前瞻研发布局：AI 分析全球市场与竞品，规划新品矩阵

作者：**SVSAC 硅谷战略顾问委员会**

核心判断

新品矩阵需要有取舍的前瞻判断；AI 更适合扩大信息视野、暴露盲点，不能代替企业决定押注何处。对企业而言，值得优先推进的是：把市场需求、竞品变化、技术可行性和企业能力放在同一组合中比较，以小步验证调整研发优先级。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：全球市场情报；产品组合；生成式人工智能；竞争分析；需求验证；创新选择

SVSAC - sv-sac.org

SVSAC | sv-sac.org

新品矩阵需要有取舍的前瞻判断；AI 更适合扩大信息视野、暴露盲点，不能代替企业决定押注何处。

我们的建议是：把市场需求、竞品变化、技术可行性和企业能力放在同一组合中比较，以小步验证调整研发优先级。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. 核心问题：从更多信息到更好的组合决定

制造企业把新品机会放进组合时，需要同时判断几个并不等价的问题：某地区是否存在可辨认的未满足需求；需求是否由足够多且可触达的买方表达；企业能否在成本、法规、质量与供应约束下交付；新品会不会替代现有产品或争夺同一研发资源；最后，客户是否愿意购买、持续使用并支付足以覆盖全生命周期投入的价格。AI 可以帮助整理外部材料或提出候选概念，但这些输出处在决策链的不同位置，不能用一个“市场洞察准确率”概括。

本篇要回答的不是模型能否生成看似合理的市场报告，而是 AI 介入前瞻研发的哪一段机制可能改善决策，所需证据是什么，何处容易发生过度外推。更精确地说，目标是比较在同一预先界定的产品–市场机会集合中，AI 辅助的信号处理与组合规划政策，相对于既有研究和规划流程，是否增加了可验证需求假设的质量，且在同等预算、研发容量和评审规则下提高最终组合的客户结果与风险调整贡献。这个比较包含多个结果，而不是单一模型分数。

“全球市场”也不能作为一个无边界样本。国家、语言、渠道、法规、产品规格与购买角色共同定义可服务市场。全球网页资料数量增加，可能只是某几个大市场或英文网站被更充分抓取；某类竞品在多个国家出现，也可能源于母公司同一产品的重复页面。若没有清晰的地域与时间框，系统可制造出看似广泛、实际集中于少数公开来源的证据。

“竞品”同样要有操作性边界。直接替代品、功能互补品、客户内部自制、旧产品升级，以及不购买或延迟采购，都可能构成决策反事实。仅比较显性同类商品会忽视客户解决问题的其他方式；反过来，把所有技术相邻产品都纳入竞争集，则会让分析泛化到不能用于产品定位。竞争范围应由具体客户任务、性能约束和替代选择决定，而不是由模型语义相似度独自决定。

SVSAC | sv-sac.org

因此，前瞻分析的目标变量应分层：资料事实是否可核验；需求假设是否在预先指定客户或场景中得到独立支持；方案在技术上是否可行；组合能否在预算、认证、模具、研发人员及渠道资源限制下实施；上市后是否产生采用、留存、贡献毛利及风险调整回报。前几项是必要中间证据，只有后几项接近商业结果。某一环节的改善不自动传递到下一环节。

本次分析提出的核心命题具有条件性：AI 只有在覆盖相关市场的资料可被追溯、信号来源与潜在需求机制相匹配、方案经过独立可行性和客户验证、组合决策显式考虑资源与蚕食，并且实际销售结果可被稳定观察时，才可能提升前瞻规划的决策质量。任何一个条件缺失，都可能使生成速度提高却不改变产品组合的实际价值。相反，如果传统流程能以较低成本完成这些环节，AI 带来的信息规模不一定值得额外成本。

2. 概念框架：信号、假设、方案与组合结果

2.1 市场信号不是需求本身

分析流程首先要区分观测信号与潜在构念。实际订单、重复采购和付费续约反映已发生的交易，但仅代表既有产品、价格、渠道和可达客户。售后咨询、搜索、问卷、评论、经销商访谈、招标文件和竞品页面能补充解释，却各自存在选择机制：搜索者不等于购买者，公开评论通常过度代表有强烈体验的人，招标文档只覆盖进入采购流程的组织，竞品页面则受营销选择与库存影响。

不同来源即便指向相同词语，也可能测量不同对象。例如“便携”在消费者评论里可能指重量，在采购文件里可能指可移动安装，在竞品宣传中可能只是一种定位标签。模型将文本聚类到同一主题能够提高可检索性，却不能抹平用途、测量时点和人群差异。聚类结果应保留原文、发布日期、作者/机构类型、地区、产品规格与引用链，才能判断多个页面是独立证据还是转载同一新闻稿。

专利公开可展示某项技术被申请人主张为发明并进入公开记录，但并不直接测量买方需求、产品可制造性、权利有效性或自由实施空间。WIPO 2026 报告将生成式 AI 专利活动更新至 2025 年，描述已公开专利家族的增长与技术分类，同时指出申请至公开存在时滞。[2] 因而，专利趋势可以成为技术活动和潜在竞争进入的一个线索，不能当作需求预测或市场规模的替身。

建立市场信号账本时，适合把记录定义为“来源–主张–时间–地区–产品/场景–证据等级”关系，而非仅存一段模型摘要。每个数值都应附带分母和口径，例如搜索热度的查询词与地域范围、评论的产品版本、专利家族的统计口径、交易

SVSAC | sv-sac.org

额的币种和时间窗。若来源无法确认发布时间、地理覆盖或版本，正确状态是“时间/范围未知”，不是由生成模型补齐缺失字段。

2.2 需求解释需要反事实和竞争性解释

一个观察到的信号只有放进可反驳的需求假设，才能用于产品决策。假设某地区相关搜索量上升，至少存在几种解释：潜在需求变强；媒体事件促使好奇而非购买；竞争者加大广告；平台调整搜索排序；产品短期缺货使客户寻找替代品；当地节假日改变查询时点。若分析只选取支持新品的解释，市场情报就会退化为对既定方案的论证工具。

同样，竞品销量增长可能来自价格下降、渠道扩张、补贴、供应恢复或长期品牌优势，而非产品功能差异。公开材料通常缺乏真实转化漏斗、退货、客户构成与单位经济性。AI 可以把多来源材料拼成一个连贯叙述，但叙述的连贯性不是识别策略。产品机会的证据强度应取决于观察设计与可比较对象，而非引用数量或摘要长度。

更实用的输出不是“市场需要 X”的结论句，而是一组有时点和边界的预测：若假设成立，某种购买角色应在特定场景选择某功能组合，愿接受某成本/价格区间，并在一定期限出现可观察行为；若数据仍只支持关注或意向，则不升级为已验证需求。竞争性解释应在同一份记录中列出可以推翻该假设的观测，而不是等项目失败后再补充。

2.3 概念生成与筛选是两个不同问题

AI 在构思阶段可能扩大候选方案数量或提高平均表达质量；这不代表它能准确挑出最值得投入的方案。宝洁现场实验将想法生成、选择和进一步展开分成多个步骤，研究人员在每个条件中约束参与者先生成五个初始想法，再由参与者选择其中一个展开。[1] 正式论文报告 AI 条件下平均生成想法质量提高，同时在选择最优想法的准确度上，未使用 AI 的团队约为 50%，AI 条件约为 37%；AI 组最终所选方案质量仍较高，说明生成质量提升和选择判断不是同一效应。[1]

这一证据对组合规划的含义有限但重要：若将 AI 输出量或候选平均评分直接用作投资排序，可能把“提供更多可选项”误当作“识别最优项”。产品组合会在多个阶段发生淘汰，且后续评审依据技术、合规、供应链、客户支付与资源机会成本。每阶段的筛选器应有单独结果记录；模型辅助产生的候选不得仅因表达完整就获得更高先验权重。

人类评审也不是自动正确的金标准。评审可能偏好熟悉品类、短期可行或容易量化的构想，低估新使用场景；不同职能专家也可能对“创新”“可行”“商业潜

SVSAC | sv-sac.org

力”采用不同标准。盲评、独立评分和事前确定的维度能够减少部分偏差，但不能证明获奖概念最终有客户需求。需要用后续客户行为验证评审规则，而非用评审一致性替代市场检验。

2.4 新品矩阵是受约束的组合选择

“产品矩阵”不是一组彼此独立的概念评分。一个方案可能共享平台零件、模具、软件或认证；也可能与现有产品蚕食同一客户、渠道、研发团队和产能。项目的边际价值取决于已选项目和未选替代项目。因此，逐项挑分数最高的 N 项通常不等于组合价值最高，尤其当项目之间共享固定成本、存在相互排斥、或成功概率受共同市场冲击影响时。

组合决策至少应分别列示单项需求证据、技术成熟度、可服务市场、研发与认证资源、客户迁移/蚕食、市场进入时间、潜在回报区间和失败损失。共用平台可降低增量成本，却会形成共同技术失效；地区变体扩大覆盖，亦会增加库存、售后和渠道协调。不能把“AI 检索发现了多个地区的机会”当作扩张组合的理由，而不计算这些依赖关系。

如果不同项目需要同一稀缺实验室或工程师，选择过程还包含时间顺序和排队成本。一个晚六个月上市的方案可能失去季节窗口；一个高预测分数项目也可能挤掉已有明确客户承诺的低不确定性改型。AI 组合排序必须和预算分配、阶段门、退出规则及延迟代价一起评价，不能只展示静态名次。

3. 可核证据与外推边界

3.1 生成式 AI 现场实验测到什么

Dell’Acqua 等在《Organization Science》正式论文报告一项预注册的宝洁现场实验：826 名商业与研发职能员工参加一日线上产品开发工作坊，分析聚焦 791 名随机分配者；个人/团队与是否使用 AI 组成四个条件。参与者在业务单元×地区形成的八个区组内分配；原文称这些为 randomization clusters，不能据此误写为仅把八个整体簇分到处理条件。团队由商业与研发人员配对，团队方案按团队整体分析；AI 工具基于 GPT-4 并通过 Azure 提供，AI 条件参与者接受统一的 AI 使用培训。[1]

评审由商业和技术背景专家独立完成，并对实验条件及提交者信息盲态；主要结果是 1–10 分的方案质量，另评新颖性、可行性、影响等维度。团队无 AI 相对个人无 AI 的方案质量约提高 0.24 个标准差；个人用 AI 和团队用 AI 相对个人

SVSAC | sv-sac.org

无 AI 分别约提高 0.37 与 0.39 个标准差。团队+AI 与团队无 AI 的成对差异在论文报告规格中未显著，因此不能简单写成“AI 让跨职能团队一定更强”。[1]

研究把任务放在早期构思阶段，所用问题来自参与者所在业务单元，结果具有比抽象文本任务更强的组织情境关联；同时，论文也明确指出没有直接观察后续开发或商业化决策。[1] 其结论适用于特定大型消费品公司、受训专业人员、一天工作坊及被盲评的近端创新方案质量。它不估计 AI 分析全球市场与竞品的准确性，不随机化产品上市，不测真实销售、毛利、份额或长期留存，也不证明企业可以取消创意筛选。

对于本次分析，最值得借鉴的是区分过程节点：方案平均质量、是否选出最好方案、展开后的评审质量和商业化结果分别记录。若只观察最终方案得分，不能据此断定人类筛选能力提高；若仅看想法生成质量，也不能断定组合选择更优。由单一企业工作坊外推到不同行业、地区和购买周期时，应把模型版本、任务类型、专家背景与市场验证流程视为边界条件，而不是泛化噪声。

3.2 专利趋势提供技术活动信号，不是需求曲线

WIPO 的 2026《Patent Trends Update in GenAI》基于 EconSight/IFI Claims 在 2026 年 3 月提供的专利数据，更新了 WIPO 2024 报告的趋势分析，观察期延伸到 2025 年。报告以已公开专利家族按最早公开年份统计，称相关家族数量从 2023 年约 1.4 万升到 2025 年约 3.78 万；2024 与 2025 年合计公开超过 5.6 万家族。[2] 该指标的对象是专利公开活动，不是授权有效权利数、产品销量或客户需求。

报告还指出，2025 年国际专利家族只占生成式 AI 专利家族公开量的一小部分，并解释国内申请与公开滞后会影响国际扩展和年份比较；一种多模态专利家族也可能对应多个数据模态，某些家族无法清楚归入单一模态。[2] 因此，排名或技术模式占比会受到申请时点、公开延迟、分类重叠和数据更新影响。把专利排名直接解释为某国将主导产品市场，或某类技术将取得销量优势，是跨越了证据所支持的推断层级。

WIPO 报告称某些专利类别与产品/应用发展方向相一致，这能帮助研究者生成技术搜索假设，但“方向一致”仍不证明因果关系或需求强度。专利申请成本、企业专利战略、学术成果转化、产品开发周期和商业化可能性各不相同。同一公司可能大量申请以覆盖防御性空间，也可能专利家族很少却通过商业秘密或开放组件形成竞争力。专利缺失不等于没有竞争技术，专利增长也不等于用户愿意购买。

SVSAC | sv-sac.org

在产品矩阵分析中，专利信号适合回答“公开技术活动在哪里增加、哪些申请人正在布局”这类有限问题。它不能替代客户订单、真实价格、渠道库存、售后成本和竞争产品采用情况；更不能自动给出自由实施或专利权有效性法律结论。具体专利权利范围和适用法域需要独立法律核查，本篇不对任何企业或技术作此判断。

3.3 两类证据放在同一因果链中的位置

宝洁研究支持一个近端、随机化的判断：在该组织的一日构思任务中，AI 可改变方案生成质量和选择过程。[1] WIPO 报告支持一个描述性判断：公开专利活动能够构成某些技术趋势的观测信号，但其统计时点和分类都有边界。[2] 前者不是市场预测研究，后者不是产品需求研究，两者不能拼成“专利热点经 AI 分析后带来市场成功”的因果链。

更合理的分析是把前瞻规划分成连续但需要独立验证的环节：来源覆盖决定信号是否可见；提取和分类决定原始材料是否被正确表述；需求解释决定信号是否代表购买动机；方案生成将假设转为可测试概念；工程与供应约束决定方案能否交付；组合规则决定稀缺资源投向；市场试验和上市行为才提供需求与商业结果。上游处理更快可能提高覆盖、降低搜索成本，但中间任何环节的误判都可阻断向下游的价值传递。

每一环节的误差也可能方向不对称。漏掉一个小众但高价值细分市场，与把高噪声热度误当成规模需求，后果不同；误把全球重复页面当成多个独立信号，会使信心虚高；将多个竞争产品合并成一个类别，可能隐藏关键替代品；将某地法规或电压规格抽取错误，则会使市场机会在工程上不可实施。前瞻工具应报告错误类型和未覆盖范围，而不只给平均检索准确率。

4. 研究方法与可反驳的分析框架

4.1 证据策略与范围

本篇采用定向范围梳理，优先核读能直接界定核心问题、实验单位、结果变量或专利计量口径的一手资料。现有重点证据包括正式出版的宝洁现场实验与 WIPO 2026 官方专利趋势报告；访问范围及未读部分在来源记录中逐项注明。本篇不是系统综述、元分析或对企业预测系统的评估，也没有开展市场访谈、问卷、CRM 建模、专利数据库再分析或产品上市跟踪。

定向梳理的优点是可以逐项解释近端结果和数据口径，限制是文献覆盖不穷尽，也可能遗漏没有公开、没有正面结果或行业语境不同的研究。本篇不从少数

SVSAC | sv-sac.org

来源估计一般性的 AI 效应大小，而是用真实研究作为机制证据，并标出从其样本与结果到全球新品组合之间的外推距离。作者提出的流程和比较设计都属于待检验建议，不应被阅读为已经实施的实践。

4.2 观察单位与信息冻结

一个可重复的后续实证研究需要先确定目标市场单位，例如“国家×购买角色×产品使用场景×季度”，而不是把网页或专利记录当作独立市场。候选机会的纳入规则、产品边界、竞争替代集合、销售渠道和目标客户规模应在分析前固定。此举不消除定义争议，但使研究者能够知道某个判断是对既定市场回答，还是在看到结果后改变了机会集合。

资料截点也必须冻结。每份来源保存原始 URL、抓取/访问日、页面发布日期、语言、地区、产品版本、引用来源、抽取模型与提示版本。模型生成摘要可以作为检索层，但主张性事实需能回到原始材料核对。若历史页面被修改或删除，快照与引用应保留；否则模型可能把后续发布的产品、价格或竞争者信息泄漏到预测时点之前。

对每条“需求证据”，区分观测事实、潜在构念与作者解释：事实可为某时间窗口的付费订单或可复核的采购文件；潜在构念是未直接观测的需求强度；解释则是对该事实如何对应需求的机制判断。预测模型可以组合指标，但应明确它预测的是搜索、询盘、试用、付费订单还是复购，不能用前者的准确率声称后者已验证。

4.3 分层验证与组合终点

信息质量可先在独立抽样的市场材料上检查：原文事实是否抽取正确、来源与地区是否正确、重复转载是否被合并、无法判定是否会拒答。样本需按语种、来源类型、国家、时间与产品领域分层；只在高资源语言或常见类别上准确，不能代表对低覆盖市场也可靠。验证集要与提示调优和模型选择数据分开，避免同一事实既用于优化又用于证明准确。

需求假设的验证应独立于模型输出。可在预先确定的目标市场和对照方案中，比较 AI 辅助规划与既有规划流程产生的机会预测，看其对后续询价、样品申请、订单转化或复购的概率校准、排序及决策收益。若机会没有进入上市或渠道，结果属于“尚未观察”，不能简单编码为无需求。若仅对被选中开发的概念观察销售，选择进入研发的机制会截断未选机会的反事实，必须同时记录筛选与退出。

SVSAC | sv-sac.org

真正的组合终点应以固定预算或资源约束下的选择政策衡量，而非单项概念评分。比较可在同一产品–市场候选池中冻结技术门槛、人员与预算，记录所选组合、被挤出项目、上市时点、客户采用、毛利、退货、服务成本和组合集中风险。不同项目共享平台或互相蚕食时，需按组合计算；比较单项 ROI 会遗漏协同与共同失败风险。本篇所核材料尚未提供这套整合研究的实施证据，本篇不报告组合收益估计。

4.4 分析价值及其边界

本篇的分析价值是把“AI 分析市场与竞品”从一个笼统功能拆成可被反驳的决策链，并将创意质量、需求证据、工程可行、资源组合及上市结果置于不同估计对象中。宝洁研究提示构思改善和选择判断可沿不同方向变化；WIPO 数据则说明专利公开活动有时间、分类和申请策略边界。[1][2] 这些证据共同支持谨慎区分，却没有验证我们提出的组合规划框架。

对管理实践而言，可操作的最低要求不是增加一份 AI 生成的趋势报告，而是记录每项投资主张的原始证据、目标市场、期限、替代解释、失效条件、投入资源和下一阶段验证结果。若找不到购买角色、愿付价格、技术规格或可观察行为，机会应保留为假设而不是写成需求事实。若市场信号可信但工程不可行，结果是技术约束；若方案被客户接受但单位经济性为负，结果是商业可行性问题；将这几种失败分开，有助于确定是否需要改模型、改产品还是停止投资。

本篇不承诺 AI 能预测未来市场，也不把生成式 AI 目前的能力外推到所有模型、企业或国家。本篇所列研究设计需要实际基线数据、独立评审、客户结果与合适的资源分派，所有阈值须在当地数据和风险承受范围下设定。后续章节将据此进一步分析信号测量、需求解释、竞品比较、组合约束与具体验证方案，持续区分事实、推断和待验证建议。

5. 跨来源信号如何形成可追溯的证据集合

5.1 记录单位必须保留来源与主张的关系

全球市场扫描常把网页、评论、专利和访谈摘要转换为主题清单，再对主题计数或排序。问题在于，一个“主题”并不是一个观测单位：同一份新闻稿可能被数十个网站转载；一个产品页面可能同时代表制造商的规格主张、经销商的本地库存和搜索引擎的索引结果；一项专利家族可能对应多件申请，而家族内部的优先权、公开时间和地域覆盖各不相同。若把每条可见记录等权累加，数据规模就会被发布机制和转载网络支配，而非独立需求证据支配。

SVSAC | sv-sac.org

可审计的记录结构至少包含四类实体：来源页面或数据表、页面中可定位的原子主张、主张所涉及的产品/市场/时间，以及该主张的来源链。记录应明确哪些信息是原文直接写明，哪些是系统抽取或翻译，哪些是分析者推断。比如“某机型在三个国家有销售”不能只保存为一个句子；应保存三个国家页面的 URL、页面更新时间、型号映射、库存/价格状态和引用来源。模型若将两个型号判为同一产品，也应保留判定依据和不确定状态，而不是用实体合并覆盖原始值。

关系图比单纯文本表更能表达重复与依赖。每一条主张可以连接到来源、被引用的上游材料、产品实体、市场与时点。同一新闻稿被八家媒体转引，图中是八个可见页面，但只有一个共同的事实起点；这八项能说明信息传播广度，却不是八次独立的顾客需求表达。相反，三家分销商分别给出同一地区的补货记录，若来自不同组织、不同交易周期并有原始单据，才可能形成更独立的交易证据。是否独立不能仅由域名不同推断。

原始主张和规范化主张也需要分层保存。原始层记录语言、数字、币种、计量单位、发布日期、产品代次和语境；规范层把同义词映射到受控字段，但不覆盖原文。系统应允许将“某类产品需求增长”拆成可核验的小主张：搜索兴趣上升、渠道询价增加、订单增长、交付价格下降等。若只有搜索兴趣这一项成立，摘要就不能把它升级为需求已增长。

5.2 来源依赖性与重复计数

市场扫描通常跨越搜索引擎、媒体、公司网站、社交平台和专利数据库，但“来源数量多”不等于证据来源多。不同网页可能转载同一监管公告、同一访谈或同一公司新闻稿；不同语言版本也可能只是同一内容的本地化改写。对同一事件出现十个 URL，至少有三种不同解释：十次独立观察、一个原始观察的十次传播，或一个原始事实被不同作者重新核验。三种情形对置信度的含义不同。

依赖图可按“事实出处”而非仅按网址聚类。对每条主张，存储直接引用、相似文本、共同作者/机构、链接目标、发布时间和关键数字一致性。自动相似度只生成候选连接；人工核验决定它是转载、衍生报道还是独立测量。来源链不清楚时，默认把记录放在同一潜在证据簇，按一个事实家族处理，并把“独立性未确定”作为数据质量标记。

有一个简单的构造例子说明重复计数的后果。假设某市场扫描找到 30 条页面，其中 18 条重复刊载同一份制造商调查，6 条是引用该调查的报道，另 6 条来自独立渠道咨询记录。若每页计一个信号，汇总为“24/30 支持需求增强”；若按共同源簇处理，24 条调查及转述页面共同指向一个调查源；六条咨询记录须分别核

SVSAC | sv-sac.org

对真实事件与覆盖重叠，若彼此独立可另形成六组，而非自动压成一个咨询来源，且各自测量对象不同。正确答案并非直接压成三个来源后计算统一比率，而是报告调查来源有多少、真实业务记录有多少、它们的客户范围是否重叠，以及是否存在共同的抽样框。

重复信号去重会面临另一类损失：对同一来源的多次引用或不同市场中本地独立结果，简单合并可能擦除地域异质性。比如跨国公司统一新闻稿同时发布多个产品线，但各国经销商的价格和交付条件不同。可以共享“公司发布”这个来源节点，但保留每个市场的本地事实节点；去重的是共同出处，不是所有在语义上相似的市场事实。合并和拆分规则应作为版本化处理规则，保留前后映射，以便复现不同分组对结论的影响。

5.3 数据库存量不等于独立信息量

AI 检索可能把数千条结果压缩成一个可读简报，使人误以为样本规模已经扩大。简报长度与独立信息量没有稳定关系。20 个独立客户事件与 20 个新闻转载的统计含义不同；对某一种语言的大量记录也不能自动替代其他语言或地区的缺失。来源覆盖应至少按目标市场、购买角色、平台、语言、时间窗和来源类型分层，标出有数据、弱数据和根本未观测的格子。

一个实际的监测表可以将“记录数”和“独立证据簇数”分开：国家 A 出现 120 页，归并后可能只有 7 个独立事实簇；国家 B 仅有 20 页，但涉及 5 家不相关经销商的报价与交付反馈。用页面量排序会优先 A，用事实簇和业务接近度排序则可能发现 B 的信息更适合预测付费行为。业务接近度并非一个无争议的分值，应通过来源类型、与决策结果的时间距离、是否包含真实交易/报价和样本覆盖来逐项说明。

报告还要列出观测缺口及可能的选择方向。没有公开评论可能是用户不使用该平台、产品主要经 B2B 采购、当地限制公开发言，也可能是产品尚无足够用户；不应把缺失编码为负需求。平台热度较高的市场可能只是平台覆盖更好；跨国比较若不控制平台渗透和语言检索习惯，比较的可能是数字足迹，而不是市场潜力。

这类架构的作用不是制造精确的“证据分数”，而是把结论依赖哪些材料呈现出来。一个早期方向判断若依赖单一公司来源，和由交易、客户、产品使用三类独立资料共同支持的判断，不应有同样的置信表达。将来源谱系保留到组合阶段，也使事后能够检查：若项目失败，失败源自需求假设错了、资料采样偏了，还是后续交付和资源条件改变了。

SVSAC | sv-sac.org

6. 跨语言与跨市场的可比性

6.1 同一指标名称可能指向不同母体

跨国扫描经常将不同平台的搜索指数、评论量、社交帖子或专利公开数放到一张表中。即使字段都叫“热度”或“市场兴趣”，平台实际母体、归一化分母、关键词分类、时间窗、语言处理和低频截断可能不同。Google Trends 官方说明其数据来自匿名化、分类并聚合的搜索请求样本，按地点和时间范围相对于总搜索量归一化，并按查询范围缩放为 0–100；同一指数并不代表相同绝对搜索量，低频词可能显示为 0。[4] 因此，国家 A 指数 80 与国家 B 指数 40 不能直接解释为 A 的潜在需求是 B 的两倍。

还需区分“词项”和“主题”。词项是用户输入的具体字符串，受拼写、复数、别名、品牌写法和本地术语影响；主题由平台把多个语言表达归到一个概念，范围可能比单一词项更宽。若某地区用一个词询问“工业制冷”，另一地区常用设备型号和应用场景搜索，将同一关键词翻译后比较，会产生系统性的覆盖差异。若改用主题，分类器的跨语种映射又可能把相邻但不同的用途合并。翻译工具的语言质量不能替代概念等价性检验。

比较之前应明确目标构念，例如“某地区对可变速工业冷却设备的采购兴趣”，再将各平台可观测变量映射到构念。每个映射应记录本地候选词、词义边界、品牌/通用词区别、专业缩写、否定与维修查询、地区方言以及同一词的多义性。工业设备页面的维修搜索可能因安装基数大而增长，不表示新品需求上升；“新型控制器”可能是售后替换部件而非新项目。搜索记录需按查询意图或网页内容区分采购前、使用后和售后支持，但分类器误差本身也应抽样检查。

6.2 归一化指数不保留绝对规模

假设平台把某个词在给定时间和地区的搜索量 (q_{g,t})，除以该地区与时段的全部搜索量 (Q_{g,t})，再按选定区间最大比例 (m_g) 缩放至 0–100：指数可近似写成 (I_{g,t}=100(q_{g,t}/Q_{g,t})/m_g)。这是一种简化示意，不是对某平台未公开算法的完整重建。即便 (I_A=80)、(I_B=40)，两地绝对搜索量之比仍无法从这两个数单独推出：A 可能有 100 次与 B 的 25 次，或者 A 有 10,000 次而 B 有 6,250 次。比例、样本抽取、选取的查询窗口及缩放基准都会改变指数。 该示意式仅在 Q 与 m 均大于零时定义；平台因样本不足显示的 0 也不等于真实搜索次数为零。

官方 FAQ 也明确，地区的相似兴趣指数并不意味着相同总搜索量；数据只覆盖抽样搜索，并对低量词、重复搜索和异常活动作过滤，且对低兴趣查询加入的

SVSAC | sv-sac.org

随机噪声更明显。[4] 对小市场或小众产品而言，指数上下跳动可能由采样噪声、隐私阈值或分母变化放大。某词从 10 升到 20，既可能是真正兴趣增加，也可能是平台样本与缩放窗口变化。将百分比增幅当作搜索人次增长，属于口径错误。

若需比较国家，应尽量获得同一平台、稳定词表、相同时间分辨率和可比市场分母，同时保存原始查询设置与每次下载。即便如此，结果首先是“相对搜索份额”或“相对兴趣”的跨期变化，不是买家人数。需要估算绝对市场规模时，应结合独立的母体总量、可触达客户数量或交易数据，并明确模型桥接所需假设。例如，某国搜索渗透率较低，平台查询指数可能低估其真实采购兴趣；又例如一项媒体事件促使非购买者搜索，热度可以上升而潜在付费者不变。

6.3 平台与测量机制变化使时间序列断裂

平台信号的可比性应按实际查询、地域和时间设置判断。Google 官方说明将趋势指数限定为相对数据而非科学民调。[4]我们建议保留查询参数、下载时间与原始返回结果，比较窗口重叠部分；发现差异时先查测量定义，不能直接认定需求变化。Schaer 等仅列为待核线索，[3]其原文未取得，不以其重抽样、窗口拼接或平台变化的具体归属判定任何序列失效。

Liu 对 Google Trends 的案例研究检查了三个关键词在 2022 年数据收集方法调整后的记录，发现部分更新后的历史查询结果出现在术语尚未形成或不符合历史起始时点的时期，或显示出明显偏低数值；作者明确说明这些案例不构成所有关键词和领域均有问题的普遍证明。[5] 这个发现是测量稳定性的反证，不表示所有趋势序列失效。它要求具体应用保存下载日期、关键词类型、地理/时间设置，并对关键峰值做外部时间核验。若趋势数据跨过平台更新断点，应优先在新旧数据重叠窗口检验水平和排序是否一致；若无法获取重叠观测，跨断点变化可能不可识别，不宜机械续接。 此外，作者复查时原先 COVID-19 异常已发生修正，其他词项又出现不足；这些日期差别意味着不能把某次下载的异常当作永久属性，也未证明 2022 更新是唯一原因。

时间窗口拼接会引入第二种尺度问题。平台为了展示长时期趋势，可能分别下载不同子区间并各自归一化。若每段被独立缩放到 100，再把各段直接连起来，就会把每段的局部峰值误当成同一绝对水平。可比拼接必须保留共同重叠时间段或外部参照词，以估计段间缩放；如没有共同标尺，正确做法是展示分段曲线和断点，而不是伪造连续趋势。

SVSAC | sv-sac.org

6.4 语言映射需要测量而非只需翻译

一个跨语种项目可以先为每个市场构造术语表，再由本地行业专家对抽取样本标注意图与对象。例如，分别区分新品采购、品牌比较、规格查询、维修替换、法规认证、求职/学习和新闻关注。若某语言的设备名称与维修用语高度重合，仅靠机器翻译和词面搜索会将售后流量错认成新产品机会。更稳健的检验是对相同的人工确认语义类别，在各语种分别报告分类精度、漏检比例和未知比例，而非只报告总体准确率。

作者提出的候选抽样方案应覆盖各目标语种与来源类型，且包括常见词、低频词、多义词和型号/品牌变体。抽样单位是独立查询意图或资料事件，不是对同一条内容调用不同模型的重复结果。对含糊材料允许“不足以判断”标签，并把误分类成本按具体决策区分：误将售后搜索当新品采购会推高市场评分；把真实采购意向判为维修查询会错失机会。两种错误不能在总体准确率中抵消。

该分类验证仍未解决更深层的外推问题。即使某地区的公开讨论被准确分为“购买前搜索”，其使用者也未必代表所有采购人，尤其工业采购常由工程、采购、财务和现场运营共同决定。平台上更活跃的工程师可能提供技术需求，却不能单独说明预算持有者愿意购买。跨语种测量必须和购买角色、渠道覆盖及实际交易阶段相连，才能从可比词汇走向可比需求。

7. 时间、生命周期与预测信息的可用性

7.1 早期搜索的含义随产品生命周期变化

产品相关搜索的含义可能随使用阶段改变。上市前可以涉及概念探索、竞品或预售信息，安装后可以涉及操作、耗材或保修。这里列举的是待检验的解释，不是已核消费者行为分布。企业应将搜索与可定位的询问内容、渠道和交易状态对应，再判断其预测意义；未取得的 Schaer 等原文不能为具体案例、比较模型或生命周期结果提供已证依据。

这不是说公开信号无用，而是说明预测关系取决于问题和决策窗口。若研发组合在概念阶段要决定是否投入两年的模具、认证和工程资源，未来数周的搜索峰值即使预测短期关注，也不能回答长期支付、续购和售后负担。反之，某成熟产品的实时查询或询价可能有助于短期补货，但不代表可以用于未来数年的产品组合战略。信号的预测领先期需长于决策所需提前期，并且在业务结果发生前可获得；否则即便统计拟合好，也没有决策价值。

SVSAC | sv-sac.org

在生命周期分析里，不能仅根据发布后观察到的销售曲线切分“早期、中期、成熟期”，再在同一数据上寻找效果最大的阶段。阶段定义可以基于预先约定的周/月数、累计销量比例、渠道铺货阶段或技术更新节点，但不同定义反映不同机制。若按结果变量的峰值位置事后切段，结论会受到选择性报告影响。合理做法是提前定义多个有意义的阶段，分别报告样本数和预测误差，并用新产品或未来时间窗确认阶段异质性。

7.2 线索何时可用于决定，而非只用于解释

预测的价值不只取决于平均误差，还取决于误差发生时是否还来得及行动。假设采购预测在产品定型前需要 18 个月，而外部搜索指标只比订单领先两周；该信号可能很适合解释订单变化，却无法改变产品开发排程。相反，行业招标、标准草案、经销商设备安装规划若能提前多季观察，尽管样本少，也可能对长周期决策更有用。

应把预测窗口与反应窗口放在同一时间线上：资料发布日期、平台采集时点、模型计算/人工核验耗时、决策委员会日程、设计冻结、认证申请、供应商锁定、首次可供货以及客户采购周期。对每个环节，问模型能否在下一项成本承诺前产生有效证据。不能及时进入决策的指标不会直接形成提前量；不能在产品上市后快速更新的长周期模型也可能被新竞争条件过时。

预测变量可能仅同步反映需求，而非提供决策前信息。若销售已增长后才出现教程查询，以同周期查询拟合销售不能证明上市前预测能力。本篇要求逐个预测原点记录实际可见信息；不能用最终整季评论数解释更早的上市决策后，称为事前预测。这是本篇的信息时点要求，不作为未读研究的已核发现。

7.3 一个预测指标可能因为决策动作本身而改变

公开信号并不总是外生的。企业看到竞争对手页面或搜索热度后加大广告、提高投放、改变渠道价格；这些行动又增加品牌搜索、评论和媒体报道。于是“AI 识别到市场热度–企业行动–热度进一步上升”的关系会形成反馈。若模型用处理后的搜索量预测销售，再把预测准确归因于 AI 洞察，就会把企业自己的营销反应混进预测信号中。

类似地，若公司把模型标记为高潜力的产品提前铺货，观察到更高销量，并不独立证明原始需求预测准确；供给和曝光本身改变了销量。对历史数据而言，销量、库存、价格、推广和渠道覆盖相互作用，单纯相关系数不能分解需求强度。预测试验与因果试验的目标不同：前者比较可用信息集下的预测误差，后者

SVSAC | sv-sac.org

比较采取某种规划或投入政策后的潜在结果。除非设计能够识别行动效应，模型的高预测准确不能写成投资政策提高销售。

7.4 预测更新与模型/数据版本

每次分析必须保存数据抽取日期、历史窗口、地域/语言设置、关键词或主题清单、模型版本、提示词、特征处理、预测软件包/参数和输出时间。对于实时平台数据，记录“当时能看到什么”比事后重新下载更重要。若历史平台会修订，旧数据视图可能无法复现；应保存授权范围内的观测快照、文件校验和、数据字典与下载日志。没有快照时，对过去决策的回放要标记为基于修订数据，而非声称完全按当时信息复现。

模型更新需要设置两种比较：滚动更新与固定版本回放。固定版本可以比较数据/市场环境变化；滚动版本反映实际运营中的更新政策。若只报告每次最新模型，无法分辨模型改进、数据修订和市场结构变更。关键指标变化时，应分别检查特征分布、平台采集方式、查询词覆盖、客户构成、价格/供给和结果定义，而不是默认“需求拐点”。

8. 需求校准与可复现预测检验

8.1 预测目标必须对应可观察的业务事件

“预测市场需求”常把多个不同目标压成一个数字。目标可能是未来三个月的搜索量、报价询问数、首批订单、年度装机量、客户复购，或新品组合总贡献。对供应链，上市早期订单常受初始铺货和渠道压货影响；对工业设备，询价阶段与项目预算审批相差数月；对耐用品，出货与最终使用、退换货和服务成本也可能脱节。研究设计需在预测发出之前固定目标事件、地域、客户群、产品版本、窗口和计量单位。

对每个候选机会，建立最小化的“预测登记表”：预测生成时间；可用资料截止点；预测对象和总体定义；预测的量或概率；不确定区间；具体行动阈值；采用或不采用该预测后要观察的结果；结果成熟日期；以及未能观察到的原因。用“高潜力”而不设概率/量纲/时间窗，不能比较预测与真实结果。若只报告排序，也需说明排序目标是购买概率、数量、贡献毛利还是研发成功概率。

初期可在同一市场范围和信息截止下比较历史销量或询价趋势、领域专家原有预测以及公开信号模型。我们建议同时保留简单可复制的统计基准和企业当前流程，并记录误差、成本及可执行提前期。若只选择明显薄弱对手，差异不能说明相对现实替代方案有增益；这一比较要求不依赖未取得文献的具体案例结果。

SVSAC | sv-sac.org

8.2 评估须模拟实际预测时点

时间序列评估应采用按时间向前滚动的预测原点。每个原点仅使用当时已发布日期之前的数据，先对全部候选方法生成预测，再揭示后续结果。随机打乱记录作训练/测试拆分会把未来时间信息泄漏到训练集，尤其当相邻周的搜索数据、营销活动和产品生命周期高度相关。若跨国家验证，还要检验在一个市场训练、另一个市场测试时是否仍有可用性；但跨市场结果不是随机实验，因为市场差异还包括品类、渠道、语言和竞争条件。

除了总体平均误差，至少需报告误差分布、系统偏差、不同预测期和市场的结果，以及在决策相关阈值附近的误差。平均绝对百分比误差在实际需求为零或接近零时会失真；均方误差会对少数峰值过度加权；绝对误差未计入缺货和库存积压成本的非对称性。组合规划需要用与误判相对应的损失函数，例如过高预估导致模具和库存浪费，低估导致错过设计窗口或延迟进入。损失权重须来自企业可核实的决策成本，而不能由模型供应商任意设置。

在历史回测中，只选择那些最终成功上市且有销售数据的新品会形成幸存者偏差：失败、取消、延期或从未上市的机会被排除，模型看起来只需预测成功产品中的销量起伏。一个可审计的机会总体应包含每个预定评估点的在研项目，记录暂停/取消、供给限制和是否有客户/订单标签。未上市产品在现行政策下的已实现销售可以是零，但若问其上市后的潜在销量，该反事实未被观察；可对其成本和技术结果做分析，但不能据此直接观测若上市后的需求。

8.3 基于明确假设的校准算例

以下为纯构造的教学算例，不来自企业数据，也不代表模型实测。设研发委员会在某时点把 20 个候选机会分别预测为“未来两年会形成至少一笔付费订单”的概率。为简化，先将 20 个候选按预测概率分成两组：10 项预测 0.8，10 项预测 0.4。之后所有机会均有固定两年观察窗，结果 (Y=1) 表示至少形成一笔已确认付款订单，(Y=0) 表示该窗口内没有付款订单；延期但仍开放的采购项目仍按窗口定义记录，并不随意剔除。

假设第一组实际有 6 项在窗口内付款，第二组有 5 项付款。高概率组的观察发生率是 6/10=0.60，预测均值是 0.80，校准误差为实际减预测，即−0.20；低概率组发生率为 5/10=0.50，预测均值 0.40，误差为+0.10。全体预测概率均值为 (10×0.8+10×0.4)/20=0.60，观察订单率为 11/20=0.55，整体差为−0.05。若只报总体差，两个方向相反的分层误差部分抵消，掩盖了高潜力层过度乐观与低潜力层偏保守。

SVSAC | sv-sac.org

这个简单表也不能证明模型完全失准。每组只有 10 项，实际频率有抽样不确定性；某些候选可能共享行业冲击或同一渠道，所以独立二项假设不一定成立。还必须检查分层结果是否由某一地区或单一产品类别驱动，预测是否在分组前被事后切段，以及概率是否由外部模型直接产生还是委员会人工改过。公开信号模型与传统流程须在同一机会集、相同标签成熟条件下比较，不能让 AI 组独享更完整的后续数据。

若实际决策不是每项都同等投入，还应先冻结政策，例如只将前 5 项送入客户验证，或仅在预计贡献超过资源成本时开模。概率校准与政策收益不同：一个概率准确但对低价值机会排序错误的模型，可能不增加组合贡献；一个总体校准误差较大但能可靠识别某类可避免损失的工具，也可能在有限资源下产生价值。需要对政策选择、实际动作和被挤出的机会进行单独观察。

8.4 多重检索与发现后验证

AI 可以在海量术语、国家、产品与模型设置中迅速搜寻关联。若检索了数千个词，再挑出与预期方向一致的十个，就会把偶然波动当作稳定需求信号。类似问题也会出现在人工选词、选择比较市场、窗口长度和指标定义中。每一次试探都扩大可挑选空间；只报告最后看起来有效的查询，会导致过度乐观的预测评价。

探索阶段可由 AI 提出候选词、来源和市场，由分析人员记录检索空间与淘汰理由。确认阶段冻结词表、产品边界、结果指标、期限和模型版本，在未见时期或市场检验；若反复查看并据此调参，应另设未见样本。这是本篇控制选择性比较的设计安排，不是对 Schaer 等未读研究的方法与结果转述。

对市场机会发现而言，确认工作还需包含独立的目标客户或商业记录，而非只换一个搜索平台重复相同平台偏差。若一个热门词在多个搜索引擎中同时增长，仍可能由全球新闻引发而非购买需求。一个可检验的确认预期应指定后续不同类型的观察，例如询价、样品申请、预算授权或付费订单。若搜索继续上升而这些业务事件未增加，便构成对“搜索指数代表需求”的反证，而不是事后改写成功定义。

本章小结。 跨市场情报首先是测量问题：来源必须去重但不抹除独立地域事实；相对热度必须和母体、平台机制及时间窗一起解释；预测需要在实际决策期限内与强基线比较。AI 能降低资料整理成本的假设仍需检验，且高覆盖与高准确并非同一指标。本篇到此只提出可重复检验的测量和预测框架，没有对任何真实企业市场或新品组合做过预测。

SVSAC | sv-sac.org

9. 竞品边界取决于客户可行替代，而非名称相似

9.1 从产品目录转向决策中的可替代项

公开竞争情报常依据产品名称、专利分类、网页描述或功能标签寻找“相似竞品”。这些方法适合扩展检索清单，却不足以定义某项产品机会真实面临的竞争。客户面对新产品时，可能购买直接竞品、延续现有产品、改造内部流程、外包任务、采购互补设备，或什么也不做。若模型只返回同类产品品牌，便漏掉真正能使买方拒绝新品的替代安排；若将所有功能相邻的产品都算竞品，又会把市场边界扩张到无法指导研发资源配置。

竞品集合应以客户任务和约束为起点，而非仅以类别标签为起点。可采用三层搜索框：第一层是对同一任务直接可替换的产品；第二层是在客户改变流程、接口或采购方式后可以替代的方案；第三层是不采购、延后、修复现有资产或内部自制等“外部选择”。每层内部再识别价格、产能、合规、切换成本、交付时间、服务网络和技术接口等限制。产品在技术上类似，不必然在采购时点可替换；反之，技术结构迥异的方案可能争夺同一预算和使用任务。

替代关系最好由可观察的选择或业务行为确认。客户在询价中把两款产品放入同一招标、产品 A 涨价时部分客户转向 B、停产后用户改用内部流程，都是不同强度的替代证据。工程描述和市场文案只能提出假设。若没有真实选择数据，可通过访谈、历史采购、功能测试或设计好的选择实验识别可接受集合，但结果仍是对应方法和样本范围内的证据。

9.2 选择集合不完整会改变估计对象

产品选择分析应明示纳入哪些替代品，以及展示顺序和默认设置。Huber 与 Puto 文献保留为待核线索，[7]本篇未取得实验原段，不确认其选项设计、份额转移或吸引效应。作为条件分析，即使原有产品销量不变，增加或删除计入集合的产品也会改变份额分母；因此集合定义变化与真实购买转移必须分别核对，不能把两者混为同一市场增长。

假设一个简化的选择实验只展示两种设备，受访者必须选 A 或 B；若再加入“低价、性能边界、维护要求高”的 C，选项可能改变对 A/B 的比较标尺。若模型在第一轮把 C 归为“非同类”，下一轮又将其纳入同一竞争集，预测份额不能直接并列。分母从“在三种被展示方案中选择”变成“在两种方案中选择”是研究设计改变，不是需求发生变化。

对一个具体机会，竞品边界可通过三套集合做敏感性比较：窄集合（直接功能替代）；核心集合（直接替代加客户常用替代流程）；宽集合（再加潜在新技

SVSAC | sv-sac.org

术、外包与不采购）。每套集合下，分别估计方案的条件选择概率、对既有产品蚕食比例、对外部选择的净新增比例，并展示排序是否变动。如果某个概念只有在窄集合下排第一，在纳入当前客户内部改造和现有资产延寿后便降到末位，原先的“新品机会”判断依赖的不是模型预测精度，而是被省略的替代品。

9.3 技术接近、商业替代与竞争响应要分开

技术相似度可以用于检索专利、规格和潜在研发路线，但它和客户可替代性并不同义。两个产品即使核心技术相近，也可能分别服务于不同工况、认证和安装基础；一种软件与一种硬件技术上不同，却可能共同替代某个高成本人工流程。技术分类还可能受专利申请人策略、数据库分类和历史存量影响，不能直接当作市场竞争边界。[2]

竞品也会响应市场信号。AI 发现某一地区的增长词项并建议新产品后，现有厂商可能降价、推出改款、扩大经销库存或锁定服务合同；客户的可替代集合随之变化。分析若把当前网页产品集固定为未来竞争格局，会忽视策略性反应。合理的前瞻输出应提出条件式预测：“若竞品保持当前价格与服务条件，方案 X 在这一客户群中可能获取某类订单”；同时列出对手可能的响应以及改变结论的条件。

替代关系还可以不对称。大型客户可能由现有系统切换至新产品，而小客户由于安装和维护能力不足，实际只能选择服务商方案。跨国公司采购可能能承受昂贵的认证和集成成本，本地小型经销商则不能。以一个全球平均“替代率”描述所有账户，会遮蔽规模、设备基础、维修能力和服务可达性的差异。组合设计应在可服务客户集合中区分可替代方案和不可服务需求，而非把所有有意向的人合并为一个市场。

10. 购买角色、组织决策和客户证据

10.1 企业账户不是单一买家

在工业与企业市场，产品使用者、技术评估人、预算持有人、采购部门、信息安全/法规人员和业务负责人可能分别参与决策。一个公开招标或网站点击代表不了整个采购组织的偏好。采购阶段也可能改变角色权重：现场工程师先定义技术要求，财务决定总拥有成本，采购与法务协商合同，管理层签署资本预算。AI 若把多位联系人、同一公司不同地区子公司或经销商账户混为一个“客户声音”，可能重复计数或错误认定购买意向。

SVSAC | sv-sac.org

Thomas 的探索性工业市场研究以购买中心的采购责任为细分依据，报告组织间存在“责任重叠”与“职位主导”两种不同模式。[8] 其证据不代表所有现代制造行业，更不是关于 AI 或订单结果的因果研究；但它支持一个有限方法点：账户层面购买结构可能不同，单一联系人标签不能自动代表整体决策结构。购买中心研究也会受到关键知情人报告误差、组织职位变化和具体采购情境影响，产品开发研究应记录资料来自谁、其实际决策责任及所处阶段。

客户语料可按角色而不是只按公司聚合。每条询价、试用反馈或抱怨应关联职位/职责、采购项目、产品版本、渠道、时间和是否代表真实采购过程。一个联系人可能同时承担工程和预算职能；一个职位名称也可能在不同国家企业里含义不同。模型抽取职位时，最好输出候选责任和置信依据，人工复核影响投资决策的高风险标注，并允许“未知”。不能因为一个账号在社交平台自称“采购经理”，就视为预算授权人。

10.2 意愿、技术认可与付款意向不是一个终点

访谈中说“有兴趣”、问卷中给高分、技术团队认可样机、经销商愿意上架、客户签意向书和实际付款，彼此相关但不能互换。访谈反映被访者在特定问题和情境中的表达；技术试验回答性能是否达标；报价接受反映在限定条件下的价格反应；付费订单才是具体约束下的购买行为之一。即便是付款，若包含返利、补贴、内部关联交易或渠道备货，也未必代表最终使用者实际采用。

可用一条客户行为阶梯记录证据成熟度，但不把层级压成线性“转化分数”：初次表达问题、确认采购场景、提交规格、申请样品/试用、接受报价、完成预算、签订单、付款交货、持续使用/复购。客户可以跳过其中某些步骤，也可在技术验证后因预算周期取消。不同阶段的流失率需要明确分母和观察窗口。无后续记录的项目是未观察或失联，不应默认视为拒绝，也不宜按成功计入。

若希望从早期概念识别付款可能性，需要检验前置信号是否在独立项目上预测更成熟的客户行为。可以把历史项目按发出预测的时间点冻结，比较低成本信号模型、销售团队原有记录与技术评审结果的增量预测力。项目最终状态至少包括付费订单、关闭且不可恢复的输单、延期开放、未完成跟进和未上市。在完整固定窗口内确认未付款可按 254 段定义记 Y=0，但不能解释为永久拒绝；观察窗未成熟或付款状态缺失则记未观察。将延期一概视为永久失败会低估长周期采购；仅分析最终付款项目又会产生选择偏差。

SVSAC | sv-sac.org

10.3 价格和采购条件构成需求命题的一部分

“客户需要某功能”若不包含价格、部署、维护、交货期、培训和服务边界，往往只是技术偏好，不是可执行的市场机会。相同功能组合在低价、现货和本地维护下可能有客户，在高价、长交期和跨国售后条件下则没有。新品矩阵判断需将客户所需性能和愿付条件同时列出；技术规格对齐不能替代单位经济性和合同可履行性。

如果使用选择实验，应把备选方案属性限制在真实可实现范围内，并让价格、服务和交货期与当地采购条件一致。实验题中的随机化只支撑被随机展示属性的边际偏好，不自动估计真实市场份额。受访者知道这是研究情境，可能愿意选择价格较高的环保、耐用或智能功能，却不会在预算受限时购买。实验结果应与实际报价、采购、试用或订单数据做外部核对。

对 B2B 采购，决策可能是团队协商而非个体加总。工程、采购、财务不同角色的意见不必平均；有些技术条件是硬门槛，不能用多人高满意度抵消。若其中一位安全负责人否决，产品没有进入选择集；若预算持有人不认可收益，即使使用者偏好很强也不成交。记录角色间的分歧和决策权，可以揭示功能要求和实际采购瓶颈，避免把所有表态平均成一个虚构的“客户偏好”。

11. 地区细分要联系可服务性和执行机制

11.1 国家层级可以提示差异，却不能替代客户层级

跨国差异可能来自语言、收入、法规、气候、基础设施、维修网络、渠道或偏好，不能一律归为固定文化属性。Falk 等文献原文未取得，仅为待核线索；[6] 本篇不确认其样本规模、调查验证、偏好类别及国内外差异比较。这里的方法要求是先明确客户单位和可观测商业条件，再检验国家内部异质性；国家标签本身不能证明具体客户的产品需求。

市场分层应先区分哪些变量是客户层异质性、哪些是国家/制度层约束。例如，电压、认证、气候和进口税可能在国家/地区层面共同作用；安装能力、预算周期、设备年龄和数字成熟度可能在客户组织层面变化；不同用途和负载工况可能发生在同一国家。模型可以识别聚类，但聚类标签只有在可解释、稳定且能帮助采取具体研发或服务动作时才有决策意义。若“高潜力市场”聚类只因数据量大或英语材料充足，便不代表可开发性更好。

SVSAC | sv-sac.org

11.2 可服务市场是需求、渠道和交付条件的交集

总潜在需求不能直接换算为企业可以获取的销量。可服务机会至少需要客户确有问题、产品满足当地可用规格、渠道可达、售后/备件能交付、认证周期可接受且价格在预算内。某地区搜索热度高但进口许可或现场支持条件不满足，短期仍非可服务市场；另一个线上信号较弱的地区若有确定经销网络和重复订单，可能更实际。

为避免把市场规模和进入能力混为一谈，可用漏斗分开：目标客户总体、满足问题/场景条件者、技术与法规可服务者、渠道可触达者、报价可接受者、在观察窗内实际购买者。每一层要给数据来源与分母；无法观察时用区间或明确未知，不能用模型补成点估计。若漏斗一层筛选排除小型组织，后续转化率只适用于被筛选出的组织，不能外推到全体客户。

一个构造的例子：某企业在地区 A 估计有 1,000 家潜在工厂，其中约 400 家工况符合产品应用；但只有 160 家能在现有认证和电气规格下部署，80 家经已有渠道可触达，20 家进入有效报价，8 家在 12 个月观察窗内采购。地区 B 在检索材料中的相关讨论较少，却有 150 家适配工况、120 家可部署、90 家可触达、25 家报价和 10 家采购。以上全是示意数字。若只按市场总体或在线热度排序，会选 A；若只看 12 个月订单数会选 B，但两地总体数量、利润、样本误差、渠道投资及后续成熟时间都不同，不能仅由 8 对 10 下结论。

合适的比较要明确目标究竟是近期入市订单、长期毛利、产品学习价值还是对现有渠道的协同。地区 A 可能值得投入认证，因为可服务客户池较大；地区 B 可能更快验证特定性能，但规模有限。若两地共享同一研发/售后团队，它们还互相争夺资源。把不同目标压成一个“市场吸引力分”会隐藏管理层的权重选择；应该展示权重敏感性和阈值依据，而不是声称模型客观选出唯一市场。

11.3 地域标签和文化解释应允许反例

如果 AI 从公开材料中发现“某国客户偏好低价/耐用/数字功能”，下一步应检查该关联是否由行业结构、收入水平、产品供给、当地采购规则或样本来源解释。城市和农村、出口工厂与内需工厂、新建产线与老旧设备的差异可能大于国家均值。国家层调查的偏好指标也不能直接应用到某一个企业账户，更不能代替当地行为数据。

可反驳的地域假设应明确其层级和反例。例如，假设某市场因维护服务稀缺而偏好更易维修的模块化方案，那么数据应显示维修网络覆盖弱的客户更愿接受可更换模块；若在维修覆盖充分的客户中差异消失，支持的是服务基础设施机

SVSAC | sv-sac.org

制，不是“国家文化”。若偏好在同国不同行业方向相反，就应分层而非保留平均结论。

12. 竞争与需求分析的跨市场验证

12.1 外推不是复制同一套模型阈值

在一个地区有效的关键词、竞品分类或需求概率模型，不会自动迁移到其他地区。平台覆盖、搜索语言、产品法规、销售渠道、经销商角色、价格币种、付款习惯、客户构成和竞争成熟度都可能变化。验证迁移时，模型所需的不只是目标地一组新文本，而是与决策有关的标签：询价是否变为项目、工程确认是否进入预算、经销商承诺是否转订单、试用是否持续使用。

迁移评估可分三层。第一层是表示层：事实抽取、术语映射、产品/组织实体识别和来源重复结构在目标地区是否仍可靠。第二层是排序/概率层：对可比的客户机会，原分数是否仍能区分不同成熟度及其实际比例是否校准。第三层是决策层：在当地的价格、服务、认证和渠道条件下，同一产品组合政策是否改善收益或减少错误投资。只有表示层通过，不表示销量预测已迁移；排序层通过，也不表示投资政策对目标市场有因果收益。

如果目标地区缺少足够结果标签，不能把在训练市场的高分当作充分证据。应先报告适用范围和数据不足，允许模型拒答或只提供检索支持，不输出确定销量。可以构造一个有限试验：在若干业务单元或客户集群中随机分派传统情报包与 AI 辅助情报包，统一市场问题、信息截点和专家资源，再盲评事实准确性、候选机会可检验性，并追踪后续客户验证和项目决定。若客户/渠道之间互相转介绍或销售人员共享信息，分派单位应提升到团队/地区集群，并记录污染。

12.2 需求差异和渠道差异可能混淆

某地新品订单少，原因可能是需求弱，也可能是未完成认证、缺货、销售代表未跟进、经销商不愿压库存或报价付款条件不合适。订单多也不一定证明终端需求强，可能是渠道备货或促销。将“市场信号”与结果相连时，至少记录供给可得性、价格/促销、渠道覆盖、报价时点、有效跟进、产品退货和最终使用。否则在供给受限时，销售表现无法代表未满足需求的准确量；在渠道压货时，出货量又会高估终端采用。

同一公司不同地区的渠道模式可能不同。直销订单、经销商库存、平台出货和最终客户使用是不同事件。若一地以直销为主、另一地以分销为主，简单比较“销售额”会比较渠道结构而不是顾客需求。产品组合需明确关键指标是 sell-

SVSAC | sv-sac.org

in、sell-through、付费使用还是复购，并把渠道间转售、退货和价格补贴纳入核对。

12.3 用一个比较矩阵揭示边界条件

下表为作者提出的研究矩阵，并非已执行的市场研究。矩阵将地区差异与购买/交付机制连接，避免用单一国家平均标签作结论。

可检验维度	地区/账户层数据	对决策的含义	反证或替代解释
搜索/公开讨论	本地术语、平台覆盖、采集时点和词项/主题类型	仅作为需求或关注的候选信号	新闻、售后、平台覆盖差异或数据修订造成峰值
直接替代选择	报价同台、历史切换、内部自制/延寿方案	定义客户决策中的竞争集合	选项清单/展示方式改变选择份额；切换成本使理论替代不可执行
购买角色	使用者、工程、财务、采购、授权者及其项目阶段	区分技术接受与预算/合同可行性	单一受访者不代表团队；关键否决者未纳入
可服务条件	规格、认证、交付、维修、渠道、币种/付款条款	把潜在需求映射为可达机会	订单差异来自供给/跟进而非需求差异
客户结果	报价、付费、交付、使用、退货、复购及窗口成熟度	对应明确的购买/使用估计对象	未成熟机会或渠道库存被误作失败/成功

若矩阵中的关键单元缺失，结论应限定在已经观测到的阶段。比如公开讨论与报价请求可支持“注意或兴趣有所增加”的判断，但没有付款、使用和服务成本就不能支持“商业需求和毛利已验证”。对跨市场决策来说，拒绝过度外推是研究结论的一部分，而不是分析失败。

SVSAC | sv-sac.org

12.4 解释全球差异时同时检验三类机制

全球市场分析容易把地区间差异归因于文化，或把相同产品卖点视为普遍偏好。更可检验的解释至少分三类：客户偏好不同；产品/渠道可供给条件不同；被观察到的信号和标签质量不同。三者可能同时发生。若某地区的线上检索很少，可能是需求弱，也可能是买家依赖经销商；若该地购买率低，可能是功能不适配，也可能是认证和付款渠道阻碍。

在资料可得时，可用同一类客户、同一场景、相近产品版本和共同采购窗口进行分层比较，再加入平台与供给差异的分解。若不能匹配，报告“地区差异包含市场组成与执行条件差异”，而不将其称作偏好差异。不同解释对应不同研发动作：功能差异需要本地化规格；服务网络缺口需要渠道/维修投资；资料偏差需要修正采样和测量；若三者不分，组合决策可能用错资源。

这些检验同样面临选择和外部效度限制。公开记录和活跃客户通常不是目标市场的随机样本；客户愿意参加试用者可能更需要新技术、更有预算或更愿意合作。愿意受访的渠道商也可能希望厂商扩大产品线。对客户实验，应记录纳入/拒绝、失访和参与选择；对历史回测，应显示未进入观察的人群及项目状态。研究结果只能推广到目标样本覆盖且机制相近的市场，不可凭“全球”一词推断所有国家。

13. 概念测试测到的是反应，不是尚未发生的市场

13.1 先问受访者看到了什么

新品概念通常不是一个单一刺激。文字说明、渲染图、短视频、可交互原型和可实际使用的样机，分别向受访者提供不同的信息。若 AI 先把市场文本总结为“客户重视快速维护”，再据此自动生成产品概念图，测试得分提高并不能单独说明客户更需要该产品：图像可能更醒目，文案可能更具体，或生成方案将性能承诺写得更清楚。研究要拆开内容改变与呈现改变，否则把“AI 生成”作为处理变量会混合多个机制。

概念测试至少有三种估计对象。第一是理解：受访者是否正确理解目标客户、功能边界和使用条件。第二是评价：在给定信息下，受访者表达的兴趣、偏好、购买意向或愿付价。第三是行为：受访者是否在真实价格、时间、预算和替代品条件下采取了可观察行动。第一项是信息质量，第二项是情境反应，第三项才更接近购买行为；即使第三项也需要区分预约、付订金、正式订单、交付和持续使用。若问卷只测第二项，研究结论就应停在第二项。

SVSAC | sv-sac.org

刺激的可信度可能系统性地改变评价。Durisin 的公开研究稿对三个概念测试研究进行探索性分析，比较概念板、视觉或视频、原型观察和原型试用。研究报告原型试用组的购买意向和愿付价格评价较高；但主要研究集中于意大利消费者、食品和家用咖啡设备等具体概念，样本规模有限，且作者明确指出没有观察实际购买。一个新奇产品可能需要实物试用，才让用户理解其功能；另一个依赖不可见性能、需要长期可靠性才能体现价值的产品，即便一次短试也不能验证核心承诺。[9] 因而，不宜把该研究的方向性结果翻译成“原型测试普遍提升预测准确率”。它更直接支持的是：测试媒介会改变测得的反应，媒介本身必须纳入研究设计。

13.2 生成内容和需求内容的因素拆分

一种可复核的比较不是简单地让模型组与人工组各写一段，再问哪段“更有吸引力”。至少可将处理拆为信息来源与呈现方式两个因素。信息来源可以是原有市场研究摘要或 AI 辅助整理的同一批证据；呈现方式可以是文本概念或图文/可交互原型。由此得到四种组合：人工整理–文本、AI 整理–文本、人工整理–图文、AI 整理–图文。每种刺激都需遵循同一事实约束，标示尚未验证的性能假设，并随机分配给预先定义的目标用户。若一个人看多个概念，应随机顺序，并把重复评价视作同一受访者内的相关结果，而非独立观测。

四格设计能检验“AI 整理的市场证据是否提高概念信息质量”与“视觉呈现是否提高兴趣”是否混淆。若 AI 组的理解正确率提高而购买意向不变，支持的是信息理解机制，不是需求增加。若视觉组的意向提高，但在提供真实价格、替代品和限制条件后差异消失，则更符合展示效果或新奇效应。若 AI 图文组远高于其他组，却没有 AI 文本的增量，这一交互结果提示生成视觉可能贡献主要差异；它仍不说明视觉内容代表真实可制造产品，也不说明后续销量。

在每项评价中同步测量事实理解、可信度、预期性能、采用条件、价格反应和当前替代方案，比把所有反馈揉成“好感分”更有诊断价值。例如受访者认为产品有吸引力，却不相信其低能耗承诺；或认为功能可信，却因维修网络不足而拒绝采用。两者要求不同的研发动作。模型将反馈归纳为一个总情绪标签，可能掩盖这种结构，尤其当文本中有讽刺、条件句、对竞争品的比较或角色分歧。

13.3 概念筛选会引入赢家偏差

对数十个 AI 生成概念排序，再只测试最高分的几个，测试结果不能无条件代表整个创意池。假设每个概念的真实平均反应相同，有限样本噪声仍会使某些概念偶然排在前面；若随后只报告这些赢家，它们在复测时平均分通常会向总体

SVSAC | sv-sac.org

均值回落。这是选择过程制造的乐观偏差，不需要模型存在故意过拟合。概念生成数量越多、筛选轮次越多，越可能挑出偶然高分者；若根据同一批评价不断改提示词并再挑赢家，偏差会继续累积。

可执行的防护不是只说“做盲评”，而是把候选产生、筛选和确认分成不同数据用途。开发集用于调整提示和评分规则；筛选集用于缩小候选；确认集仅在候选、价格、目标用户和主要结果冻结后打开。若确认测试失败，不能重新使用同一确认样本修改概念后再次宣称验证通过。跨产品线或跨国家时，确认单位还应是新的产品任务或市场簇，而不是把同一受访者多次评分当成多次独立确认。

选择效应也改变商业解释。筛选后赢家代表“按现有规则和评审群体被选中”的方案，不等于在所有可能创意里最有市场价值的方案。若评审者更熟悉渐进创新，突破式概念可能因信息不足被淘汰；若 AI 依据既有畅销品材料生成变体，筛选器可能奖励与历史偏好相似的方案。对“更好产品”的主张，至少要报告候选池的生成规则、淘汰数、评审一致性、保留门槛和确认结果，避免只展示成功样例。

14. 从概念评价到预上市预测：分清总量、时点和可得性

14.1 预测目标必须锁定在决策所需的单位

产品组合规划中常见“市场规模”一词，实际可能指潜在购买组织数、某年单位销量、累计装机量、替换需求、某渠道可触达的订单或企业可实现的贡献利润。它们不是同一变量。累计装机量会包括仍在使用的旧产品；年度新购可能包含替换但不增加净安装基数；渠道出货会受到库存压货影响；收入又同时受价格、折扣和产品组合影响。模型输出若不附目标定义、单位、地区、观察期和销售链条位置，精度数值无法解释。

设一个分析性新品案例：某类设备的目标名单有 100 个账户，以这 100 个账户为共同分母，40 个技术适配、30 个在两年窗口内有预算周期、20 个可由现有服务网络覆盖。这是三个边际计数，不能仅凭它们确定同时满足三项条件的账户数；交集可以在 0 至 20 个之间。只有额外假设三个条件相互独立，边际比例的乘积 100×0.40×0.30×0.20 才对应 2.4 个期望账户，而非 24 个实际账户。另一种合法的概率设定是：技术适配概率为 40%，在已适配者中有预算的条件概率为 30%，在同时适配且有预算者中网络可覆盖的条件概率为 20%；依链式法则相乘仍得 2.4 个期望账户，此时不需要条件相互独立。若“30”指 40 个适配账户中实际有预算的账户数，条件比例就是 30/40=75%，不能再写成 30%；若 12%已表示“适配且

SVSAC | sv-sac.org

有预算”的联合比例，再乘 40%才会重复折减。来自不同样本或时间窗的比例也不能直接拼成同一漏斗。实际名单应优先按账户级交集逐步去重：名单账户→满足适配条件→窗口内采购→渠道可覆盖，为每一步保留条件分母、实际整数计数和排除原因，并将概率模型的期望数与已核实的账户数分开。

分母还可能随着信息系统而变。市场数据供应商覆盖的大型企业，不等于目标地区全部采购组织；搜索趋势有活跃用户和平台覆盖限制；公开询价多来自愿公开需求的账户。若 AI 仅能读取可见样本，模型估计的是“可观测账户中的倾向”，不是全市场普遍需求。无法用名单框、权重或外部抽样校准恢复的缺失，不应由语言模型补成确定的市场比例。

14.2 预上市模型所依赖的结构性假设

Bass 在 1969 年提出的耐用品扩散模型，针对初次购买时间与既有购买者数量相关的机制建立，并以 11 类历史消费耐用品数据进行检验。出版社摘要报告模型对历史数据销售峰值及峰值时间预测表现良好。[10] 这项经典工作提供了一个明确的扩散机制视角：新产品采用不仅由外部触发，也可能通过既有采用者影响潜在购买者。但它不是对任意新品、工业设备或 AI 市场扫描的普遍精度保证。产品类别、潜在采用者范围、竞争格局、价格变化和渠道供给若不同，参数稳定性就需要重新验证。

模型的输入和目标也要区分。若没有观测到新产品的历史销量，采用类似产品估计创新系数、模仿系数和潜在市场规模，实际上依赖“类比产品与目标产品在关键采用机制上足够相似”的假设。AI 能提高候选类比的检索覆盖，或提取不同产品的特征；但“语义相似”不能自动保证采用过程相似。一个功能相似的设备可能面对不同的认证、安装成本、维修能力、替换周期或监管期限。所谓“类比产品”应由行业专家根据任务和采购机制审查，并记录为什么纳入、为什么剔除及其与目标产品的差异。

当预测期限较长时，参数不确定性会被累计放大。若潜在市场规模、购买率和产品替换周期均只用单点估计，模型图上看起来平滑的增长曲线可能掩盖多个互斥情景。与其报告一个长周期确定数，不如将外部采用、口碑扩散、价格/供给变化、竞争替代和产品更新分别设为情景条件，给出哪些数据能区分这些情景。情景范围并非概率区间，除非其概率构造有明确依据。

14.3 流程图不能替代状态定义

Urban、Hauser 与 Roberts 针对一款新汽车开发并应用预上市预测与测量系统。出版社摘要称，该研究追踪消费者主动搜索、经销商访问、口碑、杂志评测

SVSAC | sv-sac.org

和生产约束等因素，并以消费者转移过程组织预测。[11] 其设计启发在于不把“市场需求”视作一个瞬时总量，而是观察从知晓、搜寻、访问到选择的不同状态。但证据是针对汽车这一耐用品市场的具体模型和应用，不能未经验证移植到制造企业新产品组合，也不能推断当代 AI 模型已经复制该系统。

一张状态转移图只有在事件能被稳定观测时才可估计。比如概念浏览到技术咨询的转移率，分母应是实际获得浏览机会的账户，不是所有目标客户；技术咨询到样品测试的转移率，要明确客户主动申请还是销售代表推动；测试后到正式订单的间隔要区分仍在评估与已关闭失败。若各地区的 CRM 阶段含义不同，状态转移率可能只是填报习惯差异。AI 抽取历史记录可以辅助标准化文本，但不应自动把“感兴趣”“同意演示”映射为同一行为状态而不展示判定依据。

在制造业或 B2B 产品中，供应和交付能力本身影响可见需求。若目标产品没有样机，客户无法进入试用；若认证未完成，采购订单不能成立；若销售人员只联系头部账户，进入漏斗的客户已经过选择。此时漏斗中“未转化”不是纯需求障碍，“转化率低”也不能只靠强化营销解释。预测应将需求状态与公司能否提供样机、报价、认证、交付、售后分列，给出每项条件的来源和时间戳。

14.4 组合决策要惩罚错误的方向，而非追求对称误差

普通预测误差常用平均绝对误差或百分比误差汇总，但组合决策中高估和低估的后果可能不对称。高估某个新品机会可能占用试制线、关键工程师和认证预算，挤出已有高确定性项目；低估一个小而关键的客户细分则可能导致迟迟不投入、失去学习窗口。评价方法应联系实际决策损失，而不是仅比较总体预测分数。

一个简化示例：项目甲预测中值为 100 台、区间 60–140 台，预期单位贡献 1 万元；项目乙预测中值 80 台、区间 40–100 台，单位贡献 1.5 万元，但必须先支付固定认证成本。只比较预测销量会选甲，只比较中值乘单位贡献则甲为 100 万元、乙为 120 万元；两者都未包括认证失败、供货约束、售后责任和组合冲突。若甲的需求主要由一个尚未确认的大客户贡献，且设备规格无法覆盖其他客户，集中度风险会改变排序；若乙的高单位贡献依赖未验证的服务价格，亦不能作为确定利润。组合选择应以可行方案的决策损失分布为输入，并明确不可接受的下行情景，而非以 AI 生成的市场潜力分数排序。

“更高精度”也未必改善组合。假如模型对销量的平均误差下降，但误差主要发生在低价值项目，而对占用稀缺工程资源的项目仍系统高估，则组合选择可能不变甚至恶化。相反，若模型虽未降低总体误差，却能识别某个错误概率高、投

SVSAC | sv-sac.org

入不可逆的项目，避免重大资源锁定，其决策价值可能更高。该判断需要把预先规定的项目集、预算约束、决策规则和最终损失连接起来；只在事后挑选模型表现最佳的子集会重新引入赢家偏差。

15. 工程可行性不是市场信号的附注

15.1 需求假设必须写成可交付的产品命题

“客户希望减少停机”是需求陈述，不是可批准的产品规格。要进入组合决策，至少需说明目标工况、当前替代方案、停机成本由谁承担、改善如何测量、最低可接受效果、设备接口、认证或验证要求、可承诺价格和交付时点。缺少这些字段，AI 生成的“市场机会”可能只是把痛点语句转写为产品功能，却没有解释谁会为功能付费、谁负责部署或谁承担失败后果。

需求从文本到工程命题会发生多次翻译。客户说“更快”可能指设备单循环时间，也可能指换型准备、等待维护、订单交期或整体吞吐；“可靠”可能意味着少停机、结果可重复、备件及时或数据安全。不同解释导向不同架构和成本。生成模型善于把自然语言组织成连贯需求，但连贯并不代表语义已被共同确认。可审查的需求卡片应保留客户原句、来源、提出时间、对应使用任务、候选解释、冲突证据、可验证指标和负责确认的角色。任何对原始要求的改写应能回链，不得只保留模型摘要。

一项产品组合分析不应把工程筛选等同于“市场否决”。若材料供应周期导致当前无法实现某个性能，决策可以是推迟、改变材料、拆成模块或重新定义上市区域；也可能因为成本不可接受而停止。每一种结果都应区分客户价值不足、工程不可行、交付能力不足和单位经济性不成立。这样才能避免因一个通用“可行性分”掩盖完全不同的可干预原因。

15.2 硬门槛和可权衡属性分开

在多属性产品中，某些性能可以权衡，另一些是硬约束。若安全认证、关键接口或特定法规是进入市场的必要条件，其他属性的高分不能抵消缺失门槛。把所有维度直接加权平均，可能使一个不合规或无法制造的概念因为市场热度足够高而排在前列。相反，若把所有要求都设成硬门槛，又可能过早淘汰可通过重新设计解决的问题。

决策记录可区分三类约束：已知不可违反的外部要求；客户/合同要求但可协商或替代的条件；企业内部资源与经济边界。对第二、三类，需要说明谁有权决

SVSAC | sv-sac.org

定例外以及代价；对第一类，不用预测分数替代合规证明。若规则并未由法务、认证或工程责任人核实，不要让模型把网络文本里的“常见要求”写成强制标准。

技术可行性还受设计成熟度影响。概念阶段用渲染图得到的客户偏好，不等于产品可实现时的接口、体积、能耗、维护性和成本。将渲染效果拆成“市场属性”与“尚未通过工程验证的承诺”，有助于识别视觉概念是否借助实际不存在的性能赢得评分。确认测试应使用与计划交付版本一致的信息；若试验中途更改产品或服务条件，需重新标记版本，不能把不同版本的评价合并成一个新品结果。

15.3 工程门槛的反事实和退出价值

阶段决策常被描述为“继续”或“停止”，但存在至少四种策略：继续当前方案、投入有限验证、调整方案后再测、停止并释放资源给其他项目。决策价值不仅取决于某个概念成功概率，也取决于下一步试验能否降低不确定性，以及失败后能否及时终止。一个高潜力概念若要昂贵定制工装才能验证，可能不如一个中等潜力但可用低成本原型确认的方案；也可能反过来，低成本试验没有触及关键性能，信息价值很低。

可用期望损失表达这种比较，但数值必须来自组织自己的成本和风险，不应从公开论文或 AI 生成回答推断。设下一阶段试验成本为 C，若结果为正则进入开发的净价值为 V，若负则停止并释放资源，且试验对最终判断的误差概率已被有依据地估计，则试验价值取决于结果对后续行动的改变，而非试验本身的“准确率”。若无可靠的误差概率，做灵敏度分析：在一系列假设下，继续/调整/退出的排序如何变化，哪些未知参数对结论最敏感。未知参数应列为待核证据而不是由预测模型补齐。

早期试验的价值还受学习可迁移性限制。一个客户的样机成功可能只证明其特定接口和流程适配，不一定适用于其他地区或客户类型；一个失败可能因测试媒介、安装条件或操作培训不足，并非需求不存在。解释试验结果时，需记录产品版本、用户条件、安装环境、实验者协助程度和结果判据，避免把单次原型体验当成最终商业接受度。

16. 用三类分析情境检验预测链条

本章三种情境均为作者构造的分析情境，不是企业案例、现场观测或市场规模估计。其用途是展示同一个“AI 发现市场机会–提出新品–预测收益”链条在何处会断开，以及需要什么证据才能从文本信号向组合决策前进。

SVSAC | sv-sac.org

16.1 标准化模块：市场扩张可能受认证和替换周期约束

设想一家制造商正在考虑面向多个地区推出一种标准化传感模块。模型在各地公开资料中发现“预测维护”和“能效”相关讨论增加，并从竞品说明书抽出相似功能。若仅按讨论量和竞品数量排序，市场看起来足够大；但真正的购买机会取决于设备接口是否兼容、既有系统的替换周期、采购是否在年度预算内、安装是否会影响保修、认证是否允许跨区使用。老设备仍可运行且更换需要停产窗口时，公开兴趣上升可能很晚才转成订单。

可检验预测应是条件式的：在具有兼容接口、近期改造窗口和可服务渠道的账户中，基于 AI 汇集的多源信号所形成的候选列表，是否比固定行业筛选规则更能预测后续合格询价或付费试装？分析单位可为账户–设备群–季度，并保留目标账户名单中未联系和未进入样机流程者。若只跟踪已联系的询价账户，销售筛选导致的差异会被误认为模型识别能力。替代解释包括渠道商活动、强制能效标准改变、原设备停产和同期价格波动。

若 AI 识别的兴趣提升但合格询价不变，应检查其文本是否来自法规新闻、展会报道或售后话题，而非购买组织需求；若询价提升但试装不提升，可能是价格/认证或接口不匹配；若试装提升却未转订单，则可能是客户为了信息而免费试用、预算延后或安装成本过高。每一个断点都应映射到事件和分母，而不是把链条终点提前定义为“市场验证”。

16.2 定制设备：单个大客户的信号不代表可复制市场

设想一款高价值定制设备面向少量大型工厂。AI 从项目招标、招聘信息、技术论文和客户网页识别出扩产信号，并把相似工厂列为潜在客户。但此类机会通常受工程改造、厂房条件、客户验证流程、服务响应和资本预算影响；一家工厂提出的特殊需求，可能需要大量非重复开发。若模型把同一集团不同工厂、新闻稿转载和供应商公告当作多个独立信号，市场证据会被重复计算。

可检验比较需以采购项目或企业集团为聚类单位，先按来源谱系去重，再明确目标事件是进入技术评估、签署付费试点还是正式采购。由于项目数量少，不宜依靠过多预测特征训练黑箱模型；更有价值的可能是辅助研究员找出合同、接口和组织变更的证据，并以专家盲评比较遗漏/误报。若少数大客户贡献大部分潜在订单，报告总额之外还要展示客户集中度和单一项目失败的组合影响。

反例是某工厂因政府补贴或设备更新期限而集中采购，该需求可能是真实但不可持续的临时窗口。相似工厂没有同样补贴，就不会复制其购买行为。另一种反例是客户咨询由系统集成商发起，而非终端工厂预算持有人；被识别为“潜在客

SVSAC | sv-sac.org

户”的账户并未形成真实预算。AI 对公开扩产信号的召回可以很高，商业上仍可能因这些角色和制度条件而无效。

16.3 消费品新品：试用、复购和渠道可得性须分别观察

设想一款面向多个国家的消费品新品。模型综合公开搜索、零售评论、竞品成分和产品图像，生成不同配方/包装概念。试用测试的兴趣高，但首批零售铺货不足、价格促销强或当地竞品同期断货，则短期销量很难单独归因于概念质量。反之，概念评价普通但品牌已有渠道覆盖，也可能得到初期销售。数据中每项信号要标注地区、时间、平台覆盖、产品版本、价格和渠道可得性。

消费品概念验证可以在相同产品信息、价格区间和目标样本条件下随机分配“文本”“渲染”“样品试用”以及“AI 整理的证据简报/现行研究简报”等刺激，测量理解、实际选择、愿付价和后续购买。后续实际购买仍需真实上市或有交易约束的试购设计；问卷意向不能直接乘以目标人口得到销量。若 AI 版本提高意愿但真实试购率没有改善，可能是概念说服力而非产品接受；若试购增加但复购下降，则初次好奇与持续价值不同；若零售可得性不均，销售差距可能是铺货而非地区偏好。

16.4 组合判断的证据停止规则

三个情境共同提示：市场信号是否足以推动研发，不应由“数据更多”自动决定，而要看下一项关键不确定性是否可被合理成本的证据消除。标准模块可能需要接口兼容和采购窗口证据；定制设备需要确认项目预算及需求可复制性；消费品需要区分创意反应、试购、复购与铺货影响。对于每个候选项目，可写下一条可反驳命题、最关键的反证、下一项最小验证、失败后损失及何时不再投入。

停止规则并非固定通用阈值。它依赖错误投资和错失机会的相对代价、项目间资源挤占、未来信息到达时点和产品可撤回性。若项目投资可分期且试验能快速给出高质量信息，保留一个中等概率机会或许合理；若开发不可逆、认证窗口已近且关键客户条件未知，继续扩张预测模型可能比尽早获得独立工程/客户证据更差。没有本组织的成本、事件率和决策期限，论文只能提出需预注册的比较方式，不能给出普适的“分数达到多少即立项”。

SVSAC | sv-sac.org

17. 产品组合筛选不是把单项预测从高到低排序

17.1 预算约束改变“好项目”的定义

若研发经费、样机线、认证人员或上市窗口足够，多个正净现值项目可以同时推进；实际组合通常只能选其中一部分。于是，项目评价不再是“这项是否过线”，而是“在共同资源约束下，哪些项目构成更好的可行集合”。一项销量中位数最高的新品，若要占用稀缺平台团队两年，未必优于三项较小但可复用技术和渠道的产品。单项预测排名与组合价值排名并不相同。

Böttcher 与 Klingebiel 的 2024 年在线论文建立了带预算约束的组织创新项目选择模型，通过模拟比较单一决策者、专家委派、投票、评分平均和排序聚合等规则。作者报告，在其许多参数情境中，对候选项目进行排序再聚合优于平均分或投票，尤其当预算只能支持少量项目；若组织确实能把每个项目准确交给相关专家，委派仍可能合适。[12] 这一结果不是现场企业采用某种规则后收益上升的实证，也不表示所有组织都应改用排序。论文明确以独立项目、非策略性评价者和特定噪声结构为建模基础，并将项目间依赖、评价者互动及策略行为留给后续研究。[12] 对真实组合规划而言，这些被简化的假设恰好可能很重要。

组合会发生排序，是因为资源约束使相对比较不可回避。若一项项目的资金、工程人力和认证时间可以替代另一项项目，项目收益的绝对数值仍不够：还需比较投入时点、约束消耗和退出后资源能否释放。用“每元收益”指标也可能误导，因为研发资源不是完全可分、可即时转换的同质资金。项目需一整组特定技能或设备时，低单位成本项目也会因关键瓶颈资源不足而排挤其他项目。

17.2 评审聚合规则须与有限选择相匹配

跨职能评审者看到项目的不同切面。工程师能识别接口复用或技术债，销售能判断客户预算信号，供应链能看见交付风险，财务关注现金时序；任何一方都可能缺少完整信息。把每人给出的“价值分”简单平均，隐含分数在尺度、方差、理解和误差方向上可比。把赞成票相加，又会丢失评审者对项目相对优先级的排序信息。相反，排序汇总较少依赖不同评审者分数尺度一致，但若需要同时选择有依赖关系的项目，单纯排序仍不能识别组合协同。

Böttcher 与 Klingebiel 的模拟发现，严预算、候选多、决策者少时，排序聚合可以比平均或投票更好地区分少数高质量项目；随着可资助项目数增加、候选集变小或评价者增多，规则相对表现会变化，评分平均在部分情境下可能占优。[12] 这些是论文设定内的模拟而非企业通用规则，且模型将单项“质量”作为可比较维度，不等于技术、市场、战略和风险的多目标组合已经解决。实际委员会可把排

SVSAC | sv-sac.org

序用作意见压缩工具，而不是最终决策：排序之后仍需检查硬约束、项目成本、共用能力和情景风险。

一个可操作且可审查的做法，是在每个评审周期先冻结候选集合、信息截点、决策者名册和预算边界，再要求参与者对候选项目分别给出独立排序、关键理由和至少一条可能推翻自身判断的新证据。聚合结果可比较排名和权重敏感性，记录哪些项目因信息不足、而非因为预期价值低而落选。若模型提供候选排序，也应与人类排序分开保存；展示模型排序后再让评委打分，会把自动化建议变成新的锚点处理，无法再解释为独立的人类判断。

17.3 候选池本身决定可选上限

选择算法无法弥补候选池系统性缺少某类机会。若 AI 主要从现有产品页、已服务客户和历史项目生成新产品，候选池可能偏向改型和熟悉市场；若它更常覆盖线上声量大的消费领域，资本设备、维修工具和低频采购产品可能被低估。一个“组合评分模型表现好”的结论只对进入评分环节的候选项有条件成立。由于提议来源、经理赞助、用户群体和数据覆盖会影响候选进入概率，应另外检查未入池项目的来源和未纳入的理由。

可以将项目产生与项目选择分为两道验证。前一道比较市场扫描方法能否发现过去的未满足需求、客户任务或潜在替代方案，并让独立研究员复核是否漏掉机会；后一道才比较决策规则能否在固定候选池和固定预算下选出更好的组合。若 AI 组因产生候选更多而胜出，这可能是候选覆盖度提升；若各组候选池相同而组合不同，差异才可能来自筛选规则。将两者混成一项“AI 提升组合质量”的指标，会掩盖技术的作用路径。

18. 平台复用创造协同，也会制造共同依赖

18.1 共享技术不等于共享收益

多个新品可能共用传感器、软件底座、材料、认证、服务网络或生产工装。共用能力会降低重复投入，也可能带来一次故障或延误影响整组产品。组合矩阵至少需记录项目间的资源互补和风险传递：一个项目是否先交付接口、测试夹具或平台软件，另一个项目是否依赖该交付；如果基础平台改版，哪些型号需重认证、重新验证或改变客户承诺；若平台被一名核心供应商控制，缺货或涨价会怎样改变多个项目的可行性。

线性加总的项目净现值会漏掉依赖项。项目 A 和 B 各自估计贡献为 100 与 80，但 B 只有在 A 的接口平台完成后才可上市；若 A 失败，B 的收益也无法实

SVSAC | sv-sac.org

现。如果直接分别将 100 和 80 计入组合，还把共享平台节省的研发成本按两项目各自的满额节省重复计入，就会夸大组合价值。反向情形也存在：A 的基础投资已经完成，B 增加的边际开发成本很低，合并评估反而应包含平台复用的真实增量价值。

一个假设算例可以说明差异：平台建设费用为 120 个成本单位，产品 A 和 B 各自的独立收益为 90 和 70；若平台完成，A、B 均可上市，合计收益 160，若平台未完成，两者都无收益。只评估 A 时，收益减平台成本为−30；只评估 B 时为−50；组合一起推进时，收益 160 减共同平台成本 120 为 40。正确的组合结果不是把 A 单独净值−30 与 B 单独净值−50 相加为−80，也不能把同一 120 平台费用只分摊一次后仍声称每项分别盈利。组合结果依赖“平台完成”共同事件和项目关联。假设概率若为 p，期望贡献才是 160p−120；在 p=0.5 时为−40，而不是基准确定完成下的 40。这个差异突出需要验证的不是单个模型分数，而是共享依赖完成概率及失败后可回收的资产价值。

18.2 依赖图需要按事件而非标签构造

“共用平台”“技术相似”不是充分的依赖关系定义。更有用的依赖图将节点定义为项目、组件、技能、测试资源、认证证据和供应商；连边标明方向、条件和时间。比如项目 B 使用平台接口 A，前提是 A 在某日期通过兼容测试；项目 C 与 A 共用同一认证实验室，但排期冲突会延迟两者；项目 D 仅复用算法库而不共享硬件，A 的供应问题不应自动传导给 D。依赖图必须根据真实工作包和释放条件审查，不能从文本相似度直接推断项目绑定。

应分别识别正向协同、资源竞争和共同风险。协同可能是共享研发成果、组件采购规模或共同渠道；竞争可能是同一工程师、试验台、认证窗口和市场上市时机；共同风险可能来自一项法规变化、平台接口失败或供应商单点故障。将这些统一塞入“相关性系数”会失去可解释性。更可审计的报告为不同情景分别列出哪些结果共同变化、谁受影响、哪些成本已共享、哪些损失仍是项目特有。

18.3 组合分散化并不保证风险下降

产品矩阵看起来跨行业、跨地区，不等于统计上或运营上真正分散。不同客户群可能依赖相同的半导体、同一云平台、同一认证机构，或同一销售团队；在共同冲击下，收益会一起下降。反过来，外观相似的产品也可能由不同工艺、供应商和采购窗口支撑，风险未必共变。以行业标签、市场国家或产品类别计算多样性，可能误判实际共同依赖。

SVSAC | sv-sac.org

一种透明的压力测试可设三种状态：平台成功且需求如期、平台延期而需求仍在、平台失败且需求窗口关闭；为每个项目分别指定在各状态下的收益、成本和资源释放时间。若没有真实数据，概率应作为情景参数而非估计结果。决策者可查看组合在各状态下的总现金需求、关键人力峰值和项目可继续数量，而不是只看平均收益。组合若在“平台延期”状态下需要超出组织现金或工程能力，单项期望净现值为正也不能推出可执行。

19. 新品销售不等于新增市场：内部蚕食和净增量

19.1 从新品收入分解至组合增量

同一企业推出新品后，新品销量可以来自三类来源：从竞争对手获得客户、把原本不购买的客户转为采用、以及从企业既有产品转移客户。新品销售额本身不能分辨来源。第三类常被称为蚕食，但蚕食并不必然是失败：若旧产品利润低、客户长期留存改善、竞争者即将抢走用户，主动替代可能创造价值。关键估计对象应是组合总贡献相对无新品反事实的变化，而不是新品产品线收入。

Srinivasan 等文献仅核到书目身份，可保留为蚕食研究待核线索；[13]本篇未取得可读摘要或案例原段，不确认饮料企业案例、参数框架及其层级结果。以下组合层记账与 5200/2400 例是作者明确设定的分析，不是该文估计，也不证明 AI 已经识别内部替代。实际蚕食需以统一产品边界、期间和反事实重新判断。

19.2 两种相同新品销售、相反组合价值的账本

设一个明确假设的年度账本：新品售出 100 件，每件贡献毛利 100 单位，总计 10000。方案甲中，60 件来自旧款产品替代，旧款每件毛利 80；20 件来自竞争对手转入，20 件原本不购买。若销量逐件互斥、渠道库存和价格影响均已校正，则相对于无新品情景的毛利增量可先粗算为：新品 10000，减去旧款被替代的 60×80=4800，净增 5200 单位。方案乙仍卖 100 件，但 95 件来自旧款替代，剩余 5 件来自竞争对手或新需求，则净增量为 10000−95×80=2400。两种情形新品销售相同，组合增量差 2800。

这还不是完整利润比较。若新品研发、认证、售后、促销和渠道返利成本没有扣除，净贡献会更低；若新品带动旧款价格维护、共同服务合同或客户留存，则会有额外收益；若渠道转移被记录为厂商出货而非终端销售，销售差异可能只反映库存。上述账本用于展示归因所需字段，不是实测产品效应。实际计算还必须定义“旧款减少”的反事实：旧款销售也可能因竞争对手、季节和供应变化下降，不能把所有同期下跌都算新品造成的蚕食。

SVSAC | sv-sac.org

19.3 识别内部替代要观察对照和时间

新旧产品同时销售时，跨产品相关性不等于蚕食因果。若企业只在旧款需求高的地区推出新品，原有产品即使不受新品影响也会出现均值回归；如果同一时期停产旧款、变更价格或调整经销政策，则无法将下降单独归因于新品。可行的设计应在区域、渠道或客户组中比较不同上市时点，或者使用可信的历史控制组，并同步记录价格、促销、供给、库存和产品版本。新品分配给销售区域的规则也可能与预期需求相关，需明确其选择过程。

反事实的层级取决于问题。要估计“新品让企业总组合贡献增加多少”，分母和结果要覆盖新旧全产品线及目标市场；要估计“旧款客户中有多少换购新品”，需要客户级账户/身份链接和实际购买；要估计“新品抢走竞品份额”，通常需要竞争者或市场总量数据及其测量可信度。单一数据集往往无法同时识别三个问题，AI 抽取更多公开资料不能自动补出缺失反事实。

避免蚕食也不是唯一合理目标。企业可能有意推出低价或更新款，以防外部竞品抢占其旧款用户；此时新品造成的内部转移可能小于不推出新品的外部损失。另一种情况下，新品只是分走了旧款渠道陈列和售后资源，未增加总体采用或贡献。报告应呈现“主动替代策略”的对照情景，而非把所有新旧产品负相关都贴为负面。

20. 先发低配、等待高配：组合的时序选择

20.1 上市顺序包含融资、学习和声誉效应

项目组合不仅选择“做什么”，还选择“先做哪一个、什么时候上市”。成熟产品可以较早上市并产生现金，帮助后续研发；但低质量或功能不完整的先发版本可能影响客户对品牌、后续版本和服务能力的判断。等待更高配版本可提高产品质量，却延后收入、错过市场窗口或使资金耗尽。若新品之间存在连续依赖，顺序本身会改变组合价值。

Bhaskaran、Erzurumlu 与 Ramachandran 的理论论文研究现金受限创业企业在已可上市产品与开发更先进版本之间的顺序选择。其出版社摘要指出，较早推出现有版本可带来后续研发资金，但低性能也可能降低未来产品利润；模型结果显示，成熟企业与初创企业的最优推出策略可能不同，并取决于现金约束、产品依赖、研发成本和失败风险。[14] 这是理论模型与行业示例，不是企业现场因果估计，也不能被简化为“先上市总能更好”或“质量优先总更好”。

SVSAC | sv-sac.org

20.2 一个两阶段构造说明关键条件

假设组织只能为下一阶段投入 100 个资金单位。当前可上市版本在今期带来净现金 60，但其性能偏低；若不上市，组织等待下一期更高配版本，成功时净现金 180，失败时无收入，失败概率暂假设为 0.4。若当前版本上市，产生的 60 可覆盖一部分开发支出；但假设它会使下一代在成功时少获得 30 的净收益，代表声誉或客户体验损失。忽略折现、税费、运营及融资外部性时，等待策略的期望收益为 0.6×180=108；立即上市并继续开发的期望收益为 60+0.6×(180−30)=150；若推出低配版导致下一代成功概率下降，决策可能反转。假定成功概率下降至 0.3，则立即上市为 60+0.3×150=105，等待仍为 108。这里成功概率由 0.6 降至 0.3，相差 30 个百分点；收益相差 3 个资金单位，并非概率相差 3 个百分点。令 60+150p=108 得临界 p=0.32：低于 0.32 时等待更优，高于时立即上市更优。此处只比较假设现金收益，尚未给出开发支出时序，不能由 100 的初始资金直接断言两策略均融资可行。

这个小算例不是参数建议，概率和金额均为构造值，且将失败后可转用资源、上市运营成本、客户群扩散和竞争者行动等都省略。它的用途是显示预测误差进入策略的位置：市场预测影响当前收入估计，工程验证影响先进版本成功概率，客户反馈影响低配版对声誉或需求的变化，融资成本影响能否等待。不同团队若在同一总收益中把“提前上市现金”和“后续开发避免的资金缺口”重复计算，就会重复计入资金价值；现金流时间、利润和投资回报必须分别建模。

20.3 顺序政策需与组合对照

判断早上市是否更好，不能只比较“今年卖了多少”。对同一候选集，可预先定义至少两种策略：策略 A 在达到基础可行门槛后先上市现有版本，并保留升级路线；策略 B 不上市，等待更高配版本达到预先冻结的技术条件。结果指标可包括预定期限累计贡献、后续版本实现率、客户持续使用、退货/服务成本、资金缺口以及被挤出的其他项目。若要做现场比较，应按企业或产品平台集群分派策略，避免同一销售团队同时执行两种策略、造成客户和培训资源互相污染；在现实中不能随机的部分，要报告选择机制和可比性边界。

客户采用既可能是资金来源，也可能改变产品升级的证据。先发客户样本通常自我选择，早期用户可能更专业、容忍缺陷或希望参与共创；他们的反馈不必代表后续主流客户。若只分析已购买者中的升级率，就漏掉未购买者和退出者，也会把选择后的客户组成变化当作版本改进效果。应保留全部原始分派账户或预

SVSAC | sv-sac.org

定义目标客户总体，在固定时间点比较策略政策的结果，并把尚未成熟的购买/升级窗口作为未完成观察，而不是简单删除。

20.4 模型、专家和产品经理的职责边界

AI 可以辅助搜集异构市场信号、提取竞品和客户语料、检索可比产品、生成情景或暴露组合依赖；它不能单凭文本证明客户愿付款、工程能交付、认证可通过、供应商会供货或新品给企业带来利润。特别是当模型使用同一类公开材料生成机会、评价机会并为最终组合排序时，多个表面独立输出可能共享同一语料和同一错误，不能当成互相验证。

可以把模型输出作为决策过程中的一个候选意见，与独立工程、客户、财务和渠道证据并列。每个组合决策记录：目标市场和时间窗、候选池生成方式、可比项目、共享资源及依赖、单项与组合损益、证据未决项、情景假设和退出条件。事后复盘时应把当时可见信息和后来发生的信息分开，避免按结果反向改写当初预测。只有在同一预先规定的候选集和预算下，模型辅助流程能提高预先定义的组合目标、且未增加不可接受的资源/下行风险，才能进一步支持“AI 改善组合决策”这一主张；概念评分或预测误差下降仍不足以单独支持它。

21. 把市场扫描和组合决策拆为可检验的政策比较

21.1 先冻结候选总体，再比较筛选机制

要回答“AI 辅助是否改善新品组合”，不能只在模型认为有前景的产品中检验。分析开始前，应定义候选总体：哪些市场问题、产品类别、地区和时间窗进入本轮判断；候选由谁提出；什么条件允许进入；哪些候选因技术门槛、资料不足或战略不符被排除。候选集合需在结果未知时冻结，并保留候选来源和版本。若 AI 组生成更多概念，而对照组只评估人工提出的少数项目，最终差异同时包含候选生成覆盖和排序两个机制；有些核心问题需要整体比较，但不能把贡献归给单一环节。

一种分阶段比较可以先统一候选池，再比较筛选工具。对于同一组候选，将 AI 生成的证据摘要、传统研究简报和不提供新增摘要作为不同信息条件，随机分配评审团队；评委在不知道摘要来源的条件下按同一标准评价。然后在固定预算、人员和候选池下，将模型建议排序、常规评审排序和预先规定的组合规则分别形成“影子组合”，但暂不改变真实立项。这样估计的是不同信息/聚合政策在同一候选池下的评价差异，而非“AI 自动创意生成”的总体效果。

SVSAC | sv-sac.org

第二阶段才比较候选发现。由 AI 辅助扫描与常规扫描分别建立候选池；为避免候选数量本身决定胜负，可同时保留“等数量候选”比较和“实际工作流全部候选”比较。等数量比较关注排序/质量，实际工作流比较关注每单位人力可以发现多少可核机会；若后者的候选更多但每项质量较低，组织可能仍受验证和工程容量限制。两个估计对象应并列，不把其中一个称为另一个的替代。

21.2 实验单位和干扰边界

若不同团队共享同一市场数据库、工程专家、产品平台或供应商信息，将个人随机分派为 AI/非 AI 条件会产生污染。评审者可能把 AI 摘要转给对照组，或两个组把方案提交给同一工程师。若项目间共享平台，组合也会互相干扰。处理单位应根据可共享资源而定：可以是产品线、市场问题、业务团队或地区集群；若同一人员跨组工作，需记录重叠并在分析中采用集群级分派或将污染作为偏离方案记录。

一个拟议的两因素试验，可在足够多个独立业务集群内随机分配“AI 辅助信息整理/常规整理”和“组合层聚合规则/现行规则”，形成四个试验臂。项目集群分配后，集群内所有候选都遵循同一流程，避免单项项目随机化导致团队成员把流程习惯带到对照项目。主要估计应是按最初分配计算的意向处理效应；未采用模型建议、工具不可用、评委绕开规则等属于实施结果，不应在主分析中删除。另可报告实际采用分析，但它受到采用者选择影响，需要不同的因果识别假设。

处理“AI 信息”时，要锁定模型版本、提示、检索范围、来源时点、可见字段和人工训练；处理“评审方式”时，要锁定候选池、预算、各项目成本和禁止条件。除目标处理因素外，评审时长、专家数量、摘要长度、业务权限和截止日期尽量保持一致。若 AI 组额外配备更多研究员或高层审核，这就变成资源包的比较，而非纯工具比较。作为真实管理政策包比较可以接受，但结论必须明确包含的资源和职责。

21.3 结果变量与盲态评价

组合质量不能由模型自身的信心分数评价。至少要预先定义候选层、组合层和市场层三类结果。候选层可以观察来源事实准确率、适配证据完整性、技术硬门槛、客户任务可检验程度、经独立评审的需求证据；组合层包括固定预算下的风险调整贡献、资源峰值、风险集中和被挤出的替代项目；市场层则包括真实报价响应、付费试点、正式订单、交付、使用和复购。各层时间窗口不同，早期工具实验不能直接使用尚未发生的后期变量作为结果。

SVSAC | sv-sac.org

独立评委可在隐去流程条件、模型标签和团队身份后，按预先冻结的量表评价事实支持、需求清晰度、可验证性、技术可行性和方案新颖度。评分规则应说明“无证据”“与证据冲突”和“可接受但尚未验证”的区别；不能让评委把流畅、完整或行业术语丰富误认为事实正确。至少对部分材料进行双人盲评，并报告分歧处理方式、缺失率和评委一致性。评委本身的领域经验也要记录，因为一个方案对产品经理有吸引力，未必对采购者或现场工程师同样有价值。

若真实立项结果短期无法观察，可先比较预测校准和决策质量代理指标，但明确其代理地位。例如估计的客户需求概率可以和预注册的独立客户验证结果比较；工程可行性判断可以与盲态技术审查结果比较；组合选择可与事后同口径的候选表现追踪。只报告离线精度或专家评分，最多说明信息处理或筛选结果改变，不能据此声称产品商业成功。

22. 将客户验证嵌入组合研究，而非事后补一轮问卷

22.1 选择谁来验证本身是设计问题

样机或客户验证往往只对候选项目中已经通过内部筛选的一小部分进行。若 AI 与常规流程选中的候选质量不同，直接比较两组被测试项目，可能是在处理后选择的样本中比较，无法估计整个筛选政策的效果。对外部客户的访问也会受销售人员、渠道伙伴和客户关系影响；被邀请且愿意参加试用的人，可能更有兴趣、更有预算或更容易合作。

如果安全、预算和客户条件允许，可以先在预先划定的候选总体中随机分配信息政策，再预先抽取一定比例的项目进入独立客户验证。验证项目的抽样概率应留存；如果高成本项目更常被验证，应按抽样概率加权估计总体表现，并同时报告未验证项目范围。若验证完全取决于 AI 分数或经理偏好，未验证候选的结果不可从已验证样本直接外推。通过分层抽样覆盖高/低评分、成熟/早期需求、不同地区和产品类型，可识别模型只对熟悉案例有效的可能性，但小样本不适合生成大量后验子组结论。

客户验证不应把一次“喜欢”或“需要”直接当作购买。可按成本递增设置行为层级：对具体问题的确认、技术规格交换、样机测试、含价格报价、付费试点、正式采购、交付、持续使用。每一级都有其自身分母和失访原因。报价后无回应可能是采购周期未结束、联系人变更或报价不合适；这些状态不能一律归为拒绝。研究可在固定日期标记仍开放项目，说明观察期及未成熟比例，避免只用已关闭项目算转化率。

SVSAC | sv-sac.org

22.2 有真实约束的选择比口头意向更接近需求，但仍有边界

客户选择实验可以让客户在新品、当前产品、竞争替代和“不采购”之间选择，并随机改变可行属性，如价格、交期、服务和规格。随机展示有助于估计这些属性变化对实验选择的影响；但在假设问卷中选择高价方案，并不等于实际组织已经分配预算或通过采购审批。若能使用付费样机、可退款订金、真实采购申请或被随机提供的试点名额，行为成本更接近现实，仍需区分谁支付、费用是否可退、是否由渠道代付以及试用结束后是否转正式合同。

对 B2B 项目，随机化单位常不能是个人联系人。一个客户组织内的采购、工程和使用者会交流意见；若把同一账户的人分在不同产品刺激条件，信息会相互污染。可按采购项目/账户簇分派或先由企业内部角色分别回答、再以预先定义的决策规则聚合。不能因为样本中有许多联系人，就声称拥有许多独立客户项目。若目标市场只有少量大型账户，应把推断边界限定在这些账户和项目，采用情境/案例证据而不制造伪大样本。

22.3 验证成本应进入信息价值

为确定一个市场命题，工程团队可能需要制造样机、完成认证、接入客户设备或安排现场服务。概念评分越高就投入越多，会形成自我实现的验证不平衡：高评分项目获得更多验证资源，容易拿到更多证据；低评分项目没有机会证伪市场价值。合理比较应设置对所有候选均适用的最低验证包，或明确分配不同预算策略并比较其信息和决策结果。

验证资源本身可视作有成本的信息获取选择。对候选项目 i，下一项试验可在成功、失败或不确定结果下改变立项动作；试验价值并非“多收集了多少数据”，而是预期减少多少错误投资或错误拒绝，减去试验现金、工程工时、客户中断和机会成本。若项目本身已经超过硬性技术/安全门槛失败条件，继续客户试验可能没有意义；若需求和技术机制都未知，一项局部测试或许只能解决其中一项。每次验证前写出“可能结果–对应行动–尚未被检验的假设”，能避免试验完成后按结果重新解释成功标准。

23. 组合政策效果的构造性比较与误差分解

23.1 分开候选级预测和组合级政策

设一个纯构造的产品组有四个候选项目，项目投入为研发费用和瓶颈工程月；两种决策流程分别在总预算内挑选两个项目。AI 流程给项目的收益预测可以

SVSAC | sv-sac.org

更准确，但如果两项候选都依赖同一认证团队，组合仍可能不可执行。反过来，个别项目预测误差较高，但组合选择保留了不同技术路线和客户类型，实际结果可能更稳健。因而需分别评估：（a）每个候选的市场/收益预测误差；（b）固定资源下，组合决策选中集合的事后贡献；（c）因共享依赖、蚕食或风险相关性造成的组合调整后结果。

一个简单的四项目账本可显示差异。项目 A、B、C、D 的假设贡献分别为 120、100、80、70，单项资源成本为 2、2、1、1 单位；贡献暂按可加处理，除下面明确给出的 A–B 依赖外，不另设共享成本或时序约束。A 与 B 共用一项稀缺测试平台，两者同时投入会多耗 1 单位并使 B 延迟一季。总预算为 3 单位，按单项“贡献/成本”比排序依次为 C 的 80、D 的 70、A 的 60、B 的 50；按此顺序逐项填入预算，会选 C+D，用 2 单位取得贡献 150，余下 1 单位无法容纳 A 或 B。但这不是给定账本下的最优组合：A+C、A+D、B+C、B+D 的成本均为 3，贡献分别为 200、190、180、170，A+C 确实最优。A+B 加上平台占用成本为 5，超过预算；A–B 延迟不能使没有 B 的 A+C 变得不可行。该反例表明单项比率贪心可能漏掉可行且总贡献更高的组合。若真实业务中要否定 A+C，必须另有针对该组合、可核验的容量、时序或需求约束，不能用另一组合的依赖代替。所有数值均为构造参数；真实选择仍需逐项登记依赖、订单窗口、蚕食及非可加贡献，不由这个简化账本推出普遍排序规则。

23.2 把预测质量连接到决策损失

给每个项目设真实结果 Y 与预测值 $\hat{Y}$，并记录选择动作 A。预测误差可以写为绝对误差$|Y-\hat{Y}|$，但政策损失需比较实际选中组合的结果与可行预算下的最佳组合，且最佳组合必须在同一候选集和信息条件下定义。若 AI 预测有误但仍选中同一组合，政策结果可能没有变化；若很小的预测误差恰好使一个项目越过资源约束的边界，组合差异却可能很大。只看预测精度会遗漏阈值附近的选择错误，只看已选项目收益则会忽略没有被选中项目的机会成本。

在真实比较中，潜在结果只有一种被观察到：项目被分配到的决策流程和实际立项政策决定了可见结果；未选项目不会获得同样开发、营销和客户验证资源。用“未选项目后来没有销售”作为失败标签是逻辑错误，因为它们未上市就没有销售观察。可以对未选项目做阶段性工程/客户验证以获得部分结果，但其结果不是完整上市回报；对高成本、不可同时测试的项目，应报告部分识别区间或不确定性，不把模型预测当作已观察真实值。

SVSAC | sv-sac.org

23.3 固定资源和被挤出项目的反事实

比较两个组合政策时，预算必须既明确总额，也说明项目成本的时间结构。项目 A 今年需要大额工装、以后维护成本较低；项目 B 每年需要少量软件和客户支持；单一年度预算占用和全周期现金负担不同。共享人员如果原本闲置，其机会成本可能低；如果正挤出别的项目，则需要用被挤出项目的可实现贡献衡量。不能把员工工资作为现金成本、再把同一工时的机会成本全额重复扣除，除非两项分别对应实际现金和不同的可替代产出。

决策政策可以在预算内最大化某一预先选定指标，也可以在下行约束下追求风险调整回报。两者不是同一目标。若管理层设置最低现金流、最大客户集中度或不能同时承担的共同风险，这些限制需在结果揭晓前冻结。事后用“战略价值”解释一个未达财务门槛的项目，也许是合理决策，但不应改写成模型预测准确；应记录战略目标、评价规则和授权者，以便评估目标是否达成。

24. 预测随市场变化更新时，保留时间边界和失败记录

24.1 预测版本需要在新信息到来前存档

市场扫描不是一次性测量。新法规、竞品上市、原料变化、客户预算调整和供应中断会改变需求及可服务条件。更新预测可以改善运营，但如果每次预测都覆盖上一版本，团队无法判断哪个信息在何时改变了结论，也无法做公正回测。每次预测需保留生成时间、信息截止日、检索范围、模型/提示版本、候选定义、价格与供给假设、概率分布和建议动作；以后只能在新版本中更新，不应回填旧预测。

若训练数据包含预测截止之后的销售、市场评论或竞品网页，回测可能泄漏未来信息。若后续才知道某项目成功，团队再从项目当时公开材料中挑选支持性文本，便会产生事后叙事。时间切分要以当时可见信息重建，而不是以当前能搜到的信息重新生成“历史预测”。对网页存档或新闻时间戳不可靠的来源，应将资料记录为可能回溯修订并降低其证据等级。

24.2 更新频率应取决于决策事件，而非模型刷新按钮

高频更新不等于更好的决策。若目标市场的采购预算一年才审一次，每日刷新公开声量可能只制造噪声；若监管或竞品发布能够在几周内改变产品可行性，较低频更新则可能错过关键窗口。更新周期应与价格调整、投标、认证、样机试

SVSAC | sv-sac.org

验、采购会议和供应条件变化绑定。每次更新时间都需写出可能触发重新决策的事件，且说明何种变化只是信息更新、何种变化会改变产品组合动作。

模型漂移也需要区分数据分布变化与关系变化。搜索量下降可能是平台算法或术语变化；竞争品页面突然增多可能是数据抓取覆盖改善；客户同一问题文本变少可能表示问题解决，也可能是投诉转移到私域。若没有独立监测来源，把文本频率的下降直接读作需求降低，可能将测量断点误报成市场转折。应将来源覆盖、模型抽取稳定性和业务结果分开监测，至少用固定基准样本重复运行来发现抽取规则变化。

24.3 反馈标签受到决策影响

预测会影响哪些产品被开发、谁被联系、哪些市场投放资源，进而影响可观察结果。模型推荐的项目得到更多广告和工程支持，销量提高后再次被训练系统视为“需求高”，形成政策反馈；未推荐项目没有上市，其需求标签缺失。只用已经发生的销量更新模型，会强化旧决策带来的选择，而不是自动学习整个候选市场。市场信号的生成过程与公司行动共同决定了标签。

可以保留小比例安全、合规、成本可控的验证资源给未入选项目，或使用分层抽样审查来减少完全由模型分数决定的结果观察；但这会增加验证成本，需要独立预算和停止条件。若出于资源不足没有机会验证某些项目，应在数据中保留“未被观察”标记，不将其视作零需求。在线学习或自动重排项目前，需记录哪些数据受模型建议影响、哪些来自外生变化，并在独立时间窗/市场簇中确认更新后策略。离线预测精度提升不证明新政策对组合利润有正向因果效果。

24.4 重新评估应允许撤销原判断

更新制度如果只允许增加信心、不允许下调或撤回项目，模型会变成持续投资的正当化工具。对每项关键假设，可设可观察的反证：目标客户缺少预算窗口、接口适配率低于预设范围、认证时间超过上市期限、竞品推出可比功能或客户试用不转成付费。反证发生时，决定可以是改区域、改规格、推迟、分拆、退出或只保留平台研发，不必继续维持“原市场预测正确”。

评估回测应同时报告上调、下调和撤销的比例、误报机会成本、漏掉后被证实的机会、预测校准和组合结果；只报告“发现的新机会数”会奖励过度生成。对小样本产品线，年度误差波动很大，单次预测变化不适合被解读为稳健学习；应跨预先定义的多期/多类任务积累证据，同时考虑产品更新、竞争结构和来源体系变化。最终能被解释的不是模型是否更会生成市场故事，而是更新是否让组合决策在固定约束和明确目标下更可靠。

SVSAC | sv-sac.org

25. 证据综合：AI 不是组合决策改善的唯一解释

25.1 不同研究回答的是不同层次的问题

现有证据不能被简单排成“弱证据到强证据”的单一序列，因为它们估计的对象并不相同。P&G 现场随机实验比较的是特定企业工作坊中不同任务协作条件对早期概念质量的影响；专利景观描述某类技术活动在已公开专利家族中的分布；概念研究观察人们对刺激物的反应；组合理论模型比较一组明确规则在给定假设下的筛选表现；本次分析提出的全球信号扫描则还需要经过产品定义、工程约束、渠道、采购和交付才能形成实际商业结果。前一层的证据可以作为后一层的机制线索，不能自动替代后一层的结果观测。

比如，信息检索覆盖增加可能让团队发现以前未纳入的可替代品；这会改变候选集合。更好的候选集合仍不保证团队对客户价值判断正确。即便需求判断较准，工程认证和售后能力也可能使目标市场不可服务。即便某项产品获准上市，销量也可能主要来自内部产品替代，而非组合净增长。因而“AI 提高市场分析效率”“AI 提高新品质量”“AI 提高新品收入”和“AI 提高产品组合利润”是不同命题，须各自指定分析单位、分母、时间窗和反事实。

来源 1 的随机化证据适用于工作坊内概念任务的近端结果，并没有追踪上市、现金流或长期组合贡献。来源 12 则说明在其明确的模拟条件下，聚合规则会影响预算受限的选择质量；这不是制造企业部署 AI 的现场结果。将两者结合可以提出“概念生成与组合筛选需要分别检验”的假说，却不能将一项研究的盲评得分和另一项模型的模拟收益拼成已经实现的端到端效应。证据链中的空缺本身就是研究结果：目前不能从已读材料得出 AI 全球扫描带来利润提升的通用结论。

25.2 一项非 AI 随机研究说明流程能力可能是替代机制

判断 AI 是否是改进来源，必须面对一个容易被忽略的竞争解释：团队可能因为开始明确写出可证伪假设、收集相反证据并及时停止项目而表现更好；工具只是同时进入了这套管理改变。Camuffo 等人的大型复制与扩展研究提供了相关但边界明确的对照。研究整合四项随机培训试验，涉及 759 家企业，其中包含原始 116 家企业的试验样本，不能把 759 解释为 759 家新的独立复制企业。各试验设置在米兰、都灵和伦敦；实验组与对照组都接受创业训练，但实验组增加了科学假设形成与检验的训练。作者报告组合样本中的项目终止差异约为 9.8 个百分点，并以累计收入等变量分析后续表现。研究对象是创业企业的决策训练，不是 AI 或成熟制造企业的新产品组合；收入不是利润，研究也没有分别随机化每一种

SVSAC | sv-sac.org

行为机制。本篇只将它用于说明“结构化检验可以改变决策结果”是一条需要与 AI 信息增益区分的竞争机制，而不是把其效应量套用到本篇场景。[15]

这个对照改变了实验问题的表述。若 AI 组同时获得更清晰的假设记录、更多研究人员或更严格的评审，比较回答的是一套资源与流程政策是否有效，而非算法的信息处理是否有效。反过来，如果只要求两组用同样的流程，AI 最可能影响的是信号发现、资料整合或证据追溯，而其作用可能小于“引入结构化验证”本身。两种政策都可有管理价值，但需要不同的估计对象和成本核算。

Camuffo 等报告的项目终止增加也不能被直接解释成“筛掉的项目更差”。终止行为的因果变化与筛选质量是两个层次：如果项目本来就不可行，及早终止可能节省资源；如果终止的项目其实有高价值，增加终止率则可能有害。要判断方向，需要对未继续项目保留一定独立信息，或用后续可观察结果、部分识别界限以及明确的机会成本规则，而不能把“更多停止”本身当成正确决策。该研究提供流程方法可能改变终止与转向模式的证据，但没有替本篇证明任何一项被放弃新品的反事实收益。

25.3 将影响拆成信息、程序和执行三条路径

可检验的机制图至少应分开三条路径。第一条是信息路径：系统增加此前未见的来源、减少重复资料、发现地区差异或揭示竞争替代，使团队更准确地描述状态。第二条是程序路径：系统促使团队把主张写成可检验假设、记录反例、区分事实与推断，从而改变会议讨论和停止规则。第三条是执行路径：团队据此改变访谈、试制、渠道投入、价格或上市时点，进一步影响市场结果。三条路径可以同时存在，也可能相互抵消。例如更全面的信息暴露出市场机会，但新增证据的复核成本耗尽了工程测试时间；更严格的程序提高了退出速度，却错过了高不确定、长周期的基础创新。

把每条路径都当作中介变量并做统计分解，需要额外假设。若“信息质量”是在分配后由团队记录的，它可能同时受任务难度和参与程度影响；若团队自选采纳模型输出，则采纳者可能本来就更重视分析。简单回归“结果对采纳量”不会识别采纳的因果效果。较稳妥的第一步是随机分配政策包，并将可观测的资料覆盖、假设质量、独立验证结果、资源消耗与组合终点按时序记录；再用预先设计的因素随机化或中介实验区分机制。若无法随机化某一环节，应把机制分析标为探索性并说明可能混杂。

一项端到端评估还要记录失败路径而不只记录成功项目。包括：在候选池进入前被过滤掉的方向、因数据缺失而无法评分的方向、通过内部评审却未获得客

SVSAC | sv-sac.org

户试验的项目、试验后未上市的项目，以及已经上市但因渠道或供货中断而观察不足的项目。每个状态转换都可能由政策影响，因此只分析最终上市项目会造成处理后选择。候选总体和各状态的流转率应与主效应同时报告，才能判断所谓“更高成功率”究竟来自更好的识别，还是只来自把困难项目排除在分母之外。

26. 识别 AI 信息增益与管理流程改变

26.1 先定义可区分的处理因素

若核心问题是“AI 资料整理是否增加信息价值”，AI 处理与决策程序应尽量可分。可提出一个 2×2 构造试验：信息条件为常规资料包或 AI 辅助资料包；程序条件为现行评审或结构化假设检验程序。四个臂共享同一候选总体、预算上限、目标客户范围、评审时间、技术资料和客户验证额度。AI 资料包记录模型版本、检索来源和生成时间；结构化程序要求团队先写出需求命题、可反证预测、证据等级和停止条件。没有 AI 而采用结构化程序的组，能够估计管理纪律改变的作用；使用 AI 但沿用现行程序的组，能够观察信息工具在既有组织流程中的增量；两者结合的组能检验互补或替代关系。

对每个预先冻结的团队簇，设 A 为是否获得 AI 资料包，B 为是否接受结构化检验程序，Y 为固定观察窗内的团队结果。一个简单的意向处理模型可写为：

$Y = \alpha + \beta A + \gamma B + \delta(A \times B) +$ 区组固定项 $+ \varepsilon$。

β 不是所有条件下 AI 的“固有效果”，而是当 B=0 时 AI 与常规资料的差异；β+δ 才是当 B=1 时的 AI 差异。γ 是没有 AI 时程序变化的差异；δ 表示两种处理的互补或替代。若随机化充分、结果完整且分析遵守最初分配，交互可用于检验分配政策包之间的异质效应；它仍不自动说明 AI 通过哪项中间行为产生结果。若实际采用受到分配后选择影响，按采用者重算不再是随机比较。

26.2 一组构造结果显示为何不能只看“AI 组对照组”

以下均为假设数字，单位为每个团队在固定预算下的盲态证据质量评分，不是企业数据。常规资料、现行评审的均值为 60；AI 资料、现行评审为 66；常规资料、结构化程序为 68；AI 资料、结构化程序为 76。AI 在现行程序下差 6 分，在结构化程序下差 8 分；程序在无 AI 时差 8 分，在 AI 条件下差 10 分。差异中的差异为(76−68)−(66−60)=2 分，提示两种处理在这个构造中互补。若研究者只比较 AI+结构化组与常规+现行组，会得到 16 分差异，却把信息工具、程序改革及交互全部混在一起。

SVSAC | sv-sac.org

评分只是近端结果。若这四组的预算、评审时长或样本构成不一致，均值差可能来自资源差异；若盲评者能从文本风格识别 AI，结果测量也可能受预期影响。评分提高也不等于更多项目通过客户真实付款，更不等于项目组合贡献上升。因此同一设计应并列记录固定分母下的信息错误、反证捕获、客户验证启动、工程时间和项目决策，再将长期结果作为成熟后续指标。近端与远端结果可以方向不同，不能用一个复合总分掩盖这种冲突。

26.3 随机化单位必须跟随共享信息和资源边界

如果同一评审团队负责四个条件，团队成员会把 AI 检索策略、假设模板和竞品清单带到常规组，造成交叉污染。若四组项目争用同一名行业专家、同一认证实验室或相同客户账户，则一个项目的处理会改变另一项目获得的资源，构成干扰。把项目行当作独立样本会低估不确定性，也会把组织层面的政策效果错算成项目层面的处理效应。

可行的分配单位是共享关键资源的团队、事业部或预定义的项目簇；区组应尽量依据基线项目类型、地区、风险等级和资源容量建立，并在区组内分配。若可分配集群数量很少，就应限制统计主张，报告随机化分布或设计支持的区间，不用大量项目数伪造有效样本量。若企业只能给一个团队使用 AI，研究可采用逐步推广或等待名单设计，但推广顺序必须随机或有可信外生依据，并记录时间趋势和同时发生的组织变更。若无法随机分配，比较应降为观察性方案，明确不可消除的选择偏差。

公司如果希望直接知道“全面推广的效果”，小范围随机试点仍未必能识别。局部处理时，AI 组可能获得稀缺专家的额外帮助；全面部署后专家会被多个团队同时调用，资料复核和工程资源出现拥堵。局部试验估计的是当前覆盖率下的直接效应，扩张政策还包含拥堵、学习和竞争反应。要向全面部署外推，需要多水平的覆盖率设计、不同团队饱和度或明确资源队列模型，不能把局部平均处理效应乘以公司团队总数。

26.4 采纳、覆盖和推广的三个政策估计量

至少要报告三个互不替代的量。其一是分配效应：被分到 AI 政策包的团队与对照团队之差，衡量在现实不完全采用下的意向处理效果。其二是在相关性、排除限制、单调性和处理定义等条件下，以随机分配为工具识别的依从者局部平均效果；它不是所有实际采纳者的平均效果；若随机邀请同时带来培训和管理注意力，排除限制不成立。其三是全面覆盖效果：在资源和干扰都可能变化的情况

SVSAC | sv-sac.org

下，所有目标团队采用后的系统结果。不同估计量对应不同管理决策，不能用“AI有效”一句话替代。

以采纳率为例，如果随机分配团队中只有一部分实际打开工具，分配效应会被不采用稀释，但这可能正是现实部署的政策结果。管理层可以把可用性、培训和使用门槛纳入部署包，但增加这些要素后，成本和处理定义也改变。若仅报告使用者的较好结果，会受“愿意尝试的人更擅长分析”这一选择影响。按使用量给奖励还可能提高低价值调用；需要观察工具使用和决策质量的关系，却不能仅凭其相关性声称增加使用会提高结果。

27. 从点预测转向组合风险：两个构造性敏感性检验

27.1 相关性会改变组合的下行风险

组合选择常用项目期望贡献的总和，但在预算固定、少数项目占大部分资源时，结果的相关性也很关键。设两个项目在同一市场的净贡献分别为随机变量 X 和 Y，均值各为 10 个抽象单位，标准差各为 8。若两者独立，组合贡献 X+Y 的方差为 $8^2+8^2=128$，标准差约 11.31；若二者完全正相关，方差为$(8+8)^2=256$，标准差为 16。两个方案的预期总贡献都为 20，但方差不同；仅凭均值和方差不能确定任一损失阈值下的共同下行概率。假设两项都依赖同一监管许可、同一关键零件供应商或同一客户资本预算，单项目评分不会揭示这一风险共因。

这只是一个构造计算，不是风险概率估计。真实的协方差需要历史项目、外部冲击或可解释的情景模型；少量已上市项目可能无法识别尾部相关性。把相关系数设为零只加入零协方差假设，不等于独立；即使协方差为零，也不能由此推出尾部冲击相互独立或共同下行概率很小。缺乏依赖数据时，零相关仍是未经验证的输入，不应默认为事实。反过来，假设完全相关也可能过度保守。应报告多个依赖情景下的组合排序、现金缺口和资源峰值，并说明何种决策会在敏感性边界上改变。若排序对一个未经验证的相关系数极为敏感，优先价值可能是获得共因数据或分散实验，而不是继续提高单项目预测精度。

27.2 概率校准不等于行动排序

假设有两项产品机会，A 的失败概率为 0.40，失败损失为 100；B 的失败概率为 0.20，失败损失为 30。只按失败概率，A 看起来风险更高；按预期损失，A 为 40，B 为 6。但若 A 有低成本缓解措施：投入 5 单位可将失败概率降至 0.10，则预期避免损失为$(0.40-0.10)\times100-5=25$；若 B 的缓解成本为 2，将失败概率从

SVSAC | sv-sac.org

0.20 降至 0.05，则避免损失为(0.20−0.05)×30−2=2.5。若两项措施需要同一瓶颈团队，且只能执行一项，按构造的期望损失规则应先处理 A。

但如果 A 风险无法通过干预改变，降低概率的工具只提供更精确的预警，并不创造同样的可避免损失；而 B 的技术修正可能更容易实施。预测模型估计未干预风险的能力与干预后风险下降幅度不同。后者是因果反应，不能从风险分数、相关工单或历史“高分项目被修复”数据直接推得，因为历史上哪些项目得到修复本身受严重程度、团队能力和预算影响。要使用预期可避免损失排序，应在可行范围内对行动随机化或利用可信设计识别效果；否则报告合理区间并保留工程判断，不把风险预测包装成投资回报。

同样，单位工时收益排序并不总是资源分配的最优解。任务不可分、先后关系和共享设备可能使简单的“收益/工时”比率失效。例如每项都要占用整周且只能选一个，两个各需六周的任务无法通过把每项需求按比例拆分来满足容量；在依赖网络中，一个净收益较低的基础项目可能是其他项目的前置条件。预算优化必须写出可行集合、互斥规则、先决关系和目标函数；如果这些没有被观测，算法给出的精确最优解只是对不完整模型的精确求解。

27.3 资源边界下的阈值反转

再考虑一个明确假设的三项目候选集合。项目 A 的期望贡献为 12，消耗 8 个稀缺工程周；项目 B 贡献 10，消耗 5 周；项目 C 贡献 7，消耗 3 周。资源上限为 11 周且项目不可拆。可行组合包括 A、C（贡献 19、用 11 周），B、C（贡献 17、用 8 周），以及单独 A 或 B。总贡献规则下选 A+C。若 A 与 C 同时需要同一资格审查工程师，组合会增加 4 周，因而超出容量；此时 A+C 不能执行，B+C 成为候选的最佳可行组合。若这 4 周审查可由不占上述 11 周瓶颈池的独立闲置资源完成，且不延误截止期，则该项额外冲突消失，A+C 恢复可行；不能仅因同一瓶颈池内有人空闲，就在总需求 15 周、总容量 11 周不变时宣称可行。资源约束的真实机会成本取决于可用容量与被挤出活动，而不是依赖图中出现同一个岗位名称就自动增加成本。

这个例子说明，组合边界的改变可能令排序发生跳跃。决策记录应保留资源日历、需求量、空闲容量、替代项目以及项目不可拆条件；敏感性分析应扫描容量和共同任务持续时间，指出结论在哪个区间改变。只有在具体组织的工程队列和项目可行性数据中估计这些输入后，才能讨论某一企业的组合选择。模拟结果不能被写成“AI 更优”的证据，除非比较的是实际分配的政策并测量了实际后果。

SVSAC | sv-sac.org

27.4 真实选项和退出价值需要非对称计账

早期试验能否降低不确定性，取决于结果是否会改变下一步动作。若无论试验结果如何都必须继续满足既有合同，试验提供的决策灵活性有限；若客户可选择付费试点或终止，且工程团队有可转用能力，试验可能创造退出或扩张的选择价值。将这种价值纳入组合账本时，不能把“未花掉的后续预算”同时当作项目收益和现金流节省，也不能把已经发生的研发沉没成本作为继续投入的理由。

可以把下一阶段的预期价值写为试验后不同结果下最优动作的加权值，减去试验成本和等待造成的时机成本。关键并非公式形式，而是列清每个结果下可选动作、动作对现金/资源/客户承诺的后果及是否真正可逆。若试验结果在两种情况下都导致同一决定，信息价值可能接近零；若结果会把组合从不可行方案切换到可行替代，试验可能有较高价值。未经观测的转向概率、客户响应或退出成本必须以区间或构造情景呈现，不宜用单点精算制造确定性。

27.5 竞争者反应可以改变可捕获贡献，而不改变品类热度

市场需求与企业可捕获需求不能混为同一个数字。新品发布会令品类搜索、媒体讨论或客户询价增加，但竞品也可能同步调整价格、规格、交货期、捆绑服务或经销商激励。即使总品类需求上升，焦点企业获得的单位数和单位贡献仍可能下降。静态市场扫描常以竞品当前产品页或过去价格作为基线，却没有明确“竞品看到我们的进入后会怎么做”的条件假设。

一个完全构造的阈值例：预计品类在观察期内 100 个单位，焦点企业计划获得 35 个单位，每单位贡献 40 个抽象货币单位，固定上市投入为 1000。无竞争响应时，贡献为 35×40−1000=400。若竞争者降价导致焦点企业份额降为 30 个单位，且焦点企业为守住订单把单位贡献压至 30，则贡献变为 30×30−1000=−100。这里没有声称竞争者一定降价；两种结果是情景，不是概率已知的预测。若管理层基于自身历史或可比竞标记录暂把反应情景概率设为 30%，则期望贡献为 0.7×400+0.3×(−100)=250，但仍有一个有损情景。该概率只是演示输入，不能当作由检索量或 AI 生成解释所估出的真实概率。

决策阈值也可写清楚：固定投入 1000 下，单位贡献 40 时需至少 25 个单位才回本；单位贡献 30 时至少 34 个单位。竞争反应若同时压低价格和份额，就会跨越该边界。只预测总品类规模，即使误差很小，也可能错过最相关的边界变量。研究应把需求证据拆成品类量、焦点产品可获份额、价格/折扣、客户获取成本、渠道覆盖和竞品响应，分别标注数据来源与不确定性。若份额只来自概念问卷，

SVSAC | sv-sac.org

不能将其当真实购买份额；若竞品响应只由一次价格网页抓取推断，也不能当因果反应。

在适当市场中，可以记录不同地区或账户的竞品报价变化、焦点企业价格决策、投标输赢和客户替代路径，并预先规定窗口。若产品进入在市场间分期发生，可比较早进入与晚进入区域，但进入时间往往与管理者预测相关；除非时间差有随机或可信准实验依据，估计仍受进入选择和市场趋势影响。若竞争者能跨地区同步调价，各地区就不是独立处理簇，地区比较会低估干扰。最稳妥的早期用途可能是对价格和份额边界做压力测试，而不是给一个确定的“AI 预测市场份额”。

27.6 判断新增信息是否会改变组合动作

把所有不确定性都列出来会产生一张无法执行的风险清单。实际可用的优先规则是：选出有可能令可行组合改变、同时仍可采取不同动作的未知量。对于每项未知量，至少列出当前可行行动、若观测到低值/中值/高值时的行动、获取该信息的现金与资源、等待成本、以及信息到达时项目是否仍可调整。若不同结果下行动完全相同，追加试验可能只提高叙述精度；若不同结果改变是否投入、优先市场或技术路线，信息才有潜在决策价值。

这仍不意味着必须将所有未知转换为货币。安全、质量、客户承诺和法规约束可以是不可违反的约束，而非可用预期收益抵消的负成本。对于这些约束，研究目的可以是证明候选方案是否通过门槛，或明确哪些资料缺失使结论无法通过，而不是估算“违规期望损失”并和销售收益相加。风险敏感阈值应由有权限的业务和工程负责人预先确认，且说明不可接受后果；模型输出只能提供情景，不授予风险接受权限。

实际企业可在组合评审前生成一页决策表：每个候选的可服务市场定义、关键未决假设、证据更新时间、最便宜且能区分假设的验证、如果验证为负的退出动作、共享资源占用、以及适用的客户/工程硬门槛。表中的每项数字应标注是观察、预测还是管理层假设。评审结束后记录实际选择和偏离规则原因。下一周期再比较这些事前预测与成熟后的结果，才能知道模型是否帮助辨别了有价值的不确定性，还是仅令讨论更快收敛。

SVSAC | sv-sac.org

28. 跨市场外推需要重新定义目标总体

28.1 一个“全球均值”可能对应多个不同的决策对象

全球市场分析常把多国数据汇总成一个总分或市场规模，但组合决策通常需要在明确的产品、渠道和时间窗口中分配资源。若一项新品只在部分国家可以认证、由当地经销商销售，或需要不同售后能力，则“全球消费者”的平均偏好并不是可执行的需求目标。总体必须说明目标市场权重来自人口、当前销售、可服务账户、预期利润还是战略覆盖；这些权重会改变排序，而且不是统计模型可以替管理者选择的中性常数。

设两个产品概念 A、B 在市场甲和乙中的假设选择率分别为 A：60%、20%，B：40%、50%。如果试验样本中甲乙各占一半，汇总选择率为 A40%、B45%，B 领先；若企业下一阶段的可服务机会中甲占 80%、乙占 20%，按该业务权重计算，A 为 52%、B 为 42%，A 领先。两套权重都可以在数学上正确，却回答不同问题：前者是样本组成下的平均选择，后者是构造的可服务市场加权预测。除非随机抽样概率、市场规模和可服务性都能从独立资料中确定，不能把样本均值称为全球需求，更不能因样本大的地区自动获得更高产品决策权重。

当“市场”划分本身由语言、渠道、支付方式或法规共同决定时，分层单元也应围绕实际执行差异设计。国家标签不一定是适当层级：同一国家不同渠道可能价格、售后和采购权限迥异；跨国同一工业客户集团可能执行相同的全球规格。外推分析应列出预测对象、候选可达账户、渠道覆盖和资源边界，分别给未覆盖区域留出“未估计”状态。把未观测市场填入模型的相似市场预测，属于额外的迁移假设，应单独检验。

28.2 分层估计不自动修复样本选择

假设访谈样本主要来自愿意参加线上调研的客户，按国家重加权只能调整已测量的国家构成；它不会自动恢复未出现的渠道客户、采购经理或缺少英语能力的用户。若客户是否进入样本取决于产品兴趣、数字成熟度或公司关系，即使国家内样本数很大，选择偏差也可能存在。加权需要知道样本进入概率或有可信的选择机制变量；如果某类目标客户进入概率为零，权重没有定义，无法仅靠模型补回。

可把外推拆为几个逐步条件：目标市场中哪些单位具有资格；各单位进入扫描、访谈或概念测试的概率；处理条件是否在不同市场具有相同含义；结果的度量是否可比；在市场间是否存在共享供应、渠道或竞争干扰。每一项都可能失败，且失败后处理方式不同。资格不清时，先定义可服务总体；抽样概率未知

SVSAC | sv-sac.org

时，报告样本条件结果或敏感性范围；处理含义不一致时，拆分成市场特定政策；结果不可比时，不汇总均值；存在跨市场干扰时，估计局部政策而非独立单元效果。

一项定向的外部验证可在开发前冻结模型和规则，选择一个未参与开发的市场簇作为时间外/地域外确认集。先做映射核验：来源是否能稳定取得、单位/价格币种是否统一、关键产品属性是否有本地含义；随后盲态评价同一类需求命题及其反例；最后在可服务机会中比较后续行动或客户行为。若外部市场需要重新训练映射、改写特征或改变评分阈值，结果应标为适配后的政策，而不是原模型的直接复制成功。若市场数量过少，留一市场验证只能给有限稳健性证据，不能代表所有全球地区。

28.3 部分汇聚与完全分组的决策代价

在市场样本稀疏时，完全分组会导致各地区估计噪声很大；完全汇总则会抹掉真正重要的异质性。层级模型的部分汇聚在某些条件下能折中，但它依赖市场参数来自可交换或近似共同分布的假设。对于政治、法规、基础设施、购买力和售后体系明显不同的市场，统计上的相似性不保证商业机制可迁移。若使用部分汇聚，须报告超参数、市场间变异和预测区间，并比较完全汇总、部分汇聚、完全分层等替代模型对决策的影响。

一个构造的两市场排序例已显示简单平均可以因权重反转；再加入估计误差后，决策的不确定性还取决于市场内样本量与市场间异质性。若 A 和 B 差异小于市场间不确定区间，强行给出单一“全球赢家”会夸大排序精度。可先定义最小决策差异：小于该差异时，组织视为并列并通过区域试点解决；超过差异且在权重敏感性范围内仍同向时，才考虑统一推广。这个边界应由预算、切换成本和错配后果决定，不能从模型标准误自动推导。

28.4 市场权重变化带来政策而非纯统计变化

市场权重并非静态样本参数。公司在某个地区投放渠道资源，会扩大该地区可观察账户和订单；撤出服务能力又会减少可实现需求。若模型用历史销售权重代表目标市场，过去的资源分配会通过权重继续影响未来组合。要评估新政策，应预先固定政策比较使用的权重，并另做权重变化敏感性；若部署本身改变可服务市场规模，则需把渠道投资和产品分析作为联合政策来评价。

报告可以并列展示三种视角：固定当前可服务集合、按管理层事先批准的战略市场组合加权，以及在各市场权重变化范围内的最坏/最好情景。若选择只在一种权重设定下成立，决策条件就应明说。若权重由尚未签约的经销商覆盖预测得

SVSAC | sv-sac.org

出，要把该预测作为模型输入的不确定性，而非已实现渠道事实。对政策变化后的结果，也需避免把“进入更多市场后看到更多销售”误解为同市场内产品需求提升。

28.5 对未覆盖市场给出部分识别区间

缺少某一市场的观察，不代表全球比较必须立即停止，但缺失部分不能由模型的单点预测悄悄填满。可先在合理量尺上给出结果界限。以下仍是构造算例：三个市场的目标权重分别为 50%、30%、20%。前两个市场里，概念 A 的行为选择率为 60%、40%，概念 B 为 50%、30%；第三市场没有可用的行为数据。只把已观测市场的贡献按权重相加，A 为 0.5×0.60+0.3×0.40=0.42，B 为 0.5×0.50+0.3×0.30=0.34。因为选择率只能在 0 和 1 之间，第三市场对 A−B 差异的贡献最宽可落在−0.20 与+0.20 之间。因此全球 A−B 差异的界限为 0.08−0.20 至 0.08+0.20，即−0.12 至+0.28。现有证据不足以确定全球赢家，尽管两个已观察市场都偏向 A。

如果有独立工程理由认为第三市场中 A 相对 B 的差值不超过±0.10，其加权贡献仅为±0.20×0.10=±0.02，全球界限缩窄为 0.08±0.02，即+0.06 至+0.10，已能确认方向；若先获得有效数据，证据支持 A−B 差异在该市场落入+0.02 到+0.10 的范围，则全球差异为 0.084 到 0.10，方向同样稳健。这里的数值范围必须由可核机制、历史数据或试点设计支持；不得因为模型对未观测市场有很高置信度就把宽界限替换成窄界限。部分识别的作用不是给出伪精确点估计，而是明确哪些未知输入足以翻转决策。

对不同决策目标，缺失市场的结果界限还可能不同。若目标是当地付费客户数，则只观察选择率仍不足以推断客户数量，还需可服务账户总量、渠道覆盖和购买周期；若目标是组合贡献，则还要考虑价格、成本、退货、内部替代和工程占用。量尺应该直接连接所选决策结果，不能先界定态度评分再把它默认为利润范围。上面的 0 到 1 边界只适用于构造的选择率指标，不适用于现金贡献。

28.6 把确认样本资源投向可能改变决策的未知项

当总预算只能验证一两个市场时，优先抽取样本量最大的地区不一定有价值。应先问：哪一未知项会使候选排序或资源配置改变？若一个未覆盖市场的权重很小、即使结果在全范围两端也不改变现行决策，新增调查价值有限；若该市场权重高，或合理结果范围会令 A/B 排序翻转，直接验证可能比继续细化已覆盖市场的评分更有信息价值。

SVSAC | sv-sac.org

可以用完全透明的离散决策表进行判断。假定只有 A 或 B 可获得下一年度资源，并把目标总体规范为 100 个机会单位；已观察 80 单位贡献差为+8，未覆盖 20 单位的每单位结果差介于−1 与+1，因此其总贡献差为−20 至+20，总体差为−12 至+28 个标准化机会单位，均非实测货币或销量。若先验选择 A，而在第三市场最差情形下选择 B 可避免损失 12，市场试验的潜在价值取决于试验能否把这段区间缩小到不会跨零、以及结果出来前是否仍可调整产品计划。若结果要在模具或认证不可撤销后才成熟，信息即便准确也可能来得太迟。试验成本、时机损失和行动可逆性都应进入选择，而不是只按统计精度分配样本。

实际的试验样本量不能由这个玩具算例推出。需先指定主要结果、最小有意义差异、市场内方差、集群结构、失访、簇间干扰和多重比较规则；若关键市场不能随机抽样，只能报告样本条件结果或敏感性范围。用模型合成的市场回答可以帮助列出待测属性，却不能替代当地采购行为。若没有足够资源使全球差异在当前决策期限内可识别，经营者仍可选择一个有边界的试点或维持可逆方案，但应把它表述成风险管理决策，不宣称模型已经验证了市场潜力。

29. 决策建议与适用边界

一个可用的新品矩阵不是经过统一全球模型排序的产品清单，而是明确列出哪些需求命题有独立证据、哪些产品可在当前能力边界内交付、哪些组合选择依赖共享资源或市场权重、以及哪些关键假设尚未得到反证。AI 可能提高资料整理、归纳和候选扩展的速度，也可能使团队更快地复制同一错误、把不可比市场压成单一分数，或产生难以复核的来源链。二者的相对发生率不能凭模型流畅度推断。

因此，有限的产业启示是把 AI 放在可以被验证和撤销的环节：保留来源与时间版本、为重要主张登记反例、在候选集合固定后比较筛选政策、对未上市项目避免伪造销售标签、在组合层披露共同风险和可服务权重。组织若不能提供这些基础记录，最诚实的产出可能是更快地形成待检验问题，而不是更高的需求预测准确率。对利润和投资回报的结论，应等待企业层的固定分母比较、成熟结果和真实资源账本。

本篇研究的问题不是 AI 能否生成更多市场故事，而是 AI 辅助的全球市场与竞品分析是否能让组织形成更可检验的需求判断，并进一步改善受资源、工程和渠道约束的新品组合。现有证据支持若干有限判断：数据源的范围、重复依赖和时间版本会改变信号含义；概念反应、客户行为、工程可行性和组合商业结果是

SVSAC | sv-sac.org

不同结果层；预算与共享资源会改变单项目排序的价值；项目上市后的反馈还可能受到先前分配政策影响。上述判断足以支持更严格的验证设计，不能支持已实现企业收益的断言。

最有信息价值的后续研究应先回答识别问题。第一，比较 AI 资料处理与结构化假设检验流程，避免把管理纪律的变化全部归因于模型。第二，冻结候选总体和统一资源规则，避免只在被选中或已上市项目中计算成功率。第三，记录客户验证的抽样概率、失访、未成熟事件及预算/工程资源的实际占用。第四，检验组合在不同覆盖率下的拥堵和共因风险，避免把小范围直接效应当成全面部署效应。第五，保留预测和证据的时间版本，使需求标签不被事后信息污染。若数据无法支撑某一层结论，应停留在其可识别的近端结果。

对管理者而言，AI 更适合作为可审计的证据整理与候选假设工具，而不是替代客户选择、工程门槛或组合授权。只有当它改变了可观察的决策过程，且在同等预算、样本和执行条件下改善了预先指定的结果，才有理由说它贡献了增量价值。若收益来自更严格的假设检验、额外研究资源或组织重新分配，应该按真实政策包评价并计入成本。若错误方向无法从当前信息中识别、尾部风险对相关性假设敏感，或关键项目结果仍未成熟，负责任的结论应是尚不确定，而非用更丰富的预测文本填补证据缺口。

29.1 如何解释“没有改善”

若 AI 组与对照组在组合结果上没有可分辨差异，结论不应立刻写成“AI 无效”或“研究失败”。先查看预先指定的处理是否真正改变了信息可用性：来源覆盖、去重、事实核验时间、盲态评审对关键主张的准确性和不确定性；再查看这些近端变化是否改变了团队提出的可证伪假设、客户验证或项目选择；最后检查样本量、项目成熟度、干扰和结果缺失能否排除有管理意义的差异。近端结果没有变化，说明该实施方式没有产生预期信息差异；近端变化存在而项目政策不变，可能是组织流程或决策权没有响应；政策变化但终点不变，则可能是反事实结果接近、周期不足或机制无商业价值。每种解释都需要不同的后续设计。

反之，若候选文本更完整、评审分更高但真实客户动作和组合贡献没有改善，也不能仅以“长期收益还没到”无限期延迟检验。应事先指定哪些近期事件可作为合理中间证据、最终端点何时成熟，以及到期仍未成熟时如何处理。若置信区间仍包含较大正效应和较大负效应，应报告不确定；若设计能排除达到最低管理意义的改善，则可以有边界地说未观察到足以支持推广的增量。决定是否继续

SVSAC | sv-sac.org

投入要结合部署成本、可逆性和替代项目，不能把统计不显著等同于零效果，也不能把方向正确的点估计包装成有效。

独立复制也应复现决策定义和资源环境，而不只是重新调用一个同名模型。若跨地区复制需要重训、人工整理、额外外部数据或重新定义产品目标，实际被复制的是新政策包，应重做成本及结果评估。这样，负结果可以指出效果链在哪一环断开，正结果也能说明在哪些条件下可转移。本篇没有公司 CRM、项目级成本或真实上市结果；因此关于 AI 提高新品组合回报的结论目前仍是待识别的经验命题。

给企业的下一步建议

新品矩阵需要有取舍的前瞻判断；AI 更适合扩大信息视野、暴露盲点，不能代替企业决定押注何处。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：把市场需求、竞品变化、技术可行性和企业能力放在同一组合中比较，以小步验证调整研发优先级。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，进一步形成全球市场与竞品图谱、新品组合建议、研发投入路线图。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。

生成式人工智能降低了跨语言检索、资料归纳和概念生成的成本，但“发现了更多信号”并不等于“更准确地识别了需求”，更不等于新品上市后取得销售或利润。本篇讨论 AI 辅助全球市场与竞品分析何时可能改善新品组合决策。核心研究问题是：与既有情报与产品规划流程相比，AI 介入的信号搜集、需求解释和概念排序，能否使固定目标市场中的产品组合假设更可验证，并最终提高实际采用与风险调整后的组合价值？

文章采用定向一手证据梳理与机制分析，不含企业 CRM、销售数据库或实际市场预测试验。宝洁专业人员现场实验随机比较个人/团队与有无生成式 AI，测得的

SVSAC | sv-sac.org

是一日产品创新任务中盲评方案质量等近端结果；WIPO 2026 报告描述的是更新至 2025 年的已公开生成式 AI 专利家族。两类证据分别说明 AI 可能影响构思质量，以及专利活动可以形成技术趋势信号；它们都不能证明 AI 全球市场规划提高了新品销量、毛利或市场份额。[1][2] 本篇据此把“市场情报能力”拆成可追溯信号、需求假设、技术可行性、组合选择与实际采用五个环节，逐一说明其可观察指标、替代解释与证伪条件。提出的测量框架和研究协议均为作者分析，未在企业实施。结论不把生成量、专利热度或专家评分当作商业成功的替代变量。

参考文献

[1] Fabrizio Dell’Acqua; Charles Ayoubi; Hila Lifshitz; Raffaella Sadun; Ethan Mollick; Lilach Mollick; Yi Han; Jeff Goldman; Hari Nair; Stew Taub; Karim R. Lakhani. The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork. Organization Science, 37(4), 1217–1242 (2026). https://pubsonline.informs.org/doi/10.1287/orsc.2025.20702

[2] World Intellectual Property Organization (WIPO). Patent Trends Update in GenAI. Technology SPARK Report Series, WIPO Publication No. 1089.3/26 (2026). https://www.wipo.int/edocs/pubdocs/en/wipo-pub-1089-3-2026-en-patent-trends-update-in-genai.pdf

[3] Oliver Schaer; Nikolaos Kourentzes; Robert Fildes. Demand forecasting with user-generated online information. International Journal of Forecasting, 35(1), 197–212 (2019). https://www.sciencedirect.com/science/article/abs/pii/S0169207018300505

[4] Google Trends Help Center. FAQ about Google Trends data. Google Search Help, live official documentation (accessed 2026-10-23) (2026). https://support.google.com/trends/answer/4365533?hl=en

[5] Kerry Liu. The measurement errors of Google Trends data. Discover Data, 2, Article 7 (2024). https://link.springer.com/article/10.1007/s44248-024-00013-3

[6] Armin Falk; Anke Becker; Thomas Dohmen; Benjamin Enke; David Huffman; Uwe Sunde. Global Evidence on Economic Preferences. The Quarterly Journal of Economics, 133(4), 1645–1692 (2018). https://academic.oup.com/qje/article/133/4/1645/5025666

[7] Joel Huber; Christopher Puto. Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects. Journal of Consumer Research, 10(1), 31–44 (1983). https://academic.oup.com/jcr/article/10/1/31/1833486

[8] Robert J. Thomas. Industrial market segmentation on buying center purchase responsibilities. Journal of the Academy of Marketing Science, 17, 243–252 (1989). https://link.springer.com/article/10.1007/BF02729816

[9] Boris Durisin. ‘U can’t touch this’: Stimuli design decisions and prototype trial in concept testing. Marketing Trends Congress 2016 proceedings archive / author conference paper (2016). https://archives.marketing-trends-congress.com/2016/pages/PDF/DURISIN.pdf

SVSAC | sv-sac.org

[10] Frank M. Bass. A New Product Growth for Model Consumer Durables. Management Science, 15(5), 215–227 (1969). https://pubsonline.informs.org/doi/10.1287/mnsc.15.5.215
[11] Glen L. Urban; John R. Hauser; John H. Roberts. Prelaunch Forecasting of New Automobiles. Management Science, 36(4), 401–421 (1990). https://pubsonline.informs.org/doi/10.1287/mnsc.36.4.401
[12] Lucas Böttcher; Ronald Klingebiel. Organizational Selection of Innovation. Organization Science, 36(1), 387–410 (2025; first published online 2024-09-19). https://pubsonline.informs.org/doi/10.1287/orsc.2023.17357
[13] Sundara Raghavan Srinivasan; Sreeram Ramakrishnan; Scott Erwin Grasman. Identifying the Effects of Cannibalization on the Product Portfolio. Marketing Intelligence & Planning, 23(4), 359–371 (2005). https://doi.org/10.1108/02634500510603465
[14] Sreekumar Bhaskaran; S. Sinan Erzurumlu; Karthik Ramachandran. Sequential Product Development and Introduction by Cash-Constrained Start-Ups. Manufacturing & Service Operations Management, 23(6), 1505–1523 (2021; first published online 2020-10-05). https://pubsonline.informs.org/doi/10.1287/msom.2020.0895
[15] Arnaldo Camuffo; Alfonso Gambardella; Danilo Messinese; Elena Novelli; Emilio Paolucci; Chiara Spina. A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. Strategic Management Journal, 45(6), 1209–1237 (2024). https://openaccess.city.ac.uk/id/eprint/32437/

SVSAC - sv-sac.org