# 智能化产品升级：把 AI 嵌入产品，形成新卖点与高毛利

**作者：SVSAC 硅谷战略顾问委员会**

## 核心判断

制造企业应把 AI 功能当作一项产品投资来审视；客户愿意持续付费的体验，才是高毛利的基础。对企业而言，值得优先推进的是：围绕一个高价值使用场景，测算开发、算力、维护和售后成本，验证功能使用率、付费意愿及服务负担。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**：嵌入式人工智能；产品功能；客户价值；全生命周期成本；毛利率；因果识别

SVSAC - sv-sac.org

制造企业应把 AI 功能当作一项产品投资来审视；客户愿意持续付费的体验，才是高毛利的基础。

我们的建议是：围绕一个高价值使用场景，测算开发、算力、维护和售后成本，验证功能使用率、付费意愿及服务负担。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“模型能跑”到“产品有利润”之间有几道门槛

产品宣传中的“AI 功能”往往把多个不同性质的结果压成一个标签：芯片上运行了模型，用户看见一个智能按钮，销售页加了一项新卖点，财务报表中的毛利率随之上升。这个叙述跳过了关键中介，也把供给端的技术实现和需求端的支付行为视作同一件事。本篇的核心问题不是企业能否把模型部署到设备，而是：相对于一个功能等价且可量产的非 AI 基准，嵌入 AI 是否带来可验证的使用改善，改善是否改变了真实购买行为和可实现净售价，并在完整成本和销量反应之后提高单位贡献或毛利率？

这里至少有五类结果变量。第一是技术结果：模型在指定硬件、数据和版本下的准确度、时延、峰值内存、通信量或能耗。第二是产品功能结果：用户要完成的任务是否更快、更稳、更少误操作，且系统在异常情况下是否能拒绝或安全退回。第三是用户评价：消费者是否理解功能、是否觉得有用、信任或愿意尝试。第四是市场结果：在明确价格、产品配置和渠道条件下，是否购买、退货、续购，或接受可持续的净价差。第五是经济结果：新增实现收入扣除相关变动成本、生命周期服务与必要投资后，单位贡献、总贡献或毛利率是否变化。它们之间存在因果顺序，却不保证单调传递；技术测试很好可以与使用收益微弱并存，感知新奇也可先提高尝试但不形成留存。

单台毛利额和毛利率尤其容易混用。若一台 AI 版本售价更高但增加传感器、计算、质保与云支持成本，单位差额可能增加而差额率下降。即使两者都提高，总贡献仍会因销量下降、退货上升或高固定研发投入而变差。反过来，毛利率未提高也不必然说明项目没有价值：产品可能扩大销量、降低其他产品成本或增加用户留存；这些应以明确对照和时间范围检验，而不能用“战略价值”代替量



SVSAC | sv-sac.org

化或边界陈述。本次分析把“卖点”“高毛利”拆开，是为了防止把预期效应当作已实现结果。

这一问题还要求确定非 AI 基线。恰当比较应保持核心功能和服务承诺尽量一致，例如传统阈值控制与嵌入式声学事件识别都实现“设备检测到异常后提醒用户”；若 AI 版同时新增屏幕、传感器、自动联网和延长保修，而对照版只有基本硬件，任何价格或满意度差都无法归因于 AI 本身。工程上完全相同的对照未必可得，因而比较目标要写清：评估“AI 算法组件的增量效果”，还是“AI 产品包相对于现有产品的增量价值”。前者需要将其他功能锁定或因子拆分，后者则是现实产品决策，但结论只适用于整个组合包，不能单独归功于模型。

# 2. 核心概念与分析边界

## 2.1 嵌入式 AI 不是单一部署位置

“嵌入产品”在本次分析中指模型推理、学习或决策功能作为交付产品可使用的一部分，直接参与用户任务或产品性能；它可能在微控制器、手机伴侣、网关、云端或这些位置的组合中执行。产品界面里出现 AI 标识但后台仍由人工服务完成，未必改变产品功能；反之，用户未看见“AI”标签的设备也可能运行学习模型。研究单位应是可测试的功能配置与用户可用权利，而非营销用语。

端侧、云端和混合架构是对服务路径的选择，不是价值高低排序。端侧处理能减少部分原始数据上送、缩短网络往返并支持断网运行，但受制于存储、计算、功率、散热和升级能力；它仍可能通过遥测、诊断日志、配对应用或固件更新暴露信息。云端可以集中更新较大模型、统一管理计算资源，但引入联网依赖、服务连续性、带宽成本和持续运营责任。混合架构可以把低延迟触发留在设备、复杂分析放在云端，却增加版本兼容、状态同步、故障定位和用户对“何时联网”的理解难题。任何商业对比都应把原始传感数据、派生推断、模型参数、运行日志、更新包和控制指令分别画出流向，避免把“数据留在本地”简化为完整的隐私结论。

## 2.2 产品价值不是模型分数

本篇将功能效用定义为：在预先指定的任务、产品配置、用户群和环境范围内，AI 产品相对于功能等价基线带来的可观察结果差异。结果既可为连续量，例如完成时间或能耗，也可为事件量，例如每千次使用中的误触发、漏报、退货或人工干预。评价必须沿用真实暴露分母；把实验室中平衡抽样的准确率直接解释成家庭中每月误动作次数，会因事件基率和使用频率不同而产生数量级偏差。



SVSAC | sv-sac.org

感知价值是用户对功能效用、便利、控制感、可理解性、风险和新奇性的主观判断；支付意愿是受访者在给定描述或选择情境中的选择，未必等于实际支付。观察到的交易价还受促销、渠道抽成、捆绑、保修和供货限制影响。净售价需扣除折扣、返利和退货；可捕获价值还取决于消费者愿付的金额如何在品牌商、平台、经销商、供应商之间分配。于是，模型得分、感知有用性、调查中的溢价和实际净价分别是不同测量对象，不宜放进同一“AI 价值”指标。

利润口径也需明示。本篇用“单位贡献”指指定期间内净销售收入减去明确列入该决策账本的单位变动成本，用“毛利额/率”指依适用会计政策确认的收入和销售成本形成的财务指标。前者适合比较产品方案的增量经济性；后者受成本归类、质保计提、服务履约与收入确认政策影响。研发支出是否费用化或资本化、云支持是否进入销售成本、退货准备是否按财务政策计入，不能由研究者假设成全行业统一。本篇展示的构造账本因此是管理决策比较，不是会计意见或企业毛利报告。

# 3. 核心问题、因果结构与方法

## 3.1 可检验问题

本篇将总问题拆解为五个可以被证伪的命题。第一，在给定目标任务与环境下，AI 功能是否相对非 AI 基线降低错误、时间或资源占用，而不把风险转移给用户？第二，功能改善是否在真实使用频率和可理解的产品说明下改变用户体验，而非仅因“AI”标签或新奇感产生短期评价？第三，购买概率、净售价、退货和续购是否发生变化，并且这种变化是否在合理对照中归因于功能或产品包？第四，系统和服务成本如何随销量、在用设备数、支持年限与版本更新而演化？第五，扣除这些成本并考虑销量变化后，单位贡献、总贡献或毛利率究竟哪一个发生变化？

这组问题有一个常被忽略的分层：技术变量可由设备测试直接测量，使用结果需要真实任务暴露，感知变量来自用户而不是工程师的判断，交易结果须来自订单和退货记录，财务结果还要接入成本与会计政策。若把不同层级的数据拼在一起，研究者可能把一个模型基准和另一行业的消费者实验连接为完整因果链，然而两个研究的处理对象、用户和终点并不相同。证据综合可以提出机制假说，却不能凭多篇分别有效的研究合成不存在的联合估计。



SVSAC | sv-sac.org

## 3.2 因果链与断点

分析中的因果链可表示为：产品设计和执行架构影响系统性能；性能在目标场景中影响任务结果；任务结果及其可解释性影响感知效用与风险；感知和价格共同影响购买、退货与使用；交易变化再通过产品净售价、销量和生命周期成本形成财务结果。每一箭头都可能受替代因素影响。例如改善可能来自新传感器而非模型；试用者可能是更感兴趣的早期用户；AI 版的宣传更醒目，带来的差异来自信息披露；渠道为新品给予优先陈列，销量提升不必然是功能价值；早期退货尚未成熟，短期转化可能掩盖售后负担。

因此，至少要记录版本与环境、实际启用率、失败与回退、暴露次数、交易价格、退货原因和每台支持量。对照必须回答：被随机或比较的对象是什么，客户是否真的收到处理，销售人员是否改变推荐，平台是否同时调整排名，订单是否具有足够成熟的退货窗口。意向处理效应（按原分配比较）可保留产品上线策略的因果对象；只分析成功联网或持续使用者则容易选择出更积极、更稳定的用户。若 AI 功能只在高风险状态触发，常规时段的平均准确率也不是主要经济终点；异常事件少而后果重时，应单独报告该事件层的召回与误报成本，而不能用普通日常互动平均掉。

## 3.3 方法与证据层级

定向证据梳理的截止日为 2026 年 10 月 23 日；2026 年 10 月 28 日对七项来源的对应摘要、正文或元数据另行核对，逐项范围单独记录，不追溯改写此前已知状态。来源包括同行评议嵌入式推理基准、端侧系统基准、关于 AI 设计产品支付意愿的出版社摘要、在线销售服务随机实验的官方摘要，以及 IoT 产品制造商网络安全文件的指定部分。[3][4]来源按其真实可读范围登记：若只读摘要，正文只概述摘要明确报告的研究对象和结果，不推断未披露的样本、分配单位或估计精度；技术基准只能支持其设备、任务、版本与测量窗内的性能事实；指导文件只能支持其规定性质和具体建议范围，不能支持购买或利润效果。

随后采用机制分析，将“处理”拆成硬件及模型配置、产品说明、默认启用与服务承诺，分别讨论各自可能影响。构造算例只用于揭示单位毛利额、毛利率、销量和一次性投入之间的算术关系，所有金额和数量均由作者假设，没有企业数据含义。研究不声称完成了现场实验、消费者调查、企业访问或实证回归。未来的可执行现场设计需在后续章节提出，并将不同产业场景的处理单位、比较组、成本边界和观察窗写清，不能把方案说成已经实施。



SVSAC | sv-sac.org

证据综合的规则是先判定研究的处理和估计对象，再决定是否用于本题。某项研究若测试 AI 生成商品设计，它可以说明消费者对“AI 设计来源”标签的判断，但不直接说明商品安装了 AI 后实际使用改善；一项 AI 客服现场实验可能说明智能信息服务与交易变化相关或在其随机对照下产生效应，却不自动外推到嵌入式设备的溢价；微控制器模型基准证明特定推理任务可运行，并不直接证明产品性能耐久、消费价值或收益。对这些相邻证据，本篇保留其机制意义和外推限制，不把它们合并成单一效果数字。

# 4. 技术可行性为何是必要条件、不是商业决策建议

## 4.1 资源边界把模型指标变成产品设计约束

微控制器上的深度学习不是单纯把云端网络缩小。MCUNet 的正式 NeurIPS 论文联合设计搜索空间和推理引擎：模型结构决定激活张量与权重需求，运行时内存调度又决定模型可用的空间，因而硬件约束会反过来限制可搜索的模型。正式版报告 TinyEngine 相较 TF-Lite Micro 与 CMSIS-NN 在特定测试中减少 3.4 倍内存使用、加速 1.7 至 3.3 倍；在特定 MCU 和任务上报告 ImageNet 分类、视觉与语音唤醒性能。论文的正式指标不应与 arXiv 旧版本的 4.8 倍数字互换，也不能把模型运行更省内存翻译成整台产品 BOM 降低了同样倍数。[1]

工程意义在于，可行模型空间取决于峰值 SRAM、Flash、输入分辨率、推理时延、功率和准确度约束同时满足，而不是只看参数量或 FLOPs。算法能在开发板上运行，仍未证明量产芯片替代、温升、传感器噪声、生产容差、启动时间、断电恢复、用户操作延迟和多年供货条件都成立。MCUNet 作者也围绕特定硬件、数据集和实现进行比较；这些结果说明软硬件协同能够推进微控制器推理边界，不能直接推断所有产品都适合端侧 AI，更不能用算法资源效率代替整机节能、良率或用户收益。[1]

## 4.2 基准窗口与实际使用周期不同

MLPerf Tiny 提供一套基准框架，对关键字识别、视觉唤醒、图像分类和异常检测等任务记录准确度、延迟和能量。该版本的测量方法对推理阶段设定窗口，能耗也有特定供电和转换范围；不是一个家庭设备从传感器采样、唤醒、预处理、无线通信、待机到充电的整机寿命测试。[2] 因此，“每次模型推理的能量”只能回答其测量窗口内的推理成本。如果一年中无线连接和常开传感耗电占多数，即使模型推理降低一半，也可能几乎不改变电池更换周期；相反，若识别算



SVSAC | sv-sac.org

法允许更稀疏采样，整机收益可能远大于一次推理能耗，但要通过系统对照实测。

基准的任务指标也不可直接映射为日常故障概率。图像分类 top-1 准确度表示某一数据集下的标签命中率，不等于低频异常的误报率。唤醒词任务若用固定分布测得高准确率，仍需检查不同口音、噪声、距离、误触发背景和日夜温度是否覆盖。按用户实际一年使用频次换算误报数，需要一组明确的现场暴露率、阈值、数据代表性和产品运行状态。缺少这组数据，就只能报告基准性能和部署假设，不能为商业宣传给出“每户每年少发生 X 次错误”的推断。

### 4.3 功能比较必须守住非 AI 等效基准

同一产品功能往往存在简单规则、传感器硬件改良和人工设置等替代方案。假设一台设备用 AI 识别“水泵异常声”，不应只与完全没有报警功能的旧机相比。较公平的产品决策比较可以包含传统阈值检测、AI 端侧检测和 AI 云端检测三种实现，锁定外壳、传感器、报警渠道和支持期限；再分别比较遗漏、误报、响应时延、断网能力、设备成本和用户需要做的纠错操作。若 AI 版增加了更好的麦克风或更高等级的质保，则把这些作为单独因素或产品包标记，不应归到模型效果。

规则方案在稳定环境中可能更易验证、耗电更低、失效模式更少；当工况变化多、信号形态复杂、规则维护成本高时，模型才可能产生额外优势。AI 系统也可能通过更广的特征空间增加对罕见环境的敏感性，带来误触发与后续人工核验。比较的核心不是“AI 是否先进”，而是在目标用户和场景的约束下，AI 方案是否改善正确任务结果且增量成本值得承担。若简单规则在相同风险约束下足够好，模型就没有充分的产品理由；若模型确实改善结果，还需向后追踪用户是否愿意为该改善付费。

## 5. 价值实现链的起点：功能、理解与支付不能跳步

### 5.1 让“更智能”变成用户可验证的结果

用户通常不会为模型结构或边缘计算本身付费，而会为更少的错误操作、更快的任务完成、更低的损耗或更高的可用性付费。产品设计因而要先定义真实的任务和反事实：没有 AI 时用户或设备如何完成任务，失败一次的后果是什么，任务发生多频繁，错误是可逆还是不可逆，用户能否发现并纠正系统给出的错误建议。对低频但高损失的任务，平均满意度可能掩盖安全或财务尾部风险；对频



SVSAC | sv-sac.org

繁但低后果的便利功能，哪怕每次只省几秒也可能积累成可感知效用，但用户也可能把它视作基本功能而不愿单独付费。

AI 能否被用户注意到，和它是否真正改善任务也不是同一问题。产品说明或界面能够帮助用户理解 AI 提供的建议及其依据，却也可能提高对错误的敏感度，或让用户把系统判断误当作确定事实。有关解释性产品信息能否改善理解、信任或感知性能，是一个值得单独检验的设计问题；但这类沟通效果不能替代实际功能测试。购买前被试对产品说明的反应属于信息处理结果，而不是设备在真实任务中的运行质量，也不是生产成本。产品团队应明确要改善的是功能、理解还是披露后的选择，并在设计中分别测量，避免把三类反应合成一个“用户认可”。

## 5.2 AI 标签、新奇感与功能效用的竞争解释

一项关于“AI 设计产品”的四项研究及跟进研究的出版社摘要显示，在其研究设置中，受访者对标为 AI 设计的产品报告更高支付意愿，作者以好奇心解释部分关系，并报告知识水平与产品类型的边界条件及一项现场研究。[5] 这不是对产品内嵌推理功能的直接证据。受试者可能为“由 AI 创造”的新颖来源付费，和为设备在真实使用中更可靠、更省力而付费，是两类机制。前者可能在信息差消退后减弱，甚至因知识提高而反转；后者需要稳定功能与持续使用结果。将二者混作一类，会把短期注意力当成长期效用。

因此，产品实验应拆开功能与标签。一种可检验设计是让所有版本运行相同功能，仅随机呈现是否披露其由 AI 实现；这识别标签及说明对认知或购买选择的影响，不识别 AI 功能效用。另一组则保持产品说明和外观一致，比较 AI 和等效非 AI 实现的实际任务表现与真实交易；若无法盲化，至少记录用户是否识别处理及预期。四臂因子设计可以分别估计功能与标签及二者交互，但若销售人员随标签改变推荐、体验版没有同样售后服务，处理仍是整套展示政策，而非纯粹算法影响。下文会进一步说明可执行的实验单位、渠道污染和费用测量。

另一项关于线上直播销售平台的随机现场实验在摘要中报告，消费者在直播会话中可使用 AI 信息助手的处理组，相对于不可使用的对照组，销售提升 3.00%，产品退货率降低 12.55%。研究描述的机制涉及个性化信息提供及其潜在打断效应。[6] 这是有现实商业环境和随机处理的证据，说明 AI 工具可在用户信息获取受限、销售人员服务容量不足时改变购买与退货结果；但它是销售过程中的交互式助手，不是购买者持有数年的内嵌产品功能。其处理方式、结果窗、平台分配机制和成本结构都不能直接外推到家电、工业仪器或车载设备的溢价与毛



SVSAC | sv-sac.org

利。两项相邻证据一项支持 AI 来源标签可能影响支付意愿，另一项支持销售服务工具可能影响交易；它们共同提出机制假说，却没有证明嵌入式 AI 本身已经形成可持续的单位经济优势。

### 5.3 从订单结果到财务命题的最小对账

要验证“高毛利”，至少需要把产品版本、用户接触、报价、支付、出货、退货、在用设备、更新支持和成本事件连接到可审计的时间线。同一设备销售价格、一次性渠道返利和后续云服务收入不能随意在不同版本之间分配。销量变化和单台成本变化也需分别列出，否则平均毛利率可能因产品组合变化而提高，实际 AI 产品的边际经济性却没有改善。相反，AI 产品单件毛利率下降而销量或留存增加，也可能使企业总贡献变高。二者回答不同决策问题。

为了防止将“高毛利”作为预设，本篇在后续提出两个独立估计对象。其一是同一产品族在指定市场、价格和销售期限内，AI 产品包相对于等效基线的每台贡献差及该期间总贡献差；其二是在经财务确认的收入和销售成本口径下，AI SKU 的实际毛利率变化。第一个对象可以从随机价格/配置实验结合成本表估计，第二个对象还依赖会计分类、退货成熟度、生产良率、渠道折让与支持成本归集。研究者可报告这两类估计的边界，但不能用前者冒充经审计毛利率，也不能仅以毛利率变化替代企业整体利润。

同样，客户支付额增加不意味制造商完全捕获了新增价值。若渠道加价、平台佣金与经销商返点随功能价格提升，终端愿付价与制造商净收入之间会有楔子；若 AI 增加售后依赖，某些服务成本还会由经销商或用户承担，转移的成本并不等于社会成本消失。后续账本应区分终端购买者、渠道和生产方的边界，并对功能成本承担主体作敏感性分析。只有当功能改善、真实交易变化和成本承受关系在同一产品周期中对得上，商业结论才可讨论。

## 6. 用例必须先于模型：把产品承诺改写成可测任务

### 6.1 从“智能”标签还原任务边界

同一 AI 模型可以支撑截然不同的产品功能。机器上的异常声音分类可以提醒维护人员，家用设备的声音识别可以提醒用户，汽车座舱中的语音识别可以控制媒体或请求导航；其数据类型可能相似，但失败后果、用户注意力、容忍时延和动作权限并不相同。产品定义不应停留在“用 AI 做预测”或“自动理解”，而要明确输入从何而来、何种状态触发、输出给谁、用户接下来能做什么，以及什么条件下产品不会执行该功能。



SVSAC | sv-sac.org

可执行的用例规格至少应写出六项：目标事件及其可观察参考标签；一次“使用机会”的单位；实际使用环境和暴露范围；允许的响应时限；漏报、误报、拒答和延迟输出各自的后果；以及用户或系统纠正错误的途径。若没有这些定义，“准确度”就没有稳定分母。同一个“泵异常”分类任务可以按每秒采样窗口、每个设备日、每次维修请求或每个真实故障事件计算结果；窗口级正确率不能自动回答用户每月被错误打断几次，也无法直接说明是否能提前避免停机。

任务边界同时规定产品责任。若模型只给维护人员一个建议，工单负责人仍需确认；若系统自动停机，错误输出的后果和需要的独立保护就变了。设置在设备上的 AI 功能、应用程序里的提示和云服务中的诊断不能因同一个品牌名称而合并为同一处理。研究需要记录功能默认是否开启、用户能否关闭、断网时怎么退化、告警是否持续到有人处理，以及系统能否给出拒绝或“信息不足”。这些不是技术附注，而是用户实际购买的功能范围。

## 6.2 把使用频次、暴露机会和目标事件分开

识别低频事件的 AI 产品尤其容易因分母选错而夸大效果。假设某设备每天运行 8 小时，一个“窗口”是每秒一次分类，全年窗口总数极大；产品使用机会却未必是每个窗口。用户真正关心的可能是设备每月是否发生故障、是否在故障前收到可行动的提示，或每周是否被误报警一次。若同一真实事件持续覆盖多个相邻窗口，逐窗口计数还会把一次故障重复算作许多错误或命中。

对产品功能进行评估，应把分母按因果任务层次列开：传感器采样窗口、系统触发机会、独立事件、设备运行时间、用户交互次数和实际购买设备数。每一种指标回答不同问题。窗口级召回可以用于调试模型，但产品误报负担宜按设备日、周或用户月报告；用户收益可能按独立故障事件计算；企业经济性则要按售出或在用设备及其生命周期计算。一次设备故障可能带来多个告警，设备也可能在观察期内没有任何目标事件，分析计划要提前规定如何处理这两类情形。

为了说明基率如何改变可用性，设一个纯构造场景：在 10,000 个互不重叠的产品使用机会中，目标故障实际发生 10 次；假设算法对故障的召回率是 90%，对非故障的误报率是 1%。期望真阳性为 9 次，非故障机会有 9,990 次，期望误报约 99.9 次。即使台架测试中的召回看似较高，用户收到的提醒里真故障约占 9/(9+99.9)=8.3%。若通过更高阈值把误报率降至 0.1%，则期望误报约 10 次，提醒中真故障占约 9/(9+10)=47.4%，但漏掉的故障比例仍由假设召回决定。这些计算只是贝叶斯基率演示，所有基率和率值都是假设，不代表任何产品。这里 8.3%和 47.4%是给定机会总体与率值下、按期望计数形成的警报构成，不是有限样本随



SVSAC | sv-sac.org

机比值的期望；后者还要处理零警报。期望计数的相加只需相应边际率，不要求各机会独立。相关性会改变方差与区间，使简单独立二项推断失效；同一故障跨窗重复计数则改变事件层估计对象，须另作去重。提高阈值后仍维持 90%召回也是本例额外假设，真实阈值变化须实测。

### 6.3 构造例的产品含义与决策限制

上面的敏感性变化显示，阈值不只是模型团队的调参问题。更低阈值可能提高召回，却让用户更频繁地收到低价值警报；更高阈值可能降低打扰，也可能让功能在真正需要时沉默。选择阈值需要以产品承诺、可纠错性及后果成本为依据，而非只选择验证集上的最大 F1。可逆的便利提醒可以让用户关闭或延迟，涉及设备停机、健康或财产损失的动作则需更强的独立边界。产品层面还要区分“警告”“建议”和“自动执行”，不能把同一分数阈值无说明地用在三种动作上。

经济换算也需谨慎。若每个假告警需要用户花 30 秒确认，100 次误报对应 50 分钟注意力负担；但将其换成工资成本只有在用户时间确实可替代为可避免的支付、生产停工或客服工时，并明确谁承担成本时才有意义。消费者可能把打扰视为烦恼，工业用户可能需要停线后确认，物业管理者可能按事件派工；不同受影响人和成本口径不能互相代替。即使设备厂商不承担用户时间，它仍可能因差评、退货、投诉和保修而承担间接影响。后续应分别跟踪用户负担与企业成本，而不是把所有错误乘以一个未经解释的金额。

相应的最低评估表应包含真事件数、独立机会数、暴露时长、召回、误报、拒答、每设备周期错误次数、按错误类别分层的纠错时间、和可行动建议比例。若存在无法判定的现场事件，应将“未决”保留为单独状态，而不把它自动算成模型正确或错误。只有通过独立维修、人工复核、设备拆检等参考程序确定标签时，才可估计相应的产品事件表现；若标签只来自算法自己生成的告警或用户投诉，测量将受选择机制影响。这个要求并不意味着每个产品必须安装昂贵的全覆盖真值系统，而是应说明哪些现场结果可被观察，哪些仍不能识别。

## 7. 端侧、云端与混合部署：数据路径决定产品行为和成本暴露

### 7.1 用功能时序图而非“端侧更安全”做架构比较

架构取舍应从一个完整任务时序开始。设备采集传感信号后，先进行校准、滤波或压缩，再决定本地模型是否足以输出结果；若置信度不足，功能可能向手机、网关或云端请求补充计算；云端返回的类别、建议或控制指令还要经过设备



SVSAC | sv-sac.org

状态检查，然后展示、记录或拒绝执行。每个环节都影响时延、掉线行为、能源、隐私暴露及故障归因。仅比较“模型在 MCU 上运行”与“模型在云端运行”，会漏掉传感器、无线电、状态同步和产品界面的主要路径。

可将一个产品功能拆成六类对象：原始采样、提取后的特征、推断结果、模型参数、诊断日志和更新包。每类数据都需要明确是否离开设备、传输触发条件、保留期限、访问角色及用户可见性。原始音频不上传并不代表没有传输使用时间、设备标识或模型输出；本地生成的诊断类别可能仍识别个人状态。相反，云端处理也未必意味着持续上传全部传感原始数据，产品可能只上传事件片段或低维摘要。真正可比较的是数据流和用途，而不是架构标签。

产品连接状态需要作为功能条件写入需求。若云不可用，设备是保持旧状态、以简单规则运行、向用户明确提示暂不可用，还是在关键任务上拒绝给建议？如果云端模型持续更新，本地端是否缓存旧版本，如何判断设备端和云端的结果可兼容？更新失败或签名过期时，设备能否回滚到能运行的受支持版本，同时避免回退到已知易受攻击的版本？这些问题既影响客户所购功能，也会产生售后和网络安全成本。

## 7.2 三类架构并非线性的优劣排序

端侧执行适合时延和断网要求严格、输入数据不宜频繁离开设备、且受控任务能在资源内完成的产品功能。它把一部分推理成本从持续云服务转换成设备的计算资源和前期工程成本；模型容量小、出厂时硬件锁定、设备数量巨大时，本地化可能减轻逐次调用费用，却增加硬件规格、量产验证和多年固件维护压力。若用户购买后多年不联网，模型漂移和输入变化更难从厂商处观察，现场诊断也会变难。

云端执行适合需要较大计算、快速统一迭代或跨设备聚合信号的任务，但网络时延、服务中断、账户管理、峰值容量和数据传输会成为产品功能的一部分。对于以月费或按量收费的 AI 功能，必须说明售价包含多少服务年限、免费期结束后功能是否降级、如何停服以及用户的设备是否还能保持基本功能。若产品以一次性硬件收入出售而推理和支持成本持续按年发生，企业账面可能在出货时确认收入、现金负担却在之后累积；一次短期毛利比较会低估生命周期义务。

混合架构可以将低功耗触发、隐私敏感的预筛选或断网基础功能留在端侧，再将复杂推理放在云端。但“混合”并非自动折中，它会制造两个模型版本、两种错误分布和更复杂的回退状态。若本地筛选器将大多数片段丢弃，云端测试只在被筛选后的片段上完成，便不能用云模型的离线精度代表端到端漏检率；端侧初



SVSAC | sv-sac.org

筛漏掉的真实事件不会进入后续评估。端云结果如果意见冲突，产品需要规定谁占优、何时拒绝输出、如何向用户解释差异，以及日志如何支持售后复现。

### 7.3 架构必须按产品寿命和支持约束一起评估

对设备制造商而言，架构还受产品支持年限和硬件生命周期制约。量产之后的芯片替换、存储规格变更、供应商固件更新或地区无线认证差异，都可能改变模型时延、能耗、数据接口或更新路径。若每一批设备调用不同版本，却没有将序列号、硬件修订、模型版本、校准参数和应用程序版本关联，现场故障就会变成无法归因的总体投诉。消费者可能只看到“这个功能有时不好用”，工程组织却无法判断是硬件漂移、传感器装配、模型回归、网络延迟还是云端服务变化。

长期更新也不是单方向的能力增强。升级可修正缺陷、支持新硬件和抵御攻击；升级也可能改变输出阈值、增加算力需求、耗电或破坏设备间互操作。回滚可以帮助恢复功能，但如果回滚到存在已知漏洞的版本，网络安全风险会变大。设备状态提示、授权更新和安全能力在 NIST IoT 资料中有明确讨论范围，但这些文件属于能力基线或制造商推荐活动，不是产品商业收益证据。[3][4] 因此，公司应把功能验证与更新机制相连：新版本除了“能安装”，还要证明在指定设备批次、传感条件与任务参考上满足事前功能界限；若证据不足，应能暂停推送、限制目标群、恢复安全状态或让用户暂用等效非 AI 功能。

此处的关键并非给所有设备指定一个统一支持期限，而是把期限与出售承诺、成本和用户风险相匹配。短生命周期、可更换的小型附件与安装在工业现场的关键控制器，不应共享同一套停更决策。产品规格中的“智能升级”若依赖后端服务，应明确设备离线和服务终止后的剩余功能、数据导出或删除路径、备件和更新供应责任。成本估算要按设备在用年而非仅按销售当年计算，否则越成功的安装基数反而会带来被漏算的持续负担。

## 8. 从离线性能转向真实使用：错误频率、环境与版本共同决定质量

### 8.1 一个验证集指标不是产品运行分布

产品性能验证中的样本集往往经过采集、清洗、标签和抽样。开发集可能包含重复设备、特定供应商的传感器或同一现场的相邻时段；生产设备则可能出现新的房间、温度、材料、用户语言、背景音和网络状态。若训练/测试切分在事件窗口层随机进行，来自同一物理事件的近邻片段可能同时进入两侧，形成泄漏。模型在离线测试中看似可重复，遇到新的设备批次或季节变化时却会退化。产品



SVSAC | sv-sac.org

验证的关键不是再增加一个分数，而是判断切分是否对应未来将遇到的独立设备、现场、时间与版本。

至少应分清三种评估：按样本随机切分以检验模型学习；按设备、站点或产品批次留出以检验迁移；按时间先后评估以检验未来漂移。这些估计对象不同。旧批次的随机留出不能替代未来时间验证，跨站点留出也不能回答同一站点长期磨损或维护后的表现。若产品具有用户特定校准，个人设置数据必须与评估数据隔离；若运行中学习，训练反馈和正式评价要分开，否则新模型可能在被评价之前已经间接见过结果。

每次报告还需要冻结产品化流水线，而不仅仅记录模型权重。传感器采样率、前处理过滤、阈值、固件、模型、云服务、默认配置、用户提示和设备校准共同决定行为。仅仅替换模型版本也可能改变耗电、等待时间和系统内存；前端若把低置信度改成“请检查设备”，比显示一个预测类标签更能改变用户结果。因此，技术团队的模型评估和产品团队的整机评估应保留共同版本标识，但分别报告任务分数和实际功能结果。

## 8.2 把错误转换成用户和企业两套账

错误需要按后果分类，而不是只有准确与错误两类。漏掉一个故障、把正常声响判作故障、拒绝给出建议、建议迟到、重复提醒、用户不理解提示和自动动作过度，可能有不同处置路径。对用户而言，误报可能消耗注意、制造担忧或导致不必要的检查；对制造商而言，可能转成投诉、退货、服务工单和保修负债。漏报可能使用户承受停机或损失，制造商是否承担责任则取决于产品承诺和实际使用场景。需要双重账本：按固定暴露单位计算功能错误与用户负担；再按净收入、返工客服、换货和保修记录计算企业成本。不得在没有因果识别时，把每一笔售后成本都归因于模型。

对安全或高价值任务，错误还可能从平均效用中分离出来成为硬约束。如果设备自动切断电源，产品不能仅因为平均正确率达到某数值便直接自主运行；应定义独立的硬件保护、最大输出范围、异常拒绝、人工确认和失效后安全状态。对于可逆的小额便利功能，可以允许更多个性化或自动化，前提是用户能及时撤销并理解其作用。风险等级影响验证终点、测试场景、现场观察时间和动作权限，而不是只调整模型阈值。

## 8.3 从产品事件建立可复核的参考标签

现场真实标签并非总是现成的。用户报告、售后工单、维修检查和独立传感器测量可能在不同时间产生，含义也不同。若只把“用户投诉”当成错误标签，没



SVSAC | sv-sac.org

有投诉不等于功能正确；若只检查模型报警的设备，就无法知道未报警设备中的漏检事件。测量流程应随机抽查报警与未报警样本，按目标事件及严重度做独立验证，同时保存“无法判定”比例和抽样概率。若无法对全部产品拆检，可使用分层抽样及权重估计，但抽样机会必须能被重建，否则总体漏检和误报不能从选择性审查数据识别。

重复测量也不能误当独立用户。一个用户每天触发 30 次并不等于 30 名独立顾客；同一设备的几十万传感帧也不是几十万台设备。对于产品级效应，分配单位和误差结构可能是设备、家庭、工厂或现场，而不是帧和触发事件。分析应按实际随机化/采样单位计算不确定性，并展示设备间、用户间和环境间的异质性。若仅有少量现场或试点用户，结果就应限定为这些单位与时期；大量帧数不会自动扩展外部效度。

### 8.4 安全、隐私和产品承诺不能在功能比较外遗漏

端侧推理可减少某些原始数据传输，但安全和隐私结果仍需按数据类型和生命周期核对。设备可能用身份标识将本地结果关联账户，也可能定期上传错误日志或精度反馈；云端服务可能接收经过汇总的信号，但需要明确再识别和访问权限风险。产品团队要说明每个数据对象用于推理、诊断、持续学习或质量改进中的哪一种，谁能读取和保留多久，用户是否能拒绝非必要上传。否则，“离线 AI”可能只描述核心推理所在位置，未描述整个产品的数据使用路径。

网络安全也会反向影响 AI 功能价值。若产品没有授权更新和状态能力，企业不易修复长期运行中的漏洞；若功能更新很频繁，兼容和验证负担增加。NISTIR 8259A 将设备能力基线定位为组织识别新 IoT 设备能力的起点，NISTIR 8259 Rev.1 则描述制造商应考虑的网络安全建议活动；二者范围相关但不是同一文件，也不是所有商品一律适用的强制认证。[3][4] 它们可以提醒企业将更新与产品生命周期纳入治理，不能替代特定法域的法律分析，也不证明消费者因安全措施而支付溢价。产品价值研究需把安全功能成本与因果收益分开，若没有测量用户付费、风险降低或支持支出，不能将其直接计入项目回报。

## 9. 从感知到使用：用户是否能把功能差异归因于产品

### 9.1 价值发生在任务结果中，不发生在“AI 存在”本身

用户对产品功能的评价往往取决于结果是否看得见、是否来得及时，以及他们是否相信结果来自产品而不是外部偶然因素。若设备自动降低能耗，用户需要知道在相同房间、温度和使用时长下节省了什么；若设备主动发现异常，用户要



SVSAC | sv-sac.org

能核对告警对应的实际状态；若模型提供个性化设置，则用户需要知道个性化给谁带来便利、是否牺牲了控制权。没有可观察结果时，AI 通常只是消费者在购买前无法验证的属性。它可能成为广告卖点，却不能因此被当作已实现效用。

因此，一个“AI 功能价值”指标不应把性能、用户感受和交易揉成单一加权分数。工程上正确的故障预测可以没有可行动价值，因为现场没有备用件或维护窗口；响应很快也可能只把错误更快地推给用户；用户说功能方便，实际却不愿付费或持续使用。建议的证据链分别报告任务成功、错误和拒绝结果，用户理解与控制感，付费及退货行为，以及支持成本。不同维度可以作决策表，但不应被汇总为一个无法解释的“AI 成熟度指数”。

## 9.2 有用性、可理解性和控制权之间可能互相牵制

AI 功能提高自动化程度，可能减少用户输入或等待，也可能让用户难以理解系统为何采取行动。对于可被用户随时撤销的照片分类、个性化照明等功能，透明度可能是设置选项和退出机制；对于会改变设备设置、生成健康风险提示或触发工厂告警的功能，用户可能需要查看建议根据、置信程度与适用边界。过度解释也可能增加认知负担，并使本来简单的任务变得更复杂。设计研究要具体测量信息是否提升理解或信任，而不是把“可解释”视作越多越好。

产品的“可控性”也不是一个按钮。用户可能需要暂停自动执行、改回固定规则、纠正输入、撤销一次动作，或限制某些信息用于个性化。不同控制选择产生不同用例，也可能改变模型表现和采纳方式。如果支持手动覆盖的用户更常遇到异常，比较“使用自动模式者”的体验将忽略潜在不满；如果大部分购买者关闭 AI，出厂准确度再高也不等于产品价值。现场测试应记录默认状态、启用率、关闭时点、回退原因和覆盖频次，并保留按原产品分配分析的主结果，同时报告实际使用者结果作为不同估计对象。

## 9.3 试用、采纳和持续使用需要不同的观察窗口

新产品的初期使用可能受到促销、好奇和早期采用者特征推动。设置完成、第一次成功运行和数月后仍主动使用不是同一个终点。用户可能在购买后未完成配对，试用期间愿意尝试，在首次误报后关闭功能，或虽然持续使用但不愿在下一代设备上再次付费。若只观察应用激活、模型调用次数或试用点击，就无法区分有效使用和用户被默认设置持续触发。需要在设备销售、功能启用、功能接触、任务帮助、退款/退货、主动关闭和复购之间保持一条设备或账户级时间线。

这也会影响“留存”的解释。没有活跃调用不一定代表用户拒绝 AI：该设备可能只在低频异常时启动；调用很多也不一定代表满意，可能是设备持续误报或用



SVSAC | sv-sac.org

户反复检查。产品事件要按功能用途分为自愿调用、自动触发、系统更新、用户覆盖和客服诊断。对于按订阅销售的服务，续费是否发生及价格变化可作为观察结果，但续费意愿问卷只能作为前置指标；对于一次性出售的耐用品，下一次整机购买可能几年后才发生，短窗口内的“无复购”不能直接编码成弃用。

# 10. 信任、标识与风险感知：AI 作为产品属性如何被理解

## 10.1 “AI 设计”标签与“AI 功能”标签不是同一处理

AI 来源标签可以改变消费者对新颖性和创造者的判断；功能标签则尝试描述产品内的能力。这两个信息维度应分别操纵。Zhang 等人对 AI 设计产品的研究摘要指出，在其研究语境下，受访者对被描述为 AI 设计的产品报告较高支付意愿，并提出好奇心及消费者知识和产品类型等边界条件。[5] 这类结果无法告诉我们一个已经配备传感器和推理功能的产品，在连续使用后是否更有用或能维持溢价。产品开发者若把来源新奇包装成长期功能价值，可能在新颖感衰减后遭遇转化下降。

信息披露本身也需要拆成可测试的内容。可以将商品页随机分配为只描述任务收益、只说明 AI 机制、同时说明收益和机制、以及仅提供非 AI 基线说明。若目标是估计“说 AI”带来的纯标签效应，其他文案、图像、保修、价格和排名必须保持一致；若把功能优势和标签同时加上，识别对象是组合传播政策。对消费者调查可测理解、信任、预期表现及支付意愿，但真正的购买试验还需观察随机分配的有效报价、实际付款和退货。两种研究共同回答不同问题，不能用问卷结果替代市场结果。

## 10.2 信任也可能增加错误依赖

信任不是越高越好。若系统性能边界不清晰而界面语气过于确定，用户可能不再核查输出；若系统偶尔出错且难以解释，用户也可能把一次失败泛化为整类产品不可用。产品团队需要测量的不只是平均信任分数，还包括用户何时采纳、何时质疑、是否能识别低置信度、遇到冲突信息时是否会寻求人类确认，以及他们是否知道自动动作已经执行。使用问卷“我信任这台设备”无法代替观察用户行为，也不能说明错误依赖发生率。

对长期产品，信任的形成还会随版本变化。产品可能在现场更新阈值或模型以修复缺陷；如果公司只披露“软件更新”，用户却发现自动行为改变，客户承诺与真实运行状态会脱节。更新前后应保留产品版本、输入类型、默认设置、告警说明和用户控制项，以判定信任变化属于新功能质量、界面说明、新闻事件还是



SVSAC | sv-sac.org

其他同期变化。厂商自报的稳定运行期应与用户实际活跃设备、售后事件和退货数据对照，单一满意度截面不足以识别这种变化。

### 10.3 风险感知会因用途、用户和后果而异

AI 产品可能让用户担心数据外泄、行为误判、自动化失控、过度依赖云端或被迫订阅，也可能让用户更愿意把重复工作交给设备。风险取决于其使用场景，而不是只取决于模型的名义“风险等级”。家庭设备收集的睡眠、语音或活动数据、工业控制器的运行状态和普通娱乐产品的图像分类，各有不同数据边界和损失后果。风险问卷若只问“你是否担心 AI”，会将隐私、可靠性、安全、可解释性与产品依赖混在一个模糊项目里，难以说明哪一种设计改动会改变行为。

风险与效用还可能相互作用。若用户因 AI 功能减少操作负担获得实际帮助，他们可能愿意容忍一定的数据使用；若益处难以验证，同样的收集路径可能被视为没有必要。不同用户也可能分别采用默认选项、主动关闭、选择低数据模式或完全不购买。分析时应将这些选择作为政策结果记录，不能只对仍开启功能的用户测量风险感知，再据此声称全体客户接受。低隐私敏感用户中的采用率也不能简单外推为总体愿意付费。

## 11. 支付意愿与真实需求：从问卷回答到结账金额

### 11.1 陈述支付意愿的条件边界

消费者调查常询问受访者“愿意为 AI 功能多付多少”。这类回答可以发现价值来源或拒绝原因，但它受假设偏差、参考价格、品牌熟悉度、描述方式和受访者理解影响。若没有约束预算的实际选择，受访者可能对一个看起来有吸引力的新功能报告高意愿，却不会为它从其他消费中让出支出。支付意愿也会因情境措辞而变化：把同一功能描述为省电、预测故障、保护隐私或“人工智能”，可能引发不同联想。问卷可以作为形成假设的工具，不宜将其输出写成企业可执行的加价额度。

选择实验比单问开放金额更能显露产品属性之间的权衡，但也依赖选择集合、属性范围和统计假设。若所有选项都包含 AI，模型估出的 AI 系数就没有真实不购买或非 AI 替代基准；若选择题未包括维护费、服务期和退货条件，受访者无法表达对完整产品包的权衡。研究应加入“维持现状/不购买”选项，给出真实可比较的功能版本与价格范围，注明是否实际购买兑现，并报告样本、权重和估计条件。陈述选择只能估计所设计情景中的偏好，不是无条件市场需求。



SVSAC | sv-sac.org

## 11.2 随机价格与配置试验识别什么

更接近真实需求的设计是随机化报价、AI 功能包或产品信息，对接实际支付和退货。研究者需要提前确定随机单位：若相同用户反复进入不同价格页面，跨期学习会污染处理；若一个账户有多个家庭设备，设备层随机可能被家庭成员沟通抵消；若经销商自行打折，牌面价格不等于消费者实际支付。对每个原始分配账户，应记录有效触达、报价、购买、交付、启用、退货和服务成本，按最初分配比较整体产品政策，另行报告执行偏离。

一个完整识别设计可以采用两个正交因素：功能实现（AI 与等效规则）以及对消费者的说明（明确 AI/只说明功能结果）。再对价格随机给出若干可执行价位或折扣，估计每个功能与标签组合下的购买概率变化。实际运营可能无法给单个客户不同价格；此时可按市场、渠道周或账户簇分配，并处理同一买家跨渠道、促销活动和价格信息外溢。只随机标签而不随机功能，可以识别营销描述的结果；随机功能而隐藏模型类型，若有效盲化，则更接近功能效用；同时改变配置、售价和保修，则估计是产品发布包而非 AI 单独作用。

观察结果需要有足够成熟的退货窗和在用期。若试验截止日在用户交货后第七天，组间退货差可能尚未完全显现；若 AI 功能需要数周学习或校准，早期体验不能代表稳定状态。对未交付、未启用或中途停用的用户，仍应保留在原分配组的主分析中。另行估计“启用者效果”需说明选择条件，不能把成功配对的子样本当作随机总体。

## 11.3 一项价格对照说明不同估计量

考虑一个作者构造的两价格试验：同样的 AI 产品包随机报 100 或 110 个货币单位，实际购买率假设分别为 20%和 17%。在功能与其他承诺完全相同、随机分配有效且报价不串组时，每千个报价的预期收入分别为 20,000 与 18,700；价格提高 10 单位虽然提升购买者单价，却使该情境下收入减少 1,300。此处尚未扣除退货、渠道返利、生产和服务成本，也不证明任何真实需求曲线。它只说明“愿意多付”的买家比例与“提高价格后单位时间收入增加”并非一回事。

如果目标是 AI 功能增量效应，仍需一个等效非 AI 报价组。假设消费者对 AI 配置的购买率为 17%，非 AI 配置为 19%，而 AI 组价格高 10，则销量和价格两种作用可能方向相反。可以比较各自的实际净收入，但价格差不应与功能差混在一起。市场整体收益还受渠道和竞品响应影响，实验中的随机价格跨度也不能推演未测试的完整需求弹性。管理层应选择实际决策要用的对象：某价位下利润差、



SVSAC | sv-sac.org

给定预算的销量、特定配置的净价，或同一客户群内的增量购买。没有识别对象，支付意愿数字无法支持定价。

# 12. 产品捆绑、非 AI 对照和市场替代

## 12.1 功能包效应不能归功于单一模块

产品升级常同时改变处理器、传感器、外观、应用程序、保修、联网方式和 AI 服务。若新款售价、包装和宣传都不同，上市前后销售对比只观察到整个升级包及时间变化。要识别 AI 模块，需要在条件允许时拆分因素，例如同款硬件上随机开启算法或随机分配功能说明；如果用户安全或产品责任不允许关闭 AI，则可以在安全的离线测试或对照试验中限制动作权限，并说明剩余的可归因边界。

但消费者购置的真实单位往往就是整包。厂商决策也可能关心“AI 新款是否值得推出”，而非模型本身的纯效应。这种问题可以用整包对照，却应把传感器更换、应用改版和保修承诺一并写入处理定义，并将效果称作“产品方案的增量结果”。营销传播和功能执行还可能产生交互：功能确实有效但用户不知道怎么启用；仅标签增加关注却没有任务改善；说明能降低担忧但功能只在少数场景可用。简单地把所有因素归一化成“AI 作用”会丢失产品决策需要的具体机制。

## 12.2 用户可能替代既有商品而不是新增购买

AI 版产品的销量上升有时只是同品牌旧型号换代或竞品份额转移。若只看 AI SKU 销量，产品组合可能把旧版的销售蚕食算作新增需求。需要按用户账户或市场追踪相邻型号、替换周期、回购与新增客户，区分“新增购买”“同品牌型号升级”和“竞争产品转移”。消费者即使更偏好 AI 版，也可能只是把原本计划买的普通产品换成 AI 版，市场总量不变。对制造商而言，这种替代仍可能提高单客贡献，但必须核对旧版利润损失和新版本的售后责任。

经销和平台渠道还会改变谁控制价格及顾客关系。平台可能将 AI 功能与高级订阅绑定，品牌只获得硬件溢价的一部分；经销商可能因更高服务负担而要求返利；消费者可能从独立配件或第三方服务转向厂商生态。需要对渠道折扣、佣金、推荐展示、激活资格和服务续费做版本匹配。产品问卷不能告诉厂商实际捕获多少溢价；销售后台的净价、退货和返利数据才能回答与收入相关的问题。

## 12.3 产品更新可能改变买卖双方的功能边界

嵌入式 AI 产品不只是购置时的静态配置。OTA 更新、订阅权限、云端模型替换和地区功能限制会改变功能表现与持续使用权。用户可能以为购买时的智能



SVSAC | sv-sac.org

能力会随设备长期保留，但企业可能将某些新模型设为增值服务或在云服务到期后关闭。若销售页承诺“持续改善”，企业需要说明更新支持期、可离线功能、设备报废与数据迁移安排。否则用户支付的不是固定产品功能，而是一个依赖合同与服务持续提供的权利。

更新还是一项新的处理变化。AI 产品在一年内更新多个版本时，销售和支持结果可能反映不同模型、默认阈值与说明文案。若旧版用户与新版用户被共同分析，当前平均值掩盖了版本间差异。产品团队需记录每批设备更新时间、覆盖率、失败安装、回滚和用户选择；比较升级前后的使用或退货时，还应考虑同期市场与设备年龄。向用户统一推送更新并不自动形成随机试验，采用更新的设备可能本来就更常联网、更活跃或享有更好支持。

# 13. 增量硬件成本：从功能需求追到物料、良率和交付能力

## 13.1 应核算相对于什么产品基线的增量

“AI 硬件成本”通常被误说成新增一颗芯片的采购价。真正与决策相关的是，嵌入功能相对于可销售的非 AI 方案增加、替代或重新配置了什么。某产品可能需要更高主频的处理器、更大内存、更精确的传感器、更稳定的时钟、电源余量、散热结构、屏蔽或更高等级的连接模组；也可能因端侧运行而减少长期通信费用，或因专用推理芯片而替代原有处理器的一部分能力。成本基线应是承担同等核心任务、可以真实报价和交付的替代产品，而不是“零 AI”的虚构空壳。若同时改版外壳、电池或保修，则这些项目要作为独立差异列明，不能把整个新款与旧款成本之差记成 AI 单项。

应按产品配置建立可审计的差异表：料号、单位用量、供应商报价日期、采购数量区间、替代料、测试要求、装配工序、目标销售地区，以及“新增/替代/共用”标签。一个模块替代低阶器件时，增量成本是新旧成本之差，而不是新模块全价；若高阶处理器被用于其它功能，整机成本还应按明确的产品分配规则而非凭直觉全归 AI。价格引用至少需对应报价条件、币种、税费和交付条款；目录价不是实际采购价，样品价也不是量产价。物料成本的可信度应有范围，而不是单一精确小数。

此处的对照并不要求每种产品都制造一台“无 AI 版本”。若安全或架构原因使某功能不能关闭，可将硬件增量与功能增量分开：用工程物料清单比出不同器件，用同硬件离线对照估计算法任务作用，再用整包试验评估上市方案。三个估计对象回答不同问题。芯片基准能说明特定硬件上的资源可行性，却不说明生产



SVSAC | sv-sac.org

BOM 或售价；MCUNet 报告的是指定微控制器任务和实现中的内存/推理结果，MLPerf Tiny 测量边界也不是完整设备的前后处理、无线和电池寿命。[1][2] 从基准跃迁到成本结论必须再经过电路板、散热、电源、合格率、采购和售后这些中间环节。

## 13.2 器件选择会把推理资源转换成整机权衡

在某个任务上提高模型准确率，可能要求更大的参数存储、更复杂的输入、更多传感器采样或更频繁推理。这些需求会改变处理器、内存和功耗，进而影响 PCB 面积、电池容量、温升、外壳空间、EMC 设计、续航承诺及充电接口。若增加电池抵消推理耗电，电池与结构成本可能超过运算器件；若推理周期很长、触发稀疏，常驻电路的待机功耗又可能比单次推理更重要。单次推理能耗不等于月度能耗，更不等于电池寿命，因为网络等待、唤醒次数、用户操作和传感器常开状态共同决定实际负载。

传感器配置同样有替代关系。高分辨率摄像头不一定改善真实决策，若镜头污染、安装角度和照明条件占主导，算法无法弥补输入失真。精确传感器可能需要校准、标定工装和现场更换流程；若消费者不接受摄像或音频采集，低成本传感器也可能因采纳率低而无价值。应将传感器准确度、使用位置、校准周期、替换件可得性、隐私选择与目标任务联合建模，而非单独追求设备数据更多。

构造一个纯假设的设备配置比较：A 为基础产品，物料增量以相对基线记为 0；B 增加推理芯片与内存，单台物料增量为 18 单位；C 不增加专用芯片，却升级传感器、电池和散热，增量分别为 7、6、5 单位，合计 18。即便两种方案的 BOM 增量相同，采购风险与良率结构也不同：B 可能依赖单一芯片且软件映像需重新验证；C 可能引入多个供应商料号和更多装配/校准步骤。若 B 的设备端推理让网络费下降，须证明该任务以前确实发生了相应云端调用，并计入断网时可用性；若 C 的数据更好但用户关闭传感器，账面配置就没有产生功能价值。此例并未报告实际报价、性能或产品效果，只说明等额 BOM 不能代表相同产品风险。

## 13.3 量产良率和返修使“单位物料”不等于“单位交付成本”

采购清单按合格产品所需用料计算，但工厂要为报废、返工、测试失败和可修复缺陷付费。若一个新增模块需要额外焊接、屏蔽或校准，增量成本至少包括装配工时、治具折旧、功能测试时间、返修工位占用、报废损失和备件库存。不同部件故障还可能产生共因：新摄像头与新结构件共振或热漂移；更换板卡后固件与传感器序列号映射错位。仅将硬件物料价格乘以销量，会漏掉工艺能力和终检约束。



SVSAC | sv-sac.org

一项实用的分析先区分理论投入成本、合格品成本及售后期望成本。以下只是作者构造的大批同质投入、按期望数量归一的单次返修模型。设每台投产投入物料和生产资源 M；初次通过概率 $p_0$；进入返修的概率 r；进入返修后通过的条件概率 q；每次返修额外耗用资源 R；每台最终报废产品回收价值 S。初次通过、进入返修、直接报废三条初始路径互斥，故 $p_0 \ge 0$、$r \ge 0$、$p_0 + r \le 1$，且 $0 \le q \le 1$。最终合格比例为 $p_0 + rq$，最终报废比例为 $(1 - p_0 - r) + r(1 - q) = 1 - p_0 - rq$；在 $p_0 + rq > 0$ 且两类报废具有相同回收价值的简化条件下，净制造资源成本与合格交付数量之比为 $[M + rR - (1 - p_0 - rq)S]/(p_0 + rq)$。M 已包括报废品原有投入，不能再把同一投入作为额外“报废损失”加一遍；回收 S 须扣减，返修后通过的产品须进入分母。该比值是期望总成本除以期望合格数量，并非有限批次中随机成本比值的期望。若存在多次返修、不同回收价值或返修成本随缺陷变化，应逐路径核算；没有数据时应把这些概率与成本作为情景变量，不能把供应商的承诺当成已实现量产结果。

例如，纯假设的 1000 台投产计划中，原型阶段估算每台物料+装配为 100 单位，量产初次通过率可能处于 90%至 98%。在明确设定不返修、失败产品无回收价值、因而初次通过率也等于最终合格比例的粗略敏感性演示中，100 单位投入除以 0.90 约为 111.1 单位/合格品，除以 0.98 约为 102.0。差异约 9.1 单位，可能大于芯片方案之间的报价差。它不是某条产线的良率预测，也不允许用范围中点代替试产数据。更关键的是，如果失败集中在某一传感器批次，平均良率对下一采购批的预测价值很低；应保留供应批次、工站和缺陷原因关联，直至统计上可辨认改进或替代解释。

## 13.4 产品承诺会改变可接受的硬件配置

硬件预算需随产品级服务承诺变化。若销售承诺断网持续运行、快速响应、长期续航或高可靠告警，最低可行配置可能比一个能通过离线模型测试的开发板贵很多。反之，若 AI 只在用户主动打开的少量场景运行，则连续在线的设备和备用电池可能过度设计。消费者愿意承担的重量、噪音、热感、外观体积和换电频率，会约束可用于算力与传感的空间。工业设备还会受环境、安装、访问和校准窗口限制。成本模型应把承诺写成可验证条件：响应分位点、运行温度、可用率、断网功能、事件频率以及寿命期限；“智能”这个名称没有工程规格。

配置选择也有影子成本。把一种高阶芯片固定到单一产品，可能减少未来替代料；引入云接口可能缩小端侧资源，却增加地区连接与服务中断风险；一颗附加传感器增加数据能力，也增加装配校准和隐私同意流程。若量产数未达到预期，最低起订量、库存减值、过时料和供货保留价会把采购风险转成现金损失。



SVSAC | sv-sac.org

因此，器件决策需同时报告每台成本、固定承诺、取消代价、可替代供应和需求区间下的库存暴露，不能用一个预测销量摊薄所有固定成本后再把该单位成本当成保证。

# 14. 研发、集成、验证和非经常性工程成本

## 14.1 单位经济的分母必须与开发投入的边界一致

设计验证、模型适配、工具链、集成、数据权利、治具和量产导入若为发布功能所必需，便属于决策的增量资源，不能因已记研发而忽略。应在同一决策期并列现金支出、内部资源占用、财务确认与单位分摊；资本化或费用化不使现金成本消失。第 18 章据此区分指标，第 25.8 节再加入稀缺工程产能及替代项目，而不以财务人工费率代替机会成本。

可把开发投入拆成一次性、重复研发和量产启动三类。一次性工程包括新硬件架构、核心模型集成、安全威胁评估、目标设备调优和自动化测试框架；版本重复研发包括地区适配、不同芯片、设备变体、客户定制、法规/网络变动后的再验证；量产启动包括工厂验证、供应商审核、治具和培训。若平台模块未来可复用，不能假定每个产品都重新支出全额；若共用平台需要持续适配和维护，也不能把全部研发当作免费。应按明确可复用单元及实际使用产品分配，另报告不分摊的企业级能力建设成本。

预期销量是分摊固定开发成本的分母，而不是成本事实本身。假设总开发现金支出为 60 万单位、上市后实际合格交付分别为 1 万或 4 万台，简单均摊分别为每台 60 或 15 单位。若销量未成熟、退出市场或一半设备采用不同版本，则平均数还要变。报价前使用预测销量计算可以作为情景，但必须标为预估；试点销量、签约意向和量产出货也不能互换。核算最好并列固定总额与多种交付量情景，避免把预计规模造成的低成本误写成技术固有低成本。

## 14.2 验证成本应按“被覆盖的产品状态”而非模型运行次数计量

嵌入式 AI 的测试对象不只是模型版本。传感器、处理器、固件、校准参数、预处理、外壳、温度、网络状态、用户语言和云服务共同构成一个待验证状态。每次模型更新可能要求回归测试；同一个模型移植到新芯片或新镜头，也可能改变数值精度、延迟、噪声和结果。若把测试成本按模型推理数量估计，将忽略样品制造、设备间差异、长期运行、极端环境、用户动作及故障恢复。反过来，对每种排列组合全面穷举又可能不可行；需要用风险依据选择边界覆盖并明确未测组合。



SVSAC | sv-sac.org

测试成本可拆为数据/场景构建、参考标签独立确认、样机与设备批次、自动化夹具、性能测量、可靠性/环境试验、人员评审、失败定位、复测、产品安全及上市资格要求（适用时）。关键是避免双重计算：一个工程师等待样机的日历时间不等于全额人工工时，测试台占用天数也不等于现金费用；但限制其他项目用台的机会成本可能真实存在。项目账本应分别记录人员实际投入时间、设备占用、采购支出、排队延迟、复测次数和因延期失去的销售窗口，并只在转换成经济成本时说明估值规则。

作者建议采用两层验证：先对功能在目标任务与失效边界上做工程接受，再对完整产品在代表性用户和场景中验证结果与负担。两层可以共用设备或标签，但终点不同。工程测试回答达到哪个输入/输出规范；用户试验回答是否正确解决目标任务、产生误报负担、能否理解和控制。只通过工程测试无法证明客户价值，只有用户总体满意度也无法替代高后果误差或鲁棒性试验。上市前样本应按设备序列、组件批次和软件版本追溯，失败与无结论样本都留在分母规则中。

## 14.3 供应和产品变更使研发投入具有路径依赖

高性能器件与特定 AI 工具链绑定时，供应商停产、固件更新、芯片替代或云服务接口改变都会引发重新验证。产品团队一开始将功能放在外部 API，可缩短原型开发；进入市场后，调用价格、服务规则、地区可用性和安全审计要求可能变化。端侧部署可以降低部分连接依赖，却提高硬件资源和固件维护成本。不存在天然“长期更便宜”的架构，只有在明确任务负荷、设备年限、服务条件和更新路径下才可比较。

变更成本需记录触发事件到证据重新建立的链条：影响哪些设备与用户；是否改变输入分布、算法版本或失效模式；哪些结果仍可沿用；哪些须重新测试；谁承担召回、回滚或客户通知。某个内存升级如果不改变结果规范，可能只需重新确认功耗与兼容性；镜头光学变化即使外观相同，也可能改变识别效果，不能沿用原视觉测试。对高后果功能，允许替代料的工程决策需在量产前证明替代条件和验证门，而非故障发生后才确认。

一个可比较的产品开发账本会将平台与 SKU 层成本分开。平台共用的模型运行时、更新机制和基础测试框架可按预计复用产品分配，也可按企业能力列示；地区 SKU 的语言、标签、联网认证和电源适配应归到相应变体。若每个地区特性和器件组合不断增长，测试矩阵可能随组合而非产品数量线性增长。可采用代表组合、兼容性约束或预定义等价组，但其节省依赖覆盖假设；抽测数量减少并不



SVSAC | sv-sac.org

意味着未知交互风险消失。产品经理需看到减少组合覆盖后省下的人时、仍未测试的边界和可能的市场暴露。

### 14.4 取消、延期和重启项目的成本不应被存活项目遮蔽

成功上市产品的毛利账本往往看不到已经取消的原型和开发分支。若 AI 集成功能在设计冻结后失败，前期定制硬件、数据工程、工具许可和供应商预付款可能无法复用。只对最后发布的产品分摊开发成本，会形成幸存者偏差：产品看起来研发效率高，但风险项目的损失由公司其它产品承担。项目组合分析应记录每个候选的研发起止、阶段取消、沉没支出、可回收资产与取消决定；不能因失败版本未上市就从总体投资回报分母中删除。

延期则可能同时增加成本并改变收入时点。增加几轮验证会占用人员和设备，但如果延期让产品避开旺季缺货或修复重大缺陷，延迟也可能带来收益。一个简单的“每延迟一周损失销售额”只在需求、渠道窗口、产品竞争和交付能力符合假设时成立。要估计延期机会成本，需要用未延期的合理反事实及订单/季节性证据；否则应展示区间情景而不宣称已损失某金额。取消后团队转去其他项目产生的产出，也可能抵消原产品的机会成本，不应把同一人工时重复计入现金支出和机会成本。

构造性对照可说明区别：产品 A 花 60 万开发，计划交付 4 万台但因传感器故障只量产 1 万台；产品 B 花 90 万开发，达成 4 万台并复用一套平台。按已交付数量简单均摊，A 为 60 单位/台，B 为 22.5 单位/台。若只比较已售产品，A 的困难会显示为单台高研发负担；若把可复用平台资产或取消产品的后续价值计入，排序可能改变。两个数字不构成会计利润率，更不证明 AI 使产品 A 失败。正确做法是展示现金支出、产量实现、复用价值与失败支出各自的假设，让决策者知道哪项假设驱动排序。

## 15. 产品版本、地区配置和上市延误的经济后果

### 15.1 一个产品名称可能掩盖多个功能版本

设备型号相同，用户实际拥有的功能可能因芯片批次、固件、功能开关、订阅、地区法律/网络条件或模型版本而不同。若只以 SKU 汇总收入和售后，企业无法判断实际销售的是哪种 AI 能力，也无法将体验、退货或服务使用关联到功能状态。功能版本应有可查的定义：硬件料号、模型与运行时标识、输入预处理、阈值、云端接口、默认设置、订阅权和支持期。版本号不仅为工程追踪，也定义了销售承诺和成本归属。



SVSAC | sv-sac.org

地区变体尤其容易导致“同一产品”指标不可比。某地区可能禁用音频功能、使用不同云端区域或被迫启用离线模式；另一个地区可能有更长保修、更复杂的语言校准或额外售后渠道。按全球平均调用数、退款率或客服成本会掩盖地区组合变化。比较前应明确目标市场权重，并报告区域内效应与汇总权重；否则一个地区销量占比增长就能让综合指标改变，即使每个市场的用户体验不变。

某些设备会在同一硬件上通过软件或订阅解锁能力。此时出货量不是可服务的功能量；需区分已具备能力设备、资格设备、实际激活、付费功能、可用设备月以及有效运行时长。激活率可能受默认设置、渠道培训和配网质量影响；活跃设备存量的分母需包含断网、停用和尚未到期等状态。若订阅功能启动后被长期关闭，收入与活跃使用也不能视作同一结果。功能版图和商务计划若不对齐，销售预测会把未授权或不可用状态算入高价值用户池。

## 15.2 更新与支持的成本取决于在用设备存量

一次性研发成本发生在上市前，更新与服务成本则沿装机设备年份持续。一个产品上市第一年部署一万台，如果承诺五年安全/功能支持，后续每年仍需维持兼容、处理缺陷、支持经销、监测服务和验证新版本。用户退役与闲置设备会改变服务需求，也会留下未更新风险。按新出货量年度摊销云成本会漏掉老设备调用；按注册设备数计费又可能高估停用设备。需要建立设备安装队列并记录状态变化、调用分布、支持工单、更新成功率、服务终止与设备退役。

云调用成本也不是一个固定 token 费率。实际计费对象可能是请求、推理时长、并发、存储、出入站传输、日志、告警保障或不同地区服务等级；还会有峰值容量、故障回退和安全运营成本。企业可选用自建服务器、第三方 API、边缘代理或本地模型，不同方案的固定投入和边际成本随装机规模变化。比较时应使用实际负载曲线和服务协议，而非把某个当前公开单价乘以预测请求量当作总成本。即使单次推理近似免费，客户支持、服务故障和版本兼容仍要消耗资源。

对端侧产品，生产后的电池替换、温升投诉、固件储存空间和设备回收都可能关联到 AI 功能。若频繁推理使电池寿命缩短，成本不仅是电池材料，还包括上门服务、用户停机和品牌补偿；但因耗电而退货必须有对照或可靠故障归因，不能从相关投诉直接推因果。监测阶段可把同型号设备按批次、版本与功能暴露记录电池健康、充电频率、服务请求和运行环境，再用随机部署或可信对照识别功能增量。未经识别的风险应进入敏感性而非确认为 AI 成本。



SVSAC | sv-sac.org

## 15.3 上市时间、采购窗口和商业价值的冲突

AI 产品增加的验证和供应准备可能延后发布日期。此时应区分三种时间：工程完成日、开始可售日、具备稳定交付能力日。早期预售或小批量出货不代表可全量履约；产品发布也可能只是市场公告。成本账本应记录设计冻结、供应商承诺、生产爬坡、渠道入库及用户交付的阶段门。若固定销售窗口已经过去，错过时点的需求损失可能存在；若等待验证避免大量退货，延迟却可能提升长期价值。用工程迭代速度直接声称商业提速，跳过了生产和渠道的关键路径。

制造业的约束常跨职能：芯片供应的交期可能长于软件开发；法规样品认证有排队；经销商要给门店和客服培训；更新包要通过设备兼容测试。并行工作可以缩短日历时间，却不一定缩短总人工投入或风险暴露。可画关键路径网络，记录前置依赖、资源共享和返工回路。单纯将每条活动的时间相加会高估并行可行工作；只计算最长路径也会遗漏共享专家与设备形成的资源冲突。没有实际项目日志，不应虚构一个普遍可节省的日期比例。

一个构造例中，算法验证需 4 周、硬件认证 6 周、工厂治具 5 周、渠道培训 2 周。若彼此完全独立并行，最早可能在 6 周之后具备条件；若硬件认证必须在治具冻结前完成且培训依赖稳定功能版本，总周期会更长。若认证失败导致硬件改版，原并行计划产生返工。例子仅说明依赖关系如何改变日期，不是企业项目估时，也没有把资源人员投入视作零成本。实际计划应由阶段日志估计活动持续期分布，记录取消和返工，而不是拿最乐观的工程路径作为业务承诺。

## 15.4 版本分裂如何侵蚀毛利

沿用 14.3 节的状态与覆盖定义，本节只增加地区机会取舍：兼容矩阵区分共用组件与独立验证组合，更新影响面、用户数和后果风险决定维护负担。减少变体可节约平台成本，也可能损失地区功能、履约或环境适配。该权衡须与第 18 章一致的成本边界联结，不能把统一版本本身当作毛利增量。

定价层级和订阅还可能把毛利成本移到不同账户。硬件以低毛利出售、依赖订阅回收云推理费时，若订阅转化低、取消率高或云成本超出预测，硬件销售本身就无法支撑长期服务。若服务成本仅按平均活跃设备估算，重度用户可能造成尾部成本集中；若限制使用次数，又可能破坏价值承诺。建议按购买队列和使用强度分布报告每设备贡献、续费、服务消耗和退出状态，并对设备年限、模型调用量、地区价格与币种波动做敏感性分析。平均值应与长尾设备的预算风险并列，不替代其分析。



SVSAC | sv-sac.org

同理，更新可以创造新收入，也可能是原有购买义务的一部分。若用户认为功能持续更新是购买承诺，收费升级未必是新增变现；若合同明确提供基础功能并另售高级模型，续费可能是单独服务价值，但同时增加告知、兼容和退款工作。不能通过把研发成本分配给硬件、将服务成本转给云部门来“提升毛利”。对外和内部评估都应跨硬件、软件、云、售后与渠道建立同一设备或账户周期的经济账，并区分公司、渠道和用户承担的资源。

# 16. 云端、连接、能耗与持续推理：把服务成本放回产品边界

## 16.1 账单单位不等于经济驱动因素

云端 AI 的账单可能按调用、输入输出长度、加速器时长、并发容量、存储、传输或服务等级组合收费。单次调用价便于初筛，却不直接等于单台设备成本：调用频率由用户活动、自动触发阈值、重试、诊断、错误回退和软件设计共同决定；同时，一个请求可能触发多个模型或多次检索。工程日志应能将调用连回设备版本、任务、地区、时间、处理结果及是否重试，在隐私许可和数据最小化范围内采集。没有这种计量，财务团队只能把云总账单按出货量平均，活跃用户差异和未出货设备共用的固定容量都会被掩盖。

收费量与资源量也可能不同。服务合同可能要求预留峰值并发，而平均调用远低于峰值；突发故障后大量设备同时重连，形成峰值负载。长期合同可能以较低单价换取最低消费，实际量不足时也要支付；按量采购则可能在高峰期更贵。成本分析应分别计算实际使用量、订购容量、最低承诺、峰值冗余和故障应急资源。若某次大规模重试是模型错误导致，不能把它归为“用户需求”；若峰值是产品被大量使用，限制并发或加价可能损害产品价值。架构、服务等级与营销承诺必须共用同一负载情景。

云服务的单位成本通常随时间和规模变化：模型替换会改变输入长度、缓存、推理时长和准确度；供应商可以调价；区域迁移会改变数据传输和合规成本。模型压缩可能降低每次计算资源，却增加低置信度请求的人工升级或二次模型调用。对云端总成本有帮助的真实优化应在固定质量和服务约束下验证，而不是只报告单次推理毫秒数。若低成本版本的拒答率上升、用户反复重试或客服接管增多，表面单价节省可能是成本转移。

因此应按设备队列建立成本视图：每代产品的部署数量、在用月份、启用比例、任务频率分布、升级和停用状态、区域及故障重试。固定平台工程费、最低



SVSAC | sv-sac.org

云承诺和监控系统可按设备年、请求量或贡献分配，但每种分配回答不同的问题。产品级决策可看完整增量现金，部门绩效也许需要管理分摊；前者不能由后者取代。对于共享模型服务，另报告被多个产品共用的基础设施总成本与某一 SKU 的可避免增量，避免既把平台固定成本全部压给新产品、又把它当作零成本。

## 16.2 端云架构的经济比较应统一任务和服务边界

端侧增加制造资源并可保留部分断网任务，云端增加模型与集中更新能力但暴露于连接、中断和运维，混合路由可能省调用却增加状态一致性与追踪负担。比较固定任务结果、输入频次、时延、用户覆盖、支持期及同一设备队列/负载；新增问题是成本何时发生和由谁承担，而非再按芯片或 API 单价给架构排序。

一项可复算的架构比较可将成本写成全周期现金流：每台一次装入的器件增量 H、每台其他制造增量 B（两项不重叠）、端侧一次性工程投入 F_edge、服务固定成本 F、每个有效设备月云费 c、连接和流量成本 n、维护更新 M、故障和支持成本 S。N 台设备运营 T 个月时，云端总成本简化为 F+N×T×(c+n)+M+S；端侧方案则为 F_edge+N×(H+B)+M_edge+S_edge。该公式不是完整估值：销量分批、折现、故障关联、退役及残值都可能要加入，但它能暴露一个常见错误——将每次云调用价格和一次性芯片价格直接比较，却漏掉其中一个方案的固定服务和量产成本。 H 和 B 单位为金额/台，F、F_edge、M、S 为该队列期间总额，c 和 n 为金额/有效设备月；N×T 仅在全体设备均活跃 T 个月时成立，分批激活或退役时改用各设备有效月份之和。相同基线及已在 H/B 中的投入不得再计入工程总额。

例如纯构造地假设，云端每设备月增量服务费 0.8 单位、支持期 36 个月，端侧硬件每台增量 20 单位，忽略其他成本，则每台 36 个月云费 28.8，高于端侧器件增量 8.8。若云端用户实际激活比例只有一半、端侧需增加更高容量电池，排序可能相反；若云端最低合同使闲置成本照付，激活率下降也不会同比降低账单。该算例没有真实云报价、设备量或耗电数据，也不能决定架构。它要求后续采集真实合同、使用分布、功耗和支持支出，并对共用基础设施、设备报废前服务年数和容量承诺做敏感性分析。

混合路由还要区分技术失败和业务策略。若端侧先筛选、云端只处理置信度不足事件，整体人工复核量与云账单取决于端侧筛选覆盖和漏检后果；不能用云模型在选中请求上的成本除以全部任务数，再称全链路成本。若为了控制费用提高上云门槛，端侧无法处理的事件可能被静默拒绝。路线选择应预先定义什么情



SVSAC | sv-sac.org

况本地返回、升级云端、拒绝或请求用户确认，并在测评中记录每类路径的资源、正确性和用户结果。

## 16.3 能耗边界应与产品使用及生命周期口径一致

小模型论文的单次推理能耗或延迟只覆盖指定测试边界，不自动覆盖预处理、传感器、唤醒、网络、屏幕、待机和电池老化。MLPerf Tiny 的性能边界可用于理解基准测量窗口，但不能将单次推理指标外推为整机年度电费或电池寿命。[2] 产品功耗需在明确环境中测整机：典型待机、典型任务、连续高负载、网络重连、更新和异常循环；按用户任务频率计算设备月或设备年能耗，并报告设备分布而非只取均值。若功耗变化可被用户感知，还需测量续航、充电、发热和停机。

假设某设备每次推理增加 0.2 焦耳，典型用户每天触发 100 次，则该项推理理论上每天增加 20 焦耳。这个数无法单独说明电池寿命，因为尚未计传感器常开、无线传输、休眠状态、系统唤醒与模型失败重试；也不能直接乘以全球出货量推导环境收益或损害。对能耗价值敏感的产品，现场应从独立功率测量和日志构造任务负载分布，再估算各区域电价下的运行成本，并将电池替换和设备报废单独记账。超低频高后果事件与常频便利任务也不应共用同一平均频率。

节能收益只有相对于合理基线才有意义。若 AI 功能替代旧版一个持续运行的通信过程，可能降低网络或操作能耗；若功能新增传感与云连接，则净消耗可能增加。要验证“AI 使设备更节能”，需对相同服务水平与用户行为比较新旧方案，控制外界温度、使用强度、网络覆盖和硬件寿命，并观察用户是否因自动优化而增加使用。功能引发的反弹效应可能抵消单位任务效率。没有基线和用户行为测量，低功耗芯片指标不是整机能效结论。

## 16.4 收费、使用和价值兑现会产生不对称风险

一次性设备售价能提前收回部分硬件投入，但云端推理、更新和客服成本会随在用设备年持续；订阅则可按月回收服务费，却提高取消风险和支付摩擦。企业可能同时承担早期开发与容量投入，而只有部分用户启用或付费。若计划以订阅覆盖云成本，需分别估计合格设备池、可激活设备、实际付费账户、留存月数、使用长尾及退款；同一个设备可能共享家庭账户，设备数与付费客户数不可混算。

云端长尾特别重要：少数高频用户、重试异常或大模型路径可能占用大部分调用。按平均次数估算服务成本，会低估高分位负载和异常用户治理成本。产品可采用限额、降级、缓存或分层定价，但这些策略会改变用户可获得的功能，也



SVSAC | sv-sac.org

可能产生新的投诉和公平性问题。资源限制应被描述为产品政策变量而非基础设施细节，并测试对任务完成、用户退出与成本的联合影响。

账单发生时间和收入确认也可能错位。设备在当年出售，客户可能在多年后更新、退订或产生保修支出；提前收取的多年订阅费仍对应未来服务责任。比较不同产品方案时应将一次性研发、产线变更、收入、渠道回款、云支出和后续维修按时间列示。仅看上市季度毛利可能把持续支持负担推到未来，单纯生命周期总额又可能不区分资金占用和风险。管理层可以同时查看单期经营指标、队列现金流和情景净现值，但不能在未说明折现率、存活期和续费条件时把它们混成一个“生命周期毛利”。

# 17. 售后、保修、更新与退役：谁承担产品功能的持续义务

## 17.1 保修与服务成本需要故障分母及责任归因

内嵌 AI 产品的售后事件可能来自硬件、模型、连接、训练数据不适配、用户安装、第三方平台或使用环境。若每个投诉都归给 AI，费用会被高估；若 AI 相关失效表现为普通退货、客服咨询或电池问题，又会被低估。工单需关联产品版本、购买渠道、功能状态、事件严重度、解决时间和责任判定，另记录无法归因与争议状态。投诉数量可作运行警报，但只有经独立确认的故障与有效暴露分母才能构成可比较的产品故障率。

保修成本可写成各类事件率×单位处置成本，但事件率本身依赖设备年龄、版本、地区和使用暴露。早期产品队列可能尚未经过高故障年龄，按上市头几个月外推五年保修会偏低；同时，新固件修复也可能改变风险。较稳妥的分析按设备售出批次跟踪至指定期限，区分已观察、右删失、迁移和无法联系的状态，并分别报告实际现金支出与未决义务准备。若设备由经销商承担初步客服，品牌方的现金账单虽下降，渠道训练和返利仍是产品经济成本的一部分。

质量责任还影响功能可用边界。对于建议类输出，产品可以显示来源、让用户拒绝并保留人工确认；对于自动控制，应明确独立安全机制、异常停止与故障后状态。售后处理不仅是补偿损失，也是新版本数据来源，但客户工单样本常过度代表严重或可见失败。把投诉直接用于调优可能提升投诉用户上的性能，却无法证明无投诉设备安全。设计审计应随机检查报警与无报警设备，且将复核成本纳入持续运营成本。



SVSAC | sv-sac.org

## 17.2 更新支持期与云服务终止必须作为可计量承诺

产品承诺支持更新多久，决定企业需要维持哪些设备、工具链和安全能力。设备型号可能继续销售多年，导致同一时间内有多个硬件代际和软件分支。每条分支都需要兼容验证、漏洞处置、签名基础设施和客户沟通。停止更新可以降低未来支出，却会把网络风险和功能退化转给客户；继续支持则需预算工程值守、供应商许可和回滚能力。NIST 制造商网络安全材料可作生命周期治理背景，但它们本身不能替企业确定商业支持年限，也不能证明更新一定符合用户价值。[3][4]

支持承诺应可被审计：支持终止日期、设备仍能离线工作的功能、云端功能依赖、数据导出、账户关闭和迁移方式、更新失败的处理期限。若产品更新改变传感或自动化行为，应有版本通知和功能回退路径。假设 5 年支持期内，每年每个活跃软件分支需 0.15 个工程全职人年、测试设备占用 120 小时，若同时维护 3 条硬件分支，资源就不是一个单产品功能的零散维护。该假设算例只显示分支与持续维护相乘的结构，不能当行业人力基准；真实成本取决于产品复杂度、缺陷率、供应商与内部架构。

云服务终止可能使设备部分功能失效，即使物理硬件仍可正常使用。供应合同到期、模型版本不兼容、地区连接禁用、账户关闭或支付失败均可能造成服务断裂。对消费者而言，产品的剩余价值可能因此改变；对制造商而言，迁移、补偿、延长服务和法律咨询也会产生成本。各类退出需要分开记录，不要把设备退役、用户主动关闭、服务关停和厂商停服统称为流失。若产品承诺“离线可用”，验证应测试网络中断期间的具体功能，而不是只记录设备能否启动。

## 17.3 退货与替换对利润的影响不止退款金额

退货减少净销售收入，并可能产生逆向物流、检测、数据清除、翻新、重新包装、渠道扣款或无法再售的折价。换新机则可能重复制造成本且延长客户服务周期。AI 功能若造成误警或无帮助，用户可能因“质量不符合承诺”退货；但促销后改变主意、配送损坏或竞品降价也会产生同样结果。退货调查应采用独立原因分类并保留未决，比较组间差异时按原分配与统一成熟退货窗估计，而不只分析给出解释的客户。

如果经销商先承担退款，制造商可能稍后通过返利或扣款受影响；若平台承担退货运费，渠道合同可能重新定价。退货率应和每件退货的实际净回收、处理时长、翻新率、折价及后续再次销售连接。多次换新与一次退款的直接现金不同，客服体验也不同。只用退货订单数乘零售价作为损失会忽略可回收库存，同时可能重复扣除销售成本和退款金额。



SVSAC | sv-sac.org

一个假设的退货成本账可将 1000 台发货按退货原因划为：20 台由功能错误、30 台为配送损坏、25 台为购买后改变想法、15 台原因未明；其余 910 台暂未退。数字是示例，不能据此声称某类 AI 产品有 2%退货率。若功能错误退回的每台处理成本为 8 单位且可回收库存净值 70 单位（须注明估值时点，非同时代表售价与原始成本），赔付净额需由真实交易记录决定；原因不明的 15 台不能自动归 AI，也不能从总体质量报告删除。对随机分配的对照与 AI 组，应报告各类退货风险差和不确定性，另给出把未明原因全部归 AI 或全部不归 AI 的敏感性界。

### 17.4 网络安全与数据运营的持续支出需穿透到产品组合

AI 产品可能收集输入、推理日志、诊断和反馈数据。每种数据的收集目的、最短保留、访问权限、传输、用户删除和服务终止处理都关联长期运营。为了提高模型效果而保留更多历史数据会增加存储、访问、权利确认、删除请求和泄露事件处置工作；减少留存则可能提高重建问题和模型漂移的难度。成本与风险不能只用每 GB 云储存单价估算。需要记录责任人员、访问审核、删除任务、异常响应、供应商审查和用户请求处理资源，并报告数据对产品功能是否必要。

当多个 SKU 共用数据平台，安全和隐私基础设施可能属于共享固定成本，而某个产品新引入的敏感输入会产生边际评估和额外保护成本。按收入比例摊分可以服务部门绩效，但不回答停止该产品后企业可避免多少支出。产品级决策应区分可避免成本、共享平台成本、外部事件尾部风险和合同/用户承诺。以概率极低为由把严重数据风险记作零，或将全平台安全预算全部分配给一个新 AI 型号，都会扭曲上市决策。

持续支出同样可能通过质量反馈转化为价值。用户允许上传失败片段，可帮助定位模型问题；但不完整或偏向投诉人的数据会使优化方向偏离全部用户。修复模型可能需要重新标注、复测、分批发布并追踪回退；这段工作是运行支持的一部分，不应在模型训练结束时视为终止。若产品完全无法获取诊断数据，企业仍需设计本地日志、用户导出或返修检查，成本只会换一种形式。数据权限限制应写入可观察的维护流程，而非被当成允许无限收集的技术前提。

## 18. 毛利率、单位贡献和生命周期现金流不是同一个结果

### 18.1 先固定分子、成本范围和期间

本章将前述成本分成三种结果：按适用列报政策定义收入及销售成本的毛利率，扣明示变动成本的单位贡献，以及纳入研发、销售、云支持、退款和时间价值的生命周期现金。资本化、保修、服务收入、折扣、分摊与跨期确认依企业及



SVSAC | sv-sac.org

准则而异，须由财务负责人调节管理账与批准列报，本篇不提供通用会计或审计标准。第 25.7 节进一步区分报价政策的机会均值与买家贡献，第 25.8 节比较容量替代后的组合净值，均不能仅凭毛利率命名。

计算之前应明确实收净价是否扣渠道折扣、退货、优惠券、税费与退款；单位成本是否包含制造良率损失、物流、售后、云调用及服务承诺；研发支出如何列示；按哪个期间和哪个销售队列统计。若将 8 单位云支持计入 AI 版单位成本、却把非 AI 版的同类售后放到期间费用，比较就不一致。若平台服务同时支持多个产品，管理分摊与可避免成本也应分开。报告率时同时给出金额和数量，避免百分比变化掩盖总贡献变化。

## 18.2 单台多赚并不等于毛利率更高或总利润更多

以下使用作者构造的无量纲价格单位和产品数量。基线非 AI 产品当期售出 10,000 台，每台净收入 100、定义成本 60，单位差额 40、差额率 40%。AI 产品每台净收入假设 120，硬件与制造成本 68、已分配支持和云成本 8，单位差额 44、差额率 36.67%。AI 产品在此定义下单台差额增加 4，却因成本和售价比例不同，差额率下降 3.33 个百分点。这个算例并不意味着市面产品实际有 20 单位 AI 溢价，也不称上述单位差额为某法域经审计的法定毛利。

若 AI 版本只售 8,000 台，AI 产品期间差额为 352,000，基线为 400,000，差额少 48,000。可逐项复算：保留的 8,000 台比基线每台多 4，共多 32,000；少售 2,000 台若本来按基线每台产生 40，则少 80,000；32,000−80,000=−48,000。再扣除一次性增量研发投入 150,000，定义范围内的经济比较差额为−198,000。若两版本各售 10,000 台，AI 版当期差额为 440,000，比基线多 40,000，仍未覆盖 150,000 投入；在各单位参数固定的假设下，AI 版售 12,500 台时，12,500×44−10,000×40−150,000=0。12,500 不是销量预测，是此构造账本的盈亏平衡解。

该例说明售价更高、单位差额更高、毛利率更高和总投资回收是不同命题。分析者不能只挑一个有利指标：若论点是客户溢价，需识别有效市场价格；若论点是单位贡献，需定义可归因成本；若论点是毛利率，需按照真实财务政策确认成本；若论点是项目回报，则要包括销量差、研发支出、服务年限、退货、折现和退出。成本拆分中的“支持和云”采用每台配置值只是展示结构，不意味着所有企业都按销售台数分摊固定服务。



SVSAC | sv-sac.org

## 18.3 队列和时间价值改变盈亏平衡点

一次性销售与多年服务混在当期单位贡献中，可能让高销量看似立刻覆盖研发，却忽略安装基数后续成本。设产品每台贡献 c，新增研发现金 F，之后每设备年维护现金 m，服务年数 T、年度存活比例 q，且销量按期分批，成本现值需按队列计算。简化的盈亏平衡条件不是单一 Q=F/c，而是收入贡献现值−研发支出现值−每批设备的支持与退役负债现值≥0。若付费设备留存低，订阅收入下降；若未付费设备仍享受保修与安全支持，服务成本却不会同步消失。队列模型需区分付费资格、使用活跃和支持责任。

折现与存活率的假设对长寿命设备尤其敏感。将所有服务费按稳定平均值乘以五年，可能低估早期故障高峰，也可能高估退役后的调用。订单取消、保修索赔延迟和订阅年付预收会改变现金与利润的时点。分析可采用低/中/高使用寿命及活跃度的场景，另报告未贴现累计现金与贴现净现值，并公开折现率和时点约定。若管理层只需要短期产能决策，也可以单独看当年边际贡献，但不能把该短期指标称为长期毛利已验证。

当算法输出让客户避免停机或降低电费时，价值可能归客户，而非制造商。企业只能通过溢价、续费、差异化留存、降低保修或扩大销量捕获其中一部分。客户总价值不能自动写成公司收入。相反，功能的外部效益可能使厂商难以收费，却仍有社会价值；若核心问题是企业毛利，社会效益应另列而不能塞进商业回报数字。支付渠道、经销议价、竞争者复制和服务替代都会影响价值获取。

## 18.4 成本范围的反例与审计式核对

成本口径可以在不改变产品任何物理特性的情况下改变报告指标。例如把云支持划入销售成本会降低产品毛利率；放入研发或期间费用则毛利率数字提高，但企业现金不会因此改变。如果 AI 版服务依赖云而非 AI 版的配套应用同样使用云，必须按同一分配规则比较；如果只有 AI 版需要持续模型调用，则应核算这项增量而非只看当前账单。费用列报的变化不能被当作成本节省，也不能仅凭毛利率表面高低判断配置优劣。

管理核对可制作从现金流到产品指标的桥接表：总现金研发投入、量产制造现金、云合同支付、实际净收入、售后赔付、当期确认成本、资本化部分、未摊销服务责任及按管理规则分摊的每台成本。每项应链接原始合同、采购记录、批次及账户队列；无法归因的共享费用保留单独池，不强行全数分摊。若一个产品毛利率高但需要大量前期营销和专属客服，项目现金回报可能较弱；若毛利率稍



SVSAC | sv-sac.org

低而复用平台使后续产品成本下降，组合价值可能较高。两种结论须用不同的分母和时间范围说明。

最后，因果结论需要保留可比产品和明确版本。单看同一公司新旧产品财务报表无法排除同期涨价、芯片成本波动、渠道结构、促销、供应短缺、竞争反应与汇率变化。最有用的证据来自预先定义的产品包对照、真实价格/数量、相同财务分类的成本账及足够成熟的售后队列；条件不满足时，应收窄为“描述了某段期间的账面表现”，而非“AI 使毛利提升”。产品化的商业卖点必须经过功能、购买、成本、履约和时间几层测量后才可判定。

# 19. 从智能恒温器看嵌入式功能的反证：工程节能并不自动成为用户行为

## 19.1 预测的功能结果可能被使用者目标重写

最直接的“智能产品价值”论证常从一个物理机制开始：算法识别空屋或需求低谷，自动降低设备运行；因此节能或延长寿命；用户由此获得更低成本。机制链看上去清楚，却隐含用户目标与设计目标一致。家用恒温器的用户在节省电费、保持舒适、避免重复操作之间权衡；如果智能设置让房间太热或太冷，用户可能覆盖自动程序。此时功能并非简单失准，而可能是在实现另一个用户优先级。产品评估若只记录建议是否符合工程模型的最佳点，会把真实使用中的效用取舍误标为不采用或失败。

Brandon、Clapp、List、Metcalfe 和 Price 2026 年 8 月的作者稿报告两项加州现场实验：家庭被随机分配为有资格获得免费安装智能恒温器或对照，18 个月观察高频用电及天然气数据。作者报告对总体能源使用的估计在统计和经济意义上都难以与零区分；处理组安装比例为 73%。他们进一步用安装用户的设备交互日志分析设定值、用户覆盖和温度回调时长，指出工程节能预测依赖的使用假设未必对应家庭实际取舍。[7] 作者主页将该稿列为工作论文，并标明条件接受《American Economic Journal: Economic Policy》；这里按 2026 年 8 月作者稿和作者主页状态引用，不称其已正式出版。 该文主表采用随机资格作为安装的工具变量，报告条件下的安装效应；附录中的意向处理结果虽被提及，但该附录未完整核对。使用日志仅覆盖部分安装家庭且时间窗短于能源观察，不能当全样本机制中介证据。

这个结果提供了针对“能力可用即价值实现”的直接反例，但其边界也要保留。试验提供的是可领免费设备资格，不是所有家庭实际安装；随机处理估计首



SVSAC | sv-sac.org

先对应被招募且符合资格的家庭及这套产品包。所研究设备是 2012 年前后智能恒温器，远早于当前一代通用生成式 AI 产品，也包含远程设置、用户可覆盖和其它功能，并非一个孤立 AI 模型或硬件溢价试验。它不识别消费者愿付价格、制造商毛利，也不证明节能为零以外的舒适价值不存在。论文对使用日志里覆盖行为的机制分析是处理后的描述，不是随机操纵“覆盖”所得到的中介因果效应。[7]

## 19.2 人机交互是功能因果链中的中间状态，不是噪声

设备可以给出正确、及时、节能的程序，用户也可以有理由不按程序行动。忽视家庭到家时间、幼儿舒适、过敏、老人健康或房间分区，可能让经济模型推荐的方案在实际价值上更差。反过来，用户手动设定的温度比能耗最优值高，未必意味着产品失败；若智能设备让舒适控制变得容易，用户得到的私人效用可能提高，但节能目标没有达成。厂商不能用节能模型的最优设定替客户定义唯一结果，亦不能用“用户体验”把可量化的耗能失败抹掉。

这一逻辑也适用于工业产品。维护预警可能判断轴承即将失效，却在工厂没有备件或停机窗口时无法带来行动；自动分拣可能提高识别率，却增加一线人员的复核负担；智能相机可以调参数，但操作员可能因误报风险关闭功能。系统提示被确认、忽略、撤销或延迟，其背后有不同的任务、权限和资源约束。只看模型准确度会丢失产品能否嵌入实际工作；只看功能开启率又会把自动默认触发误当作主动接受。

因此应把交互日志设计成机制测量，而非“使用多即喜欢”的替代指标。对于可覆盖功能，记录自动输出、用户覆盖、覆盖方向、时间间隔、之后的实际任务结果和是否再次启用；对于手动确认功能，记录采纳、拒绝、编辑、未响应以及求助。对高频任务，应将每单位有效任务暴露与误报、漏报、动作时间和人工接管相连；对低频功能，应报告设备月/设备年和实际任务发生次数，而非只有帧数或调用次数。交互日志的抽样要覆盖未活跃设备与未报警时段，并预先说明哪些日志缺失、哪些用户退出不可追踪。

## 19.3 自动化承诺与控制权形成不同产品价值

“全自动”能减少用户操作，却也把误判后果和纠错成本转给了产品厂商及用户。用户越难恢复旧状态，自动化的验证要求就越高。提供预览、延时执行、确认、撤销和手动模式会增加界面与支持开发成本，却可能降低错误执行损失。某些产品可以让用户在习惯形成后逐步转到自动操作；另一些产品因风险不适合从建议升级到自主动作。功能权限不应作为模型能力的自然延伸，而要按错误代价、可逆性和用户理解制定。



SVSAC | sv-sac.org

这种权衡还影响愿付意愿。买家可能喜欢“自动调节”作为一句话卖点，却实际更愿意购买可选择自动化程度的设备。功能说明若只写“AI 自动优化”，用户可能假设自己可控制目标；现实中系统根据厂商目标调参时，信任受损。界面若显示节能预测值而不揭示假设，用户可能将工程模拟当作个人账单承诺。产品卖点的可验证内容应与用户能掌控的目标相符，例如在相同舒适边界下减少多少运行时间，且说明天气、户型和能源价格改变后为何结果不同。

厂商也可能把行为覆盖视作设计失败，因此追加 nudges、提醒或更严格的默认设置。然而，若目标只优化设备指标，提醒可能让用户承担更多注意力成本；若用户持续覆盖同一自动规则，应检验这是否表明需求定义有误。反过来，完全交由用户每次批准可能让产品不再真正节省认知劳动。管理选择需要权衡功能效用、用户控制、持续操作、错误责任与服务成本，不能预设更自动一定更智能、更有溢价。

# 20. 三类产品的嵌入机制比较：价值定义、验证单位与利润捕获

## 20.1 家用舒适/能源产品：舒适边界比模拟节能率更关键

在家用温控产品中，潜在机制可以表达为：传感器/行为输入→推荐或自动设定→设备运行时长和温度→能源账单与舒适体验→设备购买、订阅或能源服务支付。关键效用并不单一：节能是能源公司的目标，舒适是家庭的私人目标，方便是操作成本，售后还可能涉及安装兼容。作者构造的研究可将家庭或住宅作为分配单位，比较相同硬件的 AI 控制、传统定时规则及人工设置，按预先定义的温度舒适范围报告能耗、覆盖、投诉、设备能耗和退货。每户而非每小时读数是随机推断单位；每小时数据只是重复观测，需要按家庭聚类或用预先规定的家庭级汇总。

智能恒温器现场实验表明，工程节能机制在真实用户覆盖下可出现明显差距，但不代表所有设备都将无节能。作者指出当代某些设备有占用感知或自动空屋设置，可能改变用户是否按工程假设行动。[7] 这构成竞争机制：新的传感和控制确实可能减少手动覆盖，也可能让舒适目标更精细地实现，或产生更频繁的“温度追逐”。对照应同时观测能源计量与舒适端点，并避免仅筛选遵循预设调节的用户。处理后只比较“听从自动规则者”，会把潜在不采用人群从总体产品效果中剔除。



SVSAC | sv-sac.org

经济捕获取决于谁付款。能源公司可能补贴设备以获得系统负荷价值，制造商也许靠硬件差价或订阅收费，用户则可能购买便利或舒适。免费安装试验无法识别真实支付意愿；能源节省即使存在，也未必由硬件品牌捕获。设计中要分别测净价、安装/渠道补贴、订阅续费、安装成功、设备在用和服务支出。若能源收益由公用事业获得、用户未看到账单差，品牌的商业模式需依赖服务合同或渠道付款，社会效益不可直接记为厂商毛利。

## 20.2 工业预测维护：可行动窗口和停机责任决定功能价值

工业设备内嵌 AI 常以异常预测或寿命估计做卖点。其链条是信号采集→异常评分→维护决策→备件和工单→停机/故障/产出→客户续购或服务费。指标应以设备故障事件、维护窗口、工单和产品停机小时为主，而非传感器帧或模型评分。若同一设备在没有故障时高频产生数据，这些行不是独立故障样本。产品客户可能更关心“在计划停机前给出可行动提醒”，并不需要准确预测某个抽象寿命；若预警早得太多，备件库存成本上升，过晚则无法避免停机。

一个假设的应用对照可按生产线/设备簇分配 AI 辅助维护或既有规则，事先固定可处理机器类型、故障分类、行动提前量、备件可用性和停机记录。结果可包括：预测正确且行动的故障数、错误告警引发的非计划检查、漏报后损失小时、维护人员工时、备件浪费和设备可用率。若同一个维护团队覆盖多个分组，AI 告警可能挤占控制设备的维护资源；如维修人员从 AI 组学习并改变其它组流程，组间差异被污染。分配层级应接近资源共享单位，主分析保留被分配设备簇，不以“最终执行了建议的机器”选样。

商业价值的计算不能假定一次停机避免即为完整收入。制造商销售的可能是传感器、嵌入式授权、云诊断或维护合同；客户获得的是减少故障与库存，成本可能包括数据集成、校准、验证与现场支持。若 AI 版用订阅收费，应看客户可见节约的净值、实际续费以及服务履约；若设备是一次售出，制造商必须说明如何捕获下游收益。也需考虑非 AI 维护规则或更频密检修是否能以更低成本实现同等结果。若传统阈值同样减少停机，AI 的新增价值是成本或跨设备适配能力，而不是“AI 预测”本身。

## 20.3 高后果设备：准确率之外，失效责任和拒绝行为构成产品属性

医疗、交通、工业安全或健康监测设备中，错误输出的严重度和可逆性高。产品功能不只是预测或分类，而包括数据质量检查、拒绝输出、提示不确定性、人工复核、动作权限、保存记录、异常退出及售后。分母要按设备任务、患者/操作员、班次或独立事件定义，不能从大量窗口数推断独立客户可靠性。结果还要



SVSAC | sv-sac.org

区分及时识别、误警、未识别、拒绝、复核未完成、错误执行及延迟损失。一个高准确度模型可能因关键亚组输入条件不满足而频繁拒绝；拒绝能保护用户，也可能使产品在高价值时刻失效。

对安全敏感场景，随机试验不能为了研究目的剥夺合规标准护理或既有保护。可先在离线影子模式比较输出与独立参考，再通过受限建议功能、模拟器或不改变最终安全动作的用户研究评估可用性；若到真实部署，则由伦理/法规和产品风险治理确定可接受干预。具体法规义务依设备、用途、市场和时间而异，本章不把一般验证方案称作认证规则。对任一错误都应设处置路径：发现后暂停版本、通知责任方、确定影响设备范围、回滚/更正及复核，不把事故风险仅转成用户同意书文字。

高后果用途的利润也存在非线性。一个极少见但严重的产品责任事件可能远大于日常云调用节省；然而以单一极端事故的假设风险推导整个产品不能销售，同样不充分。应先列出严重危害路径、暴露设备年、保护层和故障可恢复性，再对频率和损失作情景范围。若产品价格包含专业服务、保险或现场升级，需要把相应成本和续约价值放进比较。对客户来说，置信度解释、人工审查和可追溯记录可能是价值组成；对制造商来说，它们增加硬件、数据保留和专业支持支出。评价要回答完整产品政策是否改善决策，而非抽象模型是否达到某个离线分数。

## 20.4 比较的共同骨架与不能被平均掉的差异

三个情境可放进相同账本，但不可合并成同一效用尺度。共同字段包括：可比非 AI 基线、产品功能版本、实际暴露机会、目标任务结果、错误后果、用户行动、售价和服务模式、售后资源、观察窗及退出。不同之处在于：家庭产品的核心反事实是相同舒适服务水平下的能耗与覆盖；预测维护的单位是独立设备故障和可行动窗口；高后果设备需要保护层、拒绝和人工确认及严重度分类。以“模型准确率+购买意愿+毛利率”的单一分数排列三类产品，会抹去功能类型和风险容忍的结构差异。

市场也影响价值转化。消费产品的用户与付款者通常接近，但平台和公用事业补贴可能拆分收益；工业设备中设备操作者、维护经理、采购和资产所有者的决策权不同；医疗和安全设备的使用者、付款者、专业人员及责任承担方可能完全不同。功能被消费者喜爱，不意味着采购者有预算；工程师认同预测效果，不意味着财务部门认可投入。产品化研究要追踪受益者、选择者、付款者、使用者与承担错误成本的人，并在账户或采购单位上定义结果。



SVSAC | sv-sac.org

可反证机制至少包括四类：同一任务的传统规则性能相当；用户目标与自动控制目标冲突；服务/渠道/安装条件造成的差异大于 AI 功能；或 AI 功能真实有效但制造商无法捕获客户价值。研究应预先明确哪个观测会推翻“嵌入 AI 提升产品毛利”的判断，例如目标结果无改善、激活和续费低于盈亏平衡所需水平、误报/退货/云成本抵消单台增量，或传统版本的组合贡献更高。反证不等于证明 AI 无用，而是让产品价值主张能因证据而收窄或被拒绝。

# 21. 产品级现场识别：从可用技术到可归因的客户与企业结果

## 21.1 先确定待比较产品政策，而非从现成日志挑分组

一项现场评估应先确定管理层究竟要决定什么：在同一硬件上开启 AI、将 AI 配置作为新 SKU 推出、对客户收费、通过订阅持续服务，还是用 AI 结果替代某项人工服务。处理方案不同，估计对象也不同。若研究算法模块，可以在同款硬件与相同销售条件下随机启用功能或使用非 AI 规则对照；若功能无法安全关闭，研究完整发布包可随机配置报价或市场上市策略，但结论只适用于该产品包。不能用已购买 AI SKU 者与老款客户直接比较，因为选择产品版本的客户可能本来就不同。

分配单位要匹配交互和资源共享。消费设备按账户/家庭随机，避免家庭中的多台设备被交叉使用；工业产品按维护班组、生产线或站点分配，避免共用专家把一组的方案带到另一组；涉及渠道促销则可能需按经销商、区域或周分配，并评估交易跨区。每一组应定义允许功能、版本、培训、服务和渠道激励。若控制组可从公开市场自行购买 AI 设备，需记录交叉而不从意向治疗分析中删除。

随机分配可在销量预测和效果阈值未知时仍被制定，但研究规模不能凭空指定。先用现有基线估计结果方差、设备/账户组内相关和可观察比例；对主要结果设定决策上有意义的最小差异，再做功效与可行性分析。若所需样本超过可触达市场或售后能力，应将研究定位为估计精度受限的先导，或换用多个市场/较长试验期，不得随便给“至少 1000 台”一类通用门槛。高后果结果事件太罕见时，可以先报告机制与边界，不能把无事件误写成已证明安全。

## 21.2 锁定分母、指标及版本，避免处理后挑选成功用户

主要估计应以首次随机分配的设备/账户或产品簇为分母，按意向治疗比较指定观察窗内的购买、退款、任务结果与成本。激活、正确配网、功能调用、采用建议和续费是处理后的过程，宜作为中间指标或不同估计对象报告。若 AI 组只



SVSAC | sv-sac.org

有成功激活设备参与体验评分，比较可能因掉线、安装失败或用户放弃而偏向AI。把这些未激活设备剔除会将部署摩擦从产品效果中删掉；可以单列“成功安装者结果”，但明确依赖选择条件，不称随机因果估计。

在随机前冻结功能版本和每个产品包的规格：硬件批次、固件、模型、阈值、默认状态、说明、保修、价格、地区与渠道。版本升级可能是必要修复，也会改变处理。如果试验中途更新，需记录更新时间、覆盖范围和是否依据结果适应；主要分析可保持原分配处理效应，同时另报版本敏感性，不可按表现挑选最佳版本。相同产品叫法不能保证实际功能一致。日志应能回溯序列号、用户、时间、服务路径、故障事件、退货、渠道补贴和真实净价，并限制只收集决策必需字段。

固定结果定义以避免“单位跑分”替代商业目标。家庭功能可以定义每设备日能源和舒适区间，同时记录用户覆盖和费用；工业功能可以定义每设备年故障、提前量、可行动维护和非计划停机；高后果功能则需主结果与安全护栏并列。企业商业结果还需含真实付款而非页面点击、净收入而非牌价，以及退货成熟窗口、服务支出和成本责任。价格试验的主要分析须以被随机报价者为分母；按实际购买者算毛利会条件化于处理后选择。所有主次指标、权重和缺失处理需在解封结果前写定。

## 21.3 结果盲评和来源独立性

若产品输出本身参与定义“成功”，模型可通过改变自己的评估记录制造表面改进。目标事件标签应独立于 AI 输出：例如现场设备故障由维修检查或预先约定的设备状态定义；能耗来自计量表；用户舒适可用同一采样问题与物理温度结合；退货是支付和逆向物流记录。对主观结果可采用盲态评审，隐藏处理组、模型建议和品牌标签，再按统一标准评估。双评审分歧如何仲裁、无法判定如何计入，也需提前设定。

判定标签如果从投诉工单或 AI 报警生成，就会只覆盖可见错误。应在 AI 报警、未报警和拒绝结果中按已知概率抽样复核，并记录纳入概率、失访和不可测状态；按抽样概率还原总体错误率需要观察样本代表相应目标总体且未复核组别有正概率进入样本。若自动日志无法证明实体状态，需保留人工抽检或维修后的真实参考。质量检查不是模型的一部分，而是测量系统；它也有成本并可能被处理组影响，需一并记录。

对商业结果应使用独立账务源校验支付、退款、折扣、返利、订阅取消和客服/保修支出。市场平台归因窗口与实际回款期可能不同，试验政策不能把平台标



SVSAC | sv-sac.org

签或模型预测的购买额当成净收入。财务和产品数据要按账户/设备和时间对齐，并预先确定归因规则；跨账户共享、渠道换货或家庭多设备需有去重方案。否则同一次购买可能在设备、用户、渠道三个表重复计数，账面增长不是新增成交。

### 21.4 估计、停止与外推

主要比较可按预先指定的组间均值差或比例差估计，并以实际随机分配层级聚类计算不确定性；分层随机时使用事前确定的层权重，而非按结果重新加权。对成本和交易金额长尾，可报告平均效应与分位/截尾敏感性，不因极端但真实的退货支出而删除观测。若多市场、多任务或多个模型版本同时检验，应将主终点、异质性和探索性分析分开；不能先扫描所有切片再只报告一个显著市场。

安全停止规则需要同时保护用户和不压制有价值的学习。设定严重事件立即暂停、达到某种连续故障信号时复核、成本/服务超上限时切回，以及数据完整性不足时不作效果结论等行动。具体阈值需根据产品风险、历史基线和责任要求制定，不存在跨行业统一阈值。试验可以通过阶段门从影子评估转到建议模式，再到有限自动化，但每次权限扩大都是新的风险决策；不能因为模型先前离线准确便自动进入闭环。

现场效果的外推要说明谁被招募、谁安装、谁付费、在哪些环境和版本下使用。智能恒温器研究的样本要求自有房屋、特定暖通系统、智能手机和高速互联网，且免费资格安装率为 73%；其结果不能直接搬到租赁住房、另一地区、当前自主感知设备或 AI 消费电子全市场。[7] 同样，现场随机试验可以识别被观察样本中的产品方案效果，却不自动预测竞争者降价、全国服务容量、供应链冲击和长周期复购。若现有数据只支持某一地区和一版功能，结论应锁在该边界，并安排后续复现而非外推为全行业毛利结论。

论文后续行动安排应区分三个层次：工程证明该设备可完成任务；产品试验估计同一用户和价格/服务条件下的结果差异；企业经营评价完整履约与生命周期的净贡献。只完成第一个层次时，可以说技术可行；完成第二层才可讨论客户效用、选择或续费；完成第三层并解决反事实与成本归因后，才有条件判断产品毛利是否改善。层次之间可以互相提供证据，但不能彼此替代。

## 22. 跨版本、跨市场验证：功能效果能否迁移到下一代产品

### 22.1 版本改进可能是新处理，不是重复测量

跨硬件批次、模型和配置的差异可能来自算法、校准、预处理、默认值、提示、容量或设备年龄。沿第 21 章固定处理和分母后，本章增加可迁移性检验：将



SVSAC | sv-sac.org

硬件、模型、运行时、阈值、传感器批次、功能开关、服务合同和说明绑定设备/账户，并记录自选、强推或回滚。名称与出货月不能替代这套处理版本；第25.9 节还需另辨共同更新批次与独立设备样本。

将新版用户与旧版用户直接比较，混入了设备年龄、用户选择和市场时点。愿意尽早更新的人可能更活跃、网络更好、技术理解更强；新版本也可能集中在新购买设备，而旧版更久使用。可行时，同一批可升级设备按账户或设备集群随机邀请/分阶段升级，并保留有资格但未升级的组；若安全修复需快速推送，不能为识别因果而延迟保护，应另找未受影响功能或分区上线的自然对照。对于版本差异的主估计，可以比较按原分配的升级政策，而非只比较实际完成安装者。

版本结果应按其对购买承诺的影响分层。模型更精确但云成本上升，可能提升功能却降低单台贡献；压缩模型使延迟下降而漏检增加，可能改变风险；界面更新让用户更容易覆盖，可能减低目标功能使用率但增加控制感。产品团队需同时观察任务结果、暴露、用户选择和成本，而不是更新后只看评分提升。若升级后用户目标和处理内容都变化，则它是新版本的新处理；不应把前代试验结论默认转移。

## 22.2 市场外推需定义目标总体和可比条件

版本之外，目标市场还受价格、能源/网络、住房、设备、语言、保修、服务队伍和隐私预期约束。应先选择总体——招募用户、可达市场、全部出货或未来组合——再检验覆盖及加权；高速网络、自有住房、特定型号的结果不能只换全球权重就推到无稳定网络的租户。这个迁移问题不同于第 21 章的内部随机比较，也不同于 25.7 节买家选择下的估计对象。

跨市场效果迁移可拆成两个问题：某机制是否相同，机制产生的总体效果是否相同。用户是否把舒适放在节能之前，可能在不同环境有不同排序；维护窗口和备件供给可能因工厂规模变化；渠道服务质量也可能调节产品可用性。即便模型准确度稳定，用户选择和支付路径仍可变。某地区订阅接受度较低、网络调用更贵、地区服务器需冗余时，功能价值相似也未必转成同等利润。应报告各市场内的技术、使用、付款与服务结果，并以预先确定的市场权重汇总，而不是用一条“平均毛利”推断同等可迁移。

智能恒温器现场研究的入组条件和安装率已说明“能用设备”与目标总体并非同义。[7] 更广的试验计划应分辨可接入网络、符合设备、完成安装、激活、持续使用和具备付款能力等阶段，不把每级人数相互当作同一分母。若自愿采用者与随机资格组差异很大，安装者效应不能直接代表向全体买家推广。通过再加权来



SVSAC | sv-sac.org

外推需要知道目标变量和选择概率，并假定重要效应修饰因素已观测；不存在重叠或未观测异质性时，应报告外推不能识别，而非硬给一个全市场点估计。

### 22.3 长期测试面临模型漂移、设备退出和存活偏差

产品性能可能因设备磨损、环境季节、传感器漂移和模型更新随时间变化。只在保修期初测一次会漏掉晚期维护成本与老旧设备错误；只对仍联网设备测量则排除退出或被替换设备。若设备最常因问题而脱网，存活设备的平均精度会高估真实装机质量。纵向数据需要保留设备首次部署、每次版本、维修、失联、用户停用、退役和保修事件日期，并定义状态变化后是否仍属于目标队列。

分析时间终点要匹配购买周期。一次家庭设备的退货窗口也许数周，订阅续期需要至少一个合同周期；工业设备故障可能跨季节才显现；耐用品替换和品牌忠诚更长。短期使用、故障率和长期品牌购买不能被一个固定的“留存”概念代替。若只观测到用户在某期仍活跃，未观测其后是否退订，不能把末期活跃当永久留存。对停服、用户迁移或拒绝后续数据授权，应保留删失类别并评估是否与处理相关。

产品更新会影响测量系统本身：某次日志字段更名或传感器采样频率改变，可能使故障次数、活跃设备数和能源读数在版本间不可比。应维护跨版本指标字典、数值校验与并行跑数，记录接口和口径变更。如果无法重建历史指标，可将序列切断并从新版本重新建立基线，不应以拼接制造表面稳定趋势。长期效果研究必须把模型可复现性与用户结果的时间可比性同时管理。

## 23. 盈亏平衡不是“每台利润更高”：销量替代、服务尾部与市场不确定性

### 23.1 分开存量替代与净新增需求

AI 版本的销售可来自老客户换新、普通版本客户升级、竞争品牌转入或原本不会购买的新用户。它们对企业的增量不同。若 AI 版覆盖了本来会买旧版的客户，增量贡献是新旧版本单位差额之差；如果开拓了新需求，则贡献是 AI 版的新增单位差额；若其吸引的是竞争者客户，产品组合价值还与流失品牌和渠道分成有关。只看 AI 版出货台数，会把内部蚕食和新增市场混为一谈。

设基线期间会卖出 Q0 台旧版，每台单位贡献为 m0；处理后 x 名原有客户转买 AI 版，每台贡献为 m1；另外 n 台为相对基线的净新增购买，增量开发固定投入为 F。忽略价格反馈、时间价值及其它后续成本时，处理政策的贡献变化可写为：ΔΠ = x(m1−m0)+n·m1−F。m1 须已扣可归属于每台的生命周期变动支持；另



SVSAC | sv-sac.org

有固定云服务或固定售后投入时，在式外扣 F_service，即再减一次该总额，而不能把固定总额直接从每台 m1 扣除。已含在 F 或已分摊进 m1 的同一成本不得重复扣。这个式子要求新旧台数按相同期间和目标市场比较，不能把 AI 版本所有成交都标记成新需求。

构造一组完全假设的参数：旧版单位贡献 52，AI 版扣除已估计的单位支持后贡献 49，20,000 台既有销量转为 AI；一次性投入 300,000。仅这批替换单位就少贡献 60,000。AI 版要靠净新增销量 n 弥补，门槛是 n≥(300,000+20,000×3)/49≈7,347 台。若每台持续服务成本再多 5 单位、AI 贡献降至 44，则门槛约为 8,182 台。数字不是市场预测，也不反映产品规模或需求，只展示当云尾部成本改变单位贡献时，同一研发支出需要不同的新增量。若新客户需求无法从旧版替换中识别，盈亏平衡数量就缺少实证分母。

## 23.2 需求、价格、退货及成本的不确定性要联合敏感性分析

盈亏平衡表常固定价格、使用率、退货和单位成本，再变销量。这会隐含所有变量可独立预测。现实中价格上升可能降低销量，低转化人群也许更多退货，重度用户既可能支付订阅也产生更高云费。测算应把合理变量关系也放进情景：不同价格下转化与退货、使用强度对应服务成本、版本效应对应支持工单、销量对应采购折扣与产能。给每个参数一个区间后独立抽取，会生成业务上不可能的组合；情景需保留已知相关机制，并说明未知相关性。

一个稳健结果可以报告“达到哪组条件才不亏”，而不是提供无条件销售预测。若固定开发支出 F 为 300,000，AI 单位贡献 m1 从 44 到 49，单看开发投入对应的新增销量门槛从 6,819 降至 6,123；若另需补偿被替换的客户旧版贡献，门槛需再上升。假设价格试验只能估出当前测试范围内的需求变化，不能直接把区间外的价格弹性填进预测。使用者激活率、服务年限和退货成熟度同样应以已观测数据或透明情景表示。

概率分析只有在输入分布有实证依据时才比情景表更有信息。把预测销量、价格弹性和故障率指定为正态分布后运行模拟，不能让主观假设变成客观风险。应区分随机波动、参数估计误差、结构不确定性和无法识别因素。可能的输出包括盈亏概率、极端亏损分布和对关键假设的排序，但要给出输入来源及依赖结构。尾部事件如广泛更新失败或供应中断并非每台独立，不能用独立单位的方差低估共同冲击。



SVSAC | sv-sac.org

### 23.3 价值实现不等于价值创造，更不等于财务确认

在产品讨论中，“客户节省时间/能源”经常被直接加到公司收益。实际上存在三道分离：产品是否改变客户任务结果；客户是否愿意把部分价值支付给厂商；公司是否在扣除生产、服务、渠道和资本投入后留住该价值。前两项可以成立而第三项失败。若竞品复制功能、平台控制服务费、经销商要求返利或客户用旧设备做替代，制造商利润可能没有增长。若功能节省了用户时间但只作为免费赠品，产品仍可能提高满意度却不提高单品毛利。

对价值捕获的检验需要同时观察使用和支付。随机化真实价格或功能配置能估计特定报价策略下的交易变化；售后和云账务可估计履约净成本；客户节约需要独立测量并避免与公司收入重复计入。若产品在高价位卖得更少，总收益未必增加；若云订阅有续费，收入金额需与未来服务责任并列；若减保修降低成本，需确认该功能真正减少了相关故障，而非改变客户报修行为。公司不应将替换旧款的 AI 收入全数当成品牌新增收入。

同样，成本节约不一定体现在毛利率。若自动化减少客服工时但客服仍保留相同人员编制，短期现金成本不会下降；工时释放可能用于更快处理其它问题，是能力价值，不是立即利润。若 AI 芯片将单台制造成本增加、而制造商随后涨价，毛利改善可能来自定价而非 AI 功能。评估需要把价格、数量、替代、功能使用和成本变动作为可区分组件。若不能识别组件，结论应是“发布包期间差额变化”，不能声称由 AI 功能引致。

## 24. 产品组合、渠道与能力约束：哪些功能值得持续投入

### 24.1 产品组合的选择单位是项目增量贡献与共用能力

企业通常同时评估多个 AI 产品、多个模型版本和多种非 AI 替代方案。若每个提案单独摊同一平台固定成本，可能过度惩罚共享架构；若所有产品都把平台当免费资源，则会低估总投入。项目组合应按“新增产品是否带来增量能力、占用什么瓶颈资源、能否复用已验证模块”计算，并保留平台总成本及共同风险。共享模型一旦有缺陷，会同时影响多个 SKU；复用降低边际开发费，也扩大共同失效的损失范围。

可将有限预算拆为固定平台建设、产品专属适配、独立验证、上市场景和后续维护。某个项目的估计价值取决于在预算、测试设备、产品经理、隐私/安全审查和渠道支持限制下是否可完成，而不仅是其单体毛利预测。若多个项目都依赖同一部署工程师或云容量，分别通过各自 ROI 门槛后仍可能无法同时交付。资源



SVSAC | sv-sac.org

分配决策可以先排除违反服务能力或风险底线的方案，再比较不确定区间下的增量现金和战略选择权。不能把评分排名当成精确优化结果，若输入需求与成功概率都是粗略判断。

## 24.2 共用平台会同时产生规模收益和相关风险

可复用的推理运行时、设备注册、更新验证和日志采集，可能减少后续 SKU 的重复工作；但平台故障或供应商政策改变会同时影响多个产品。若产品组合都采用同一模型服务，模型停服、单一云区域不可用或安全漏洞可成为共同冲击。组合风险不能按产品数量独立相加。需要识别关键共享组件、替代路径、版本兼容、离线能力和服务依赖，并将共同故障下的最大设备暴露作为风险情景。

平台收益也有学习效应，但必须区分可共享知识与专用例外。一个摄像头模型在同一镜头和光照下迁移到相似产品，可能保留部分标注与测试资产；若转向热成像或低照度，训练和故障边界变化，旧测试不能全复用。平台团队应记录迁移后重新验证的工作，而不是把历史投入都记作沉没成本。模块复用率可描述技术规模，却不能替代交付可靠性、质量问题或客户价值。最有意义的组合指标是边际项目为共用能力新增多少可验证用途，且是否提高共同失效暴露。

## 24.3 渠道把产品价值分配给不同参与者

品牌、经销商、集成商、云平台、公用事业、维护承包方和最终用户可能共同参与 AI 产品交付。不同参与者的收益与成本不同。经销商如果必须解释功能、安装传感器和处理账号问题，却无法分享订阅收入，可能降低推荐力度；云平台掌握关键推理能力，可能将硬件升级的价值通过服务价格收回；设备商的预测维护减少客户停机，但若维护合同按人工工时计价，服务方可能不愿快速减少现场访问。合同和激励决定功能是否被积极交付，不能把渠道当成外部噪声。

渠道试验应跟踪产品展示、报价、激活、安装、净价、折扣、返利、服务指派、退货及实际回款，并识别参与者间潜在冲突。若经销商对 AI 套装有额外佣金，销量上升可能来自更强销售激励，而非功能本身；若平台首页给新款更多曝光，平台销量也不能单独识别用户偏好。分层报告各渠道结果可以定位机制，但渠道子组若没有足够样本或是事后切片，结论只能探索。合同条款不公开时，应把无法观察的返利和履约成本列为不确定范围。

## 24.4 投资门槛应能根据反证调整

产品组合评审不应仅问“模型准确率是否提高”或“价格能否上调”，而要提出能够让方案停止、修改或扩大的证据门槛。技术阶段可要求设备任务结果超过既



SVSAC | sv-sac.org

有规则并满足拒答/安全条件；客户阶段需要在真实报价下观察足够的购买和使用，且没有不可接受的退货或支持负担；经营阶段则比较全队列净贡献和规模化服务能力。各阶段指标与预算要预先固定，同时允许风险委员会因严重事件暂停，而不是为了保持试点进度降低证据要求。

投资方案也要容纳“不明所以”的结果。如果小样本不能区分无效与估计不精确，正确动作可能是扩大或重新设计试验；如果实际需求低于成本门槛，可以缩小功能、改服务模式或停止。一个大市场宏观预测不能代替产品的可验证需求，技术团队的沉没投入也不该自动成为继续投资理由。通过产品组合重新排序是有价值的决策结果，不一定要最终推出每个 AI 版本。用明确反证保持决策开放，能避免将 AI 功能卖点当作既定业务目标。

# 25. 讨论：何时嵌入 AI 会增加产品利润，何时只是功能增加

## 25.1 技术改善、用户价值和利润之间并非单调关系

跨章证据可以合成一条有条件的机制链：新增 AI 能力先改变设备测量或执行，再影响任务结果和用户行为；任务结果要被购买者看到并理解，才可能改变购买、续费或退货；实际成交与履约成本决定公司能捕获多少价值。链上每一处都可能断开。MCUNet 证明了指定微控制器任务的技术资源方案，MLPerf Tiny 提供了受限基准条件；它们没有测量整机、用户、购买或成本。[1][2] AI 设计标签的支付意愿研究与嵌入式功能不一样；直播销售 AI 助手的现场随机试验研究的是交互服务；智能恒温器研究呈现用户行为如何削弱工程预测，但不是当前生成式 AI 或制造商利润试验。[5][6][7] 将这些来源放在一起有助于提出机制，却不能拼成一条已被证实的嵌入式 AI 商业因果链。

这一区分也解释为什么“模型分数提升”和“高毛利”之间没有必然正相关。更高的召回可能需要更大传感器、更长推理、更多上传和更多人工复核；更少误报可能改善使用体验，却令异常检测服务需要更多云计算；自动化可能减少操作时间，但用户选择关闭功能；标榜 AI 可能拉动早期兴趣，却改变不了净购买，或让高价挤掉边际买家。产品设计面对的是互相作用的系统，而非单个模型系数。相对于单任务提效，实质性增量分析应把效用、分母、选择行为、版本变化、退货、售后和价值获取联在同一个产品定义上。

因此本篇的回答是条件性的：嵌入式 AI 可能提高单台贡献或生命周期回报，前提是它相对于可销售的基线带来可验证且足够重要的用户任务改进；目标买家在真实配置和价格下购买或持续付费；新增硬件、云、验证、渠道、支持、



SVSAC | sv-sac.org

退货和资本成本不会吞噬利润；公司能而非仅客户/平台捕获这部分价值；并且该方案可在实际设备、市场与在用设备年规模履约。若只符合前两项，功能可能有客户价值而不产生更高公司毛利；若只符合技术可行，却没有用户选择与支付证据，利润主张尚未成立。

## 25.2 多目标压力测试：节能、舒适、成本和负担不能靠任意权重隐藏

家用温控说明产品升级可能追求几个相互冲突的结果。下表是一组完全假设的 1,000 户年度情景，用来展示决策选择，不是智能恒温器研究的效果估计。设基线平均电气/燃气费用合计 200 单位/户年；三种策略在不同舒适边界上得出不同节能幅度、每户舒适区间外时间和公司持续支持成本。舒适区外的小时不等价于货币价值；公司支持成本也不等于家庭的净福利。

| 方案（构造） | 相对基线能源降低 | 舒适区外时长 | 每户公司持续支持成本 | 1,000 户节能金额（按 200 单位基线） |
|---|---|---|---|---|
| 传统定时规则 | 3% | 240 小时/年 | 0 单位/年 | 6,000 单位/年 |
| AI 建议、用户确认 | 2% | 80 小时/年 | 15 单位/年 | 4,000 单位/年 |
| AI 自动控制 | 4% | 400 小时/年 | 30 单位/年 | 8,000 单位/年 |

这组情景没有一个方案在所有列支配另外两者：传统规则公司支持成本最低；AI 建议的舒适区外时间最少；自动控制节能最多、支持成本也最高且舒适指标最差。如果产品只用“节能金额减公司服务成本”排序，自动控制仍因 8,000−30,000 的 1,000 户账目更负；AI 建议为 4,000−15,000 也更负，传统规则为 6,000。可是这一粗减法并未给舒适定价，也没有加上设备成本、实际售价、能源机构补贴、安装成本或用户是否付费。产品团队无法从表格本身推出最佳方案；购买者、使用者和支付者需要决定效用优先级，经营分析再判定厂商能否捕获收益。

多目标报告应给出效果前沿及实际政策备选，让决策者明确愿意交换什么。若存在最低舒适保护边界，就把它作为约束，排除不合格控制策略；若研究提供用户对舒适和账单的偏好，应使用适当设计估值而不是作者任意指定权重；如果客户之间偏好不同，应展示分布而非只报告平均。尤其在安全功能中，严重误执行不应简单换算成用户的货币权重后被高收入预期抵消。允许用户选择自动化水平可能让不同需求兼容，但增加产品复杂度和支持成本，需放进方案而不是默认为零。



SVSAC | sv-sac.org

## 25.3 同一设备故障率可能掩盖共同更新事故

单位故障风险不够说明共享云服务和统一固件发布的组合风险。构造两个 10,000 台装机的情境，二者单台年度故障的边际概率均为 1%。情境 A 假设故障独立发生，期望约 100 台受影响，标准差约 10 台。情境 B 假设一次有 1%概率发生的共同更新故障、发生时全体设备失效，其期望影响同样为 100 台，但分布是 99%时候零台、1%时候 10,000 台，标准差约 995 台。平均发生台数不能显示极端容量压力。

进一步假设每台故障设备的直接处理成本 20 单位；超过 300 台需启动一次 200,000 单位应急支持。情境 A 通常约 100 台，不触发应急费用，期望直接支出约 2,000 单位；情境 B 共同事件发生时，直接处理支出为 10,000×20=200,000 单位，加一次性应急支持 200,000，事件总成本为 400,000；乘 1%事件概率，期望总成本为 4,000 单位，已包含直接处理期望 2,000 和应急支持期望 2,000，不能再加一次应急费用。本例把同一年度累计故障台数超过 300 作为一次性费用触发条件，并未模拟故障到达时点或实际并发队列。计算是作者构造的压力测试，不是任何企业事故概率。若共同事件会引发召回、责任或信誉损失，情境差别可能更大；若有硬件隔离、分批发布和自动回滚，相关损失会变小。应把风险共因、受影响设备暴露、恢复容量与系统恢复时间放入产品组合分析，而不是把单台独立概率乘销量。

这个构造的核心不是说设备一定会遭遇 1%共同事故，而是说明可靠性指标必须说明随机结构。单设备测试和独立同分布的故障率不能直接代表云集中式或全量更新产品。发布批次、地区服务器、设备类型和共同组件可能形成相关簇。压力测试可人为设定冲击范围和概率，回答业务需要多大回退能力；要称为统计风险估计，则须从实际故障日志或可靠度证据估计参数。若风险无法量化，应报告边界和不可识别因素，不把均值假设伪装成预测。

## 25.4 长期云承诺：出货、激活、付费和容量最低额的可复算例

共享平台的最低服务承诺可能使小型付费群不足以覆盖成本。考虑完全假设的一批 10,000 台设备，订阅月费 6 单位，购买时只有 10%用户初始订阅；每月续费留存率假设 95%，云端按活跃订阅账户每月收费 0.8 单位，供应商合同另有每月 2,500 单位最低容量费，支持 36 个月。若订户每月按 0.95 存活，则每个初始订阅者在 36 个月预计贡献的订阅月数为 $\Sigma(t=0\ldots35)0.95^t=(1-0.95^{36})/(1-0.95)\approx16.8444$。1,000 名初始订户约产生 16,844.4 订阅月，订阅收入约



SVSAC | sv-sac.org

101,066；按量云费约 13,476；36 个月最低容量费 90,000。扣除这两项云成本后约 −2,409，尚未计工程维护、支付费、售后、退款、税费或收入确认。

若初始订阅率增至 30%，初始账户 3,000，预计订阅月约 50,533.2、收入 303,199；按量云费约 40,427，加最低容量 90,000 后，扣除这两项云成本后贡献约 172,773，仍未扣其它履约支出。相同设备出货可以产生不同服务经济，因为能否激活和留存改变实际订阅月，而最低支出在假设期内照付。例子不代表任何 API 价格或市场转化率，只显示云固定承诺和付费池需要同时建模。若最低容量为平台共用合同，停止该 SKU 未必能避免其全部成本；可避免增量与内部会计分摊不同。

该压力测试还暴露续费选择偏差。长期活跃用户是处理后留下的人群，通常不能与购买当月所有用户直接比较平均满意或获利；未续费者可能仍承担保修或安全支持，早期取消的收入损失却先发生。分析应以最初销售/随机邀请队列为基数，逐月记录订阅、取消、激活、调用、支持和退款；以购买者维度核对实收净收入。若用截至某年仍活跃的设备作分母，不能把已取消或无法联系设备从成本责任中自动删除。

## 25.5 什么设计才能区分技术增益、产品流程和市场选择

有力的下一步研究不只是多找几家企业谈“AI 功能很有价值”，而要把三个解释分开：算法效果是否超过传统规则；标准化流程、传感器更新和培训是否解释了大部分变化；以及用户选择、价格或渠道给不同样本带来的差异。可在产品和资源允许范围内开展因子设计：同一硬件随机分 AI 与既有规则，另一因子随机产品说明或界面，第三因子评估服务模式/价格。安全不允许关闭功能时，用离线影子对照、非关键功能随机或不同市场发布政策识别可允许的部分。处理方案必须与正式销售中的版本、权限和服务承诺一致。

多市场试验还需要预注册目标市场权重、任务分层、购买资格和实际分配单位。若每个市场内样本过小，随机报价的总体均值可能有用，却无法估计细分市场效应；此时不应挑出最高增长的市场当作普遍适用。用户、账户、设备、渠道和售后团队可能相互影响，应将共享资源纳入分配设计。需求实验应覆盖真实价格与不购买选项，记录所有被邀请者和非购买者；产品价值实验应保留任务暴露、失败、拒答与用户退出数据；经营评价需跨产品账目连接订单、实收、渠道返利、在用服务月份与保修支出。

优先对照的主要结果包括：（1）功能是否达到目标任务及保护边界；（2）真实价格下每千次有效报价的净成交与退货；（3）非 AI 基线替代和净新增购



SVSAC | sv-sac.org

买；（4）每个最初分配账户/设备的全周期增量现金贡献；（5）用户控制和舒适/便利结果的分布；（6）共享平台尾部风险和服务恢复能力。对照和分析应分别按产品版本、市场、设备年龄、渠道和使用强度做有目的的异质性检验。每个结果可能有不同推断单位；若某个交易结果是账户级，数百万传感器行不能提高随机试验的独立样本量。

研究还应预先写明哪些结果将推翻收益主张。若 AI 单位贡献提高但销量下降使总体贡献不升；若客户节省显著但厂商未能通过价格或续费捕获；若版本升级只提高基准而真实使用/退款无改善；若重度用户云费和共同故障储备超过毛利增量；或若传统方法在同一产品边界以更低成本实现同等结果，应限制投资、改配置或停止。反过来，如果功能效用和支付效应在多个版本与市场中稳定，且服务边界清晰，仍不等于每一产品组合都应采用 AI；项目投资要考虑资本约束、替代机会和平台风险。

## 25.6 证据局限和解释的边界

当前可用证据分属不同层次。MCUNet 与 MLPerf Tiny 是指定芯片、任务和推理边界的工程/基准结果，不能推出整机毛利；NIST 设备网络安全材料是制造商或设备能力背景，不是利润证据，也不能被称为所有市场的强制认证。[1][2][3][4] Zhang 等人的 WTP 来源只核读出版社摘要，且研究 AI 设计标签，不是设备内嵌能力；Wang 等人摘要报告的是直播销售场景的 AI 信息助手，不是嵌入产品试验。[5][6] Brandon 等智能恒温器工作稿提供较直接的真实用户行为反证，但它研究特定时期的恒温器产品包、节能终点和加州合资格家庭，当前版本状态为条件接受工作论文，不足以预测当前 AI 硬件需求或制造商毛利。[7] 本节补核限于所列相关章节，不代表所有附录、原始数据或整条理论文献均已审计。

证据边界集中为：未取得企业销售账、报价、采购合同、设备队列、成本中心、云日志、退货记录或消费者试验；说明参数、情境表和公式只显示计算与识别路径，不能证明 X%溢价或 AI 版利润更高。会计、责任、安全、网络安全及数据权利仍须按产品、市场和时点判断，本篇不替代法律或审计结论。这一共同限制也适用于 25.7–25.9 新增的选择、产能和批次推断。

因此，25.7–25.9 的增量不在再次列条件，而在区分三个容易混同的对象：全报价机会的政策贡献与被吸引买家的潜在贡献；有正单位贡献的项目与占用稀缺能力后的组合收益；大量设备与少量独立发布批次的信息量。它们均未由本篇样本验证。不同研究的处理、结果、分母、时期与选择不可直接合并；未来证据需



SVSAC | sv-sac.org

分开 ITT/采用者、短期任务/长期购买、端侧技术/企业财务，并允许收益方向或幅度仍未知。

## 25.7 真实报价中的买家选择：利润 ITT 与“边际客户质量”不是同一个估计对象

构造一个能复算、但不冒充市场数据的报价试验：从符合条件的 2,000 个销售机会中，在账户层随机各分配 1,000 个机会，向一组提供带 AI 功能的真实价格方案，另一组提供传统功能方案。所有未购买账户仍留在原分母。假设 AI 组有 220 笔购买，其中 20 笔退货、200 笔保留；每笔保留交易在硬件、渠道和预计服务成本后贡献 70 单位，退货交易因逆向物流净损失 10 单位；每个分配机会的报价与支持成本为 1 单位。AI 组净贡献为 200×70−20×10−1,000=12,800。假设传统组有 200 笔购买、10 笔退货、190 笔保留，每笔保留交易贡献 60 单位，同样每次退货损失 10、每个机会成本 1，则贡献为 190×60−10×10−1,000=10,300。原始分派政策差为 2,500 单位，按所有合格报价机会计为每机会 2.5 单位。这个差值是本构造中报价政策意向处理估计量的算术示例；真实随机试验的有限样本均值差仍有抽样不确定性，不能等同已知总体效应，不是买家平均收益，更不是利润率估计。

即便两组购买率为 22%和 20%，聚合差也不能告诉我们“新增的 20 个买家是谁”或这些人的反事实支付意愿。若作出 AI 报价不会使任何本来会购买的人转而不买的单调性假设，2 个百分点的购买差可被解释为被 AI 报价诱发的净买家份额；但该假设不能由这两个购买率检验，且诱发者在未收到 AI 方案时的购买、使用和售后结果不可同时观察。若存在被 AI 方案挤出的买家，净增 20 笔可以与 60 笔新增、40 笔挤出相容。产品经理常把“AI 买家”与“传统买家”作条件比较，但两组是被处理影响后筛选出的不同人群，不能把其均值差解释成同一消费者的因果支付增量。

这个区别影响定位与毛利决策。技术方案可能吸引预算更高、设备更复杂或更早采用的人群，使 AI 组买家售后成本和使用强度上升；也可能吸引低风险客户而把长期服务需求留给未购买者。若目标是判断在一批符合资格机会中，是否应当继续发出 AI 报价，主要结果应是固定机会分母上的净现金贡献、购买、退货、服务调用及安全结果。若目标是某一类客户的产品设计，则需另有可识别的细分目标和足够重叠；不能先按购买后行为挑出高毛利用户，再将其描述为营销可精准获取的“边际买家”。

这个选择问题可以用一个更尖锐的反例展示。仍假设每组 1,000 个随机报价机会且每个未购买机会贡献为零，AI 功能报价组有 200 名购买者，购买者平均贡



SVSAC | sv-sac.org

献 50 单位，因而政策贡献为 10,000；传统配置组有 100 名购买者，买家平均贡献 80 单位，政策贡献为 8,000。AI 报价的固定分母政策贡献高 2,000，然而“已购买者平均贡献”低 30 单位。构造上可把两组解释为 100 名总会购买者和 100 名只因 AI 报价而新增的购买者：前者在传统产品上的贡献 80，AI 配置贡献 80，新增买家贡献 20。即使这种分层解释符合数据，实际试验也不会告诉研究者两臂中哪 100 个人属于“总会购买者”；身份是跨世界反事实，无法同时观察。

在单调性、排除限制和随机分配等条件成立时，工具变量式的 Wald 比率可以把每机会净贡献 ITT 除以购买率 ITT，作为被报价诱发购买者的局部平均净贡献：2,000/100=20 单位/诱发购买者。这个解释依赖 AI 报价只通过是否购买影响计量结果、没有被挤出者，并且购买与贡献都按同一资格机会定义；若报价改变未购买者的咨询成本、促销引流或售后义务，排除限制就不成立。它仍是特定报价、合格账户和实验时期的局部结果，不能当作所有 AI 客户的平均毛利，也不能用于声称某类可识别个人必然会被诱发。

若实际问题是 AI 功能是否支持溢价，报价设计可将功能与价格分开：在允许的安全配置中，随机提供 AI 或既有规则，再独立随机一个标准价或预定溢价，形成 2×2 政策臂；各臂均按最初合格机会报告购买、取消、退货、净贡献及使用结果。这样可估算功能效应、价格效应及两者交互，而不是把“有 AI 且更贵”的销售变化误归为功能价值。若市场实际只能提供一种组合，这种因子试验不是直接部署方案，但能区分机制；随机化应在客户簇或账户层执行，且同一渠道不可让销售人员把一组报价泄漏到另一组。研究要报告被邀请、分配、实际看到报价和购买的漏斗，并保留价格/功能版本，以识别处理前选择和执行不遵从。

即使随机报价设计看见 20%对 10%的购买率，单调性也不是唯一数学解释。令总会买者比例为 a，只在 AI 报价下买的比例为 c，只在传统报价下买的比例为 d，永不购买者为 n。观测只给出 a+c=0.20、a+d=0.10，因此 c−d=0.10；在非负概率约束下，a 可从 0 到 0.10，c 就可从 0.20 到 0.10，d 从 0 到 0.10。换言之，净购买增量是 10 个百分点，但 AI 诱发购买的比例可能在 10%到 20%之间，同时存在最多 10%的被 AI 报价挤出者。若把 20 个百分点的 AI 组买家都说成新增客户，或将 2,000 单位总贡献差直接除以“AI 组新增买家数”而忽略挤出，都会改变被声称的估计对象。

单调性是否可信，应由价格、功能替代、兼容性和渠道情境论证，并用购买率之外的行为诊断；实验数据不能直接标记个体的潜在购买类型。即便不追求主层效应，固定分母上的政策 ITT 仍可识别“在此报价分配制度下每个合格机会的平均贡献变化”，前提是随机化执行正确且结果完整。若决策者要求回答“哪些客户



SVSAC | sv-sac.org

被 AI 吸引”或“这些买家的贡献比原有客户高多少”，就需要额外结构假设、纵向资料或边界分析，不能把更细的模型预测分数当成这些潜在结果已被观测。

## 25.8 稀缺工程产能、替代项目与一次性开发成本

一项嵌入功能可以有正的单位贡献，却因为占用稀缺验证资源而降低公司组合收益。设一个构造项目在计划出货 10,000 台时，扣除每台可变制造、渠道和预期服务成本后，合计形成 110,000 单位增量贡献；这个数字尚未扣一次性非经常性工程（NRE）和项目占用的验证能力。方案需 80 个稀缺工程师日。若另一已准备好的升级项目可在这些同类工程师日上带来每日 800 单位净贡献，且团队确实因容量不足无法同时做两个项目，则被挤出的组合贡献为 80×800=64,000，AI 功能相对另一用途的容量调整贡献只剩 46,000，尚未扣 NRE。若替代项目并未准备好、产能本来闲置或两项目使用不同资质人员，则该机会成本会低于 64,000，甚至为零；内部工资费率本身不是稀缺产能的影子价格。

再设 AI 项目一次性 NRE 为 80,000 单位。以项目现金流视角，若没有容量挤出，项目贡献为 110,000−80,000=30,000；扣 64,000 容量机会成本后为−34,000。若财务报表需把 NRE 按 10,000 台分摊，则每台为 8 单位；经营分析可先将单台贡献从 11 减去 8，得到总额 30,000，随后只扣未含在毛利分摊中的容量机会成本 64,000，仍为−34,000。不能在已把 8 单位/台计入单台利润后，再重复扣整个 80,000，否则会双扣 NRE。相反，若毛利只呈现经常性单位成本、不含研发分摊，报告应明确这一口径，并在项目投资判断中单独扣一次 NRE。

有限产能应以可替代的边际任务评估，而不是把所有固定部门费用平均摊到产品。若两种工作可并行、团队存在学习曲线、试验失败能降低后续项目成本，影子价格也会随决策变化。一个可检验的组合研究应先记录验证队列、资质约束、已批准替代项目、实际等待和取消理由；再比较“AI 优先”“传统项目优先”或允许新增资源的政策。单产品的正净现值并不表示现在投入优于使用同一人员开发另一功能；反过来，共用平台的成本已经由其他 SKU 承担时，也不能把全平台费用再完整计入每个 SKU 的增量成本。

容量的机会成本也应做“确有替代需求”与“产能闲置”两种情境。沿用前述 80 工程师日，若团队已有可立即启动的替代功能并受同一资格人员约束，按其实际可交付边际贡献估出的 800 单位/日影子价格可构成 64,000 的反事实成本；若其他项目排期已满、没有合格任务，或替代项目要等到 AI 验证完成才具备关键输入，则同一 80 日并未挤出可实现贡献，短期边际机会成本可以是 0，而不是 64,000。只看到工时表不能判断哪种情况成立；需要记录当时可执行的备选任



SVSAC | sv-sac.org

务、队列顺序、因容量导致的延迟和被取消里程碑。内部管理费率、人员工资和替代项目账面收益分别是不同量，不能自动互换。由此构造的决策是“若不做此功能，团队会实际推进哪一个项目”，而非给稀缺二字贴固定货币价格。

在此例中，项目扣除 80,000 NRE 后尚有 30,000 安全余量，故能承受的最高替代产能影子价为 30,000/80=375 单位/工程师日。低于该值，按这一构造假设项目净值为正；高于该值则转负。该分界不是通用采购阈值，因为 110,000 贡献、10,000 台销量和 80 个工程日均为假设；它的作用是显示决策所需的信息：真实替代任务的边际贡献及该团队的可替代天数。若项目贡献和容量价格都不确定，应绘制二维敏感性区域，而非把财务部门选定的固定人工费率当作确切机会成本。

## 25.9 共同故障的有限样本：设备数量不是独立故障证据的数量

25.3 中的压力测试说明共同更新可放大损失；本节进一步说明，少量批次的“零事故”并不能确认共同故障风险接近零。构造一项发布验证：产品经过 30 个互相独立的发布批次，每批抽取 10 台设备，批次中任何设备出现一次严重的共同原因故障即记为该批发生事件。若 30 批均未观察到事件，在批次独立、固定风险率、检测完整且事件定义不变的简单伯努利模型下，事件概率的一侧 95%精确上界为 1−0.05^(1/30)，约 9.5%。即便总共检查了 300 台设备，也不能把这 300 台当成 300 个独立的“发布环境”；独立信息单位是有代表性的发布批次。

作为对照，只有在严重故障确为设备间独立、暴露条件相同且检测无漏失时，300 个独立设备中零故障给出上界 1−0.05^(1/300)，约 1.0%。如果这些设备来自仅 3 个固件构建版本、共用同一云依赖或同一供应批次，设备级公式会过度乐观。实际数据常同时包含个体器件故障、批次共同故障和发布/云端系统事故；三者需不同暴露单位和风险模型，不能用一个“每台故障率”代替。上述界限是假设算例，不是安全认证门槛，也不表明真实发布风险在该范围内。

批次数改变时，这个零事件界限变化很快：若分别有 3、30 或 100 个独立批次都未发生共同故障，一侧 95%上界约为 63.2%、9.5%和 3.0%。这里的“独立批次”假设意味着软件构建、云依赖或共同供应批次不能让多个名义批次实际上共享同一失效原因。为把上界压低，增加同一批的抽检台数帮助识别批内设备故障，但对批次事件概率的信息增加有限；增加覆盖独立软件发布、供应批次和云配置的批次数才更相关，同时成本高、版本可能漂移。设计选择因此是明确目标风险后，在批内抽样和跨批次覆盖之间分配验证资源，并维持可追溯的风险定义，而不是追求一个脱离事件结构的总设备数。



SVSAC | sv-sac.org

部署规模上升还会改变对相同概率的经济含义。若每批有 10,000 台，某个共同事故概率即使很低，其恢复需求也可能越过现场支持容量；若只有 100 台试点，短期内零故障可能主要说明样本不足。验证设计应分层覆盖设备版本、批次、地区、云配置与回滚路径，独立发布数量要与目标事故机制匹配。多个批次共用同一缺陷组件时，批次之间未必独立；此时简单二项界限也不成立，应基于更高层共同因子或压力测试报告边界。测试组需要明确记录是否观察了完整暴露期、设备离线是否算失访、修复升级后事故如何归属，并报告可识别的风险对象，而非把“测试过 300 台”写成普遍可靠性的证明。

故障概率界限本身也不是尾部经济损失估计。即使知道每次共同事件的发生率，仍需区分受影响设备数量、服务恢复时间、替代控制可用性和责任成本；事故分布可能在发生后有不同严重度。历史上未出现事故只能约束所观察暴露下的频率，不限制一个未见极端版本的最大损失。桌面演练或故障注入能测恢复路径和资源瓶颈，却不能单独估计自然事故的发生概率。决策账应分别列出概率证据、条件严重度假设、可承受恢复容量与未识别的系统性风险，避免用“零故障”同时宣称频率和损失都很低。

若产品承诺的支持期短于实际使用年限、保修期或云服务约定期，试验结论只能覆盖被观察的时段；生命周期现金尚未成熟时，应报告随访缺口，不能把短窗贡献外推为全生命周期毛利。

## 26. 决策建议：把 AI 嵌入产品不是毛利提升的充分条件

嵌入式 AI 能够成为产品价值来源，但不是因为设备拥有模型或在基准上跑得更快。它只有在明确任务上改善了用户可感知、可核查的结果，且买家愿意以真实价格选择该功能、产品能以目标配置长期履约，公司能在扣除硬件、验证、云服务、渠道和售后成本后捕获价值时，才可能提高单位贡献或项目回报。功能更好并不保证毛利率更高；单位贡献增加也不保证销量、总利润或生命周期现金为正。

证据应由工程、产品和企业层级逐步建立。基准用于筛选技术方案，独立产品验证用于测量实际任务、错误、使用和控制，随机报价或可比上市设计用于估计选择与支付，成本队列则评估部署、支持、退货和版本责任。各类证据不能相互替代。特别是，共同平台故障、用户适应、旧款替代和渠道分成会改变量产后的风险与利润，单台平均成本或首发期跑分不足以支持高毛利承诺。



SVSAC | sv-sac.org

因此，高毛利应被作为可以被证伪的经营命题：在预定义的产品基线、市场、价格、功能版本和支持期内，完整产品政策是否提高每个合格销售机会或每个最初分配设备的净贡献；其收益是否在合理的价格、采用、故障及服务成本敏感性范围内仍成立。若因果和账务证据不够，产品团队可以说 AI 技术可行或用户表达兴趣，但不应把这两件事宣传为利润已经实现。商业判断要允许缩小功能、转变服务模式、保持非 AI 方案，或在收益不足时停止投资。

## 给企业的下一步建议

制造企业应把 AI 功能当作一项产品投资来审视；客户愿意持续付费的体验，才是高毛利的基础。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：围绕一个高价值使用场景，测算开发、算力、维护和售后成本，验证功能使用率、付费意愿及服务负担。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，进一步形成智能化产品机会评估、功能与成本取舍、商业验证方案。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。

将 AI 装入产品，并不会自动把技术能力变成客户价值，更不保证企业取得高毛利。需要检验的是一条有多个可能断点的链：设备在目标使用环境中能否执行稳定且有意义的功能，功能是否改善用户结果，改善是否被识别并转化为支付、购买或留存，企业能否在价格、销量和渠道分成中捕获其价值，最后增量收入是否覆盖硬件、开发、部署、安全更新、支持、退货与停服责任。任何一环的证据都不能替代下一环。本篇以此为研究问题，对端侧资源约束基准、消费者对 AI 设计产品的反应、在线销售中的 AI 服务实验和 IoT 产品安全生命周期文件作定向范围梳理；通过因果链分析、边界比较和明确标注的构造账本，区分技术可行性、感知价值、实际需求和单位经济性。已有来源分别覆盖嵌入式推理能力、特定消费者判断或销售服务场景，并未共同证明“AI 嵌入制造品提高利润率”。本



SVSAC | sv-sac.org

篇分析提出可反驳条件：若功能改善不发生于高频或高价值任务、传统规则在成本和可靠性上足够、溢价被渠道或退货抵消，或全生命周期支持成本高于新增贡献，则智能化升级可能不创造增量经济价值。构造算例仅演示如何拆分单台贡献、比例和销量，不代表任何企业实绩。

SVSAC - sv-sac.org



SVSAC | sv-sac.org

# 参考文献


[1] Ji Lin; Wei-Ming Chen; Yujun Lin; John Cohn; Chuang Gan; Song Han. MCUNet: Tiny Deep Learning on IoT Devices. Advances in Neural Information Processing Systems 33 (NeurIPS 2020) (2020). https://proceedings.neurips.cc/paper_files/paper/2020/hash/86c51678350f656dcc7f490a43946ee5-Abstract.html

[2] Colby Banbury; Vijay Janapa Reddi; Peter Torelli; Jeremy Holleman; Nat Jeffries; Csaba Kiraly; Pietro Montino; David Kanter; Sebastian Ahmed; Danilo Pau; Urmish Thakker; Antonio Torrini; Pete Warden; Jay Cordaro; Giuseppe Di Guglielmo; Javier Duarte; Stephen Gibellini; Videet Parekh; Honson Tran; Nhan Tran; Niu Wenxu; Xu Xuesong. MLPerf Tiny Benchmark. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1 (2021). https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/hash/da4fb5c6e93e74d3df8527599fa62642-Abstract-round1.html

[3] Michael Fagan; Katerina Megas; Karen Scarfone; Matthew Smith. IoT Device Cybersecurity Capability Core Baseline. NISTIR 8259A (2020). https://csrc.nist.gov/pubs/ir/8259/a/final

[4] Michael Fagan; Katerina N. Megas; Barbara Cuthill; Jeffrey Marron; Brad Hoehn. Foundational Cybersecurity Activities for IoT Product Manufacturers. NISTIR 8259 Rev. 1 (2026). https://csrc.nist.gov/pubs/ir/8259/r1/final

[5] Hao Zhang; Xiaofei Bai; Zengguang Ma. Consumer reactions to AI design: Exploring consumer willingness to pay for AI-designed products. Psychology & Marketing, 39(11), 2171–2183 (2022). https://onlinelibrary.wiley.com/doi/abs/10.1002/mar.21721

[6] Lingli Wang; Ni Huang; Yumei He; De Liu; Xunhua Guo; Yan Sun; Guoqing Chen. Artificial Intelligence (AI) Assistant in Online Shopping: A Randomized Field Experiment on a Livestream Selling Platform. Information Systems Research, 36(4), 2358–2374 (2025). https://pubsonline.informs.org/doi/10.1287/isre.2023.0103

[7] Alec Brandon; Christopher M. Clapp; John A. List; Robert D. Metcalfe; Michael Price. The Human Perils of Scaling Smart Thermostats: Evidence from Field Experiments. Author working paper; authors' research page states conditionally accepted at American Economic Journal: Economic Policy (2026). https://pages.jh.edu/alec/therms.pdf




SVSAC | sv-sac.org