产品同质化与竞争力不足：用 AI 做差异化研发与快速迭代

作者：**SVSAC 硅谷战略顾问委员会**

核心判断

AI 研发最值得投入的方向，是把客户尚未被满足的需求转化为可验证的产品差异，而不是批量生成相似方案。对企业而言，值得优先推进的是：从一个产品族和一项明确痛点切入，让设计、工程与销售共同确定试验指标，再用样机和客户反馈决定是否扩大投入。本篇结合公开资料与分析，讨论企业可以如何识别机会、组织能力并安排下一步投入。

关键词：人工智能；产品差异化；研发迭代；组织学习；证据治理

SVSAC - sv-sac.org

AI 研发最值得投入的方向，是把客户尚未被满足的需求转化为可验证的产品差异，而不是批量生成相似方案。

我们的建议是：从一个产品族和一项明确痛点切入，让设计、工程与销售共同确定试验指标，再用样机和客户反馈决定是否扩大投入。企业可以先据此梳理自己的产品、客户、能力与投入边界，把讨论推进到可以分工、可以验证的具体任务。

对管理层来说，这个议题值得放进经营议程。围绕“先做什么、凭什么做、做到哪一步再扩大投入”形成共识，往往比同时铺开多个方向更有推进价值。下文保留相关资料、比较和推演，供企业判断自身条件与行动顺序。

1 引言：从方案稀缺转向可验证价值稀缺

1.1 产品相似不等于问题相同

讨论产品同质化，首先需要辨别观察对象。两款产品在外观上相近，并不能证明其价值相同；同样，颜色、包装或命名不同，也不能证明客户获得了有意义的新价值。对采购者而言，稳定交付、维修便利、兼容现有设备、降低使用负担，可能比可见造型更能影响选择。本篇因此不把差异化等同于形式上的新奇，而把它视为企业、客户、使用情境与替代方案之间的关系。差异必须指向某类客户的具体问题，且其收益需要与转换成本、购买风险以及实际可获得性共同考察。

这一界定具有方法上的意义。如果研究仅以 AI 能够生成多少设计草图来评价研发成功，实际上只测量了供给端的输出速度。企业可以在一天中得到大量概念，却仍然无法回答哪一项值得开模、哪一项会引发售后问题、哪一项能够被目标客户理解。方案数量增加甚至可能放大筛选负担。由此可提出一个待检验判断：当验证能力没有同步提高时，生成成本下降可能使企业积累更多未经验证的研发库存，而不是获得更快的有效学习。

我们关注具备产品开发和交付职能的制造企业，尤其是需要在成本、交期、质量和客户定制之间取舍的团队。这里的企业情境是分析边界，不是已经调查得到的代表性样本。对于河北制造企业的具体人员结构、数据成熟度和研发投入，本篇没有可供估计的企业原始资料，故不陈述省域普遍采用率、平均提效比例或确定的投资回报。这种边界并不妨碍提出可操作的研究框架，但要求把实施建议与观察事实严格区分。

SVSAC | sv-sac.org

1.2 AI 可能提高表现，也可能放大趋同

Noy 和 Zhang 的写作实验涉及 453 名受过大学教育的专业人员。MIT 研究机构报道给出任务耗时平均下降 40%、质量评分提高 18%，同时指出任务不要求特定企业背景，评价未包括专门事实核查。[1] 上述数字来自 MIT 机构报道而非论文原页，两者不能等同。时间与评分改善不能证明事实可靠性同步提高，更不能直接换算工业设计的良率、研发周期或新品毛利。企业试验须把事实核查和工程返工工时计入，检验表面节时是否转移了成本。

Doshi 与 Hauser 的作者预印本摘要报告，故事创意辅助改善个体作品评价，同时使作品集合更相似。[2] 本节依据的是作者预印本摘要，正式版全文未能取得，因而不补写样本、量表或正式版稳健性。另列的 Meincke 等头脑风暴讨论与 Lee、Chung 回复属于另一组作者的往返，不应暗示它们就是 Doshi、Hauser 的回复。[8,9] 两篇原页均未能读取。本篇保留“个体质量与集合多样性须分别测量”的核心问题，不以未经核对的争论替代制造研发证据。

对于制造企业，这一矛盾提出了一个值得检验的机制问题：如果不同团队向相似模型输入近似问题，并使用相似评分表筛选答案，其产品方向是否会趋于集中？这是从文献引出的推断，而非本篇观察到的行业事实。验证该机制需要追踪输入信息、模型版本、人工构思、候选方案和筛选过程，不能仅在发布后的产品上计算视觉相似度。

1.3 核心问题与贡献边界

本篇围绕三个相互连接的问题展开。第一，AI 在哪些研发环节可能扩大可探索空间，在哪些环节可能压缩差异？第二，企业应如何判断效率收益是否转化为客户价值，而不是把工作量转移给工程师、审核人员或售后团队？第三，在缺少大规模数据与专门算法团队时，企业如何建立可复查、可停止、可渐进扩展的试验？

本篇的贡献是提出一套可被企业数据反驳或修正的分析框架，而非宣告 AI 必然提升竞争力。文章将事实主张关联到来源编号，将机制解释表达为条件命题，将企业动作表述为建议。真实部署效果需要后续研究检验。这样的定位避免将研究文稿包装为已经完成的企业咨询项目，也避免以工具演示代替研发证据。

SVSAC | sv-sac.org

2 研究设计与证据边界

2.1 定向证据梳理的方法

本批检索执行于 2026 年 10 月 23 日，围绕生成式 AI 的任务绩效、创意多样性、知识检索与风险治理进行定向查找。优先检查期刊或会议原始页面、作者公开版本及官方制度文件。检索以核心机制为导向，尚未实施完整数据库覆盖、双人筛选或预注册流程，故本篇不称为系统综述，也不报告虚构的初检文献数量与排除流程。

每条来源记录包括题名、作者或机构、发表年份、永久链接、实际可读取的材料、支持范围和外推限制。出版商正文不可访问时，记录可读的作者版本或机构摘要，同时保留版本差异；没有阅读全文的来源不据以推断其方法细节。来源是否存在、正文是否可读以及是否足以支持某项具体主张，是三种不同的核验任务。本篇不把搜索结果出现过题名等同于阅读全文。

2.2 证据单位与研究类型

证据单位首先是研究中的具体任务、样本和结果变量，而不是论文标题中的宽泛技术名词。写作随机实验可以为短时写作绩效提供证据，客服的分期部署研究可以为特定工作环境的采用效果提供证据，但两者对新产品开发的适用性均需另行论证。创意评价研究可以提示方案生成机制，不能替代目标客户对具体产品的选择试验。

Brynjolfsson、Li 与 Raymond 正式发表的客服研究使用 5172 名人员的数据，报告每小时解决问题数平均提高 15%，并指出不同经验和技能群体的效果存在差异。[3]本篇统一使用正式发表版本的口径，不混用早期工作论文的样本和结果。这个案例支持对岗位异质性的关注，不支持以单一平均数预测制造企业各岗位的收益。

技术方法类文献与治理文献还具有不同的证据地位。Lewis 等提出的检索增强生成架构将模型生成与外部检索结合。[4]它为知识调用机制提供技术参照，不证明企业内部文档天然完整、最新或有权使用。NIST 风险管理框架提供治理、情境识别、测量和管理的组织语言。[5]这些指导性文件并不是特定项目合法或商业成功的认证。

2.3 事实、推断和建议的表达规则

本篇采用三层陈述。事实层只陈述能够由已核验材料支持的研究发现或制度内容。推断层解释这些发现可能如何影响企业研发，必须说明前提、替代解释和

SVSAC | sv-sac.org

失效条件。建议层提出企业能够实施的动作，并说明评价方式；建议本身不构成有效性的经验证明。

例如，某研究观察到 AI 使用者的输出更相似，属于事实层；“使用同一模型且缺少独立客户资料可能导致候选方案趋同”属于推断层；“在团队见到模型建议之前保存各自的独立方案”属于建议层。三者之间存在逻辑联系，但证据强度不能自动传递。保留这一分层有助于企业把失败反馈用于修正机制，而不是把执行偏差当作唯一解释。

2.4 缺失数据与可识别性

目前没有企业级流程日志、成本账、客户选择数据或产品质量追踪资料，因而不能估计研发投入的回报率，也不能比较采用 AI 与未采用 AI 企业的平均绩效。更不能把虚构企业名称、项目团队和访谈对话写入案例。后文的情境示例应明确标注为分析性情境，所有数字示例均须标注假设。

即使将来获得采用前后的资料，因果识别仍然存在困难。企业可能同时更换负责人、调整产品组合、改善供应商或进入新市场。销量变化也可能来自价格、促销和季节因素。研究因此需要记录共同变化，并根据实施条件选择随机试验、分期引入或匹配比较，而不是把采用前后差额全部归功于 AI。

3 概念框架：差异化的四个检验条件

3.1 需求相关性：差异要对应真实使用问题

需求相关性要求企业描述谁在何种条件下遭遇何种损失。泛称“年轻客户喜欢智能化”不足以指导研发；更可检验的表述应说明使用频率、操作环境、失败后果和当前替代办法。模型可以帮助组织访谈提纲、归类反馈和提出遗漏问题，但模型推演的用户画像不能直接视为真实用户需求。

本次分析建议把需求材料保留为可追踪的证据单元。每项材料记录采集场景、时间、对象范围、原始表述与分析标签，并区分客户明确表达的问题、研究者归纳的主题以及模型补充的假说。三类记录可相互关联，但不应合并为同一个“已验证需求”。这样能够在产品方向发生争论时回到原始信息，而不是只比较不同人员对模型摘要的理解。

3.2 可感知性：技术差异要能转化为选择理由

技术参数改善并不自动意味着客户愿意改变选择。若某项性能超过实际使用需求，进一步提升可能只增加成本；若采购者无法观察改进，企业还需找到可信

SVSAC | sv-sac.org

的展示或验证方式。因此，研发评价不仅要问性能是否变好，还要问改进在何种情境下可被识别，以及谁有权决定是否采购。

我们建议在概念阶段区分使用者、采购者、维护者和渠道方的评价标准。对于同一功能，使用者可能关注操作负担，采购者关注总成本，维护者关注故障定位，而渠道方关注解释难度和退货风险。AI 可以整理这些标准之间的冲突，却不能替企业决定所有利益相关方的权重。权重选择应成为可审查的管理决策。

3.3 可交付性：差异必须穿过工程与供应约束

可交付性强调设计不能停留在图像或文字上。一个概念是否能由现有工艺制造、是否依赖不稳定材料、是否增加装配误差，均需通过工程检查或实物验证。生成系统给出的尺寸、材料或性能承诺若没有对应试验依据，只能作为待核方案。把渲染效果直接当作工程图或把解释流畅当作物理可行，会使迭代在更晚阶段付出修正成本。

因此，本篇把约束审查放在候选方案进入详细设计之前，并建议把不满足安全、法规或关键接口条件的方案剔除，不以其他维度的高分进行抵消。对于非硬性约束，可保留权衡空间并记录理由。此处提出的是评审结构，不是某行业通用的技术标准；具体门槛必须由实际产品和适用规范决定。

3.4 可持续性：首轮成功要形成可复用能力

差异化能否持续，取决于企业能否把验证结果和失败原因转化为下一轮可用的信息。若研发团队只保存最终方案，不保存被排除的方案及原因，后续人员可能重复已知错误。若企业只能依赖某位员工的私人提示词和账号，知识并未形成可靠的组织资产。

我们建议将迭代记录至少连接到需求、方案、约束、测试、决定和结果六类对象。模型版本、检索材料和人工修改属于产生方案的背景信息，也应在影响复现时记录。组织学习并不要求无限保存所有个人信息或全部对话；《个人信息保护法》要求处理目的明确、直接相关并遵循最小必要原则。[7]记录的范围和保留期限需要围绕研发目的设计。

4 待检验的机制命题

4.1 搜索扩展与筛选瓶颈

命题一：在需求证据和评价能力保持不变时，AI 增加候选方案数量，未必提高最终入选方案的客户价值。其可能路径是生成速度超过筛选和测试能力，使团

SVSAC | sv-sac.org

队更依赖表面完整性或容易评分的指标。相反，如果团队能够提高验证吞吐量并保留不相似的候选方向，生成扩展才可能带来更有价值的搜索。

验证这一命题应同时记录生成数量、有效去重后的方向数量、进入测试的比例、被拒绝的理由和测试结果。只比较“人均方案数”会遗漏关键中介过程。本篇不预设这些指标的最佳阈值，因为适合工业配件的方案规模与适合消费品外观的方案规模可能不同。

4.2 知识调用与事实约束

命题二：外部检索与企业资料接入只有在证据有效且权限清晰时，才可能改善研发判断。若系统检索到过时规格、不同型号或未经批准的供应商承诺，增加上下文反而可能提高错误答案的可信度。因此，应将检索是否命中与命中的材料是否适用分开评价。

建议在每条关键技术主张旁保留材料标识、版本、适用型号和核验责任人。对于没有证据的内容，系统输出应允许保留“未知”或转入人工确认，而不要求每个字段都生成看似完整的答案。这一建议由知识治理问题推导，不是 RAG 原始研究已经验证的企业控制效果。[4]

4.3 协作增益与责任转移

命题三：AI 对某岗位节省的时间可能被其他岗位的复核、返工或解释成本抵消。因此，效率评估应跨越完整研发流程，至少覆盖需求澄清、概念生成、工程复核、试制、测试与变更协调。某岗位局部时间下降与企业总周期缩短可以同时成立，也可以相互背离。

我们建议把人工工作时间与日历等待时间分别记录。前者反映资源占用，后者反映流程瓶颈。AI 若缩短文档起草，却不影响等待样件或等待客户确认的时间，企业可能只得到局部人效改善。识别这一边界有助于决定下一项改进应投入工具、协作规则还是试验能力，而不把所有延误都解释为模型能力不足。

4.4 多样性保护与商业选择

命题四：保留独立信息来源能够扩大探索空间，但多样性本身不保证商业成功。一个方案与其他方案不同，可能因为它识别了被忽视的客户问题，也可能因为它违背基本约束。企业需要把“与已有方案不同”与“在目标情境下更有价值”分别评价，最后才进行资源分配。

据此，本次分析建议在概念搜索阶段保留人工独立构思、客户观察与模型建议的来源标签；在评价阶段开展适合产品的盲评或约束测试；在决策阶段明确成

SVSAC | sv-sac.org

本、风险和价值的权衡依据。该流程的目标是让企业有机会检验差异的来源和价值，而不是为了追求形式上的多样而平均分配研发预算。

5 需求证据如何进入研发决策

5.1 从反馈文本到观察单位

客户反馈常以不完整的语言进入企业。例如，“不好用”可能指操作步骤多、说明不清、握持不舒适，或者产品与既有流程不兼容。若直接把这些表述输入模型并要求生成新品方向，模型需要自行填补问题背景。结果可能语言具体，证据却并未增加。需求分析的第一步应保留原始表述，并明确哪些信息尚未获得，而不是把补全后的叙述视为客户已经说过的话。

本次分析建议以一次可以界定情境的需求事件作为初始观察单位。事件记录至少说明产品类别、任务、使用环境、遇到的障碍和目前的处理方式。相同客户在不同场景下可以对应多个事件；不同客户对同一事件类型的反馈可以归入同一主题。事件数量、客户数量和文本条数须分别统计，因为一位客户多次投诉不能自动代表多位客户，而同一段长文本也可能包含多个问题。

以一款需要频繁拆洗的容器为分析性情境。用户表达“清洗麻烦”时，研究者应进一步确认困难发生在拆解、刷洗、晾干还是复装环节；困难是否来自结构、污物类型或使用条件；是否已有替代工具。若障碍主要发生在复装，简单更换易清洁材料未必解决问题。AI 可协助提出追问或归类，但应把“可能卡在复装”标记为假说，在实际观察前不得改写为已识别的主因。该情境说明的是识别逻辑，不对应某家企业的调查发现。

5.2 编码规则与人工复核

需求编码的目的不是给所有文本贴上尽可能多的标签，而是使不同人员对同一类问题形成可比较的记录。编码表应给出类别定义、纳入条件、排除条件与边界例子。例如，“操作负担”可以包括完成任务所需动作和认知步骤，但不把运输破损纳入该类；“可靠性”需要区分偶发故障、持续性能下降和错误使用造成的问题。类别之间若存在交叉，可使用多标签，但必须保留事件级的总数，避免在汇总时把一件事情重复算作多件。

一种可执行的试验是从已经取得适当处理依据的反馈中抽取校验集，由两名业务人员分别编码，再讨论分歧并修订定义。模型在规则冻结后处理另一批材料，其输出与人工复核结果比较。研究应同时报告漏掉的问题类型和错误新增的

SVSAC | sv-sac.org

问题类型，而不只报告整体一致率。对于研发而言，把罕见但严重的安全反馈错误归入一般外观意见，可能比漏掉多条重复的颜色偏好更重要。

校验集不能由模型先标注，再以同一模型的自我评分作为准确性依据。这样的评价容易形成闭环，无法判断模型是否共同误解了材料。同样，人工人员的意见也并非天然正确。出现争议时，应回到原始记录或补充观察，而不是因为人工或模型具有某种身份就自动采纳其判断。这里的复核对象是证据与分类规则之间的关系。

5.3 样本覆盖与沉默客户

容易取得的反馈未必代表目标市场。售后记录更容易包含已经购买并愿意投诉的客户，销售人员可能更多接触询价积极的客户，线上评论则受平台展示和发言意愿影响。将这些材料混合后生成“客户需求排行榜”，若没有交代来源构成，会把可见性误当成需求普遍性。

本次分析建议先列出目标使用情境，再核对每类情境是否有材料。对于缺少证据的客户群，不宜让模型生成若干虚构访谈来填补。可将该群体标为证据缺口，决定是否补充访谈、现场观察或渠道核验。若资源不足以补齐，则应限制结论的适用范围。承认证据缺口比提供完整却没有来源的市场画像，更有助于控制错误投资。

一种反例有助于说明覆盖问题。已有客户反复要求增加某项功能，新进入者却可能因为产品基础操作复杂而从未购买。若企业只分析现有用户，新增功能可能进一步提高复杂度。此时，现有反馈内部即使高度一致，也不能证明新增功能有利于扩大客户群。研发需要把“让已有客户更满意”和“使未购买者愿意采用”作为不同的核心问题。

5.4 需求优先级不等于出现频率

频率能够说明问题在已观察材料中的出现次数，却不能单独决定研发顺序。优先级还与损失严重程度、受影响客户范围、可解决程度、交付成本及战略目标有关。若仅以文本频次排序，易被模型识别的高频词可能挤压表达含糊但后果严重的问题。对少量高损失事件，企业需要专门审查，而不能让其在平均评分中消失。

建议把频率与严重性分别列示，再由决策者说明取舍。对于无法比较的事项，不必强行压缩为一个综合分数。例如，更换一种影响手感的材料与修正一种可能导致失效的结构问题，适用的判断路径本就不同。硬性安全要求应先按相应

SVSAC | sv-sac.org

规则处理，偏好改进则可以进入客户选择与成本权衡。该分流是研究建议，具体风险判定必须由具备产品知识的人员承担。

6 从创意数量转向候选组合质量

6.1 候选方案的最小描述

若团队提交的方案一部分是产品图片，一部分是长篇营销文案，一部分只有一个功能名称，评审很难比较它们。我们建议为候选方案建立最小描述：目标情境、拟解决的问题、改变的产品要素、预期价值、关键约束与尚待验证的假设。描述保持简洁，允许未知项存在；目的不是让每份方案看起来同样完整，而是使差异能够接受检验。

候选方案也需要版本边界。改变颜色与改变拆装结构，不能一律记为“新增一个创意”。企业可把方案分为方向与变体两层：方向对应不同的问题解决机制，变体对应同一机制下的参数、形态或材料变化。例如，把清洗步骤减少作为方向，改变表面颜色通常只构成变体；若颜色具有明确的状态识别作用，则可能涉及新的功能机制。分类须依据方案内容，不能仅凭名称不同。

6.2 有效方向数的操作定义

本次分析将“有效方向数”暂定义为：在规定时间窗内，具有不同问题解决机制且达到最低信息要求的候选方向数量。最低信息要求包括能指出目标问题、说明主要改变并列明至少一项可检验假设。它不意味着方案已经可制造或已被客户接受，只排除无法判断内容的空泛标题。

去重可以先由文本或图像相似检索提出疑似重复组，再由评审者依据解决机制确认。相似度阈值不宜脱离具体产品直接套用，因为共同使用某个零部件可能只是必要约束，并不说明产品价值相同。反过来，文字描述完全不同，也可能只是用不同措辞重述同一个功能。研究应保存合并与拆分的理由，抽样检查分类一致性，并在阈值变化时报告有效方向数是否显著改变。

方案组合的多样性还须说明相对于何种参照。相对于本企业旧产品、相对于公开竞品与相对于同一轮候选集合，得到的是不同指标。一项方案可能相对本企业新颖，却在市场上已经常见；也可能与本轮其他方案高度不同，但与目标客户需求无关。因此，组合评价至少应并列呈现内部方向分布、外部替代方案对照与需求关联证据，避免用一个相似度分数代表所有层次。

SVSAC | sv-sac.org

6.3 生成来源与信息隔离

为判断 AI 是否改变探索方向，可以保留三类候选来源：人工独立构思、依据客户资料的人机共同构思、模型直接建议。标签用于分析，不能作为质量等级。人工方案可能重复或不切实际，模型方案也可能提出值得测试的组合。研究关心的是不同来源带来的增量信息及其经过筛选后的表现。

如果试验要识别模型建议的作用，人工独立构思应在参与者接触本轮模型输出前完成并锁定，否则所谓人工方案可能已经受到模型锚定。锁定不意味着后续禁止协作，而是保留一个可比较的起点。参与者若曾在其他渠道接触同类模型结果，也应在可行范围内记录，避免把分组标签理解为完全没有信息交叉的现实。

信息隔离也存在代价。严格隔离可能降低团队正常协作效率，且与实际部署情境不同。因此，研究可区分探索机制的短期受控试验与检验业务效果的现场试验。前者尽量控制输入，后者保留日常协作并记录干扰。两种设计回答不同问题，不能因为其中一种更方便，就声称它同时解决所有因果识别困难。

6.4 选择机制可能消除生成阶段的差异

即使候选组合保持多样，后续评审也可能把它们压缩到熟悉的方向。评分表若高度偏重既有供应商能立即实现的功能，具有新价值但需要小规模验证的方案可能提前退出。反之，过度奖励新颖性又可能选出工程风险过高的方向。生成与选择因此需要分开研究：最终方案趋同可能来自模型输出，也可能来自企业的评价标准。

我们建议记录评审前后的组合变化，比较每种方向的保留比例与淘汰原因。若新方向主要因缺少证据被拒绝，下一轮应判断是该方向价值不足，还是团队尚未设计合适试验；若主要因为触犯明确约束，则无需为了多样性继续投入。评审记录应容许“暂缓并补证”这一结果，而不只保留通过与否决，以免把未知一律转化为失败。

7 研发迭代的评价设计

7.1 主要结果变量的选择

评价 AI 辅助研发，首先需要确定要改善的业务问题。若问题是需求澄清慢，主要结果可设为形成经过业务确认的需求说明所需时间；若问题是试制返工多，则应观察因设计问题导致的返工，而不是只测文档产出速度。不同结果变量对应不同的干预位置，不能在试验结束后选择最有利的一项作为唯一成功标准。

SVSAC | sv-sac.org

本次分析建议在试验前确定一个主要结果和少量约束指标。例如，以“通过规定验收条件的候选方向数”为主要结果，约束指标包括工程复核工时、严重错误数量和额外试制费用。这里的通过条件需预先写明，并由实际产品知识确定。若主要结果改善但严重错误增加到不可接受程度，项目不应仅凭平均分提高而宣布成功。

7.2 观察单位与分组干扰

随机分配的单位可以是任务、研发小组或时间段，选择取决于信息传播方式。如果同一团队在处理不同任务时共享 AI 生成的检查表，按任务分组就容易发生干扰。此时，对照任务可能间接受益，估计结果不再是完全没有 AI 与使用 AI 的差异。可以考虑按团队分配，但团队数量少时，估计的不确定性又会增大。

因此，研究方案应同时说明选择某种单位的理由和该选择的局限。对任务难度进行事前分层，可以减少复杂度失衡，却不能保证所有未观察差异被消除。对同一任务让同一人先不用 AI 再用 AI，会有学习和顺序效应；若反过来，AI 建议又可能影响后续独立完成。简单的交叉设计不应被描述为天然无偏，需要根据任务可重复性安排顺序和平行任务。

7.3 同时测量增益与代价

设每个任务的总人工耗时为需求整理、方案形成、工程复核和返工四部分之和。若 AI 使方案形成少用两小时，却增加三小时的工程复核与返工，则总人工耗时增加一小时。这一算例仅是算术假设，不代表某企业观察值。它说明局部节省与全过程收益在逻辑上并不等价。

同样，总人工耗时降低不保证日历周期缩短。一个团队可能节省半天资料整理时间，但仍需等待下一次设备试验窗口。研究应分别记录主动工作、排队、外部确认和不可压缩的工艺等待。若等待是主要瓶颈，继续优化生成速度的边际收益可能有限；若工作量过大导致评审排队，减少文档负担又可能间接缓解等待。具体路径需要流程记录支持。

7.4 停止、扩展与否定性结果

试验开始前应明确停止条件，例如出现不能及时控制的关键数据泄露风险、模型反复遗漏必须检查的约束，或复核成本明显超出可承受范围。条件必须能够被实际观察，不能只写“风险过高时停止”。停止后需要保留事件与处置记录，以区分技术不适用、数据基础不足和流程设计不当。

SVSAC | sv-sac.org

扩展条件也不应仅是参与人员觉得方便。可以要求主要结果达到事前设定的业务改善幅度，同时关键质量指标没有明显恶化，并且总成本在预算范围内。阈值取决于企业损失结构，本篇不提供对所有企业通用的百分点。若样本不足以判断结果，则应报告不确定，而不是将统计上未显著等同于没有影响，或将小样本正向均值当作可靠成功。

否定性结果仍有研究价值。若工具对成熟工程师帮助有限，却改善了新成员需求整理，这可能支持缩小部署范围；若生成更快但验证更慢，则应重新设计评审流程；若多个对照方式都未发现可用收益，则需要承认当前场景不适合，而不无限追加提示词与模型调用。研究框架应允许项目被停止，否则所谓试验只是为既定采购决定寻找理由。

8 工程验证：把概念差异转化为可交付证据

8.1 设计变量与性能目标不能混写

产品概念进入工程阶段，首先需要把可改变的设计变量与希望实现的性能分开。壁厚、连接方式和材料牌号可以是设计变量，使用寿命、拆装时间和单位成本则是评价结果。若提示词仅要求“更轻、更便宜、更耐用”，模型可以在语言层面同时满足全部目标，但真实设计需要处理目标之间的约束关系。研究者应追问每项建议改变了什么变量、预期通过什么机制影响结果，以及有哪些条件可能使作用方向逆转。

以可拆洗容器继续作分析性情境。增加卡扣可能减少复装步骤，也可能提高局部应力或增加污物滞留位置；减薄结构可能降低重量，却使装配容差更敏感。这里不预断任何具体结构必然存在上述问题，而是展示需要检验的作用链。工程评价的任务不是寻找一段能解释方案优点的文字，而是设计足以暴露其失效条件的检查。

设计空间也需要明确边界。若试验只覆盖特定材料、温度范围与使用次数，结果不能自动推广到其他材料、极端温度或长期使用。模型生成的新组合尤其容易跨出已有验证范围。企业应把设计变量的可用区间、未经验证区间和明确禁用区间分别记录，避免系统将数据缺失理解为设计自由。

8.2 从语言合理性到可观测的失败模式

AI 能够提出听起来合理的技术解释，但解释并不是测量。工程师需要把解释转换为可观测量：例如装配所需力、可见磨损、泄漏情况、步骤完成时间或关键

SVSAC | sv-sac.org

尺寸变化。每项量应说明测量方法、条件和判定标准。若团队对“易用”或“稳定”的含义没有形成一致定义，模型写得再完整也无法使测试结果可比较。

失败模式的描述应尽可能具体。与其写“存在质量风险”，不如说明哪个部位在什么条件下可能发生何种失效、怎样发现以及产生什么后果。这种描述能够指导测试资源分配，也有助于区分结构缺陷与使用误解。若某个问题只能通过不现实的试验才能观察，应说明可识别性的限制，而不能把没有发现问题等同于没有问题。

生成系统可扩展检查项，但不能独自批准清单是否充分。Vaccaro 等区分两种比较：人机组合相对人单独工作的增强，以及相对人、AI 中较强一方的协同。其预注册综述含 106 项实验、370 个效应量，组合平均未超过较强一方，任务类型之间有差异。[10]这些是跨任务汇总，不是工厂复核员效果。企业应同时记录无人复核的模型输出、可比人工流程及人机流程，并计入核查成本；只显示人机优于人工，仍不能声称优于所有可行替代方案。

8.3 仿真、代理模型与实物试验的分工

不同验证手段提供的证据不可简单互换。仿真结果取决于模型结构、边界条件和输入参数；代理模型通过既有样本近似某种输入输出关系；实物试验受到样件、设备和测量误差影响。三者可以相互补充，但若它们共同依赖错误参数，也可能一致地给出误导性结果。多种工具意见一致并不自动构成独立证据。

Frazier 的贝叶斯优化教程讨论了利用代理模型及其不确定性选择下一次昂贵评价的机制。[11]这种思路提示，在试验成本高且变量范围可界定的条件下，可以研究如何更有效地分配测试预算。然而，它不意味着任何研发任务都适合该方法。若测量目标经常变动、样本来自不兼容工艺，或无法可靠记录设计变量，先引入优化算法可能只是在优化记录错误。

在本次分析框架下，顺序试验应保留每轮选择下一个试验点的依据，区分寻找当前较优方案与减少未知区域的不确定性。最终拟交付方案仍需接受适合其用途的确认。代理模型预测的优值不能直接写成已经达到的产品参数，也不能用优化器推荐的位置替代对安全边界的判断。

8.4 变更控制与交付责任

研发迭代越快，越需要辨别哪一个版本正在被测试、报价和生产。若结构已经修改而试验记录仍对应旧版，企业可能以旧证据支持新产品。AI 自动生成的说明、物料清单和测试摘要若分别引用不同版本，会使这种错配更难被发现。本篇因此将版本一致性作为交付证据的组成部分，而非文档整理的附属工作。

SVSAC | sv-sac.org

建议对影响关键性能的变更建立最小追踪：改变内容、原因、影响范围、所需复测与批准记录。并非每次文字调整都需要重新做全部测试，但不进行复测的理由也应可解释。决策者应明确哪些变化只涉及表述，哪些改变工程对象。对于尚未完成验证的版本，应避免在对外资料中使用暗示已验证的措辞。

9 经济评价：效率、现金与机会成本

9.1 成本边界决定决策建议

AI 项目的价格容易被看见的是软件订阅或模型调用支出，而项目经济性取决于更多环节。本次分析建议把成本分为初始接入、持续运行、业务复核、工程验证和退出迁移。初始接入包括资料清理与流程配置；持续运行包括许可、计算和维护；业务复核包括对需求与生成内容的检查；工程验证包括样件和测试；退出迁移包括导出资料、重新配置和替代工具衔接。这里给出的是核算边界，不是市场报价。

成本是否新增，需要以没有本项目时会发生什么作为参照。企业原本就必须做的试验不应全部计作 AI 新增成本，因新方案数量增加而额外实施的试验则需要记录。同样，既有员工参与试点未必产生当期额外工资支出，但会占用原可用于其他工作的时间。现金支出与资源占用都重要，却回答不同问题。

一个容易忽略的反例是，AI 减少了需求整理工时，但团队把节省时间用于改进测试，因此工资和现金费用均未下降。企业可能获得更多学习或更高可靠性，却不能将节省工时直接记作已经实现的现金利润。相反，若减少工时确实避免了加班、外包或新增招聘，才可能形成对应的现金节约。研究应说明具体实现路径。

9.2 一个透明的假设算例

以下算例仅用于检验核算逻辑，全部数量与单价均为假设，不代表任何企业、行业均值或工具报价。设一轮研发有 20 项可比任务。原流程每项需求整理 2 小时、概念形成 3 小时、工程复核 2 小时、返工 1 小时，共 8 小时，合计 160 小时。AI 辅助流程每项分别为 1 小时、1.5 小时、2.5 小时和 1.5 小时，共 6.5 小时，合计 130 小时。净减少 30 小时，而非只计算前两项所显示的 50 小时节省。

再假设所有工时使用每小时 100 元的统一机会成本口径，则 30 小时对应 3000 元资源价值。这只是便于说明的统一影子价格，实际核算应根据岗位和受约束资源分别估计。若本轮工具与运行新增支出为 1200 元，额外测试的材料与设备

SVSAC | sv-sac.org

现金支出为 2500 元（不包含前述已计人工，避免重复核算），在暂不考虑其他收益和初始接入成本的条件下，净资源收益为 3000- 1200- 2500 = - 700 元。

该负值不能证明 AI 项目必然不可取，因为算例没有给出测试带来的质量收益或后续复用价值；同样，不能凭空给这些收益加上一个足以使结果转正的数字。正确做法是标明未计量项，提出需要获得的证据。如果假设没有现金工资节约，且 1200 元与 2500 元均为当期新增现金支出，则本轮现金变化为支出增加 3700 元。资源收益与现金变化不同，并不构成矛盾。

在相同假设下，若所有新增支出保持 3700 元，工时影子价格为 100 元，则仅依靠人工时间价值抵补新增支出需要净减少 37 小时。这是算术上的资源收支平衡条件，不是实际市场回报门槛。接入成本、不同岗位单价、折现、税费和质量风险一旦加入，结论会相应改变。

9.3 质量收益与损失不能重复计入

若 AI 使错误减少，可以研究避免返工、退货或停机的价值，但每项收益必须对应可辨别的损失路径。已经在工时项目中扣除的返工减少，不应再以“效率提升”重复计入；同一批退货既减少直接损失又影响未来客户选择时，也需要区分可观察的当期成本与尚未确认的长期效果。

以新品延后上市为例，不能简单把缩短天数乘以预计日销售额当作利润增量。销售额不是利润，需求也可能只是从后续时间提前发生。若竞争窗口、渠道上架或客户项目节点存在明确时限，提前交付才可能改变可获取的需求。研究需要检验是否新增销量、是否挤占旧品、单位贡献如何变化，以及是否伴随质量风险。

对于低频高损失风险，单次试点没有事故并不意味着预期损失为零。企业可以进行情景分析，列出不同发生概率与损失规模下的结果，但概率不能凭模型口头给出后就当作真实估计。若缺少可靠数据，应呈现范围、依据和决策敏感性，并说明哪些结果由假设主导。

9.4 有限预算下的阶段决策

项目早期与规模部署的经济问题不同。小试阶段可以接受暂时较高的单任务成本，因为其目的之一是获得是否值得扩展的信息；规模阶段则需要检验持续运行是否可承受。把试点成本简单线性外推到全公司，可能高估共享基础设施成本；把试点的少数热情参与者表现外推到所有岗位，又可能高估收益。

建议按证据成熟度设置阶段预算。第一阶段验证数据能否使用及关键任务是否可被可靠辅助；第二阶段比较流程结果和全过程成本；第三阶段再研究跨团队

SVSAC | sv-sac.org

扩展。每个阶段应有明确的决策问题和结束条件。阶段推进不是默认购买更多服务，而是根据新增证据决定继续、收缩或退出。本篇不预设哪一种部署形式最便宜，因为现有资料不足以比较具体供应商的总成本。

10 数据与知识治理：研发记录的适用边界

10.1 数据对象与使用目的

研发资料中可能同时存在客户联系人信息、产品规格、供应商报价和设计知识。它们不适用完全相同的处理规则，也不应只用“内部数据”一个标签管理。对每类资料，企业需要先识别拟用目的、可访问人员、接收系统和保留安排，再判断是否适合用于当前模型任务。模型能够接收某种文件，不构成企业有权提供该文件的依据。

我们建议把技术上可访问、业务上相关与具有适当处理依据分别作为检查项。例如，客户联系方式可能存在于售后记录中，但识别清洗障碍通常不需要把号码传给生成系统。删去无关字段既减少不必要的数据暴露，也降低模型受无关信息干扰的机会。不过，字段删除是否足以构成匿名化，需要根据实际可识别性判断，不能仅因姓名被遮盖就宣称资料已匿名。

10.2 保留期限与知识复用

《个人信息保护法》第十九条规定，除法律、行政法规另有规定外，个人信息保存期限应为实现处理目的所必要的最短时间。[7]因此，研发可追踪性不能被解释为无限保留个人信息。企业需要区分对验证有必要的技术结果、用于联系客户的信息以及可能识别个人的原始叙述，并分别评估保存需要。

一种可研究的组织安排是，在合法适当的前提下，以技术问题及测试结果形成可复用知识，同时对原始个人信息设置独立访问控制和到期处理。这个安排是治理建议，不意味着自动满足全部法律义务。若资料涉及跨境、敏感信息或特定行业限制，还需结合实际数据流与适用规则核验；本次分析不凭工具部署在本地或云端这一单一属性作出合法性结论。

10.3 知识库更新与错误传播

知识库中的内容也有生命周期。已经废止的规格、因特定客户而设的例外条款、失败方案中的假定，若没有适用标签，可能在下一次检索中被当作一般规则。系统输出越流畅，使用者越可能忽略资料的时效和适用条件。错误传播由此不只是模型“编造”的问题，也可能是检索忠实地复述了不适用的记录。

SVSAC | sv-sac.org

建议为关键资料设置状态、版本、生效范围与核验来源，并使检索能够排除或明确标注已废止内容。对错误输出，应追查它来自检索遗漏、资料过期、适用性判断失败还是生成过程偏离。不同原因对应不同修正。若一律通过改提示词处理，企业可能掩盖需要清理数据或修改权限的根本问题。

10.4 治理强度与研发负担的权衡

治理控制也会占用资源。要求每次普通文本改写都进行完整审批，可能使低风险任务失去试验价值；对关键技术承诺只进行随机抽查，又可能不足以控制损失。控制强度应围绕任务后果与证据需要设计，而不是对所有模型输出采用相同流程。NIST 框架的四项功能可作为组织讨论的参照，但具体控制仍需结合任务落实。[5]

一种分层方式是区分内部草拟、进入工程评审的材料与形成对外承诺的内容。分层并不改变事实应有的准确性要求，而是明确不同用途需要的确认强度和责任人。研究应记录控制措施增加的工时及发现的问题，以便判断它是否确有价值。只有把治理成本与错误损失同时纳入，企业才能避免以“更安全”或“更高效”这样的抽象词替代实际权衡。

经济算例可复算，但不构成企业成效证据；治理建议须接受实际数据流与产品情境检验。

11 两类案例及其对研发机制的不同启示

11.1 GM 与 Autodesk：约束搜索的概念验证

Autodesk 在 2018 年 5 月发布的合作材料，将 GM 座椅支架项目明确称为初始概念验证。材料称，工程师设定连接点、强度和质量等参数，软件产生超过 150 个候选，最终展示方案相对原部件减重 40%、强度提高 20%，并将 8 个组件合并为 1 个 3D 打印件。[12]这些是厂商报告的项目指标，本篇没有取得独立测试数据，也不据此认定规模量产或成本收益。

该案例首先提示，生成设计的输入并不是一句宽泛的创新要求。连接关系、载荷与制造条件限定了可搜索空间。因而，企业能否表达约束，可能与模型生成能力同样重要。本篇由此提出的推断是：如果设计知识未被转化为可检查的条件，扩大候选数量可能只是扩大工程师后续排除无效方案的工作量。这个推断需要通过不同约束表达方式的对照试验验证，不能由厂商个案单独确证。

其次，部件合并改变的不只是形状，还可能改变装配、采购和维修的任务分布。减少组件有可能减少接口和装配动作，也可能使局部损坏后的更换范围扩

SVSAC | sv-sac.org

大。公开材料所报告的质量与强度改善，不能回答这些全寿命周期问题。研究中应分别列出已报告的设计指标与仍需证据的交付后果，避免把“部件更少”直接翻译为“总成本更低”。

最后，本案例的生成设计不能无条件等同于当下的聊天式大语言模型应用。两者可能共同扩大方案搜索，但其输入结构、可行性约束与输出验证方式不同。企业采用产品图片生成或文本建议时，不能直接借用工程生成设计的性能指标作为预期效果。比较时应以任务机制为单位，而不是把所有带有 AI 名称的工具视为同一种干预。

11.2 A-Lab：闭环试验及其证据更正

A-Lab 论文及其更正列为自动化实验研究的待核文献线索。[13,14] 本篇未取得两者正文，出版社入口要求身份验证，大学存储库也不可读。因此，不以检索片段中的计数、材料名称、训练资料或更正内容作为已核实的案例证据，也不据此计算成功率或判断材料新颖性。只有核对相应原段及其版本关系，才能将具体结果纳入比较。

在研发机制上，应区分生成候选与验证候选：前者提出可试方向，后者需要与任务相适应的物理试验或评价，再决定是否进入下一轮。这是我们提出的分析区分，不是对尚未读到正文的 A-Lab 平台运行情况的核实。若制造企业拟建立连续试验与表征反馈，设备、材料处理、测量可靠性与领域知识都是需要另行评估的条件；增加软件账号本身不能证明这些条件已经具备。

评价标准还应区分“对象已实现”与“对象对科学领域新颖”，以及“系统给出成功标签”与“证据足以排除替代解释”。这些是本篇对成果判定的要求，不作为未读更正具体内容的转述。研发指标应将对象、参照范围和证据标准写进定义；缺少其中任何一项，都不宜仅凭成功数量评价研究或部署价值。

11.3 案例之间可以比较什么

上述材料不能组成可汇总的效果样本。GM 材料是厂商合作发布的概念验证，A-Lab 在本次分析中仍只是未完成原文核查的文献线索，不能与前者形成已经核实的案例对照。本篇仅以候选产生、约束表达、验证标准和失败反馈作为后续收集可比证据的维度；不合并百分比或成功数量，也不推算 AI 平均收益。

这类机制比较能够帮助企业提出更具体的问题。例如，若团队缺少可执行的约束，优先工作可能是整理工程条件；若已有仿真却缺少实物确认，应检查模型误差；若试验持续进行却没有回写失败原因，则需要改善记录和决策联系。具体优先级仍需流程证据，不能从两个案例推出所有制造企业的统一顺序。

SVSAC | sv-sac.org

12 评价指标、估计目标与不确定性

12.1 指标必须明确分子、分母和预算条件

第 7 节以通过验收的候选方向数作为主要结果的例子。本节进一步限定其含义：分子为经过预先约定的审核与测试后符合条件的方向数，方向去重按第 6 节规则执行；若要比较单位任务产出，分母为事前分配的全部任务数，包括未完成任务；若要比较资源效率，则另外报告每百工时产生的合格方向数。两种比率不能混用。

同一指标需要在可比预算下解释。若 AI 组获得更多工程复核、额外测试设备或更多专家时间，它产生更多合格方向可能反映资源增加。研究可以固定总预算来比较产出，也可以固定交付目标来比较成本；若两者都不固定，就需报告投入变化并限制因果解释。不能将没有记录的额外人工视为免费。

合格方向数也不等于最终商业价值。一项高价值方向可能比多项轻微改进更重要，因此应在主要指标之外记录客户验证结果和后续阶段表现。但这并不意味着可以在看到结果后随意改变主要指标。企业可为下一阶段修订问题，并把修订原因记录下来，使探索性分析与事前检验保持区别。

12.2 分配效果与实际使用效果

如果研发团队被随机分配使用某工具，最直接可识别的对象通常是“提供工具及配套安排”的效果，而不是每一次模型调用的纯技术效果。有人可能不使用分配的工具，也有人可能在对照组通过其他渠道接触 AI。应记录这些行为，但不能简单删去不遵从者后把剩余比较称为随机试验结果。

本次分析建议以原始分配组比较作为主要分析，明确它包含培训、使用障碍与流程适应在内的实施效果。若研究实际使用强度与表现的关联，应标为次要分析，并讨论主动使用者可能在动机、技能和任务选择上不同。未取得足以识别使用因果效应的设计时，不应将关联系数解释为多调用一次模型必然带来的收益。

12.3 小样本与多层数据

任务、候选与团队形成多层数据结构。一个团队生成 100 个方案，并不等于获得 100 个相互独立的企业观察。若这些方案共享同一需求、人员和评审者，其误差可能相关。把候选数量直接作为样本量，容易夸大估计精度。分析单位应与分配单位及误差结构相匹配，并报告实际独立团队或任务的数量。

在团队数量较少时，任何统计修正都不能凭空制造信息。研究可提供各团队结果、差值范围与对关键假设的敏感性；若设计允许，可以使用遵循原分配规则

SVSAC | sv-sac.org

的随机化推断。采用某种区间估计或重抽样方法前，需判断其条件是否满足，不能将软件默认输出视为稳健性证明。样本量规划应依据主要结果的基线波动、最小业务重要差异和分组结构，当前缺少这些资料，不宜给出一个万能任务数。

12.4 未完成任务与记录缺失

未完成任务和结果记录缺失需要区分。如果在规定期限内未形成可验收方案，按预定规则将合格方向记为零，可能是合理的业务结果；如果方案已经完成但测试记录丢失，则结果状态未知，不能同样记零或直接删除。两种处理都会改变比较方向，必须事前说明并在报告中列出数量。

缺失边界须与结果单位一致。本篇主要结果是每项固定任务的合格方向数，可能大于 1；不能把未知任务简单赋值 1 就称宽松上界。若截止前提交方向清单完整，仅验收标签缺失，下界令这些方向均未通过，上界令它们均通过。若整项清单缺失，则只有在事前确定共同方向上限或有可验证有限上限时才能给数值上界；没有这种依据就应报告上界无法识别，并按透明假设作敏感性分析。不能用任意上限维持正面结论。

12.5 可证伪条件与解释竞争

对第 4 节命题，证伪不应被理解为只有观察到负面效果才算。以筛选瓶颈命题为例，如果生成数量增加、验证投入不变，而合格方向和客户验证结果仍持续改善，同时没有不可见工时转移，那么应修正“筛选能力限制收益”的具体解释。反之，若多样性下降但客户价值上升，则需检验收敛是否反映更准确的需求聚焦，而非直接判定趋同有害。

同样，工具效果可能来自一般流程规范化，而不是 AI 生成能力。若引入 AI 的同时统一了需求模板、明确了责任人并增加评审频率，观察到的改善可能由这些改变驱动。可设置只改流程、不提供 AI 的比较组，在资源允许时区分二者；若无法实施，就应把干预描述为组合方案。准确界定干预，比将所有改善归于一个技术标签更能指导后续投资。

13 三类产业情境中的差异化机制

13.1 比较逻辑：技术类别不是分析单位

“制造业”不足以预测 AI 研发工具的作用。至少需要把产品的需求稳定性、规格重复性、变型频率、失效后果和验证成本纳入比较。以下三类情境是用来推导可检验机制的分析性情境，分别代表标准化零部件、面向消费者的产品以及按项

SVSAC | sv-sac.org

目交付的定制设备；它们不对应具名企业、现场调查或实测绩效。三者都可能使用生成模型、检索系统或优化算法，但 AI 所处的决策节点和错误代价不同，故不应把某一类场景的效率结果直接移植到另一类。

Verganti、Vendraminelli 与 Iansiti 提出，算法承担更多创造性问题求解后，人的设计活动更需要聚焦于识别哪些问题值得或能够解决；其框架由 Netflix、Airbnb 等前沿企业案例阐释，属于理论分析而非制造企业因果效果估计。[15] 这一视角提示，比较时可沿一条因果链追问：客户或工程信息如何进入系统，系统输出改变了什么决策，变更如何进入制造和服务，最后由什么证据判定价值。若中间环节没有可观察记录，“AI 帮助差异化”的说法就难以区别于同时发生的流程改造、市场变化或管理关注度上升。由此，三个情境比较的核心不是排列行业先进程度，而是辨别哪种机制更可能成为约束，以及何种证据可以区分竞争解释。

13.2 标准化零部件：接口稳定，价值常藏在边界条件

设想一家供应多种工业设备通用紧固件、阀件或传动部件的制造商。其零件具有明确接口和长期质量要求，设计更改需要面对材料批次、工艺窗口、认证文件与客户既有设备兼容性。差异化未必来自肉眼可见的新造型；减少装配停顿、降低磨损、缩短故障定位时间，可能更接近客户的选择理由。本篇没有把这些价值视为该假设厂商已发现的需求，而是提出可现场验证的问题。

在此情境中，AI 较可能先作用于已有工程资料的检索、失效记录的归类、规格变更的影响筛查，以及在限定变量内提出候选设计。作用路径可以表述为：资料被整理并标注型号与版本，检索减少工程师寻找既有试验的时间；候选方案围绕明确接口和工况生成；工程人员核对约束后，试验比较磨损、装配、稳定性或成本指标；结果再反馈到同类规格。若资料缺少失效条件，模型可能把不同设备上的“成功配置”错误合并；检索看似高效，实际却将适用边界抹平。因此，关键中介不是答案生成速度，而是资料的适用性标记是否足以阻断错误迁移。

该情境的反证也很重要。若零件规格变化极少，标准和既有设计已覆盖绝大多数需求，那么 AI 带来的新增搜索空间可能很小，改进图纸版本治理或测试排程也许更有收益。若问题来自供应商原料波动而非设计知识不足，语言模型难以替代材料检测和供应商过程控制。企业应比较“AI 辅助检索/设计”与“只改进文档索引、工程变更流程”的结果，分别记录检索时间、约束遗漏、复核工时、测试通过率和客户现场失效，而不能只数生成的候选件。

SVSAC | sv-sac.org

13.3 消费品：感知价值、使用情境与组合趋同

设想一类需要频繁使用、购买决策受外观与体验影响的家用产品。客户群体可能按清洁习惯、空间限制、操作能力、价格敏感度或使用频率呈现不同需求。产品评价既涉及可测的耐用性和能耗，也涉及用户是否能感知差异、是否愿意为之支付以及是否愿意改变现有习惯。对这类产品而言，单纯把线上评论摘要成“热门功能”有选择偏差：留下评价的人不等于所有购买者，表达次数也不能当作总体偏好比例。前文提出的需求事件与沉默客户处理原则适用于此处，模型生成的细分画像只能作为待检验假说。

一种可能机制是 AI 降低概念表达、文案探索和视觉变体生成的边际成本，使团队更快比较不同价值主张；另一种可能机制是团队共享相似的流行语料与审美默认值，使不同品牌在短期内都集中于易被模型生成、也易被内部评审接受的选项。前一种机制扩大探索，后一种机制产生组合趋同，两者可以并存。判断哪一条占主导，应保留输入来源、候选方向和评审前后组合，邀请目标用户对具体使用任务作盲评或情境选择，并观察选择是否转化为可重复的购买或使用行为。点赞、概念偏好和实际采购是不同结果变量，不宜互相替代。

消费品还面临上市节奏与售后反馈的共同作用。若系统把退货理由快速聚类，产品团队可能提前发现某个尺寸或操作步骤的问题；然而，若类别规则不稳定，少数但严重的安全或兼容问题可能被高频的颜色偏好淹没。一个可证伪判断是：在反馈覆盖和评审资源相当时，辅助分类是否提高了问题定位的可复核性，而不是只增加摘要速度。替代解释包括平台算法改变、促销人群变化、供应质量提升和新营销方案。若无法记录这些变化，销量提高不能被单独归于 AI。

13.4 定制设备：配置复用与项目独特性的张力

设想一家依据客户现场条件设计非标生产设备的企业。项目共享若干子系统，但空间布局、工艺节拍、接口、控制策略和验收条件可能各异。这里的差异化不是给标准产品增加装饰选项，而是将客户约束翻译为可交付的系统配置。系统工程资料、报价、设计变更、调试记录和验收条款构成跨部门知识链。AI 可能协助搜索旧项目中的模块、提取需求差异、生成检查清单或提示接口冲突；价值取决于它能否区分可复用规律与只对单一现场成立的例外。

可检验的作用链是：先把客户要求拆成可追踪的需求项与验收依据，再从有权限的历史项目检索相似配置，工程师比较接口和适用条件，最后由具体负责人确认本项目设计并经过验证。若模型直接复制旧项目中的布局，未标记的现场差异可能传入新设计；若为避免风险而每个项目都从零开始，企业又无法获得模块

SVSAC | sv-sac.org

复用带来的学习收益。企业可记录检索命中后被采纳、被修改、被拒绝的比例及原因，追踪变更单、调试返工和验收问题，并访谈客户工程接口人。需要注意，项目规模、客户成熟度和合同边界往往同时影响这些指标，故简单比较 AI 项目与非 AI 项目容易有选择偏差。

定制设备的独特之处还在于交付现场会产生新的信息。调试问题可能暴露的是设计不完整，也可能来自客户工艺改变、安装误差或操作培训不足。若将全部现场问题编码成模型的“设计缺陷”，系统会让研发组织过度修正产品；若将它们全部归为现场例外，则同类问题可能重复发生。记录中应将事件起因、责任边界和证据状态分开，并允许“原因未定”。当模型参与需求整理或交付文档时，还需检查客户合同中的保密、使用和披露约定。

13.5 跨情境判断与外推边界

三类情境揭示了一个共同问题：AI 的性能并不是脱离组织条件的属性。零部件场景中，型号与规格治理影响检索准确；消费品场景中，样本覆盖和审美选择影响客户价值；定制设备场景中，项目差异与模块复用边界影响工程责任。相同模型在不同资料结构下可能产生相反结果。生产周期更短或方案更多，不足以证明差异化增强；还需确认被改变的决策、交付质量与目标客户结果。

适当的研究设计可以先在每种情境内选取一类具体任务，绘制现有流程和决策权，再确定主要结果、硬性约束及资源投入。跨情境比较应先比较作用机制和实施条件，不宜直接合并估计平均效果。若企业希望将一种情境的经验推广到另一产品线，至少要检查关键约束、数据粒度、客户决策者和试验可得性是否相似，并对可能改变效应方向的差异作出记录。外推是有条件的判断，不是试点成功后的默认步骤。

14 组织实施：把工具放进有责任边界的工作流

14.1 责任配置需要覆盖决策链

AI 项目通常跨越研发、工艺、质量、采购、售后、信息技术、法务或知识产权管理。若只有一个“AI 负责人”，但没有人对需求来源、工程放行和对外承诺负责，模型输出就可能在交接处变成无人承担的事实。我们建议为每类试点明确四类责任：业务问题负责人决定要解决的结果与停止条件；资料责任人确认资料来源、版本和访问范围；技术/工程责任人定义约束并批准工程证据；试验负责人记录分配、指标、偏离和复盘。法务、信息安全和知识产权人员按触发条件参与，而非被抽象地要求审批每一次交互。

SVSAC | sv-sac.org

责任表还应区分建议生成、证据核查和最终决定。模型可以被允许草拟需求分组，却不应因此成为需求的权威来源；工程师可以接受模型给出的检查项，但批准产品变更仍需按企业既有质量体系确定的人承担；采购可依据核验后的性能资料谈判，却不应把未经验证的模型输出写入对客户的保证。此种配置不是增加形式层级，而是将“谁有权决定”和“依据何种证据”在流程中显式化。

14.2 从任务入口到决策留痕

试点入口应定义允许和禁止的任务类型、可用数据类别、输出用途以及遇到未知时的升级路径。员工若不知道哪些材料不能输入，单纯开展通用培训不足以控制实践行为。可为低风险的文字归纳、会议准备等任务提供简明操作规则；对可能改变安全、性能、成本承诺或客户合同的输出，则设置独立核对点。需要核对的对象是输出所依据的原始资料、所适用的版本和未解决假设，而不是只确认“有人看过”。

记录粒度需要服务于复现和责任追踪。建议保存试点目的、模型或服务版本、输入资料的来源标识、生成内容被如何修改、评审人与决定、下游测试和结果。个人信息和商业秘密仍需按目的、权限与期限管理；为可审计而保留记录不等于可以无限收集完整提示词或员工活动。可研究以项目编号、去标识摘要及访问审计记录替代冗余的个人数据。记录设计本身也应接受成本评估：如果每项低风险任务都需耗费大量时间填写表格，员工可能转向不可见渠道，降低组织对实际使用的了解。

14.3 跨职能协作的具体接口

有效协作不是把所有职能拉入一次会议，而是设置能够传递决策信息的接口。客户或销售角色提供问题发生的市场情境及承诺边界；设计与研发角色提出不同解决机制；工艺与供应链角色识别制造窗口、材料和供给风险；质量角色定义可接受的证据与追溯要求；财务角色区分现金成本、资源占用与可验证收益；信息技术和数据团队管理账号、连接器、版本和访问权限。角色可以由少数人员兼任，但每项责任必须有人认领，尤其是小企业中的兼职岗位。

会议应围绕具体决策对象组织：哪些需求证据足以进入概念阶段，哪些方案进入样件测试，哪些变更必须复核，什么结果触发暂停或退出。会前由不同职能分别标注不确定项，会议中记录分歧的证据基础和被采纳的权衡，之后将决定与版本关联。若讨论只留下“大家一致同意 AI 方案”，就不能判断一致是因为数据充分、权责清楚，还是因为资深人员的意见没有被提出。

SVSAC | sv-sac.org

14.4 能力建设与激励扭曲

能力建设不等同于让每个人学会更复杂的提示写法。研发人员需要辨别模型回答与外部证据的差别，理解版本和测试记录，知道何时应要求补证或停止；管理者需要识别局部效率与全流程效应的区别；数据人员需要理解产品语义和适用范围；评审者需要有权拒绝证据不足的承诺。培训可使用本企业脱敏后的典型失败模式，但若案例材料不允许复用，就应使用明确标注的模拟样例，而不能未经许可上传真实客户内容。

激励结构也会影响记录质量。若团队只按方案数量、调用量或上线速度考核，成员可能增加低成本候选、降低复核投入，或把错误隐藏在“快速试验”名义下。若只惩罚失败，则团队倾向提交熟悉方案而不探索。可将激励与被验证的学习、风险及时暴露、跨职能复用及交付结果相联系，同时保留失败结果的描述空间。这是关于激励可能改变记录和探索行为的机制推断，不是本篇获取的企业观察或普遍经验规律；可检验方式包括比较不同考核安排下的低质量候选比例、复核投入、风险迟报率和跨团队复用，并控制任务难度与人员构成。这样的指标不自动消除操纵，但可降低单一容易计数的目标替代实际价值的风险。

14.5 小型试点、扩展与退出

组织实施宜从一个边界清楚、反馈周期可控的任务开始，例如历史故障资料的检索辅助或一个低风险部件的概念筛选。启动前明确基线流程、资料权限、预期使用者、硬性约束、工时口径和停止规则；过程中保存偏离与不确定结果；结束时由未直接负责工具采购的人复核结果解释。这个安排的目的在于让失败和收益都能被发现，而不是预先保证试点必然扩大。

扩展应要求证据跨过新的边界：不同团队是否能在不依赖原倡导者的情况下复用流程；数据版本和权限是否稳定；新增审核及维护成本是否可承担；目标客户或工程指标是否仍有效。退出也要被设计：如何导出研发记录、删除不再需要的输入、恢复原流程、处理供应商变更，以及防止关键知识只留在个人账号中。若这些条件没有讨论，试点可能形成事实依赖，却没有形成可控制的组织能力。

三个情境均为分析性构造；责任框架和分阶段动作属于可检验建议，不宣称已由企业部署验证。

SVSAC | sv-sac.org

15 知识产权、资料来源与技术部署边界

15.1 将权利问题拆分为输入、过程与产出

研发团队谈到“AI 知识产权风险”时，常把不同问题压缩成一个结论。实际上至少有三类对象：输入材料是否有权访问并用于当前目的；模型服务如何处理输入、日志和输出；最终方案、数据和说明文件由谁控制、如何验证和披露。三类问题的法律基础、合同条款和业务影响不同。某供应商声称“企业级安全”，不能据此推定其训练、留存、跨境处理或输出权利安排符合企业所需；本段不基于宣传语判断某产品或具体数据流合法。

可操作的资料清单应逐项记录来源、权利或合同依据、可使用范围、是否含个人信息或保密内容、适用产品与版本、允许进入的系统、接收方以及删除或复核节点。来源可能包括自有设计文件、客户提供的图纸、供应商样本、公开标准、购买的数据库和由员工整理的故障记录。其“在公司电脑上”并不自动等于公司有权把它用于任何模型。对确有必要使用的资料，先判断能否提取技术特征或脱敏摘要；若必要上下文无法在权限范围内保留，则把任务移出该系统或改用经过批准的部署方式。该做法是风险筛查流程，不是法律意见。

15.2 专利与技术秘密：公开换取排他权和保密控制的取舍

中国《专利法》区分发明、实用新型与外观设计，并规定职务发明、合作或委托发明的申请权归属以及专利权行使边界。[17]据此，企业对 AI 参与形成的设计不应只保存最终图形，而应保留提出技术问题、选择约束、修改方案、完成验证和作出关键技术决定的过程记录，并依劳动、合作、委托及服务合同核验权利安排。过程记录能帮助建立事实链，但不能单凭提示词或日志自动判定谁是发明人、谁拥有申请权，具体判断须结合适用法律和项目事实。

是否申请专利还涉及披露时点、可保护主题、竞争者可观察性与商业价值。公开申请可能换来一定范围的法定排他权，但也暴露技术方案；保留为商业秘密则依赖信息不为公众所知、具有商业价值且采取相应保密措施等条件。2025 年修订的《反不正当竞争法》第十条规定了以不正当手段获取、披露、使用商业秘密等行为，并定义商业秘密要件。[19]企业若将尚未公开的工艺参数输入外部服务，除核对合同许可外，还需评估保密措施是否被破坏、服务商及分包方能否访问、日志多久保留、内容是否用于模型改进。这里不能推断某次输入必然使秘密失去保护；实际后果取决于信息、控制措施、协议和法律适用。

商业秘密管理与研发复用存在张力。若保护措施是“任何人都不能看到任何资料”，跨部门诊断和知识复用会受阻；若资料广泛进入模型提示与共享库，访问范

SVSAC | sv-sac.org

围又可能超出必要岗位。按资料敏感度配置权限、划定训练与推理用途、记录导出和共享、对项目结束后的访问定期复核，是可以试行的控制选项。控制是否有效，需通过访问日志、越权事件和知识可用性一并评估，不能用签署保密协议替代技术限制，也不能将所有内部技术记录一律认定为商业秘密。

15.3 著作权、来源许可与训练/检索材料

《著作权法》保护符合独创性并以一定形式表现的作品，涵盖工程设计图、产品设计图、计算机软件等类别；法律还规定作品权利归属与许可转让等制度。[18]因此，团队使用公开图片、竞品说明书、付费数据库、客户图纸或供应商软件时，应核实具体许可和合同限制，而不是仅凭“网页可访问”就默认能够复制、上传、微调或再发布。公开获取、内部阅读、模型检索、训练使用和对外发布是不同利用场景，应逐项确认授权范围。

模型生成的图像、文字或工程提案不能自动证明其独立性、权属清晰或不与第三方权利冲突。研发记录宜保留输入材料出处、许可状态、模型输出的人工改动、可识别的近似内容、检索到的依据和最终采纳理由。对高价值设计，企业可在公开、申请或交付前按其流程开展来源审查与专业评估；本篇不把相似度工具的“未命中”当作不侵权证明，也不认为模型服务合同可以消除第三方权利问题。

15.4 生成服务规则与组织内部部署

法规适用范围必须准确表达。《生成式人工智能服务管理暂行办法》针对利用生成式人工智能技术向中国境内公众提供服务等情形，并对提供者的服务规范、个人信息保护、输入和使用记录等提出要求；国家网信办的官方答问说明，企业、科研机构等研发或应用相关技术但未向境内公众提供生成式人工智能服务的，不适用该办法相关规定。[20]这不等于内部使用免于个人信息、数据安全、知识产权、劳动、合同或行业规则约束。企业应按自身是服务提供者、组织使用者还是开发者的具体角色识别规则，而不是把“内部部署”或“云端调用”作为唯一法律判断条件。

2025 年施行的《人工智能生成合成内容标识办法》对符合所列网络信息服务规则情形的服务提供者规定显式和隐式标识要求，用户在网络信息内容传播服务发布生成合成内容时也有相应声明和标识义务。[21]办法的适用主体和场景不能直接扩张为所有内部研发草稿都必须采用相同形式的标识；但内部文件保留生成来源、模型版本和人工修改记录仍有助于追踪版本与复核依据。若生成内容进入公开传播、客户材料或相应网络服务，组织应由责任部门按具体角色、渠道和标准核实要求，不能只依靠员工对规则的概括记忆。

SVSAC | sv-sac.org

部署选择还包含技术边界。公共云服务、专有实例、本地推理和检索增强系统在数据控制、性能、维护责任、更新速度及成本上可能各有权衡。离线或本地部署可减少某些外传路径，却不会自动解决资料权属、访问管理、模型错误或备份删除问题；云服务也不能仅因外部运营就被断定不可用于研发。评估应以数据流图和供应商合同为起点，检查哪些内容离开控制域、用于何种目的、保留多久、由谁访问、如何导出删除，以及发生服务中断时如何维持工程流程。对客户或员工个人信息仍须按既有法律要求处理目的、必要性和期限；对涉及国家秘密、重要数据或特定行业义务的项目，需要额外的专业核验。

15.5 从法条到可审查的项目动作

法律文本规定边界和责任，并不能代替项目级事实调查。企业可把每项试点拆成“输入许可—处理路径—产出审查—部署角色—保存与退出”五项检查，并由对应职能提供可追踪的依据。若资料来源不明，先不将其输入；若合同条款不清，要求确认服务范围、保留与再利用方式；若输出将成为专利申请、客户承诺或公开内容，安排相应专业审查；若部署范围变更，则重新核对数据流。此类分流能够把需要法律判断的问题及时转交有资格人员，也让研究者知道哪一项控制影响了周期和成本。

本次分析没有审查具体模型服务合同、企业数据流、客户协议或知识产权作品，也没有对任何组织作合规评价。上述官方法条和规章只支持文中所述的一般规则及适用范围，不足以对单个输入、输出或部署方式作合法性结论。此限制应保留在后续案例和建议中。

16 研发组合、试验资源约束与流程标准化对照

16.1 研发预算不是生成预算

AI 可能降低产生候选方案的成本，却不会同步降低样件、测试、客户验证、认证和工程评审的资源需求。若预算按生成数量扩张，新增候选会争抢有限验证能力。此时，企业的决策问题不是“还可以生成多少”，而是哪些不确定性最值得花钱消除，以及不同研发方向的试验资源如何分配。应把候选视为需要购买证据的选项，而非已经形成价值的资产。

组合层面的选择需同时观察价值潜力、不确定性、可逆性、战略相关性和证据成本。高不确定而验证代价低的方向，适合先做廉价测试；高潜在价值但试验昂贵的方向，可能需要阶段性投入并设置停止节点；风险低、竞争差异弱的方向，也许仍值得通过小改良维持客户兼容，却不应冒充突破性创新。上述分类只

SVSAC | sv-sac.org

是决策结构。权重和门槛需由企业根据损失、资源和战略明示，不能由一个综合分数掩盖风险容忍度的管理选择。

组合层面的多样性也要与市场组合和能力组合区分。不同外观但依赖同一客户需求、同一工艺和同一供应商的方案，可能只形成表面多样；功能机制不同但共用测试设备的方案，既可能节省资源，也可能造成测试瓶颈。若所有预算集中在同一个预测看似高收益的方向，组合可能对需求误判或法规变化过度敏感。保留方向差异有期权价值，但须以可检验的假设和清楚的资金上限为条件，不是把资源平均分给所有点子。

16.2 试验队列与共享资源的排程

企业应把实验室设备、模具、领域工程师、供应商样件和客户现场窗口视为稀缺资源。排队时间与主动测试时间需要分开：模型加快了设计准备，若所有候选同时涌入同一测试台，交付等待反而可能变长。优先级可以考虑每单位稀缺资源所能减少的关键不确定性、结果对决策的影响，以及失败是否能及时终止较大投入。估算应以实际历史记录或透明假设为依据，不可把模型自行给出的成功概率当作已校准的概率。

试验计划还要控制重复与信息价值。若多个候选只是同一机制的小幅参数变化，可以先设计能区分关键因素的试验；若候选代表不同机制，先测其中一个的极值可能无法为其他方向提供信息。对样件数量有限的项目，可在预设范围内选择一部分探索当前可能较优区域，另一部分用于检测模型外推和测量稳定性。无论采用何种分配，选择下一个试验点的原因、未测方向和预算消耗都应记录，使事后能判断是方案失败还是资源分配使重要假设未被检验。

16.3 一个可复现的示范比较设计

为回应“AI 是否比一般流程标准化更能提升研发结果”这一问题，可设计一个分析性方案，但当前没有真实企业样本或实施结果。为使方案具体可复现，示例假定试验观察期为连续 12 周，基线期为此前 12 周；这是用于说明设计的参数，不是适用于所有产品研发的建议周期，真实项目应在分配前按开发节奏登记。研究对象限定为有重复任务、可定义验收条件且能够按团队分配的产品线任务。先把标准化方案定义为需求记录模板、版本编号、评审时点与最小测试记录，并为所有组保留最低必要的安全和质量控制。随后按产品族和实施前团队能力形成区组，在每个区组内将团队随机分配至四种安排：现行流程、仅流程标准化、标准化加 AI 辅助、AI 辅助但不增加标准化。若实际团队无法接受其中某一条件，研究应改用可行的比较设计并明确其因果识别限制，而不是称其为随机试验。

SVSAC | sv-sac.org

随机分配单位设为团队，目的是减少同一团队在不同任务间分享模型提示和检查表造成的组间污染。每个团队在研究期内接收事前确定的可比任务清单，任务数量是固定的分配设计要求。主要结果定义为每个分配团队在这批任务中通过预定工程及客户相关验收的方向总数除以分配任务总数；未按时完成的任务按预先规则计入失败，已完成但资料缺失则标记为未知并另作边界分析。主要因果估计量是按原始分配计算的 AI 提供效应和流程标准化效应的团队平均差异，并报告两因素交互项，涵盖培训、流程适应和实际采用在内。对照组若接触模型或流程材料，需记录交叉使用；不得因不遵从而在主要比较中删去团队。

分析规则应在分配前登记。区组用于控制产品族和基线能力差异；基线任务难度与前一阶段的同口径结果可作为事先指定的协变量，任何事后增加变量均需标为探索性。由于随机化单位是团队，推断必须尊重团队层级，不能把团队内每项候选当作独立样本扩大精度。设 μ_{as} 表示 AI 因子 a 与标准化因子 s 对应的团队平均结果，a、s 取 0 或 1；先按各区组预先确定的团队总体占比汇总四格均值，再以 $\frac{1}{2}[(\mu_{10}-\mu_{00})+(\mu_{11}-\mu_{01})]$ 定义 AI 边际提供效应，以 $\frac{1}{2}[(\mu_{01}-\mu_{00})+(\mu_{11}-\mu_{10})]$ 定义标准化边际效应，并报告 $(\mu_{11}-\mu_{01})-(\mu_{10}-\mu_{00})$ 的交互。这里对另一个因子的两种水平各取二分之一权重；即使实际四格团队数不等，也不改为事后按便利样本量加权。若团队数及分配结构支持设计型或模型型区间，应预先确定估计方法、区组处理与适用条件，并报告不确定性。少团队时，可按实际区组分配枚举或抽样允许的再分配，检验“四种安排对每个团队的结果均相同”的 Fisher 联合尖锐零假设；该联合检验不能替代 AI 边际平均效应为零的一般检验。若反演为区间，必须另加并说明恒定加性效应等额外结构，该区间不能自动解释为一般异质平均效应的区间。若条件不满足，就报告尖锐零假设检验、团队结果和界限，承认平均效应区间证据不足，不制造看似精确的区间。所有路径均报告分配团队数、任务数与缺失状态，不以单个显著性阈值决定结论。可用团队数和基线波动目前未知，因此不编造通用样本量；实施前应据真实基线评估最小可识别差异，必要时承认证据只能作试点判断。

以下基于区间的决策规则，仅适用于所比较平均效应的区间可由设计和假设支持的情形；不能形成此类区间时，应判为证据不足，不按点估计决定扩展。决策规则可分三类。若 AI 组相对仅标准化组的主要结果差异达到事前由管理者与工程负责人确定的最小业务重要改善，关键质量约束未恶化、全过程资源可承受，才考虑扩展 AI 功能；若区间仍覆盖无实际价值改善与有价值改善，结论应为不确定并决定是否追加数据；若差异上界低于事前门槛，或出现无法接受的硬性失效，则不扩展当前方案。门槛依据企业基线、损失和机会成本确定，本篇不

SVSAC | sv-sac.org

替企业指定数值。比较 AI+标准化与仅标准化可检验 AI 的增量作用；仅标准化与现行流程的差异则估计一般流程规范化的作用。若 AI 与标准化存在交互，组合收益不能简单拆成各自独立贡献。

16.4 工时、成本和主要结果的边界

主要因果比较使用固定分配任务数作为分母，回答“提供某种安排后，每批相同任务的合格产出有什么变化”。每百工时合格方向数则将实际消耗工时作分母，而工时会受到干预本身影响，属于资源效率的次要结果；它不能替代按原始分配比较，也不能被解释为同一批固定任务的独立因果结果。

工时至少应覆盖培训、系统接入、数据整理、需求澄清、概念形成、工程复核、试验执行、返工、跨职能协调和维护支持。初始接入与培训可报告未摊销的本期资源占用，同时提供按明示使用周期和任务量摊销的敏感性结果；两种口径均应说明采用年限、预计任务量和残值假设，不能把一次性工作隐去在“平台已经建好”的前提中。研究还应分开统计现金支出、内部人员时间、外部专业支持和设备机会成本，避免一项成本在两种科目重复计算，也避免把现有员工工时记作零成本。

若 AI 组产生更多合格方向但使用了额外专家、测试设备和客户试用机会，结果应同时呈现产出与资源变化。若效率比率上升但主要结果下降，说明团队更省资源却交付较少合格方向，不能仅以单一比率宣布整体改善。反过来，单位任务工时增加也可能对应更充分的验证和更低的现场失效；这需要后续质量与客户结果判断。指标之间的张力是管理信息，不应通过合成为一个总分而消失。

16.5 与一般流程标准化的机制对照

流程标准化和 AI 可能互补，也可能分别起作用。标准化通过明确输入、交接、版本和评审规则降低过程变异，使工作可以比较与追溯；AI 则可能扩展检索、归纳和候选生成。若原流程含糊，模型输出可能放大不同人员对任务的不同理解；若先统一了任务定义，观察到的改善又可能来自标准模板而非模型。反过来，过度标准化可能把研发限定为可计数的例行任务，降低对异常信号和新客户问题的响应。研究应测量实际执行程度和非预期适应，而不是把标准化假定为天然有益。

上述四组比较为区分两种机制提供起点，但还不能消除所有混淆。管理者可能更愿意把 AI 分配给能力较强的团队；团队之间可能交换材料；标准化实施质量可能随负责人变化；市场需求可能在研究期间移动。团队随机化、区组和共同任务可以降低部分偏差，记录污染、人员变化和并行项目可帮助解释其余偏差。

SVSAC | sv-sac.org

若不能随机分配，可按分期引入或匹配团队开展准实验，但必须把选择机制、共同趋势或其他识别假设写明，并将因果主张相应收窄。

直接制造业证据也提醒研究者谨慎扩展结论。Luo、Tang 与 Zhang 基于 2017—2023 年上市制造企业非平衡面板，报告 AI 与制造企业创新质量正相关，并在摘要中指出创新网络嵌入的中介作用及知识多样性的调节作用。[16]该研究为“组织连接和知识组合可能与 AI 创新表现相关”提供直接产业层面的证据，但本篇仅核对出版商摘要，未核查全文的变量构造、识别策略和稳健性检验；样本为上市公司，结果变量是研究中的创新质量指标，也不等于企业新品差异化、客户选择或利润。其结果不能解决上述四组设计中的随机分配问题，更不能提供适用于任意中小企业的部署收益。

16.6 研究边界与后续检验

研发组合涉及多年期产品路线和多个项目间的资源重配，短期任务试验无法完整估计其长期价值。即使随机设计显示某工具改善了特定任务，也不证明组织会把释放的时间用于更有价值的创新，或竞争者无法快速复制。后续研究需要将任务级过程日志与产品级质量、客户留存、毛利和替代品变化连接，并在合理时间窗内观察副作用。缺少此类数据时，研究者只能报告阶段性学习和适用边界。

本篇由此把 AI 差异化研发视为一个同时约束信息、决策、知识产权、验证容量和组织责任的问题。第 13–16 节形成了可进一步检验的产业比较与实施设计，但它们仍是机制分析和建议，不构成企业实证结论。

17 客户选择验证：差异是否改变了真实选择

17.1 满意、偏好与购买并非同一结果

“客户喜欢这个方案”可能指受访者给出的高分、在两张概念图中选择其中一张、愿意支付某个假设价格、接受样机试用，或实际购买并持续使用。它们位于选择链条的不同阶段。概念评价反映特定问法和信息下的态度；二选一表示相对偏好，但未必形成购买承诺；意向支付受假设环境影响；实际采购还受预算、可得性、渠道和组织审批约束；持续使用则加入产品可靠性、学习成本及替代方案的影响。将这些指标统称为“客户验证”，会掩盖差异究竟改变了什么。

Lusk 与 Schroeder 在质量不同的牛排选择实验中比较假设性与实际支付条件，发现受试者在假设条件下报告的总体购买概率通常更高，因而估计的总支付意愿偏高；质量变化的边际支付意愿在两种条件下则通常没有统计显著差异。[22] 这是一个具有真实支付约束的特定食品实验，不是制造业新品研究，也不证明所

SVSAC | sv-sac.org

有假设调查都会产生同样偏差。它支持更窄的方法判断：企业不能把假设购买率直接当作销售预测，且偏差可能体现在总体购买基准而不是每一项属性的边际效应。

一项方案也可能在没有价值差异时获胜。若新方案标价更低、展示位置更好、销售人员投入更多或消费者只看见有限替代品，选择结果无法单独归因于产品特征。因此，验证问题应写成比较命题：对于相同目标客户、同一使用任务、相近价格和可比接触条件，加入某项设计差异是否提高选择、减少放弃，或改善后续使用结果？与其询问“你觉得新功能有用吗”，更有识别力的问题是：当新功能占用空间、价格或操作时间时，客户仍愿意放弃哪个现有属性来换取它？

17.2 参照产品和使用情境决定选择含义

差异必须相对于某个可用替代物来解释。客户可能在现有产品、竞品、维修旧品、自行改造、外包任务或“暂时不购买”之间选择。把实验选项限制为几种企业自己设计的概念，会使“胜出的方案”只表示它在这个受限集合中较受偏好。验证时需要记录替代项是否真实可得、参与者是否认识该替代品，以及当前选择的转换成本。对工业采购而言，设备停机风险、既有接口和审批周期可能让“继续使用旧件”成为强替代方案，即使它的性能较差。

同一产品在实验室、日常环境和极端工况下也可能表现不同。一个轻量手持工具在短时展示中更受欢迎，连续使用后可能因握持疲劳而被弃用；一个设备界面在熟练工程师的受控演示里操作流畅，对缺少培训的轮班人员却可能增加误操作。验证设计至少要标出目标人群、任务、环境、观察时长和重要替代方式。若只覆盖易招募、熟悉科技的客户，结果不应自动外推到主要采购者或维护者。

可以把验证分为递进层级，但不把层级本身当成可靠性的保证。第一层用概念选择淘汰明显不相关的方向；第二层让客户在具有真实约束的任务中使用原型，并观察完成时间、错误、求助和放弃；第三层在合法且可执行的条件下引入真实价格、预算或采购流程，记录是否成交、退回或复购。前一层可以节省后一层的资源，却不能替代它。不同层级回答不同的问题，结果应分别报告，不能把概念偏好率乘以市场规模作为收入预测。

17.3 选择与使用行为之间的反证

若差异化主张是“减少客户的操作负担”，那么销售访谈中的偏好和实际操作中的完成成本应方向一致。若客户选择新方案却完成任务更慢，可能存在新奇偏好、展示引导、学习曲线或测试任务不代表真实工作等解释。若购买增加但退货也增加，价值可能只对一类用户成立，或销售承诺扩大了适用人群。若客户复

SVSAC | sv-sac.org

购，却未能减少停机或维护投入，购买行为可能受合同约束、渠道关系或替换成本影响。反证不是将所有非预期结果解释为“用户还没适应”，而是检验价值命题在哪个环节失效。

一个可执行的比较可将现有方案与新方案按可行条件随机安排给客户团队或终端使用者，预先规定主要任务指标和观察期限，记录价格、供应和培训差异。对无法随机安排的企业采购，可比较自然形成的报价或试用批次，但需说明客户自我选择如何影响结果。研究还应记录拒绝和退出者：只分析完成试用的人会高估接受度。价格变化可作为独立处理，若产品改动与折扣同时发生，就无法判断成交改善来自功能还是优惠。

客户异质性不能在结果不理想时才被发明。企业可事前基于实际任务、维护职责、使用频率或采购约束定义分组，再报告各组方向和样本覆盖。若总体平均差异接近零，但高频用户明显获益、低频用户承担额外成本，这可能支持窄市场定位；若分组是事后挑选且没有重复样本，应当标成探索性结果。分组结果是否值得转化为不同产品版本，还要进入第 18 节的架构和变体成本分析。

17.4 愿意选择何时才构成差异化价值

真实选择依然不是竞争力本身。客户可能选择更便利的方案，却不愿支付足以覆盖变体开发和服务成本的价格；也可能只因供应商关系而暂时采购。企业需要把客户增益和价值获取分开衡量：客户侧观察任务结果、风险或体验是否改变，企业侧观察贡献毛利、服务成本、退货、续约和替换成本。二者不必同步改善。若客户价值提升但竞争者能迅速复制、企业没有相应交付能力，商业价值可能被竞争或渠道分走。

差异化主张也应明确谁承担验证责任。设计团队可以提出“减少步骤”的机制，用户研究负责界定任务和人群，工程负责确认样件与交付条件一致，销售与财务记录价格和成交约束。模型可帮助把访谈整理成候选假设，却不能充当未曾接触的用户样本，也不能用合成画像证明选择概率。只有当来自真实对象的可观察行为支持一项具体主张，且替代解释得到适当检查，企业才有依据决定继续投入或缩小目标市场。

18 产品架构与变体管理：把共用能力和可感知差异同时设计

18.1 架构不是外观分类，而是变化如何传播

产品架构决定功能如何分配给部件、接口如何连接，以及一项改动会不会牵动其他部分。竞争者可以用不同外观提供相似功能，也可以用近似外形采用不同

SVSAC | sv-sac.org

系统关系。若企业只按 SKU、颜色或营销名称统计产品多样性，可能高估独立价值方向，也低估内部结构变更的影响。

本篇将部件自身的改进与部件连接关系的改变分别分析：前者可以保留原接口，后者可能要求重新分配功能、责任和验证任务。Henderson 与 Clark 的文献保留为架构创新的待核理论线索。[23] 本篇未取得其摘要或正文，故不将光刻设备行业细节或既有企业知识失效的具体结论作为已核实案例。上述区分是本次分析的分析安排，不能据此预测某次架构变化必然淘汰在位企业或证明 AI 的效果。

在研发中，模块化常被理解为把部件做成可替换单元，但关键不是模块数量，而是接口稳定性与依赖传播。可观察的问题包括：替换一个功能模块时需要重新验证多少相邻部件；客户定制能否局限于一个边界清晰的模块；跨模块变化是否需要多个工程团队同步修改；变体是否占用独立物料、测试和售后流程。一个图纸上有很多部件的产品，可能因强耦合而难以独立更换；少数功能模块也可能通过标准接口实现较大变型空间。

18.2 平台共用如何支持差异，也如何制造隐性耦合

平台把经验证的基础模块复用于多个产品或客户，理论上可降低重复设计和制造成本，同时把研发资源集中到客户可感知的差异点。但复用成立需要接口稳定、功能边界清楚、供给条件可维持，且共用模块在各目标情境中没有互相冲突的性能要求。若一种平台承担太多互斥需求，工程团队可能通过参数开关、例外接口和手工配置逐步叠加复杂度，表面上共用、实际维护却越来越昂贵。

对标准部件或消费产品，平台可以共用底盘、动力单元、控制板或材料规格，同时改变客户能感知的服务、人体工学或功能模块。对定制设备，共用安全控制、传动单元或软件功能块可能帮助缩短交付，但现场布局、工艺节拍和接口约束仍需项目级确认。平台适用性应以变体之间的差异矩阵表达：哪些参数可配置、哪些模块可替换、哪些改变触发系统重测，以及哪些客户要求无法由当前架构承接。矩阵中的“未知”不能默认当作兼容。

一个反例是过度延展的平台。为了复用，设计团队把低频且高风险的特殊条件塞入通用模块，使所有产品都负担更大尺寸、更高成本或更复杂的维修；专属子系统可能反而更经济。相反，过度分拆也会产生接口治理、库存、质量验证和供应协调成本。决策应比较额外定制自由度的客户收益与架构耦合、认证、库存和售后成本，而不是把“标准化”或“模块化”当作天然优点。

SVSAC | sv-sac.org

18.3 架构变更如何影响研发周期与责任边界

架构决定一次变更的波及范围。若市场部门希望增加一个客户功能，而该功能要求控制、结构和制造工艺同时更改，表面上只是一个选项，实际可能触发多个子系统的设计复核、供应商重新报价和系统验证。反过来，若变化能限制在接口稳定的独立模块，团队可能只重测模块及必要集成场景。是否真的减少周期，取决于接口验证是否足够、测试设施是否排队以及变更是否触及安全或法规边界。

Brusoni、Prencipe 与 Pavitt 研究飞机发动机控制系统的技术发展，指出多技术企业需要掌握超过其内部实际制造范围的知识，以处理技术发展不同步和产品层面难以预测的相互依赖，并协调专业供应商与系统集成。[24]其含义不是企业必须把所有知识内制，而是不能将“外包某部件”误作“无需理解它与系统的关系”。对于 AI 支持的模块搜索或配置生成，关键质量标准应包括依赖是否显式、例外条件是否保留、负责系统集成的人能否复核接口。

若模型提出新的模块边界，研发流程仍须明确接口由谁维护、跨模块故障由谁协调、局部优化如何接受系统检验。这是我们提出的组织设计问题，不是对软件样本结果的复述。MacCormack、Rusnak 与 Baldwin 的文献可作为组织与产品架构关系的待核线索，[25]但本篇未取得其正文，撤回对匹配软件样本、组织松紧和模块性结果的具体断言，不据此建议制造企业采用某种团队形式。

18.4 变体数量与价值密度

变体管理可从三种数量区分：面向客户提供的可选组合数、工程上需要维护的唯一配置数、在供应链和服务网络中形成的物料与流程数。一个配置器可以生成上千种组合，但如果多数组合共享同一结构和客户价值，商业上的有效差异远少于组合数；反过来，两种看似相近的选择可能因维修方式、认证或使用场景不同而需要独立管理。每增加一个变体，企业都应追问它是否服务可识别的客户群、是否改变购买或使用行为，以及由此触发哪些生命周期成本。

“价值密度”可以作为内部比较概念：在相同资源约束下，哪些配置为明确客户问题带来较大收益，又能维持较低的变更和交付复杂度。它不宜被压缩为一个脱离产品语境的标准分。企业可以对每个候选变体估计工程修改、工装、测试、库存、培训和售后负担，再和客户选择或实际使用证据并列讨论。若精确成本不可得，使用区间并指出主导假设，比将未知成本记为零更有决策价值。

产品架构也影响竞争者复制路径。一个易于复制的独立模块可能很快被对手替换或仿制；由工艺、校准、软件、设备接口和服务经验共同形成的系统表现，

SVSAC | sv-sac.org

复制成本可能更高。但复杂性本身不是壁垒：它也可能增加故障、维护和供应风险。企业需要识别哪些接口应开放以便客户采用，哪些依赖必须牢固管理，以及交付的总体性能是否确实超过各单模块之和。第 19 节将进一步讨论这种差异能否被企业保留。

第 17 节的方法判断来自一项真实支付选择实验，不能外推为制造新品的效果。第 18 节中，已读 Brusoni 等的摘要仅支持其所述的知识与生产边界关系；架构创新与软件比较文献仍为待核线索，不作为已核实的案例依据。接口维护、模块责任与系统验证是我们提出的分析问题，不是 AI 改变模块化的实测结论。

19 竞争模仿、差异持续性与价值获取

19.1 当工具普及后，什么仍然可能不同

若竞争者都能调用相似的大模型、生成式设计工具或代码助手，单纯“使用了 AI”很快就不再是可区分属性。更有解释力的问题是：工具进入产品开发之后，哪些投入会转化为客户看得见、买得到且竞争者难以同步复制的结果？答案未必是模型本身。可供检验的候选来源包括专有且经授权的产品使用反馈、独特的过程数据、对失效模式的积累、与客户系统的接口知识、较短的部署验证周期，以及把设计正确地带入供应、校准、维修和持续更新的交付能力。这些是待比较的机制，而非天然形成的护城河。

技术工具的可得性降低了某些构思和试制活动的门槛，却不会自动提供场景定义、可靠资料、试验设施或实施责任。一个竞争者可以在数小时内生成更多方案，但若不知道现场停机由何种条件触发，也无法将候选设计接入客户设备，方案数量就不构成已兑现的差异。相反，企业若持续记录经过验证的操作条件、材料批次、失效结果和维修处置，并获得合适的使用许可，可能形成可用于缩小搜索范围的反馈资产。价值来自数据与任务、测量和决策之间的连接；数据量本身不说明结果更好。

这项机制存在反例。历史数据可能记录了旧产品的局限，因而把新设计锁定在既有模式；使用记录受到维修策略、客户环境或选择性上报影响；记录越细，隐私、保密和保存成本越高；竞争者也可通过测试、逆向分析或供应商转移获得相近信息。过度依赖内部样本还可能忽略新用户和罕见任务。企业要判断反馈资产是否构成优势，必须比较它对候选质量、现场失效率、更新周期或客户续用的增量贡献，并检查替代数据源及获取成本。

SVSAC | sv-sac.org

19.2 模仿路径、互补能力与可占有价值

本篇把功能、系统、服务与商业安排共同形成的价值主张作为差异化分析对象，并区分客户得到价值与制造企业取得回报。Teece 的文献保留为创新获利研究的待核理论线索；[26]由于正文未取得，本篇不将对其框架、资产分类及结论的转述计为已核实依据。在本篇的拟议评估中，复制可能性、制造能力、认证、渠道和维护责任须分别列入待检验条件：有用的功能未必带来本企业利润，专利或保密也不能单独证明销售结果。

模仿有多种路径。对手可以复制界面或性能参数，也可从客户任务重新设计功能，或绕过原有设计，以替代技术解决相同问题。可见部件容易比对，隐含在工艺窗口、材料选择、校准方法、软件更新和现场响应中的知识可能需要较长时间积累；但这些知识若依赖少数人员、未形成可复用流程，也会随离职、供应变化或规模扩大而流失。因此，所谓“难以复制”应拆成对手必须取得什么、取得要花多久、复制后需要补齐哪些互补投入，以及客户是否愿意等待和支付。不能仅凭复杂或专有标签断言壁垒存在。

可占有性也可能和客户采用冲突。封闭接口或专有格式能够保护某些知识，却可能提高客户转换与集成成本，降低采用意愿；公开接口可能扩大装机基础和兼容服务，却让部分模块更易替代。选择权取决于客户系统结构、互补品市场、售后依赖和标准演进。若产品要进入客户既有生产线，开放可验证的接口可能比隐藏协议更能扩大市场；若其差异来自需持续调校的核心流程，企业可能选择保留特定内部参数，同时向客户提供可审计的性能与安全信息。这些都是情境性取舍，不能化约为“开放”或“封闭”更优。

19.3 差异持续性是动态命题

一项优势要持续，至少要经受竞争者回应、客户需求变化和企业自身交付扩张三种压力。竞争者可能快速跟进功能，使早期领先窗口缩短；用户可能改变任务或对成本更敏感，使原有差异失去重要性；企业扩大产量后，原本稳定的小批量工艺可能产生波动，令服务品质下降。产品在发布时有差异，不表示这种差异一年后仍有价值，更不表示利润优势可持续。

因此，时间维度应纳入验证：识别进入市场、对手可观察、客户转换、服务更新和部件替换所需的周期；分别记录功能优势、客户结果和企业收益何时变化。若竞品功能追平后客户仍因更短的故障恢复时间而留存，差异可能从功能层迁移至交付层；若客户采用了产品但不愿续约，则初期优势可能只降低了试用门

SVSAC | sv-sac.org

槛。若更新速度加快同时增加兼容故障、培训和质量返工，频繁发布也可能损害总体价值。任何有关持续性的判断都需要与可能的竞争跟进和使用期变化对照。

专有学习也可能被过度宣称。企业累积了更多日志，不等于获得更多可解释信息；大量训练数据不会自动给出因果诊断；AI 优化出的参数在一个设备、原料或班次上有效，不保证迁移后仍成立。数据反馈价值需要通过数据质量、情境标注、许可范围、测量稳定性和重复验证来评估。若出现重大工艺变化、供应商更替或客户环境迁移，原先关系应视为待重新检查，而非永久资产。

19.4 从优势主张到可反驳的竞争判断

企业可把竞争优势主张拆成连续的机制链：某种 AI 能力或数据资源使候选方案更适配目标任务；产品与服务因此改变可测量的客户结果；客户在可比替代方案中选择并持续使用；企业通过价格、复购、服务或转换成本取得足以覆盖投入的回报；竞争者在给定时间内难以用更低成本复制。每一环都可能断裂。如果效果没有重复出现，优势源可能只是样本运气；如果客户结果改善但无支付意愿，价值未必能被企业获取；如果竞争者很快达到同样表现，先行者也可能只有短暂窗口。

分析时应同时保留反向因果与选择解释。表现好的企业可能更有预算购买数据、模型和工程设备，因此观察到的 AI 使用与产品绩效关联，不足以说明 AI 造成了优势。客户可能选择原有供应商而非更优产品，原因是认证、合同或更换风险。特定项目成功也可能源自负责团队经验而非新工具。比较应记录既有能力、客户基线、项目难度和资源投入，必要时用阶段性部署或合适的对照，避免将先行者优势误读为工具效应。无论用何种设计，结果只能支持其覆盖的人群、任务和时期。

竞争力因此不是某个模型版本或一项专利的静态属性，而是能否持续将识别需求、获取反馈、组织工程变更、完成集成和形成客户可验证结果连接起来。这个判断不意味着所有环节都要内部完成，也不意味着速度必然优先于安全、成本或耐久性。外部供应商、客户共同开发或开放接口可以缩短开发周期；同时，它们也会带来依赖、知识边界和价值分配问题。有关选择的关键，是比较具体的替代安排及其全周期影响，而非使用“生态”“平台”一词代替机制分析。

SVSAC | sv-sac.org

20 从试点到产品组合：有限资源下的选择与相互依赖

20.1 试点单位的效率不能直接回答组合预算问题

第 16 节提出的四臂团队分配方案，可以估计在限定任务和观察期内流程标准化、AI 工具及其组合对任务结果的平均影响，但它不直接决定公司应把下一单位验证资源投向哪个项目。组合管理面对的是多个有不同潜在价值、证据质量、资金需求和时间窗口的项目。单个项目的平均处理效应，即使估计可靠，也不能代替“是否值得挤占另一项研发”的机会成本判断。

项目之间至少有三类资源竞争：稀缺的领域专家时间；无法同时开放的试验线、设备、测试台和认证时段；以及有限预算对原型、数据接入和供应商配合的支付能力。项目还占用管理注意力和客户试用额度。若试验设施是瓶颈，增加构想数量可能只会让排队时间变长，令更晚项目错过客户窗口；若专家在多个项目上分散参与，名义上的并行未必形成有效并行。组合评价要把资源的可替代性和排队效应纳入，而不是只比较项目商业价值评分。

可以用决策期限和资源消耗路径澄清选择。项目甲可能需要较小的原型投入即可在短期验证一个客户痛点；项目乙有较高潜在价值，但依赖昂贵的工艺测试、供应商定制和漫长认证。若两者共享同一测试台，甲的快速验证可能挤掉乙的关键测试周期；若甲的结果可复用到多个项目，先做甲又可能降低组合不确定性。应记录每个项目的最早可判定日期、下一阶段所需资源、等待成本与可复用信息，而非仅按当前预测收入排序。

20.2 不确定性不等于低价值，也不等于可忽略的风险

早期项目往往缺少可比销售数据，评分很容易把“未知”压成一个看似精确的风险分数。估值应区分市场需求、技术可行性、生产可重复性、知识产权边界、客户集成和竞争响应等不确定来源。不同不确定性有不同验证办法：客户需求可通过真实选择和采购约束验证；技术性能需由测量方案确定；量产稳定性要依靠跨批次和供应条件检验；权利和部署问题则需按具体来源、合同和司法辖区核查。用一次演示同时给这些维度打高分，无法说明哪类风险已经下降。

不确定性也有可逆和不可逆部分。方案设计和软件仿真在早期较易调整；专用模具、生产线改造、客户认证和长期采购承诺则可能形成难以回收的成本。把全部预算一次性押在尚未验证的方案上，会将未知转为沉没成本；但若每个项目都无限期做小试，也可能错过规模化、认证或市场进入时机。阶段投入的作用是购买信息并保留选择权，前提是下一阶段的关键问题足够明确，新增证据会改变决策。若无论结果如何都继续投入，所谓阶段门就只是手续。

SVSAC | sv-sac.org

20.3 组合中的相关风险与共同失败原因

项目排序不能只看每个项目各自的预期价值，还需看它们是否暴露于同一失败来源。若多个产品线依赖相同模型供应商、同一敏感数据来源、共同的关键材料或相同客户行业，它们表面上分散，实际可能在供应中断、合规变化或行业需求下滑时同时受损。若许多 AI 试点都依赖同一支中央数据团队，团队拥塞会让项目表现相关下降。组合的多样性不是标题、产品数量或市场名称不同，而是主要风险驱动因素是否真正不同。

相反，过度分散也会损害学习。把预算摊到很多彼此无关的小试点，可能没有任何一个项目获得足够样本、专家注意力或完整部署周期。项目间存在可复用知识时，集中投向一组关联项目可能更高效，例如共享传感器验证方法、质量追溯接口或客户测试协议。关键是区分可共享的通用能力与重复发生的风险：前者支持协同，后者可能造成共同脆弱。组合图可按客户群、底层技术、供应链、数据依赖、试验资源和交付能力标注，并对集中度作定性或有据可查的量化评估。

20.4 比较选择方案，而不是追求虚构的通用阈值

组合评审可为每个候选项目列出目标客户、差异化主张、当前证据层级、主要反证、下个决策点、资源消耗、可复用成果和放弃成本。之后比较不同资金方案：资助高价值高不确定项目的下一步判别性测试；优先扩展已有客户行为证据但交付能力不足的项目；保护能够降低多个项目共同不确定性的试验基础；或者暂缓竞争窗口较远、需要大额专用投入的方向。各方案会形成不同的短期收入机会、信息获取能力和风险集中度。

Cooper、Edgett 与 Kleinschmidt 报告了对 205 家美国企业新产品组合实践与绩效的调查，调查内容包括企业采用何种组合方法、对方法的满意度及其所报告的绩效结果。[27]这是行业问卷研究，可说明组合决策实践和企业报告绩效曾被系统调查；仅据其调查设计不能得出某种流程对绩效的因果效果，也不能直接替代当前企业的基准值。文献可作为提出组合评审问题的背景，而非为 AI 项目设定胜率、投资回报或停止门槛的普遍依据。

因此，不应从一篇跨企业调查、一个试点或既有经验中编造统一的通过率、样本量、收益门槛和“最优”组合比例。可用数据不足时，评审应显式记录分布范围、证据来源与未知项，并说明何种新证据会改变排名。对每个项目而言，明确下一次信息更新的时间和资源成本，比给出精确但缺乏依据的投资回报率更可信。

SVSAC | sv-sac.org

20.5 从团队比较到组合级学习

第 16 节的示范研究设计适合比较相同范围内任务的流程和工具条件。依第 16 节在区组内将团队随机分配到工具与流程条件后，应按团队原分配估计处理效应，并把培训、接入、数据准备、概念复核、测试返工和协调所需工时纳入范围；固定任务分母可降低完成量变化对主比较的干扰。组合层还需记录由试验产生的跨项目学习：例如一项试验是否生成了可复用的材料性能数据、接口测试脚本或客户情境分类；这些产出何时被其他项目实际采用，是否减少了后续成本或失败。

共同学习不应简单按“资料已存档”计算。可检验的复用应有下一项目的使用记录和结果差异；同一模型提示词被复制多次，不代表知识已经迁移。也应记录负面结果的可复用性：某材料在某温湿度范围失效，可能帮助排除相邻项目的错误方向，但仅在边界条件相符时才有价值。组合层面的知识扩散需要判断适用范围、验证新情境以及保留来源，否则容易将局部有效误作普遍规律。

资源分配应允许组合随证据而变，而不是把早期商业评分永久冻结。客户真实选择、模块化试验、供应报价、竞争产品变化或试点成本，都可能更新项目的排序。重大更新要留下为何改变配置的依据，使日后能够区分“根据新信息理性转向”与“为维护已有投入而持续追加”。这种记录帮助形成可追踪的项目学习，但不保证每次选择正确。

21 把试验规则写到可执行：任务入组、验收盲评与争议处理

21.1 先界定任务与差异化方向

第 16.3 节把团队作为分派单位，将固定任务清单的合格方向数除以分配任务数作为主结果，并比较 AI 提供、流程标准化及其交互。本节把这个设计转成操作协议，不改变四格边际效应、交互定义、区组权重或推断边界。真实研究的周数、团队数、任务数与验收门槛均需依据现场条件登记；以下分数和步骤只是示范，不能冒充普遍阈值或已收集结果。

纳入任务须在分配前具有可辨认的客户或工程问题、明确的预期交付物、能够预先写出的最低验收证据，并可由参加团队在研究期内处理。仅有“提升竞争力”“更智能”之类口号而无可观察任务，暂不纳入。安全事件、法定紧急修改、客户停线等不能等待研究安排的任务，应走正常业务流程，不得为维持试验而延误。资料授权不足或泄露可能不可接受的任务，也不得进入需要共享测试素材的试验臂；相应限制需在团队分配前登记。

SVSAC | sv-sac.org

任务清单由不参与结果评分的项目协调员汇集；工程、客户研究和质量代表按同一条件核对。一个“任务”应包含使用情境、问题边界、交付物和主要验收条件。若两个需求面对同一使用者、同一情境、产生同一交付物且由同一验证条件裁决，例如同一手持部件的颜色微调和标识位置调整，可合并为一个任务中的设计要求，避免把一个方向拆成若干易过关的小任务。若需求对应不同操作情境、需要独立验证或失败后可分别部署，例如设备上可单独更换的传感模块与全新的测量原理，则应拆为两个任务，分别保留依赖关系。边界存在争议时，先按能否独立验收、能否单独交付和失败后是否单独撤回三个问题裁决；理由进入清单版本记录。

“方向”则指对任务的一种非重复解决路径，至少记录设计选择、拟改变的客户结果、关键假设和可证伪测试。仅改变命名、颜色或提示词，若未改变机制、交付物或验收结果，按同一方向合并。若采用不同的技术原理、接口安排或服务方式，而且每种路径均能独立通过或未通过验收，才保留为不同方向。任务拆分调整的是分母单位，方向合并拆分调整的是分子识别；两者不得在看到处理组结果后追溯修改。为控制评审资源，可在分配前按核心问题设置每任务可提交方向的共同上限；若现场无法设定一致上限，则所有在截止时间前提交的方向均进入评估，并把生成和评审工时作为次要资源结果记录，不可只截取有利候选。

21.2 锁定任务分母并完成分层分派

协调员生成带时间戳的任务台账，至少包括唯一编号、任务说明、产品族、难度等级、基线流程状态、团队所需技能、预定验收文件、数据权限、相关任务及纳入/排除理由。经独立复核后导出只读版本并计算校验值；之后所有改动另起版本，记录修改人、时间、原因、是否已揭示分组。任务清单、每个团队接收的任务数、观察窗口和主要结果的缺失处理规则须在揭示分派前固定。分母是研究开始时分配给团队的任务，不是最后完成的任务，不因延期、撤项或未出现候选方案而缩小。

先依第 16.3 节的产品族与基线能力形成区组；只有在同区组的团队确实可以被交换分派且四种方案均可执行时，才在该区组内安排现行流程、仅标准化、标准化加 AI、AI 但不增加标准化四种条件。若企业不允许某组不接受最低安全质量流程，则不得为了形式上的 2×2 而削弱保护，可将“现行流程”定义为已存在的最低要求，并重新说明可识别的处理对比。可执行分派由未参与任务评分的统计人员根据预先备案的随机种子和区组名单生成，保存代码、输入清单和输出记录；在任务台账锁定并团队基线登记完成前，不向项目负责人揭露组别。团队数

SVSAC | sv-sac.org

不足以填满区组时，应依据可行分配数合并合理区组或承认设计不支持所需对照，不得事后调换团队以改善四格平衡。

团队在观察期内收到固定任务分配表。临时跨团队援助、使用对照组禁止的模型或流程、团队成员离职、设备停机、客户改变需求等事件均记录日期、任务编号、持续时间和影响范围。主要分析仍按原始团队分派进行；记录偏离是为了描述实施和解释结果，不是按依从情况删选团队。若某任务必须因紧急安全或客户需要撤出试验，协调员填写临时撤销单并保留原分母：结果按预先规则记为未通过，或若只因验收材料确实不可得而记为未知并作边界分析。管理者可要求暂停干预，但须记录决定人、时间、触发理由、受影响任务和恢复条件；不得将撤销任务从台账删除。

21.3 盲态验收与可追溯评分

分派团队提交候选时，需使用相同结构的验收包：匿名任务编号、候选方向、需求到设计的对应关系、图纸或原型版本、测试条件、结果证据、已知不适用范围及尚未解决风险。材料管理员移除团队、处理组、模型供应商、提示词风格和其他可直接表明分组的字段，保留评判安全、功能和来源所需的内容。完全盲态未必可能：模型专有格式、生成文本或方案痕迹可能泄露来源。因此评审者需记录其是否猜测处理组及判断依据，分析报告盲态是否受损，不能宣称“盲审”即保证无偏。

示范 rubric 按每个方向评分，量表可设 0 至 3 级：0 表示没有满足证据或与要求冲突；1 表示有合理解释但关键证据缺失；2 表示达到预先登记的最低要求并附可复核证据；3 表示在最低要求之上呈现额外可验证收益。评分维度至少包括：任务/客户结果对应、工程可行性与接口兼容、测试结果与可重复性、交付和维护约束。涉及法定安全或质量硬约束的项目另设否决项，不用其他维度高分抵消。哪些维度为关键门槛、每个等级的例子，以及达到何种组合算“合格方向”，须由工程、客户与质量负责人在看见分组前用历史样件或模拟材料校准并登记；本篇不指定统一分数。评分依据必须引用验收包中的页码、测试编号或数据行，不能仅写“看起来新颖”。

每个方向由两位未参与该方案开发、且不知道处理组的评审者独立评分。平台先分别保存原始评分，双方提交前看不到彼此结果；同时指定第三名不参与项目执行的裁决人。两人对通过/不通过意见不一致，任何安全否决项判断不一致，或关键维度评分跨越预定合格线时，进入裁决。裁决人先独立查看同一匿名材料与 rubric，再对争议点逐条给出书面决定及引用证据。若争议来自任务边界、要

SVSAC | sv-sac.org

求冲突或证据包缺件，裁决人不得靠主观平均分“抹平”；应将该方向标记为暂缓/未知，由数据管理员按事先流程补件，且补件不可包含处理组身份。若规则本身无法裁决，独立方法负责人记录分歧并将该项列入不确定性分析，不得在揭盲后新设例外。

为便于复核，可报告各维度一致率和分歧类型，但不把单一一致性统计当作评分有效性的证明。第三人裁决解决主结果分类；原始两份评分仍保留，供敏感性分析使用。全部评分冻结后，数据管理员将通过/未通过/未知状态回连团队分派表。分派键与结果表的连接由未参与评分的分析员在锁定评分后执行，并保存连接日期与版本。方向增加、合并、拆分、改名或撤回均保留历史版本，主分析依分配前登记的识别规则计算。

21.4 按原分派估计并如实处理信息不足

主结果继续采用第 16.3 节的团队平均合格方向数除以固定分配任务数，再依事先确定的区组权重汇总四格均值。方向去重规则和任务分母均在揭盲前冻结。四格边际 AI 提供效应、流程标准化效应与交互按 16.3 公式报告，不因实际格子人数不齐而改权重。区组团队分配允许的随机再分配可用于检验“所有四种安排对每个团队的结果都相同”这一 Fisher 联合尖锐零假设；它不等同于一般 AI 平均效应为零。若将尖锐零假设检验反演成区间，必须另行说明恒定加性效应等结构假设；不能把该区间说成一般异质平均效应区间。若设计与样本量不能支持可信的平均效应区间，就报告结果、缺失和界限并判为证据不足，不能按点估计决定扩展。

未按截止时间提交方向的已分配任务保留在固定分母；依 16.3 的预先规则，未完成任务贡献零个合格方向。资料缺失、撤销、无法判定和安全暂停应分别记录，不追溯删除分母。对已登记但验收未知的方向，可分别全判通过和未通过；对清单本身缺失的任务，则按第 12.4 节说明有限上界的依据，无法支持时明确上界未识别。每百工时合格方向数、总方向量、完成时间及复核工时可作次要结果，但工时受处理影响，不替代固定任务分母的原分派比较。报告保留团队和任务流转、交叉使用、未知原因、盲评猜测和协议偏离。

22 跨情境验证：区分可迁移机制与局部效果

22.1 迁移问题不只是复制一遍试验

同一 AI 辅助方式在不同情境中可能通过不同机制影响研发。对标准化零部件，瓶颈或许是成熟规格搜索与公差验证；对消费品，差异可能取决于用户选

SVSAC | sv-sac.org

择、价格和使用周期；定制设备则更容易受现场布局、接口和单次项目需求影响。因而“AI 有效吗”不是可直接迁移的问题。应先说明处理是什么、作用于何种任务、通过哪个环节改变输出，以及哪些条件必须保持相同。

跨情境验证可先将机制拆为处理链：输入资料与权限、模型/工具及版本、使用任务、团队知识、试验基础、产出验收、客户行为和企业捕获。研究者比较各场景中哪些链节保持一致、哪些发生改变，再决定是否共享某个效应参数。例如，两个产品族都把 AI 用于设计空间搜索，但其中一个能用仿真快速筛选，另一个依赖昂贵的实物认证；搜索速度提升并不必然转成同样的交付结果。应分别估计各场景的结果，再报告异质性和可迁移条件，避免直接平均成一个“制造业 AI 效应”。

22.2 情境与替代解释识别矩阵

情境轴	可能变化的机制条件	需记录的替代解释	可比较的证据与边界
任务重复性与验收标准	任务是否可模板化、模拟或复用历史数据	简单任务更易被选入 AI 组；成熟任务自然更快	分别按任务类型报告任务清单、基线结果与验收证据；不把筛选差异当处理效果
产品架构与认证	模块接口是否稳定，变更是否触发系统级重测	复杂产品的工程师和投入本来更多	记录基线架构、重测范围和认证队列；说明效应针对模块还是系统级结果
客户及采购制度	用户是否直接选择、组织审批与合同是否介入	供应商关系、折扣、库存或采购周期改变成交率	同时报客户行为阶段与交易条件；概念偏好不能替代购买和持续使用
数据权限与来源	训练、检索、日志是否完整且有权使用	高数据可得性企业本来能力更强	记录来源、覆盖、缺失、权限和部门投入；仅有相关性时不作

SVSAC | sv-sac.org

			因果归因
团队经验与模型素养	团队是否能校验建议并识别失效	资深人员被优先安排到难任务或AI试点	记录基线能力、培训、实际调用和核查行为；分层分析需事先定义
试验资源与交付路径	仿真、样件、产线、客户试用和认证是否可用	排队、供应延迟或并行项目共同影响结果	报告实际可用资源、等待时长、延期原因；有限试验能力属于实施情境的一部分
竞争阶段与替代品	对手产品、开放标准或替代技术是否变化	行业需求或竞品上市同时改变选择	对齐日历时间与市场变化；内部任务结果不等于竞争优势

该矩阵是登记与比较的框架，不是每项研究都要收集的无边界清单。启动前应按主要核心问题选出关键机制轴及其可观察指标；对不会影响该对比的字段可不采集。替代解释要有具体的可证伪预测：如 AI 组任务在基线就更简单，则应在分派记录中体现；若仅在某类任务改善，可检验该类任务是否确实需要大量搜索，而不是事后把分组边界改到有利结果。

22.3 何时可以综合，何时必须保留异质性

两个研究之间只有在处理、结果、分析单位、客户阶段与关键条件大致可比时，平均结果才有共同含义。不同团队可以分别使用“同一个 AI 工具”，但一个用于重写技术文档、一个用于搜索设计组合，处理内容并不相同。即使任务名称相同，若一组可直接模拟、另一组必须经客户现场验收，结果也不具同一解释。研究综述应先列出这些差别，再决定定性综合或量化汇总。

反向证据应保留其设计和边界。Brynjolfsson、Li 与 Raymond 在单一客户支持企业的分期引入研究中，报告生成式 AI 辅助与每小时解决问题数量提高相关，并发现工人间收益异质；作者特别限定为单一企业的中期效果。[3]Dell’Acqua 等在顾问任务实验中则报告，AI 在其能力范围内的任务可改善质量和速度，超出能力范围时可能降低表现。[30]两项研究不能直接回答制造研发的平均效应，但合在一起说明平均值可能掩盖任务边界、技能和使用方式差异，也说明“员工更

SVSAC | sv-sac.org

快”与“企业产品更有差异”属于不同结果。制造情境需要新的、与交付和客户结果相连的比较，而不是把知识工作者实验当作产业结论。

跨情境研究还要区分复制失败与理论失败。如果同一机制在缺少关键条件时没有作用，例如没有可执行仿真数据，就可能是边界条件不满足；如果条件齐备但结果方向仍与机制预测相反，则是对理论的反证；若指标未测到预期构念，可能是测量失败。报告应明确此三种解释所需的证据，不能把所有不支持结果归到“实施不到位”。

23 模型版本变化、复现与可转移性的边界

23.1 “同一 AI 处理”要精确到什么层级

组织内部常以“使用某模型”描述处理条件，但这一标签可能同时掩盖模型版本、系统提示、检索资料、接口工具、温度与采样设置、权限、数据清理、人工复核以及实际调用方式的差异。研究中若只登记供应商和产品名，就无法判断观察到的变化来自模型升级、提示流程、数据覆盖还是团队适应。复现首先是把处理定义为可检查的配置，而不是期待模型名称本身足以重建实验。

一项评估至少需要保存：供应商和可获取的模型标识、访问日期、版本或部署端点记录、系统与用户指令的版本、检索库快照或不可逆校验值、输入数据来源及清理规则、工具调用与权限配置、随机性设置、界面和人工复核步骤、输出保存格式、验收 rubric 及分析代码版本。对于不允许留存原始客户数据的环境，应保存获准的匿名化样本、字段结构或访问审计记录，而不能为了复现违反数据许可。若供应商不披露权重、后台更新或服务器端配置，应明确标成不可观测，而不是虚构“模型完全相同”。

23.2 复现、重复运行和概念性重复不是一回事

复现一个既有结果至少有几种含义：用同一输入、同一可固定配置和同一分析代码重算；在相同模型端点上重复调用以估计输出波动；用新版本和固定测试集比较变化；在另一团队、产品族或客户环境重做同一机制检验。第一种偏向计算可重复性，第二种测量随机输出稳定性，第三种研究版本敏感性，第四种才接近跨情境概念重复。若服务端模型无法冻结，严格的同一模型复现可能做不到，研究者可用时间戳、端点和已留存响应实现可审计复算，但仍应承认后台状态未知。

D’Amour 等对机器学习流水线的研究指出，多个模型可以在同分布留出测试上表现相近，却在部署分布或稳健性测试中出现明显差异，并以多个应用领域展

SVSAC | sv-sac.org

示这种欠规格问题。[28]该研究不是生成式 AI 研发实验，但它提示评估者不要只保存平均准确度或主结果表；训练/检索资料、测试构成、预处理、选择规则和部署场景都决定证据含义。两次结果差异也不自动意味着哪一模型更优：需看任务覆盖、误差严重性和客户后果。

23.3 版本切换如何设计对照

若企业需要决定是否升级模型，可用分层、时间匹配的版本比较作为未来研究方案。预先选定一组不向模型开发或调参过程公开的保留任务，再选一组用于日常调试的开发任务；每个版本在相同任务包、相同资料权限、相同系统指令和相同验收规则下运行。涉及随机输出时，对可复现实验参数保持固定，并以预登记的重复调用次数估算输出变异；次数须依据预算和基线波动确定，不设一个放之四海皆准的数字。新旧版本输出用匿名编号混排，交由不知道版本的评审者按同一 rubric 评分。

这不是无缝的版本随机试验：版本本身可能改变界面、延迟或工具兼容性；若旧版本在比较时已被服务商停用，比较只能在两者可同时访问期间进行；若任务包泄露到开发和训练中，保留集也会被污染。报告应记录升级日期、服务商变更公告、失败和超时、调用成本、数据权限及人工修复。比较任务结果之外，还应检查工程集成失败、回滚、操作负担和客户侧结果。对不可同时运行或不能设置同一权限的版本，明确其不可比之处，改用时间序列或分期上线时说明同期需求变化、团队学习和其他软件更新等替代解释。

23.4 外推需要跨过的验证台阶

同一版本在受控任务集上通过，不能保证在新材料、新班次、新客户接口、新环境或新团队里可靠；新版本在某一测试上提升，也不能自动视作全产品组合改善。D’Amour 等的研究是外推边界的警示：相近的测试表现可能共存于不同部署行为。[28]企业应保留固定锚点任务以观察随时间的变化，同时安排不同产品族和未知情境的外部验证；每次都报告锚点与新情境之间的差距，以及新增失败类型。

“可迁移”应拆为效果方向、大小、适用条件和实现路径。若某项候选搜索能力在标准零件仿真任务有效，但在定制设备现场验证受安装条件影响，可能只迁移了搜索步骤，未迁移整个交付效果；若新消费品的客户偏好改变但购买不变，可能迁移了概念选择机制，却没有迁移价格和渠道结果。研究者应把迁移对象限定到某个机制链节，并说明哪些环节仍需本地重测。不同模型版本之间、不同团队之间和不同市场之间的迁移是三个不同问题，不宜用“可复制”一词一并概括。

SVSAC | sv-sac.org

模型更新也应作为可能改变处理的事件记录，而不是背景噪声：若供应商变更了生成模型、检索机制或安全过滤，研究者需辨别这是产品版本变化还是运维配置调整，并保存能够确认的说明。变化发生在观察期中时，按变化前后报告暴露，不把整段时间压成“使用 AI”；若调整与研究组或任务难度同时发生，版本比较本身就可能受混淆。

24 文献机制综合与后续验证安排

24.1 证据回答的是不同层级的问题

现有证据并未构成直接证明“AI 让制造企业实现产品差异化”的因果链。概念框架的出版商摘要讨论设计问题与生成能力的关系；[15]制造企业面板研究摘要报告 AI 与创新质量的正向关联及中介、调节模式，但全文识别策略未核，创新质量不等于客户选择或利润；[16]客户选择实验摘要区分假设选择与真实支付；[22]知识与生产边界的研究摘要可支持有限的组织机制讨论。[24] 架构创新、组织与产品架构以及创新获利的另三篇文献仍是未读原文的理论线索，[23,25,26]不列为已取得的历史案例或比较结果。概念、关联和实验各有支持范围，不能相加为同等强度的制造业因果证据。

更近期的工作场景研究也各自对应局部结果。客户支持企业的分期引入研究发现员工每小时解决问题数平均提高且效果异质[3]；顾问任务随机实验则发现 AI 在测得的能力范围内提升表现，超出该范围会降低表现[30]。前者测员工工作流效率，后者测知识工作任务质量与速度，均未直接测制造产品的差异化、采购或利润。与面板正相关并置时，合理的综合不是选择其中一个结果代表所有情境，而是提出待检验的机制链：工具改变搜索、创作或分析过程；过程影响候选质量、成本与周期；验收和交付把候选转为可用产品；客户在替代品中选择并持续使用；企业通过互补能力获取足够回报。每一环节有独立结果变量，也会有不同失效条件。制造企业面板摘要[16]报告 AI 使用与创新质量正相关，但只凭摘要无法核实变量构造、识别策略和稳健性，且选择更可能采用 AI 的企业本就可能拥有更多研发投入、数据和组织能力。因此该关联既不能视为相反证据已被排除，也不能被当作本项目试点的先验效果大小。更合理的做法是把企业层面关联当作提出可检验问题的线索，随后以任务分派、工程验收和实际客户结果识别不同因果环节。

SVSAC | sv-sac.org

24.2 相反解释及其可区分预测

把上述证据放在一起，至少需要区分四种机制解释。第一，效率机制认为 AI 压缩搜索或制作时间；若如此，主要变化应先出现在完成时长或单位工时产出，客户侧结果不必同步改变。第二，能力补足机制认为工具帮助经验较少者完成既有任务；若成立，收益可能集中在基线能力较低或缺少范例的任务上，且未经复核的质量不一定相同。第三，边界误判机制认为用户不能准确识别模型能力范围，导致在错误任务上过度依赖；若成立，失败应与任务类型、验证负担和调用后人工检查不足相关。第四，互补资产机制认为价值主要来自系统集成、生产、认证、渠道和服务；若成立，早期候选数量和速度提升并不会自动带来交付成功或客户支付。

可辨认机制的设计应预先指定不同层次的指标，并在结果不支持时保留替代解释。例如，候选方向增加但盲评合格率下降，可能是生成量上升而筛选能力不足；合格率提高但客户实际选择不变，可能是设计改善未触及主要采购条件；客户采用增长而企业毛利下降，可能是服务与认证成本吃掉了价值；速度提高而后续返工上升，可能是前端加速把问题推迟到昂贵阶段。只有把过程数据、独立验收、客户行为和成本结果连接起来，才能判断差异发生在链条的哪一处。此处是未来研究的识别逻辑，不是对本次分析已有企业样本的因果陈述。

24.3 制造研发的后续研究路径

更有解释力的后续研究可以分阶段开展，但每一阶段解决不同的因果和外部效度问题。第一阶段在能够按团队分派的研发场景中，按第 21 节锁定任务清单与固定分母，使用第 16.3 节的四格设计，测量预先定义的可接受方向、质量约束、全口径工时与失效边界。团队而非任务是分派单位；团队内任务是重复观察，不可当独立随机样本。试验所需团队数与最小可识别差异应用真实基线和可行区组结构估算；若不够，只报告受限的试点证据，不假装验证了无差异。

第二阶段应把视角移到产品与客户端。对通过工程验收的候选，沿用相同价格、替代品、任务情境和接触条件，记录客户在概念选择、原型试用、真实采购、退货、复购和持续使用各阶段的转化；无法随机化时，说明选择机制和比较对象。需建立候选方向、工程变更、验收任务与客户结果之间的可追溯关系，避免把一次性原型评分直接等同于商业差异化。结果不支持时，须区分客户未感知、采购约束阻断、产品交付失败和价值无法获取。

第三阶段研究跨版本与跨产品族迁移。预先规定模型版本和配置变更的登记结构；固定保留任务用于版本内比较，不被调试流程反复调用，外部验证集来自

SVSAC | sv-sac.org

不同团队或情境。新版本在锚点任务表现改善，但外部情境失效增加时，报告为版本特异的取舍，不以单一平均分掩盖。若能开展多企业或多场景研究，应事先声明哪些共同工具、任务特征和验收口径可比较，哪些结果只能在各场景内解释。异质性分析围绕预先定义的任务和能力指标进行，避免结果揭示后再切分子组。

此外，未来研究宜同时登记预先提出的受益机制与可观察的失败预测。例如，若主要假设是 AI 减少早期搜索成本，则在最终验收不变时应先看到候选探索工时下降；若成本下降但方向质量也下降，效率假设部分成立而净产品价值未必改善。若主要假设是模型建议帮助低经验团队，则应事先说明经验指标及如何测量学习，而不是只在平均值为正后寻找获益子组。把这些预测预先列出可减少事后改写理论的空间。最后，纵向观察需要把上线时间、竞争者响应、产品架构变化和互补资产投入纳入观察窗。若差异仅在初期领先，随后对手跟进或客户换用替代品，竞争优势是暂时的；若客户持续选择但利润没有改善，价值捕获机制需进一步解释；若差异随服务和反馈积累增强，则要检验哪些投入可复制、哪些来自情境专属经验。这里提出的是可以逐步累积的后续行动安排，而非声称一个试验即可识别全部效果。

25 讨论：从生成能力到可验证的竞争价值

25.1 对核心问题的回答

我们提出的三个问题可以作条件性回答。第一，AI 可能扩大搜索、归类、建模和方案构造的空间，但是否扩大有效差异，取决于输入信息是否涵盖异质客户与真实约束、选择机制是否保留不同解题路径，以及工程验证能否识别模型看似合理却不适配的候选。若多个团队向相似工具提供同质资料、采用同一筛选标准，生成数量上升并不代表客户面对的价值选项增加。现有证据支持把这一机制作为待检验命题，而不是宣告制造产品趋同已经发生。

第二，效率是否转化为客户价值，需沿研发链条分别检验。工作时长缩短只表明特定任务可能更快；候选通过技术验收表明满足约定的工程条件；真实客户选择和持续使用表明产品差异触及了使用或采购行为；企业毛利、复购、维修支出和可持续交付才关系到价值获取。任一环节改善，都不自动证明后续环节同步改善。生成、验收、采用和获取回报之间既可能相互强化，也可能断开。

第三，缺少大型数据平台或算法团队并不妨碍形成可复查的本地试验，但试验结果要严格限定于具体任务、团队、版本和客户情境。本篇给出的团队分派协

SVSAC | sv-sac.org

议、方向验收和版本记录属于研究设计建议，尚未执行。若企业无法随机安排，也仍可开展分期引入或匹配比较，但选择机制和同期变化须明确；有限样本若不能支持平均效应区间，应承认证据不足，而不能用点估计替代可靠性。

25.2 相对“多生成”和“单任务提效”的增量

单纯统计生成方案数，混合了模型产出量、团队筛选规则、候选重复率与评审资源。方案更多可以带来更多可选路径，也可能增加去重、核验、集成和淘汰负担。若未观察最终保留的独立方向及其客户结果，“生成越多越好”没有说明价值发生在哪里。本篇的补充是将差异化作为多个结果层次：候选是否对应独立问题，工程上是否可交付，客户是否在真实替代品中选择，企业能否在持续服务中取得回报。生成量仅能作为过程指标或搜索投入结果。

单任务效率研究则主要回答某个岗位或任务能否更快完成、产出质量是否变化。客服部署研究报告每小时解决问题数提高，但作者把结论限定于单一企业、特定职业及中期效果；不同经验员工收益不同。[3]管理顾问任务随机实验显示，任务落在测得 AI 能力范围内时表现可改善，超出范围时则可能变差。[30]人机协作元分析还指出，组合表现未必超过人和 AI 各自较强的一方，任务类型会影响结果。[10]这些发现使“部署后员工更快”不能直接替代“产品更有差异”这一问题。本篇的增量在于明确连接任务效率、候选差异、盲态工程验收、实际客户行为和企业价值获取，要求分开测量并检验何处断链。

因此，相较于单任务平均效应，企业更需要知道效率释放后时间和预算流向何处：是否转向更多独立客户问题，是否减少返工，还是只增加待评审清单；共同约束是否使团队更快收敛到标准答案，还是使低频但关键的用户问题被筛掉。这些问题不能靠工具调用量回答，而要靠固定任务分母、可审查的候选定义、外部验收和随访结果共同判断。

25.3 相反证据与替代解释如何改变论证

不同来源处于不同分析层级。制造企业面板摘要报告 AI 与创新质量正相关，[16]但全文识别细节未核，“创新质量”也不是产品差异或利润；高绩效企业原有资源和能力可能构成替代解释。客服分期部署与顾问任务实验测的是员工和任务结果，不能充当制造产品竞争力的重复证据。[3,30] Autodesk 与 GM 材料只支持厂商所述初期概念验证，不证明量产、总成本或客户选择。A-Lab 论文及更正未完成原文核查，[13,14]不参与材料发现成功率或自动化科学效果的比较。

更深一层的替代解释来自采用者选择、团队能力、任务组成与互补资源。更成熟的团队可能更会使用 AI，也更容易完成复杂项目；AI 试点项目可能被优先

SVSAC | sv-sac.org

安排资源；更大的方案数量可能只是问题范围变宽，而非同一任务的搜索效率提高。模型建议带来的短期速度也可能伴随人工核查或下游纠错增加。要区分这些解释，研究需要先锁定任务与分派，记录基线能力、实际调用、复核工时、候选去重和失败原因，再把结果连接到客户试用和交付成本。只有符合识别条件的对照才可支持因果判断；跨行业相关性或前后变化只能支持更有限的推断。还需区分“工具可用”“任务适用”和“产品有商业差异”三个决策对象。模型在一个任务上生成答案的可行性，只支持工具与任务的技术匹配；通过盲评，才支持候选满足既定工程标准；只有客户在真实替代品、价格与采购条件下选择，并且企业能承受交付成本，才支持商业差异化判断。把这三个对象合并成一个试点成功率，会使企业无法知道下一步应买工具、改流程、补工程验证，还是停止产品方向。

因此，支持性与不支持性发现都应保留。AI 提高了某类任务速度，但盲评质量没有改善，说明效率未转化为合格方向；技术验收提高而购买未变，说明工程价值未通过客户选择；购买改善但毛利下降，提示交付成本或价值分配问题；某版本在固定测试集提高而新情境错误增加，说明内部得分不能代表部署鲁棒性。与其事后将失败都归因于“组织尚未准备好”，不如预先为各机制列出可观测的成功与失败模式。

25.4 收敛何时合理，差异何时没有商业价值

产品组合趋同并非总是坏事。若客户共同重视的安全、可靠性、接口或可维护性约束明确，经过多种独立试验后没有证据表明其他方案带来额外客户收益，收敛到标准模块或稳定规格可能降低质量波动、库存复杂度和维修成本。在此情形下，竞争力可能来自可靠交付和总拥有成本，而不是表面设计新奇。但企业仍需保留能发现少数关键需求的渠道，避免以平均需求掩盖小群体高损失场景。

反过来，差异本身也没有商业价值。若客户无法感知某项属性、使用结果没有改变，或客户虽喜欢却不愿承担相应价格，差异只能增加研发和变体管理成本。即便客户真实获益，如果相关改进容易被竞争者复制，企业又无法凭生产、服务、认证或渠道捕获回报，也不应将其直接称为可持续竞争优势。价值判断需同时看客户净收益、企业净收益和差异维持周期；高新颖性或高专利数都不能代替这三个判断。

收敛和差异化应按问题选择，而非预设孰优。标准件可把资源集中在一致的接口与可靠性上，同时对关键边界工况保留专门方案；消费品可减少缺少需求依据的装饰变体，将有限实验资源用于实际使用和复购行为；定制设备可复用经过

SVSAC | sv-sac.org

验证的平台模块，但对现场布局和工艺变化保留独立工程验收。这些是由前文机制导出的分析性启示，不是已比较三类行业后得出的平均效果。

25.5 可识别性、外部效度与研究局限

本篇没有制造企业原始数据、已执行试验或采购跟踪，不能估计差异化新品、利润或竞争位置的因果效应。证据范围随来源而异：客服正式版与顾问实验依据相关摘要及引言，[3,30]写作数字依据 MIT 机构报道，[1]制造商面板依据出版商摘要，[16]GM 依据厂商概念验证原文。[12] A-Lab 论文及更正未取得正文，[13,14]相关具体结果不作为实证支持；尚未读到的理论文献也不支持本篇对其案例细节的断言。本篇因此是有边界的机制分析与拟议协议，原创性与跨行业适用性仍需系统比较，不能汇总出制造业 AI 提效幅度。

另一个限制来自定向文献选择：本篇没有系统检索所有领域和所有语种，发表的正向结果也未必代表未发表或企业内部无效试验的分布。即使来源均可核验，按机制选取的研究仍会遗漏未检索到的证据。故本讨论适合作为研究设计和论证边界，不适合据以声称完整统计了领域内所有正负效果。未来工作的可识别性取决于分配单位、共同干预和测量边界。若以团队分派、却把团队内多项任务当独立样本，会低估不确定性；若处理组和对照组同时改变流程、工具、培训和资源，难以识别单一作用；若删除未完成任务，可能高估合格率；若模型在试验期间更新却不记录，试验处理本身就不稳定。第 16、21 和 23 节提出的规则是为了让这些问题可以被检查，但不保证一个小型企业试验必然有足够统计能力，也不消除跨团队污染。

外部效度同样有限。一个版本、一个产品族或一个采购阶段的结果不自动推广到新客户、新供应商、不同认证制度或下一代模型。由同行评议知识工作实验得到的能力边界，需要在制造任务上重测；由内部工程验收得到的改进，也需要经过真实客户和量产条件验证。研究应清楚标注结果迁移到了机制链哪一节，避免从“通过任务验收”跳到“获得市场优势”。未来复制失败应能区分理论边界、实施条件、版本漂移、样本构成和测量失配，而不是统一归为偶然波动。

后续研究最有价值的方向不是把试点数字做得更大，而是建设可连接项目任务、模型版本、验收记录、客户选择、售后结果与竞争跟进的纵向证据，并公开主张的适用范围。试验之前确定可能的替代解释、实际业务门槛和停止条件；观察后保留负面结果与未知项。这样的积累才可能逐步判断，哪些差异能由 AI 帮助生成，哪些能通过工程和组织能力交付，哪些最终能为企业和客户共同保留价值。

SVSAC | sv-sac.org

25.6 用分层证据决定继续、转向或停止

研发决策不必等待一个覆盖所有市场、版本与产品族的总分，但每一步投资都应对应一个尚未解决的问题。技术可行性阶段可以确认候选是否在规定条件下产出；工程阶段检查安全、接口和重复性；客户验证阶段比较选择、真实试用和购买；商业评估阶段检查利润、售后负担、模仿速度与价值捕获。只有前一阶段的证据支持进入后一阶段，才能避免把概念展示越级说成客户验证或市场成功。

如果工程验收通过但客户偏好没有变化，下一项研究应针对需求与替代品，而不是立即扩大模型许可；如果客户结果改善但交付复杂度上升，决策对象是变体架构与生命周期成本；如果竞争者迅速复制且企业没有互补资产，可能应缩短投入期、转向服务或不再把该特征视为长期差异。若某个客户群获益明显而总体平均接近零，则可以比较限定市场的规模与支持成本，但不能事后选择有利群体并称总体有效。

阶段最低业务意义、不可接受风险和追加试验成本，应在查看分组结果前结合企业基线、采购周期和替代方案确定；本篇不提供跨行业门槛。追加信息的价值在于它能否改变继续、转向或停止决策，不能仅因样本更大而认定更好。低成本试验即使得到负面结果，若能排除高额错误投入，也可能值得进行；但测试排队、客户窗口和专家时间同样有成本。比较时应说明哪些不确定性可能被新试验减少，以及哪个决策会随结果改变。若现有可行试验无法区分关键替代解释，正确做法是缩窄决策对象或暂缓该项投入，而非反复以同一试验“证明 AI 有效”。

26 决策建议

本篇关心的是：在 AI 工具日渐可得时，制造企业能否把生成能力转化为有客户价值且可持续交付的产品差异。现有文献与案例尚不足以作出普遍肯定的经验结论。较稳妥的研究结论是一个有条件的判断：AI 可以成为扩大研发搜索和缩短部分工作任务的手段，但其竞争贡献取决于信息是否真实多元、筛选与验收是否可靠、工程接口和生产服务能否承接，以及客户是否在可比替代方案中选择并持续使用。只增加候选数量，或仅在孤立任务上提高速度，都不构成已验证的差异化优势。

有限的产业启示是把研发投资从“买到什么模型”转向“要解决哪个客户问题、怎样判断通过、需要哪些资源、何时停止或升级”。企业可以优先选取重复、可定义、风险可控且结果可观察的任务，建立不扩张固定任务分母的比较；同时将工程、试验、复核、返工、数据接入与客户交付成本纳入判断。对缺乏足够团队或

SVSAC | sv-sac.org

客户数据的企业，试点适合用于检验可行性与识别边界，不应包装成通用收益保证。对客户价值尚未证实、交付成本高于可捕获回报或安全约束不满足的方向，停止或收窄投入是合理结果。竞争者采用同类工具的压力可以成为评估理由，却不能替代本企业的任务适配和机会成本分析。若工具把一个瓶颈从概念搜索移到工程审核、验证排队或售后纠错，管理者应以全流程变化判断是否真的释放资源。

我们提出的是供后续数据检验的分析框架和研究协议，不是对河北或其他制造企业 AI 收益的调查报告，也不预测特定项目必然成功。最终判断应由可复核的工程证据、真实客户行为和完整成本共同支持，并允许结果否定 AI 投入的原始假设。只有在这些条件成立时，AI 辅助的快速迭代才有理由被称为差异化研发，而不是更快地产生尚未验证的设计。

给企业的下一步建议

AI 研发最值得投入的方向，是把客户尚未被满足的需求转化为可验证的产品差异，而不是批量生成相似方案。因此，我们建议把下一步工作收敛为一项有明确交付物的专项任务。

具体起点是：从一个产品族和一项明确痛点切入，让设计、工程与销售共同确定试验指标，再用样机和客户反馈决定是否扩大投入。在此基础上，管理层可以按证据充分程度、资源投入与业务影响确定优先级，先验证关键假设，再安排后续合作与投入。

要把上述判断用于实际经营，可以围绕具体产品或项目、目标市场、现有能力和决策难点，进一步形成差异化机会图谱、AI 研发试点方案、新品验证路线图。这类工作需要把行业研究、企业诊断与合作条件核验衔接起来，让分析结果成为管理层可以使用的决策依据。

资料与分析范围

下列说明列明资料范围、分析依据与适用条件，便于读者辨别文中的事实、推演和建议。市场进入、交易安排及适用规则仍需结合具体项目核实。

产品同质化并非单一的设计问题，而是需求识别、方案搜索、工程约束与商业反馈之间的信息失配。在企业普遍获得相似生成工具的条件下，提高方案生成速度可能同时加快模仿与趋同。本次分析以定向证据梳理和机制分析为方法，讨论 AI 如何在差异化研发中发挥作用，并提出可检验的企业实施框架。公开写作实验和客服部署研究提示 AI 可提高特定任务绩效，创意研究则提示个人表现改

SVSAC | sv-sac.org

善可能伴随集合多样性下降；这些发现均不能直接等同于制造企业的新产品利润提升。本篇据此将差异化界定为特定客户在真实约束下感知、愿意选择且企业能够持续交付的价值差异，区分生成、筛选、验证与组织学习四种作用机制，并以可追踪需求、独立创意来源、约束测试、全过程成本和持续反馈作为实施条件。本篇未获取企业原始数据，不报告真实企业采用 AI 后的因果效果；所提度量与试验设计属于研究建议。由此形成的核心命题是：AI 对竞争力的贡献取决于企业是否将生成能力嵌入保留异质信息和检验客户价值的研发系统，而非单纯增加方案数量。

SVSAC - sv-sac.org

SVSAC | sv-sac.org

参考文献

[1] Noy, S.; Zhang, W.. Experimental evidence on the productivity effects of generative artificial intelligence. Science 381, 187–192 (2023). https://www.science.org/doi/10.1126/science.adh2586

[2] Doshi, A. R.; Hauser, O. P.. Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances 10, eadn5290 (2024). https://www.science.org/doi/10.1126/sciadv.adn5290

[3] Brynjolfsson, E.; Li, D.; Raymond, L.. Generative AI at Work. The Quarterly Journal of Economics 140(2), 889–942 (2025). https://academic.oup.com/qje/article/140/2/889/7990658

[4] Lewis, P. et al.. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Advances in Neural Information Processing Systems 33 (2020). https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[5] NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0). (2023). https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

[7] 全国人民代表大会常务委员会. 中华人民共和国个人信息保护法. (2021). https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm

[8] Meincke, L.; Nave, G.; Terwiesch, C.. ChatGPT decreases idea diversity in brainstorming. Nature Human Behaviour 9, 1107–1109; Matters Arising (2025). https://www.nature.com/articles/s41562-025-02173-x

[9] Lee, B. C.; Chung, J. J.. Reply to: ChatGPT decreases idea diversity in brainstorming. Nature Human Behaviour 9, 1110–1112; Matters Arising (2025). https://www.nature.com/articles/s41562-025-02195-5

[10] Vaccaro, M.; Almaatouq, A.; Malone, T.. When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour 8, 2293–2303 (2024). https://www.nature.com/articles/s41562-024-02024-1

[11] Frazier, P. I.. Bayesian Optimization. INFORMS TutORials in Operations Research, 255–278 (2018). https://pubsonline.informs.org/doi/10.1287/educ.2018.0188

[12] Danon, B. / Autodesk. How GM and Autodesk are using generative design for vehicles of the future. (2018). https://adsknews.autodesk.com/en/news/gm-autodesk-using-generative-design-vehicles-future/

[13] Szymanski, N. J. et al.. An autonomous laboratory for the accelerated synthesis of inorganic materials. Nature 624, 86–91 (2023). https://www.nature.com/articles/s41586-023-06734-w

[14] Szymanski, N. J. et al.. Author Correction: An autonomous laboratory for the accelerated synthesis of inorganic materials. Nature 650, E1 (2026). https://doi.org/10.1038/s41586-025-09992-y

[15] Verganti, R.; Vendraminelli, L.; Iansiti, M.. Innovation and Design in the Age of Artificial Intelligence. Journal of Product Innovation Management 37(3), 212–227 (2020). https://onlinelibrary.wiley.com/doi/abs/10.1111/jpim.12523

SVSAC | sv-sac.org

[16] Luo, T.; Tang, H.; Zhang, J.. Artificial intelligence and the innovation quality of manufacturing firms. Journal of Manufacturing Technology Management 37(5), 885–904 (2026). https://www.emerald.com/jmtm/article/37/5/885/1330669/Artificial-intelligence-and-the-innovation-quality

[17] 全国人民代表大会常务委员会. 中华人民共和国专利法（2020 年修正）. Official statute; reproduced by China National Intellectual Property Administration (CNIPA) (2020). https://www.cnipa.gov.cn/art/2020/11/23/art_97_155167.html

[18] 全国人民代表大会常务委员会. 中华人民共和国著作权法（2020 年修正）. Official statute; Ministry of Commerce intellectual-property law database (2020). https://ipr.mofcom.gov.cn/law/detail.shtml?id=1969

[19] 全国人民代表大会常务委员会. 中华人民共和国反不正当竞争法（2025 年修订）. Official statute; reproduced by China National Intellectual Property Administration (CNIPA) (2025). https://www.cnipa.gov.cn/art/2026/5/20/art_104_206437.html

[20] 国家互联网信息办公室等. 生成式人工智能服务管理暂行办法. CAC department rule and official scope clarification (2023). https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

[21] 国家互联网信息办公室等. 人工智能生成合成内容标识办法. CAC interdepartmental measure (2025). https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm

[22] Lusk, Jayson L.; Schroeder, Ted C.. Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. American Journal of Agricultural Economics 86(2), 467–482 (2004). https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0092-5853.2004.00592.x

[23] Henderson, Rebecca M.; Clark, Kim B.. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. Administrative Science Quarterly 35(1), 9–30 (1990). https://www.jstor.org/stable/2393549

[24] Brusoni, Stefano; Prencipe, Andrea; Pavitt, Keith. Knowledge Specialization, Organizational Coupling, and the Boundaries of the Firm: Why Do Firms Know More than They Make?. Administrative Science Quarterly 46(4), 597–621 (2001). https://journals.sagepub.com/doi/10.2307/3094825

[25] MacCormack, Alan; Rusnak, John; Baldwin, Carliss. Exploring the Duality between Product and Organizational Architectures: A Test of the 'Mirroring' Hypothesis. Research Policy 41(8), 1309–1324 (2012). https://www.sciencedirect.com/science/article/abs/pii/S0048733312001205

[26] Teece, David J.. Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing, and Public Policy. Research Policy 15(6), 285–305 (1986). https://www.sciencedirect.com/science/article/pii/0048733386900272

[27] Cooper, Robert G.; Edgett, Scott J.; Kleinschmidt, Elko J.. New Product Portfolio Management: Practices and Performance. Journal of Product Innovation Management 16(4), 333–351 (1999). https://onlinelibrary.wiley.com/doi/10.1111/1540-5885.1640333

[28] D'Amour, Alexander; Heller, Katherine; Moldovan, Dan; Adlam, Ben; Alipanahi, Babak; Beutel, Alex; Chen, Christina; Deaton, Jonathan; Eisenstein, Jacob; Hoffman, Matthew D.; Hormozdiari, Farhad; Houlsby, Neil; Hou, Shaobo; Jerfel, Ghassen; Karthikesalingam, Alan; Lucic, Mario; Ma, Yian; McLean, Cory; Mincu, Diana; Mitani, Akinori; Montanari, Andrea;

SVSAC | sv-sac.org

Nado, Zachary; Natarajan, Vivek; Nielson, Christopher; Osborne, Thomas F.; Raman, Rajiv; Ramasamy, Kim; Sayres, Rory; Schrouff, Jessica; Seneviratne, Martin; Sequeira, Shannon; Suresh, Harini; Veitch, Victor; Vladymyrov, Max; Wang, Xuezhi; Webster, Kellie; Yadlowsky, Steve; Yun, Taedong; Zhai, Xiaohua; Sculley, D.. Underspecification Presents Challenges for Credibility in Modern Machine Learning. Journal of Machine Learning Research 23(226), 1–61 (2022). https://www.jmlr.org/papers/v23/20-1335.html

[30] Dell'Acqua, Fabrizio; McFowland III, Edward; Mollick, Ethan; Lifshitz, Hila; Kellogg, Katherine C.; Rajendran, Saran; Krayer, Lisa; Candelon, François; Lakhani, Karim R.. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. Organization Science 37(2), 403–423 (2026). https://pubsonline.informs.org/doi/10.1287/orsc.2025.21838

SVSAC - sv-sac.org