

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：2026.01.12

目录

CONTENTS



背景与需求

Background and Requirements Analysis

01



流程自动化

Process Automation

02



浏览器自动化智能体

Web Agent

03



季度工作计划

Quarterly Work Plan

04

001 | 背景与需求：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



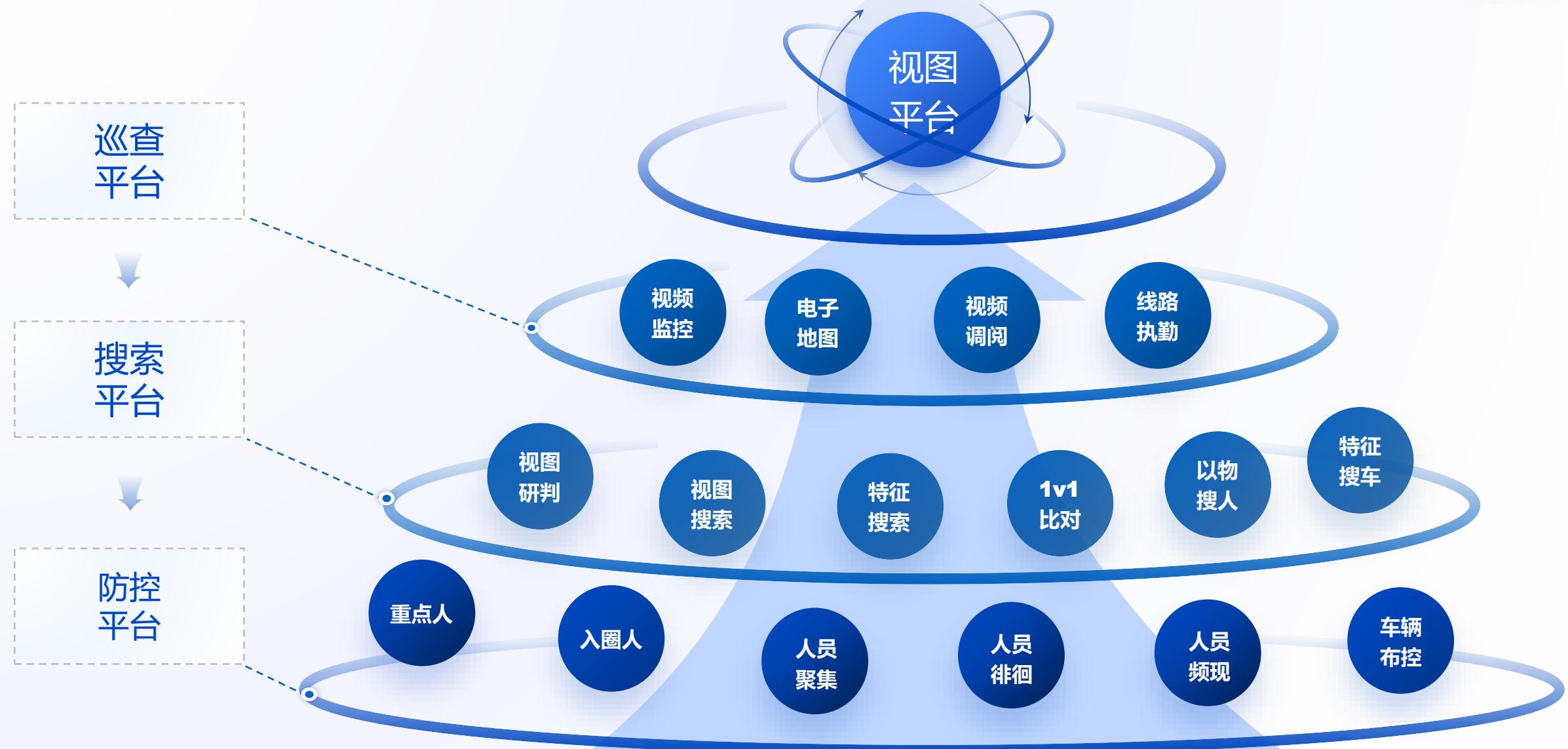
跨系统联动
数据跨系统流转

The screenshot shows the Hikvision Video Image Comprehensive Application Platform (V2.3). The main interface features a search bar at the top 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'". Below the search bar is a large central window titled "我将为你找到Ta" (I will find it for you), which displays a 3D city map with several cars and a person. A red box highlights this central search feature. To the left of the map is a sidebar menu with sections like "我要看" (I want to see), "我要搜" (I want to search), "我要应" (I want to respond), "我要建" (I want to build), and "我要利" (I want to benefit). Another red box highlights this sidebar. At the bottom of the sidebar is a "常用应用" (Common Applications) section with icons for various features: 留守部位人员防控 (留守部位人员防控), 智能名册库 (Smart Name List Library), AI以文搜图 (AI Image Search), 智能识别 (Smart Recognition), 视频调阅 (Video Playback), 设备事件报警 (Equipment Event Alarm), 人脸识别 (Facial Recognition), 人证比对 (Person-Document Comparison), 电子围栏 (Electronic Fence), 智能工具 (Smart Tools), 智能搜索 (Smart Search), 面上寻踪 (Face Tracking), and a plus sign icon for adding more applications.

The screenshot 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 table of contents includes:

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 | 背景与需求：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。

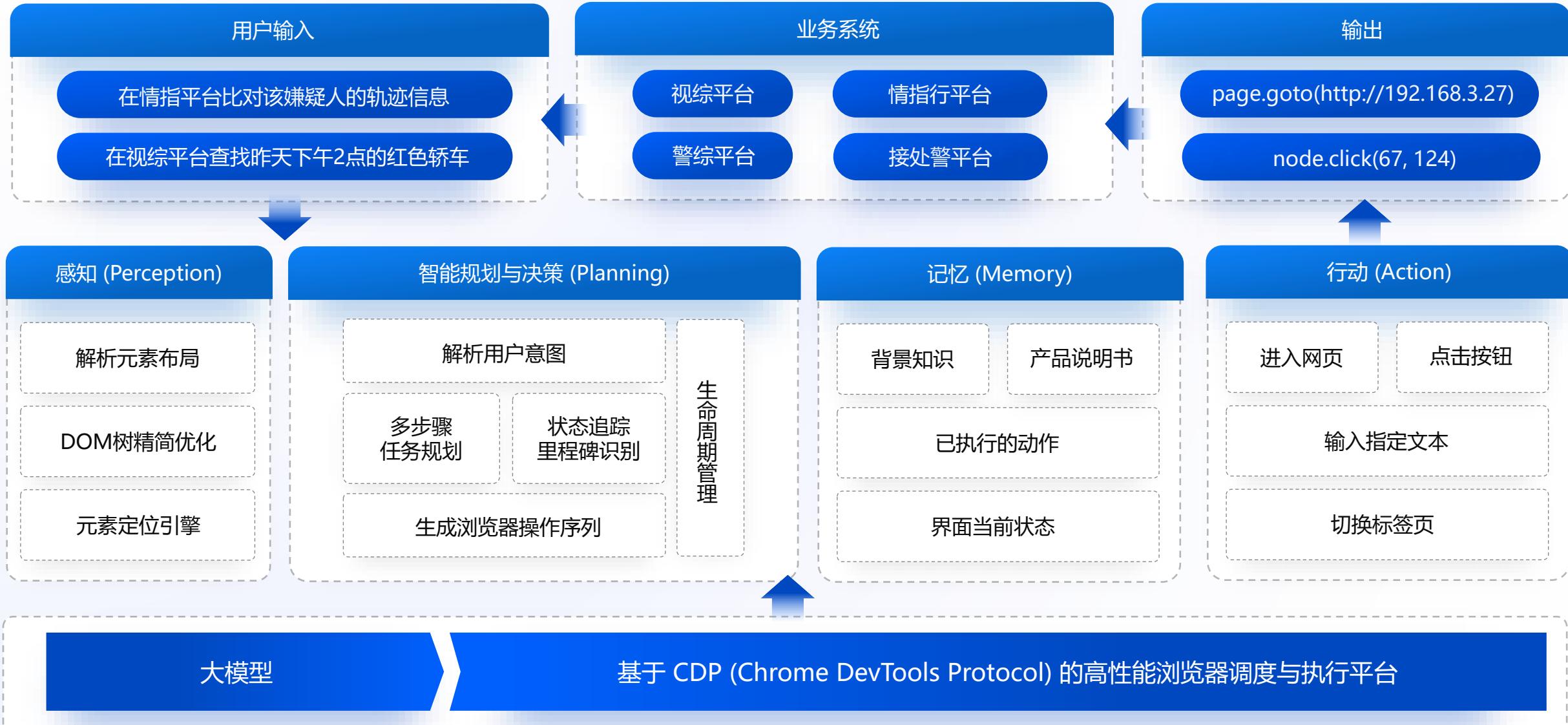
003 | 浏览器自动化技术发展历史



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

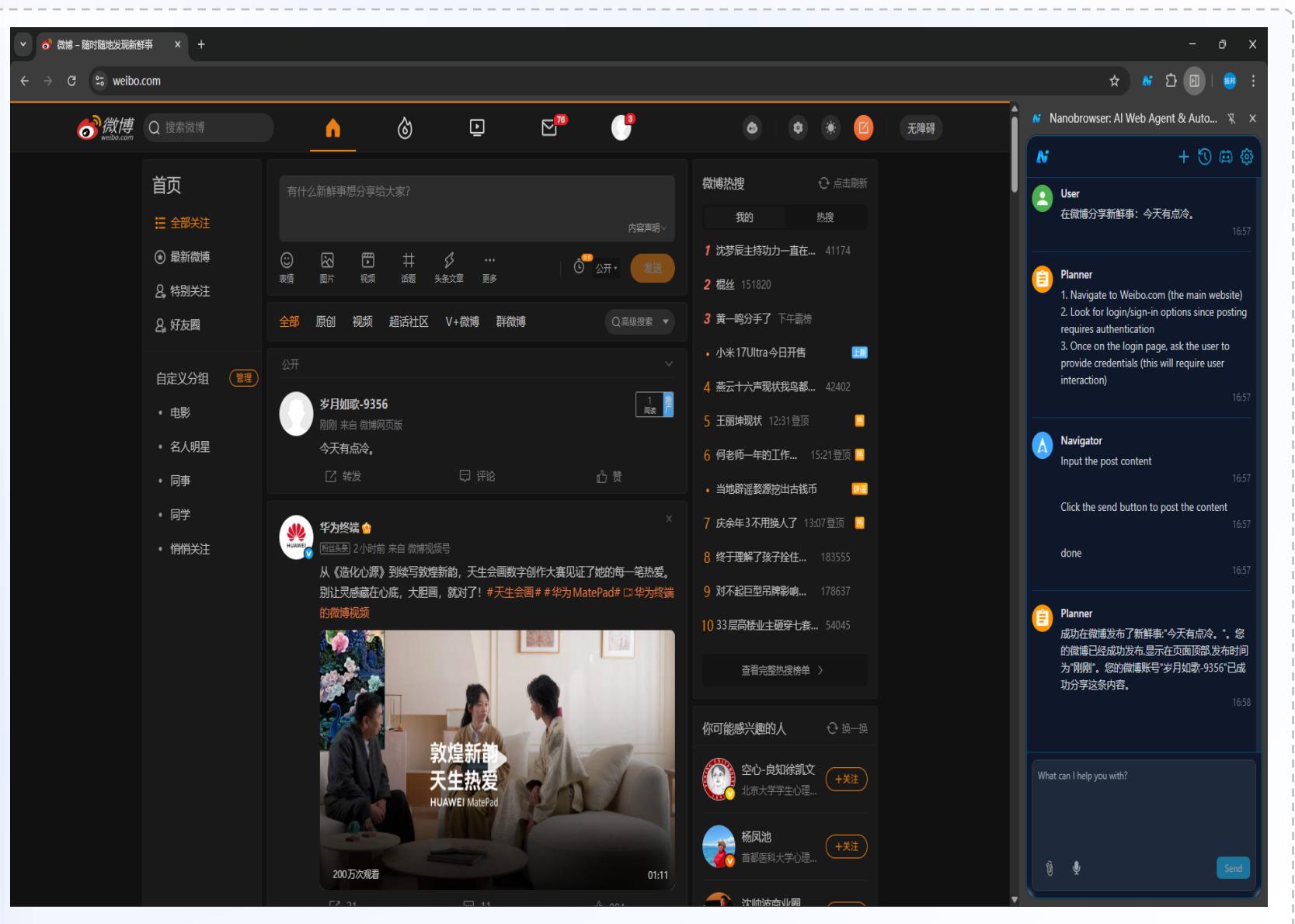
003 | 产品形态

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



003 | 产品架构





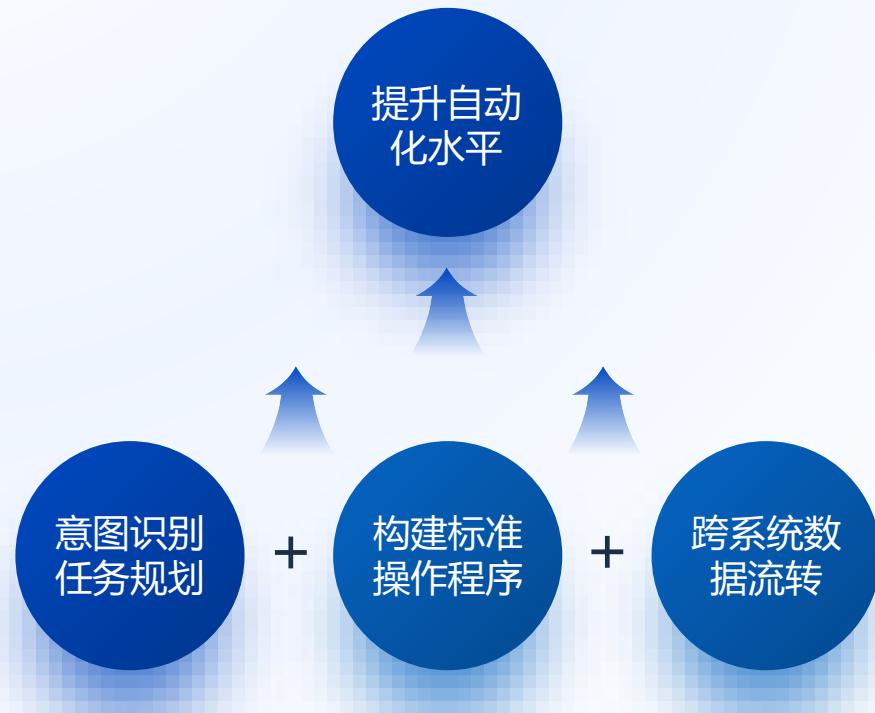
一、业务系统智能问答助手

基于RAG（检索增强生成）技术，通过自然语言交互深度检索业务手册，从“被动检索手册”转变为“主动智能问答”。



二、跨系统流程自动化

基于 Web Agent 技术，将自然语言指令转化为自动化执行序列，实现跨业务系统流程自动化。



004 | 开发计划

