



AI MARKETING / EVIDENCE-GATED
ROLLOUT

14天验证、30天试点、90天规模化：AI 营销落地路线图

每一次扩张都先获得证据许可：验证问题，跑通真实闭环，再复制可运营、可审计、可回退的工作单元。

14天消除错误问题

用基线、最小任务集与风险边界决定是否试点

30天验证真实闭环

把客户、责任、例外、人工控制和业务写回同时纳入

90天证明可重复

由第二团队独立运行，验证成本、监控与运营接管

四项决议管理投入

继续、调整、暂停或终止，不用庆功会替代资源取舍

BOARD BRIEF 01

出版说明

竹势 AI 营销智库出品

本报告提供一套可直接用于企业 AI 营销落地的董事会级管理系统。它将场景选择、业务验证、真实试点、跨团队复制、治理接管与退出纪律连接为连续的证据许可轨道。报告中的时间节点用于建立管理节奏；企业应依据风险等级、数据准备度、任务频率、采购周期、合规审查和核心系统改造复杂度调整日历，但不得取消阶段闸门、责任确认与停止机制。

报告中的企业结果均按公开披露口径说明。公司披露、平台案例或供应商测量的结果不被包装成独立行业结论；无法公开核验的经营结果明确标注“未披露”。示例公式和阈值制定方法用于帮助企业设计本企业标准，不代表行业平均值。

版权所有 © 2026 竹势 AI 营销智库。允许企业内部用于管理讨论、试点设计与培训；公开转载应保留报告名称与出品方。

- 01 执行摘要
- 02 核心判断
- 03 第一章 证据许可制落地轨道
- 04 第二章 场景选择与问题章程
- 05 第三章 14天验证
- 06 第四章 30天真实业务试点
- 07 第五章 价值证据树与五类指标
- 08 第六章 阶段门与四种决议
- 09 第七章 数据、集成与治理递增
- 10 第八章 人、责任与运营接管
- 11 第九章 90天规模化
- 12 第十章 停止、回退与资产回收
- 13 第十一章 案例卷宗
- 14 第十二章 董事会决议页
- 15 管理工具附录
- 16 来源索引

执行摘要

竹势 AI 营销智库出品

EVIDENCE LICENSE / 证据许可原则

速度不是把全部工作塞进九十天，而是让每一笔资源、每一次授权和每一次集成都由更高等级证据解锁。

14天消除错误问题 30天验证真实闭环 90天证明可重复与可接管

企业 AI 营销落地最常见的失败，不是模型无法生成内容，而是组织把“看起来能用”误当成“值得投入”，把沙盒准确率误当成业务价值，把一次成功演示误当成可复制能力。采购、培训和演示可以在数天内完成，却无法自动回答五个董事会问题：问题是否值得解决，基线是否可信，AI 是否产生可识别贡献，风险是否在容忍范围内，价值能否被另一个团队重复获得。

“证据许可制落地轨道”把项目资源分成连续但可中断的三笔投资。14 天阶段购买的是认知：确认问题、基线、最小任务集、离线评测和风险边界。30 天阶段购买的是真实性：让 AI 进入真实客户、真实工作、真实责任和真实例外，跑通端到端最小闭环。90 天阶段购买的是可复制性：把经过证明的闭环封装为标准工作单元，完成运营接管、成本模型、监控审计、培训通关和跨团队复制试验。每一笔资源都必须由上一阶段的证据许可。

这一轨道的反共识是：日历本身不创造速度。真正的速度来自第一天就写清成功证据、反指标、责任人、异常处置、停止条件和回滚路径。项目因此可以更快排除伪需求、坏数据和不可承受风险，也能避免因沉没成本继续投入。对于高监管发布、低频高价值任务、核心系统改造或长品牌周期，14/30/90 应扩展观察期；但仍应保留“问题许可—真实闭环许可—复制许可”的逻辑。

董事会应掌握的四句话

一、先证明问题和基线，再证明技术。二、先跑真实最小闭环，再谈平台化。三、规模化复制工作单元，不复制个人技巧。四、停止低价值试点是投资纪律，不是失败。

阶段	购买的认知	最低交付	核心决议	典型回退
14 天验证	问题是否值得解决，AI 是否有可识别贡献	问题章程、基线、最小任务集、离线评测、风险清单	进入 / 重构 / 停止	回到数据修复、任务重构或人工流程优化
30 天试点	真实闭环能否在责任与例外下稳定运行	工作流、RACI、人工控制点、业务与反指标、异常处置、回滚方案	继续 / 调整 / 暂停 / 终止	退回影子运行、缩小权限、恢复原流程
90 天规模化	价值能否重复、风险能否运营化、成本能否承受	标准工作单元、运营手册、监控审计、成本模型、培训通关、复制试验	规模化许可 / 有条件许可 / 延长验证 / 退出	停止扩张，保留已验证单元或切回人工

BOARD BRIEF 03

十个核心判断

竹势 AI 营销智库出品

1. 落地速度由证据设计决定，而非日历压缩。

当成功、反指标和停止条件在项目后期才被讨论，团队会反复争论“是否有效”。第一天定义证据，才能并行安排数据、评测、流程和治理。

2. 首个场景必须同时满足“小到可归因、真到能暴露摩擦”。

太大的场景无法识别贡献；太假的场景不会暴露权限、例外、客户反应和跨角色责任。

3. 14 天的目标是消除错误问题，而非做出漂亮演示。

若基线不存在、数据不可用或任务频率不足，应先修复测量与数据，不应以生成样例替代业务证据。

4. 30 天试点的对象是端到端工作闭环。

真实试点必须覆盖输入、执行、审核、发布或触达、结果回收、异常和复盘；仅测单点输出质量不足以支持经营决策。

5. 价值必须同时包含效率、质量、增量经营结果、长期资产与风险。

只看节省工时会鼓励低质量自动化；只看收入会忽略短观察期和归因噪声；只看准确率会忽略客户和组织结果。

6. 阶段门是资源配置会议，不是成果展示会。

评审必须允许继续、调整、暂停、终止四种决议，并明确下一笔预算、责任和风险暴露。

7. 90 天复制的是可复制工作单元。

工作单元包含输入、角色、标准、接口、监控、例外和处置；单个操作模板或账号无法承载治理与结果。

8. 集成深度应随证据递增。

在价值未证实时连接核心系统会制造沉没成本；但身份、权限、隐私和监管底线必须从第一天满足。

9. 业务 owner、技术 owner、风险 owner 和运营接管人不可或缺。

批准、执行、复核、接管和事故责任若没有名字，试点就会停留在“创新团队的项目”。

10. 可回退性是扩张速度的前提。

影子运行、人工审批、版本冻结、日志和恢复原流程让组织敢于授权；没有回退能力，风险部门只能选择阻止。

BOARD BRIEF 04

第一章 证据许可制落地轨道

竹势 AI 营销智库出品

01 / 14 DAYS

验证问题与贡献

基线 · 最小任务集 · 离线评测 · 风险边界

许可：进入真实试点

02 / 30 DAYS

跑通真实业务闭环

客户 · 责任 · 例外 · 人工控制 · 可回退

许可：复制工作单元

03 / 90 DAYS

证明可重复与可运营

第二团队 · 成本 · 监控 · 接管 · 再许可

许可：受控规模化

每一道门都允许继续、调整、暂停或终止；日历不能替代证据。

核心判断：AI 营销落地应被视为一组逐步行权的投资，而非一次性批准的技术项目。每一阶段只购买足够的信息来决定是否配置下一笔资源。

实物期权思想指出，在高度不确定的投资中，保留延迟、扩张、收缩或退出的选择具有经济价值。AI 项目同时存在技术能力、数据质量、客户反应、流程适配和监管风险，适合采用小额、可逆、分阶段的承诺。14 天验证相当于购买“继续研究”的权利；30 天试点购买“进入真实运营”的权利；90 天规模化购买“复制与平台化”的权利。管理动作不是追求一次性净现值精确计算，而是明确每一道门的新增信息价值、不可逆成本和退出代价。

Cooper 的 Stage-Gate 思想强调把创新过程分段，并在闸门处依据质量、商业价值与行动计划配置资源。应用到生成式 AI，需要三项修正：第一，模型和数据会快速变化，因此闸门不应只审文档，还应审可复现测试、日志和异常；第二，探索性任务不能被僵化阈值扼杀，应使用“证据充分度+风险容忍度”的组合判断；第三，业务 owner 必须承担结果责任，技术团队不能独自为经营价值背书。

Popper 的可证伪原则要求一个主张必须存在可能被事实推翻的条件。AI 试点章程因此不能写“提升营销效率”，而应写成可被推翻的假设，例如：“在指定内容审校任务中，AI 辅助流程在保持合规错误不增加的条件下，减少从 Brief 到可发布初稿的人工处理时间；若质量复核不通过率上升或品牌禁区触发超过容忍范围，则假设不成立。”管理层由此能区分验证学习与事后解释。

Ries 的验证学习将创业活动的进展定义为对关键假设的证伪或强化。企业 AI 场景比创业产品多出数据治理、权限和既有流程约束，因此最小可行对象不能只是前端原型，而应是“最小可评测任务集”和“端到端最小闭环”。二者分别承担 14 天和 30 天的学习任务。

轨道层	核心问题	增加的真实性	新增不可逆成本	许可证据
问题许可	是否值得解决	真实历史样本、真实基线	极低	问题频率/价值、基线可信、最小任务可测
技术许可	AI 是否有贡献	盲测、离线评测、对照	低	质量、时间、错误类型、风险边界
运营许可	能否进入真实工作	真实角色、客户、例外、权限	中	端到端结果、异常处置、人工控制、回滚
复制许可	能否跨团队重复	第二团队/第二区域/第二产品	较高	效果重复、成本稳定、运营接管、审计可用
平台化许可	是否值得深度集成	核心系统与长期运营	高	组合价值、接口稳定、治理能力、全生命周期成本

边界：对于医疗、金融建议、重大广告发布、个人信息密集处理和核心交易系统，监管审查、模型验证和安全测试可能远超 90 天。此时应扩大日历而非取消证据轨道。对于低频高价值任务，例如年度品牌重塑或重大新品命名，30 天内可能没有足够结果样本，应使用历史回放、专家盲评、过程证据和延长观察窗口。

第二章 场景选择与问题章程

竹势 AI 营销智库出品

经营问题 谁的哪项结果需要改变
最小闭环 一类客户 · 一条任务链 · 一个结果
基线与反事实 没有 AI 时会怎样
停止条件 何时不再投入下一笔资源

核心判断：首个闭环的优先级由“可归因价值密度”决定。它需要足够高频、结果可观察、责任边界清晰、风险可控，并能在短周期内获得对照；高热度、跨系统、长周期的大场景通常不适合作为第一笔投资。

Goldratt 的约束理论提醒管理者，局部效率只有在作用于系统约束时才会转化为整体结果。若销售线索的瓶颈是跟进速度，批量生成更多内容可能只会扩大无效流量；若内容审批是瓶颈，优先自动化初稿却不改审批流程，周期仍不会缩短。因此问题章程必须先描述系统目标、当前约束、下游承接能力和不可恶化的反指标。

Drucker 的结果观把管理重心放在外部结果，而非内部活动。AI 项目常用“生成数量、账号开通、培训人次”表示进展，但这些是活动，不是经营结果。问题章程应把活动指标降级为诊断信息，把客户响应、转化、响应时效、质量、风险和资产沉淀放到结果链上。

筛选维度	关键问题	优先信号	降级或延后信号
价值密度	每次任务影响何种成本、收入、客户或风险	高频、可量化、靠近经营约束	低频、价值依赖长期品牌效应
可观察性	何时能看到结果，是否有历史基线	结果周期短、日志完整、可建立对照	无基线、数据口径频繁变化
可归因性	AI 贡献能否与渠道、促销、人员差异分开	任务边界清晰、可做交叉或分组比较	多变量同时大改、无法保留对照
流程完整性	是否可覆盖输入至结果回收	有明确 owner、输入输出可结构化	跨越过多部门、下游无人承接
风险可控性	最坏后果是否可恢复	可影子运行、可人工审批、可撤回	不可逆发布、重大监管或安全后果
复制潜力	成功后能否形成标准单元	同类任务多、接口稳定、差异可参数化	高度依赖个人关系与一次性创意

场景一页章程必须包括：经营问题、系统约束、目标用户或客户、任务频率、当前基线、AI 贡献假设、五类价值证据、反指标、风险等级、数据来源、角色与事故责任、进入条件、停止条件和二次假设。章程的目的不是获得所有人的口头支持，而是让不同部门对“什么事实会改变我们的决定”形成书面一致。

场景边界测试

首个闭环应“小到可以解释结果、真到可以暴露摩擦”。若它只能在演示数据和创新团队手中运行，说明还不够真；若它同时改变渠道、组织、价格、产品和系统，说明还不够小。

第三章 14天验证：证明问题、基线、贡献与风险边界

竹势 AI 营销智库出品

D1-2 校准问题 owner、基线、边界
D3-5 建立任务集 正常、边界、事故样本
D6-10 离线验证 贡献、错误、人工负担
D11-13 风险与成本 权限、数据、回退
D14 证据门 继续、调整、暂停、终止

核心判断：14 天阶段是证据准备期，不是缩短版实施。它的产出是一个可进入真实试点的决议包；若数据、任务定义或基线不成立，按时停止比按时上线更有价值。

第 1—2 天确认问题章程与责任。业务 owner 说明为什么当前约束值得解决，财务或经营分析负责人确认价值口径，风险 owner 确认不可触碰边界。第 3—4 天建立基线：从历史样本中测量处理时间、返工、质量、转化、异常和单位成本，并记录分母。第 5—6 天建立最小任务集：覆盖常规、边界、失败和敏感样本，避免只选择“模型擅长”的题。第 7—9 天完成离线运行与盲评。第 10—11 天开展红队与风险测试。第 12 天复核数据与归因。第 13 天形成进入、重构或停止建议。第 14 天召开问题许可门。

Deming 强调，没有对系统与变异的理解，数字会误导管理。AI 输出质量具有任务、输入、模型版本和评审者差异。14 天评测应记录样本构成、版本、评审准则和评审者一致性；不能把一次平均分当成稳定能力。对于创意任务，应同时保留专家质量评审和真实受众实验的后续计划。对于分类与抽取任务，可使用精确率、召回率或错误成本，但阈值必须由业务风险决定。

最小任务集不追求统计代表整个行业，而是覆盖本场景的关键风险面。建议至少包括：高频标准任务、长尾例外、缺失数据、相互矛盾资料、品牌禁区、敏感个人信息、外部事实需要核验的任务、超出权限的动作请求，以及必须转人工的情形。样本量取决于任务异质性和决策风险；报告不设虚构的统一数量。

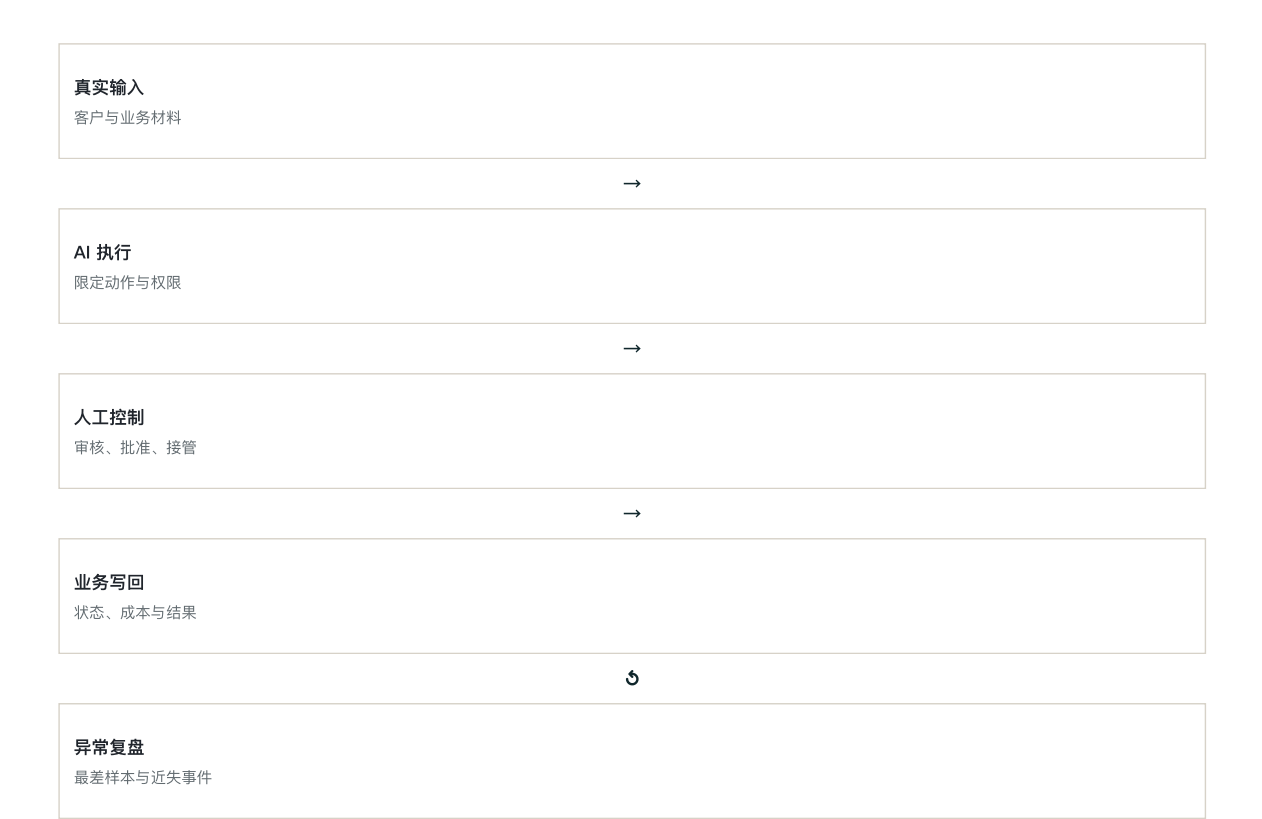
离线评测应尽量采用盲评或成对比较。对内容类任务，可让评审者在不知道生产方式的情况下，比较可发布性、事实准确、品牌一致、说服力 and 编辑工时。对流程任务，应记录机器处理时间和总周期，防止“模型秒出结果、人工花更久纠错”的假效率。对高风险任务，应单独报告最坏错误和不可接受错误，而非用平均分掩盖。

14天节点	关键任务	主要交付	责任主位	未通过时动作
D1-D2	问题章程、约束、责任与风险分级	签字版章程	业务 owner	缩小问题或停止
D3-D4	历史样本与基线重建	基线表、口径字典	经营分析/数据 owner	先修测量与数据
D5-D6	最小任务集与评审标准	任务集、盲评规则	业务专家	重构任务边界
D7-D9	离线运行、对照与编辑工时记录	输出包、评测记录	技术 owner+业务评审	更换方案或人工增强
D10-D11	红队、隐私、权限和事实核验测试	风险清单、控制建议	风险 owner	限制权限或终止
D12-D13	证据树、归因与决议建议	证据包	项目负责人	补证或提出二次假设
D14	问题许可门	进入/重构/停止决议	决策委员会	按决议配置下一笔资源

14 天进入条件建议采用“必要条件+整体判断”。必要条件包括：问题价值和 owner 明确；基线可复核；最小任务集覆盖主要风险；AI 在至少一个关键指标上显示可识别贡献；不可接受风险有控制或能通过缩权避免；30 天真实闭环具有数据、人员和回滚条件。任何一项缺失都不应被“演示效果好”替代。

第四章 30天真实业务试点：跑通端到端最小闭环

竹势 AI 营销智库出品



核心判断：30 天试点必须让 AI 进入真实责任系统。它不是扩大离线样本，而是验证 workflow 在真实输入波动、真实客户反应、真实审批、真实例外和真实系统约束下是否成立。

一个完整闭环至少包含七个环节：任务触发、上下文与数据读取、AI 执行、人工控制、对外发布或内部交付、结果回收、异常与复盘。任何环节靠人工临时搬运都可以在试点中保留，但必须显式记录为“人工桥接”，并计算成本和出错点。隐形人工会制造“自动化率”幻觉。

高风险对外动作应采用影子运行或双轨运行。影子运行中，AI 产生建议但不直接对外，团队比较 AI 建议与原流程结果；双轨运行中，一部分低风险任务进入受控真实流程，另一部分保留原流程作为对照。随着证据增加，可逐步扩大权限。摩根士丹利 2024 年发布的 AI Debrief 工具在客户同意前提下生成会议记录与行动项，邮件由顾问编辑后自行发送，并写入 Salesforce，体现了“自动记录+人工决定外发”的控制结构。[S8]

RACI 不足以覆盖 AI 事故责任，因此建议增加 O（Operational owner，运营接管人）和 X（Exception owner，异常负责人）。业务 owner 对经营目标负责；技术 owner 对系统可靠性、版本和接口负责；风险 owner 对政策、隐私和控制设计负责；运营接管人对试点结束后的日常运行负责；异常负责人在服务中断、错误发布、数据泄露或客户投诉时拥有停机与升级权。

30 天指标应同时观察业务指标和反指标。内容效率提升若伴随事实错误、品牌返工或受众负反馈增加，不能视为通过。客户服务处理量提高若复联率、投诉或复杂问题转人工延迟恶化，也不能视为通过。Goodhart 定律提醒：当指标成为目标，它就可能失去衡量作用。因此应设置成组指标，并定期检查团队是否通过挑选简单任务、延迟记录异常或改变分母来“优化”数字。

试点周次	真实度升级	主要活动	关键证据	控制点
第1周	真实输入、有限用户	配置 workflow、角色培训、影子运行	基线复核、输入缺陷、人工编辑负担	不外发或全部审批
第2周	低风险真实任务	小比例上线、异常日志、客户或内部用户反馈	周期、质量、采用、异常类型	白名单、限额、可撤回
第3周	扩大样本与例外	处理长尾任务、系统接口、故障演练	反指标、回滚时间、责任响应	权限分层、人工接管
第4周	稳定性与交接	重复运行、成本核算、运营演练、阶段评审	可重复结果、单位成本、运营准备	版本冻结、审计导出

试点的最低完成定义：至少一个真实业务闭环稳定运行；所有关键动作可追溯；人工控制点和转人工规则有效；异常有责任人与处置时限；业务与反指标均被记录；可在约定时间内恢复原流程；运营接管人完成至少一次独立运行。若只有创新团队能操作，试点尚未完成。

第五章 价值证据树：效率、质量、增量结果、长期资产与风险

竹势 AI 营销智库出品

E1 效率 端到端周期与人工负担
E2 质量 一次通过、返工与错误
E3 增量结果 客户行为、收入与成本
E4 长期资产 任务集、知识、接口与能力
E5 风险 事故、近失、回退与暴露

领先指标说明运行状态，结果指标说明经营改变，反指标防止把成本转移给客户与一线。

核心判断：单一 ROI 无法承载 AI 营销早期决策。管理层需要一棵从活动到经营结果的证据树，区分领先指标、结果指标、反指标、归因边界和观察周期。

效率证据回答“同样结果是否消耗更少资源”，包括人工触达时间、编辑时间、等待时间、返工次数、总周期和单位任务成本。必须测总流程而非模型响应时间。质量证据回答“结果是否达到可用标准”，包括事实准确、合规、品牌一致、客户问题解决率、专家评分和一次通过率。增量经营结果回答“是否改变客户或现金结果”，包括额外合格线索、响应率、转化、留存、客单、媒体效率或服务成本，但必须说明归因设计。长期资产证据包括结构化知识、可复用 workflow、数据质量、员工能力、实验记录与客户洞察。风险证据包括不可接受错误、隐私暴露、越权、歧视、投诉、审计缺口和供应商依赖。

观察周期必须与因果链匹配。30 天可以较好观察内容周期、客服响应、审核返工和低频率不高的转化先导指标，却不能强行证明长期品牌资产、客户终身价值或年度收入。对长周期结果，可在 30 天阶段记录领先指标和实验设计，并把最终判断移交至后续季度。

归因强度从低到高可分为：前后比较、匹配历史、同期对照、分组实验、交叉实验和随机实验。企业不必对每个场景做学术级实验，但应清楚知道证据处于哪一级。促销、渠道预算、季节、人员变化和产品供给会同时影响结果，应在证据包中记录。

价值类	领先指标示例	结果指标示例	反指标示例	典型观察期
效率	自动完成比例、等待时间、编辑工时	端到端周期、单位任务成本	返工、隐形人工、队列积压	日/周
质量	规则通过率、事实核验覆盖	一次通过率、解决率、专家盲评分	严重错误、品牌偏离、客户误导	周/月
增量经营结果	响应速度、有效互动、MQL质量	增量收入、转化、留存、服务成本	退订、投诉、低质线索、渠道蚕食	月/季度或更长
长期资产	结构化记录率、复盘完成率	可复用单元数、知识命中、能力通关	知识过期、版本漂移、人员依赖	月/季度
风险	审批覆盖、日志完整、转人工及时	事故率、损失额、监管发现	近失事件、权限绕过、供应商锁定	持续

示例价值公式（不代表行业标准）： $\text{净可验证价值} = \text{可归因增量毛利} + \text{可确认成本避免} + \text{风险损失期望值下降} + \text{经批准的资产价值代理} - \text{全生命周期增量成本}$ 。
全生命周期成本至少包含模型与平台、集成、数据准备、人工复核、运营、监控、安全、培训、供应商管理和退出成本。资产价值代理须由管理层批准，例如“可复用工作单元替代未来重复设计工时”，不得把抽象知识随意货币化。

第六章 阶段门：继续、调整、暂停、终止

竹势 AI 营销智库出品

继续 证据达到下一阶段许可
调整 改变范围、流程或假设
暂停 等待关键依赖或风险修复
终止 回收资产并释放资源

核心判断：闸门的职责是配置下一笔资源和责任，不是确认团队努力。四种决议必须都可被选择，否则评审会退化为庆功会。

继续：关键必要条件满足，证据方向一致，新增风险在容忍范围内，下一阶段的 owner、预算与回退方案明确。调整：价值假设仍值得验证，但任务边界、数据、流程、技术、指标或控制需要重构；调整必须写明新假设与补证期限。暂停：外部依赖、监管审查、数据修复、资源或关键人员尚未就绪；暂停要有恢复条件，不能成为无期限搁置。终止：问题价值不足、贡献不可识别、风险不可承受、成本结构不成立或组织无意接管；终止同时启动资产回收与复盘。

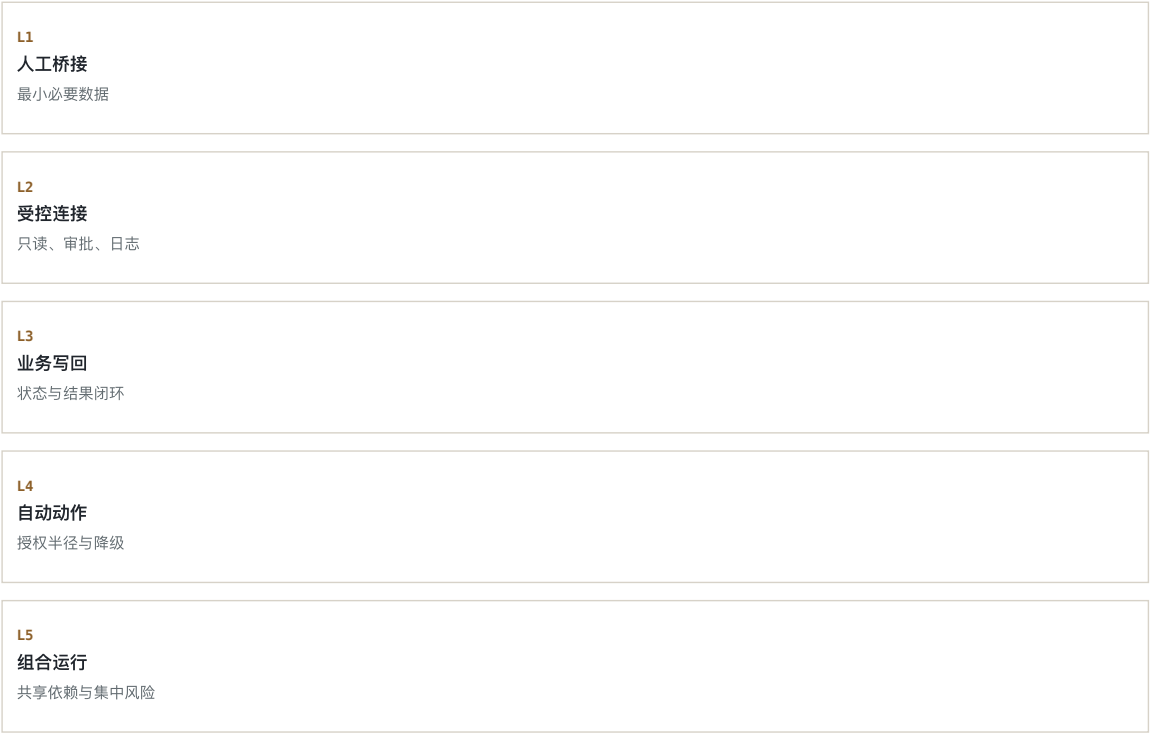
Kahneman 与 Tversky 对损失厌恶和沉没成本的研究说明，决策者更难承认已经投入的资源无法回收。阶段门应由未直接拥有项目成功叙事的人参与，采用预先承诺的停止条件，并在评审材料中把“已花成本”与“未来增量成本”分开。决策应基于未来，而非为过去辩护。

探索项目需要避免假精确。对于创意、洞察和策略任务，可使用多证据组合，而非规定一个看似科学的单一分数。例如，专家盲评改善、编辑工时下降、严重错误不增加、真实受众小样本信号改善，可共同支持进入真实试点；但若证据相互冲突，应选择调整或延长，而非平均后掩盖。

评分域	建议问题	红色信号	绿色信号
问题与基线	问题是否仍是系统约束，口径能否复核	价值低、分母不清、基线被重写	owner确认、历史样本稳定
贡献与归因	AI 是否产生可识别增量	只展示样例、无对照、人工代做未记录	盲评/对照支持，贡献链清楚
风险与控制	最坏错误是否可防、可发现、可恢复	不可逆、无日志、无人停机	审批、限权、监控、回滚有效
经济性	单位价值能否覆盖全生命周期成本	复核成本高、供应商价格敏感	成本透明、规模后不恶化
可运营性	业务是否愿意并能够接管	仅创新团队会用、责任空缺	运营 owner通关、SOP与值班明确
可复制性	第二团队能否重复结果	高度依赖个人和一次性数据	工作单元标准化、差异可配置

第七章 数据、集成与治理递增

竹势 AI 营销智库出品



核心判断：价值证据不足时应保持技术承诺可逆；监管、安全、身份与隐私底线则不能延期。

14 天阶段优先使用脱敏历史样本、只读数据和隔离环境，目标是验证任务与风险，不应为了演示而连接核心系统。30 天阶段按闭环需要增加最少接口，并采用白名单、限额、人工审批和完整日志。90 天阶段只有在价值可重复后，才投资稳定接口、监控、权限自动化、灾备和供应商管理。集成深度与证据成熟度同步，可以避免在错误场景上形成高昂沉没成本。

NIST AI Risk Management Framework 将治理、映射、测量和管理视为持续循环，而非上线前一次检查。对营销场景，映射应包括客户、员工、数据主体、渠道平台和监管主体；测量应覆盖有效性、可靠性、隐私、公平、安全和可解释；管理应落实到权限、日志、人工控制、事件响应和供应商条款。[S16]

欧盟 AI Act 的风险分层与逐步适用提醒跨国企业：同一营销流程可能因用户画像、自动决策、员工监控或敏感数据处理而进入不同合规义务。中国企业还需结合个人信息保护法、数据安全法、网络安全法、生成式人工智能服务管理暂行办法、广告法及行业规则。报告不提供法律意见；项目章程应由合规或法务确认适用边界。

阶段	数据策略	集成策略	治理最低要求	退出设计
14天	脱敏历史、只读、数据字典	文件/API沙盒、人工导入可接受	风险分级、用途限定、访问控制、评测日志	可删除数据与环境
30天	最少真实数据、分层授权	有限生产接口、人工桥接显式记录	审批、白名单、限额、监控、事件响应	一键停用、恢复原流程
90天	数据质量责任、保留与更新机制	稳定接口、版本与依赖管理	审计导出、模型/供应商变更、灾备、定期复评	数据迁移、合同退出、资产归属

第八章 人、责任与运营接管

竹势 AI 营销智库出品

BUSINESS 业务 owner 问题、结果与取舍
TECH 技术 owner 能力、连接与可靠性
RISK 风险 owner 控制、事故与许可
OPERATE 运营接管人 标准、监控与日常处置

核心判断：项目团队能跑通不等于组织能运营。规模化许可的必要条件，是业务运营者在没有原项目核心成员临场救火的情况下独立运行、识别异常并完成回退。

March 的探索与利用理论说明，组织需要同时发现新方法和稳定复制既有方法。14 天与早期 30 天偏探索，应允许多方案和快速证伪；90 天偏利用，需要标准、培训、监控和成本控制。两种逻辑若混在一个团队中，容易出现两类失败：探索过早被合规化，或试点长期处于“实验特区”而无法接管。建议设立清晰的移交条件和双环学习机制：一方面改进执行动作，另一方面检视目标、规则和假设本身。

培训通关不以听课时长计，而以真实任务表现计。运营人员需在受控环境中完成任务启动、结果复核、异常识别、人工接管、数据纠错、日志查询和回退演练。管理者需能解释证据树、成本台账和四种决议。风险与技术团队需完成事故桌面推演。

心理安全影响异常上报。若员工认为报告 AI 错误会被视为抵制创新，近失事件会被隐藏，管理层只能看到漂亮指标。业务 owner 应明确：发现高价值失败样本、主动停机和提出停止建议属于项目贡献。绩效设计应奖励问题发现和资产沉淀，避免只奖励自动化率。

角色	不可转移责任	30天最低动作	90天接管证据
业务 owner	经营问题、资源与结果	确认指标、处理跨部门阻塞	批准标准单元与预算
技术 owner	可靠性、版本、接口与成本	维护运行、记录故障与版本	监控、变更、灾备可运营
风险 owner	边界、控制、事故升级	审批控制点、复核近失事件	定期审计与再评估机制
运营接管人	日常运行、排班、SLA	参与试点并独立值守	独立运行与回退演练通过
数据 owner	口径、质量、访问与更新	修复缺陷、维护数据字典	质量监控和责任机制
财务/经营分析	价值口径与归因	核对成本和分母	持续价值台账与预算规则

第九章 90天规模化：复制可治理工作单元

竹势 AI 营销智库出品

可复制工作单元
一个可独立运行、审计和回退的经营能力

输入

数据与触发

角色

执行与批准

标准

质量与边界

接口

系统与写回

监控

成本与异常

处置

暂停与回退

核心判断：规模化对象是“可复制工作单元”，不是更多账号、更多内容或个人操作技巧。工作单元是一份可运营的契约，定义输入、角色、标准、接口、监控、例外和处置。

标准工作单元包含十二项：业务目标与适用边界；触发条件；输入及质量要求；步骤与系统接口；AI 与人工角色；输出标准；人工控制点；例外分类；服务水平；监控与反指标；成本结构；版本、变更与退出。跨业务差异大时，应复制原则、接口和控制骨架，让内容标准、渠道规则和数据字段参数化，而不机械复制流程细节。

90 天不应等到最后才开始复制。建议第 31—45 天稳定第一单元并形成草案；第 46—60 天由运营团队接管并进行故障演练；第 61—75 天在第二团队、区域、产品或渠道开展复制试验；第 76—90 天比较效果衰减、成本变化、例外差异和治理负担，形成规模化许可。复制试验应选择“相似但不相同”的环境，才能检验标准化边界。

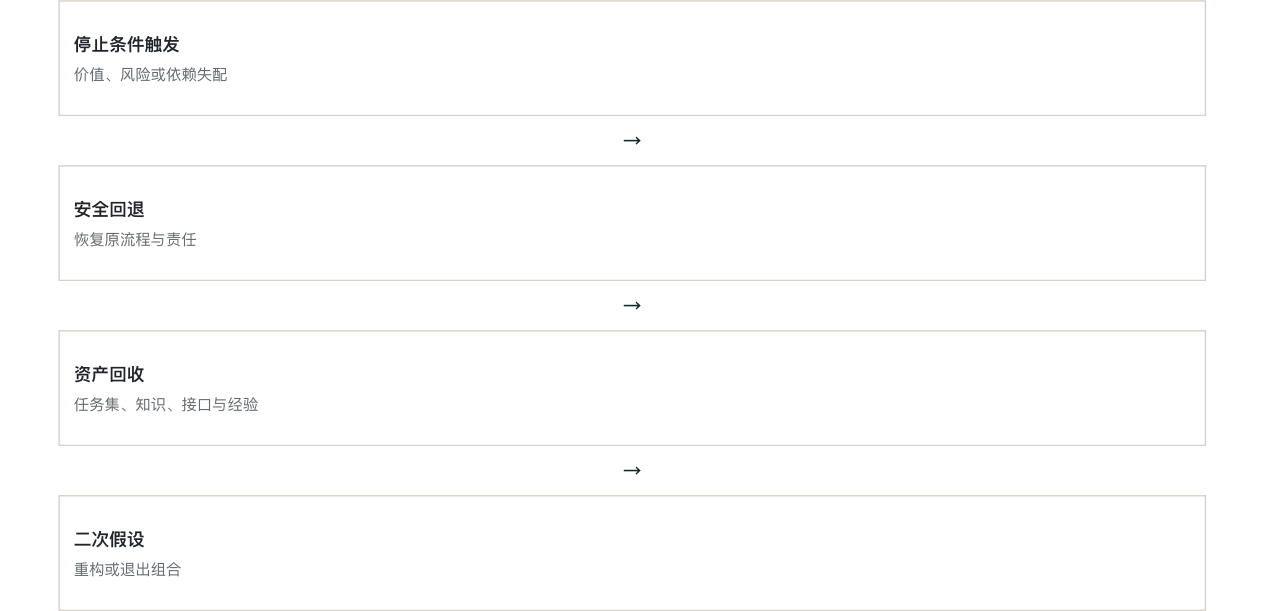
成本曲线可能随规模改善，也可能恶化。更多任务会摊薄固定集成成本，但长尾例外、人工复核、模型用量、渠道 API 限制和支持成本可能上升。成本模型应按标准任务、复杂任务、异常任务分层，记录单位经济性和边际运营负担。

P&G 的随机现场实验以真实产品创新任务考察生成式 AI 与团队协作。2025 年 NBER 工作论文披露 776 名专业人员的预注册现场实验；2026 年正式发表版本报告 791 名专业人员。结果显示，AI 可帮助个人达到传统双人团队的表现水平，AI 团队产生最高质量方案，同时缩小职能知识边界；但该实验聚焦特定创新任务，不能直接外推到所有营销执行或长期市场结果。[S1][S2] 管理含义是：规模化应复制“任务设计、角色组合和评测”，而非只发放工具。

90天区间	目标	交付	许可问题
D31–D45	稳定第一工作单元	标准草案、监控、成本分层	结果是否可重复而非偶然
D46–D60	运营接管	运营手册、培训通关、故障演练	离开项目团队后能否运行
D61–D75	跨团队复制	第二单元部署、差异清单、效果对比	哪些可标准化，哪些需参数化
D76–D90	规模化评审	审计包、组合成本、资源计划、许可建议	扩大后价值、风险与成本是否仍成立

第十章 停止、回退与资产回收

竹势 AI 营销智库出品



核心判断：退出机制决定组织是否敢于试验。没有清晰回退路径，风险部门会阻止真实授权；没有资产回收纪律，停止项目只会留下散落账号、数据和依赖。

停止条件应在项目开始前写入章程，并分为即时停止与阶段停止。即时停止包括重大数据泄露、未经批准对外发布、严重客户误导、越权操作、无法恢复的系统影响或触发监管底线。阶段停止包括问题价值下降、AI 贡献无法识别、人工复核成本抵消收益、单位经济性不成立、异常长期高于容忍范围、业务拒绝接管或供应商风险不可接受。

回退不是简单关闭系统。它包括冻结新任务、保存证据、通知相关角色、切换原流程、撤回或更正外部内容、修复数据、撤销权限、关闭密钥、核算影响、完成根因分析和重新进入条件。高风险流程应在上线前演练回退时间。

资产回收包括：可复用数据字典、评测任务集、工作流模块、控制规则、失败样本、风险分类、接口文档、培训材料、成本记录和二次假设。即使项目终止，这些资产也能降低下一轮试验成本。复盘应区分“假设错误、数据不足、执行失败、环境变化和治理不允许”，避免把所有终止归为技术失败。

退出对象	必须回收	必须关闭/撤销	二次利用
数据	数据字典、质量缺陷、经批准样本	临时副本、超期数据、无用途授权	后续场景基线与评测
技术	接口文档、版本记录、故障与成本	密钥、账号、临时环境、自动任务	通用连接与监控模块
流程	SOP、异常分类、人工控制设计	无人负责的自动动作	其他低风险工作单元
组织	培训记录、责任教训、接管缺口	临时权限与非正式责任	能力地图与岗位设计
证据	正负结果、归因限制、停止理由	选择性展示材料	二次假设与组合决策

第十一章 案例卷宗：成功、受限与边界

竹势 AI 营销智库出品

十个案例只回答一个问题：什么证据足以获得下一阶段许可？

企业自报、政府案例、现场实验和运营披露按各自口径使用，不把局部结果外推为行业成功率。

案例1：宝洁“Cybernetic Teammate”随机现场实验（2025—2026）

主体与情境：研究团队与宝洁在真实新产品创新挑战中开展预注册现场实验，工作论文披露 776 名专业人员，正式期刊版本披露 791 名。动作：参与者被随机分配为个人或双人团队，并分别有无生成式 AI 支持。机制：AI 提供跨职能知识与快速方案迭代。结果：研究报告 AI 支持的个人可达到传统团队水平，AI 团队质量最高，并改善知识整合与部分体验。归因限制：随机设计支持任务内因果判断，但任务属于产品创新，不代表长期销量、品牌资产或所有营销流程。可复制部分：真实任务、随机或对照设计、角色组合、盲评与对专业边界的测量。[S1][S2]

案例2：摩根士丹利财富管理 AI Debrief（2024）

主体与情境：金融顾问会议记录、行动项和 CRM 写回涉及客户同意、监管与事实准确。动作：工具在客户同意后生成会议记录、行动项和可编辑邮件草稿，由顾问决定是否发送，并将笔记保存至 Salesforce。机制：把低风险记录自动化与高风险外发决策分离。结果证据：公司披露产品发布与流程设计，未在该公告中披露收入或顾问生产率的独立量化结果。归因限制：官方发布说明能力与控制，不足以证明经济收益。可复制部分：客户同意、人工编辑、系统写回和责任保留。[S8]

案例3：Klarna 客服 AI 助手（2024—2025）

主体与情境：全球支付与购物平台面对多语言、高量客服。动作：2024 年上线 AI 助手处理对话，并扩大内部 AI 使用。公司披露首月 230 万次对话、约占客服聊天三分之二，并声称处理时间下降、重复咨询减少；2025 年公司继续披露单笔客服成本相较 2023 年第一季度下降 40%。机制：高频标准问题自动处理，复杂问题转人工。归因限制：结果为公司披露，受到需求结构、人员、产品与流程变化影响；后续公司也重新强调人工服务，说明客户体验与成本之间需动态平衡。可复制部分：高频任务、分层转人工、持续观察投诉与复杂问题。[S9][S10][S11]

案例4：海尔智家客服大模型（2024）

主体与情境：家电企业客服需要快速查询、准确指引与减少人工负担。动作：公司在客服领域应用大模型辅助问题处理。结果：海尔智家 2024 年年报披露平均会话时长缩短 15%。机制：知识检索与问题指引减少查询和处理时间。归因限制：公司年报未在该段披露样本、对照组、质量与投诉变化，无法把企业整体经营结果归因于该应用。可复制部分：以会话时长为效率证据，同时必须补充一次解决率、转人工、投诉和事实准确反指标。[S3]

案例5：海尔智家 AI 原生数字伙伴与营销应用（2025—2026）

主体与情境：企业希望把 AI 从单点工具扩展到客服、工艺和营销传播。动作：推出数字伙伴、轻应用和企业级应用；官方 2026 年披露在智能工艺、客服工作群组、营销种草视频传播等场景实践。结果：工艺规划案例披露编制时长由数小时压缩至分钟级、准确率提升 20% 以上；营销种草场景未披露可核验经营结果。归因限制：不同场景不可互相借用结果，营销效果应单独测量。可复制部分：员工从执行转为审核、低代码场景开发、企业级应用分层。[S4][S5]

案例6：湘钢集团钢铁行业大模型（2024—2025）

主体与情境：高监管、高可靠性的流程制造，强调质量、能耗和闭环控制。动作：与华为合作，把行业模型与现有系统集成，应用于视觉和预测场景。湖南省政府材料披露 23 个视觉类和 9 个预测类场景，关键工艺推荐准确率超过 95%、吨钢综合能耗下降约 4%、质量判定效率提升超过 50%。机制：行业数据、系统集成和现场闭环。归因限制：属于政府案例材料，具体基线、持续周期和各指标分母未完整公开；制造结果不能直接外推营销。可复制部分：先建立高价值单元，再逐场景扩展；强监控、系统集成与专业人员审核。[S6]

案例7：P&G/AI 团队实验对“规模化人数”的反例

同一宝洁实验同时揭示限制：AI 改善个人和团队表现，并不等于减少所有协作。研究显示 AI 改变知识共享和互动结构，最佳表现来自任务、人员与 AI 的组合。管理含义：若企业只按账号数部署，而不重构角色、评审和协作，可能获得更多产出但没有更好的决策。可复制部分是团队设计和任务边界，而非“一个人替代两个人”的口号。[S1][S2]

案例8：大规模人机广告协作实验（2025）

主体与情境：Aral 与 Ju 的现场实验平台让 2,310 名参与者组成人人或人机团队，完成广告创作，并在约 500 万次曝光的活动中测试。动作：随机配置团队与 AI 特征，记录消息、文本、图像和广告结果。结果：论文报告人机团队提高每名员工生产率，文本质量受益，但人人团队的图像质量更高；最终广告表现整体相近，且不同 AI 特征与人的组合影响结果。归因限制：预印本、特定平台与任务，尚不代表企业长期组织绩效。可复制部分：细粒度日志、分模态质量、真实投放验证、观察互补而非只看平均产量。[S7]

案例9：中山六院生殖中心智能客服（2025）

主体与情境：医疗服务咨询具有高频、敏感和错误成本高特点。广州市天河区政府公布的典型案例材料称，项目通过 AIGC 与自然语言处理提供 24 小时服务，问答准确率为 98.5%。机制：标准咨询自动响应、持续优化。归因限制：政府新闻未披露评测样本、错误定义、转人工规则和患者结果；“准确率”不能替代医疗安全与责任。可复制部分：高监管场景必须把转人工、敏感问题禁答、审计和持续评测放在第一阶段。[S12]

案例10：钛动科技出海营销 AIGC（2025）

主体与情境：出海营销需要跨语言创意、渠道适配和快速测试。天河区政府材料将钛动科技列为“人工智能+”典型案例，并称其较早出海营销场景落地 AIGC。结果证据：公开材料未披露具体客户、样本、增量收入或成本口径。归因限制：只能证明场景实践与政府遴选，不能据此推断 ROI。可复制部分：将跨语言素材生产与真实投放反馈连接；在企业试点中必须补充对照和单位经济性。[S12]

经典管理思想的当代修正

思想	原始要义	AI情境修正	管理动作
Drucker 结果观	组织以外部结果而非内部活动证明价值	生成量和使用率只作领先信号	把客户、现金、质量与风险写入章程
Popper 可证伪	主张需存在可推翻条件	AI假设必须附停止条件和反指标	预先写明何种事实会改变决定
Ries 验证学习	通过最小实验验证关键假设	最小对象需包含任务集和真实闭环	14天测任务，30天测运营
Deming 系统与变异	理解系统和波动再管理数字	模型、输入与评审存在版本变异	记录版本、分层样本与误差类型
Goldratt 约束理论	改善系统约束才改善整体结果	局部生成速度可能扩大下游堵塞	先识别获客、审批或跟进约束
Cooper Stage-Gate	分阶段评审并配置资源	闸门要审可复现证据、日志与风险	允许四种决议并绑定下一笔资源
March 探索/利用	探索新知与复制既有能力需平衡	试点团队与运营团队逻辑不同	设立接管门和双环学习
实物期权	不确定下保留扩张、收缩与退出选择	集成和采购应随证据递增	小额可逆投资，延迟不可逆承诺

第十二章 董事会决议页

竹势 AI 营销智库出品

董事会只做四项决议

继续配置资源 · 调整假设与范围 · 暂停等待依赖 · 终止并回收资产

董事会不需要批准每个技术细节，但应对四项事项作出明确决议：问题许可、真实试点许可、规模化许可和退出/回退。每项决议都应记录证据、反对意见、风险接受人、下一笔资源和复评日期。

决议一：问题许可

确认问题是当前经营约束；基线可复核；AI 贡献假设可证伪；风险边界和 owner 明确。决议选项：进入 30 天试点 / 重构后复评 / 停止。

决议二：真实试点许可

确认真实闭环、RACI-O-X、人工控制、数据用途、反指标、异常与回滚方案。决议选项：继续 / 调整 / 暂停 / 终止。

决议三：规模化许可

确认价值可重复、单位经济性可承受、运营接管通过、第二团队复制有效、审计和供应商管理可用。决议选项：全面许可 / 有条件许可 / 延长验证 / 退出。

决议四：退出与资产回收

确认停机、原流程恢复、数据和权限撤销、外部纠正、资产回收、根因复盘和二次假设。终止项目不自动否定 AI；它否定的是当前问题—方案—控制组合。

BOARD BRIEF 16

管理工具附录

竹势 AI 营销智库出品

工具1：试点场景一页章程

字段	填写要求
经营问题	用客户、现金、成本、质量或风险语言描述，不写技术目标
当前约束	指出系统瓶颈及下游承接能力
范围与排除	任务、用户、渠道、地区、数据和不做事项
基线	时间范围、样本、分母、数据 owner 与缺陷
AI贡献假设	可被事实推翻的完整句子
五类价值证据	效率、质量、增量结果、长期资产、风险
反指标	不可因追求主指标而恶化的指标
角色责任	业务、技术、风险、数据、运营、异常负责人
进入条件	进入真实试点所需必要证据
停止/回退	即时停止、阶段停止、恢复原流程与资产回收

工具2：14天任务日历

日	任务	输出
D1	确认经营问题与系统约束	问题草案
D2	确认角色、风险与停止条件	签字章程
D3	抽取历史样本	样本清单
D4	重建基线与口径	基线表
D5	设计标准与边界任务	任务集V1
D6	定义盲评与错误分类	评分规则
D7	离线运行A	输出与日志
D8	离线运行B/对照	成对比较
D9	记录总流程时间与人工编辑	效率证据
D10	红队与敏感任务	风险样本
D11	权限、隐私、事实核验测试	控制清单
D12	证据树与归因复核	证据包
D13	形成决议建议与二次假设	决议草案
D14	问题许可门	正式决议

工具3：最小任务集与基线表

任务ID	任务类型	历史频率	当前周期/成本	当前质量	AI输出	人工编辑	严重错误	结论
示例-01	标准高频	填写真实数据	填写	填写	填写	填写	填写	通过/调整/停止
示例-02	边界例外	填写真实数据	填写	填写	填写	填写	填写	通过/调整/停止
示例-03	敏感高风险	填写真实数据	填写	填写	填写	填写	填写	通过/调整/停止

工具4：证据树

层级	问题	证据
经营结果	客户/现金/成本/风险是否改变	结果指标与归因设计
行为与流程	关键岗位是否采用、闭环是否运行	日志、周期、异常、人工桥接
输出质量	是否达到可用与合规标准	盲评、规则、严重错误
AI贡献	相对基线或对照有何增量	成对比较、分组、交叉或随机实验
前提条件	数据、权限、角色和系统是否成立	数据质量、责任签字、控制测试

工具5：30天 RACI-O-X

工作项	业务Owner	技术Owner	风险Owner	数据Owner	运营接管O	异常X
问题与指标	A	C	C	C	I	I
workflow配置	C	A/R	C	C	I	I
真实任务审批	A/R	C	C	I	R	C
异常停机	I	R	A/C	I	R	A
价值评审	A	C	C	R	C	I
运营接管	A	C	C	C	A/R	R

工具6：人工控制点与异常清单

控制点	触发条件	人工动作	日志	回退
对外发布	所有高风险内容/抽样低风险内容	核验事实、品牌、法律与渠道	审批人、版本、时间	撤回/更正
客户触达	敏感、投诉、复杂购买决定	接管对话并确认上下文	转人工原因与时延	恢复人工队列
数据写回	关键字段、批量更新、删除	预览差异并批准	前后值与操作者	恢复快照
预算/投放	任何资金或受众重大变化	仅生成草案，授权人执行	建议、批准、执行结果	恢复原配置
系统异常	超时、错误率、接口失败	停用自动任务、切换原流程	故障与影响范围	版本回滚

工具7：阶段门评分卡

域	权重制定方法	证据等级	否决条件
问题价值	由董事会按当前约束设定	主张/样例/对照/真实结果	问题不再重要
贡献	按场景核心价值设定	离线/真实/复制	贡献不可识别
风险	按最坏后果设定	控制设计/演练/运行证据	不可接受风险无控制
经济性	按单位经济性设定	估算/实测/规模实测	边际成本长期高于价值
运营性	按接管复杂度设定	参与/独立运行/跨团队	无人接管或无法回退

工具8：停止与回退清单

- ☐ 冻结新任务并保存当前版本、输入、输出和日志
- ☐ 按事故等级通知业务、技术、风险、法务和客户责任人
- ☐ 切换原流程并验证服务连续性
- ☐ 撤回、更正或人工复核已产生的外部影响
- ☐ 撤销临时权限、密钥、账号和数据副本
- ☐ 核算影响、完成根因分析和监管/客户报告义务
- ☐ 回收任务集、数据字典、控制规则、接口和失败样本
- ☐ 记录重启条件或正式终止理由

工具9：90天可复制工作单元模板

模块	内容
目标与边界	解决何种问题，适用/不适用范围
触发与输入	触发事件、字段、质量和数据责任
角色	AI、执行、批准、复核、运营、异常
步骤与接口	顺序、系统、版本、人工桥接
标准	质量、时限、格式、事实与品牌要求
控制	审批、限额、白名单、抽样、转人工
例外	类别、识别规则、升级与处置
监控	主指标、反指标、日志、告警与复评
成本	固定、可变、人工、风险与退出成本
变更与退出	版本管理、再验证、停机与资产归属

工具10：成本与价值台账

周期	任务量	可归因价值	平台/模型	集成摊销	人工复核	运营与治理	异常损失	净值	归因等级
填写	填写	填写	填写	填写	填写	填写	填写	公式计算	前后/对照/实验

工具11：运营接管清单

- ☐ 运营负责人和备份人员已命名
- ☐ 独立完成真实任务启动与复核
- ☐ 识别至少三类异常并正确升级
- ☐ 完成停机与原流程恢复演练
- ☐ 能够查询日志、成本、版本和权限
- ☐ 数据更新、知识过期和模型变更有责任人

- ☐ SLA、值班、供应商支持和事故联系人明确
- ☐ 月度价值复评与季度再许可已进入经营节奏

工具12： 董事会四项决议模板

决议	事实基础	可选结果	必须记录
问题许可	章程、基线、任务集、风险	进入/重构/停止	反对意见、下一笔资源
真实试点许可	闭环、RACI、控制、回滚	继续/调整/暂停/终止	风险接受人、复评日期
规模化许可	复制、成本、接管、审计	全面/有条件/延长/退出	范围、预算、停止条件
退出与回收	事故/价值/成本/责任证据	回退/终止/二次假设	资产、责任、纠正与重启条件

专题深化：把证据轨道嵌入经营节奏

竹势 AI 营销智库出品

一、把阶段门嵌入预算与经营会，而不是另建一套创新仪式

证据许可制只有进入现有经营节奏，才会成为管理系统。建议把问题许可门放入月度经营会或专项投资会，把真实试点许可与业务单元的预算和责任确认连接，把规模化许可纳入季度资源配置。这样做能避免两套治理并存：一套由创新团队展示项目，一套由经营团队分配资源。每次评审材料应控制在一套事实包内，包括基线、证据树、风险、成本、反对意见、下一阶段资源和停止条件。演示可以作为补充，不能替代数据与责任。

预算也应按证据成熟度分层。14 天预算主要用于业务专家时间、历史样本整理、评测与风险测试；30 天预算用于最少真实接口、运营人员、人工控制和异常演练；90 天预算才承担稳定集成、监控、培训、支持和跨团队复制。若企业在第一阶段就购买大批账号或长期平台合同，项目团队会产生“必须找到用途”的反向激励。更稳健的做法是把合同里程碑与证据门绑定，并保留缩减、迁移和退出条款。

二、如何为低频、高价值和长周期营销任务调整节奏

并非所有营销任务都能在 30 天内产生足够样本。品牌定位、年度传播、重大新品、渠道拓展和战略客户营销具有低频、高价值、长反馈的特点。此类场景仍可使用证据许可制，但要改变证据组合：14 天阶段更多依赖历史项目回放、专家盲评、失败模式和可用性测试；30 天阶段可验证关键子任务，例如洞察整理、版本生成、合规审查、销售协同和决策周期；最终市场结果需要跨季度观察。管理层应明确哪些结论只能是“方向性证据”，不能把短期点击或内部评分包装成长期品牌价值。

对于低频任务，最小任务集可以由多个历史事件构成，并采用时间切片回溯：团队只使用当时可获得的信息，让 AI 和原流程分别完成任务，再由不知来源的专家评审。回溯能提高学习速度，但仍存在历史选择偏差和“知道答案”的污染风险。进入真实试点后，应保留前瞻性记录，避免只选择成功案例。

三、场景组合管理：允许单个试点停止，保护整体学习

董事会通常同时面对多个 AI 营销候选场景。组合管理的目标不是让每个项目都成功，而是让有限资源获得最多可用知识和经营价值。可把场景分为三类：防御型，主要降低合规、服务和运营风险；延展型，改善现有流程的效率与质量；探索型，测试新的客户体验、渠道或商业模式。三类场景的证据门应一致，风险容忍和观察周期可以不同。

组合层面应避免高度相关的风险。例如，多个试点若依赖同一外部模型、同一数据源和同一接口，表面上是十个场景，实质上是一个集中风险。组合台账应记录共享依赖、关键人员、数据域、供应商、监管类别和退出成本。规模化许可不仅要看单个单元通过，还要看并发运行后的成本、权限复杂度、监控负担和事故关联。

停止一个试点可以提高组合回报。被终止项目释放的业务专家、数据和预算应明确回流到优先级更高的场景；评审会应报告“停止带来的避免成本”和“回收资产”，而非只报告损失。这样才能弱化团队对终止的抵触，并把证伪转化为组织学习。

四、证据质量的分级与冲突处理

建议将证据按决策强度分为五级。一级是主张和样例，只能支持继续探索；二级是可复现离线评测，可支持进入受控试点；三级是真实流程运行，支持运营许可；四级是同期对照或实验，支持更强的增量判断；五级是跨团队、跨周期重复，支持规模化。证据等级不是对研究方法的炫耀，而是限制结论边界。一个真实客户案例可能比大样本问卷更贴近本企业流程，但不足以证明普遍因果；一个随机实验因果更强，也可能因任务差异而难以外推。

当证据冲突时，不应简单加权平均。应先检查口径、分母、样本构成、模型版本、评审者和时间窗口，再判断冲突是否来自真实权衡。例如，效率改善而质量下降，说明流程把成本转移给客户或审核人员；内部满意度提高而客户转化不变，说明价值停留在员工体验；离线质量高而真实采用低，说明工作流、信任或激励存在约束。冲突本身通常比平均分更有管理价值。

五、模型与供应商变更后的再许可

AI 能力、价格和服务条款会变化。已获得规模化许可的工作单元不能因此永久免检。以下变化应触发不同程度的再许可：基础模型或主要版本变化；知识库、检索或安全策略重大调整；关键数据源变化；自动化权限扩大；供应商托管地点、条款或价格变化；渠道平台规则变化；事故或近失事件显著增加。再许可不必重复全部 90 天，但必须针对变化影响重新跑关键任务集、风险测试、成本与回退演练。

企业应保留“黄金任务集”和“事故任务集”。黄金任务集代表正常业务质量，事故任务集包含过去真实发生或差点发生的高风险错误。每次版本变更都要比较退化、改善和新错误类型。只有平均分提高而事故任务恶化时，不应自动上线。版本台账需记录上线时间、适用工作单元、批准人、回退版本和观察期。

六、把长期资产纳入经营账，而不制造虚假估值

AI 营销项目会沉淀任务集、数据字典、客户语言、内容规范、异常分类、接口与运营能力。它们确实有长期价值，但价值容易被夸大。建议采用“资产可用性”而非随意货币化：是否有明确 owner，是否结构化，是否有版本与更新时间，是否被第二个工作单元复用，是否降低后续设计或培训时间，是否能在供应商退出后迁移。只有满足这些条件，资产才可进入规模化证据。

长期资产也有负面面。过期知识会系统性制造错误；大量低质量内容会污染品牌资产；高度绑定单一供应商的工作流会增加退出成本；过度依赖少数专家维护会形成关键人风险。因此资产台账必须同时记录维护成本、过期风险、可迁移性和弃用条件。规模化许可应确认企业有能力维护这些资产，而不是只确认它们已经被创建。

七、从首个闭环到场景群：规模化边界的三次验证

第一轮复制验证任务是否可重复，第二轮验证组织是否可接管，第三轮验证组合是否可治理。很多企业只完成第一轮：同一个项目团队把流程复制到另一个部门，于是得出“已经规模化”的结论。真正的组织复制应让目标部门承担自己的数据、审批、异常与成本责任，并在项目团队退出后维持运行。组合复制则进一步检验多个工作单元并发后，是否出现共享接口拥塞、权限膨胀、监控盲区、模型成本激增或供应商支持不足。

跨区域或跨品牌复制时，应把差异拆成不可变控制、可配置参数和必须重设计三类。不可变控制包括隐私、审计、事故升级和关键审批；可配置参数包括语言、品牌语气、渠道格式、服务时限和数据字段；必须重设计的内容包括当地法规、客户旅程、组织权责和核心系统接口。把所有差异都做成参数会产生复杂而脆弱的系统，把所有流程都重新开发又会失去规模经济。标准工作单元的价值正是帮助管理层判断哪一层应统一。

八、董事会如何识别“假快”和“假规模化”

假快通常有五个信号：用模型响应时间替代端到端周期；用精选样例替代任务集；用项目团队人工修补却不记录；跳过基线和反指标；在权限、数据和运营责任未就绪时宣布上线。假规模化也有五个信号：账号数量增长但真实任务量不增；产出数量增长但客户结果不变；第二团队依赖原团队救火；成本和异常没有按任务分层；核心流程无法在规定时间内回退。

董事会可以要求每个项目回答三个反事实问题：没有 AI 时本期结果会怎样；若更换团队或模型，结果是否仍成立；若明天停止系统，业务能否连续运行。不能回答并不等于项目必然失败，但说明证据不足以支持更大承诺。评审材料还应展示最差样本、近失事件和未解决依赖，防止只报告平均值和最佳案例。

九、90天之后的持续许可机制

规模化许可不是终点。进入常态运营后，建议建立月度运行复核、季度价值再许可和年度组合重评。月度复核关注任务量、质量、异常、人工负担、成本与版本变化；季度再许可重新判断价值、风险和运营能力；年度组合重评则决定扩张、合并、替换、降级或退出工作单元。对高风险场景，事故、法规变化或重大模型变更应触发即时复评。

持续许可要避免把运营团队拖入无休止汇报。数据应尽量来自运行日志、审批记录、成本账单和业务系统，人工只补充解释与决策。监控指标应有明确用途：某个指标若连续多个周期未影响任何决策，应考虑删除；某类事故若重复发生，应升级为控制规则或设计变更，而非持续依赖提醒。最终目标是让证据进入日常经营，而不是制造一个独立的 AI 管理官僚体系。

十、实施中的关键会议与最小信息包

14 天阶段至少需要三类短会：问题校准会、评测复核会和许可门。问题校准会只讨论经营约束、基线和假设，不做产品演示；评测复核会逐项检查任务集、错误类型、人工编辑和风险；许可门只做资源与责任决议。30 天阶段增加每周异常复盘，重点查看最差样本、人工桥接和反指标，不把会议时间消耗在逐条展示正常产出。90 天阶段增加运营接管会和复制差异会，前者确认谁能独立运行，后者确认哪些差异需要参数化或重设计。

每次会议的最小信息包应保持一致：本阶段决策问题；相对基线的变化；证据等级与归因边界；最严重错误和近失事件；全生命周期成本更新；未解决依赖；建议决议及其反对意见；下一阶段预算、责任、停止条件和复评日期。统一信息结构能减少部门之间的语言转换，也能防止团队在证据不足时用更多页面掩盖不确定性。

十一、对中小企业与大型集团的不同适配

中小企业通常人员少、系统简单、决策快，可以缩短责任确认和采购环节，但更容易缺少历史基线、专职数据和风险人员。其策略应是选择高频、靠近现金结果的单一闭环，以人工桥接替代过早集成，并由老板或业务负责人直接承担问题 owner。大型集团拥有更多数据与专业职能，却面临跨部门权责、采购、合规和系统依赖，14/30/90 的日历可能需要拉长；其优势是可以建立更强对照、独立评审和第二团队复制。

两类企业都不应省略闸门。中小企业省略闸门会把有限资金投入错误场景；大型集团省略闸门会让多个部门在缺乏责任和价值证据的情况下扩大技术债。差异应体现在证据获取方式、会议层级和日历长度，而非是否需要基线、反指标、回退和运营接管。

十二、管理层的最后检查：任何扩张都应回答六个问题

在批准下一阶段前，管理层应逐一确认：第一，当前证据证明的是输出更好、流程更快，还是经营结果改变，结论边界是否被如实表达；第二，最差样本和近失事件是否进入评审，而非只展示平均值；第三，人工复核、数据准备、异常处理和供应商支持是否计入成本；第四，业务、技术、风险、数据与运营责任是否落实到具体姓名和替补；第五，系统能否在约定时间内停止并恢复原流程；第六，第二个团队是否可以依据标准工作单元独立复现。任何一个问题只能用“未来会补”回答时，都说明扩张许可仍缺少必要证据。

这六个问题也构成对项目叙事的压力测试。一个项目可以在效率上通过、在经营增量上保持待观察；可以在单团队运行上通过、在跨团队复制上暂缓；可以在低风险任务上扩大、在高风险动作上继续影子运行。许可应允许这种分层，而不是用“成功”或“失败”概括全部事实。分层许可既保护创新，也保护企业不因局部亮点承担不必要的系统性风险。

PUBLISHER

关于竹势 AI 营销智库

竹势 AI 营销智库出品

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。

本报告的判断框架、责任边界和实施工具，旨在帮助管理团队用受控方式验证 Agentic Marketing 的价值，并在规模化之前建立必要的证据、权限与问责机制。

shichangbu.ai

专题中心 · 返回顶部 ↑

竹势 AI 营销智库出品

来源索引

竹势 AI 营销智库出品

[S1] Fabrizio Dell’Acqua 等, A Field Experiment on Generative AI Reshaping Teamwork, NBER Working Paper 33641, 2025。预注册现场实验；工作论文摘要披露 776 名宝洁专业人员。

[S2] Fabrizio Dell’Acqua 等, A Field Experiment on Generative AI and Teamwork, Organization Science, 2026。正式发表版本摘要披露 791 名宝洁专业人员。

[S3] 海尔智家股份有限公司, 2024 年年度报告, 2025-03-27。客服大模型平均会话时长缩短 15%，公司披露口径。

[S4] 海尔智家股份有限公司, 2025 年半年度报告, 2025-08-28。自研数字营销大模型与用户触达、转化和沉淀等公司披露。

[S5] 海尔集团, 海尔智家：从数字化企业到 AI 原生企业跃迁探索, 2026-06-22。数字伙伴、客服、营销传播与工艺案例，公司披露口径。

[S6] 湖南省人民政府, 湖南省贯彻落实国务院“人工智能+”行动相关材料, 2025-10-30。湘钢行业大模型场景与指标，政府案例材料口径。

[S7] Harang Ju, Sinan Aral, Collaborating with AI Agents: Field Experiments on Teamwork, Productivity, and Performance, 2025。2,310 名参与者与约 500 万曝光的预印本研究。

[S8] Morgan Stanley, Launch of AI @ Morgan Stanley Debrief, 2024-06-26。客户同意、会议记录、行动项、人工编辑邮件与 Salesforce 写回。

[S9] Klarna, AI assistant handles two-thirds of customer service chats in its first month, 2024-02-27。首月 230 万次对话，公司披露口径。

[S10] Klarna, Q1 2024 results, 2024-05-30。AI 助手客户覆盖与年度化节省估算，公司披露口径。

[S11] Klarna, Q1 2025 update, 2025-05-19。客服单笔成本变化，公司披露口径。

[S12] 广州市天河区人民政府, 广州第二批“人工智能+”典型案例公布, 2025-07-08。钛动科技出海营销与中山六院智能客服案例，政府新闻口径。

[S13] McKinsey & Company, The State of AI: Global Survey 2025, 2025-11-05。多数组织仍处于试验或试点阶段，调查口径。

[S14] MIT CISR, Grow Enterprise AI Maturity for Bottom-Line Impact, 2025-08-21。企业 AI 成熟度与财务表现研究简报。

[S15] Robert G. Cooper, Stage-Gate 方法相关著作与机构材料。用于阶段化创新治理思想，不作为 AI 成功率数字来源。

[S16] NIST, Artificial Intelligence Risk Management Framework 1.0, 2023；Generative AI Profile, 2024。治理、映射、测量和管理框架。

[S17] European Union, Artificial Intelligence Act, 2024 及后续适用安排。风险分层与治理义务。

[S18] 国务院, 关于深入实施“人工智能+”行动的意见, 2025。推动 AI 融入组织、业务与产业应用。

[S19] 工业和信息化部等部门, 中小企业数字化转型支持专项行动方案（2025—2027年）。强调真实场景、数字化转型与生态建设。

[S20] Peter F. Drucker, The Practice of Management；Karl Popper, The Logic of Scientific Discovery；W. Edwards Deming, Out of the Crisis；Eliyahu Goldratt, The Goal；James March, Exploration and Exploitation in Organizational Learning；Daniel Kahneman, Thinking, Fast and Slow。经典思想用于管理机制解释。

BOARD BRIEF 19

结语

竹势 AI 营销智库出品

让正确项目更快，让错误项目更早停止

证据许可制把创新速度、经营纪律与可逆风险放进同一套管理节奏。

企业不应以部署数量证明 AI 决心，也不应以一个月收入结果否定所有长期价值。更成熟的做法是把不确定性拆成连续可验证的问题：问题是否值得解决，基线是否可信，AI 是否有贡献，真实闭环是否稳定，风险是否可恢复，价值是否能被另一个团队重复。14 天、30 天和 90 天提供的是讨论这些问题的共同节奏。

证据许可制的最终目标不是让项目更容易通过，而是让正确项目更快获得资源，让错误项目更早停止，让风险项目在可回退条件下学习，让成功单元在运营和治理接管后再扩张。当企业能把继续、调整、暂停和终止都视为正常决议，AI 营销才真正进入经营管理，而不再停留在技术热潮。