竹势

FIELD REPORT / ENTERPRISE EXECUTING AGENTS

从“AI小龙虾”到“AI市场部”

企业执行型智能体应用报告

个人智能体能跑通一项任务；企业系统必须能回答谁授权、用了什么证据、异常如何停止、结果由谁负责。

企业化分水岭是可问责

判断系统能否被正式委托，而不只看它会多少工具

八道门决定能否规模化

把身份、上下文、权限、证据与责任接进同一执行链

自主性必须逐任务授权

用价值、损失、可逆性与敏感度决定人工控制点

365 天形成经营闭环

从边界验证、受控试点走向多场景组合运营

FIELD RECORD 01

执行摘要

现场判断 / FIELD DISPATCH

个人智能体证明了机器可以完成任务；企业真正要建立的，是一套能够交责、停机、复盘并持续改进的执行制度。

结论先行

个人执行型智能体把“AI会说话”推进到“AI会动手”，因此快速激发了管理者对数字员工的想象。但个人体验中的流畅执行，不能直接等同于企业生产力。企业真正购买的不是工具调用次数，而是可控地完成经营任务、保留决策证据、在异常时停机、在结果偏差时复盘，并由明确责任人承担最终判断。

企业化的分水岭，是组织能否把智能体放入一套可管理的业务系统：目标清晰、身份可验证、上下文可信、权限最小化、流程有状态、操作有证据、能力持续评测、结果有人负责。缺少其中任何一项，个人智能体越“能干”，企业风险敞口反而越大。

本报告提出“个人能干活—企业能交责”八道门、“三重跨越”、企业执行型智能体五级阶梯、任务自主性决策矩阵、AI市场部最小可行运行架构、营销场景投资组合与14/30/90/180/365天迁移路线。它们共同构成企业在未来12—18个月的决策底图。

董事会问题	本报告答案
为什么个人智能体容易成功?	任务边界短、上下文由个人补齐、权限由个人承担、错误可被即时纠正。
为什么企业化容易失败?	组织缺少统一目标、身份、状态、权限、审批、证据与责任设计。
先做什么?	选择高频、可逆、证据充分、数据敏感度可控的单流程试点。
何时扩张?	当流程稳定、异常可发现、结果可归因、责任可执行时再提升自主等级。
何时停止?	权限漂移、证据缺失、人工绕过审批、业务指标无改善或事故率上升时立即停止扩张。

十条核心判断

核心判断	经营解释
1. 工具调用不是托付依据	企业可托付任务的前提，是智能体拥有可验证身份、受控权限、真实业务状态和完整证据链。

核心判断	经营解释
2. 个人智能体的价值上限是个人带宽	它能显著压缩检索、草拟、整理和机械操作时间，但难以独立承接跨部门、跨系统、跨责任人的长期流程。
3. 企业化失败主要发生在接口处	模型能力往往不是首要问题，身份、知识、权限、数据质量、流程状态、人工审批和系统连接的断点更致命。
4. 多 Agent 不是默认答案	当一个智能体加稳定工作流可以完成任务时，增加角色数量通常只会增加协调成本、错误传播和审计难度。
5. 自主性必须按任务定价	价值越高、错误损失越大、可逆性越差、证据越弱、数据越敏感，自主等级越应保守。
6. AI市场部首先是一套运行机制	它以经营目标、业务对象、状态机、角色技能、知识记忆、审批权限、评测监控和经营结果构成闭环。
7. 人工审批不是低级阶段	在品牌发布、广告预算、客户承诺、价格政策、个人信息处理和危机回应中，人工审批是成熟治理的组成部分。
8. 价值衡量必须连接经营结果	调用量、生成量和节省工时只能说明活跃度；真正的价值要看速度、质量、转化、收入、风险和组织学习。
9. 事故管理要设计在上线之前	预防、发现、暂停、回滚、取证和问责必须成为产品能力和运营制度，不能依赖临时人工救火。
10. 未来12—18个月的最佳顺序是先治理后扩张	先建立最小治理底座和一个可衡量闭环，再扩展连接器、流程跨度和自主等级。

十条判断共同指向一个结论：企业应把执行型智能体视为“受约束的经营执行单元”，而不是更聪明的聊天窗口。

FIELD RECORD 02

第一章 热潮改变了预期，但没有自动解决企业责任

四个信号同时出现，才值得进入企业试点

能力增长并不自动等于可委托。

01
能连续执行
不止回答问题

02
能调用工具
触达真实系统

03
能留下证据
过程可复核

04
能接受约束
异常可停止

个人执行型智能体的真正突破，是把自然语言意图转化为可观察动作；企业经营的真正难题，则是让这些动作在组织规则内持续正确。

结论

个人执行型智能体的真正突破，是把自然语言意图转化为可观察动作；企业经营的真正难题，则是让这些动作在组织规则内持续正确。

证据

个人用户可在一个会话中补齐上下文、授权工具并即时纠错，任务结束后责任自然回到本人。企业任务跨越多人、多系统和多时段，任何一个隐含假设都可能形成风险。

经营含义

管理层不应以“演示是否惊艳”判断企业价值，而应检查动作是否绑定目标、对象、状态、权限、证据与责任。

行动

将所有候选智能体场景先放入八道门和自主性矩阵，未通过治理门槛的场景不得进入生产环境。

1.1 热潮背后的三项真实变化

第一，交互入口发生变化。管理者开始用自然语言直接表达业务意图，系统负责拆解步骤并调用工具，软件使用从“人找功能”转向“目标驱动”。

第二，自动化边界发生变化。传统自动化依赖固定规则，执行型智能体可以处理半结构化信息、生成中间产物并在有限范围内选择下一步。

第三，组织预期发生变化。企业不再满足于AI提供建议，而是要求其进入内容、社媒、广告、CRM、会议和复盘流程，承担可观测的执行工作。

变化	个人体验	企业要求
交互	一句话发起任务	意图必须转化为受控工单与可审计任务
执行	调用浏览器、文件、消息等工具	连接器需要身份、权限、速率、范围和审批策略
记忆	保留个人偏好与会话上下文	区分个人记忆、组织知识、业务状态与证据档案
协作	用户本人持续盯进度	多个角色共享状态，明确交接、升级与责任人

1.2 “会做事”与“做对事”的差距

智能体能够连续完成操作，并不代表任务被正确理解。企业任务通常包含隐含的业务口径、品牌边界、客户承诺、预算约束和合规要求，这些约束必须被结构化。

企业可接受的执行结果，不仅要看最终产物，还要看输入来源、判断依据、审批记录、执行日志和异常处置。没有证据的正确结果仍然不可复用，也无法在事故中自证。

因此，企业应把“动作成功率”降为基础指标，把“目标达成率、证据完整率、越权率、回滚能力和责任闭环率”提升为核心指标。

- 动作完成不等于业务完成。
- 业务完成不等于经营价值兑现。
- 经营价值兑现不等于可以放宽权限。

1.3 “小龙虾”真正改变的是委托入口

场景 | OpenClaw 官方文档把它定义为自托管的智能体 Gateway：一个网关把 WhatsApp、Slack、Teams 等消息渠道接到编码智能体，并统一管理会话、路由和节点。安全说明同时明确，获得相应工具后，助手可以执行 Shell 命令、读写文件、访问网络服务和对外发送消息。[S29][S30] 这解释了“养龙虾”为何比普通聊天机器人更具冲击力——管理者不再只是索取答案，而是在即时通信里发出可以落到系统动作上的委托。

个人体验把五个角色压缩在同一个人身上：提出目标、补充上下文、批准权限、观察过程、承担后果。企业环境恰好相反，同一项营销动作可能跨越内容、品牌、法务、渠道、客户数据和预算。聊天入口越顺滑，越容易让这些原本分离的责任在一次自然语言指令中被无意合并。

赫伯特·西蒙在《管理行为》中把组织决策理解为在有限理性下分配决策前提与权限。放到执行型智能体中，提示语只是表面指令；真正决定行动的是目标、事实来源、角色权限、审批规则和停止条件的组合。[S31] 企业若只保留聊天记录，而没有把这些决策前提写进系统，便无法判断智能体究竟代表谁在行动。

边界｜OpenClaw 的官方定位首先面向希望自行掌控数据的开发者和高级个人用户；其文档没有宣称默认配置能够满足大型企业的身份、审计、隔离与合规要求。它证明了个人执行入口的可行性，不证明任何企业可以直接把个人 Gateway 当作生产平台。

管理动作｜在讨论部署方式之前，先列出五张表：谁能发起任务、智能体以什么身份执行、可以读取哪些数据、可以调用哪些动作、异常由谁接管。五张表不能落到系统配置和责任人，试点只能停留在隔离环境。

FIELD RECORD 03

第二章 个人智能体的价值上限，由个人带宽与责任边界决定

个人执行型智能体最适合成为高能力个人的“执行放大器”，不适合未经改造就成为组织的“责任承担者”。

结论

个人执行型智能体最适合成为高能力个人的“执行放大器”，不适合未经改造就成为组织的“责任承担者”。

证据

个人场景中，用户同时扮演目标定义者、上下文提供者、授权者、监督者和最终责任人；企业场景必须把这些职责拆分给不同角色和系统。

经营含义

个人智能体的价值可以很高，但价值形态主要是节省认知切换和机械操作时间，而不是自动形成跨团队生产力。

行动

先识别个人副驾可以稳定完成的任务，再补齐公共知识、共享状态、审批权限和运营监控，逐步转为团队能力。

2.1 个人智能体最擅长的四类工作

第一类是资料与信息处理，包括汇总、分类、摘要、转写、格式转换和初步分析。

第二类是草拟与改写，包括邮件、文章、报告、脚本、会议纪要和初版方案。

第三类是轻量工具操作，包括建立日程、整理文件、填写表单、生成任务和调用低风险接口。

第四类是个人工作流串联，即在单一责任人持续监督下完成若干连续步骤。

任务类型	典型价值	个人前提	企业化缺口
信息处理	减少查找与整理时间	用户能判断来源与结论	缺统一来源标准与证据保留

任务类型	典型价值	个人前提	企业化缺口
内容草拟	提高初稿速度	用户掌握品牌与业务背景	缺品牌知识、审批和发布权限控制
轻量操作	减少机械点击	用户对账号和动作负责	缺服务身份、最小权限和操作审计
个人流程	降低切换成本	用户全程监督	缺共享状态、交接规则和升级机制

2.2 企业化失败通常发生在六个断点

身份断点：智能体使用个人账号、共享密钥或无法区分调用主体，导致权限和责任无法追溯。

知识断点：智能体依赖散落文档和会话记忆，无法判断版本、有效期、适用对象与冲突规则。

流程断点：任务状态停留在聊天文本，未写入统一工单、内容对象、客户对象或活动对象。

协作断点：多个部门各自使用智能体，没有统一任务队列、交接标准和审批时限。

证据断点：来源、草稿、修改、审批、执行和结果没有连续记录。

责任断点：事故后无法确定谁定义目标、谁批准权限、谁审核内容、谁负责业务后果。

2.3 授权会在熟悉中自然膨胀

数据｜Anthropic 对 Claude Code 与公开 API 中数百万次人机交互的匿名分析显示，绝大多数 Claude Code 单轮工作很短，中位数约 45 秒；但 2025 年 10 月至 2026 年 1 月，最长一档会话的 99.9 分位数由不足 25 分钟升至超过 45 分钟。经验更丰富的用户更常使用自动批准，同时也更频繁地主动中断系统。[S32]

这组结果揭示了企业治理中的行为机制：授权不是一次立项决定，而会随着熟悉感和成功经验持续扩张。员工在资料整理上获得稳定结果后，很容易把同一个助手接到邮箱、客户系统或发布工具；系统名称没有变，风险结构已经完全改变。更高的中断率也说明，成熟用户并非简单地“越来越放心”，而是在放宽执行的同时保持干预。

边界｜研究样本主要来自软件开发与特定产品，极端分位不能代表一般企业，更不能直接推断营销任务的准确率或损失概率。单轮时长也是自主性的近似指标。可迁移的结论只有一个：实际授权会随使用改变，因此静态的上线审批不足以覆盖后续风险。

管理动作｜为每个智能体建立授权履历，记录从只读到写入、从单步到多步、从草稿到外发、从人工批准到条件自主的每一次变化。升级必须引用近期成功率、人工推翻率、异常类型和回滚演练；熟悉感、使用人数和厂商版本更新都不能单独成为升级理由。

FIELD RECORD 04

第三章“个人能干活—企业能交责”必须通过八道门

“个人能干活—企业能交责”八道门任何一道门缺失，规模化都会把局部效率变成系统风险。

01 经营目标	02 可验证身份
03 业务上下文	04 最小权限
05 流程编排	06 证据留痕
07 持续评测	08 责任归属

八道门不是技术组件清单，而是企业把执行权交给智能体之前必须逐项成立的责任条件。

门	必须回答的问题	最低交付物
1 经营目标	任务必须连接明确经营目标、指标、期限和预算边界	目标卡、指标口径、优先级、停止条件
2 可验证身份	每次行动必须能证明由哪个人、智能体、服务身份或组织单元发起	身份映射、凭证生命周期、责任人
3 业务上下文	使用经过版本管理的品牌、产品、客户、政策和历史信息	知识来源、版本、有效期、适用范围
4 最小权限	只获得完成当前任务所需的最小数据和动作权限	白名单、范围、时限、额度、环境隔离
5 流程编排	任务以状态机运行，明确输入、输出、异常、交接和审批	流程图、状态字典、SLA、异常路径
6 证据留痕	保留来源、推理依据摘要、调用、修改、审批、发布和结果	事件日志、版本、证据包、审计查询
7 持续评测	上线前后持续评估正确性、稳定性、风险和业务结果	离线集、在线监控、红队、漂移阈值
8 责任归属	每个任务、权限、审批和经营结果都对应可执行责任人	RACI、事故等级、问责与改进机制

3.1 八道门的通过逻辑

八道门应按“目标—主体—依据—能力—过程—证据—质量—责任”顺序检查。前一项未成立，后一项的成熟度无法弥补。例如，权限控制再精细，也不能修复错误经营目标；日志再完整，也不能替代责任归属。

试点阶段不要求所有能力达到集团级成熟度，但每道门必须有最小可用机制。企业可以在小范围内使用人工表单、人工审批和简化日志，但不能完全缺失。

升级自主等级时，应重新逐门评审。新增连接器、扩大数据范围、延长流程跨度、增加自动发布或提升预算额度，都视为重新授权。

评审状态	定义	处理
通过	有明确机制、责任人、证据与运行记录	可进入下一阶段
条件通过	风险可控，但依赖临时人工补偿	限定期限整改，不得扩大范围
不通过	关键条件缺失或无法验证	停止上线或降级为建议/草拟模式

3.2 八道门的董事会问法

董事会不需要讨论底层技术细节，但必须要求管理层用经营语言回答：智能体要完成什么目标、以谁的身份、依据什么信息、能碰哪些系统、在哪些节点停下来、出了问题如何找到证据、能力如何持续验证、最终由谁承担责任。

当回答依赖“通常不会”“大概可以”“模型应该知道”或“供应商会处理”时，说明企业尚未通过相应门槛。

- 目标是否可以在季度经营会上复盘?
- 权限是否可以在人员调岗或项目结束时自动回收?
- 证据是否可以在不依赖供应商的情况下导出?
- 事故是否有明确的停机权和升级路径?

3.3 八道门要接受任务级考试

实证 | EnterpriseBench 在一个模拟企业环境中设置 500 项软件、财务、人力和行政任务，刻意保留多来源数据、权限层级和跨职能流程。2025 年论文报告，即使最强的受测智能体，完整任务完成率也只有 41.8%。另一项 TheAgentCompany 基准让智能体操作公司网站、编写程序并与同事沟通，最有竞争力的基线也只自主完成约 24% 的任务。[S33][S34]

两项基准的共同价值，不在于给所有产品排出永久名次，而在于还原了企业任务的复合难度：取到正确数据只是第一步，系统还要理解权限、保持跨步骤状态、协调其他角色并验证最终结果。单项工具调用成功率很高，仍可能在完整任务上失败。

詹森与麦克林的委托代理理论指出，委托会同时产生监督成本、约束成本和剩余损失；企业不能仅以代理人“愿意执行”判断委托是否有效。[S35] 执行型智能体把代理速度提高了，但没有消除目标偏差、信息不对称和结果归责。八道门的作用，正是把这些代理成本转化为可检查的控制。

边界｜基准环境并非真实企业，也不会覆盖每种营销系统、组织习惯和数据质量；模型能力还在快速变化。41.8% 或 24% 不能直接作为采购验收线。它们足以否定另一种做法：拿几个演示任务的成功，替代生产流程中的端到端验证。

管理动作｜八道门的验收对象必须是“任务×环境×权限”组合，而不是模型名称。每个候选场景至少准备正常、信息冲突、权限不足、外部系统失败和高影响异常五类测试，并同时计算业务完成率、证据完整率和安全失败率。

FIELD RECORD 05

第四章 企业化需要完成三次跨越，而不是增加更多智能体

“个人能干活—企业能交责”八道门任何一道门缺失，规模化都会把局部效率变成系统风险。

01 经营目标	02 可验证身份
03 业务上下文	04 最小权限
05 流程编排	06 证据留痕
07 持续评测	08 责任归属

企业生产力的形成依赖三次跨越：个人到团队、单任务到端到端流程、工具到经营系统。

结论

企业生产力的形成依赖三次跨越：个人到团队、单任务到端到端流程、工具到经营系统。

证据

个人工具的输出通常停留在文件或对话；团队生产力要求共享对象和状态；经营系统进一步要求与预算、客户、收入、风险和组织节奏连接。

经营含义

只增加Agent数量，会把个人层面的不确定性放大为协作层面的复杂性。

行动

每完成一次跨越，企业都要新增公共能力，并设置独立评审门。

4.1 第一跨越：个人到团队

团队生产力的关键不是让更多人使用同一个工具，而是建立共同目标、统一对象、共享状态、版本规则、角色分工和交接机制。

企业应明确哪些信息属于个人偏好，哪些属于组织知识；哪些动作可以在个人工作区完成，哪些必须进入团队工单或业务系统。

团队层的最小能力包括统一身份、角色权限、公共知识库、任务队列、审批队列、共享日志和运营看板。

新增公共能力	作用	最低标准
统一身份	区分人、智能体、服务账号与外部系统	唯一标识、责任人、凭证可回收
共享业务对象	避免状态散落在聊天中	内容、活动、客户、线索、任务有唯一ID
协作与交接	明确谁在何时接手	负责人、截止时间、SLA、升级路径
统一证据	支持复盘和审计	来源、版本、审批、执行、结果可串联

4.2 第二跨越：单任务到端到端流程

单任务优化往往只改善局部速度。端到端流程需要处理上游输入质量、跨步骤状态、异常分支、人工审批、外部系统失败和下游结果回收。

企业应先把流程画成业务状态机，再决定哪些节点使用模型、规则、人工或外部服务。智能体是流程中的执行角色，不应成为流程本身。

端到端流程的合格标准，是任何时点都能回答“当前处于什么状态、下一步由谁执行、需要什么证据、失败后如何恢复”。

4.3 第三跨越：工具到经营系统

经营系统将智能体动作与目标、预算、客户、渠道、收入、风险和组织节奏绑定。它不仅追踪做了什么，还解释为什么做、产生了什么结果、是否值得继续。

当智能体能够触达真实客户、品牌账号、广告预算和CRM时，企业必须引入经营级看板、成本归集、效果归因、风险阈值和季度复盘。

AI市场部的完成标志不是功能齐全，而是管理层能够像管理一个部门一样设置目标、分配权限、审阅关键决策、查看结果并调整资源。

4.4 多智能体的价值来自任务分解，而非角色数量

案例 | Kantar 的 People Team 需要为覆盖 60 个国家的 HR 咨询智能体整理分散、多语言、重复且格式不一的政策材料。团队没有让一个通用智能体包办全部工作，而是把定位文件、分类、翻译、去重、格式整理和法律政策核对拆成专门步骤；公开案例称，10 个智能体协助团队在六周内把 4,000 份材料整理为 400 份政策文档。[S36]

机制 | 这套做法有效的前提是每一步都有明确输入、输出和交接对象。定位智能体只负责找资料，翻译智能体处理语言统一，去重与格式步骤生成可审阅版本，领域人员决定政策含义。多智能体没有取消责任，而是把一个模糊的大任务改造成可以逐步验证的生产线。

迈克尔·哈默的流程再造观点提醒管理者：把旧流程原样自动化，往往只是更快地复制浪费。[S37] 企业从个人到团队的第一步，不应是为每个岗位复制一个机器人，而应先删除无价值交接，明确业务对象和验收条件，再决定哪些步骤交给规则、工作流、智能体或人。

边界 | 数据来自微软客户案例，没有披露人工投入、准确率、返工比例和长期维护成本；HR 政策整理也不同于高频外部营销。可复制的是任务分解、统一对象和领域审核结构，不是“10 个智能体”这个数量。

管理动作 | 为每次智能体交接建立四字段契约：交付对象、必需证据、验收规则、失败去向。新增一个角色前，必须证明它降低了某项等待、返工或专业判断成本；仅让流程图更复杂的角色应当删除。

FIELD RECORD 06

第五章 五级阶梯决定能力、权限与升级节奏

自主性不是一个开关由错误损失、可逆性、证据完整度和数据敏感度共同决定。

A 建议型	B 草拟型
C 执行前审批	D 条件自主
E 禁止自动化	

等级	核心能力	权限与数据	关键KPI	升级门槛
L1 个人副驾	检索、摘要、草拟、个人整理	无生产系统写入；个人范围数据	个人效率、采纳率、修订率	输出稳定且低风险
L2 专业执行者	按角色完成受限任务并调用低风险工具	服务身份；白名单；非关键数据	任务成功率、返工率、证据完整率	可重复、可审计、权限无漂移
L3 人工批准工作流	跨步骤执行，关键节点人工批准	受控写入；审批队列；回滚机制	端到端周期、审批时效、异常率	流程稳定、责任明确、回滚演练通过
L4 协同智能体团队	多个专业角色共享状态并协同	角色隔离；共享对象；协调策略	协同成功率、交接损耗、系统性错误率	协调价值大于复杂性成本
L5 受治理的经营系统	围绕目标持续运行、监控、复盘和优化	经营级权限、预算阈值、持续评测	收入/转化、单位经济性、风险调整收益	治理、价值和组织能力长期稳定

5.1 每一级的管理原则

L1强调个人判断，企业不应误把使用活跃度当作组织能力。L2开始要求服务身份、工具白名单和证据保留。L3把人工审批嵌入流程，是大多数营销生产场景的主力形态。

L4只适用于角色分工清晰、共享状态可靠、协调收益明确的复杂流程。L5则要求持续经营目标、预算控制、风险阈值和管理节奏，通常需要跨部门共同治理。

升级应以证据为依据，而不是以厂商版本或模型能力为依据。每次提升权限、数据敏感度、流程跨度或外部影响范围，都必须重新评估。

升级信号	降级或停止信号
连续多个周期达到质量、时效和证据门槛	错误率上升或无法解释
人工审批集中在少量可预测节点	人工频繁绕过流程或补救
异常可自动发现并在SLA内恢复	出现隐性失败、重复执行或状态漂移
业务指标改善且单位成本可控	仅增加生成量，经营结果无改善

5.2 五级阶梯与营销任务映射

营销任务的自主等级不应按部门统一设定。洞察摘要可能达到L2或L3；品牌发布、广告预算变更、客户承诺和危机回应通常长期停留在L3；部分低风险CRM字段更新可在条件满足时进入L4。

企业应为每个任务建立“自主等级卡”，至少包含任务目标、输入数据、外部影响、权限、审批人、证据要求、异常阈值和降级规则。

任务	建议等级	说明
竞品信息汇总	L2	允许自动采集与整理，结论进入人工复核
内容初稿与多平台改写	L2–L3	生成可自动，发布前审批
社媒定时发布	L3	需品牌审核、账号权限和撤回机制
广告优化建议	L2–L3	建议自动生成，预算和账户修改人工批准
CRM摘要与提醒	L2–L3	写入范围受限，客户触达需审批或规则约束
危机公关回应	L1–L3	仅建议与草拟，最终发布由授权高管与法务批准

5.3 能力等级不能由主观提效感决定

实证 | METR 在 2025 年对 16 名长期维护大型开源项目的开发者开展随机对照研究，共观察 246 项真实任务。参与者可以使用当时先进的 AI 工具时，完成任务反而平均多用 19% 的时间。研究团队在 2026 年更新中明确表示，后续工具很可能已有改善，但新的样本选择和计时问题使幅度难以可靠估计。[S38]

这项研究与营销工作并不等价，却揭示了一个普遍偏差：熟悉界面、快速生成中间产物和减少手工输入，会制造明显的“正在变快”感；验证、纠错、上下文解释和并行智能体管理的成本却容易被遗漏。自主等级若依据使用者感受升级，企业会把隐性返工当成生产率。

边界 | 样本只有 16 名高经验开发者，任务来自复杂开源代码库，结果不能外推到内容草拟、客户服务或销售跟进；19% 也只是 2025 年初工具的阶段性观察。它不证明 AI 普遍降低效率，只证明企业需要自己的任务级基线。

管理动作｜每次等级升级至少保留一个对照窗口：记录无人辅助、建议模式和拟升级模式下的全周期时间、返工、人工接管和业务结果。效率指标从需求进入开始，到结果被下游接受结束；生成第一稿的时间不能代表任务周期。

FIELD RECORD 07

第六章 自主性应由任务风险结构决定

自主性不是一个开关由错误损失、可逆性、证据完整度和数据敏感度共同决定。

A 建议型	B 草拟型
C 执行前审批	D 条件自主
E 禁止自动化	

同一智能体在不同任务上应拥有不同自主等级；不存在“一次授权、全局自主”。

维度	判断问题	对自主性的影响
业务价值	任务对增长、效率、客户体验或风险控制的贡献	高价值提高投入优先级，但不直接提高自主权
错误损失	错误造成的财务、品牌、客户、法律或安全损失	越高越需要审批、限额和双人复核
可逆性	错误发生后能否快速撤回、回滚或补偿	越难逆转，自主等级越低
证据完整度	输入来源、判断依据、过程日志和结果能否验证	证据越完整，越可能采用条件自主
数据敏感度	是否涉及个人信息、商业秘密、账号凭证、价格和合同	越敏感越需要隔离、脱敏和严格授权

6.1 五种自主策略

建议型：智能体只提供分析和建议，不生成可直接执行的最终产物。适用于战略判断、危机决策和高不确定任务。

草拟型：智能体生成草稿、方案或操作包，由人完成关键判断和最终提交。适用于内容、广告、客户沟通和报告。

执行前审批：智能体完成准备、校验和模拟，进入审批队列后才调用生产工具。适用于发布、预算、客户触达和关键数据写入。

条件自主：在白名单、额度、时间窗、数据范围和异常阈值内自动执行，越界立即暂停。适用于稳定、可逆、证据充分的重复任务。

禁止自动化：即使技术可行，也因责任、伦理、法律或不可逆风险不允许自动执行。

策略	典型条件	营销例子
建议型	高不确定、高损失、证据不足	品牌战略、重大危机定调
草拟型	需要专业判断与品牌把关	新闻稿、客户方案、价格解释
执行前审批	动作可执行但外部影响明显	社媒发布、广告预算调整
条件自主	规则稳定、低损失、可回滚	低风险字段更新、日报分发
禁止自动化	不可逆或责任不能委托	虚假承诺、违规采集、未经授权对外声明

6.2 自主性评分卡

评分卡用于比较任务，不应机械地生成单一分数。企业至少要记录每个维度的理由、证据和责任人，并由业务、IT、安全与法务共同确认。

当错误损失或数据敏感度达到高等级时，即使业务价值很高、可逆性较好，也应保留人工批准。证据完整度不足时，不得采用条件自主。

维度	低	中	高
业务价值	辅助性、低频	重要但非关键	直接影响收入、客户或重大风险
错误损失	内部返工即可修复	影响单次活动或少量客户	重大财务、品牌、法律或安全后果
可逆性	可立即撤回	可补偿但成本较高	难以撤回或不可逆
证据完整度	来源和过程不完整	关键环节可追踪	全链路可验证、可导出
数据敏感度	公开或低敏	内部经营数据	个人信息、商业秘密、凭证或合同

6.3 客户入口获得执行权后，架构必须围绕结果重建

案例｜Microsoft.com 的 Ask Microsoft 最初用于回答产品、价格和支持问题，随后扩展到产品评估、试用与注册。随着流量和知识源增加，原架构出现延迟，团队改为由多个领域子智能体分别处理 Microsoft 365、Azure、价格和试用等知识，再由总入口编排。2026 年公开案例称，Microsoft 365 站点测试的响应延迟最高下降 61%，Azure 站点试用发起增加 16%，人工聊天升级最高减少 70%；使用该入口的访客注册服务概率为其他访客的 10 倍。[S39]

机制｜关键改变不是“回答更聪明”，而是把客户意图接到可观测结果：产品信息由特定知识域负责，复杂问题可以升级人工，试用流程有明确转化动作，离开试用申请的访客还可进入邮件跟进。智能体由此成为客户旅程中的一个执行节点，而不是漂浮在网站上的问答框。

边界｜案例由微软发布，平台供应商同时是使用者，测试方法、基线、成本和客户构成没有完整公开。“10 倍更可能注册”是相关关系，不等于智能体单独造成 10 倍增量；更积极的访客本来就可能更愿意使用问答入口。

管理动作 | 外部营销智能体上线前，要把每类问题映射到权威来源、允许动作、升级对象和结果事件。至少按访客意图报告答案解决、人工升级、试用发起、有效线索和后续转化，并设置延迟、错误承诺和重复触达的护栏。

FIELD RECORD 08

第七章 AI市场部的最小可行架构由九个运行层组成

最小可行运行架构不是堆叠更多 Agent，而是把目标、状态、行动与结果接成闭环。

经营目标	**业务状态**
角色与技能	**知识与记忆**
连接与编排	**权限与审批**
监控与评测	**经营结果**

运行层	核心职责	最低产物
1 目标层	年度/季度目标、活动目标、指标、预算、优先级、停止条件	目标卡、KR、预算边界
2 业务状态	内容、活动、客户、线索、广告、任务的唯一对象与状态	状态机、对象ID、事件流
3 角色与技能	人类角色、智能体角色、技能包、责任边界	岗位卡、技能目录、RACI
4 知识与记忆	品牌、产品、行业、客户、历史经验与个人偏好	版本、来源、有效期、访问范围
5 连接器	内容平台、CRM、广告、协作、文档、数据和身份系统	认证、限流、白名单、健康检查
6 流程编排	步骤、分支、并行、重试、补偿、人工节点与SLA	工作流、状态转换、异常路径
7 审批与权限	最小权限、四眼原则、额度、时间窗、环境隔离	策略库、审批队列、凭证管理
8 监控评测	质量、稳定性、安全、成本、漂移与业务指标	在线监控、离线集、告警、红队
9 经营结果	速度、质量、转化、收入、成本、风险和组织学习	经营看板、复盘、资源调整

7.1 业务状态是企业协同的中心

聊天记录不是可靠的业务状态。企业必须为内容、活动、客户、线索、广告计划和任务建立结构化对象，记录当前状态、负责人、输入、输出、审批和异常。

多智能体协同的基础不是“互相对话”，而是对同一业务对象进行受控读写。任何角色都应通过状态变化进行协作，而不是依赖隐含上下文。

状态机还承担审计与恢复功能：当任务失败时，系统能够判断已完成步骤、待处理步骤和可执行补偿。

对象	关键状态示例	必须保留的证据
内容	选题—Brief—草稿—审核—排期—发布—复盘	来源、版本、审核、发布链接、表现
活动	立项—策略—资产—执行—复盘—沉淀	预算、责任人、审批、渠道结果
线索	新建—评分—分配—跟进—商机—成交/关闭	来源、触达、负责人、阶段变更
广告	建议—模拟—审批—执行—监控—复盘	账户快照、建议、审批、变更、效果

7.2 知识、记忆与状态必须分开治理

知识是可复用、可引用、可版本管理的组织资产；记忆是对个人或团队偏好的连续记录；业务状态是当前任务和对象的真实进展。三者混在同一会话中，会导致版本冲突、权限泄露和状态失真。

企业应为知识设置来源、所有者、有效期和访问范围；为记忆设置保存范围和删除策略；为状态设置唯一对象、状态转换规则和责任人。

类型	保存内容	主要风险	治理要点
组织知识	品牌规范、产品资料、政策、案例、方法	过期、冲突、越权访问	版本、有效期、来源、审批
个人/团队记忆	偏好、历史协作、上下文摘要	过度保存、跨人泄露	范围、期限、可查看、可删除
业务状态	任务阶段、输入输出、审批、异常	状态漂移、重复执行	唯一ID、事件日志、幂等与补偿
证据档案	来源、调用、修改、发布、结果	不完整或不可导出	连续链路、只读保存、审计查询

7.3 连接器必须被视为生产基础设施

连接器不只是API封装，而是智能体进入真实业务世界的边界。连接器需要明确认证方式、调用主体、权限范围、速率限制、失败重试、幂等性、日志和停机能力。

采购验收时，应要求供应商展示连接器在凭证过期、权限不足、接口超时、重复请求、数据结构变化和外部系统故障时的行为。

生产环境不应使用个人长期密钥或共享账号。服务身份应有明确所有者、期限、轮换和回收机制。

7.4 数据治理决定智能体能否成为团队能力

案例 | Regal Rexnord 是拥有 3 万多名员工的工业制造企业。其最初在另一平台开发的政策问答智能体，因为知识接入有限，只能回答约 40% 的员工问题；迁移并连接 SharePoint 等知识源后，企业级 RRX GPT 每月处理超过 2,000 次查询，公司估算每年节省 2,400 小时。面向客户的 RRXy 每周服务超过 1,000 名用户，满意度保持在 80% 以上，并能把复杂问题连同上下文交给人工。[S40]

团队还把难以直接读取的多标签 Excel 数据抽取为结构化 CSV，存入统一存储并建立搜索索引；500 多名客服人员使用另一智能体检索产品与流程信息，企业估算生产率改善约 12%。这里的性能提升首先来自数据处理和知识连接，而不是更长的提示语。

杰伊·加尔布雷思把组织视为信息处理系统：任务不确定性和相互依赖越高，组织越需要增加横向信息能力。[S41] 智能体架构同样如此。知识、状态和权限分散在不同系统时，模型只能靠猜测填补缺口；把对象、来源和交接结构化，才会形成可复用的团队能力。

边界 | 结果来自微软与客户联合案例，没有独立对照或全成本披露；节省小时与生产率均为企业估算。Regal Rexnord 还设置了包含审计、法务和业务的治理委员会，并隔离 HR、法律、财务环境，这些配套条件不能从结果中删除。

管理动作 | 架构评审先检查数据加工链：原始来源、结构化方式、版本责任、访问范围、刷新频率、答案引用和人工转交。若智能体表现改善只能靠持续扩大上下文或人工暗中补资料，问题在信息系统，不在模型参数。

FIELD RECORD 09

第八章 人机协同的核心是分权、升级与责任设计

角色	核心责任	关键权力
董事会/经营层	确定风险偏好、投资边界、重大停止条件	不直接审批日常任务，但要求季度价值与风险复盘
CMO/市场负责人	拥有业务目标、场景优先级、品牌标准和最终经营结果	批准场景、关键内容、预算与客户影响动作
业务部门	提供真实流程、数据口径和结果反馈	承担场景效果与一线执行责任
IT/数字化	平台架构、身份、连接器、数据集成、可用性	保证技术底座、环境隔离和运行稳定
安全/风控	权限策略、监控、事故响应、红队与审计	拥有高风险暂停权和风险验收权
法务/合规	个人信息、广告、知识产权、合同与监管边界	批准高风险规则和对外承诺
供应商	按合同提供产品、实施、支持与证据	不得替代企业承担经营与法定责任

8.1 CMO是业务所有者，IT是平台所有者

CMO负责回答“为什么做、做什么、如何衡量”；IT负责回答“如何连接、如何运行、如何恢复”。安全与法务分别对风险与合规拥有独立意见。

当业务把所有责任推给IT，智能体容易成为技术试验；当IT被排除在外，业务容易出现影子AI、共享账号和不可审计连接。

最有效的治理机制是业务牵头的跨职能委员会，设置固定决策节奏和清晰升级路径，而非临时项目群。

事项	主责	会签/参与
场景价值与优先级	CMO/业务负责人	财务、IT、安全
数据与系统接入	IT/数据负责人	业务、安全、法务
自主等级与审批策略	业务负责人	安全、法务、IT
上线验收	业务+IT共同	安全、法务、供应商
事故暂停与恢复	安全/业务共同	IT、法务、管理层
季度价值复盘	CMO/经营层	财务、IT、安全

8.2 人工审批要被产品化

人工审批不能依赖群聊中一句“可以”。合格的审批应展示任务目标、关键输入、拟执行动作、影响范围、证据摘要、风险提示、回滚方式和有效期限。

审批人应拥有真正的拒绝权和修改权；审批结果应绑定具体版本，后续内容或参数变化应触发重新审批。

企业应跟踪审批等待时间、拒绝原因和重复修改点，用于优化流程，而不是简单减少人工节点。

审批类型	适用场景	最低信息
业务审批	品牌主张、活动策略、客户沟通	目标、受众、版本、责任人
预算审批	广告预算、优惠、采购	金额、账户、期限、预期、上限
合规审批	个人信息、广告、版权、危机	依据、风险、适用范围、留存
技术审批	生产连接、权限扩大、数据写入	身份、权限、环境、回滚、监控

8.3 人工接管是一项产能，而不是一个按钮

案例 | 英国业务流程服务商 Capita 拥有约 3.4 万名员工。公司先从个人生产力工具起步，再进入邮件分流、消防风险评估和车队路线等执行场景；公开案例记录，三个月内已发生 7 万次智能体交互，其中邮件智能体每天分流数千封咨询，使响应时间下降 60%。公司同时建立智能体网络，并强调每个交接都受明确责任和监督约束。[S42]

机制 | 邮件分流把高频、规则较清楚的识别和路由交给系统，让人工把时间放在同理心、例外判断和解决问题上。但自动化越稳定，人工越少接触常规案例，接到的工作便越复杂；如果转交只给出一段摘要，而没有原始邮件、已做动作、风险信号和可用权限，人工反而更难恢复情境。

利泽恩·班布里奇在“自动化的反讽”中指出，自动化接管常规任务后，人类留下的往往是最难、最少发生、也最缺乏练习的异常；系统仍要求人保持理解并迅速接管。[S43] 对客户触点而言，人工在环不能只计算审批人数，还要设计异常工作台、排班、权限和训练。

边界 | Capita 数据来自供应商客户故事，未披露质量误差、客户满意度、人工总量和实施成本，不能据此推导固定 ROI。邮件分流的风险也低于合同承诺、价格变更或危机沟通。可复制的是“按任务分权并准备接管”的结构。

管理动作 | 为每类异常定义接管包：客户原始请求、权威资料、系统已采取动作、未完成步骤、风险原因、可选处置和最晚响应时间。每季度抽演一次高影响长尾事件，验证人是否真的能在目标时限内理解、停止、修复和通知。

FIELD RECORD 10

第九章 事故治理必须覆盖预防、发现、暂停、恢复和追责

事故通常沿同一条链放大

目标含混 → 上下文过期 → 权限过宽 → 动作不可逆 → 证据缺失 → 无人负责

治理的任务，是在链条前端设置边界，在中段建立监控，在末端保留暂停、回滚与问责。

风险	典型表现	首要控制
越权操作	使用超出任务需要的系统、数据或动作	最小权限、白名单、时限、额度、环境隔离
错误发布	内容、对象、时间、渠道或版本错误	版本锁定、预览、四眼审批、撤回机制
数据泄露	敏感数据进入错误环境、人员或外部服务	分级分类、脱敏、访问控制、出境与留存规则
品牌风险	不当表达、虚假承诺、敏感议题误判	品牌规则、禁区、事实核验、高风险人工批准
重复执行	重试或状态异常导致重复发送、重复扣费	幂等键、状态锁、去重、补偿事务
隐性失败	流程表面完成但结果未落地或对象错误	结果回读、对账、健康检查、异常告警
证据缺失	无法还原输入、审批、调用和结果	不可变日志、版本、时间戳、证据导出
责任悬空	供应商、业务、IT之间互相推诿	RACI、事故等级、停机权、合同责任

9.1 四道防线

第一道防线是设计控制：任务范围、数据范围、工具白名单、预算限额、时间窗和审批策略。

第二道防线是运行监控：异常调用、权限变化、内容风险、重复执行、成本突增和结果对账。

第三道防线是快速暂停：业务负责人、安全负责人和平台运维都应有明确停机路径，关键连接器应支持单独隔离。

第四道防线是事故复盘：保全证据、界定影响、恢复服务、通知相关方、修复控制、验证整改并更新评测集。

阶段	关键动作	责任人
预防	权限、审批、沙箱、测试、培训	业务+IT+安全

阶段	关键动作	责任人
发现	监控、告警、对账、用户反馈	运营+安全
暂停	隔离连接器、冻结任务、撤销凭证	安全/IT，业务会签
恢复	回滚、补偿、验证、分阶段恢复	IT+业务
追责与改进	取证、根因、责任、整改、复测	管理层+法务+安全

9.2 必须预设的停止扩张信号

当出现权限范围持续扩大但缺少正式复审、人工通过群聊绕过审批、日志无法串联、同类事故重复发生、业务指标无改善却持续增加调用量、供应商无法导出证据或企业无法独立停机时，应立即停止扩张。

停止扩张不等于停止所有使用。企业应将场景降级到建议型或草拟型，收回生产权限，完成根因整改和重新验收后再恢复。

- 发生重大越权、泄露或错误发布。
- 证据链不完整，无法确认影响范围。
- 权限、知识或流程版本发生漂移。
- 人工审批被形式化或批量“秒过”。
- 单位经济性持续恶化。
- 关键责任人离岗后无人接管。

9.3 提示注入会把外部资料变成动作指令

实证 | NIST 的 AI 安全团队使用 AgentDojo 测试间接提示注入：智能体在处理邮件、文件、旅行和银行等模拟任务时，会遇到伪装成普通数据的恶意指令。五类注入任务单次测试的平均攻击成功率为 57%；当每类攻击重复尝试 25 次时，平均成功率升至 80%。研究还发现，为一个环境设计的新攻击可以迁移到其他环境。[S44]

OpenClaw 官方安全文档对风险的描述十分直接：获得工具和渠道权限后，助手可以执行命令、读写文件、访问网络和发送消息；官方沙箱默认关闭，即使启用也被明确说明并非完美安全边界。[S30][S45] 营销智能体读取网页、邮件、附件和客户输入时，外部数据与控制指令处在同一语言界面，传统“不要相信输入”原则由此升级为“不要让输入直接决定动作”。

詹姆斯·里森关于组织事故的研究强调，严重事故往往由多层防线中的缺口在特定时刻对齐，而非单一人员失误。[S46] 对智能体而言，模型拒绝、输入过滤、工具白名单、最小权限、审批、速率限制、异常监控和可回滚动作需要共同工作；任何一层都不应被描述为绝对防护。

边界｜NIST 结果来自模拟环境和特定模型，不代表任何生产系统都有 80% 被攻破概率；重复攻击设置用于揭示概率系统的累积风险。企业仍需用自己的工具、数据和权限验证。报告引用这些数字，是为了说明一次通过的红队测试远远不够。

管理动作｜把来自网页、邮件、共享文档、CRM 自由文本和第三方插件的内容统一标记为不可信数据；高影响动作只接受结构化、经过策略校验的参数。安全测试必须覆盖重复攻击、跨渠道攻击、权限升级、数据外传和清除痕迹，并验证监控能否及时发现。

FIELD RECORD 11

第十章 营销场景应按价值、可行性和风险组成投资组合

事故通常沿同一条链放大

目标含混 → 上下文过期 → 权限过宽 → 动作不可逆 → 证据缺失 → 无人负责

治理的任务，是在链条前端设置边界，在中段建立监控，在末端保留暂停、回滚与问责。

场景	价值	可行性	风险	数据要求	建议等级	优先级
洞察研究	中高	高	低中	公开/内部资料	L2–L3	首批
内容策划与草拟	高	高	中	品牌、产品、历史内容	L2–L3	首批
社媒排期与发布	高	中高	中高	账号、内容、日历	L3	第二批
广告分析与建议	高	中高	中	账户与效果数据	L2–L3	首批
广告预算/参数执行	高	中	高	账户、预算、策略	L3	谨慎
CRM摘要与线索分配	高	高	中	客户、线索、规则	L2–L3	首批
客户自动触达	高	中	高	客户授权、旅程、内容	L3–L4	条件试点
销售协同与话术	中高	高	中	产品、客户、阶段	L2–L3	首批
会议纪要与行动项	中	高	低中	会议、任务、组织	L2–L3	首批
复盘与知识沉淀	高	高	低中	过程、结果、证据	L2–L3	首批

10.1 首批试点应满足六个条件

频率足够高，能够在14—30天内产生多轮运行数据；任务边界清晰，输入和输出可定义；错误可逆，影响范围可限制；关键证据可获取；业务负责人愿意投入时间；结果能够连接速度、质量、转化或收入指标。

不建议首批选择跨越多个部门、依赖大量历史脏数据、需要深度系统改造或直接触达大量外部客户的流程。试点的目标是验证运行机制，而不是证明技术无所不能。

优先选择	暂缓选择
竞品日报、内容Brief、初稿、复盘	重大危机自动回应
广告账户诊断与优化建议	自动修改大额预算
CRM摘要、线索评分与跟进提醒	未经授权的批量客户触达
会议纪要、行动项与周报	跨国敏感数据自动流转

10.2 场景组合要形成闭环

单点内容生成难以证明经营价值。更优的试点组合是选择一个小闭环，例如“洞察—内容—审批—发布—线索—复盘”，每个环节只纳入最小必要能力。

闭环中至少要有一个结果指标与经营目标连接，例如有效线索、商机推进、获客成本、内容转化、活动报名或客户响应。

试点组合不宜超过三个核心场景，避免同时引入过多连接器和责任人。

10.3 网站智能体只有接入线索系统才形成经营闭环

案例｜Salesforce 在 2025 年 2 月把 Agentforce 部署到官网，使用 2,000 多个网页和 400 条产品记录回答产品、价格与购买问题，并在访客希望联系销售时创建 Sales Cloud 合格线索。公司公开数据称，上线后处理超过 10 万次会话，覆盖 11 种语言，生成 3 万多条新线索，机会资格判断时间同比缩短 40%。[S47]

机制｜定价问题从结构化产品目录读取，开放性产品问题使用经过维护的网站内容；访客身份、地区和前序输入提供情境，转交销售时把会话与线索对象连在一起。这个设计把“回答一次问题”接到“形成可跟进对象”，也让营销、产品内容和销售团队共享同一事实源。

边界｜Salesforce 既是平台供应商也是案例主体，指标由公司自行披露；3 万条线索不等于 3 万个新增客户，时间同比变化也可能受流程、流量和团队调整影响。报告没有公开对照组、获客成本、线索质量和最终收入，因此不能把全部改善归因给智能体。

管理动作｜网站智能体的经营账必须沿漏斗连续记录：合格回答、人工升级、线索创建、重复合并、销售接受、商机推进、收入与投诉。CMO 同时检查源内容新鲜度和错误承诺；如果会话量增长而销售接受率下降，应先收紧入口和知识，而不是追求更多对话。

FIELD RECORD 12

第十一章 ROI必须衡量经营结果、风险调整与组织学习

事故通常沿同一条链放大

目标含混 → 上下文过期 → 权限过宽 → 动作不可逆 → 证据缺失 → 无人负责

治理的任务，是在链条前端设置边界，在中段建立监控，在末端保留暂停、回滚与问责。

指标维度	示例指标	管理目的
速度	从需求到初稿、从线索到首次响应、从异常到发现	缩短周期但不牺牲质量
质量	一次通过率、修订率、事实错误率、品牌一致性	减少返工和品牌偏差
转化	内容到线索、线索到商机、商机推进、客户响应	连接真实业务结果
收入与成本	增量收入、毛利、获客成本、单位任务成本、平台总成本	形成单位经济性
风险	越权率、事故率、证据完整率、恢复时间、审批绕过率	计算风险调整收益
组织学习	知识复用率、规则更新周期、重复错误率、员工采纳	形成可持续能力

11.1 不要把调用量当作价值

调用量、生成量、活跃用户和节省工时只说明系统被使用，无法证明业务结果。生成更多内容可能增加审核负担，自动化更多步骤也可能放大错误。

价值评估应采用基线对照：明确试点前的周期、成本、质量和转化，再比较试点后的变化。对于收入结果，应区分直接贡献、辅助贡献和不可归因部分。

财务测算应纳入模型调用、平台、集成、运维、数据治理、人工审核、培训和事故准备成本。

层级	只看活跃度的指标	应升级为
使用	调用次数、生成篇数	完成合格任务数、采纳率、返工率
效率	节省工时	端到端周期、等待时间、单位成本
效果	曝光、点击	有效线索、商机、转化、毛利贡献

层级	只看活跃度的指标	应升级为
治理	无事故	证据完整率、越权率、恢复时间、控制有效性

11.2 风险调整收益框架

高收益场景如果伴随不可接受的泄露、越权或品牌风险，不应被视为成功。企业可将预期收益减去平台与运营成本、人工审核成本、失败补偿成本和风险资本占用，形成风险调整收益。

风险成本不必追求精确财务模型，但必须进入决策。至少要记录事故概率、影响范围、恢复成本、客户补偿、法律与品牌后果。

当风险调整收益连续两个评审周期为负，或收益主要来自压缩必要审核，应降级或停止。

11.3 高增长案例必须接受归因审查

案例｜Asymbl 是 2023 年成立的劳动力编排公司。2025 年业务快速增长时，一名销售开发代表难以覆盖大量入站与出站线索；公司把客户数据、合同、评分和机会统一到 Salesforce，并让名为 Theodore 的销售智能体进行资格判断、会议安排、外呼培育和记录更新。供应商案例称，该系统每周处理超过 1,000 条线索，带来 150 万美元成本节省，并使潜客互动提高 427%。[S48]

机制｜案例中最重要的条件不是 24×7 运行，而是数字与人工 SDR 使用同一客户对象、评分、销售路径和预测模型。智能体处理覆盖与跟进，人类负责复杂关系和商机判断；统一记录减少了外部工具、表格和演示文档之间的状态断裂。

边界｜所有数字来自 Salesforce 客户故事，没有独立审计、对照组和详细成本口径。Asymbl 同期处于高速增长，并完成系统整合，增长、互动和节省不可能全部归于智能体；小型科技服务公司的销售结构也不能直接外推大型消费品牌。高倍数结果更应提高证据要求，而不是降低。

卡普兰与诺顿提出平衡计分卡，目的在于把学习能力、内部流程、客户结果和财务表现连接成可检验的因果链。[S49] 智能体 ROI 也应按同样顺序展开：团队是否掌握新能力，流程周期与质量是否改善，客户行为是否改变，最终是否形成风险调整后的利润。跨层跳跃的指标不能互相替代。

管理动作｜对任何高回报案例建立归因表，分开记录平台替换、流程清理、数据统一、市场增长、人员变化和智能体执行的贡献。试点扩张只引用企业自己的增量证据；供应商案例用于提出假设和设计指标，不作为预算承诺。

11.4 中国企业化路径：保留“小龙虾”的执行力，重建生产边界

实践｜中兴通讯 CDO 崔丽在 2026 年世界移动通信大会·上海公开表示，公司已在内部部署 Co-Claw，并针对 OpenClaw 的个人形态加强合规、安全、可靠性和内部系统集成；系统进入文档管理、合同分析、故障工单、事件创建和研发协作等高频任务。她同时提到，中兴在营销侧以“大企业中的敏捷小团队”推进流程重构和跨系统连接。[S50]

这段实践提供了中国企业从个人工具走向生产系统的清晰方向：保留自然语言入口和持续执行能力，但把身份、知识、权限、系统连接与运行责任重新接回企业。硬件或本地部署可以改善数据控制和成本可见性，却不会自动解决错误目标、越权动作、知识过期和跨部门问责。

边界｜中兴披露了部署范围和建设方向，没有公布任务准确率、成本、业务收益和事故数据，因而只能视为企业化实践片段，不能计为量化成功案例。其电信与制造背景、既有数字化基础，也决定了其他企业不能照搬系统规模。

理查德·鲁梅尔特把好战略的内核概括为诊断、指导方针和连贯行动。[S51] “建设 AI 市场部”若只是一串采购与上线日期，就不是战略。企业应先诊断一个真正限制增长或客户体验的瓶颈，再确定授权原则，最后让数据、流程、组织、技术和风险动作围绕同一瓶颈推进。

管理动作｜把 365 天路线改写为四次经营评审：是否解决了原瓶颈、公共能力是否可复用、风险是否受控、继续投入是否优于其他选择。新增场景只有在前一阶段形成可审计证据后进入；无法证明价值或控制成本的场景按计划退役。

FIELD RECORD 13

第十二章 未来12—18个月应按“先闭环、后扩张”推进

阶段	目标	负责人	交付物	评审门	退出条件
14天	确定一个高频、可逆、证据充分的场景；完成八道门最小检查	CMO/业务负责人	场景卡、基线、RACI、风险清单、试点流程	目标与责任清楚；无生产越权	无法定义结果或责任人缺位
30天	跑通人工批准工作流，形成首轮数据	业务+IT	工作流、审批、日志、周报、问题清单	连续运行、证据完整、异常可恢复	审批绕过、状态不一致、结果不可验证
90天	形成小闭环并验证业务价值	CMO+财务+IT	经营看板、单位经济性、评测集、控制测试	指标改善且风险可控	价值无改善或事故重复
180天	扩展到相邻场景和第二连接器	跨职能委员会	平台化身份、知识、监控、供应商SLA	公共能力可复用	每个场景仍靠定制救火
365天	建立受治理的AI市场部运行体系	经营层	年度目标、预算、岗位、审计、季度复盘、扩张计划	成为稳定经营能力	治理成本高于价值或控制失效

12.1 14天：定义可验证的最小试点

选择一个业务负责人真正关心的高频任务，确认基线、目标、输入、输出、错误边界、审批人和停止条件。

仅接入最小必要数据和工具。所有生产动作默认人工批准，先验证状态、证据和责任链。

完成一次桌面演练：错误输入、权限不足、外部系统失败、重复请求和紧急停机。

- 交付物：场景卡、流程图、数据清单、权限清单、RACI、基线与指标。
- 评审门：任何人都能说明当前状态、下一步、证据与责任人。
- 退出条件：无法获得可靠数据或业务负责人无法投入。

12.2 30天：把演示变成可重复运行

在真实但受限的业务环境中运行多轮，记录每次输入、输出、审批、异常和结果。优化重点是稳定性、状态一致性、审批体验和证据完整，而不是追求更多功能。

建立每周运营会，CMO查看业务结果，IT查看运行健康，安全查看风险事件，流程负责人处理异常和知识更新。

- 交付物：运行日志、周报、异常清单、审批SLA、质量评测。

- 评审门：连续运行，错误可发现、可暂停、可恢复。
- 退出条件：频繁依赖人工隐藏补救，系统表面成功但结果无法对账。

12.3 90天：验证端到端价值

将试点扩展为小闭环，至少连接一个上游输入、一个执行动作和一个结果回收。建立基线对照和单位经济性。

正式确定哪些节点可进入条件自主，哪些长期保留人工批准。所有提升自主性的决定都要基于历史运行证据。

- 交付物：经营看板、评测集、控制测试、价值复盘、升级建议。
- 评审门：速度、质量、转化或成本至少一项显著改善，风险指标不恶化。
- 退出条件：价值无法归因、风险持续上升、责任人无法承担。

12.4 180天：建设公共能力而非复制烟囱

扩展到相邻场景时，应复用身份、知识、业务对象、审批、日志和监控，避免为每个场景重新搭建独立机器人。

建立供应商管理、版本变更、权限复审、知识治理、红队测试和事故演练制度。

- 交付物：平台化能力目录、统一策略库、服务SLA、季度审计。
- 评审门：新增场景的边际实施成本下降，公共能力复用率上升。
- 退出条件：每次扩展仍需要大量人工定制，或治理能力跟不上连接器数量。

12.5 365天：形成受治理的AI市场部

将AI市场部纳入年度经营计划、预算、岗位、绩效与风险管理。明确哪些能力自建、采购或外包，保留企业对身份、数据、证据和停机的控制权。

每季度复盘场景组合，淘汰低价值流程，提升高价值稳定流程的自主等级，并将事故与成功经验写入知识、规则和评测集。

- 交付物：年度目标、预算、组织架构、场景组合、审计报告、扩张计划。
- 评审门：价值、风险和组织能力三者稳定。
- 退出条件：治理成本持续高于价值，或关键能力被供应商锁定且不可迁移。

管理者工具一：董事会决策清单

主题	董事会必须问的问题
战略	该场景连接哪个经营目标？不做的机会成本是什么？
价值	基线是什么？90天内看哪些业务指标？
边界	哪些动作永远需要人工批准？
责任	业务、IT、安全、法务和供应商各自负责什么？
数据	使用哪些数据？是否包含个人信息、商业秘密或凭证？
权限	智能体以什么身份执行？权限如何授予、复审和回收？
证据	能否导出来源、版本、审批、调用、发布和结果？
安全	谁有停机权？重大事故多久内发现和隔离？
财务	总成本包括哪些隐性成本？风险调整收益是否为正？
扩张	出现哪些信号才能提高自主等级？哪些信号必须停止？

董事会无需审批每个技术选择，但应要求管理层按季度提交价值、风险、能力与供应商依赖四张表，并对重大权限扩大、敏感数据接入和条件自主升级进行审议。

管理者工具二：CMO场景立项卡

字段	填写要求
经营目标	明确收入、线索、转化、速度、质量或风险目标
任务边界	起点、终点、输入、输出、频率、受众与渠道
当前基线	周期、人工成本、质量、转化、事故与返工
自主策略	建议型、草拟型、执行前审批、条件自主或禁止
数据与知识	来源、所有者、版本、敏感等级、访问范围
权限	工具、对象、动作、额度、时间窗、环境
审批	节点、审批人、SLA、拒绝与重提规则
证据	必须保留的来源、版本、调用、审批和结果
异常	暂停、重试、回滚、补偿、升级与通知
成功/停止条件	明确14/30/90天评审门和退出条件

管理者工具三：企业采购验收清单

验收域	关键检查
身份与权限	是否支持人、智能体、服务账号分离；凭证可轮换、到期和回收
知识治理	是否支持来源、版本、有效期、权限和冲突处理
业务状态	是否有结构化对象、唯一ID、状态机和事件日志
连接器	是否支持幂等、限流、错误处理、回读与健康检查
审批	是否可按任务、金额、渠道、数据和风险配置
日志与证据	是否可导出完整证据链，且企业可独立留存
评测	是否支持离线集、在线监控、质量与安全指标
事故处置	是否支持单任务、单连接器、单身份和全局停机
可迁移性	数据、配置、知识、日志和流程能否导出
服务边界	SLA、版本变更、漏洞响应、分包商与数据处理责任是否明确
成本	平台、调用、集成、运维、审核、培训和退出成本是否透明
验收	是否以真实业务流程和异常演练验收，而非功能演示

采购否决项

- 无法区分调用身份；生产权限依赖个人账号；证据不可导出；无法单独停用连接器；数据与知识无法迁移；重大版本变更缺少通知与回退。

管理者工具四：上线前控制测试

测试项	测试方式	通过标准
目标偏差	输入一个看似合理但不符合经营目标的请求	系统拒绝或要求确认目标
知识冲突	提供两个不同版本的品牌规则	优先使用有效版本并提示冲突
权限不足	请求访问未授权客户或账户	明确拒绝并记录
重复执行	重复提交同一发布或写入请求	幂等处理，不重复执行
外部故障	模拟接口超时或返回异常	进入重试/暂停/人工处理，不误报成功
审批变更	审批后修改内容或参数	触发重新审批
敏感数据	输入超范围个人信息或凭证	阻断、脱敏或升级人工

测试项	测试方式	通过标准
紧急停机	触发高风险告警	在规定时间内冻结任务与凭证
证据导出	抽取一个完整任务	可还原来源、版本、审批、调用、结果
恢复演练	从失败状态恢复	状态一致，不重复、不丢失

管理者工具五：季度价值与风险评分卡

维度	指标	本期	基线	趋势	管理动作
经营价值	有效线索/商机/收入贡献				
效率	端到端周期、单位任务成本				
质量	一次通过率、事实错误率、返工率				
风险	越权、泄露、错误发布、重复执行				
证据	证据完整率、审计抽样通过率				
运营	可用性、失败恢复时间、审批SLA				
组织	采纳率、知识复用率、重复错误率				
供应商	SLA、版本稳定性、可迁移性				

评分卡应由CMO牵头，财务、IT、安全和法务共同审阅。任何自主等级提升都应引用最近一个季度的运行证据。

FIELD RECORD 14

附录A 单智能体、工作流、多智能体与AI市场部的区别

最小可行运行架构不是堆叠更多 Agent，而是把目标、状态、行动与结果接成闭环。

经营目标	**业务状态**
角色与技能	**知识与记忆**
连接与编排	**权限与审批**
监控与评测	**经营结果**

形态	核心特征	适用问题	主要风险	治理重点
单智能体	一个角色理解任务并调用工具	边界清晰的专业任务	上下文过载、权限过宽	身份、知识、工具白名单
工作流	固定或半固定步骤与状态转换	可重复的跨步骤流程	状态错误、异常路径缺失	状态机、审批、幂等、补偿
多智能体	多个角色共享对象并分工协同	复杂、并行、专业分工明显的任务	协调成本、错误传播、责任模糊	共享状态、角色边界、协调策略
AI市场部	目标、业务状态、角色技能、知识、连接、治理、评测与结果闭环	持续经营与组织协同	系统性风险与治理成本	经营目标、权限、证据、责任、复盘

优先选择最简单、可审计的形态。只有当复杂性带来的业务价值明显高于协调和治理成本时，才采用多智能体。

FIELD RECORD 15

附录B 任务自主等级模板

自主性不是一个开关由错误损失、可逆性、证据完整度和数据敏感度共同决定。

A **建议型**	B **草拟型**
C **执行前审批**	D **条件自主**
E **禁止自动化**	

字段	说明
任务名称/对象	明确唯一任务和业务对象
目标与指标	经营目标、成功标准、截止时间
自主策略	五种策略之一
允许动作	工具、数据、对象、额度、时间窗
禁止动作	明确负面清单
审批节点	审批人、SLA、版本绑定
证据要求	来源、调用、修改、审批、结果
异常阈值	错误、成本、质量、权限、数据异常
暂停与回滚	停机人、恢复条件、补偿方式
复审周期	按月、季度或重大变更触发

FIELD RECORD 16

附录C 责任分配RACI示例

活动	董事会/经营层	CMO	业务	IT	安全	法务	供应商
场景立项	I	A	R	C	C	C	C
自主等级	I	A	R	C	R	C	C
数据接入	I	C	C	A/R	R	C	C
内容/发布审批	I	A/R	R	C	C	C	I
上线验收	I	A	R	R	R	C	C
事故暂停	I	A	C	R	R	C	C
季度复盘	A	R	R	C	C	C	I
供应商退出	A	R	C	R	C	C	C

A=最终负责，R=执行负责，C=会签或咨询，I=知会。企业应根据自身组织调整，但不能出现关键事项无人A或无人R。

FIELD RECORD 17

附录D 术语表

术语	定义
执行型智能体	能够理解目标、规划步骤、调用工具并根据结果继续行动的AI系统。
服务身份	供智能体或自动化服务使用、可独立授权和审计的非个人身份。
业务状态	业务对象在流程中的真实阶段、属性、负责人、输入输出与异常。
证据链	从来源、版本、调用、修改、审批、执行到结果的连续记录。
最小权限	仅授予完成特定任务所需的最小数据、工具、动作、范围与期限。
条件自主	在明确白名单、额度、时间窗、证据和异常阈值内自动执行。
幂等	同一请求重复发生时不会造成重复业务结果。
补偿	流程部分失败后，用反向或替代动作恢复业务一致性。
AI市场部	以经营目标为牵引，融合业务状态、角色技能、知识、连接器、流程、权限、监控和经营结果的运行系统。

FIELD RECORD 18

附录E 来源索引

以下来源用于支撑本报告对企业执行型智能体、风险治理、信息安全、隐私保护、人工智能管理与营销经营的通用判断。索引仅以纯文本呈现。

- 国务院：《关于深入实施“人工智能+”行动的意见》，2025年。
- 工业和信息化部等部门：《中小企业数字化赋能专项行动方案（2025—2027年）》，2024年。
- 国家互联网信息办公室等七部门：《生成式人工智能服务管理暂行办法》，2023年8月15日起施行。
- 全国网络安全标准化技术委员会：GB/T 35273–2020《信息安全技术 个人信息安全规范》，2020年。
- 全国信息安全标准化技术委员会：GB/T 22239–2019《信息安全技术 网络安全等级保护基本要求》，2019年。
- 全国信息安全标准化技术委员会：GB/T 43697–2024《数据安全技术 数据分类分级规则》，2024年。
- 国家标准化管理委员会：GB/T 41867–2022《信息技术 人工智能 术语》，2022年。
- ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system, 2023.
- ISO/IEC 23894:2023, Information technology — Artificial intelligence — Guidance on risk management, 2023.
- ISO/IEC 27001:2022, Information security management systems — Requirements, 2022.
- NIST: Artificial Intelligence Risk Management Framework (AI RMF 1.0), January 2023.
- NIST: Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, July 2024.
- NIST: Cybersecurity Framework 2.0, February 2024.
- MITRE: ATLAS — Adversarial Threat Landscape for Artificial–Intelligence Systems, continuously maintained.
- OWASP Foundation: OWASP Top 10 for Large Language Model Applications, 2025 edition.
- OECD: OECD AI Principles, updated 2024.
- European Union: Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, 2024.
- World Economic Forum: Global Cybersecurity Outlook 2025, January 2025.

- Microsoft: Responsible AI Standard, Version 2, 2022; related updates and transparency materials through 2025.
- Google: Secure AI Framework (SAIF), 2023; enterprise implementation guidance through 2025.
- Anthropic: Responsible Scaling Policy, updated versions through 2025.
- OpenAI: Preparedness Framework, updated version published in 2025.
- Salesforce: State of Marketing, 9th Edition, 2024.
- McKinsey & Company: The state of AI in early 2024: Gen AI adoption spikes and starts to generate value, May 2024.
- Deloitte: State of Generative AI in the Enterprise, quarterly research series, 2024—2025.
- Accenture: Technology Vision 2025, 2025.
- Gartner: Market and research notes on agentic AI and AI agents, 2024—2025.
- 中国信息通信研究院：人工智能治理、可信人工智能与大模型安全相关报告，2024—2025年。
- 中国广告协会及相关行业机构：互联网广告、品牌安全与营销合规相关规范和公开材料，持续更新。

V1.1 新增来源索引

- S29 | OpenClaw, OpenClaw Docs / Overview, accessed 2026-07-20. 自托管 Gateway、消息渠道、会话与多智能体路由说明。
- S30 | OpenClaw, Gateway Security, accessed 2026-07-20. Shell、文件、网络、消息权限及不可信输入威胁模型。
- S31 | Herbert A. Simon, Administrative Behavior, 1947. 有限理性、决策前提与组织权限的经典著作。
- S32 | Anthropic, Measuring AI Agent Autonomy in Practice, 2026-02-18. 对 Claude Code 与公开 API 中数百万次交互的匿名分析。
- S33 | Vishwakarma et al., Can LLMs Help You at Work? EnterpriseBench, Proceedings of EMNLP 2025. 500 项模拟企业任务评测。
- S34 | Xu et al., TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks, 2024. 模拟软件企业中的专业任务基准。
- S35 | Michael C. Jensen & William H. Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, 1976. 委托代理与代理成本经典论文。

- S36 | Microsoft Customer Story, Kantar Reimagines Data Prep and Puts Copilot Studio Agents to Work, 2025-11-18. 企业自行披露的数据整理案例。
- S37 | Michael Hammer, Reengineering Work: Don't Automate, Obliterate, Harvard Business Review, 1990. 流程再造经典文章。
- S38 | METR, Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity, 2025-07-10; experiment update, 2026-02-24. 16 名开发者、246 项任务的随机对照研究及后续限制说明。
- S39 | Microsoft Customer Story, Microsoft Uses Copilot Studio to Reshape Customer Experience and Drive Higher Engagement, 2026-02-27. Ask Microsoft 官网营销与销售案例。
- S40 | Microsoft Customer Story, Regal Rexnord Goes All In with Microsoft Copilot Studio, 2026-03-10. 企业知识、客服、人工转交与治理案例。
- S41 | Jay R. Galbraith, Designing Complex Organizations, 1973. 任务不确定性、相互依赖与组织信息处理的经典著作。
- S42 | Microsoft Customer Story, Capita Uses Microsoft Copilot to Transform Service Delivery for Clients, 2025-09-10. 企业自行披露的邮件分流与智能体网络案例。
- S43 | Lisanne Bainbridge, Ironies of Automation, Automatica, 1983. 自动化与人工异常接管的经典研究。
- S44 | NIST CAISI, Strengthening AI Agent Hijacking Evaluations, 2025-01-17, updated 2025-12-19. AgentDojo 间接提示注入与重复攻击实验。
- S45 | OpenClaw, Gateway Sandboxing, accessed 2026-07-20. 沙箱默认状态、适用范围与边界说明。
- S46 | James Reason, Human Error, 1990. 多层防线与组织事故机制的经典著作。
- S47 | Salesforce Customer Story, How Salesforce Deployed Agentforce to Answer Questions and Capture Leads on Its Website, accessed 2026-07-20. 官网智能体与线索闭环案例。
- S48 | Salesforce Customer Story, Asymbl Scales Like a Company 10x Their Size with Agentforce, accessed 2026-07-20. 企业及供应商自行披露的销售智能体结果。
- S49 | Robert S. Kaplan & David P. Norton, The Balanced Scorecard—Measures That Drive Performance, Harvard Business Review, 1992. 平衡计分卡经典研究。
- S50 | ZTE, CDO Cui Li at MWC Shanghai 2026: Unlocking Value and Embracing Uncertainty in the AI Era, 2026-06. Co-Claw 企业内部部署说明。
- S51 | Richard Rumelt, Good Strategy/Bad Strategy: The Difference and Why It Matters, 2011. 诊断、指导方针与连贯行动的战略内核。

FIELD RECORD 19

结语：企业最终交付的不是任务，而是受约束的执行权

自主性不是一个开关由错误损失、可逆性、证据完整度和数据敏感度共同决定。

A 建议型	B 草拟型
C 执行前审批	D 条件自主
E 禁止自动化	

个人执行型智能体让管理者第一次直观看到AI“能动手”的可能性。企业真正需要完成的工作，是把这种动作能力纳入经营目标、组织角色、业务状态、权限审批、证据评测和责任体系。

未来12—18个月，领先企业不会以拥有最多Agent取胜，而会以更清晰的任务边界、更可靠的公共能力、更强的证据链、更快的异常处置和更严格的价值纪律建立优势。

> **AI市场部的成熟标志，不是AI做得更多，而是企业知道哪些事可以交给AI、交到什么程度、凭什么相信、何时叫停，以及最终由谁负责。**

竹势 AI 营销智库

2026 年 7 月 · V1.1 内容深化版

PUBLISHER

关于竹势 AI 营销智库

竹势 AI 营销智库关注企业增长、营销组织变革、智能体运营与 AI 治理，帮助企业老板、CEO、CMO 和数字化管理者把复杂技术议题转化为可讨论、可比较、可执行的经营决策。

本报告的企业化门槛、任务授权框架和迁移工具，旨在帮助管理团队把个人执行型智能体纳入正式经营体系，并在规模化之前建立必要的身份、上下文、权限、证据与问责机制。