



企业 AI 试点为何难以停止

2026 企业 AI 项目组合：淘汰、限期修复与有边界扩大投入的结果与管理含义

报告类型
公开旗舰报告

语言
简体中文公开版

资料截止日期
2026-07-14

版本
v2.2 / 2026 双语公开版

研究与编辑机构
太华智库

编辑复核
公开版编辑与证据
边界复核

使用边界

本报告使用截至 2026-07-14 的公开资料，并以完全虚构的 6 个试点条件展示 1 个有边界扩大投入、3 个限期修复、2 个淘汰的示例结果。示例不代表真实企业，也不替代本地数据与必要的财务、法务、安全和合规复核。



企业 AI 试点为何难以停止

结果、依据与行动

目录

先读管理层结论与 6 个虚构项目的 1 / 3 / 2 示例判断，再核对证据、三道门、仪表盘与 90 天行动。

01	给管理层的结论	09	组合仪表盘：分别看价值证据与共享承载
02	为什么 2026 年更需要组合决策	10	90 天可逆执行节奏
03	看似冲突的数字，测量的不是同一件事	11	2026—2027 年重点观察
04	局部生产率变化可以测量，但不能直接等同于企业价值	12	证据、推断、假设与建议边界
05	证据较弱的试点项目为何仍可能存活	13	更新与纠错政策
06	先换比较单位：从工具转向一条责任清晰的工作流	14	来源
07	三道门：证据、转化、控制	15	延伸阅读
08	三种项目处置状态：淘汰 / 限期修复 / 扩大投入		

证据边界 调查口径不合并为单一采用率或成功率；任务层效果不直接外推为企业利润；P1-P6 的分类基于完全虚构的示例条件，不代表真实企业结果。

- 结果优先阅读路径**
- 先看 1 个扩大投入、3 个限期修复、2 个淘汰，以及 P1-P6 的逐项理由。
 - 再核对不同调查口径和任务研究，避免把采用、投资回报、利润影响与局部生产率混为一谈。
 - 最后用三道门、组合仪表盘和 90 天节奏，把示例结论转成可验证的本地行动。



企业 AI 试点为何难以停止

2026 企业 AI 项目组合：淘汰（Kill）/ 限期修复（Repair）/ 扩大投入（Scale）的判断结果与管理含义

如果你负责企业 AI 预算、业务流程、技术交付或风险管理，而试点数量增长快于集成、复核和变更能力，这份报告正是为你准备。它只回答一个组合决策：哪些项目可以扩大投入，哪些只能限期修复，哪些应停止当前的资源优先权？

给管理层的结论

本报告判断，2026 年许多企业需要检查一项组合风险：试点项目增长快于可比较证据和共享交付能力。发生这种错配时，弱证据项目可能继续占用集成、复核和变更资源，证据更强的项目则被迫等待。

在下述完全虚构的 6 项目组合中，1 个项目符合有边界扩大投入条件，3 个只能进入限期修复，2 个应停止当前对稀缺资源的优先占用。该分布只用于说明判断标准，不代表任何真实企业、行业基准或项目建议。

6 个虚构试点项目：示例判断

P1–P6 是说明性标识。所有证据状态、约束条件和判断结果均为虚构。

试点项目	公开判断	关键依据	下一步边界
P1 客服回复草拟	限期修复	只观察到响应时间变化；缺少可信对照，重开率、资深人员复核和错误建议未完整记录	不扩大现有范围，补齐对照与质量负担；明确的具体缺口到期未关闭，则停止当前优先占用
P2 销售会前研究	限期修复	只有主观节时，CRM 写入尚未获批	保持只读，确认业务是否实际接纳结果；不得扩大系统权限
P3 发票异常预检查	限期修复	已有比较证据，但漏报严重程度尚未关闭	限于只读预检查；漏报风险得到验证与控制前不得获得付款权限
P4 内部知识检索	有边界扩大投入	可比较结果已在至少 2 个复核周期重复出现；已指定负责人，只读权限过滤器和运行负担保持稳定	仅在经过权限过滤的标准作业程序（SOP）语料库内扩大；持续监测来源新鲜度、权限过滤和异常情况
P5 合同问题识别	淘汰（停止当前优先占用）	只有观察证据；漏检影响不对称，法务复核构成当前瓶颈	释放当前扩大投入优先权；保留测试与问题分类资产，风险或复核能力变化后再评估
P6 代码变更智能体	淘汰（停止当前优先占用）	变更建议数量增加，但生产系统写入、回滚和权限边界仍构成硬性阻断	不进入生产环境扩大；只有控制缺口关闭并出现新证据后才能重新提出

企业 AI 试点为何难以停止

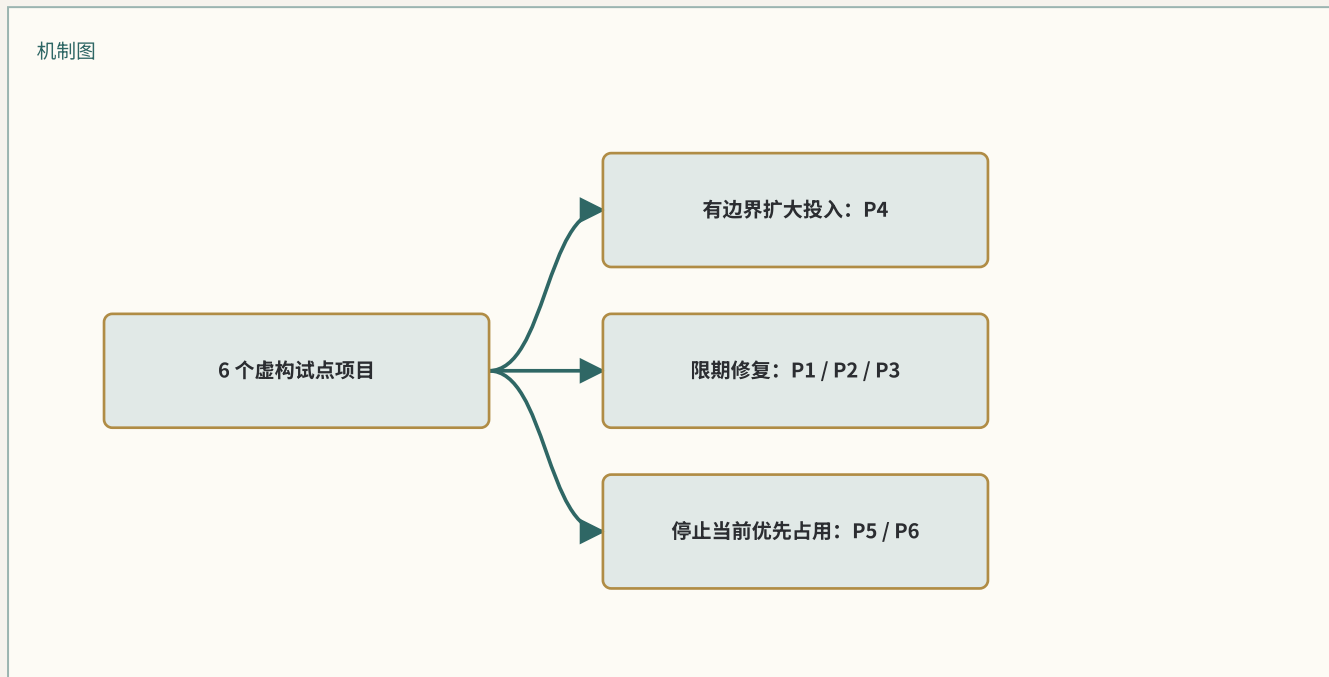


图 1 | 虚构项目组合的示例判断。“淘汰”只表示取消当前组合周期内的资源优先权，不是永久禁止，也不是对 AI 或项目团队的整体否定。

管理含义与主要风险

- 扩大投入需要结果证据、运行负担和控制条件同时稳定。仅凭使用量或一次演示，不足以获得扩大投入资格。
- 限期修复需要明确的具体缺口、负责人、截止日期和默认退出条件，否则会变成无限期保留。
- 系统写入、权限、回滚和高影响漏报属于硬性约束，不能由潜在收益抵消。
- 组合判断需要纳入共享团队承担的集成、复核、返工、异常处理和维持成本。
- 停止当前优先权后，应保留对照设计、问题分类和控制记录，供条件变化后的重新评估使用。

为什么 2026 年更需要组合决策

BCG 2026 年高管调查显示，2,360 名受访高管所在企业预计 AI 支出将从 2025 年约占营收 0.8% 上升到 2026 年约 1.7%；94% 的受访者表示，其所在组织即使在 2026 年没有回报，仍会继续投资。该调查测量的是计划、信心和压力，不是经审计的 ROI。它不能证明停止已经变难，但提示受访组织中的停止、缩减或延期决定可能面临更高的组织与声誉阻力。

McKinsey 2025 年 AI 状态调查覆盖 1,993 名受访者，其中 88% 表示所在组织至少在一个职能中经常使用 AI，近三分之二尚未进入企业级规模化；39% 报告了任何企业级 EBIT 影响。**IBM 2025 CEO Study**给出另一组口径。在 2,000 名受访 CEO 的组织中，25% 的 AI 项目达到预期 ROI，16% 实现企业级规模化。另有 50% 的 CEO 表示，快速投资造成了技术割裂。



企业 AI 试点为何难以停止

这些数字不能相互平均，因为样本、问题和分母不同。它们支持的管理含义有限但重要：企业继续投资的意愿，可能强于可归因结果的成熟度。管理层因此需要一套跨项目一致的续投标准，而不能让每个试点项目自行选择成功口径。

看似冲突的数字，测量的不是同一件事

来源	样本或方法	测量对象	可支持的判断	不能支持的判断
U.S. Census 2026	全国代表性 BTOS AI 补充调查	企业是否在业务职能中使用 AI	在官方企业框架中，使用深度仍有限	项目 ROI 或管理质量
NBER <i>Firm Data on AI</i>	近 6,000 名四国高管	企业使用、过去影响和未来预期	广泛使用可与有限的已实现影响并存	全球企业失败率
McKinsey 2025	1,993 名在线调查受访者	组织使用、规模化和 EBIT 影响	受访组织的规模化与影响不均衡	全国代表性采用率
IBM CEO Study 2025	2,000 名 CEO	项目是否达到预期 ROI	CEO 感知的价值转化缺口	经审计的标准化 ROI
Microsoft / NBER / METR	现场实验或任务研究	特定任务中的生产率	部分场景记录到可测局部效果，且结果具有异质性	整个企业利润

U.S. Census 2026 年工作论文估计，2025 年 11 月至 2026 年 1 月约 18% 的企业在至少一个职能使用 AI；在职能采用企业中，57% 只覆盖三个或更少职能；在报告员工任务使用的企业中，约 65% 只覆盖三个或更少员工任务。该调查的问法在 2025 年 11 月发生变化，因此不宜与旧序列直接拼接。

NBER 2026 年 *Firm Data on AI*基于美国、英国、德国和澳大利亚近 6,000 名高管。69% 的受访企业表示正在积极使用 AI；超过 90% 报告过去三年没有观察到就业影响，89% 报告没有观察到生产率影响，超过 80% 对两者都没有观察到影响。该结果来自高管自报和四国样本；它不能证明企业没有任何局部收益，却能提醒管理层区分“在使用”和“已改变企业结果”。

因此，本报告不制作统一的采用率或成功率图表。把不同分母画成同一条趋势线，会制造比原始研究更强的结论。

局部生产率变化可以测量，但不能直接等同于企业价值

任务研究显示，部分工作场景确实出现了可测变化。NBER 客服现场研究对 5,179 名客服人员进行分批上线比较。使用生成式 AI 后，每小时解决的问题数平均提高 14%；收益主要集中在经验较少、技能较低的员工。Microsoft Research 三项现场实验覆盖 4,867 名开发者，报告已完成任务量约提高 26%，标准误为 10.3 个百分点。这些研究支持一个有限判断：AI 在部分结构化工作中能够产生可测改善。

但不同任务的变化并不同向。METR 2025 年开源开发者研究覆盖 16 名熟悉大型开源代码库的资深开发者和 246 个真实任务；使用 2025 年初的 AI 工具后，完成时间平均增加 19%，参与者却在事后认为自己加快了 20%。该结果只能作为历史反例：METR 已明确标注这项 2025 年结果过时，不能代表当前工具影响；其 2026 年后续更新也未给出一个可替代的可靠当前估计。

管理上，不宜从这些研究中抽取一个平均生产率数字；更有用的做法是区分三层效果：



企业 AI 试点为何难以停止

- **任务效果：**一个具体任务是否更快、更准；
- **工作流转化：**局部改善是否减少了整个流程的周期、错误或等待；
- **项目组合价值：**该改善是否值得占用共享的接入、复核、风险与变更能力。

一个试点项目可以在第一层成功、在第二层无效、在第三层仍不值得优先。组合决策因此需要同时保留收益证据、共享成本和机会成本。

证据较弱的试点项目为何仍可能存活

不同参与者面对的收益和损失并不对称。AI 产品团队可以从使用量、功能上线和项目发起人认可中受益；工作流负责人需要处理边缘案例和人工兜底；CIO 承担长期集成与维护；CISO 和法务部门可能承担低概率但高影响的事件；CFO 则被要求证明这些分散变化是否进入利润表。

当每个部门自行选择成功指标时，项目比较可能退化为叙事竞争。图 2 概括了这条机制链。

该循环可以用委托—代理问题提供一条解释路径。项目发起人更接近试点项目，也更早掌握演示和局部数据；共享团队更了解长期运行负担，却往往在预算批准后才进入。信息不对称可能使证据缺口在决策界面中不易被识别，也可能使组织把“停止”误读为失败。



企业 AI 试点为何难以停止

机制图

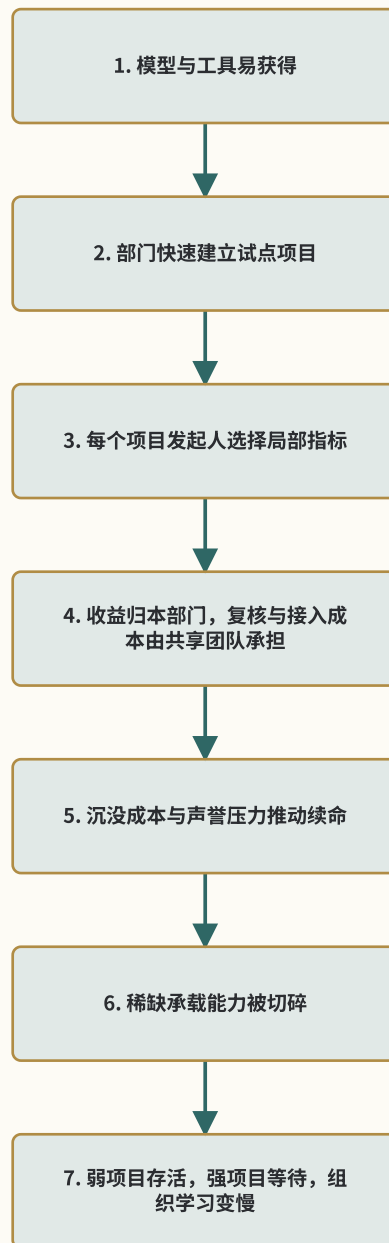


图 2 | 试点项目存续机制链。 该图提出一项待验证的机制假设：局部指标、共享成本和沉没成本可能共同延长弱证据项目的存续时间，并挤压证据更强项目所需的共享能力。具体强度需由企业项目历史、成本台账和退出记录验证。



企业 AI 试点为何难以停止

先换比较单位：从工具转向一条责任清晰的工作流

每个试点项目在争取共享交付能力前，应提交同一张一页式合同。合同以一条端到端、可验收的业务工作流为分析单位；模型、席位和功能只是输入。

本文所称“承载能力”，指集成、复核、数据、安全和变更等共享交付能力；“业务采纳且下游有效的结果”，指成果已被业务使用，并在后续环节继续有效；“硬性门槛”指不能由其他高分抵消的边界；“系统写入”指改变业务系统状态；“回滚”指把流程恢复到已知安全状态。

字段	最低要求	不足时的处理
工作流	开始事件、结束结果、负责人	无法命名则淘汰或退回备选项目池
业务采纳且下游有效的结果	已被业务使用，并在后续环节继续有效	只有原始输出则进入限期修复
基线 / 对照项	同期、分批、相近队列或稳定历史线	只有主观节时不得提出 ROI 主张
测量窗口	起止日期、任务量、异常定义	无期限试点项目不占用集成承载能力
运行负担	人工复核、返工、接入、监测、变更与备用方案	缺项需补齐或保留为不确定成本
负责人	业务、技术和风险负责人	任一缺失不得扩大投入
停止规则	错误、权限、复核或维护越界条件	测试前未写明则不启动

这张合同不要求不同项目使用相同阈值，而是要求它们用同一组字段说明结果、运行负担和风险。高频、低风险的知识检索与低频、高影响的发票控制智能体可以采用不同门槛，但都需要说明证据等级和停止条件。

三道门：证据、转化、控制

第一道门 | 可比较证据

管理团队先检查是否存在具名工作流、可验收结果、基线或对照项、测量窗口、异常定义和三类负责人。只有使用量、消息量、生成量或演示速度，仍可说明项目具有学习价值，却不足以与具备结果证据的项目争夺扩大投入预算。

第二道门 | 转化证据与运行负担

本报告不展开企业级 ROI 模型、完整反事实设计或加权优先级排序。组合层只比较三项：结果相对基线处于哪一证据等级；实际承载团队确认的运行负担；以及项目占用共享能力后被延后的其他机会。

字段	组合层记录	未知时的处理
转化证据	未知 / 已观察 / 已比较 / 已重复	不得进入扩大投入
运行负担	未知 / 低 / 中 / 高；项目比较前以小时/月、集成天数或成本区间定义本地档位，并由实际承载团队确认	未定义档位或未获确认时进入限期修复
机会成本	集成资源额度、安全复核名额、验证小时数及被延后的项目	由预算负责人明确确认



企业 AI 试点为何难以停止

详细的价值、成本与因果归因需要企业本地数据和单独的工作流测量包。缺少基线时，不应使用行业平均数补出一个精确 ROI。

第三道门 | 控制与约束适配

NIST 生成式 AI 风险管理框架 (GenAI Profile) 将治理、内容来源、部署前测试和事件披露列为四项主要考虑事项，并在建议行动中讨论持续监测与事件响应。NIST 2026 AI 智能体标准倡议 将智能体可靠性、互操作性、安全和身份置于重点；其启动材料还指向 NCCoE 的软件与 AI 智能体身份授权工作，以及智能体访问外部系统和内部数据时的上下文。

本框架因此把权限、可追踪性、人工升级、回滚和业务连续性纳入上线前检查。对于涉及系统写入、客户沟通、资金动作、受监管判断或生产变更的试点项目，尚未关闭的控制缺口被定义为硬性阻断条件；该框架判断不替代企业自身的法务、安全或合规复核。

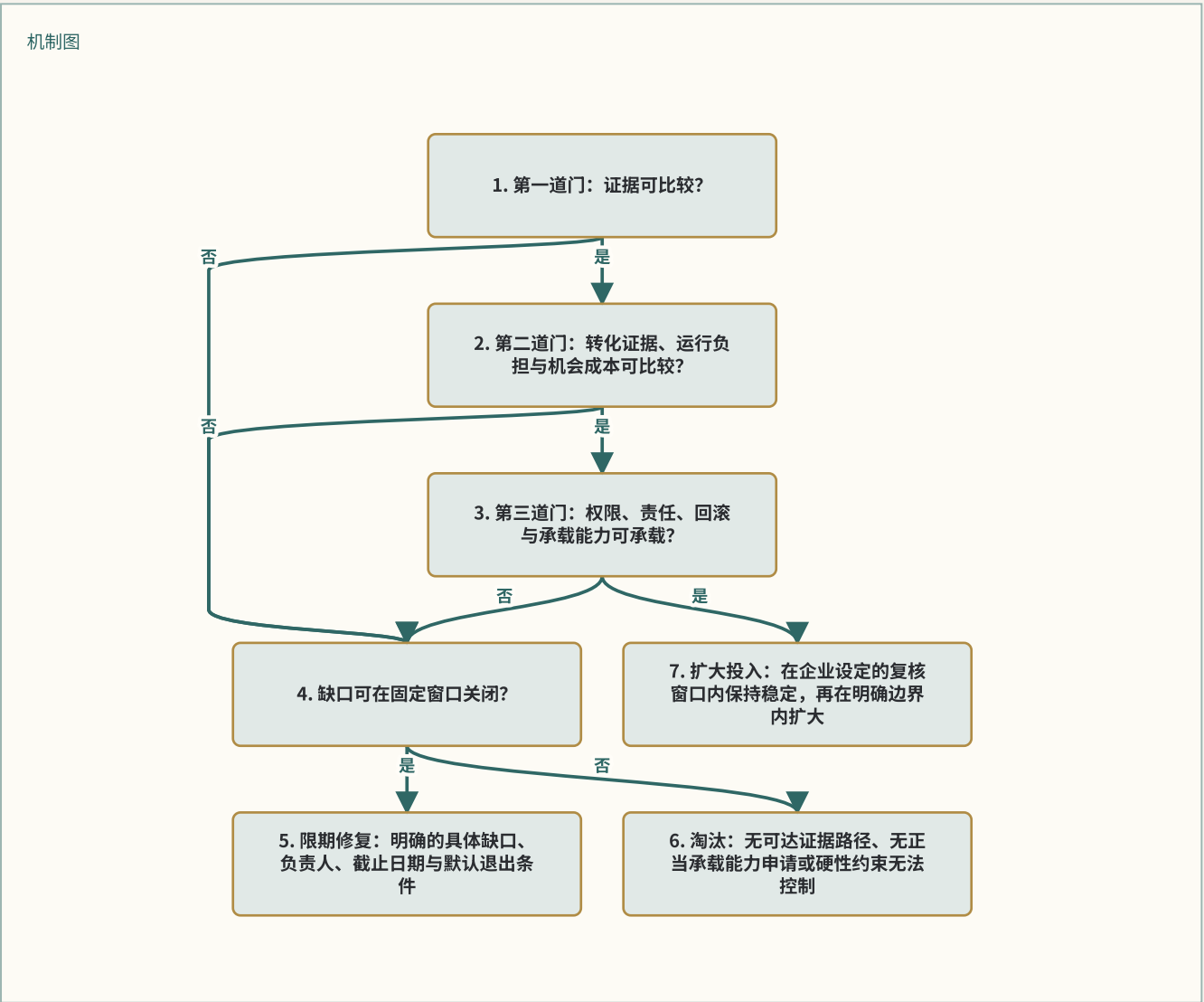


图 3 | 三道门与三种项目处置状态。任何一门发现缺失证据或硬性阻断条件，都可能进入限期修复、退回备选项目池或停止；只有三门均通过才讨论有边界的扩大投入。连续 2 个复核周期仅是本文采用的审慎示例，不是行业基准；企业需按 workflow 风险和本地承载能力设定。



企业 AI 试点为何难以停止

三种项目处置状态：淘汰 / 限期修复 / 扩大投入

“继续保留”容易成为没有期限的中间状态。限期修复则需要同时写明：哪一个缺口、由谁修复、何时复核、什么证据算关闭、到期未关闭是否自动淘汰。

状态	适用条件	必须留下的记录
淘汰	没有可达证据路径；硬性约束无法控制；机会成本明显更高	终止理由、已学到什么、可复用资产、何时可重新提出
限期修复	价值仍可能成立，且一个明确、具体的证据或控制缺口可在固定窗口关闭	负责人、截止日期、所需证据、默认退出条件
扩大投入	转化证据、运行负担和控制条件跨至少两个复核周期稳定，且机会成本已由预算负责人确认	有边界范围、承载能力预算、监测、回滚、下次复核

在本框架中，淘汰取消的是当前项目对稀缺承载能力的优先占用，不构成对 AI 或团队的整体否定。若新证据、成本下降或控制改善改变条件，项目可用新合同重新进入备选项目池。

组合仪表盘：分别看价值证据与共享承载

价值与证据

指标	推荐定义	常见误读
试点项目转化	想法 → 测试 → 已比较 → 扩大投入的数量与停留时间	项目越多代表创新越强
有效采纳率	业务采纳且在下游仍有效的结果 / 总结果	输出量等于价值
复核负担	每 100 个结果的专家复核、异常处理与返工小时数	人工参与自动等于安全
证据深度	未知、已观察、已比较、已重复在项目组合中的分布	单次前后对比等于因果证据

承载能力与控制

指标	推荐定义	常见误读
集成拥堵	等待接入天数、共享依赖项、维护队列	API 接通等于完成集成
控制完整性	权限失效、不安全操作、回滚成功率，以及按严重程度分层的每 1,000 次运行事件数	低报告率或没有事故等于控制条件成熟
逾期末决项目率	超过限期修复截止日期且仍未产生所需证据的试点项目占比	持续运行等于保留长期选择权
重新分配速度	从淘汰决定到承载能力重新分配的时间	终止本身已经释放价值

这些指标仍需本地定义。它们的用途是让管理层看见项目组合学习和瓶颈，而不是制造另一个综合成熟度分数。



企业 AI 试点为何难以停止

指标完整性规则

1. 测试前冻结纳入比较的项目或任务范围、分子、分母和排除规则；任何改变都需要保留版本并重启比较窗口。
2. 由工作流负责人确认哪些结果已被业务采纳并在下游继续有效；项目发起人不得自行确认分母、成本或控制条件。
3. 记录排除案例、隐性工作、人工绕行方案与险情；缺失项记为“未知”，不得记为零。
4. 每个项目使用不可重置的试点编号，防止改名后清零限期修复截止日期或历史事件。
5. 控制完整性同时报告事件严重度、报告完整度和回滚成功率。
6. 重新分配速度需要与重新分配后的已实现结果配对，避免奖励草率终止。

为避免组合机制只偏向容易测量的低风险项目，可为高不确定性项目保留一个有上限、会到期的验证窗口。该窗口不能绕过生产环境的硬性门槛，也不能自动续期。

90 天可逆执行节奏

第 1-10 天 | 盘点

列出所有 AI 试点项目、负责人、工作流、当前权限、共享依赖项、累计支出、已有证据和下一次复核。若当前集成队列已明显拥堵，可暂缓新增投入承诺；低风险想法探索仍可继续。

第 11-30 天 | 标准化

所有项目使用同一张一页式证据合同。使用指标与业务结果分开记录；缺失基线、复核成本、权限和回滚直接列为待关闭缺口。

第 31-60 天 | 有边界测试

在共享承载能力确实稀缺的企业中，优先向证据就绪或高学习价值的试点项目分配资源。可使用影子模式、只读、决策支持、分批上线和可恢复的人工流程，并记录负面结果与未观察到效果的结果。

第 61-75 天 | 压力测试

在形成最终决定前，用成本上升、事故、关键人员离开、审计要求或需求变化检验临时结论。若项目无法安全降级、回滚或保持明确责任，结果应退回限期修复或淘汰。检查范围还应包括分母、激励、负责人缺位和共享成本转移，而不能只看模型准确率。

第 76-90 天 | 项目组合决策

对每个项目形成淘汰、限期修复或扩大投入决定。限期修复需要截止日期和默认退出条件；扩大投入需要有边界范围、承载能力预算、监测、回滚和下次复核。将释放的承载能力明确重新分配，避免它被无形消耗。

2026—2027 年重点观察

以下是方向性观察清单，不是概率预测。



观察 1 | 从采用叙事转向项目组合问责

- **方向性预期：**若预算和董事会压力延续，CEO、CFO、COO 与技术和风险负责人更可能共同决定项目应继续投入、限期修复还是停止。
- **领先指标：**公开案例开始披露工作流负责人、验证负担、完整运行成本、淘汰决定与承载能力重新分配。
- **证伪条件：**大型企业持续只按席位、消息量或试点项目数量扩大，且没有出现明显的预算、事故或维护约束。
- **复核日期：**2026-10-15；2027-01-15。

观察 2 | 智能体身份与授权进入扩大投入门槛

- **方向性预期：**系统访问权限增加后，身份、最小权限、可重建日志、人工升级和回滚更可能成为客户与架构审查的前置条件。
- **领先指标：**NIST 倡议形成可实施标准；采购、审计或网络保险要求智能体身份和行动轨迹。
- **证伪条件：**大多数智能体部署长期停留在低风险、无系统写入的范围，相关标准不影响企业采购和上线。
- **复核日期：**2026-10-15；2027-02-15。

观察 3 | 高质量淘汰逐步成为学习指标

- **方向性预期：**若项目组合拥堵上升，成熟团队可能开始记录淘汰质量、可复用学习和重新分配速度。
- **领先指标：**转型仪表盘披露已取消的试点项目、限期修复截止日期和复用的控制措施，而不只披露采用情况。
- **证伪条件：**组织绩效和融资叙事继续单向奖励项目数量与扩大，没有可见的停止机制。
- **复核日期：**2027-01-15。

证据、推断、假设与建议边界

事实。本报告引用的调查、官方企业数据和现场实验支持以下事实：AI 使用与支出意愿上升；不同调查记录的企业级规模化与可归因影响并不均衡；任务效果具有异质性。NIST 2026 AI 智能体标准倡议将智能体可靠性、互操作性、安全和身份置于重点；其启动材料也指向身份与授权工作，以及智能体访问外部系统和内部数据时的上下文。

推断。本报告推断，多个试点项目争夺共享承载能力时，组合选择和跨职能责任可能成为约束。来源支持集成、复核、数据、变更和治理摩擦，但不能证明这些因素在每一家企业都是第一瓶颈。

假设。局部指标、成本转移、沉没成本和项目发起人激励可能提高弱试点项目的存活率。该机制需要用企业自己的项目组合历史记录、会议记录和成本台账检验。

建议。使用统一证据合同、三道门和固定期限的限期修复；向管理层呈现决定、证据等级、未决风险、边界范围、下次复核与默认退出条件。该建议是一项可逆的管理实验，不构成会计、法律、合规、投资或特定供应商建议。



企业 AI 试点为何难以停止

不确定性。 本报告没有接入任何具体企业的项目级基线、跨项目对照和完整经常性成本台账，因此只能提供组合判断标准，不能替任何企业决定哪个项目应当扩大投入。

更新与纠错政策

- 来源截止日期：2026 年 7 月 14 日。
- 若主要来源修订样本、口径或结论，应重新核对相关页面。
- 若出现更强的企业级因果证据、公开负面案例或智能体标准，应触发证据更新。
- 若后续真实项目组合结果与本报告的判断不符，应保留原判断、记录偏差，并以带日期的复核或更正说明更新，而不是静默改写。
- 本版本为公开免费报告，版本号 v2.2。若实质性证据改变结论，将保留原判断，并发布带日期的更正说明。读者在应用时仍需使用本地数据，并完成企业自身的财务、法务、安全与合规复核。

来源

- [NBER, Firm Data on AI \(2026\)](#)
- [U.S. Census Bureau, The Microstructure of AI Diffusion \(2026\)](#)
- [BCG, As AI Investments Surge, CEOs Take the Lead \(2026\)](#)
- [McKinsey, The State of AI 2025](#)
- [IBM, CEO Study 2025](#)
- [Microsoft Research, The Effects of Generative AI on High-Skilled Work \(2025\)](#)
- [NBER, Generative AI at Work \(2023\)](#)
- [METR, Early-2025 AI Study of Experienced OSS Developers](#)
- [METR, 2026 Uplift Update](#)
- [NIST AI 600-1, Generative Artificial Intelligence Profile](#)
- [NIST, AI Agent Standards Initiative \(2026\)](#)
- [NCCoE, Software and AI Agent Identity and Authorization](#)

延伸阅读

以下资料未用于核心事实判断。

- [OECD, Empowering SMEs in the Age of AI \(2026\)](#)
- [BCG / MIT SMR, Responsible AI Needs More Than Good Intentions \(2026\)](#)
- [Anthropic, The 2026 State of AI Agents](#)
- [BCG, CEOs and Boards Are Aligned on AI in Theory but Divided in Practice \(2026\)](#)