

Agent 爆发

为什么 2026 年 Agent 真正出现

只有当模型能力、工具接口与成本三条线一起过线，智能体才真正从演示走向可用。



模型过线

Claude Opus 4.5 等强模型把编程、工具调用和长任务推进到可用阈值。



成本下降

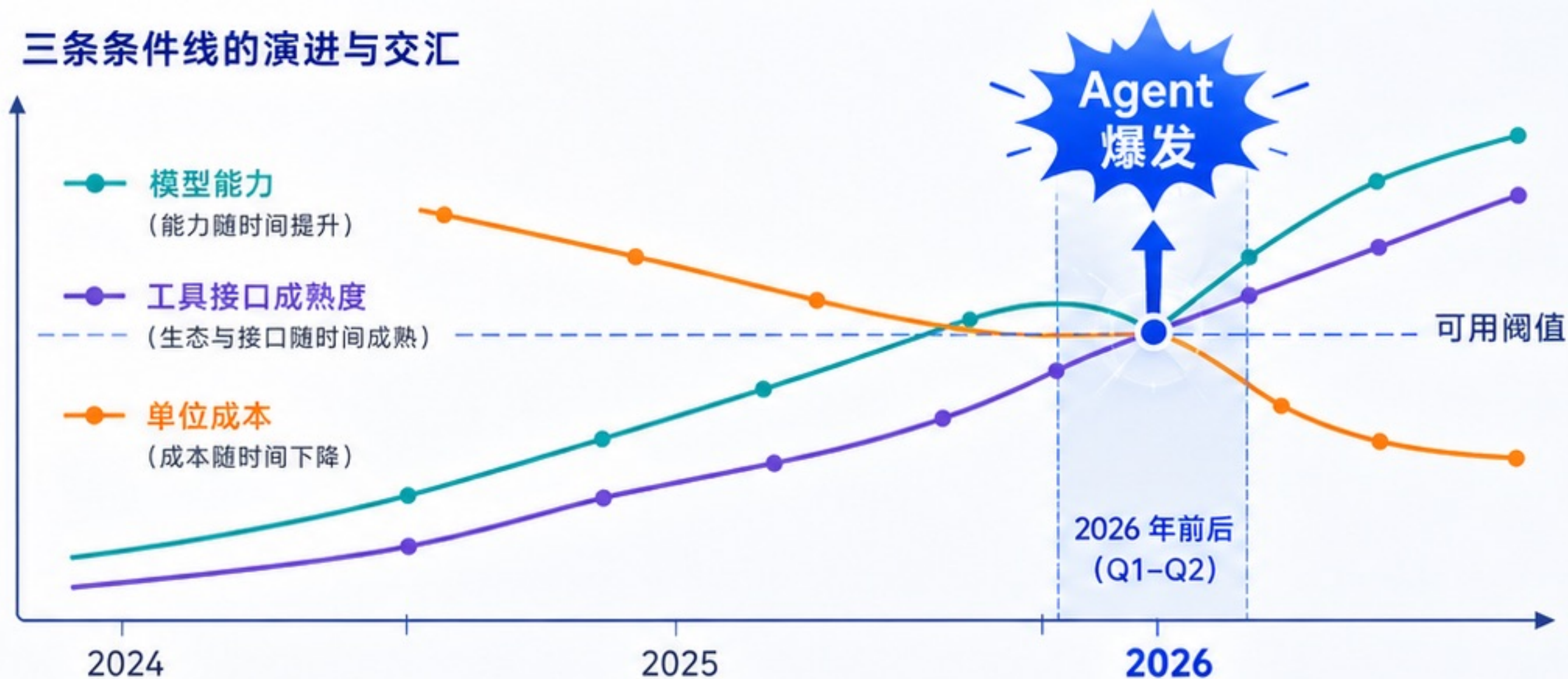
更便宜的模型也逐渐能规划、调用工具、处理多步任务。



接口成熟

MCP、Skills、浏览器 / 电脑使用能力，让模型终于接上真实世界。

三条条件线的演进与交汇



Agent 不是突然诞生，而是三条条件线在 2026 年前后同时交汇。

智能体定义

目标、计划、工具、反馈

智能体不是只会回答，而是围绕目标持续行动，并在结果中修正自己。



不是聊天机器人

聊天机器人回答问题；
智能体接收目标，并尝试完成任务。



问答

VS



完成任务



关键边界

能做事以后，权限、错误、审计和责任
都会变得更严肃。



权限



错误



审计



责任



Agent = 围绕目标行动、调用工具、读取反馈，并把结果交回人审查。

OpenClaw

个人智能体第一次真正出圈

它让很多人第一次感觉：AI 不只是聊天，而是在替我做事。



为什么出圈

像发消息一样交代任务，
门槛低，普通人也能立刻上手。



核心感受

不是回答一句话，
而是接住任务、继续往下做。



课堂定位

这是很多学生第一次
真正感受到 Agent 的时刻。



交代任务

OpenClaw
Personal Agent



邮箱

清理收件箱



日历

发邮件



文件

安排会议



浏览器

整理文件



终端 / 命令

打开网页



差旅 / 登机

办理登机



从聊天入口 → 任务执行 → 持续跟进



OpenClaw 让很多人第一次直观感受到：AI 不是聊天，而是在替我做事。

为什么它让人上头



聊天入口

用户通过 WhatsApp、Telegram 等入口发消息，不必一直坐在电脑前。



真的会做事

清理收件箱、发邮件、管理日历、办理登机、运行命令、调用技能。



情绪体验

它像一个随身数字助理：有记忆、能响应、能把任务接着做下去。



低门槛入口 + 后台持续做事



真正让人上头的，不只是会说，而是像随身数字助理一样持续替你做事。

能跑起来，不等于能长期托付



维护成本

环境、账号、API、插件、技能、系统权限都要持续维护。



上下文成本

要让它懂你，需要记忆、规则、文件和大量边界说明。



稳定性成本

多步任务里任何一步失败，都可能让整个 workflow 偏航。



退潮原因

新鲜感过去后，许多人发现它强大但难养、难管、难放心。



让 Agent 跑起来只是第一步；真正难的是持续维护、持续理解你、持续稳定。

权限越大，风险越大



权限范围

它可能读写文件、运行命令、发消息、连接邮箱和日历。



攻击面

技能、插件、外部连接与身份凭证，会共同扩大风险暴露面。



课堂判断

个人助理不是玩具；它是一种拥有身份和权限的新执行者。

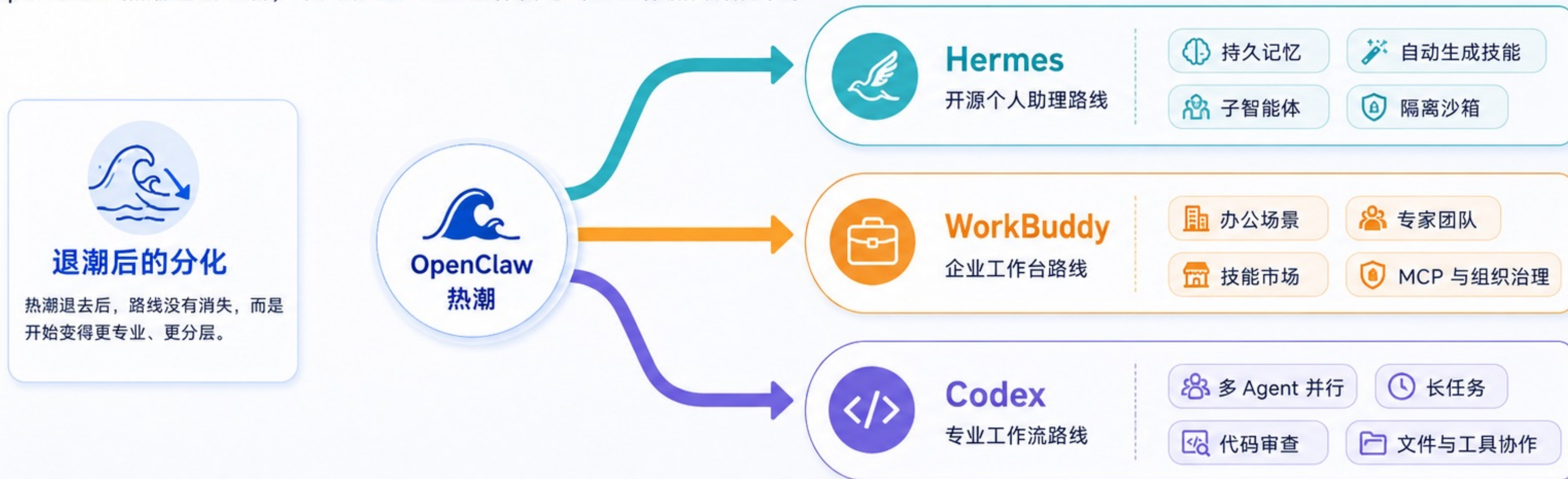


当 Agent 拥有身份、工具和权限时，安全问题就不再是附带问题，而是核心问题。

退潮后的路线

Hermes、WorkBuddy、Codex

OpenClaw 热潮退去之后，个人助理、企业工作台与专业工作流开始分化。



退潮不是结束，而是分化：个人自托管、企业治理、专业工作流三条路线开始清晰。

MCP

让智能体接上外部工具

Anthropic: 到 2026 年 MCP 已有 10000+ 活跃公共服务器，并被 ChatGPT、Cursor、Gemini、Copilot、VS Code 等采用。



MCP 让模型不再只会回答，而是能通过统一接口接上真实工具世界。

把流程经验打包给智能体



什么是 Skill

把某类任务的规则、步骤、参考文件打包成可复用的能力。



为什么重要

MCP 给工具连接，Skills 给流程经验，让 Agent 不用每次从零摸索。



教育例子

批改作文、检查数学步骤、生成教案、整理课堂观察，都可以做成技能包。



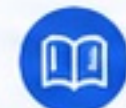
安全提醒

技能包会影响 Agent 行为，来源、权限和审计都要检查。



MCP = 工具连接

连接外部工具和服务，
扩展能力边界。



Skills = 流程经验

封装可复用的规则与步骤，
传递领域流程智慧。



Skill 的价值，不只是多一个功能，而是把可复用的流程经验交给智能体。

Agent 是一种新身份

身份

它能代表用户执行动作，因此不只是软件功能，而是带权限的数字身份。

权限最小化

只给完成任务必要的账号、文件、工具和网络访问。

审计与供应链

保留日志、审批关键动作，不随便安装陌生技能包。



理解 Agent 的第一步，不是把它当工具，而是把它当作带权限的数字身份来管理。

为什么最先成熟



反馈清晰

代码能运行、测试能失败、错误能定位，Agent 有明确反馈闭环。



工具完备

仓库、终端、测试、CI、代码审查天然适合自动化。



阈值已到

Cursor、Claude Code、Codex 等把编程从补全推进到任务协作。



人的角色

人从写每一行代码，转向定义目标、审查 diff、管理任务。

代码智能体工作流



编程场景先成熟，不是因为最简单，而是因为它最容易形成清晰的反馈闭环。

工作方式

从副驾驶到同事



当 AI 从补全器走向协作者，人的核心能力也从“亲自做”转向“定义、分工、审查与协作”。

中国模型：进入 Agent 阈值

DeepSeek、豆包/Seed、千问/Qwen、GLM、Kimi、MiniMax：从低成本推理、开源生态、多模态到 Agent workflow。

为什么说进入 Agent 阈值？

- 会推理
- 会调用工具
- 会处理多模态
- 会组织 workflow

国产模型竞争，正在从“能聊天”走向“能完成任务”。



中国模型正在从“参数竞争”走向“任务交付竞争”。

企业工作台：从个人助理到组织流程

AI 真正进入学校和企业，不只是更聪明，而是更可治理。



教育启示

学校真正需要的不是放任每个人乱接工具，而是可治理的 AI 工作台。



AI 进入组织，关键不只在能力，更在治理能力。

成本现实：Agent 会消耗更多 token 和算力

教学场景里，不能只看单次效果，还要看长期规模成本与管理成本。

为什么更贵



反复规划

长任务需要多轮思考与拆解。



工具调用

观察结果、再调用工具会放大 token 消耗。



纠错循环

失败后重试与修正继续增加算力开销。

成本仪表盘



教学影响

视频、多模态和长任务不能只看效果，还要看规模成本。



管理问题

谁付费、谁限额、谁审计，会成为学校 AI 采购与使用的硬问题。



AI 越能做事，成本越需要被看见；可持续使用，必须把效果、预算与治理一起考虑。

最后带走的一句话 · 写给老师

拥抱 AI，但不外包判断

大模型不是魔法，也不是末日；真正重要的是学会定义任务、审查结果、保护边界，并保持自己的思考能力。



如果只带走一件事



大模型不是魔法，也不是末日。
它是一种正在进入工作流的强大工具。
未来的关键能力不是“会不会问 AI”，
而是会定义任务、会审查结果、
会保护边界、会和 AI 共创。



写给老师的话



请把这句话带给学生：
未来不是不用 AI 的人赢，
也不是盲目依赖 AI 的人赢，
而是能指挥 AI、审查 AI、
约束 AI，并保持自己思考能力的人赢。



定义任务

清晰目标与边界，
让 AI 做对的事。



审查结果

验证、质疑与评估，
让结果可靠可信。



保护边界

守住隐私、规范与伦理，
让使用安全可控。



与 AI 共创

发挥各自优势，
实现更大价值。



未来属于会使用 AI、也会约束 AI，并始终保持独立思考的人。