application of GUI Agent in Security Process Automation

汇报人：李振邦

汇报时间：2026.01.07

目录

CONTENTS



公安业务 系统

Security Management System

01



流程自动 化

Process Automation

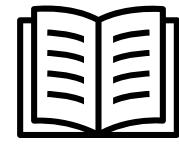
02



浏览器自动 化智能体

Web Agent

03



后续工作 计划

Next Steps

04

001 | 公安业务系统中缺少智能化的辅助工具



业务系统复杂度高、流程衔接自动化程度低、重度依赖人工经验

传统交互模式下公安业务应用系统的局限



操作步骤多、路径长
交互繁琐



专业性强
使用说明书800+页



各平台相互孤立
需人工衔接

新型交互模式



自动执行任务
业务系统流程自动化



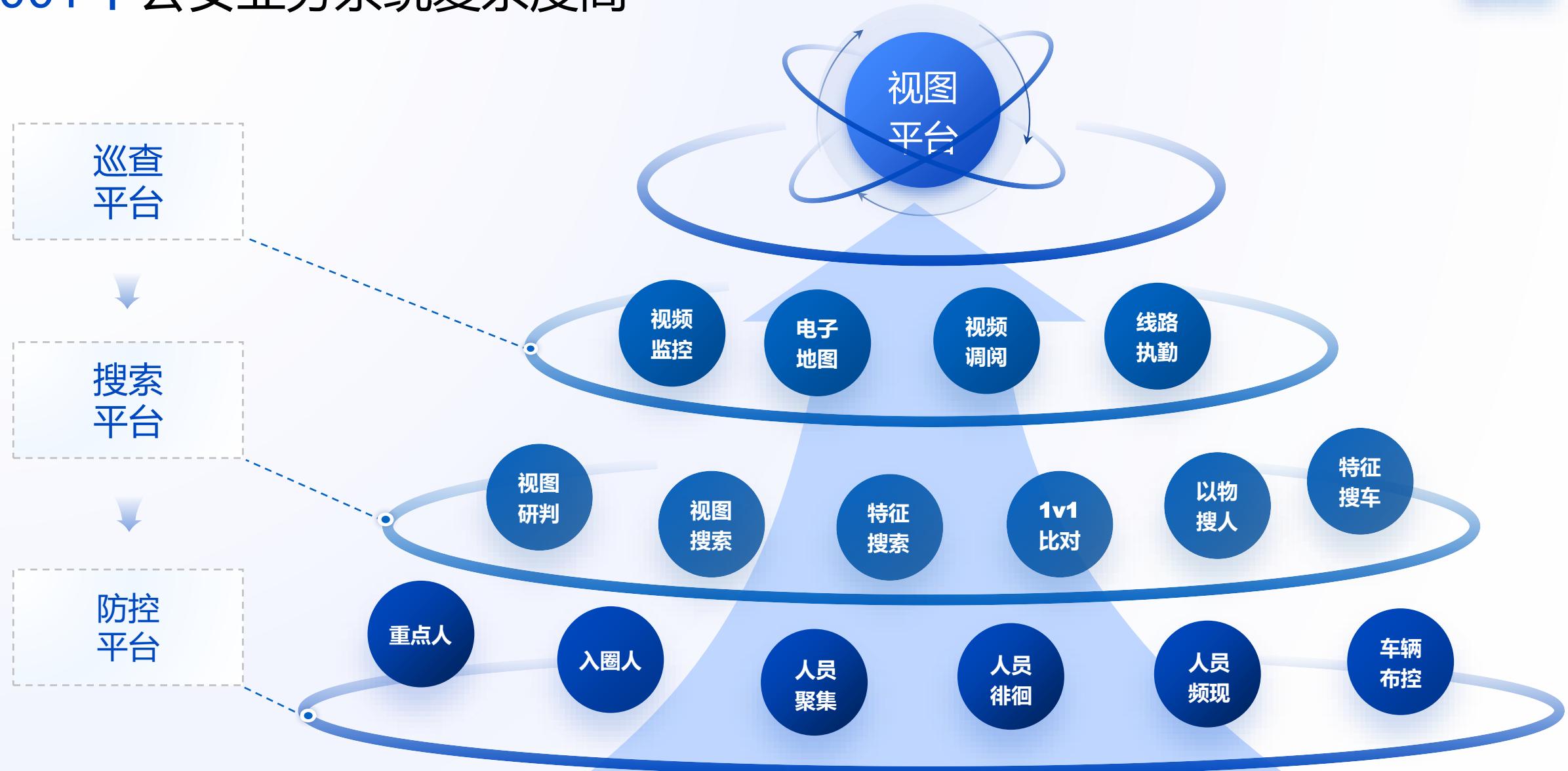
跨系统联动
数据跨系统流转

The screenshot shows the Hikvision Video Image Comprehensive Application Platform (V2.3). At the top, there's a navigation bar with links like '首页', '视频搜索', '智能搜索', '通知提醒', '多维搜索', '智能统计', and '视频智能分析'. Below the navigation is a search bar with placeholder text '可输入车牌号、车身颜色或人员特征。如：输入“白色衣服女”' and a '搜索' button. A large red box highlights a feature titled '我将为你找到Ta' which includes a car icon and a '上传图片' button. Another red box highlights a sidebar titled '常用应用' containing icons for '重卡部位人员防控', '智能名册库', 'AI以文搜图', '智面识人', '视频闯入', '设备事件报警', '车辆单元数据应用', '人证比对', '电子围栏', '智能工具', '智脑决策', '面上寻踪', and a '+' button.

The image shows the table of contents for the '海康视频图像综合应用平台产品v2.3使用说明书 (7).pdf'. The contents are as follows:

5.15 搜索.....	792
5.16 地图.....	793
5.17 人脸布控.....	795
5.18 车辆布控.....	796
5.19 消息.....	797
5.20 人员档案.....	798
5.21 车辆档案.....	799
5.22 个人中心.....	800

001 | 公安业务系统复杂度高



002 | 流程自动化软件



流程自动化的典型产品形态为 RPA (Robotic Process Automation)，通过模拟人类在计算机界面上的操作逻辑，在不改变企业现有 IT 架构的前提下，实现跨系统、跨平台的业务衔接，能够极大地提升数据流转的效率与精准度。





RPA软件的核心技术之一是浏览器自动化技术，用于模拟人类用户在Web环境下的交互行为，实现复杂业务流程的自动化处理。

○ 问题回顾01 | Review of annual issues one

浏览器自动化技术的作用？

通过对DOM（文档对象模型）的解析，识别网页上的输入框/按钮/菜单等元素，执行点击/输入等动作，无需业务平台内部API即可完成多系统间数据搬运和流程流转

○ 问题回顾02 | Review of annual issues two

与公安业务平台的关系？

目前绝大多数主流的公安业务平台（如视频监控系统 VMS、综合安防集成平台等）均为Web应用。面临挑战：系统多、接口闭塞、人工手动切换操作效率低。

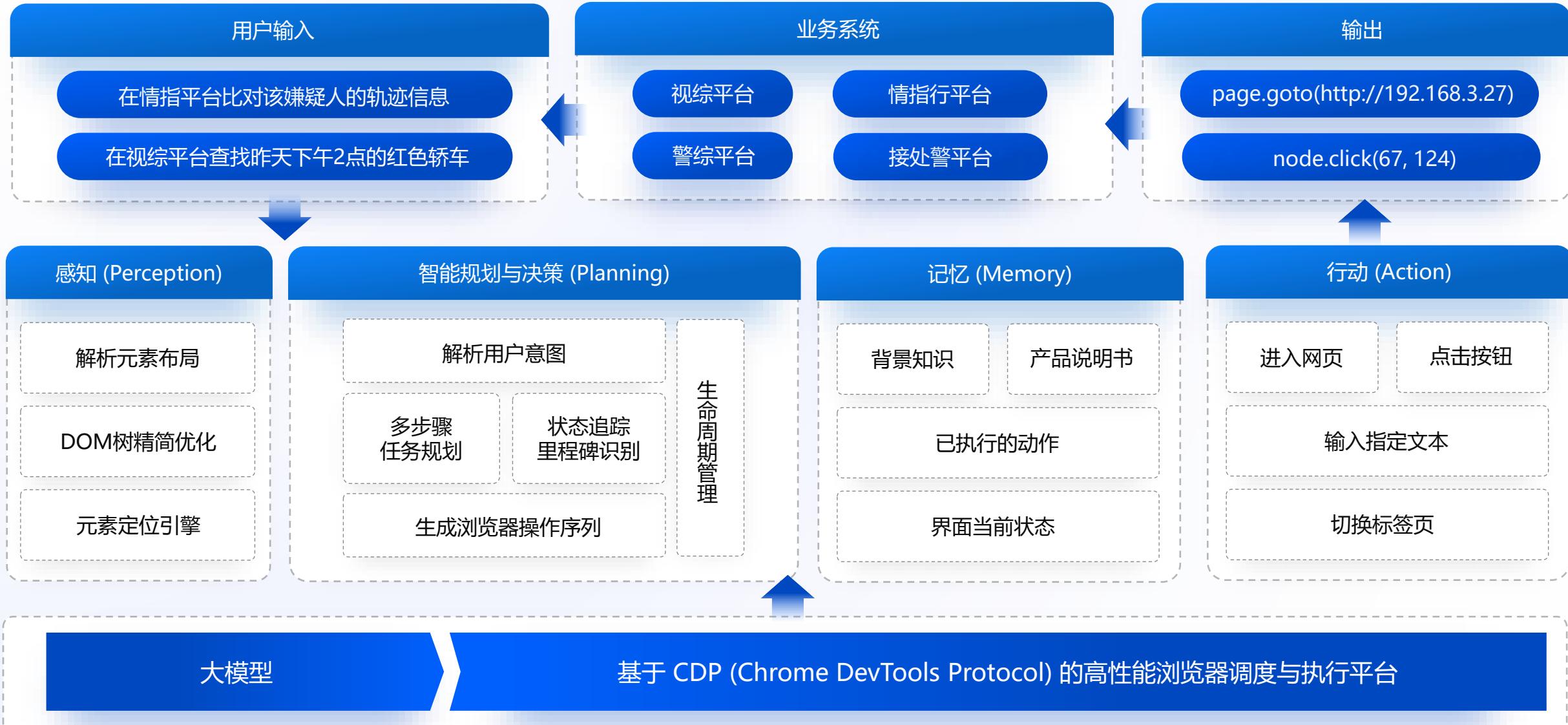
003 | 浏览器自动化技术发展历史



浏览器自动化技术经历了从初期依赖元素定位符的静态脚本编写，到中期基于底层通信协议的深度控制，再到当前利用大语言模型实现自然语言指令解析与动态环境自主适应的演进历程。



003 | Web Agent: AI驱动的浏览器自动化框架



003 | GUI Agent



运行平台	GUI Agent	Web Agent	Mobile Use Agent	Computer Use Agent
概念介绍	高层技术概念，按平台可分为后续三类具体Agent 泛指能够像人类一样通过视觉感知图形界面，并执行点击、输入等交互行为的 AI 代理。	浏览器环境(Chrome, Edge 等 Web 平台) 专注于网页自动化，能够理解网页 DOM 结构，跨页面执行复杂的工作流。	移动端操作系统(Android, iOS) 模拟人类在手机屏幕上的操作，通过截图理解 UI 元素，实现手机App 自动化任务。	桌面端操作系统(Windows, macOS, Linux) 具备全局操作系统权限，能跨窗口、跨软件进行通用操作，直接“接管”计算机鼠标键盘。
代表产品	UI-TARS AgentS3	Microsoft/playwright-mcp browser-use.com	OpenGLM 豆包手机	Claude Computer Use Microsoft UFO ²

003 | 各 GUI Agent 可用性



Agent 名称	操作系统	任务类型	测试结果	耗时	使用模型	备注 / 表现
CUA	Windows 11	基于视觉 (OS)	成功	150s	glm-4.5v	表现稳定。
CUA	Ubuntu 24.04	基于视觉 (OS)	失败	250s	glm-4.5v	文本输入错误（武林林传），且错误误判为成功。
UI-TARS-desktop	Windows 11	基于视觉 (OS)	成功/失败	88s	doubao-1-5-thinking	结果1成功；结果2因未输入回车键且误判网络卡顿而失败。
UI-TARS-desktop	Windows 11	基于 DOM (Web)	成功	29s-48s	doubao-1-5-thinking	速度较快，适用于浏览器任务。
UI-TARS-desktop	Ubuntu 24.04	基于视觉 (OS)	失败	-	doubao-1-5-thinking	未执行有效操作却误判为成功。
UI-TARS-desktop	Ubuntu 24.04	基于 DOM (Web)	失败	-	doubao-1-5-thinking	未执行有效操作却误判为成功。
Microsoft UFO2	Windows 11	基于视觉 (OS)	成功	240s	qwen-vl-max	耗时较长，但任务完成成功。
browser-use	Ubuntu 24.04	基于 DOM (Web)	成功	109s	Qwen3-VL-235B	Ubuntu 平台上唯一推荐的 Web 任务 Agent。
Agent S3	Windows 11	基于视觉 (OS)	失败	9min+	gpt-5-2025-08-07	陷入死循环（反复聚焦地址栏），无法确认操作成功。

003 | 各 GUI Agent 可用性



基于视觉感知的 Computer Use Agent 尚处非成熟阶段，而基于DOM驱动的 Browser Use Agent 具有更好的执行成功率与响应速度优势，是目前解决基于Web的公安业务系统平台自动化难题现实最优技术路径。

测评
结论

	Windows 11	Ubuntu 24.04
Web 任务	browser-use UI-TARS	browser-use
OS 任务	CUA UFO2	✗ 无可用推荐 (全失败)

Web 任务 仅浏览器/DOM 操作

OS 任务 涉及桌面/全局视觉)



打造以“智能体”为核心的公安业务流程自动化系统

从“人工经验依赖”向“智能自主编排”的演进

交互意图理解难

难以准确捕捉用户真实意图，存在大量繁琐重复的人工操作。

流程衔接断层多

各系统孤立，缺乏自动化衔接，依赖人工跨系统、跨模块流转。

平台操控门槛高

系统逻辑复杂，用户需具备丰富的平台操作经验才能解析系统反馈并手动规划后续步骤。

人工驱动
依赖经验
操作繁琐

AI驱动
自主编排
高效闭环

智能意图识别

工作人员通过自然语言与系统交互，大模型深度理解操作意图，实现人机交互的极简化与智能化。

自动化任务编排

自动调度任务流，实现业务系统调用与任务执行的无缝闭环编排。

专家经验知识库

将资深用户的平台使用经验转化为智能体的内置指令

004 | 已完成的阶段性成果



产品选型-Nanobrowser

成熟开源

- GitHub 开源的 Web Agent 项目，收获 10,000+ Stars
- 功能完整：通过浏览器侧边栏聊天界面，通过自然语言下达任务指令，完成网页任务自动化

部署方便

- 浏览器插件形式安装
- 不挑硬件，兼容老旧PC，性能占用低。不挑系统，兼容国产麒麟系统。不跳环境，无需安装包和系统权限。

模拟装机

在模拟攒机网站，自动选择CPU、硬盘等硬件，配置电脑装机方案



自动检索

自动进入百度百科
搜索执行词条

公安系统

通过自然语言描述
操作步骤，能够自动进入系统、加载
监控列表，查询点位，播放监控画面

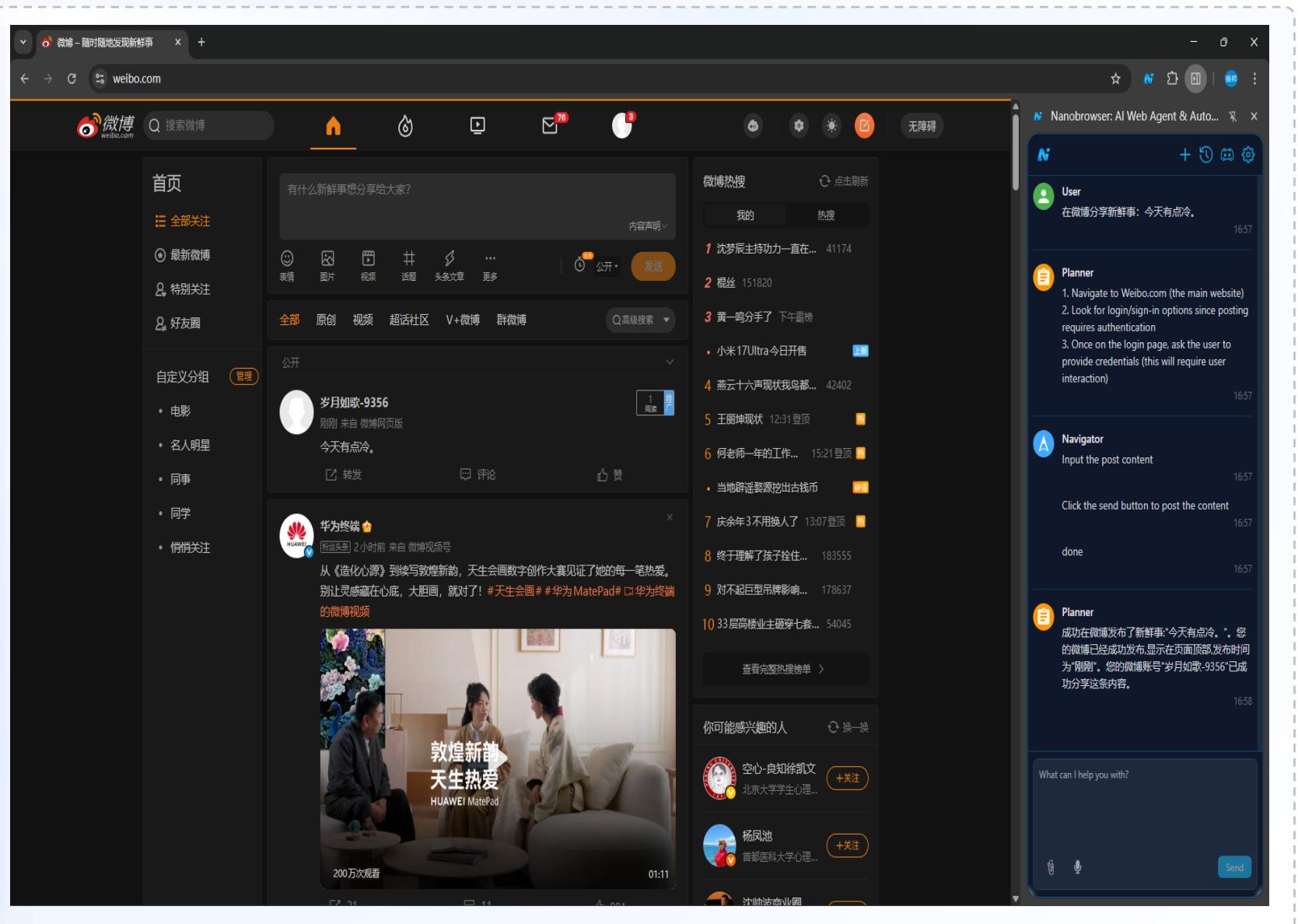
003 | 产品形态

Browser Extension

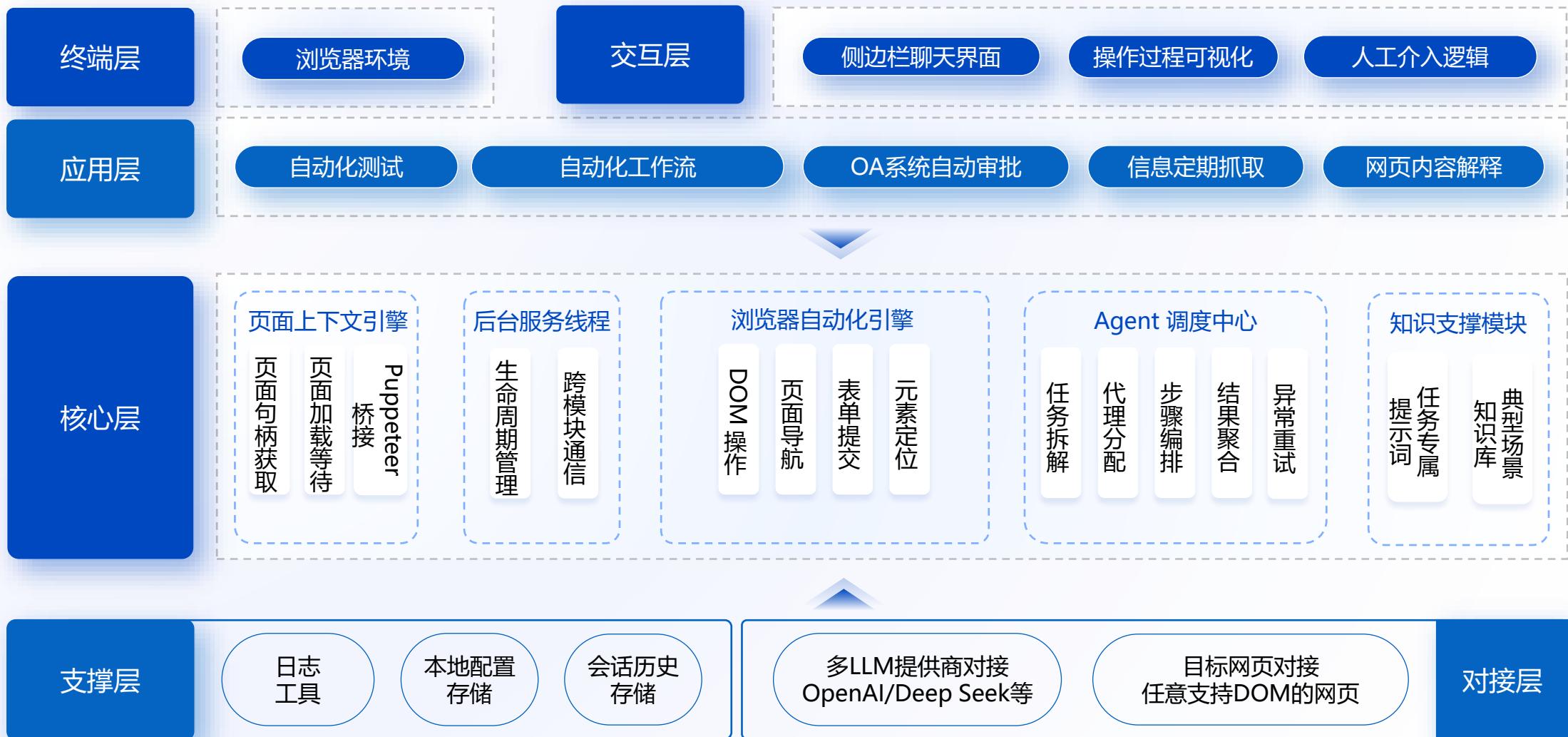
开箱即用 仅依赖现有浏览器

- 面向用户的、封装了 Web Agent 技术的完整应用。
- 免代码环境。无需 Python / Pip，零门槛上手。
- 不仅仅多出了一个聊天侧边栏，而是能够自主操作浏览器。

01



004 | 产品架构



004 | 开源产品的不足



01

FIRST

适配问题

开源产品未针对公安业务场景特定需求进行定制化开发

前端界面定制化改造

后端用户管理功能开发

跨系统数据流转功能开发

02

SECOND

在业务平台的表现

在公安业务平台上，Agent能力有待提升

现状：通过详细的自然语言进行步骤描述，来执行任务。
目标：通过简练的自然语言进行需求描述，来执行任务。

现状：仅进行了一个任务的验证
目标：能够完成多种任务

03

THIRD

本地模型集成

GUI Agent 调用本地大模型存在兼容性和稳定性问题

Tool Call

500 Error

04

FOURTH

Human In the Loop

通过人机协作的主动澄清机制消除指令模糊性

指令模糊

信息缺失

主动澄清

信息补全

工具调用

确定性执行

004 | 特定业务平台的SOP流程知识库构建：技术路线



必要性

公安业务平台具有高封闭性与复杂逻辑。Web Agent 拥有通用知识，但无法操作专业性强的特定系统。必须通过结构化的实操经验，将通用 Web Agent 转化为“行业专家”。

SOP

一份标准操作程序 (Standard Operating Procedure, SOP) 是一种结构化的文档，提供清晰的、逐步的指令，用于完成特定流程或任务。

技术路线

手动编制

产品说明书
RAG

Trace2SOP

Agent
自主探索

具体方案

第一阶段

快速完成数个指定任务-已验证可行性
-计划基于此方案建立首批演示范例

特点

解决了 Web Agent 操作专业平台能力的“从无到有”，但扩展性差（依赖纯人力）。

第二阶段

说明书里全面介绍了系统的各项功能

特点

利用了现有文档，但依赖度高。不是每个平台都能找到说明书

第三阶段

专家操作一遍，后台录制键鼠轨迹，
LLM 解析轨迹生成 SOP，作为后续类似任务的参考

特点

适合重复性任务

第四阶段

学习页面间的跳转逻辑

特点

极低的人工干预成本

004 | 开发计划



2026.01.18

基于手动编制SOP方案，
完成视综/MRAG平台的
10项流程自动化任务

2026.02.08

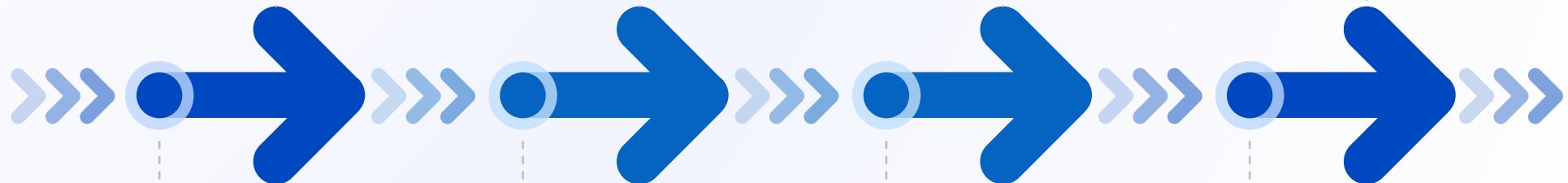
实现产品说明书RAG功
能，基于此功能完成10
项流程自动化任务

2026.02.28

跨系统数据流转功能开
发，基于此功能完成3
项跨系统自动化任务

2026.03.31

用户试用与反馈。开发
Agent自主探索功能。



2026.01.11

解决本地模型集成问题

2026.01.31

实现Trace2SOP功能，
基于此功能完成10项流
程自动化任务

2026.02.15

实现 Human In the Loop
基于此功能完成10项意图模
糊的自动化任务

2026.03.15

实现后端用户管理功能
开发与前端界面定制化
改造

开发周期
三个月