竹势

039 / SKILL-COMPONENT FOUNDRY

AI MARKETING SKILLS / CONTRACT · TEST · GOVERN

技能不是提示词

企业 AI 营销能力的工程化方法

把隐性经验封装成能力契约：可发现、可授权、可测试、可组合、可升级，也可安全退役。

划清六层能力边界

知识、提示词、工具、技能、工作流、智能体各归其位

签订十一格技能契约

输入输出、权限依赖、失败与证据完整

建立七层测试证据

从单元测试一直追到经营结果

完成 90 天有限发布

影子、回放、灰度和目录治理后再扩权

shichangbu.ai

竹势 AI 营销智库出品

BOARD BRIEF 01

执行摘要 | 十项高管判断

竹势 AI 营销智库出品

SKILL ENGINEERING THESIS / 董事会判断

技能的价值不在于写得多，而在于能被发现、授权、调用、验证、组合和升级；它必须是一份可执行、可审计的能力契约。

判断	董事会含义
一	技能的管理对象是“可重复兑现的能力”，不是提示词文本。提示词可以是实现材料，但不能单独承担权限、错误、证据和版本责任。
二	技能必须以业务任务为入口。高频、相对稳定、可验收且责任边界清楚的任务，优先于“看起来聪明”的开放式任务。
三	粒度决定复用经济性。过粗会形成黑箱和平台锁定，过细会造成编排成本、调用噪声和责任碎片化。
四	技能契约至少覆盖输入、输出、上下文、先决条件、权限、工具、依赖、资源、失败语义、证据和服务水平。
五	目录不是清单，而是企业能力控制面。名称、语义、所有者、信任等级、兼容版本和运行证据必须可检索。
六	最小权限必须落实到每次调用。敏感数据、外发内容、预算调整、客户触达和账号操作需要分级授权与人工门。
七	跨技能组合要遵循类型兼容、状态传递、幂等、超时、重试和补偿规则；内部步骤不应伪装成可复用技能。
八	没有测试证据的技能只能进入实验区。单元、契约、回归、安全、红队、业务验收与线上观测共同形成发布依据。
九	技能、提示词、工具接口、模型和政策规则应分别版本化。灰度、影子、回放、回滚和弃用窗口是经营连续性的保障。
十	董事会不应审批每个技能，但应批准风险偏好、信任梯、发布门、重大事故阈值和 90 天资产化投资。

这十项判断把“会不会写提示词”的个人生产力问题，转化为“企业是否拥有可持续能力资产”的治理问题。技术团队关注调用和运行，营销团队关注业务验收，风险团队关注权限与证据，管理层则必须确保这些视角进入同一份能力契约。

本报告不把任何一家厂商的 Skill 文件格式视为行业统一标准。OpenAI、Anthropic、GitHub、MCP 及云平台展示了不同的能力封装方式；企业需要在这些实现之上建立中立的注册、契约、测试、授权与生命周期治理。

BOARD BRIEF 02

第一章｜经营命题：把散落经验变成企业能力资产

竹势 AI 营销智库出品

REUSE
复用覆盖
一次封装，多团队稳定调用

CHANGE
变更成本
接口稳定，内部可迭代

RISK
风险半径
权限、数据与动作都有边界

营销组织的经验通常分散在员工记忆、提示词收藏、SOP 文档、供应商模板、自动化脚本和各类 SaaS 配置中。它们能够在熟练员工手里产生结果，却缺乏稳定入口、责任边界和可验证的输出。员工离开、模型升级、接口变化或渠道规则调整时，组织只能重新摸索。

技能工程要解决的不是“怎样让模型更聪明”，而是怎样把一个业务能力定义为可被发现、受控调用、组合执行和持续改进的组件。能力一旦进入企业目录，就必须能回答：谁拥有它，何时允许调用，依赖哪些数据与工具，失败时如何停止，结果如何验收，升级后如何证明没有破坏既有业务。

因此，技能资产化的经济价值来自三条路径：减少重复设计与返工；降低关键人员和单一供应商依赖；把运行证据转化为可持续学习。它既是 AI 平台工程，也是营销运营制度。只做技术封装会失去业务验收，只做业务模板会失去安全和可维护性。

散落载体	当前价值	主要缺陷	工程化处理
员工经验	情境判断强	不可发现、不可继承	提炼决策规则与例外边界
提示词	启动成本低	输入输出不稳定	纳入技能实现并建立回归集
SOP	流程可见	无法直接执行	拆分为能力契约与跨技能工作流
脚本/API	执行确定	业务语义不足	补充描述、权限、证据和所有者
供应商服务	可快速交付	锁定和知识产权不清	明确接口、数据归属和退出包

董事会问题｜未来十二个月，哪些营销能力必须由企业自己拥有，哪些可以采购，哪些必须保留人工判断?

BOARD BRIEF 03

第二章 | 六层能力边界：知识、提示词、工具、技能、工作流与智能体

竹势 AI 营销智库出品

01	02	03	04	05	06
知识	提示词	工具	技能	工作流	智能体

六类对象经常被混用，导致采购和治理责任错位。知识回答“事实和规则是什么”；提示词组织模型的注意力与表达；工具提供外部系统动作；技能承诺在给定条件下完成一个可验收能力；工作流协调多个技能和人工节点；智能体围绕目标选择与调度能力。它们可以组合，但不能相互替代。

层级	核心问题	最小治理对象	典型营销例子
知识	知道什么	来源、有效期、访问域	品牌规范、产品事实、渠道规则
提示词	怎样引导模型	模板、变量、示例、模型适配	把 Brief 转成公众号初稿
工具	能调用什么	接口、认证、权限、限额	读取 CRM、创建企微草稿
技能	能稳定完成什么	能力契约、测试、版本、证据	品牌审校、投放诊断
工作流	怎样跨步骤交付	状态、分支、人工门、补偿	活动 Brief 到复盘闭环
智能体	如何围绕目标选择行动	目标、规划、路由、监督	营销运营智能体

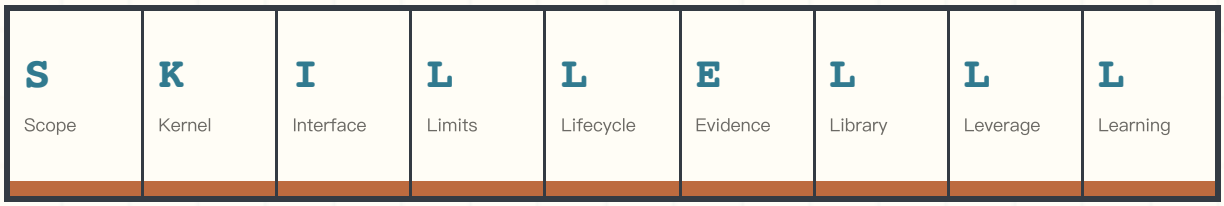
边界判断的关键是“独立可验收”。品牌审校可以被单独调用并输出问题清单、风险级别和证据，因此适合成为技能；“先查品牌库、再比对禁词、再生成修改建议”只是技能内部实现步骤，不应全部暴露为目录项。活动从 Brief、内容、审批、发布到复盘涉及多种责任和状态，应定义为跨技能工作流。

误判	后果	纠正
把提示词当技能	无法声明权限与失败语义	将提示词放入实现层，外部以契约调用
把 API 当技能	技术接口缺少业务语义	增加业务输入、验收与证据
把工作流当技能	黑箱过粗、难以复用	拆出可独立验收能力
把智能体当技能库	路由不可解释	目录与调度器分离

BOARD BRIEF 04

第三章 | SKILL-9 能力工程链

竹势 AI 营销智库出品



SKILL-9 是本报告提出的能力资产化主链。九个维度不是线性瀑布，而是一组必须共同闭环的经营控制点。Scope 确定业务边界，Kernel 固化最小专业内核，Interface 提供稳定契约，Limits 声明不能做什么，Lifecycle 管理版本与退役，Evidence 形成测试和运行证据，Library 建立目录与发现，Leverage 衡量复用与经营价值，Learning 负责从事故和结果中更新。

维度	关键问题	强制产物	失控征兆
Scope	为谁解决什么任务	场景、责任、验收边界	一个技能包打天下
Kernel	不可替代的专业判断是什么	规则、示例、算法、工具编排	只有长提示词
Interface	调用者如何正确使用	输入输出 schema、错误码	靠口头说明
Limits	哪些场景禁止或降级	权限、数据域、人工门	默认全权限
Lifecycle	怎样发布、升级、弃用	版本、兼容窗、回滚包	上线即永久
Evidence	凭什么信任	测试集、审批、日志	只看演示
Library	怎样被发现和路由	目录、标签、语义描述	重名与漂移
Leverage	是否产生复用经济	覆盖率、周期、返工、成本	调用多但无业务价值
Learning	如何持续变好	复盘、数据反馈、变更提案	事故不改系统

SKILL-9 的管理意义在于防止“局部技术成功、整体经营失败”。例如内容生成技能可能在单次测试中表现优秀，但如果没有 Limits，它可能读取不应访问的客户数据；没有 Evidence，品牌团队无法证明升级没有破坏语气；没有 Library，其他团队会重复创建近似技能；没有 Learning，线上返工只会变成人工补救。

BOARD BRIEF 05

第四章｜技能发现与粒度：从高频任务到能力候选

竹势 AI 营销智库出品

ELEVEN-CELL SKILL CONTRACT / 十一格技能契约

目标 · 输入 · 输出 · 上下文 · 先决条件 · 权限 · 工具 · 依赖 · 失败语义 · 证据 · SLA

候选技能应来自真实工作，而不是从模型功能列表倒推。优先筛选高频、相对稳定、可验收、输入可获得、责任边界清楚且具有跨场景复用潜力的任务。开放式战略判断可以被技能辅助，但不宜在证据不足时承诺自动完成。

筛选维度	高优先特征	低优先或暂缓特征
频率	周级或日级重复	一年一次且强情境
稳定性	规则变化可跟踪	目标和口径持续漂移
可验收	结果有标准或对照	只能凭主观喜欢
输入可得	数据字段和来源明确	依赖未授权隐性信息
风险	可回滚或可人工门控	不可逆高声誉动作
复用	多团队共享同一能力	单人一次性技巧

技能粒度二维判定矩阵

粒度可由“业务独立性”和“实现耦合度”两个轴判断。业务独立性高、实现耦合度低的能力最适合作为企业级技能；业务独立性低的步骤应留在技能内部；高度耦合且跨责任域的任务通常应由工作流协调。

业务独立性 / 实现耦合	低耦合	高耦合
高独立性	企业级技能：品牌审校、线索评分	领域技能包：投放诊断，需声明平台依赖
低独立性	内部函数：格式清洗、字段映射	工作流步骤：跨系统同步、审批等待

过粗损失	表现	经营后果
黑箱	一个技能覆盖研究到发布	难测试、难授权、难替换
锁定	依赖单一平台内部对象	退出成本高
责任混合	策略、执行、审批同体	事故责任不清

过细损失	表现	经营后果
调用膨胀	每个小步骤一个技能	成本和延迟上升
语义噪声	目录出现大量近义名称	错误路由
状态碎片	中间结果散落	补偿与审计困难

BOARD BRIEF 06

第五章｜十一格技能契约卡

竹势 AI 营销智库出品

契约卡是技能进入企业目录的最低门槛。它不是技术说明书的缩写，而是业务、平台和风险团队共同签字的能力承诺。每一格都需要机器可读字段与人类可读解释，避免“系统知道但管理者不知道”或“文档写了但运行时无法执行”。

契约格	最低声明
1 输入	字段、类型、必填、来源、敏感级别
2 输出	结构、质量标准、允许空值、可见范围
3 上下文	品牌、客户、活动、历史状态的选择规则
4 先决条件	授权、数据新鲜度、账户状态、依赖健康
5 权限	数据读取、写入、外发、预算、账号动作
6 工具	允许调用的接口、模型、执行环境
7 依赖	知识库、模型、第三方服务、政策规则
8 资源	超时、预算、并发、token、人工时间
9 失败语义	错误类别、可重试性、降级、补偿
10 证据	输入快照、决策理由、输出、审批、日志
11 SLA	成功率、时延、恢复、支持与责任人

“失败语义”必须具体到调用者能够采取动作。品牌知识缺失、模型拒答、工具超时、权限不足、输出不合规、业务规则冲突属于不同错误类别。把它们统一返回为“失败”会诱发盲目重试，既增加成本，也可能重复触达客户。

失败类型	是否重试	降级策略	证据要求
输入不完整	否	请求补充字段	缺失字段清单
依赖超时	有限重试	切备用或转人工	调用链与超时点
权限不足	否	走授权流程	主体、资源、动作
质量未达标	可重生成	替换模型或人工修订	评分与差异
外部动作不确定	禁止自动重试	先查询真实状态	幂等键与外部回执

BOARD BRIEF 07

第六章｜注册、目录、发现、路由与五级信任梯

竹势 AI 营销智库出品

T1
只读建议

T2
受限准备

T3
低风险执行

T4
人工批准

T5
禁止自动化

技能目录是企业 AI 能力的控制面。名称只是入口，真正用于发现和路由的是任务语义、适用对象、输入条件、风险等级、证据等级、成本和版本兼容。目录必须阻止近义重名、过度营销式描述和能力边界漂移。

目录字段	用途	治理规则
唯一标识	跨系统引用	不可复用、退役后保留墓碑
业务名称	人类检索	使用动宾结构，禁止夸大
语义描述	模型路由	包含适用与不适用场景
所有者	责任追踪	必须是岗位而非个人昵称
信任等级	授权和发布	由证据与风险共同决定
兼容范围	组合判断	声明输入输出和依赖版本
运行证据	持续评估	显示近期开销、失败和事故

五级信任与权限梯

级别	允许范围	典型证据	营销示例
T0 实验	隔离数据、不可外发	开发测试	新创意评估
T1 建议	只读、输出供人参考	基本单元与业务验收	投放建议
T2 草稿	可生成可编辑对象	契约、回归、安全测试	企微跟进草稿
T3 受控执行	限额写入、人工门	灰度、回放、审计	排期发布、CRM 更新
T4 自动执行	低风险可逆动作	长期线上稳定与事故记录	内部标签、日报生成

信任等级不是技能的永久荣誉，而是特定版本在特定数据域和动作范围内的许可。一个品牌审校技能可在内部文档上达到 T4，但面对高监管产品公开发布时仍可能只能达到 T2。目录应展示“版本—场景—权限”的组合，而非单一星级。

BOARD BRIEF 08

第七章｜权限、安全与依赖供应链

竹势 AI 营销智库出品

最小权限原则要求每次调用只获得完成当前任务所需的最小数据、工具和时间窗口。技能注册时声明最大权限，运行时再按主体、场景和动作缩小。永久令牌、共享账号和全库检索会把一个局部技能变成横向风险入口。

控制面	工程要求	审计证据
身份	人、智能体、技能与服务账户可区分	主体链路
数据域	品牌、客户、区域、项目隔离	访问资源清单
动作域	读、写、发布、删除、预算分开	动作与结果
时间域	短期令牌、按需授权	签发与失效时间
网络/文件	白名单、沙箱、路径约束	连接与文件摘要
人工门	敏感动作双人或角色审批	审批人、理由、版本

依赖供应链同样需要治理。技能可能引用第三方脚本、模型、连接器、知识包和规则库。企业应保留依赖清单、来源、许可证、哈希、维护者、漏洞状态和替代方案。未经锁定的远程依赖会在企业不知情时改变能力行为。

供应链风险	控制	退出要求
第三方脚本变化	版本锁定、签名、扫描	保留可验证副本
模型行为变化	模型版本与回归集	可切换兼容模型
API 变更	契约测试、适配层	数据导出与替代接口
知识产权不清	许可证与成果归属	源文件、提示、测试资产移交
供应商停服	业务连续性演练	降级流程与手工操作册

BOARD BRIEF 09

第八章 | 组合工程：类型、状态、幂等、超时、重试与补偿

竹势 AI 营销智库出品

类型兼容	状态传递	幂等	超时	重试	补偿

技能可组合的前提不是“都能返回文本”，而是输出类型、业务语义和状态责任兼容。内容 Brief 技能输出的受众、目标、渠道和禁区，需要以结构化字段传给内容生成；品牌审校应返回风险项及定位，而不是只给一段意见。类型兼容可以减少隐式猜测和上下文丢失。

组合规则	必须回答的问题	错误示例
类型兼容	字段和枚举是否一致	把自由文本当预算数字
状态传递	谁拥有当前业务状态	把聊天记录当发布状态
幂等	重复调用是否产生重复动作	重复创建客户或群发
超时	多久视为失败	等待第三方接口无限挂起
重试	哪些错误可安全重试	对不确定外发自动重试
补偿	部分成功如何撤销或修复	预算已改但记录未写回

技能内部步骤由技能所有者负责封装和优化，跨技能工作流则必须显式管理状态、责任和人工门。判断标准是：步骤是否需要被其他场景独立调用、是否有独立业务验收、是否跨越权限或责任边界。没有独立验收的内部步骤暴露为技能，只会把实现细节泄露到组织目录。

场景	内部步骤	跨技能工作流节点
品牌审校	分句、查禁词、比对语气	审校结果进入发布审批
投放诊断	拉数、计算异常、生成建议	建议审批后形成执行包
线索评分	字段清洗、规则计算	高分线索分配销售并触发跟进

BOARD BRIEF 10

第九章｜七层测试证据金字塔与线上观测

竹势 AI 营销智库出品

技能发布不能依靠少数漂亮示例。测试证据应从最小确定性单元逐层上升到真实业务结果：单元测试验证规则与转换；契约测试验证输入输出和错误；回归测试保护既有行为；安全测试验证权限和数据边界；红队测试探索诱导、越权和间接提示注入；业务验收验证岗位可用性；线上观测验证真实分布下的稳定性。

层级	测试对象	发布证据
1 单元	规则、解析、评分、模板	覆盖率与失败样例
2 契约	schema、错误码、幂等	消费者兼容结果
3 回归	历史金标与关键边界	版本差异报告
4 安全	权限、数据泄露、依赖	安全测试记录
5 红队	诱导、注入、越权组合	攻击路径与缓解
6 业务验收	真实岗位任务	验收人、标准、结论
7 线上观测	成功、返工、成本、事故	监控趋势与告警

对于生成式输出，测试不应追求每次字面一致，而应定义必须满足的结构、事实、品牌、合规和业务效果约束。可采用规则检查、模型评审、人类抽检和结果指标的组合，并保留评审模型版本与校准样本。

证据类型	适合判断	局限
确定性规则	字段、禁词、格式、权限	不能覆盖开放语义
金标对照	事实与关键建议	维护成本高
模型评审	大规模语义质量	存在偏差和漂移
人工抽检	高风险与细腻品牌判断	速度和一致性有限
业务结果	真实经营价值	归因滞后、受外部影响

BOARD BRIEF 11

第十章 | 技能版本栈、兼容性与安全发布

竹势 AI 营销智库出品

VERSION CABINET / 版本柜

技能 · 提示词 · 工具接口 · 模型 · 政策规则

影子和回放先积累证据，灰度缩小影响半径，兼容窗口保护调用者，退役流程清理僵尸资产。

技能行为由多个可变层共同决定：技能契约、提示词、工具接口、模型、知识与政策规则。如果只给“技能”一个总版本，事故发生后无法定位变化来源。企业应分别记录各层版本，并为组合生成不可变运行清单。

版本层	变化例子	兼容判断
技能契约	新增必填字段、改变错误语义	主版本变化
提示词/规则	措辞、示例、评分阈值	回归测试决定
工具接口	API 字段、认证、限额	契约测试
模型	供应商或模型快照	质量与安全回放
知识	品牌规范、产品事实	有效期与来源
政策	监管、平台、内部审批规则	即时生效与例外管理

兼容性矩阵

变更	向后兼容	发布动作
新增可选输出字段	通常是	次版本、消费者验证
删除字段	否	主版本、迁移窗口
错误码细分	视消费者而定	契约测试
模型替换	不保证	影子回放与灰度
政策收紧	业务上可能不兼容	立即控制并通知

影子运行让新版本接收真实输入但不执行外部动作；回放使用历史输入比较结果；灰度限制用户、数据域或流量；回滚必须包含契约、实现、依赖和配置；弃用应设置兼容窗口、替代技能和最终停用日期。对于公开触达和预算动作，禁止只凭离线评分直接全量发布。

阶段	允许动作	停止条件
影子	不写外部系统	关键差异超阈值
回放	历史数据离线比较	回归或合规失败
灰度	小范围受控执行	事故、返工或成本异常
扩大	按信任梯放量	SLA 下降
全量	持续观测	触发回滚阈值

BOARD BRIEF 12

第十一章｜治理角色、变更评审、事故与退役

竹势 AI 营销智库出品

技能资产需要清晰的经营责任。所有者对业务价值和生命周期负责；审批人对特定风险授权负责；平台团队保障运行、目录和基础控制；使用者遵循输入与验收要求；品牌、法务、安全、数据等风险角色定义边界并参与高风险变更。

角色	主要责任	不得转移的责任
技能所有者	范围、指标、版本、退役	业务结果与维护预算
审批人	风险接受、发布或敏感动作	审批理由和权限
平台团队	运行、目录、日志、基础安全	平台控制有效性
使用者	正确输入、结果复核、反馈	不得绕过限制
风险角色	政策、审查、事故参与	风险标准与例外记录

变更评审应按风险和兼容性分层。文案提示的低风险改进可以由所有者在回归通过后发布；权限扩大、外部写入、新数据域、模型供应商替换、重大契约变化必须进入跨职能评审。事故复盘关注系统条件，不以“员工没有仔细看”作为唯一结论。

治理机制	触发	核心产物
目录评审	新增或重名技能	唯一性与边界结论
变更评审	权限、契约、模型、政策变化	风险与兼容决定
事故复盘	错误外发、越权、重大返工	时间线、根因、控制改进
退役治理	低使用、替代、风险不可控	迁移、数据处置、墓碑记录

BOARD BRIEF 13

第十二章 | 营销场景的风险分级与技能形态

竹势 AI 营销智库出品

CONTENT / BRAND
Brief、审校与分发

ADS / LEADS
诊断、评分与培育

SOCIAL / RISK
触达、舆情与升级

营销场景的风险不是由“是否使用 AI”决定，而由数据敏感度、动作可逆性、外部影响、金额、监管和客户承诺共同决定。同一技能在不同上下文下可能需要不同信任等级。

场景	主要数据	主要风险	建议等级	技能形态
内容 Brief	内部策略资料	可逆、内部	T2–T4	结构化生成与人工确认
品牌审校	品牌与产品事实	影响公开内容	T3–T4	规则+语义审校
投放诊断	广告账户与成本	建议可逆，执行涉预算	诊断 T3，执行 T1–T2	建议与执行分离
线索评分	客户与行为数据	可能影响销售优先级	T2–T3	可解释评分
企微跟进	客户身份和对话	外部触达与承诺	T1–T2	草稿优先、敏感语句人工门
活动复盘	多源经营数据	内部决策	T3–T4	证据聚合与归因边界
舆情升级	公开信息与危机材料	声誉与响应时效	T2–T3	分级、证据、人工决策

投放诊断应与投放执行分离：前者可以读取数据、识别异常并生成建议，后者涉及预算、出价和人群变更，需要更高权限、幂等键、审批和回执。企微跟进也应先从草稿和任务提醒开始，在企业证明客户授权、话术边界和撤回机制后再逐步提高自动化。

内容 Brief 契约摘要

要素	说明
输入	目标、受众、渠道、事实包、禁区
输出	结构化 Brief、缺口清单、证据引用
停止	事实不足或目标冲突

品牌审校契约摘要

要素	说明
输入	待审内容、品牌版本、产品事实
输出	问题定位、风险级别、修改建议
停止	品牌规范过期或高监管声明

舆情升级契约摘要

要素	说明
输入	来源、传播量、主体、时间
输出	事件簇、可信度、建议级别
停止	证据冲突或涉及重大危机

BOARD BRIEF 14

第十三章｜八个完整案例：成功、受限与治理启示

竹势 AI 营销智库出品

调用可靠

成功率与错误类型

工作改善

周期、返工与覆盖

经营有效

收入、成本与客户风险

案例 1｜OpenAI Codex：Skills 让代理超越单次编码提示

2026 年 OpenAI 在 Codex 产品演进中把 skills 作为扩展代理能力的重要机制，官方材料强调技能能够把重复工作和特定工作方式封装给代理使用。这个变化的价值不在于增加一段系统提示，而在于让指令、脚本和资源以可复用单元进入代理工作环境。

对企业营销而言，可迁移的机制是“按需加载专业能力”，避免把所有规则、模板和脚本永久塞入一个超长上下文。品牌审校、研究报告生成、投放分析等能力可以按任务发现和加载，降低上下文冲突，并让能力资产与具体对话分离。

其公开资料主要说明产品能力与使用方式，并未披露企业营销场景的量化收益。企业不能据此直接推断稳定性、安全等级或跨厂商兼容性；仍需自行建立契约、权限、回归和运行证据。

管理含义是：把厂商 skills 当作实现载体，而不是治理终点。目录标识、业务所有者、风险分级和版本清单应保留在企业控制面。

要素	核验口径
主体与时间	OpenAI，2026 年 Codex 产品资料
动作	以 skills 扩展代理的专门能力
结果	官方披露功能扩展；营销量化结果未披露
边界	产品实现不等于企业治理标准

案例 2｜Anthropic Agent Skills：开放规范与渐进式加载

Anthropic 的 Claude Code 文档把 skills 描述为可由模型在相关任务中加载的指令、脚本和资源，并说明其遵循 Agent Skills 开放标准、同时增加自身扩展。这样的设计把技能从单次提示中抽离，形成可移植的文件夹级能力包。

渐进式加载有直接的工程意义：目录元数据用于发现，详细指令在被选中后加载，脚本和资源在需要时调用。它降低了把全部企业方法论一次性注入模型的成本，也减少无关说明干扰。

但文件组织本身不能回答谁能执行外发、脚本是否可信、依赖是否锁定、测试是否通过。企业若只复制技能包而没有代码审查和权限隔离，反而会扩大供应链和提示注入风险。

因此，开放格式带来可移植性，但信任必须由企业证据建立。采购时应要求源码、依赖、许可证、测试资产和退出包。

要素	核验口径
主体与时间	Anthropic，Claude Code Skills 官方文档
动作	开放技能格式、按需加载指令脚本资源
结果	形成跨工具复用基础；企业结果未披露
边界	格式开放不代表安全与质量自动成立

案例 3 | GitHub Copilot Agent Skills：团队级专业行为复用

GitHub 官方文档将 agent skills 定义为包含指令、脚本和资源的文件夹，Copilot 在相关任务中加载，以提升特定任务表现。其团队价值在于把项目规范和专业行为从个人提示迁移到仓库或组织可管理资产。

迁移到营销组织，可以把品牌、内容 QA、数据分析和交付规范放入受版本控制的技能包，使不同项目和成员获得一致能力入口。代码仓库式评审也为变更比较、责任追踪和回滚提供基础。

公开文档没有证明技能在所有模型、所有任务上自动等效，也没有提供营销成果数据。技能描述不准确时仍可能错误路由；脚本执行还需遵守环境权限。

治理启示是把技能变更纳入类似代码评审的流程，但评审参与者不能只有工程师，品牌、业务和风险所有者必须共同验收。

要素	核验口径
主体与时间	GitHub，Copilot Agent Skills 文档
动作	团队可管理的指令、脚本、资源封装
结果	支持专业任务复用；量化营销结果未披露
边界	仍依赖模型路由和环境治理

案例 4 | MCP：把工具、资源、提示分开，避免能力概念混乱

Model Context Protocol 的正式规范把服务器能力区分为 resources、prompts 和 tools：资源提供上下文数据，提示提供模板化消息与工作流入口，工具使模型调用外部系统。工具以名称和输入 schema 被发现和调用。

这一区分为企业技能工程提供了基础语言。技能可以组合 MCP 工具和资源，但技能本身还要增加业务验收、权限边界、失败语义和生命周期。一个“创建 CRM 任务”工具不等于“线索跟进技能”，后者还需要客户状态、话术规则、责任人和回执。

MCP 是互操作协议，不是营销能力质量标准。协议能够让连接更一致，却不会自动解决工具是否安全、描述是否准确或业务结果是否有效。

企业应把 MCP 适配器纳入技能依赖层，保留工具版本、服务器身份、授权范围和调用证据，避免把协议兼容误认为业务可替换。

要素	核验口径
主体与时间	MCP 规范，2025–11–25 版
动作	区分工具、资源与提示并提供发现调用机制
结果	增强连接互操作；业务结果未披露
边界	协议不负责企业技能验收

案例 5 | 阿里云百炼：插件与工作流能力封装的中国平台样本

阿里云百炼等企业智能体平台通过插件、知识库、工作流和应用编排，把模型与企业数据、API 和业务步骤连接。对中国企业而言，这类平台降低了接入国产模型、云资源和本地业务系统的门槛。

工程化启示是把插件视为工具层，把工作流视为跨步骤协调层，再由企业技能契约定义“完成什么营销任务”。例如一个公众号内容技能可以调用知识检索、文案生成和文档输出插件，但必须独立声明品牌版本、事实证据和审校门。

平台官方资料更侧重产品能力，具体客户营销结果、失败率和跨云迁移成本通常披露有限。企业需要在采购阶段验证数据导出、提示与工作流资产归属、接口限额和替代路径。

该案例说明中国企业可以利用本土云平台快速构建，但目录、版本和证据最好保持平台中立，避免能力资产完全锁在某一控制台对象中。

要素	核验口径
主体与时间	阿里云百炼，官方产品与文档
动作	插件、知识库、工作流与应用编排
结果	提供企业构建机制；具体营销收益未披露
边界	跨平台可移植性需企业验证

案例 6 | 腾讯云智能体开发平台：连接、知识与流程的本土治理挑战

腾讯云智能体开发平台面向企业提供模型、知识、插件和流程能力，并与腾讯生态连接。营销场景中，企微、内容、客户服务和数据分析具有天然协同潜力。

但生态连接越深，权限边界越重要。企业需要把读取客户资料、生成草稿、发送消息、创建任务和修改状态拆成不同动作权限，不能因为连接成功就默认自动执行。

官方资料能够证明平台提供相关构建能力，但公开材料通常不足以证明某个技能在企业真实客户触达中的长期稳定性。对外发送必须结合客户授权、平台规则和企业审批。

管理含义是先以 T1–T2 建议和草稿能力落地，积累回执、返工和投诉证据后，再决定是否升级到受控执行。

要素	核验口径
主体与时间	腾讯云，智能体开发平台官方资料
动作	模型、知识、插件与流程连接
结果	支持本土生态构建；营销量化结果未披露
边界	客户触达需额外授权与证据

案例 7 | 百度智能云千帆 AppBuilder：应用模板不等于技能资产

百度智能云千帆 AppBuilder 提供面向企业的模型、知识库、组件和应用编排能力，帮助组织较快搭建问答、分析和流程应用。其价值在于降低原型和场景应用开发门槛。

企业技能工程需要进一步把应用中的可复用能力抽离出来。若品牌审校、报告生成和线索分析全部封在一个应用页面中，其他团队难以调用，版本也难以独立升级。

公开资料表明平台提供构建机制，但并不代表应用内部能力已经具有统一契约、幂等和兼容窗口。具体项目的业务结果应以客户验收和运行日志为准。

该案例提醒管理者：应用是用户体验和工作入口，技能是可复用能力组件；两者需要关联，但不宜等同。

要素	核验口径
主体与时间	百度智能云千帆 AppBuilder 官方资料
动作	模型、知识、组件和应用编排
结果	降低应用搭建门槛；具体结果未披露
边界	应用封装可能掩盖技能边界

案例 8 | Salesforce Agentforce：标准动作、主题与企业数据的商业平台实践

Salesforce Agentforce 通过 topics、actions、数据与信任控制，把代理可执行能力配置到销售、服务和营销相关场景。标准动作和自定义动作使企业能够把 CRM 数据和业务操作纳入代理执行。

其机制说明技能工程必须与业务系统的真实对象和权限模型结合。线索评分、跟进建议和任务创建不能只输出文本，而要明确 Lead、Contact、Campaign 等对象的读取和写入边界。

厂商客户案例可能披露效率或体验改善，但不同客户的数据基础、流程成熟度和实施范围差异很大，不能把单个结果作为普遍承诺。本报告不以厂商营销数字推导企业收益。

治理上，企业应把平台 action 映射到中立技能契约，并保留数据模型、审批、日志和退出导出安排。

要素	核验口径
主体与时间	Salesforce，Agentforce 官方文档与客户资料
动作	主题、动作、CRM 数据和信任控制
结果	形成业务系统内代理能力；结果依客户而异
边界	平台动作不等于跨平台可移植技能

BOARD BRIEF 15

第十四章｜指标树、采购退出与 90 天阶段门

竹势 AI 营销智库出品

0–30
盘点与契约
筛选候选，冻结边界

31–60
影子与回放
建立测试证据链

61–90
有限发布
目录治理后再扩权

技能资产的指标必须同时覆盖工程健康、使用质量、风险和业务结果。调用量只能说明被使用，不能说明产生价值；成功率也可能掩盖大量人工返工。董事会应观察从“可调用”到“业务结果”的完整链条。

一级指标	二级指标	管理解释
可靠性	调用成功率、超时率、依赖故障	能力是否可用
质量	返工率、一次验收率、事实错误	输出是否可接受
效率	周期时间、人工分钟、单位成本	是否降低吞吐瓶颈
复用	复用覆盖、团队数、重复技能减少	是否形成资产经济
风险	越权、外发错误、投诉、回滚	是否在风险偏好内
版本	采用率、旧版残留、回滚时间	变更是否可控
业务	线索响应、内容表现、转化、预算效率	是否支持经营目标

采购和退出安排应覆盖能力资产而非只覆盖软件账户。合同需要明确技能源码或可读定义、提示词与规则、测试集、运行日志、知识产权、第三方许可证、数据导出格式、迁移协助、停服通知、兼容窗口和替代接口。

采购条款	最低要求	退出证据
资产归属	企业定制规则、测试和数据归属明确	完整导出清单
可移植性	开放接口和中立 schema	替代平台验证
成本	模型、调用、存储、人工运维可分解	历史用量与预测
供应链	依赖与许可证透明	依赖清单和副本
停服/变更	通知期与兼容窗口	迁移与降级演练

90 天技能资产化阶段门

阶段	核心动作	交付物	停止条件	董事会/高管门
0–14 天 资产盘点	收集任务、提示词、SOP、脚本、供应商能力	能力地图、重复与风险清单	找不到所有者或真实任务	批准候选范围
15–30 天 候选筛选	按频率、稳定、验收、风险、复用评分	首批 10–20 个候选	输入不可得或无法验收	批准工程化投资
31–45 天 契约化	完成十一格契约、权限和依赖	契约卡、目录草案、测试设计	权限边界不清	批准进入测试
46–60 天 影子测试	历史回放、单元、契约、回归、安全	证据包与差异报告	关键回归或安全失败	批准有限发布
61–75 天 有限发布	T1–T2 小范围真实使用	运行指标、返工与事故记录	超阈值风险或无业务采用	批准扩容或修订
76–90 天 目录治理	定级、版本、所有者、退役与采购规则	正式目录与季度治理机制	无法持续维护	决定规模化、暂停或退出

90 天的目标不是把所有营销经验都变成技能，而是建立一条可重复的资产化生产线，并证明首批能力在真实业务中能够被发现、正确调用、稳定验收和安全升级。停止条件是治理能力的组成部分：当输入、所有者、权限或验收无法成立时，应暂停，而不是用更多提示词掩盖结构性缺陷。

高管决策	批准依据	拒绝或暂停依据
扩大技能目录	复用、质量和风险证据稳定	目录膨胀、重复能力
提高自动化等级	长期稳定、可回滚、低事故	外部动作不确定
更换平台或模型	兼容测试与退出包完备	资产不可导出
增加预算	单位经济和业务结果改善	仅调用量增长

经典思想的当代转译

1. David Parnas：信息隐藏

Parnas 在 1972 年讨论模块化设计时，核心不是把系统按流程步骤切块，而是围绕可能变化的设计决策隐藏信息，使变化不会扩散。技能工程应隐藏模型、提示、脚本和工具编排细节，对外暴露稳定能力契约。修正之处在于生成式系统的行为并非完全确定，因此除了接口稳定，还要暴露证据、置信和失败语义。管理含义是把“实现可替换”写入采购和架构。

对营销管理者而言，这一思想的价值不是学习软件术语，而是获得一套判断能力资产是否真正可维护、可审计、可替换的标准。它要求业务团队参与契约和验收，而不是把技能治理完全交给平台团队。

2. Bertrand Meyer：契约式设计

契约式设计用前置条件、后置条件和不变量描述软件组件责任。迁移到 AI 技能时，前置条件包括授权、数据新鲜度和上下文完整性，后置条件包括结构、事实、审校与回执，不变量包括权限和品牌禁区。生成式输出需要概率性验收，不能机械照搬确定性断言。

对营销管理者而言，契约式设计首先改变需求评审方式。业务方不能只说“生成一份好文案”或“给出靠谱诊断”，而要共同定义可检查的前置条件、输出结构、禁止事项、异常返回和验收阈值。平台团队负责把这些要求变成机器可执行的约束，业务所有者则必须对“什么结果可进入下一环节”承担判断责任；两者缺一，契约就会退化成技术字段表或主观审美表。

在经营管理上，契约还提供了跨部门争议的裁决基线。输入缺失、上游数据过期、模型输出偏离、人工审核超时，应分别归属不同责任，而不能都算作“AI 不稳定”。董事会和管理层应要求高风险技能在上线前明确责任分界、拒绝服务条件和证据留存要求，使返工、事故和供应商争议能够沿契约定位，而不是依赖会后解释。

3. W. Edwards Deming：质量是系统结果

Deming 强调大多数质量问题来自系统而非个体。技能事故不应只归咎于使用者“没有看清”，而要检查目录描述、输入设计、权限、测试和监控。管理含义是建立闭环证据和系统改进，而不是追加人工检查。

Deming 的系统观要求管理者把技能质量看成设计与运行共同产生的结果。品牌审校频繁漏检时，优先检查规则覆盖、样本分布、上下文装载、版本漂移和人工门位置，而不是简单增加审核人数；投放诊断反复误报时，应分析数据延迟、指标口径和阈值校准，而不是责备某位运营人员“不会提问”。只有把缺陷还原到系统条件，改进才会形成可复用能力。

相应的治理机制应围绕稳定过程建立：每次失败进入统一缺陷分类，重大偏差触发根因分析，回归样本吸收真实事故，版本评审同时查看质量趋势与业务损失。管理层需要防止以平均成功率掩盖关键场景失效，也不能用一次优秀演示替代长期过程能力。技能所有者的绩效应包含缺陷关闭速度、重复事故率和证据完整度，而不只是调用量。

4. 语义化版本 SemVer

语义化版本以主、次、修订版本传达兼容性。AI 技能可以借用这一语言，但模型和政策变化可能在接口不变时改变行为，因此必须增加行为回归和运行清单。版本号是沟通信号，不是质量证明。

语义化版本对营销技能的管理价值，在于让兼容性成为发布承诺。输出字段删除、风险等级含义改变或授权范围扩大，应视为破坏性变化并进入主版本；新增可选能力可进入次版本；不改变契约的缺陷修复才适合修订版本。业务系统据此决定是否自动升级、并行验证或继续锁定旧版，避免“后台悄悄更新”使审批流和下游报表突然失效。

但生成式技能不能只看接口版本。模型替换、系统提示调整、知识库刷新和政策规则变化，即使没有改变字段，也可能改变事实准确性、语气、拒答边界和成本。因此每次发布还应绑定行为基线、依赖清单和适用政策版本，并设置兼容窗口、影子流量与回滚条件。管理者应把版本号理解为变更沟通协议，再用回归证据证明行为仍然可接受。

5. 最小权限原则

最小权限源于信息安全工程，要求主体只获得完成任务所需权限。AI 技能将主体、工具和数据组合在一起，更需要按调用收缩权限。管理含义是把“能做”与“允许做”分开，并对敏感动作设置人工门。

最小权限原则在营销场景中必须落实到“每一次调用需要什么”，而不是只按岗位授予长期宽权限。生成活动复盘可能只需读取活动数据；形成企微跟进草稿需要读取客户摘要但不需要发送权限；真正发送消息则应临时取得指定客户、指定模板和限定时间窗的授权。把读取、生成、提交审批和执行拆开，才能让高价值自动化与风险控制同时成立。

管理层还应关注权限的生命周期与组合效应。单个技能看似低风险，但与客户名单导出、外部网络访问或批量发布技能组合后，可能形成新的泄露或声誉路径。目录评审应记录数据域、工具白名单和不可组合项；临时授权到期自动收回，离岗与供应商退出立即失效，敏感动作保留双人复核和不可抵赖日志。权限治理的目标不是让系统“什么都不能做”，而是让每个动作只在明确责任和有限暴露面内发生。

6. 软件供应链与可追溯构建

现代软件供应链治理强调依赖清单、来源、签名和可复现性。技能包包含脚本、模型、连接器和知识资源，同样存在依赖污染与许可证风险。企业应保存运行时物料清单，使事故能够定位到具体版本。

软件供应链思想提醒企业，技能资产的风险不只来自模型输出，也来自被引入的脚本、连接器、模板、第三方 API、开源包和知识资源。采购或安装一个技能时，应知道它依赖什么、由谁维护、以何种许可证分发、是否访问外网、更新渠道是否可信，以及依赖失效后会影响哪些业务链路。没有这些信息，所谓“可复用技能”实际上是一组无法审计的隐性依赖。

治理上应为每个生产技能保存可追溯物料清单、来源校验、批准版本和替代方案，并把关键依赖纳入漏洞与停服监测。第三方更新不得直接进入生产，需先在隔离环境完成安全扫描、契约测试和业务回放；供应商退出时，企业应能导出契约、测试集、配置和运行证据，切换到替代模型或工具。董事会关注的不是每个开源包名称，而是关键营销能力是否存在单点失效、许可证争议和不可退出的经营暴露。

具名专家与机构视角

专家/机构	原始视角	本报告转译
David Parnas	模块应围绕易变设计决策隐藏信息	技能对外暴露稳定契约，内部实现可替换
Bertrand Meyer	前置条件、后置条件和不变量形成责任契约	把授权、验收、错误和边界写进技能
W. Edwards Deming	质量取决于系统设计与持续改进	用证据链和复盘替代事后责备
NIST	最小权限与安全控制体系	按调用限制数据、工具和动作
OpenAI	Codex skills 扩展代理可执行专业工作	厂商实现可作为载体，不等于企业标准
Anthropic	Agent Skills 封装指令、脚本和资源并按需加载	支持渐进式披露与可移植
MCP 社区	工具、资源、提示分别提供动作、上下文和模板	技能应在协议层之上增加业务契约
GitHub	Agent skills 支持团队共享专业行为	将技能变更纳入版本和评审
OWASP	LLM 应用面临提示注入、过度代理等风险	红队、最小权限和外部内容隔离
SLSA/OpenSSF	供应链完整性需要来源与构建证据	技能依赖应可追溯、可验证

专题深化｜让技能工程进入真实经营系统

一、能力发现必须从工作证据出发

企业第一次盘点营销技能时，最常见的错误是让各部门填写“希望 AI 会什么”。这种愿望清单会迅速膨胀，因为每个人都可以把岗位职责改写成一个宏大技能名称。更可靠的方法是抽取过去八到十二周的真实工作证据：任务单、会议行动项、内容修改记录、投放诊断、客户跟进、复盘报告和供应商交付。只有能够找到真实输入、实际操作者、验收动作和重复频率的任务，才进入候选池。

发现过程应同时记录“人工为何需要判断”。如果一个任务的价值主要来自事实检索、格式转换和规则核对，工程化空间通常较大；如果价值来自利益权衡、政治判断、关系维护或不可逆承诺，技能更适合提供证据和建议，而不是替代责任主体。这样的区分能避免把高层判断包装成自动化承诺。

候选池还要识别同义任务。市场部称为“品牌校对”，法务称为“宣传合规初筛”，电商团队称为“商品页风险检查”，它们可能共享事实核验和禁用声明能力，却拥有不同审批和监管边界。目录设计应保留共用内核，并通过场景配置表达差异，不能简单复制三个近似技能。

二、技能所有者要拥有预算与退役权

把技能所有者写成“某位最懂提示词的员工”会形成新的关键人风险。所有者应是能够调动维护资源、决定适用范围、接受业务指标并推动退役的岗位。品牌审校技能通常归品牌治理负责人，投放诊断技能归营销运营或增长负责人，平台团队则对共用运行底座负责。

所有者必须拥有停止使用的权力。当法规变化、知识源失效、供应商接口不稳定或返工率持续超阈值时，继续维持表面可用会积累更大风险。目录中的“暂停、降级、退役”应与“发布”同等正式，且保留替代流程、在途任务处理和历史证据。

预算责任同样重要。技能运行成本不只是模型 token，还包括知识维护、测试更新、平台监控、人工抽检、事故处置和供应商支持。没有总成本视图的技能可能看似调用便宜，却把大量返工转移给品牌、销售和法务团队。

三、描述质量决定错误路由概率

智能体或使用者通常先读取技能名称和简短描述，再决定是否调用。描述如果使用“全能增长”“智能营销大师”之类宣传语言，无法提供边界信息，路由器只能依据模糊相似度做选择。高质量描述应说明任务、对象、必要输入、输出形态、适用场景和明确禁用场景。

例如，“分析投放并给出建议”仍然过于宽泛。更好的表述是：“读取指定账户在给定时间窗内的计划、素材与转化数据，识别异常并生成不含自动执行的优化建议；不用于修改预算、出价或定向。”这段描述同时改善人类选择、模型路由和风险审查。

描述漂移需要版本治理。技能实现升级后，如果能力范围扩大而目录描述未变，使用者可能在不知情时触发新动作；反过来，描述宣称的功能已经被取消，也会造成错误期待。变更评审应把描述、契约和实现作为一个发布包检查。

四、上下文不是越多越好

营销技能往往需要品牌规范、产品事实、活动目标、客户状态和历史结果。把所有可访问资料一次性塞入上下文，会增加成本、冲突和泄露面，也会让模型难以判断哪一版事实有效。上下文契约应定义选择规则：按品牌、区域、产品、时间和任务类型检索最小必要内容。

上下文还要区分权威层级。经批准的产品说明应高于历史公众号文章，当前价格政策应高于旧销售话术，监管规则应高于创意偏好。冲突时技能不应自行“综合”，而应返回冲突证据和需要决策的字段。

上下文快照是审计关键。企业需要知道某次输出使用了哪一版品牌规范、哪些产品事实和何时抓取的外部资料。只记录最终文本而不记录上下文版本，事故复盘就无法判断是模型错误、知识过期还是调用者选择错误。

五、业务验收要从“好不好”变成可执行标准

开放式营销产出很难用单一准确率评价，但这不意味着无法契约化。内容 Brief 可以检查目标、受众、渠道、事实、禁区和成功指标是否完整；品牌审校可以检查问题定位、依据、风险级别和修改建议是否齐全；活动复盘可以检查证据来源、归因边界和下一步责任是否明确。

主观质量应通过锚定样本校准。品牌团队可以提供“可直接采用、需小改、不可采用”的代表性样本，并说明差异原因。评判模型和人工评审都基于这些锚点训练一致性，避免不同验收人各用一套隐性标准。

验收标准还要包含拒绝条件。事实来源不足、数据时间窗不完整、客户授权不清、监管分类不确定时，技能能够正确停止，比勉强生成一个流畅答案更有价值。停止率不是天然负面指标，应与输入质量和风险避免共同解释。

六、人工门必须有明确的判断任务

很多企业在高风险节点设置“人工审核”，却没有告诉审核人要检查什么。结果是审批按钮成为形式流程，审核人面对长篇输出只能凭直觉点击通过。有效人工门应提供风险摘要、关键差异、证据链接、待确认字段和建议动作，让人承担明确判断。

人工门也不能无限叠加。品牌、法务、业务和老板逐层审批同一内容，会把 AI 节省的时间全部消耗在等待上。应根据风险类型分配唯一主审批人，其他角色只在触发特定条件时加入。低风险可逆动作可以抽检，高风险不可逆动作采用强制审批。

审批结果必须回流技能学习。通过、驳回和修改不是流程终点，而是标注数据。企业应记录驳回原因、修改差异和责任角色，用于更新规则、回归集和目录描述。

七、线上观测必须连接业务状态

仅监控 HTTP 成功或模型是否返回内容，无法说明技能完成了业务任务。线索跟进技能返回 200，但 CRM 任务没有创建、客户已经退订或销售没有接收，都属于业务失败。观测应贯穿输入校验、工具调用、外部回执、状态写回和人工接管。

关键指标需要按技能版本、业务场景、数据域和用户群切分。全局成功率可能掩盖某个区域知识过期、某类产品事实错误或新版本在高监管场景下返工激增。分层观测让回滚可以只影响问题范围。

日志也要控制敏感信息。审计需要足够证据，但不意味着长期保存全部客户对话和原始附件。企业应定义脱敏、摘要、哈希、保留期限和访问权限，在可追溯与数据最小化之间取得平衡。

八、技能组合需要业务事务思维

跨技能工作流经常出现部分成功：内容已经生成并审批，但发布接口失败；投放建议已批准，但预算修改成功后日志写回失败；客户触达已发送，但 CRM 状态没有更新。若没有事务和补偿设计，系统会在重复执行时产生二次发布或重复触达。

每个外部动作应生成幂等键和回执，并在重试前查询真实状态。补偿不一定等于技术回滚：公开内容可能无法完全撤回，客户承诺也不能靠删除记录消失。此时补偿是发送更正、通知负责人、冻结后续动作并保留证据。

工作流所有者需要定义“最终一致”的时间窗口。某些分析任务可以等待数据迟到，客户投诉升级则需要分钟级响应。超时、重试和人工接管策略应反映业务后果，而不是套用统一技术默认值。

九、模型替换不是普通配置变更

企业常把模型路由视为降低成本的开关，但模型替换会改变事实遵循、格式稳定、拒绝行为、工具调用和语言风格。即使技能接口没有变化，业务行为仍可能发生重大漂移。模型版本必须进入运行清单，并触发代表性回放。

回放集应覆盖常见任务、边界任务、历史事故和高价值客户场景。比较不仅看平均分，还要看最坏情况和风险类型。一个轻型平均文案质量更高，却在监管声明上更容易过度推断，不能直接替换高监管场景。

多模型路由同样需要解释。路由规则应基于任务复杂度、数据敏感度、时延、成本和合规限制，而不是只看价格。关键输出应记录实际使用模型，便于复盘和成本归因。

十、知识产权要覆盖隐性资产

技能采购中最容易被忽略的是提示词、评分规则、测试集、样例、字段映射和失败处理。这些材料可能比一段代码更能决定能力效果。如果合同只约定“交付一个可用智能体”，企业在更换供应商时可能拿不到真正的能力资产。

企业定制内容与供应商通用框架应分开约定。供应商可以保留通用方法和底层产品权利，但企业自己的品牌规则、客户字段、历史标注、测试结果和运行日志应可完整导出，并明确不得用于其他客户训练。

开源组件也不等于没有义务。技能中的脚本、模型和数据包可能具有不同许可证和使用限制。采购评审应要求依赖清单和许可证说明，避免在公开营销产物或商业再分发中产生争议。

十一、可移植性要通过迁移演练证明

“支持 API”“兼容 MCP”或“采用开放技能格式”只是可移植性的必要条件，不是充分条件。真正迁移还涉及身份、权限、数据对象、工具语义、状态存储、模型差异和日志格式。企业应选择一到两个关键技能进行替代平台演练。

迁移演练至少验证：契约能否导出，知识能否保留来源和版本，测试集能否复用，工具适配成本多大，历史运行证据能否查询，旧平台停用后是否有在途任务。只有完成这些验证，退出安排才不是合同文字。

可移植性不意味着追求最低共同能力。企业可以利用平台特色，但应在中立契约层保留业务语义，并把平台专属优化标注为可替换实现。

十二、规模化的前提是目录治理而非数量增长

当首批技能证明价值后，组织容易用数量衡量进展，要求每个团队快速提交几十个技能。结果往往是重复命名、无人维护、缺少测试和低使用率。规模化目标应是提高关键任务覆盖和复用，而不是扩大目录条目。

目录应定期清理低使用、无所有者、长期不兼容和被替代技能。对近义技能，可以合并公共内核并保留场景配置；对高耦合能力，可以升级为领域技能包；对一次性项目组件，应退出企业级目录，放入项目空间。

季度治理会议需要同时看新增、采用、事故、退役和经济价值。只有新增没有退役，说明组织在堆积数字债务；只有调用没有业务结果，说明技能可能成为新的形式主义。

董事会审议清单

审议事项	必须看到的证据	典型红旗
首批技能范围	真实任务、所有者、验收与风险	愿望清单替代工作证据
自动化等级	信任等级、回滚与人工门	用演示效果替代运行证据
平台采购	资产归属、依赖、迁移演练	只承诺 API 可用
规模化预算	复用覆盖、返工、单位成本、业务结果	只报告技能数量和调用量
重大事故	时间线、根因、控制改进、责任	只处罚使用者、不改系统

十三、从技能组合推导组织边界

技能目录会反向暴露组织设计问题。若一个技能必须同时读取多个部门数据、调用多个账号并由四个角色连续确认，说明企业可能把责任拆得过碎；若所有高价值能力都依赖一个“万能运营智能体”，说明组织又把判断和执行过度集中。技能组合图可以作为组织接口图，帮助管理层看见职责交接、等待和重复审批。

能力边界与岗位边界不必一一对应。一个岗位可以调用多个技能，一个技能也可以服务多个岗位，但每次调用必须有明确责任主体。营销运营人员可以使用品牌审校技能，销售也可以使用同一能力检查客户材料；两者的输入数据、审批规则和输出去向可以不同。企业应复用内核，同时保留场景策略。

当技能跨越业务单元时，平台团队不能自动成为业务所有者。平台团队负责可用性、目录和基础安全，业务共同体需要指定主所有者并建立协商机制。否则出现质量争议时，工程团队会被迫决定品牌、客户和渠道规则，形成责任倒置。

十四、技能经济性必须计算“避免成本”与“选择权”

技能资产的收益不只来自减少人工分钟。标准化输入和输出减少了跨团队沟通损耗；版本与回滚降低了变更事故；目录和复用避免重复采购；中立契约为企业保留更换模型、工具和供应商的选择权。这些收益在单次调用成本表中看不见，却决定长期总拥有成本。

经济性评估可以把成本分为构建、运行、维护、风险和退出五类。构建包括业务梳理和测试集，运行包括模型、工具和人工门，维护包括知识和政策更新，风险包括事故、投诉和合规审查，退出包括迁移和替代平台验证。只有五类成本都被记录，管理层才能比较自建、采购和外包。

高复用技能应获得更稳定的维护预算，低复用但高风险技能则可能仍有战略价值，例如危机舆情升级和高监管内容审校。投资组合不应只按调用量排序，而要结合风险避免、关键业务覆盖和替代难度。

十五、数据新鲜度是技能契约的一部分

营销事实变化速度不同。品牌使命可能多年稳定，产品价格和促销政策可能按周变化，广告账户数据按小时更新，舆情信号甚至按分钟变化。技能契约需要声明每类数据允许的最大陈旧时间，并在超过阈值时停止、降级或请求刷新。

数据新鲜度还要区分“系统更新时间”和“业务有效时间”。一个文件刚被上传，不代表其中事实仍然有效；一个政策文档很久没有修改，也可能仍是当前有效版本。知识治理应记录生效日期、失效日期、批准人和替代关系，而不是只看文件修改时间。

当多源数据冲突时，技能应按照权威层级和时间规则处理，并保留冲突证据。对于价格、产品承诺、监管声明和客户状态，不能让模型通过语言流畅度自行选择。冲突本身应成为异常输出，触发业务确认。

十六、营销技能需要面向反事实与边界样本测试

常规样本通常来自成功历史，因此容易掩盖技能在异常输入下的脆弱性。反事实测试会改变一个关键条件，例如把受众从普通消费者改为未成年人，把产品从一般消费品改为医疗相关，把内部分析改为公开发布，观察技能是否提高风险等级、收紧声明或停止执行。

边界样本包括极短输入、相互矛盾的品牌规则、超出预算上限的投放建议、客户已经退订却仍被列入跟进、舆情来源只有匿名截图等。一个可信技能必须在这些情形下给出明确失败语义，而不是用补写和猜测制造完整答案。

历史事故应永久进入回归集。事故修复后，如果只修改提示词而不保留触发样本，后续模型或规则升级可能重新引入同类问题。回归集因此是组织记忆的一部分，也需要访问控制和版本管理。

十七、红队要模拟真实营销攻击面

营销技能接触大量外部内容：网页、评论、私信、附件、竞品材料和用户生成内容。这些内容可能包含诱导模型忽略规则、泄露上下文或调用工具的指令。红队测试应模拟间接提示注入，验证外部文本是否被当作不可信数据处理。

攻击面也包括业务诱导。销售人员可能要求技能绕过审批快速发送，供应商可能在素材中嵌入追踪链接，客户可能诱导系统披露其他客户报价，内部人员可能通过更改标签获得更高优先级。安全测试不能只关注技术黑客，还要覆盖真实激励和流程绕行。

红队结果应转化为具体控制：输入隔离、工具白名单、敏感字段遮蔽、二次确认、异常告警和使用者培训。只保留一份攻击报告而不修改技能契约和运行策略，不能形成风险降低。

十八、服务水平必须区分“技术可用”与“业务可用”

技术 SLA 通常关注响应时间、错误率和恢复时间，但营销技能还需要业务服务水平。例如品牌审校在大促前必须在十分钟内返回，舆情升级在严重事件中需要两分钟内给出证据摘要，活动复盘可以在数据完整后一个工作日交付。不同任务的时效价值差异很大。

业务可用还包括正确降级。当模型服务不可用时，系统可以切换备用模型，但如果备用模型未通过高监管场景回归，就只能降级为规则检查或人工处理。追求表面在线而牺牲风险边界，会让 SLA 成为错误激励。

服务水平应与支持责任绑定。技能所有者负责业务规则和验收，平台团队负责基础运行，供应商负责其服务承诺，使用团队负责及时提供完整输入。没有责任拆分的 SLA 会在事故时变成相互推诿。

十九、目录发现需要同时服务人和机器

人类使用者习惯按岗位、场景和产物搜索，智能体路由器更依赖语义描述、输入条件和风险标签。目录应提供两套互补视图：面向人的业务地图，以及面向机器的结构化元数据。两者必须引用同一唯一标识，避免出现两套事实。

机器发现字段应包含明确的正例和反例。正例说明何时调用，反例说明相似但不适用的任务。例如“品牌审校”适用于已形成的内容，不适用于从零制定品牌战略；“线索评分”适用于已授权的潜客记录，不适用于未经同意抓取的个人信息。

路由错误需要被观测和纠正。企业可以记录用户取消调用、改选技能、人工驳回和后续替换，形成描述优化依据。路由器升级同样要版本化，因为它会改变技能被选择的概率，即使技能本身没有变化。

二十、退役是能力资产成熟度的检验

成熟目录一定会持续退役技能。业务策略变化、渠道消失、平台接口停用、规则被公共能力替代或维护成本超过价值时，继续保留旧技能会增加错误调用和安全暴露。退役不是删除文件，而是一个受控迁移过程。

退役计划需要确定替代技能、最后可用日期、兼容窗口、在途任务、历史记录、数据处置和调用者通知。目录中应保留墓碑记录，告诉调用者该能力为何停止、应该迁移到哪里，避免同名技能被重新创建。

董事会和高管应关注退役率与目录净增长。长期零退役通常意味着治理没有真正运转；大量紧急退役则说明前期契约、测试或供应链审查不足。稳定的新增、合并、升级和退役节奏，才表明企业把技能当作长期资产管理。

实施手册 A｜内容 Brief 技能的工程化拆解

内容 Brief 是适合作为首批技能的场景，因为任务高频、输入相对明确、输出可结构化，而且外部风险低于直接发布。工程化时应先规定业务问题：把活动目标、受众、渠道、产品事实、资源约束和禁区转化为可被创意、文案、设计和投放共同使用的标准 Brief。它不负责替管理层决定战略目标，也不应在事实缺失时自行补全产品承诺。

输入契约至少包含目标层级、目标受众、期望行为、核心事实、渠道、时间、预算或资源限制、品牌版本和审批角色。输出需要使用稳定字段，并把“已知事实、推断、待确认事项”分开。对缺失输入，技能应返回缺口清单，而不是通过通用常识制造看似完整的 Brief。

测试集应覆盖新品发布、促销活动、品牌内容、B2B 白皮书、线下活动和高监管品类。业务验收关注跨角色是否减少反复确认，而不仅是文字是否专业。线上指标包括 Brief 一次通过率、下游返工次数、从需求提出到创作启动的周期，以及因事实缺失触发停止的比例。

实施手册 B｜品牌审校技能的双引擎设计

品牌审校同时需要确定性规则和语义判断。禁用词、产品名称、价格格式、法定声明和视觉规格适合用规则或结构化校验；语气、承诺强度、品牌人格和上下文一致性需要模型与人工样本。把全部问题交给模型会降低可解释性，把全部问题写成规则又无法覆盖开放表达。

输出不能只给总分。每个问题应包含定位、问题类型、依据、风险等级、修改建议和是否必须人工确认。对于事实或法规不确定的内容，技能应标注“需要权威来源”，不能用模型生成的替代措辞掩盖证据缺口。

版本管理必须关联品牌规范。品牌更新定位、产品命名或禁区后，旧规范不能继续静默使用。技能运行记录应保存规范版本和规则版本，灰度阶段比较新旧版本的误报、漏报和人工修改。高监管场景即使长期稳定，也应保留最终人工责任。

实施手册 C｜投放诊断技能的建议与执行隔离

投放诊断的核心是把账户数据转换为可验证的问题假设，而不是自动给出“提高预算、换素材”的泛化建议。输入应声明平台、账户、时间窗、归因口径、目标、预算约束和异常阈值。输出需要区分观测事实、可能原因、所需补充证据和建议实验。

诊断技能应保持只读，执行能力单独注册。这样可以让更多人员安全使用分析，同时把修改预算、出价、定向和素材状态限制在更高信任等级。执行技能必须拥有幂等键、变更前快照、审批、平台回执和回滚或补偿方案。

业务验收不应以建议数量衡量，而应看有效问题发现率、无效建议率、从异常出现到确认的时间、被采纳实验的结果和人工分析时间。模型更换时要用历史账户回放，特别检查对小样本、归因延迟和季节性变化的误判。

实施手册 D｜线索评分技能的可解释与公平边界

线索评分会影响销售资源分配，因此需要把预测、规则和经营策略分开。模型可以估计转化倾向，业务规则可以反映战略客户、区域覆盖和产品适配，但不能把无法解释的个人属性或代理变量静默引入评分。输入字段、来源、授权和保留期限必须在契约中列明。

输出应包含分数、等级、主要贡献因素、数据缺口和建议动作。销售人员必须能够发现错误并反馈，不能把分数当作不可挑战的真相。对新行业、新渠道或分布变化，技能应降低信任等级并请求重新校准。

指标需要同时观察转化、覆盖和分配影响。只提高高分线索转化率可能是因为系统缩小了覆盖范围，导致潜在客户被忽略。管理层应看召回、销售接受率、跟进时效、异议和不同业务群体的表现，并定期审查特征合法性。

实施手册 E｜企微跟进技能的客户授权与承诺控制

企微跟进直接面对客户，风险来自身份、授权、频率、语气和商业承诺。第一阶段适合生成草稿、任务提醒和对话摘要，由员工确认发送。只有在客户同意、话术范围明确、退订机制有效且运行证据稳定后，才考虑对低风险服务消息进行受控自动化。

技能应读取最小必要客户上下文，不应为了“更懂客户”检索无关历史资料。输出需要标记使用的客户状态、最近互动、允许承诺和禁止承诺。涉及价格、交付、合同、投诉、退款和敏感个人信息时，必须转人工。

线上观测除了发送成功率，还要监控撤回、客户负面反馈、退订、重复触达、承诺纠纷和销售接管时间。任何不确定外发都禁止盲目重试，应先查询渠道回执。事故复盘要检查授权、路由、幂等和人工门，而不是只检查文案。

实施手册 F｜活动复盘技能的证据与归因纪律

活动复盘技能容易生成结构完整却缺乏证据的报告。工程化首先要定义数据来源、时间窗、基线、目标和归因口径。技能应把事实、解释和建议分开，并对缺失数据、跨渠道重复计算和外部因素明确标注。

输出可以包括目标达成、渠道表现、内容与人群信号、成本、异常、学习和下一步实验，但每个关键结论要关联证据。不能因为活动后指标上升就直接宣称活动造成增长，也不能用模型叙事替代因果判断。

业务价值来自缩短复盘周期和提高学习回流速度。指标应关注数据完整时间、报告一次通过率、建议被纳入下一轮计划的比例、重复错误减少，以及复盘结论在后续技能和工作流中的使用。

实施手册 G｜舆情升级技能的证据优先

舆情升级的首要任务不是生成公关回应，而是把分散信号聚合为事件、判断来源可信度、估计传播与利益相关方影响，并在规定时间内交给责任人。技能应避免仅凭单条截图或匿名转述提升为重大事件，也不能因为声量暂低就忽略高严重度指控。

输入需要保留来源、时间、主体、原始内容、传播关系和已知事实。输出应包含事件摘要、证据等级、尚未证实内容、风险维度、建议升级级别和下一步取证。对回应口径的建议必须与事实确认和审批分离。

测试应使用历史危机、误报、同名主体、讽刺表达、旧闻翻炒和跨语言传播样本。线上指标关注发现到确认时间、误报与漏报、升级准确性、人工接管和证据完整度，而不是简单追求告警数量。

实施手册 H｜企业技能目录的运营节奏

目录上线后需要固定运营节奏。每周处理新增候选、路由错误和紧急变更；每月审查使用、返工、成本、事故和旧版本残留；每季度进行合并、信任等级调整、供应商依赖审查和退役。没有节奏的目录会很快退化为静态展示页。

目录治理委员会不应成为新的重审批机构。低风险、兼容变更由所有者和平台自动化证据驱动；高风险、权限扩大和不兼容变化才进入跨职能评审。会议材料应突出例外和趋势，避免逐条朗读技能清单。

管理层最终需要看到能力组合是否支持战略重点。新品增长、品牌一致性、销售响应、私域运营和危机管理各自需要哪些关键技能，哪些已经达到受控执行，哪些仍依赖人工，哪些存在单一供应商风险。目录由此成为能力投资地图，而不只是技术资产库。

管理专题｜高管如何避免五类技能工程误区

第一类误区是把技能建设变成提示词征集比赛。员工提交大量个人模板，平台把它们包装成可点击卡片，却没有统一输入、验收、权限和所有者。短期看目录很丰富，长期看每个模板都依赖创建者解释，模型升级后也没人负责回归。纠正方法是先确认业务任务和证据，再决定哪些提示词值得成为实现材料。

第二类误区是把技术可调用等同于业务可复用。一个 API 能返回文本，不代表不同团队能够用同样口径验收；一个脚本能修改账户，不代表企业允许所有场景执行。复用需要稳定语义、兼容字段、权限策略和运行证据。没有这些条件，所谓复用只是共享代码。

第三类误区是先追求自动化等级，再补治理。企业为了展示“AI 能自主执行”，过早开放发布、触达和预算动作，事故后再增加层层审批。更合理的路径是从建议、草稿和影子运行积累证据，逐步提高信任等级。治理不是速度的对立面，而是可持续提速的基础。

第四类误区是把所有风险交给最终审核人。人工审核无法弥补错误数据、过期知识、越权工具和不确定重试。审核人也不可能在几分钟内重新完成系统本应做的全部核验。有效治理要把风险前移到输入校验、权限、规则、证据摘要和异常停止。

第五类误区是用技能数量、调用次数和节省 token 作为主要成绩。真正的能力资产应减少重复建设、缩短周期、降低返工、扩大关键任务覆盖并控制风险。一个调用量很高但持续需要人工重写的技能，可能只是在制造新的工作。

高管还需要区分“能力拥有”与“平台拥有”。企业可以使用外部平台运行技能，但应掌握业务契约、测试集、知识版本、运行证据和退出安排。否则平台更换时，组织会发现真正可迁移的只有几段说明文字，而不是稳定能力。

技能工程的组织成熟度可以分为四个阶段：个人技巧阶段依赖少数高手；团队模板阶段开始共享但缺乏契约；平台组件阶段具备调用和版本；企业资产阶段进一步拥有权限、证据、所有者、经济指标和退役机制。管理层应明确当前阶段，不要用平台功能清单冒充组织成熟度。

在个人技巧阶段，最重要的动作不是立即搭建复杂平台，而是收集真实任务和修改记录，识别哪些经验可以明确表达。团队模板阶段的重点是统一命名、输入和验收，淘汰无法复现的技巧。平台组件阶段要建立契约测试、身份和日志。企业资产阶段则把预算、风险和战略覆盖纳入治理。

董事会关注的不是每个技能的技术细节，而是能力组合是否支持战略。若企业把增长重点放在大客户获客，就要看到研究、内容、线索识别、销售跟进和复盘之间是否形成可控链路；若重点是品牌升级，就要看到品牌知识、创意、审校、发布和舆情是否共享同一事实与权限体系。

高管决策门应当少而有力。首批候选范围、提高自动化信任等级、引入新的高风险数据域、替换关键平台、发生重大外发事故和扩大规模化预算，属于高管层应审议的节点。普通提示调整、兼容性修复和低风险知识更新应由所有者在证据门内处理。

技能所有者的绩效不能只看上线速度。应同时考察采用、一次验收、返工、事故、成本、旧版本清理和业务结果。平台团队的绩效则关注目录可用性、契约测试覆盖、日志完整、恢复时间和跨平台适配。风险团队应关注政策转译速度、例外质量和控制有效性，而不是审批数量。

组织还应防止“影子技能”。员工为了绕过目录限制，可能继续在个人账号、私有脚本或供应商控制台中运行高价值能力。治理不能只靠禁止，而要让正式目录更易发现、更快获得低风险授权，并提供把个人成果转化为企业资产的通道。对确有价值的影子能力，应完成来源审查和契约化后纳管。

当技能表现下降时，排查顺序应从输入分布、知识新鲜度、工具接口、模型版本、政策规则、路由和使用方式逐层展开。直接修改提示词可能暂时改善几个样本，却掩盖真正根因。运行清单和分层指标能够减少这种“凭感觉调参”的维护方式。

企业技能工程最终会形成新的运营语言：能力有所有者，调用有主体，输出有验收，动作有权限，变化有版本，发布有证据，事故有回放，退役有迁移。它让营销方法论、数字平台和组织责任进入同一个可执行体系。

这套体系不会消除人的判断。相反，它把人的判断集中在目标、边界、例外、审批和学习上，把可重复的检索、转换、核对、分析和执行交给受控组件。企业因此获得的不是一个“更会聊天”的系统，而是一种可以积累、组合、审计和改进的营销生产能力。

管理专题｜把技能资产纳入年度经营与内控

年度经营计划应把关键营销能力与经营目标对应，而不是单列一个抽象的“AI 建设预算”。例如提高大客户线索响应，需要研究、线索识别、评分、销售任务创建和跟进建议等能力共同作用；提升品牌一致性，需要品牌知识、Brief、内容生成、审校和发布控制共同作用。能力投资只有与经营链路对应，才能判断缺口与优先级。

预算编制时可以把技能分为公共底座、业务关键、风险控制和探索四类。公共底座由平台统一投入，业务关键由受益部门承担主要预算，风险控制由企业层面保障持续维护，探索技能设定明确期限和退出条件。这样的分类避免所有成本都压给技术部门，也避免业务部门只享受收益而不承担维护责任。

内控审计不需要阅读每一段提示词，但应抽查契约是否完整、权限是否与目录一致、版本是否可追溯、重大动作是否有审批和回执、旧版本是否按期退役。审计样本可以优先覆盖公开发布、客户触达、预算、个人信息和高监管声明。

管理层应要求重大技能具有业务连续性方案。方案包括依赖不可用时的降级路径、人工操作手册、备用模型或工具、在途任务清单和恢复后的对账。连续性演练比文档声明更有价值，尤其是大促、发布会、危机响应和季度结算前的关键时段。

技能资产还应进入人员与岗位设计。员工不再以“会不会某个模型”作为主要能力标签，而是要能够定义任务、判断证据、管理例外、审查结果和改进契约。平台人员则需要理解业务对象和责任，不只是维护 API。培训应围绕真实技能和事故样本，而不是泛化工具演示。

供应商绩效也应与技能证据挂钩。交付一个可演示页面不能视为能力完成；供应商需要提交契约、依赖、测试、权限说明、运行清单和迁移材料。持续服务应对回归、政策更新、故障响应和兼容窗口负责。只有把这些要求写入验收和付款节点，企业才可能真正获得资产。

跨部门争议应通过契约解决，而不是依赖临时会议。业务方认为输出不合格时，应指出违反哪一项验收；平台方认为输入不足时，应返回缺失字段；风险方要求停止时，应关联具体政策和适用范围。契约使争议从角色权力转向可核验条件，同时保留必要的管理判断。

技能工程也需要控制文档负担。契约字段应由平台自动带出运行数据和依赖信息，业务人员只维护必须由人判断的范围、验收和边界。对于低风险内部技能，可以使用轻量模板；高风险外部动作则需要完整证据。分级治理比所有技能套用同一厚重流程更有效。

季度复盘时，管理层可以选择三类样本深入审查：调用最多但返工高的技能，调用不多但风险关键的技能，以及新版本采用缓慢的技能。第一类可能需要重构契约或粒度，第二类需要确认持续投资，第三类可能存在兼容、培训或信任问题。

当企业能够用同一套语言讨论能力边界、权限、证据、版本和经营结果时，AI 营销才真正进入组织能力建设阶段。此前的提示词、工具和平台并没有失去价值，它们被放入了更清晰的责任结构中。技术变化仍会持续，但企业不必每次从零开始，因为业务契约、测试资产和运行证据可以跨实现延续。

能力资产化还需要处理跨地域与跨品牌差异。集团可以共享契约骨架、测试方法和平台控制，但品牌事实、语言、渠道、客户授权和监管要求应通过配置域隔离。总部不应把单一市场的规则直接复制到所有区域，分支机构也不应为了本地便利复制公共技能。公共内核与本地策略分离，能够兼顾规模经济和适用性。

对于多品牌企业，技能目录应显示品牌适用范围和继承关系。集团级禁区、数据政策和安全控制作为不可覆盖基线，品牌级语气、产品事实和审批人可以扩展。任何本地例外都要有有效期和批准人，避免临时放宽永久化。运行证据应按品牌切分，使一个品牌的稳定表现不会掩盖另一个品牌的高返工。

最终，技能工程的成熟标志不是系统可以完成多少动作，而是企业能够清楚回答每个关键动作为何被允许、基于什么事实、由哪个版本完成、谁验收、失败如何处理以及怎样继续改进。这样的透明度让管理者敢于授权，也让员工知道何时依赖系统、何时质疑系统、何时必须接管。

当这些制度稳定后，技能目录会成为企业营销能力的动态地图：它记录现有优势、关键缺口、风险集中点和下一步投资。新模型和新平台出现时，企业可以替换实现、复用契约与测试，而不必再次把组织经验交给个人摸索。能力由此获得持续性，AI 才真正成为企业生产系统的一部分。

企业在完成首轮 90 天建设后，应把技能工程纳入常态经营：年度确定关键能力组合，季度调整信任等级与投资，月度清理旧版本和重复能力，周度处理异常与路由反馈。经营节奏与技术节奏一旦脱节，目录就会重新退化为静态清单。持续治理的价值，在于让每次业务变化、模型变化和政策变化都能通过同一套契约、证据和责任机制进入系统，而不是依靠临时通知和个人记忆。

因此，董事会最终批准的不是一批孤立功能，而是一套能力治理制度：业务可以提出需求，平台能够工程化，风险角色能够设限，使用者能够理解，所有者能够维护，审计能够追溯，供应商可以替换。只有这套制度持续运转，企业营销经验才不会随着人员、模型和平台变化而反复流失。

BOARD BRIEF 16

来源索引

竹势 AI 营销智库出品

以下索引用于公开报告阅读，只列机构/作者、标题、日期或版本、章节与采用口径，不嵌入第三方可点击链接。产品能力信息均按官方文档口径处理；未公开的企业营销结果明确标注为“未披露”。

机构/作者	标题	日期/章节与口径
OpenAI	Introducing the Codex app	2026-02-02；Skills 与安全配置
OpenAI	Codex for every role, tool, and workflow	2026-06-02；技能与插件扩展
Anthropic	Extend Claude with skills	访问版；Skills 定义、结构与加载
GitHub	About agent skills	访问版；指令、脚本、资源与开放标准
Model Context Protocol	Specification	2025-11-25；工具、资源、提示与生命周期
Model Context Protocol	Tools	2025-11-25；工具名称与输入 schema
David Parnas	On the Criteria To Be Used in Decomposing Systems into Modules	1972；信息隐藏
Bertrand Meyer	Object-Oriented Software Construction / Design by Contract	1988/1997；契约式设计
W. Edwards Deming	Out of the Crisis	1986；质量系统
Semantic Versioning	Semantic Versioning 2.0.0	2013；兼容版本语言
NIST	SP 800-53 Rev.5	2020；最小权限与审计控制
NIST	AI Risk Management Framework 1.0	2023；治理、测量与管理
OWASP	Top 10 for LLM Applications	2025；提示注入与过度代理
SLSA	Supply-chain Levels for Software Artifacts	v1.0；来源与构建完整性
OpenSSF	Scorecard documentation	访问版；依赖与项目安全信号
Salesforce	Agentforce Developer Guide	访问版；topics、actions 与信任控制
阿里云	百炼官方产品与开发文档	访问版；插件、知识与工作流
腾讯云	智能体开发平台官方文档	访问版；知识、插件与流程
百度智能云	千帆 AppBuilder 官方文档	访问版；组件与应用编排
AWS	What are AI agents?	访问版；代理与环境交互定义

PUBLISHER

关于竹势 AI 营销智库

竹势 AI 营销智库出品

竹势 AI 营销智库聚焦 AI 市场部与 AI 营销，面向企业创始人、CEO、CMO及市场增长团队，构建“6+1”知识架构。六大核心模块组成“道、法、术、器、技、例”六脉体系：“道”负责认知刷新，研究 AI 时代的营销本质与战略判断；“法”关注体系建设，沉淀可持续、可复制的方法论；“术”提供落地路径，把策略转化为具体行动；“器”评测工具与平台，帮助企业选择合适的能力组合；“技”分享可以立即应用的实战技巧；“例”通过标杆案例验证方法与成效。“+1”专题围绕 AI 营销的重要趋势与关键经营问题，推出系统、深入、可下载的专题白皮书。

竹势智库通过专业研究、实践经验与管理框架连接认知、决策和执行，帮助企业看清方向、减少试错，逐步建设能够创造真实经营价值的 AI 市场部。