OpenClaw 未来可能方向研究报告

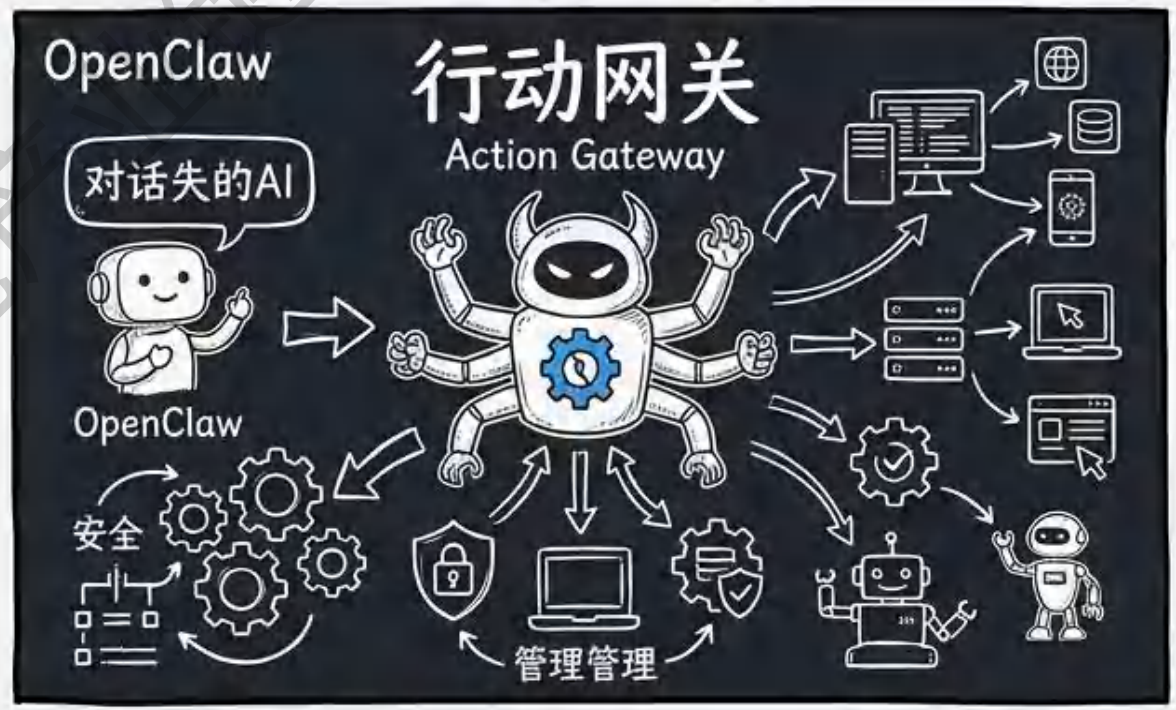
从“会聊天的助手”到“可治理的行动型操作层”

— 清新研究团队

— 基于公开资料、官方文档与论文研究

— 截至 2026年3月 的阶段性判断

Hero Image



人工智能产业链联盟

星主： AI产业链盟主

知识星球

微信扫描预览星球详情

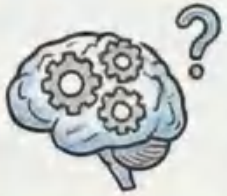
@清新研究团队简介

沈阳：清华大学新闻学院/人工智能学院双聘教授、博导

- 领导学术研究团队近30人。指导大数据、AI、人形机器人等多个产业团队。
- 团队坚持：整体主义、实证主义、社会建构、进步主义。

视频号：@清新研究；公众号：@清新研究

六大研究方向：



1. AI大模型理论与哲学

2. AI文艺

3. AI应用

4. 新媒体与网络舆论

5. 大数据

6. XR应用

邮箱：124739259@qq.com | 微博：@清新研究 | 公众号：@清新研究

核心结论先看

- **OpenClaw 的价值不在单点功能，而在定义代理操作层**

— 它把聊天入口、工具执行、记忆与工作区编成可运行的代理链路

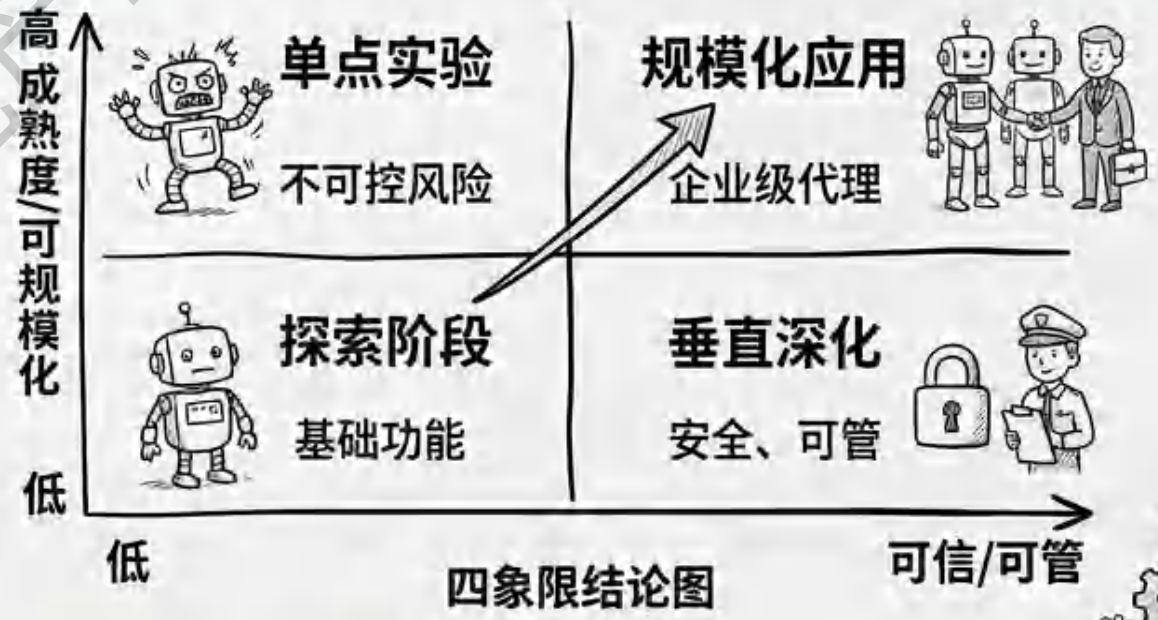
— 未来主线将从“更会做事”转向“更可信、更可管、更可规模化”

— 安全、企业化、记忆、多代理、浏览器行动接口、混合推理与可观测性将成为将成为关键方向

当前 focus：“更会做事” → 未来主线：“更可信、更可管、更可规模化”

— 未来主线将从“更会做事”转向“更可信、更可管、更可规模化”

— 安全、企业化、记忆、多代理、浏览器行动接口、混合推理 与可观测性将成为关键方向



四象限结论图

OpenClaw 到底是什么

它更像一个 self-hosted agent gateway，而非普通聊天机器人

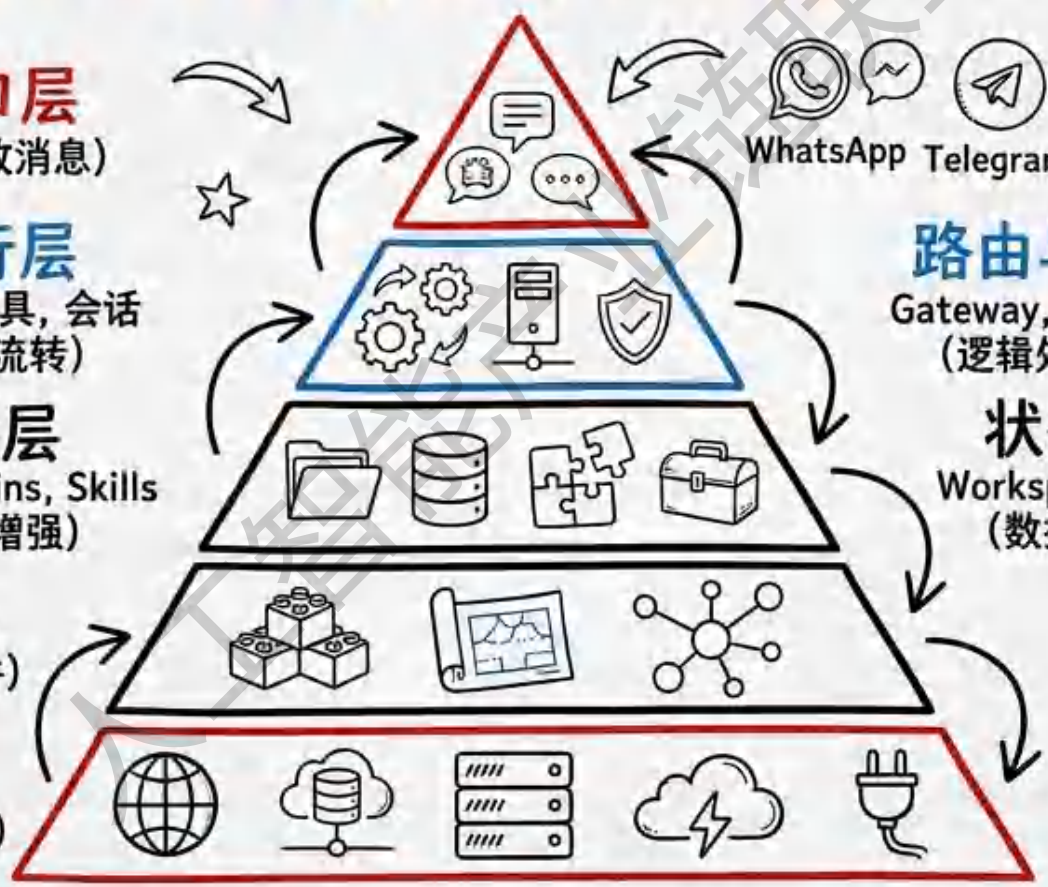
消息入口层
（用户交互，接收消息）

路由与执行层
Gateway, Agent, 工具, 会话
（逻辑处理，流程流转）

状态与扩展层
Workspace, Memory, Plugins, Skills
（数据持久化，能力增强）

核心架构层
（系统基石，核心组件）

宿主环境层
（服务器，操作系统，硬件设施）



WhatsApp Telegram Discord iMessage chatbot

路由与执行层
Gateway, Agent, 工具, 会话
（逻辑处理，流程流转）

状态与扩展层
Workspace, Memory, Plugins, Skills
（数据持久化，能力增强）

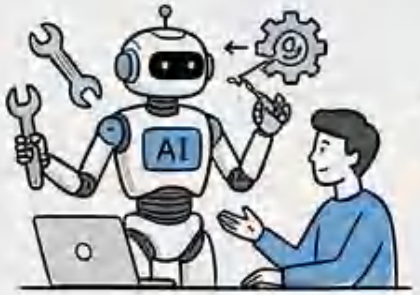
OpenClaw　　传统聊天机器人

@ 清新研究团队 | 2026年3月 | 数据来源：https://docs.openclaw.ai/

为什么说它是一个信号，而不只是一个项目

AI 正从“回答问题”走向“完成任务”

✗ 回答问题 → 完成任务 ✓



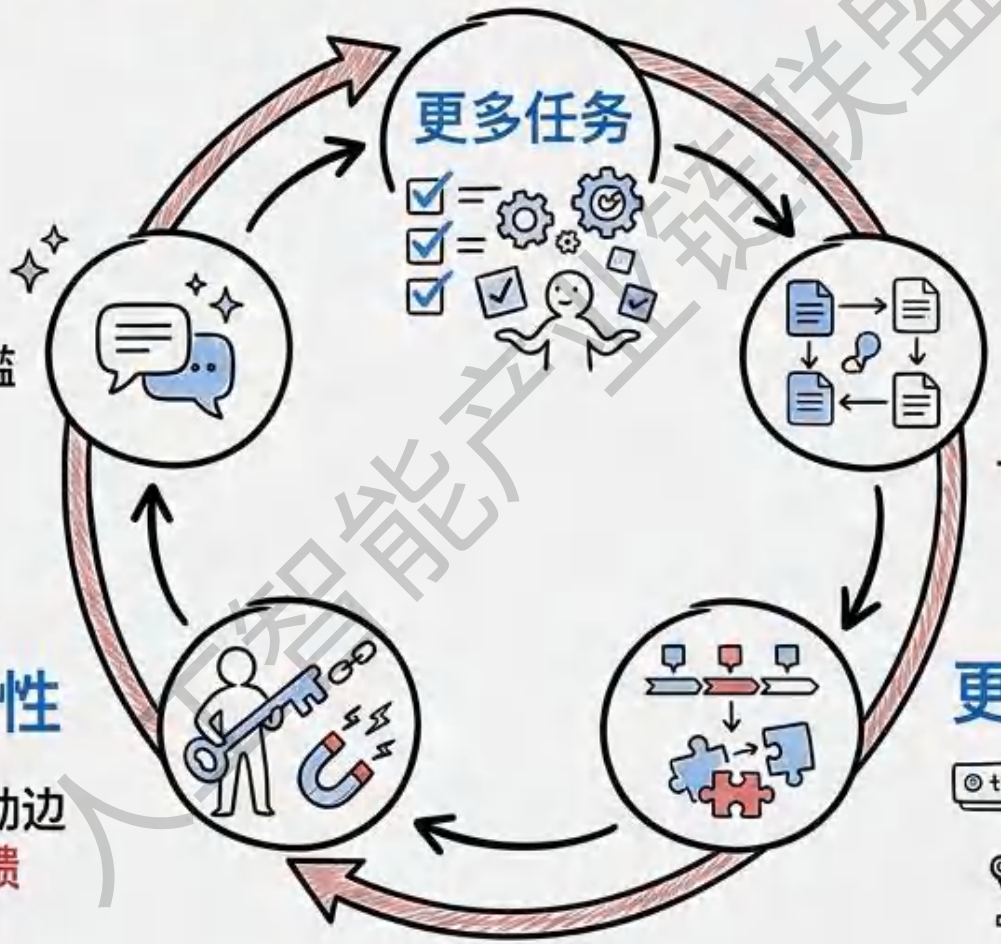
旧范式

- 旧范式关注回答质量，
- 新范式关注任务完成率与权限正确性

新范式

- OpenClaw 把 Agent 从研究概念压缩为为可安装、可演示、可扩展的产品形态
- 它定义了一套新的工程语言：网关、配对、记忆、工具、技能、审计

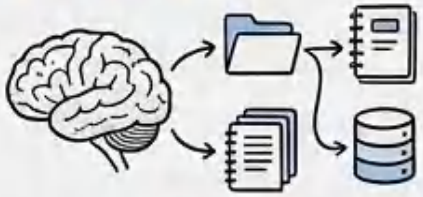
OpenClaw 的底层飞轮



熟悉入口

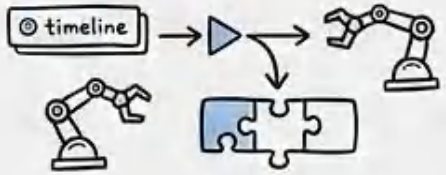
— 聊天入口降低使用门槛

更多记忆



— 工作区与记忆积累长期上下文

更强编排



更高粘性

— 技能与插件扩展行动边界，形成生态正反馈

它已经证明了什么

不是万能，但足够说明 Agent 化的方向成立

能力矩阵：入口、信任、编排、生态

入口：通用行动

— 聊天入口可以成为通用行动入口

信任：本地优先

— 本地优先是一种信任策略，而非保守路线

编排：护城河转移

— 真正的护城河开始从模型本身转向编层与执行层

生态：Agent 链条

当前约束

今天的问题，决定明天的路线图

安全是第一性约束

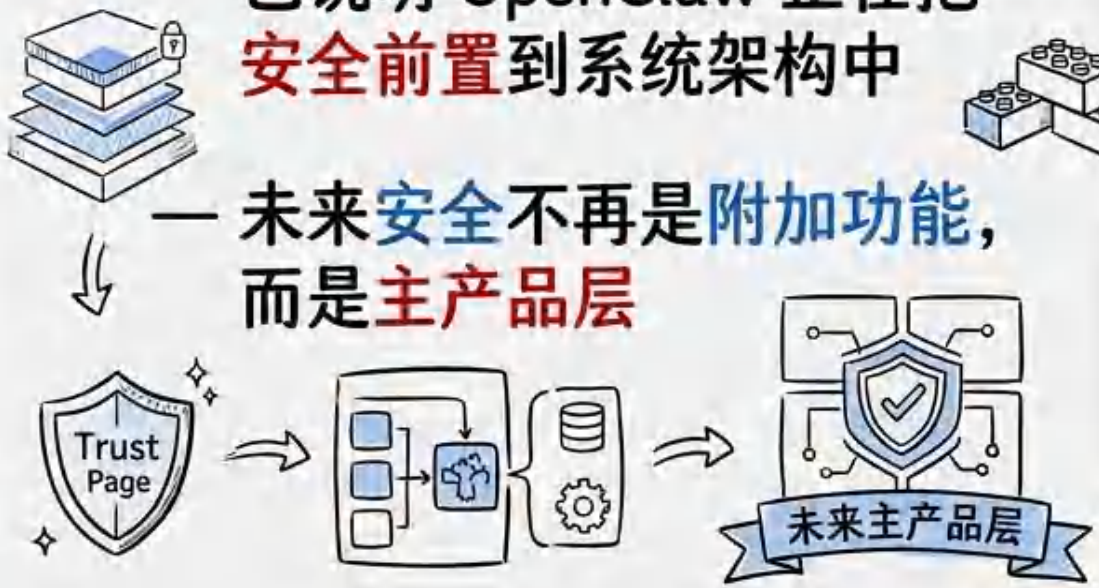
Agent 一旦能执行动作，风险就从回答错误升级为**权限错误**

左风险面

- 提示注入
- 恶意网页
- 恶意技能
- 凭证泄露与**越权执行**会**叠加**出现

右防护面

- Trust 页面与 Threat Model 已说明 OpenClaw 正在把**安全前置**到系统架构中
- 未来**安全**不再是**附加功能**，而是**主产品层**



可演示不等于可托付

真实工作流要求稳定、恢复与低门槛

技术门槛山峰

— 多步任务中的状态漂移、误触、重复执行仍是核心难题

稳定性检查清单

— 非技术用户在安装、配置、成本控制上存在明显障碍

— 未来必须强化 onboarding、模板、失败恢复与可视化配置

本报告原创核心概念

用五个概念解释 OpenClaw 的演化逻辑

— 行动界面层 交互与操控

— 权限编排面
安全与调度

— 记忆账本 演化轨迹
演化轨迹

— 宿主可信带
环境安全保证

— 代理监理层 合规与审计

@ 清新研究团队 | 2026年3月

团队首创概念

行动界面层

高熟悉入口 + 可执行上下文 + 低摩擦任务下发

聊天气泡与执行箭头构成的行动入口

— 未来最有价值的界面，不一定是最复杂的控制台，而是**最自然的任务发起面**

— OpenClaw 选择聊天入口，意味着它在定义一种**新的行动界面**

— 这层做得越顺，Agent 越可能进入**高频日常使用**

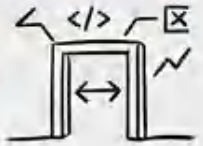
团队首创概念

权限编排面

工具边界 + 执行审批 + 风险分层 + 渠道隔离

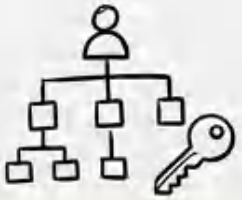
— 权限不再只是给与不给，而是要条件化、场景化、可审计地授予

条件化、场景化、可审计地授予



— tools.profile、allow/deny、pairing 与 browser profiles 已是雏形

— 谁先把权限编排做成熟，谁才可能进入组织场景



多层闸门与不同风险等级的动作流

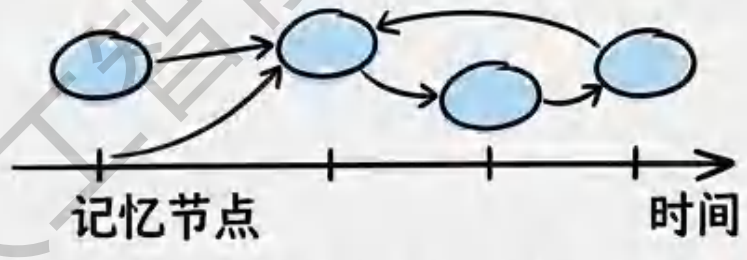
团队首创概念

记忆账本

可追溯记忆文件 + 语义索引 + 生命周期管理 + 纠错机制

— 记忆必须可见、可校正、可清理，而不是模型黑箱残留

— Markdown source-of-truth 是好的起点，但还不是终点



— 未来记忆会从可搜索文档集合升级为长期认知基础设施

团队首创概念

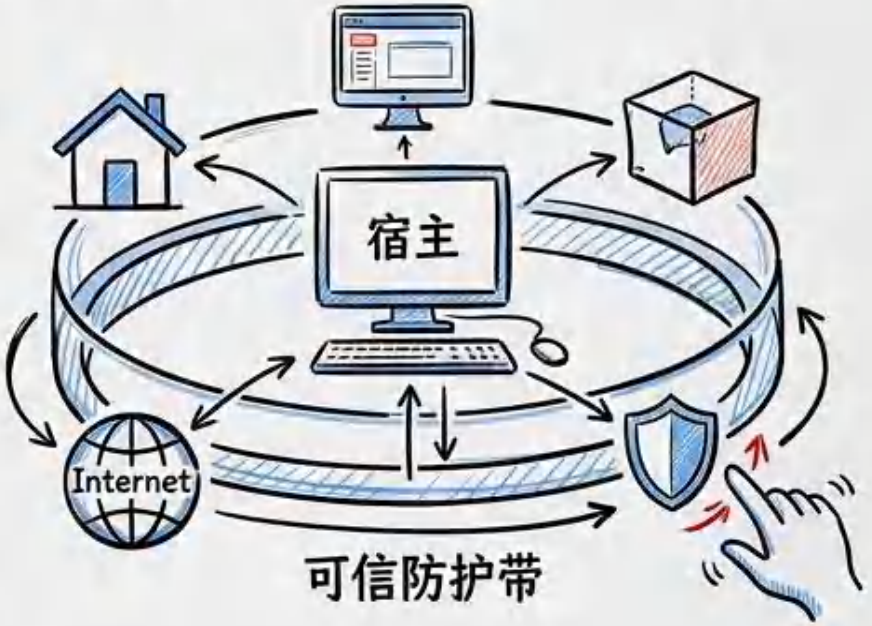
宿主可信带

本地优先环境 + 沙盒执行 + 可信浏览器/终端 + 凭证边界

— 用户真正需要的不是抽象安全，而是可感知的控制权

— self-hosted、workspace、sandbox、isolated browser 共同构成可信带

— 谁能建立更强宿主可信带，谁就更有机会赢得长期信任



@ 清新研究团队 | 2026年3月

代理监理层

团队首创概念

可观测日志 + 回放回滚 + 异常预警 + 人类兜底

— 未来 Agent 产品之间的差距，不只在谁最会做事，也在谁最容易被监督

— 监理层决定组织是否敢把更多动作交给代理

— 它会成为 Agent 进入现实世界的通行证

未来可能方向

从爆款项目到可治理基础设施

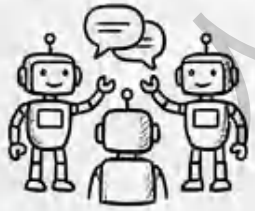
— 主线一：安全产品化

- 安全产品与目纲件
- 机制与证项责的结构

— 主线二：企业控制平面

- 企业控控制平面
- 分布分布控制系统

— 主线三：记忆、多代理、混合推理与监理层



方向一：安全将从功能旁注升级为产品主轴

- **权限细粒度化、隔离执行、供应链治理与安全可视化**

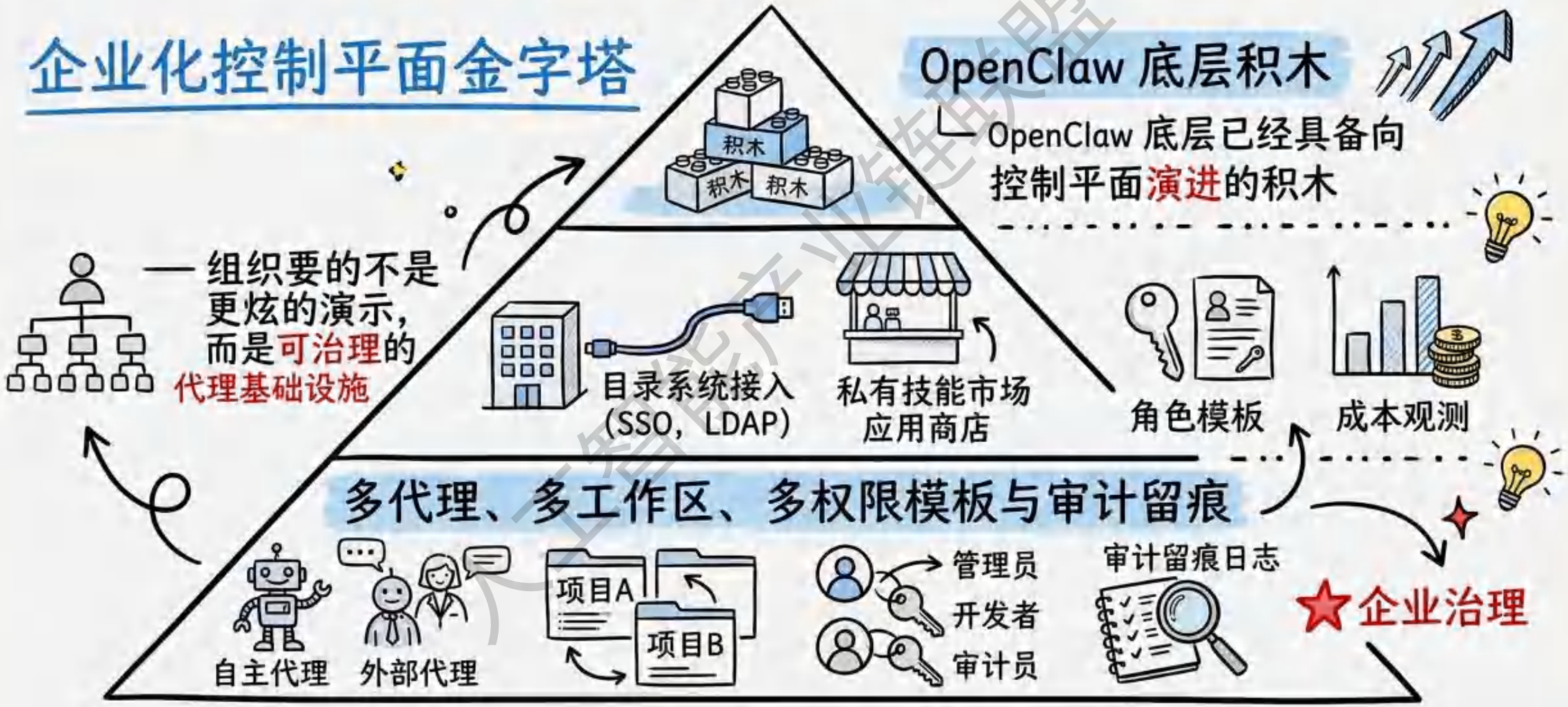
— 高风险动作进入强制审批与隔离环境

— 技能市场将出现签名、信誉、权限、权限标签与企业私库

— 安全将从文档层进入用户可感知的体验层

从默认配对到审计中枢的路线图

方向二：企业化控制平面会成为第二增长曲线



方向三：记忆将升级为长期认知基础设施

从 Markdown 记忆文件走向多层、可治理、可演化的记忆系统

未来记忆需要：

— 实体摘要，

— 时效管理，

— 冲突检测

和版本历史

四层记忆金字塔

记忆一旦沉淀，将成为比模型更高的迁移成本

方向四：从单代理走向多代理协作与“代理组织”

一个网关，多种角色代理，任务拆解与监督汇总

多代理环形协作图

多代理环形协作图

- 万能代理不是长期最优，角色分工才更符合真实工作流
- 多代理需要共享记忆、任务分发、冲突仲裁与反成本控制
- OpenClaw 的 multi-agent routing 已给出雏形

方向五：浏览器会被重构为“高风险行动接口”

— 浏览器是最通用的行动层，也是最集中的风险源

— 未来会出现更多隔离 profile、动作回放、站点级策略与可信页面包装

不是更像人点击网页，而是更受约束地完成高价值动作

— 谁能平衡“能干事”与“不乱来”，谁就最接近真实落地

方向六：本地优先 + 混合推理，将成为成本与隐私平衡点

本地做低风险重复任务，云端做高复杂关键决策

— Agent 的成本结构决定它不能长期只依赖最贵模型

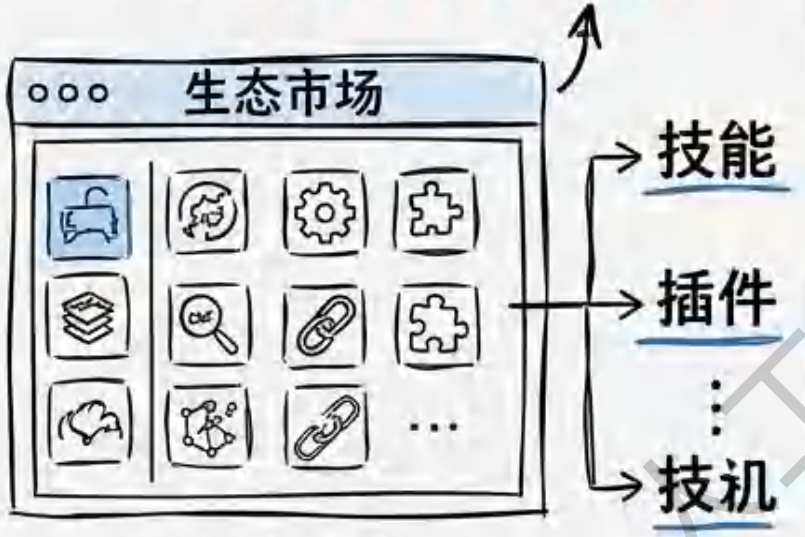
— 未来会出现规划模型、执行模型、审核模型的任务分层

— 混合推理是走向规模化部署的现实选择

方向七与八：生态平台化 + 代理监理层成熟

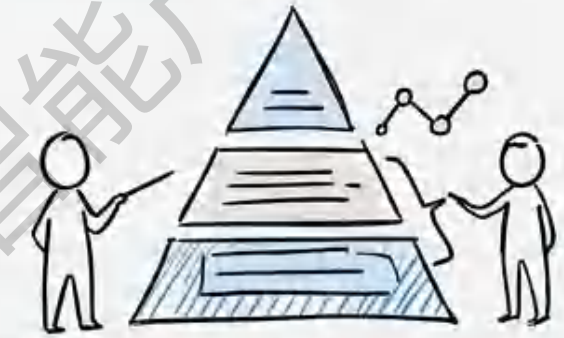
谁组织扩展生态，谁提供可观测与回滚，谁更接近基础设施

— 技能、插件与 MCP 连接器会一步平台化和分层化



生态市场

— 评测体系将一能力分数转向能力+安全+长时任务+可观测的组合



安全+长时任务+可观测的组合

— 代理监理层会成为组织准入的核心门槛

监理控制台

@ 清新研究团队 | 2026年3月 | 数据来源：https://arxiv.org/html/2602.17753v1

最终判断：OpenClaw 正在提前定义“代理原生软件时代”

- **它未必赢得所有终局，但大概率已经赢得了“定义问题”的先手**

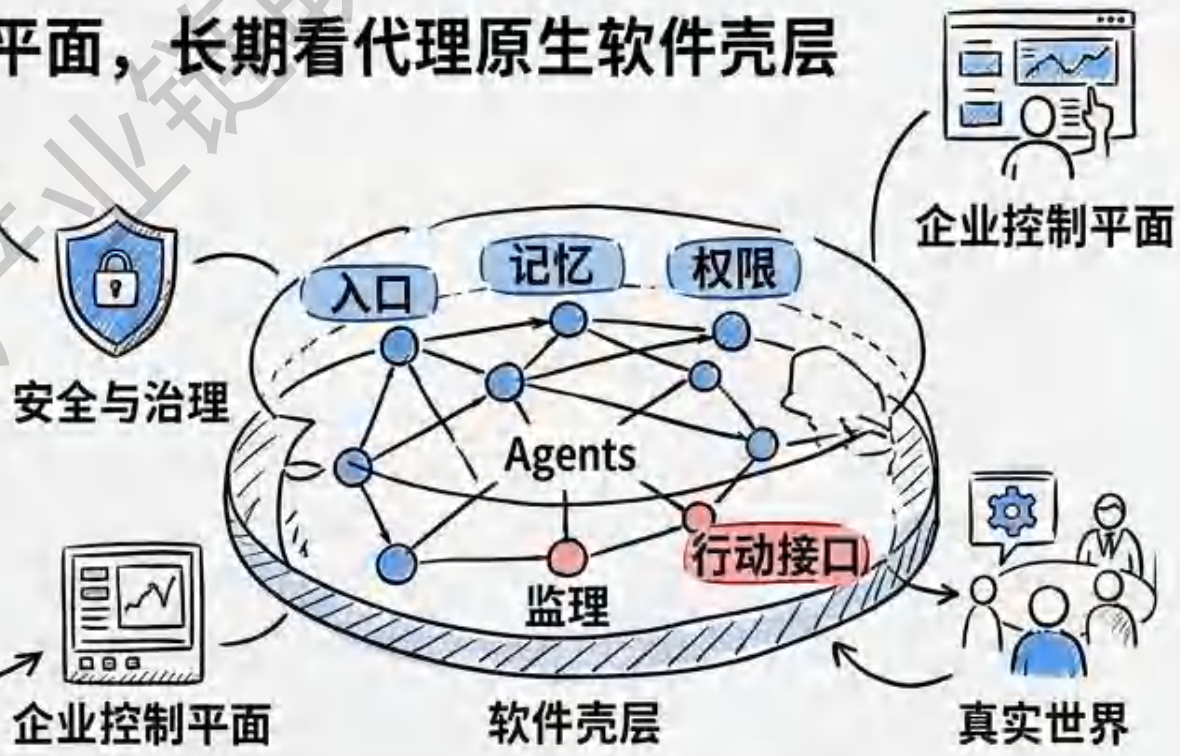
— 短期看安全与治理，中期看企业控制平面，长期看代理原生软件壳层

— 真正值得关注的是它塑造的新秩序：

入口、记忆、权限、监理与行动接口

— 真正值得关注的是它塑造的新秩序：
入口、记忆、权限、监理与行动接口

— 对行业而言，研究 OpenClaw 就是是在研究 Agent 如何进入真实世界



未来代理网络与软件壳层的全景结尾图

AI人工智能产业链联盟

#每日为你摘取最重要的商业新闻#

更新 · 更快 · 更精彩

Zero

AI音乐创作人
水墨动漫联盟创始人
百脑共创联合创始人
人工智能产业链联盟创始人
中关村人才协会秘书长助理
河北北大企业家分会秘书长
墨玫星辰智能科技有限公司CEO
河北清华发展研究院智能机器人中心线上负责人
中关村人才协会数字体育与电子竞技专委会秘书长助理

主要业务:AI商业化答疑及课程应用场景探索，各类AI产品学习手册，答疑及课程

欢迎扫码交流

提供：学习手册/工具/资源链接/商业化案例/行业报告/行业最新资讯及动态

人工智能产业链联盟创始人

邀请你加入星球，一起学习

人工智能产业链联盟报告库

星主：人工智能产业链联盟创始人

每天仅需0.5元，即可拥有以下福利！
每周更新各类机构的最新研究成果。立志将人工智能产业链联盟打造成市面上最全的AI研究资料库，覆盖券商、产业公司、研究院所等...

知识星球

微信扫码加入星球

感谢观看

THANK YOU

执行人：杜靖洋

提议人：沈阳

具体实施：OpenClaw

资料搜索及整理：ZeeLin Desearch

PPT生成：Gemini