开发周期长、打样慢：AI 辅助设计、仿真与打样提速

作者：**SVSAC 硅谷战略顾问委员会**

核心判断

AI 打样提速的竞争价值，在于更快取得能被客户和工程端共同验收的样机，而不是更快生成图纸。对企业而言，值得优先推进的是：拆分需求澄清、设计、仿真、加工与验证的等待时间，先处理关键瓶颈，再引入适配的 AI 工具。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：工程设计；生成式设计；代理模型；物理验证；工程样机；研发周期

SVSAC - sv-sac.org

AI 打样提速的竞争价值，在于更快取得能被客户和工程端共同验收的样机，而不是更快生成图纸。

我们的建议是：拆分需求澄清、设计、仿真、加工与验证的等待时间，先处理关键瓶颈，再引入适配的 AI 工具。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：把“算得快”与“样机通过得快”分开

1.1 研究对象和终点

设计软件出现“几分钟生成数百种方案”时，直观上容易把这种速度理解为产品开发提速。然而，一个设计对象从数字候选到可验收样机，需依次通过需求解释、几何建模、分析模型建立、仿真收敛与可信度检查、工程审查、加工排产、样件制造、测量和试验。候选生成只是其中一段。若新方法让候选池扩大十倍，却使设计评审、网格建立和加工准备量同比上升，最终可验收样机的时间未必变短。反过来，即使推理阶段的时间节省很小，若模型帮助团队提前排除注定无法加工的候选，减少一轮跨部门返工，也可能降低总历时。

因此，本篇以“需求和验收条件已经定义后，获得首个通过预定工程验收门槛的样机所需的日历时间”为主要结果变量。该定义不等同于项目上市时间，也不把单个组件的数值仿真收敛视作产品验证。样机“通过”必须指向事先指定的任务：例如承载能力、疲劳寿命、尺寸与配合、公差、表面粗糙度、热变形或装配干涉中哪些项目需要验证，哪些阈值来自设计输入或既有工程规范。具体阈值不能由 AI 输出自身设定，否则模型可能通过改写验收口径制造表面成功。

开发历时可分解为一条带并行和返工的流程，而非简单串行求和。令阶段 (i) 的处理时间为 (p_i)，排队等待为 (q_i)，因未通过而回退的次数为 (r_i)，每轮返工时间为 (u_i)。单阶段有效占用大体为 (p_i+q_i+r_i u_i)，但阶段之间存在并行、资源共享和依赖关系，因此项目总历时取决于关键路径及其变化，不是各软件运行时间的总和。AI 若只压低一个非关键路径上的计算时间 (p_i)，端到端结果可能几乎不变；若它把更多工作推入共享加工队列，甚至可能延长其他项目的关键路径。

SVSAC | sv-sac.org

1.2 可检验核心问题

本篇围绕三个可检验问题组织分析。第一，在任务定义、物理模型和资源条件保持可比时，AI 方法是否减少达到同一验收标准所需的独立仿真调用、实体试件数或返工轮次？第二，计算侧的减少是否降低从需求基线锁定到样机验收的日历时间，而不是只把工程师从一项工作转去执行另一个排队任务？第三，这一转化受哪些边界条件调节：设计空间是否被数据覆盖，是否可将候选稳健地转换为 CAD/CAE 实体，目标附近是否有可验证精度，样件制造和检测能力是否足够，失败结果是否及时反馈给下一轮设计？

这些问题要求明确分析单位。一个“候选”可能是神经网络输出的边界形状、参数化 CAD 特征序列、一个有限元网格、一个制造试件或一个经测量的样机。只有同一对象在设计版本、材料、制造路线、测试条件和判定标准上能被连接，才能将计算阶段的改进与物理验收联系起来。否则，候选数增加与样机通过率上升之间的关系只是并列统计，并不能说明前者造成后者。

1.3 证据范围和推论边界

本篇采取定向范围梳理：优先选取能观察到设计表示、AI 与仿真衔接、代理模型或物理试验闭环的原始研究，并以技术综述帮助识别研究类型和共性限制。检索与核对围绕计算性能、几何可执行性、物理/制造验证及端到端周期四个不同层面展开。研究不追求对所有工程学科进行系统检索、文献筛选双人复核或效应量合并，也未收集企业项目日志。因而，本篇可以综合已发表研究中的机制与可迁移问题，但不能估计 AI 普遍能缩短多少研发天数、节省多少成本，或断言某类企业应达到某个速度目标。

我们特别区分证据层级。基准数据集或解析算例可检验算法在已知任务上的表现；数值仿真研究可检验特定物理模型下的搜索与预测；作者所述工业概念设计案例能说明方法进入设计流程的可能性；真实实体试验才检验材料、边界条件、加工偏差和测量链共同作用后的性能；多项目对照或经审计的项目历时数据才可能支持端到端周期影响。一个层级的结果不能自动替代后一个层级。后文涉及的机制解释均明确区分论文事实、本篇推断、建议和分析性假设。

SVSAC | sv-sac.org

2 端到端周期的工程机制：提速会落在哪里

2.1 开发不是单一计算任务

工程师提出的“设计很慢”，可能对应不同约束：需要反复重建几何、有限元计算昂贵、已有数据太少、CAD 和 CAE 之间接口脆弱、审批队列过长，或实体样件一次加工后才暴露装配冲突。把这些约束统称为“设计搜索慢”，容易对准错误的环节。生成模型可能提高方案空间覆盖度，却不一定减少参数化重建；代理模型可能降低单点求解成本，却增加训练、验证和版本管理；自动 CAD/CAE 脚本可能减少重复操作，却因边界条件设错而更快地产生误导性结果。

端到端改进需要沿着一条可观测的因果链成立：AI 或自动化工具改变候选提出、排序或分析的方式；这直接缩短关键路径处理时间，或使每次迭代选中对象的信息量更高、重复工作更少；由此降低被实体制造的失败候选数、跨工种返工轮数或项目关键路径时间；最终在相同验收定义下更早出现可接受样机。每一箭头都可能断裂。候选更丰富可能造成审查负担增加；仿真误差较小可能与加工误差方向相反；较少的样件可能是预算被削减而非学习效率提高；较短的阶段周期也可能伴随后续返修或资格认证时间上升。

这一因果链还受资源队列影响。若一个设计组能在夜间并行运行数百个仿真，瓶颈可能从 CPU/GPU 转移到工程师审查、共享许可证、几何修复或试验设备预约。并行计算减少了某一批次的执行时间，却不会自动增加下游资源容量。当 AI 方法提高每周候选通量时，应同时观测设计评审在制品数量、等待时间、样件排产等待、试验夹具占用和决策延迟。只报告平均单次仿真耗时，会遗漏拥塞效应：到达速率接近服务能力时，队列等待可能非线性增长，新增候选反而延长从启动到验收的历时。

2.2 设计表示本身是一个门槛

生成器输出的“形状”未必是可编辑、可约束、可制造的设计文件。体素和三角网格适合神经网络学习，但通常还需转为边界表示、参数化特征或可交换的中性格式，并满足几何封闭、最小壁厚、孔隙连通性、刀具可达性和装配接口等要求。若转换是人工完成，计算阶段省下的时间可能被返建 CAD 抵消；若转换是自动的，拓扑错误、曲面断裂、不可达内腔等失败会影响候选能否进入后续流程。故“生成速度”必须以候选能否无人工修补或在预定修补工时内转化为工程实体来界定。

DeepCAD 的研究将 CAD 模型表示为草图与拉伸等操作序列，并训练网络重构和生成这一类参数化命令；论文对模型生成、重构误差和数据集内外表现开展

SVSAC | sv-sac.org

实验。它说明了学习生成“CAD 程序”而非单纯图像或网格的路线，也显示表示选择影响何种编辑语义可以被保留。[2] 但该工作的评估对象是 CAD 操作序列及其形状质量，不等同于结构安全、材料性能、制造约束或实体样机验收；从有效命令序列到可制造工程设计之间仍有独立的设计审查和验证环节。若系统训练集以既有 CAD 模型为主，它也可能偏向历史建模习惯，而不覆盖新的连接方式或特种工艺。

这形成一个容易被忽视的权衡：生成输出越贴近机器学习模型偏好的表示，越容易优化和快速批量采样；越贴近企业已有 CAD/CAE 系统、设计规范和制造设备的接口，越容易落地但可能限制搜索空间。对真正的周期影响，应比较从“需求可计算化”到“可提交仿真/制造”的总转化成本，而不是只比较模型采样时间。对复杂自由形状而言，设计表示和几何修复可能是主瓶颈；对成熟参数化零件而言，候选表达可能不是瓶颈，仿真审查和实物验证反而更慢。

2.3 将生成任务变成工程目标

生成模型拟合设计样本，不自动满足强度或装配要求。Regenwetter 等的综述保留为工程生成模型的待核文献，[1]但其正文未取得，本篇不据题名转述它的覆盖范围或结论。我们提出的评价要求是：候选多样性之外，还需记录约束满足、性能估计所用求解器、独立实物验证及转入下一工序的人工时间。这是作者的工程分析，不是该未读综述已经证明的提速效果。

如模型只学习已有几何分布，它有理由生成“看起来像”历史零件的设计；但样本相似性与目标性能不是一回事。即便训练样本来自高性能产品，标签可能缺少失败设计或未测工况，导致生成分布在已经通过的区域过度集中。若任务目标是降低质量并同时维持刚度、疲劳强度、热性能和加工可达性，单一综合分数会遮蔽权重选择。AI 能让某个目标值更快改善，但它不能替代工程团队决定哪一项约束不可交换、哪一项性能存在可接受折衷。

2.4 资源节省和信息价值并非同一指标

仿真提速的潜在价值并不只在少用计算资源。更重要的可能是，在固定的工程时间与试验预算下，能否多回答一个关键不确定性问题。例如，设计候选的失效可能来自边界条件、材料参数还是几何拓扑？如果代理模型只预测一个汇总性能指标，难以区分这些解释；增加候选计算数也不会自动改进试验设计。相反，模型若能选取一个物理结果最能区分竞争假设的样件，虽然计算量下降不大，却能减少后续错误方向的开发投入。

SVSAC | sv-sac.org

因此，至少需要分别报告三类“效率”：计算效率（每一物理问题所需 CPU/GPU 或求解器时间）、工程流程效率（从输入到可审查模型的人工工时和返工量）、证据效率（单位实体样件或测试周期带来的决策信息）。如果研究只测第一项，不应将结论写成“加快产品开发”；如果它只给出生成候选数量，不应把数量视为证据效率；如果报告了已制造样件，也需要说明每个样件是否来自相同验收流程及失败样件是否进入分母。

3 设计生成与 CAD/CAE 联接：从候选到可复核模型

3.1 学习何种对象，决定能够提速什么

工程设计 AI 可被粗略分为三种计算任务。前向预测给定设计及工况，预测应力、温度、流量等响应；逆向设计根据目标响应寻找几何或参数；生成模型则从数据分布采样可能设计，必要时再按目标或约束筛选。三者常在一个工作流内组合，但效率、误差和责任边界不同。前向代理主要替代或近似部分求解；逆向方法需要处理多解性及约束满足；生成模型增加候选覆盖，不意味着已完成设计验证。

若生成器每秒输出千个几何体，但其中可解析、可网格化的比例低，或只有极少数满足约束，吞吐量本身没有工程意义。更合适的漏斗指标包括：原始生成数、几何解析成功数、进入 CAE 的比例、通过预筛选的比例、独立复核的比例、进入试制的比例、通过指定实体测试的比例，以及各层筛选消耗的人时和日历时间。漏斗的分母需固定：若团队只把通过算法筛选的对象送测，不能用送测样件的通过率代替所有生成候选的可行率。

3.2 CAD/CAE 自动化的增益与隐性复核

Yoo 等人提出一个深度学习驱动的 CAD/CAE 概念设计框架，包含二维生成、潜在变量降维、潜在空间实验设计、CAD 自动化、CAE 自动化、迁移学习和可视化分析七个阶段，并以概念车轮设计说明框架可以把生成设计与工程性能估计衔接起来。[3] 公开作者稿摘要将目标定位为帮助工程师和工业设计师共同审阅大量三维 CAD 候选及其估计性能，并筛选后续详细设计的概念方案。该研究的重要含义是“自动生成”需要跨越表示、参数化 CAD 和工程分析的接口；其案例范围则止于概念设计筛选，不能当成生产轮毂的制造验证或缩短实体样件验收时间的企业试验。

自动化接口减少重复点击、脚本化模型建立或大批量仿真提交，确实可能压低流程中的人工等待。但它也会把原先可见的人工检查隐藏在代码默认值里，例

SVSAC | sv-sac.org

如载荷方向、固定边界、接触关系、材料牌号、网格尺寸和求解收敛阈值。一旦默认值错误，运行更快意味着错误更快扩散。工程上应把“可执行”与“可置信”分开记录：脚本无报错、CAD 模型能生成、网格能够划分，只能证明操作成功；它们不证明模型表达了正确的物理问题，也不证明结果达到用途所需精度。

由此可设计一个可证伪的比较：将常规工程师流程与 AI/自动化流程应用于同类、预先定义的设计任务，保持验收要求、CAD 版本、求解器和制造路线不变。记录操作和等待时间、错误发现位置、回退次数，并由未参与模型搭建的工程师盲审边界条件与几何。若自动流程仅缩短求解提交，而该步骤不在关键路径、节省又完全被后续等待抵消，则没有端到端提速；若其直接缩短关键路径处理时间，即使审核与返工不变，也可能提前验收；若它降低了建立模型的中位工时，但更高比例的候选在独立复算中改变排序，就说明表面速度伴随可信度成本。

3.3 数字模型误差的来源不可混为一谈

“仿真不准”不是单一故障。至少应分开以下来源。第一，物理建模误差：方程、材料本构、边界条件、接触关系或载荷谱不恰当。第二，离散误差：有限元网格、时间步长或几何近似有限。第三，代理误差：机器学习近似与其仿真标签之间的偏差。第四，参数不确定性：材料批次、摩擦系数、温度、润滑或操作条件未知。第五，制造偏差：打印、铸造、加工或装配后几何及材料属性偏离名义设计。第六，测量误差：试验夹具、传感器、校准、取样和数据处理带来的变异。一个 AI 代理的预测与训练标签高度一致，只能说明代理复现标签；若标签来自简化有限元，模型不会自动修正其中的物理建模误差。

平均验证误差小，不保证目标最优点附近准确。模型可能在常见区间表现很好，却在强度或疲劳约束边界系统性乐观；优化器反复选择预测最优的候选，会放大这种选择偏差。预测值与真实值之差本身不是形式化的优化后悔值；对最小化问题，后悔值需比较所选方案与真实可行最优方案的目标值，后者在工程中往往未知。本篇据此讨论选择导致的性能落差，不声称已估计后悔值。应针对被选候选及约束边界作独立高保真分析或实物试验，同时检查可行性，不能只报告全空间随机留出集上的平均误差。

即使模型以“物理知情”作为卖点，也不能据此假设物理约束已被满足。Krishnapriyan 等人在 NeurIPS 2021 对 PINN 进行若干对流、反应和扩散方程问题测试，展示软 PDE 正则项在某些非平凡参数条件下会使训练更难，模型仍可能无法恢复相关物理现象。[6] 这是一项关于若干数学基准问题的反例，不能泛化为

SVSAC | sv-sac.org

PINN 或物理机器学习整体无效；它所支持的更窄结论是：把物理残差写入损失函数不等于精确施加物理约束，训练可优化性与结果正确性仍需独立检查。对应工程开发，物理先验可能减少数据需求，也可能因权重、边界条件或方程简化而把系统导向不可信的解。

关于误差的报告也需和任务相配。预测连续应力值时，均方误差不能单独说明失效风险；约束分类任务要报告假可行与假不可行，且通常假可行代价更高；逆向设计需检查目标可达性、多解与候选分散度；生成几何需报告拓扑修复、特征重建和不可制造率。若样本是按设计族随机划分，测试集中可能含有与训练样本近乎相同的结构；更严格的检验应按参数区间、拓扑族、材料牌号、工况或时间批次留出，以接近真实外推任务。

3.4 目标附近的可行性和约束管理

对样机开发而言，预测值接近目标不够。比如结构部件同时要求质量低于上限、在极限载荷下不屈服、疲劳寿命达到目标、关键孔位处于尺寸公差内，且能按指定设备加工。候选必须在多约束交集内可行；每个约束可能由不同计算工具、试验方法和责任人判断。AI 把一个指标提升到最好，仍可能让其他硬约束失效。尤其当训练数据未覆盖强约束边界，系统给出的置信度也可能失准。

Bastek 等人针对桁架超材料的逆向设计，构建深度学习框架，从目标各向异性刚度寻找结构单元，并用前向模型与物理约束处理“给定性能可能对应多种结构”的多解问题；论文包括仿真生成的结构单元训练、测试和对骨组织刚度的匹配演示。[4] 这一研究展示了在特定参数化设计空间内，将昂贵结构计算转为训练后快速搜索，并通过多个不同拓扑匹配同一性能目标。需要严格区分：它给出的是数值/计算设计证据，不是制造出的骨植入体经临床或力学试验通过，也不能把训练后毫秒级生成解释成样机开发周期缩短。若某一目标刚度在所选结构族中不可达，生成模型无法突破设计空间和材料假设本身。

工程流程可将自动候选排序和正式放行分离。第一层用廉价近似模型排除显然不合格或重复候选；第二层对低置信度、靠近约束边界或与已有数据分布相距较远的方案调用更高保真求解；第三层对关键候选进行独立复核和物理验证。这样的分层做法不是无条件更快，它的收益取决于代理模型是否能可靠识别“值得升级计算”的点、升级计算是否有可用资源、实体试验是否有明确判定标准。如果所有候选都需要同等昂贵复核，则代理只增加额外模型维护成本。

SVSAC | sv-sac.org

4 代理模型与仿真驱动搜索：省下的计算能否换来少一轮试制

4.1 代理模型替代的是一次计算，不是一次验证

当有限元、流体或多物理场仿真运行成本高时，代理模型从有限的高保真样本中学习输入变量到响应的近似映射。它可以在候选排序、设计空间探索或多轮优化中减少反复调用昂贵求解器的次数。代理训练与验证的代价仍需计入：样本生成、数据清理、超参数选择、局部补点、模型失效诊断和版本维护。只有在代理复用多轮、被替代的高保真调用占主要耗时、且候选任务与训练分布相近时，成本摊销才可能有利。

训练样本是怎样产生的，决定了代理能够回答什么。若用线弹性有限元生成数据，它只学习了给定材料参数、几何理想化和边界条件下的线弹性响应；若实体部件存在接触、塑性、疲劳、热循环或制造缺陷，代理不会自动捕捉这些机制。代理输入如果省略装配间隙或工艺路线，输出也不能承担相应判定。代理模型加速的是“依照某个模拟世界的规则回答问题”，而非让模拟世界本身更真实。

对拟合准确度的评估，至少要区分随机样本内插与实际设计外推。前者回答模型在相似训练样本附近能否重现标签；后者回答模型在新拓扑、新材料或工艺窗口中是否保持校准。随机划分常使同一参数化族的近邻结构分布在训练和测试集，导致结果较乐观。留出测试应当对应将要使用的边界：如果未来新方案来自不同拓扑族，就需要按拓扑族留出；如果工况随季节或载荷谱变化，就需要按时间或工况留出。验证集不能被反复用于选模型后仍被报告为独立测试。

4.2 自适应试验的价值在于减少无信息样件

当物理实验昂贵，贝叶斯优化、主动学习或其他自适应设计会根据已观察数据决定下一次试验点。它们可能优先探索预测不确定区域，或优先验证更可能满足可行性条件的参数。有效机制不是“AI 更聪明”，而是以每次物理试验的信息增益或可行概率为依据，重新安排有限实验预算。其效果须与合理的基线比较，例如领域专家设计、均匀覆盖、已有 DOE 方案或历史试验规则，并对候选总量、失败定义、试件制造次数、测试时间和中途策略更新保持透明。

本节所依据的是 Fadhel 等人 2026 年 BEAM 作者预印本的实验设置与结果，书目所列正式会议卷页未及独立确认。研究用 37 次既往失败试验初始化代理模型，在 GRCop-42 铜合金定向能量沉积中选择参数，再进行真实打印及质量分析。[5] 其报告的三个月实验部署、多个激光功率下可行配置，与合作团队此前数月未成功的探索形成历史比较，不是随机并行试验。观察终点是特定材料、设备

SVSAC | sv-sac.org

与参数域内找到可行打印配置，而非通用产品从需求冻结到首件验收的历时。初始知识、专家约束、设备和时期均可能影响差异，因此不能从该历史比较推出通用 AI 提速率，也不能把巨大离散搜索空间当成实际执行过的穷举对照。

自适应实验的另一类风险是标签定义过粗。若只记录某次打印“成功/失败”，多个潜在原因会被压成同一个二元结果：粉末流量不稳、熔池不连续、裂纹、孔隙、尺寸偏差或检测漏判。如果下一轮模型不知道失败是由哪个机制造成，就可能选中重复失败点或误以为参数成功。更细的结果记录有助于诊断，却会提高检测和标注成本。选择何种响应变量，是实验设计与工程判断的共同决定，不宜把所有指标压缩为模型方便使用的一个标签。

Gongora 等人的 BEAR 研究将贝叶斯优化与自动高通量实验结合，在一类参数化结构的韧性探索中，相对于网格搜索，报告识别高表现结构所需实验数约减少近 60 倍。[7] 这是一项关于特定实验搜索任务和“实验数”的比较，说明主动选点可能提高有限试验预算的信息效率；它没有证明样机开发日历历时缩短 60 倍，也不能外推到其他材料、设备或验收链。试验数降低是否转成更短历时，仍取决于实验排队、测量时间和试验后工程决策。

4.3 反例：局部模型提速可能延长整个项目

设想一个工程项目原本每周只能完成十个高保真仿真，因求解器占用而进度受限。部署 AI 代理后，团队每周生成并初筛五百个候选，但需由一位分析工程师确认候选建模假设、修复几何并审查边界条件。若工程师只能审核二十个，候选通量远超审查服务能力；队列开始堆积，设计决策仍需等待。若另一个共享设备的工装加工排期固定为三周，新增方案也无法并行转为实物。这个例子不是企业数据，而是说明局部处理速率提高后，系统瓶颈会迁移到容量有限的下游资源。候选生成量不能代替关键路径历时。

另一种反例是代理模型误差导致的返工。代理在训练集上平均误差低，但对最优候选附近的应力低估；团队因此选择薄壁结构，CAD/CAE 再次审查后发现失效风险，退回设计阶段重新加厚并重做疲劳模型。此时软件运行快了，关键路径却多一次建模审批。即使最终方案性能提高，首个合格样机的到达时间也可能更晚。判断提速是否成立，需以固定验收条件下所有失败迭代和重做历时为准，不能只计最终设计成功后的顺利阶段。

反过来，如果物理验证资源充足、样件可快速加工，传统逐轮专家搜索可能显得较慢；AI 排序在高保真仿真和实体试验前减少无望候选，就可能显著改善样

SVSAC | sv-sac.org

机到达时间。只有在具有足够预算验证该机制时，才能断言实际收益。否则更准确的结论是“存在减少无信息试验的机制，净周期效果尚未识别”。

5 可检验的评价框架：结果、过程与边界

5.1 主结果须落在实体验收上

如果研究目标是评估“打样提速”，主要结果应是达到相同验收要求的时间，而非候选生成时间或求解器耗时。建议采用三个互补结果：第一，需求基线冻结至第一个通过预定验收测试的样机所经历的日历天数；第二，同一时间段内通过验收的工程样机数；第三，完成验收所消耗的工程师工时、实体样件数和关键测试资源。任何组合指标都应公开组成项，避免把“更少时间、更多返工、降低样本覆盖”折成一个不可解释的分数。

过程结果包括候选进入下一阶段的分层比例、各阶段处理与排队时间、模型校准和独立复核结果、失败及其分类、几何修补人工量、试验计划更改以及未完成项目退出原因。对于每次更新，应保存设计版本、生成模型版本、代理模型训练数据截止点、求解器配置、制造路线和测试程序的时间戳。否则无法判断周期变化来自算法、人员熟练度、设备可用性、项目范围收窄，还是验收门槛改变。

5.2 反事实比较必须控制任务与验收口径

最具说服力的验证可以在多个相似设计任务中，将常规流程与 AI 辅助流程分配到不同团队或项目，预先确定相同的验收标准和预算，并按分配组比较达到首个可接受样机的时间。实际研发项目数量通常少，任务差异大，随机化也可能困难；可行替代方案包括相同产品族的配对项目、分阶段引入和预先设定的差异中差异分析。但无论使用何种设计，都需记录同时发生的工具升级、团队学习、人员配置和外部供应商变化，并为不同的项目难度与验证风险留出比较空间。

一个任务可被认为更难，可能因为几何新颖、材料数据薄弱、载荷谱复杂、工艺路线未定或验收测试周期更长。若 AI 组被分配更成熟、可复用程度高的任务，未调整的历时差异会高估工具影响；若 AI 只被用于高风险新设计，则会低估潜在收益。因此应明确任务纳入规则，按产品族、成熟度、验证类型和关键资源需求形成可比较层次，并在分析前锁定主要结果定义。退出项目不应从分析中静默删除：被放弃的候选可能是有效止损，也可能是工具未能适用，应分开记录原因及其决策规则。

SVSAC | sv-sac.org

5.3 把估计精度与工程决策联系起来

一个可行的现场研究可估计 AI 流程相对基线的“通过样机历时差”，同时报告不确定区间。设每一项目的起点为验收需求冻结日，终点为首个通过验收的样机签署日期；未通过或观察期内未完成的项目应按预先定义的删失规则保留，而不是当作零天或完全忽略。若多个项目使用同一设备或评审组，它们结果并不独立，分析需反映项目、团队、产品族或设备层级上的聚类。若样本量不足以稳定估计平均差异，应报告不能识别净周期效果，而不是用一两个成功案例推断普遍收益。

周期改善也要与质量/风险结果一并评估。若 AI 流程更早提交样机但关键测试项目覆盖减少、后续资格验证失败率增高或设计变更率上升，短期“到样机”速度可能牺牲后续流程。可以报告达到验收结果的同时，追踪后续资格验证、返修、材料/制造偏差和变更。研究期间应区分可逆的前期原型试验和高后果的正式设计放行，不能因为算法能快速枚举方案便建议降低验证深度。

计量链是比较结果可信度的必要条件。NIST 的金属增材制造模型验证项目提出以校准和表征过的粉末、熔池及零件尺度测量数据支持模型验证，并研究如何把仿真输出映射到测量系统输出、比较两侧不确定度及空间配准信息。[8] 该项目页面描述的是持续开展的计量与参考数据工作，并非 AI 缩短企业样机周期的因果试验，也不是特定产品的强制认证条款。它对本次分析的启示较窄但关键：仿真预测和测量结果不能直接对照一个裸数值，需说明测量对象、设备校准、空间对应和不确定度；否则所谓模型误差、制造偏差或验收差异可能无法区分。

5.4 本篇分析边界

本篇的重点是工程样机历时的因果链，而不是 AI 创造力、竞争优势、在线生产控制或企业数字化成熟度。已检索研究提供了若干机制片段：生成模型可产生工程表示；CAD/CAE 框架可自动连接部分设计环节；神经代理可近似特定仿真或实验映射；自适应实验能在特定材料工艺中以实际测试反馈选点。它们并没有共同构成“AI 已普遍缩短打样周期”的证据。评估范围也不是全领域系统综述，且我们没有取得项目级企业日志或开展独立复现实验。

因此，本篇可以提出操作性判断：只有在最终验收定义保持稳定、AI 环节位于关键路径或能减少关键路径上的失败迭代、模型误差由独立证据约束、设计输出可进入已有 CAD/CAE 与制造流程、并且下游人力和设备具备承接能力时，计算提速才具有转化为打样提速的合理路径。这个条件链是理论和文献支持的研究

SVSAC | sv-sac.org

命题，不是已实测的充分条件。下一阶段仍需对数字候选、仿真代理、人工审批、实体加工与正式验收之间的时间戳和回退链进行联合研究。

6 从二维约束到三维工程定义：可编辑不等于完整表达

6.1 参数化草图保留了什么

工程 CAD 的关键不只是几何坐标，而是元素之间的约束和生成顺序。两条边相交、保持平行、同心或距离固定，表达了设计者希望编辑时维持的关系；这些关系将尺寸变化传递给草图、特征、零件和装配。若模型只记住最终曲面或三角网格，几何外观可以重现，却未必能恢复设计意图。一次 AI 生成若无法保留约束，后续工程师改孔距、壁厚或接口尺寸时就可能需要重建特征树，局部节省便被修改成本抵消。

SketchGraphs 为这一层问题提供了原始数据与任务基准。Seff 等人从公开 Onshape 文档中抽取超过 1500 万个参数化二维草图，表示为由几何基元和设计者施加的约束构成的图，并测试草图生成及给定未约束几何时推断可能约束等任务。[9] 论文强调，通用三维网格数据通常不保留模型的构造路径，而几何约束图更接近草图构建过程。[9] 这使它成为训练“能生成可继续编辑草图”的研究基础，但并不意味着这些草图自动构成完整三维零件，也不表明学到的约束与具体装配功能、公差要求或制造规则一致。数据来自公开模型，论文分析的是 2D sketch；它的规模不能直接转换为工程设计时间节省或下游可制造率。

更深的区别在于“合法约束”和“适合设计目的”的差异。对某个图形，几何求解器可能找到满足平行、重合、相切约束的一组坐标；但模型仍可能过约束、欠约束或在局部解之间跳变。欠约束会让编辑产生意外自由度，过约束则可能使参数变更失败；即便数学上完全约束，也不保证孔位、基准和功能面选得正确。因此，自动补约束的准确率应同时衡量约束可解性、自由度和设计意图保持，而不能只比较与训练标签的逐条匹配。历史草图标签也未必是唯一合理解：作者可能只标注部分约束，或用外部基准隐式完成约束。模型推断不同约束组合不一定就是错误。

6.2 特征树和装配约束是另一个放大层

二维草图可通过拉伸、旋转、扫掠、倒角、圆角、孔加工等特征生成三维实体。特征树不是随意的操作记录；先建立基准面，再以基准定位孔阵列，与先拉伸实体后重新选择边界创建孔，具有不同的参数依赖。当特征顺序改变，设计变量的敏感性和更新失败模式也改变。模型即使生成相同终态几何，如果缺少特征

SVSAC | sv-sac.org

语义，仍可能降低后续修改能力；若特征可编辑却依赖隐含参考面，也可能在几何重构时发生引用丢失。

装配层的风险更不适合用单件几何相似度表达。孔位存在、尺寸合理不等同于多个零件在公差范围内可装配；一个紧固件周围的工具空间、插入方向、维修可达性、线束路径和热胀间隙都可能是组件级条件。AI 候选若只优化零件刚度和重量，忽略装配约束，可能将几何复杂性从单件移到工装或装配步骤。因而，适宜的验证单位应随任务逐级扩展：草图是否可解，特征树是否可重建，单件是否满足几何接口，装配体是否可定位和运动，最后才是实体样机在装配工序中的可重复性。

一个有用的诊断不是问“生成的 CAD 能否打开”，而是问三项操作：第一，工程师能否在既定接口上改变关键尺寸并保持设计关系；第二，系统能否对参数边界内的小扰动稳定重建几何与装配；第三，失效时是否指出受影响的约束、特征或引用，而不是静默地产生外形相近但语义已变的零件。若 AI 增加初始模型建立速度，却让改版、派生规格和变更影响分析变慢，项目总体周期仍可能恶化。这里的机制预测可以用成对工程任务比较：给定同一基准需求，记录 AI 与常规 CAD 建模在初建、变更、重建失败、装配干涉排除上的人时及返工次数，而非只计第一版模型生成时间。

7 容差链：设计输出如何在实体组装中被放大

7.1 名义尺寸不是可验收的尺寸

CAD 模型通常表示名义几何；加工、成形和装配得到的实体则含尺寸偏差、位置误差、表面粗糙度、形状误差和材料状态变异。某一个孔的直径在公差内，并不能单独证明轴孔配合或多孔定位关系可接受，因为配合由两个零件及基准体系共同决定。对于多层装配，单件尺寸偏差沿功能尺寸链累计，最终间隙或干涉可能由多个看似合格的局部尺寸共同产生。若优化设计改变了支撑结构、基准可达性或加工装夹方式，制造误差分布本身也可能变化；因此旧产品的过程能力数据未必适用于新拓扑。

这类链条可通过一个非实测的简化算例说明。假设两块支架由两个定位销连接，设计中心距相同；每块支架中心距误差分别为随机变量 (e_1,e_2)，销孔配合误差为 (e_3,e_4)。装配偏差可写作 (Delta=e_1-e_2+e_3-e_4)。若各误差独立且以零为均值，方差为四项方差之和；若由同一夹具或同一刀具引入相关偏移，协方差项也会进入总方差。单纯把四个标称尺寸逐个“都检查合格”不能推出组件误差

SVSAC | sv-sac.org

满足要求。该算例只演示误差传递，未使用企业尺寸数据，也不能替代实际基准、公差和相关结构的设计分析。

当 AI 候选采用传统设计中少见的有机曲面、格构或薄壁结构时，测量与定位也更复杂。高密度点云与名义 CAD 的对齐方法会影响偏差图；遮挡、表面反射、局部特征尺度和基准选择会影响测量不确定度。若只是把最大偏差当唯一误差量，某个小范围离群点可能掩盖整体变形，也可能把测量伪影当成结构缺陷。对验收有意义的比较应定义被测特征、基准坐标、滤波和拟合规则、测量不确定度及功能限值，并保留原始点云或量测记录以便复核。

7.2 容差分配不是优化器可自选的自由变量

尺寸链要求在功能目标和制造能力之间分配公差。AI 系统可以枚举尺寸方案或估计装配偏差，但决定哪些功能尺寸关键、成本和良率如何权衡、是否采用统计公差方法，涉及真实过程分布与失效后果。最坏情况叠加常用于确保边界条件下仍可满足要求，但可能提高加工成本；统计叠加可能更经济，却依赖过程稳定、相关性和分布假设。对尚无量产历史的新结构，过程分布未知，直接沿用旧零件的标准差可能使预测精度虚高。

因此，产品首件或工程样机时期的任务不是让 AI“计算一个公差”，而是识别模型假设缺少的证据。哪些特征要测量？测量工具能否分辨该误差？每个接口需几何基准还是装配功能基准？实体试制能否提供足够重复件来估计变异？如果只制造一件样件，观测到的一个偏差不能区分稳定系统偏差与随机波动。若样件数量由高成本制造限制，则应将不确定性显式带入验收，而不能把一个数值估计伪装为量产能力。

7.3 AI 何时可能降低、何时可能提高容差风险

一种可能的收益是，生成系统早期把关键接口和禁布区域编码为硬约束，让候选在进入详细设计前保留相同定位基准，降低变更导致的接口重构。另一种可能是，模型为提高局部性能改动结构骨架，改变尺寸链长度、基准面可加工性或装配顺序；即便名义仿真合格，实体偏差对功能尺寸的敏感性也升高。二者方向相反，不能仅凭“AI 支持制造约束”的标签判断净效果。

一个可检验的比较应预先定义功能尺寸与其验收规则，在同一制造路线、材料、设备和检验程序下，对 AI 辅助和常规设计的样机测量误差分布、尺寸链最坏值/统计值、返修工时、装配成功率及检验工时进行比较。对于尚无足够重复件的任务，可先将结果限定为设计可测性和单件装配成功，而不宣称过程能力或量

SVSAC | sv-sac.org

产适用。若样机图纸在制造后继续更改，需区分设计方主动修改与对制造偏差的补偿；否则“最终合格”会掩盖模型引发的额外调试。

8 CAE 链条一致性：CAD、网格和边界条件必须描述同一问题

8.1 几何传递产生的离散问题

从 CAD 到 CAE 通常要经历表面修复、特征抑制、几何简化、网格划分和材料/接触属性赋值。每一步都可能改变分析对象。窄槽、尖角、微小圆角在几何清理中被抹除，可能降低网格规模，却同时移除应力集中；薄壁若被简化成中面，需用正确厚度和偏置恢复结构刚度；复杂曲面由有限单元近似后，局部网格误差可能改变峰值应力。几何转换成功只是流水线能继续，不能证明仿真与 CAD 名义设计一致。

工程分析的困难往往集中于局部而非全域误差。总体位移场变化很小，仍可能在孔缘、焊趾、接触面或承载接口处改变疲劳判断；网格收敛也需结合所用结果量，某一全局能量或位移收敛不保证局部峰值应力收敛。优化器尤其可能利用几何离散的弱点：只要局部特征小于网格分辨率，结构看似“减重”而分析模型未正确捕获薄弱截面。此类伪优解通常不会被平均代理误差发现，必须对候选附近重新加密网格、采用独立求解设置或通过实体测量和试验检查。

8.2 载荷、约束、接触和材料数据的版本一致性

CAE 模型的另一个风险是工程上下文错误。相同零件在不同支承、连接刚度、冲击方向、温度和载荷谱下，响应可以不同。若训练数据使用固定支承和静载，生成候选却用于存在间隙、冲击或循环载荷的装配，代理无法用几何相似性弥补条件缺失。材料牌号还涉及热处理、构建方向、批次和后处理；若 AI 只读一个标准弹性模量，结论只能对应该理想化输入，而不能覆盖屈服、疲劳或各向异性性能。

更容易被忽略的是输入版本不匹配：几何更新了但网格缓存未更新，载荷方向已改而模板仍引用旧坐标系，接触对因面编号变化指向错误表面，材料卡片来自另一个版本。自动脚本会使同一错误批量复制。流程必须把几何哈希或版本 ID、网格策略、求解器版本、边界条件和材料来源连成分析记录，关键结果能够从签署版 CAD 重算。AI 输出的“推荐方案”不应被直接混入当前正式设计分支而不保留原始版本和差异。

SVSAC | sv-sac.org

8.3 模型验证的比较对象要写清楚

验证模型不能只有“预测对实测”。最少要说明：预测的物理量是什么，仪器实际测量什么，二者如何通过坐标变换、传感器模型或数据处理联系；是在名义几何、打印几何还是加工后几何上比较；试验载荷和边界能否复现分析输入；材料和温度是否相匹配；误差是否包括测量重复性与仪器校准不确定度。NIST 的增材制造计量项目将粉末、熔池及零件尺度实验、校准/表征、空间配准和从模拟输出到测量系统输出的映射作为持续研究内容。[8] 这个项目展示了工业模型验证所需的证据链复杂度，而不是说所有此类参考测量已完成，更不意味着一个数据集即可替代某类构件认证。

对 AI 代理来说，正确比较还需要至少三条线：代理预测与对应高保真仿真结果之间的误差；高保真仿真与经过计量设计的物理试验之间的差异；名义设计与实际制造几何之间的偏差。若只检查第一条，代理完全可能忠实复现一个未通过实物验证的仿真模型。若第一条较差，代理方案可被拒绝或重训；若第一条好而第二条差，应该怀疑物理建模或边界条件；若第二条好而第三条偏大，则制造过程可能是主要问题。把三类误差拆开，才能判断下一轮资源应该用于算力、试验还是工艺修订。

8.4 验证成本应按决策风险分级

并非每个候选都需要相同精度的求解或相同规模的实物测试。低风险概念件可先用粗粒度分析剔除明显不合格方案；承载关键接口、边界脆弱或代理置信度低的候选则应进入更高保真和独立复核。风险等级需要依据失效后果、模型外推距离、约束裕度和制造路线成熟度，而不能只根据模型自身信心。若模型的置信区间在域外未校准，低方差并不意味着低风险。

这种分级可以被评估为“单位验证资源改变设计决策的频率”，而非只看最终结果。例如，记录每次高保真求解或实体试验是否导致候选排序变化、约束判定改变、几何修改或放弃方案；如果耗费较高的测试从未改变决策，它可能过度；如果未经测试的候选在正式验收中频繁反转，则初始验证层级过低。必要的安全和资格试验不应因 AI 预筛表现好而删除；可优化的是把试验安排到信息价值高、风险边界清楚的候选，并确保终验完整。

SVSAC | sv-sac.org

9 制造可行性与工艺变更：优化对象和可做对象之间的距离

9.1 可制造性是设计空间条件，不是末端打分

制造约束会参与定义什么方案“存在”。增材制造需要考虑最小壁厚、悬垂、支撑、构建方向、粉末/树脂清除、热残余应力和后处理；减材加工要考虑刀具可达、装夹、刚度、刀路、内腔和基准可测性；铸造关注拔模、分型、补缩、孔隙和加工余量。约束常与目标相互作用：减少材料可能让某些结构失去打印稳定性，追求连续光滑曲面可能使刀具路径变复杂，采用新工艺减少部件数却引入额外热处理及检验步骤。

因此，所谓生成模型找到数学可行设计，指的是给定设计空间、仿真模型和编码约束之后的可行。若设计空间中没有表示支撑可达性或制造方向的变量，模型就无法优化这些因素；若约束只用简化惩罚项表示，候选可能在数值目标上得分高但越过真实工艺边界。将工艺约束留到后处理阶段可能出现“优化得越好，越难修复”的现象：修复改变厚度、拓扑或载荷路径后，原本的性能优势需要重新分析。

9.2 实体样件研究揭示了仿真之外的步骤

Okorie 等支架拓扑优化文献尚待原文核验。[10] 本篇撤回其具体算法比较、PLA 制造/压缩设置及约 20%减重的确定归属，不将它作为已获得的实物试验证据。可以独立提出的工程问题是：计算候选进入实体验证时，CAD/CAE 设置、网格、制造方位、试件一致性与加载条件如何记录并核对？这些条件决定比较对象，不因文献题名出现“实验”就自动得到验证。

在一个条件化的材料桥接问题中，即使某聚合物样件通过静态压缩，也不能推出金属零件符合疲劳、温度、振动、环境或资格要求；换材料和工艺需另论证适用性。这是本篇的假设推论，不是对未核支架论文测试结果的转述。任何实际结论均须注明通过了哪项具体测试，并另记录从需求冻结到同标准验收的历时，不能用减重或数值优化代替。

9.3 工艺切换会重写时间账本

AI 生成的复杂设计可能只在增材制造中可行；原团队若设备、材料、供应商或后处理能力不足，改工艺会引入新供应链和验证任务。反向也成立：算法为传统加工输出高复杂曲面，可能造成长刀路、专用夹具、装夹变形或检验点难以布置。即使一次样件很快印出，支撑去除、热处理、表面加工、尺寸检测和装配仍可能比打印本身更久。工艺路径变化不仅影响制造阶段时间，还可能改变设计约

SVSAC | sv-sac.org

束、材料性质、零件数量、失效模式和验收策略，因而不能只将样件交期视作一项固定外生参数。

有必要区分“验证样件”与“代表性制造样件”。低成本树脂或塑料件可能验证包络、孔位或装配关系，但不能检验金属材料强度、疲劳或热性能；软模、快速铸造或可加工替代材料也有类似边界。若以不同材料样件缩短早期反馈，应清楚列出哪些属性可外推、哪些无法外推，并把后续代表性样件仍需完成的试验计入总周期。否则项目会把早期粗略验证的成功误认为工程验证已经结束。

9.4 变更传播的成本

进入物理样件阶段后，设计变更会在模型、图纸、程序、工装、物料和检验规程之间传播。一次孔位更改看似局部，可能要求重算载荷传递、重建装配基准、更新加工夹具及复核测量程序。AI 若提出大量非局部几何变化，传统基于尺寸参数的变更影响分析可能不再足够；工程人员需要追踪哪些功能接口和试验结果依然有效。相反，若候选保持功能基准不变并在参数化特征层修改，能够复用已有的夹具和测试流程，开发节奏更可能维持。

这种机制可按每个设计变更记录“变更源–受影响文件–需重算项目–需重做工装/样件–验收重置范围”。分析单位不是单一一次改版，而是项目中每个版本的父子关系。一个容易混淆的结果是：AI 方案的单个样件更快，但因设计变更增加，最终验收时间更长；只记录最后版本加工用时会漏掉被弃用样件和重复验证。另一方面，若较早试制让设计问题提前暴露，前期样件数可能上升但总返修和后期验证风险下降。判断净效果必须依赖完整历时路径，而不能预设“少打样必然更快”或“多试制必然浪费”。

9.5 从物理案例得到的有限推论

样机速度的评价对象是适配后的实体在预定功能与验证条件下是否通过，以及到达该状态的时间。已读 SketchGraphs 摘要涉及二维草图的关系约束，桁架研究相关段涉及指定空间的逆向映射，NIST 项目相关段讨论测量与模型的对应。[4,8,9] Okorie 支架文献仍是待核线索，不纳入已获制造/静力试验证据。[10] 这些不同层级不能汇总为提速率。检验计算收益能否进入实体验收，需联结设计版本、分析版本、工艺路线与验收时间；这是拟议测量安排，不是本篇已完成的企业验证。

SVSAC | sv-sac.org

10 多保真计算与主动试验：信息价值取决于目标之间的可转移性

10.1 低成本观测并不天然是高成本结果的缩小版

样机开发通常同时使用解析估算、粗网格仿真、精细仿真、材料试片和实体部件测试。它们的成本、等待时间和物理代表性不同，因而可构成多保真观测层级。真正困难的不是把廉价观测与昂贵观测放进同一个数据库，而是判断二者在当前设计区域、响应变量和失效机制下是否相关。若低保真模型遗漏了高保真模型中占主导的接触、热耦合或缺陷机制，增加廉价数据可能让代理更加确信一个错误方向。

多保真方法通常利用不同成本级别的响应共同拟合目标函数，或者先用低成本模型筛选候选，再把昂贵资源放在少数点上。其有效性依赖一项常被简写的条件：较便宜的观测对决策所关心的昂贵结果具有足够预测信息。这个条件不能由“低保真”和“高保真”的标签本身保证。可以用同步采样的一组配对点估计两级模型在候选区域的偏差、相关程度及其随设计变化的结构；如果两种保真度在主要约束边界附近排序频繁反转，低成本层适合排除明显差解，却不适合确认赢家。

Folch 等的多保真异步优化文献保留为待核方法线索；[11]本篇未取得原文，不将匿名 BASF 数据、硬币/软包电池设置或所称机制遗漏作为已核事实。本篇的条件分析是：若实验成本、等待、容量与信息准确性不同，下一项实验须同时考虑预期信息与资源可用性；廉价观测不能仅凭价格被当成可替代的高保真证据。要验证这种选择是否缩短产品开发，仍需对应的真实历时、失败和验收终点，而不能把假设目标函数上的改善直接记作日历收益。

10.2 保真层级需由失效机制定义

同一项目里的“保真”可能对应不同维度。细化网格改善空间离散，却不一定改善材料卡片；增加材料试片数量降低材料参数的不确定性，却无法检验装配边界；把静载改为动态载荷提高工况代表性，也可能仍然没有考虑实际制造缺陷。将这些差异压成一条由低到高的直线，会误导资源配置。更合适的表达是为每个证据层级列出它覆盖的物理现象、尚未覆盖的假设、成本、周转时间以及可以支持的决策。

在一个承载件问题中，粗网格模型或许能排除明显超重设计；精细局部网格可帮助判断应力集中；材料试片可约束弹性或强度参数；实体部件测试才可能检验几何、材料、边界与制造缺陷的联合影响。层级不是简单替代：一个经过验证

SVSAC | sv-sac.org

的材料测试也不会自动验证部件载荷工况。若 AI 优化的目标靠近某一约束边界，资源应优先投向可能改变可行性判断的证据，而不只是提升平均预测精度。

10.3 主动选择要优化决策，而不只是预测误差

主动试验常以减少达到某个预测精度所需的观测量为目标，但工程研发的目标通常是找到可验收方案，或者尽早证明候选无法满足约束。平均均方误差下降不一定改进候选排序，尤其当高分候选集中在模型数据稀疏、代理不确定性高的区域。反过来，模型在整体空间上的拟合误差尚未很低，只要能可靠排除大多数不可行方案，也可能对有限试验预算有用。

因此，主动选点策略至少要区分三类信息需求：缩小可行边界的不确定性、比较接近候选的工程价值排序、估计某个高风险设计是否违反关键约束。对于前两类，模型可能需要在当前最优附近探索；对于第三类，应考虑最坏方向误差或概率约束，而不是只取后验均值。选点算法的奖励函数还要考虑实验结果何时到达及其是否能改变下一步决策。如果某一高成本测试要等数周返回，而低成本试验可以当日完成，单纯按单位成本最大化信息量会忽略等待期间可并行启动的工作。

10.4 提前设定何种观测可以改变决定

一项试验具有决策价值，意味着可能的结果会使方案选择、模型信任或资源投入发生变化。若无论试验结果如何都继续当前设计，则该试验的边际信息价值有限，除非它是安全或资格验收的强制证据。相反，某项低成本试验即使不直接验证最终性能，只要能区分两个关键竞争机制，也可能值得优先安排。价值评估应在结果出现之前明确候选行动、可能观测及各自后果，避免团队事后把任何结果都解释成支持既定路线。

在工程试验中，测量量也可能不是最终响应的直接代理。材料试片孔隙率、局部刚度或小部件热变形，只有在与最终部件尺度性能关系可验证时才可帮助筛选。数据采集速度快不等于决策信息高。对于未建立尺度映射的观测，适当的表述是“提供了某类中间机制信息”，而不是“确认了产品性能”。

10.5 异步试验中的“下一步”由等待状态决定

当不同设备同时运行，实验结果不是在同一时刻到达。若高保真测试尚未结束，低保真结果可能已提示某个区域不值得继续；若高保真结果即将返回，启动一项无法与之并行解释的新试验可能只是增加队列。异步算法的价值在于利用可用结果推进决策，但必须区分“设备空闲可启动”和“工程上已有足够信息更新方

SVSAC | sv-sac.org

案”。例如，低成本模拟在新工况下返回后，若它与已验证的高保真范围不再重叠，自动更新候选排序会把未经校准的关系当成事实。调度策略因而需要同时读入设备状态、候选间信息重叠、前一轮参数是否已改变及复核人员是否可用。

同时运行多类实验还可能使不同版本的设计在不同条件下被测量。若后续更新使用了较新的模型或工艺版本，而早期样件仍按旧版本测试，跨轮比较需要保留版本映射，不能用时间相近替代设计相同。对实体试验来说，试验顺序还可能带来设备热状态、夹具磨损或操作熟练度变化。若这些变化与 AI 方案使用时段重合，观察到的性能差异就可能来自实验条件，而非候选本身。将设备校准、运行顺序和版本信息纳入数据结构，是分析异步实验的前提。

11 排序风险：平均拟合正确仍可能选错样机

11.1 预测误差与排序误差回答不同问题

代理模型的均方误差针对响应值预测，而研发决策往往关心候选之间谁更好、谁满足硬约束。若十个候选响应彼此差异很大，较大的绝对误差也可能不改变排序；若两个候选几乎相同，极小误差也可能颠倒优选方案。更关键的是，模型错误集中在最高排名区域时，整体平均误差可能掩盖决策损失。因而模型评价应同时报告目标变量预测误差、约束越界的错误判定、候选成对排序一致性，以及按预定规则选出的首位候选在独立高保真或实体验证中的表现。

排序结果还受设计空间定义影响。如果候选集合只包含模型能够生成的形态，未覆盖的结构族不会进入比较；如果工程师在看到建议后手动修改候选，算法输出与最终测试对象不再是同一方案。分析应保留原始候选、修改差异、选择理由及每次验证的版本号。这样，最终表现差异才可以分解为模型排序、工程修订和制造变化，而不是笼统归于“AI 有效”或“AI 失效”。

11.2 最优点附近更容易出现选择性乐观

当研发人员从许多候选中挑出预测最高的一项时，选中者往往同时包含真实性能和模型噪声中向上的部分。候选越多、预测不确定性越高，这种选择性乐观越值得关注。团队若反复调整生成提示、设计变量和筛选阈值，直到获得一个看起来优秀的结果，却只报告最终版本，会进一步低估搜索过程产生的尝试数量。候选“胜者”的分数需由未参与筛选的模型、独立仿真设置或实体测试确认；确认前应称为预测领先方案，而非经验证的最佳设计。

保留全部迭代并非为了要求每个概念都完成实体试验，而是为了识别筛选规则是否系统性排除难以建模但有价值的方案。对被淘汰候选记录淘汰时使用的信

SVSAC | sv-sac.org

息、模型版本、约束原因及是否抽样复核，可以评估错误排除概率。若只测试通过 AI 筛选的设计，就无法发现模型是否把某类有潜力但陌生的结构排得靠后或直接排除，产生验证选择偏差。

11.3 候选确认应使用与排序独立的证据

如果同一代理负责生成、筛选和“验证”候选，流程中的重复使用不能产生独立确认。确认可以包括独立高保真求解、采用不同网格/边界处理的复算、独立工程师复核或按既定加载条件实施的物理测试；应根据候选风险和决策后果选择，而非统一要求所有方案制造实体样件。关键是确认证据所依赖的误差来源与筛选模型不完全相同，并且测试条件明确。

一个可检验的研究结果是“首位候选在独立确认后保持可行且保持相对优势的比例”，但该比例不能脱离候选池和确认程序解释。如果只确认少数安全成熟设计，外推到全空间不成立；如果确认结果因制造失败而缺失，不能把缺失视为性能通过。可同时报告候选排序变化和未完成验证的原因，使决策信息量与实体资源限制各自可见。

11.4 在约束边界比较候选时，报告“错选的代价”

两个候选的预测值很接近，但其中一个在某一载荷方向可能不满足硬约束；将它们的排序差一位与普通性能误差等量计分，会淡化不可接受的风险。可预先定义分层错误：一类是性能次序颠倒但两者仍可行；另一类是模型把不可行候选判为可行；还有一类是模型把可行候选误判为不可行而导致机会成本。三类错误的后果不同，测量方法也不同。前者可用独立测试的成对差值评估，第二类需按约束定义报告假通过率，第三类需对被排除候选进行抽样验证。

阈值应来自项目需求或适用的工程标准，而不是根据模型验证集事后挑选。若验收边界随设计负责人讨论而改变，应记录变更日期、批准理由和受影响候选，并对原始门槛与新门槛分别呈现，而不能用新标准重算旧候选后声称预测持续稳定。还应区分设计阶段的安全裕度和正式放行条件：代理模型给出的概率可用于决定何时追加试验，不构成对系统安全的许可。

SVSAC | sv-sac.org

12 原型资源、并行队列与关键路径：把计算速度换算成日历时间

12.1 研发工作流是带容量约束的网络

一个候选设计从提出到被接受，可能经过 CAD 建模、分析审查、样件制造、后处理、计量、性能测试和工程签署。各步骤存在依赖关系，也会竞争同一名工程师、设备或外部供应商。用一条“算法用时”代表全流程，会遗漏等待和返工；但把所有任务历时简单相加，也会忽视并行工作。日历时间应依据实际依赖网络估算：串行步骤相加，同期可并行步骤取完工时间较晚的一支，资源冲突则会把原本并行的任务排入队列。项目关键路径可能随 AI 引入而移动，例如原来受高保真计算限制，后来受人工确认、样件加工或测试台可用性限制。

评估时应分别记录处理时长与排队时长。工程师收到任务到开始审查的等待，不是模型推理时间；打印完成后的热处理等待，也不等于打印时长。数据准备、模型调用、几何修复、求解配置、程序审核、测试准备、失败分析和重制都属于研发所消耗的真实资源。若 AI 增加了候选生成数量，服务能力有限的检查节点可能成为瓶颈。排队论中的利用率机制提示，当某一资源接近满负荷时，小幅增加到达量就可能显著增加等待；具体程度依赖任务到达和服务时间的变异，不能套用一个通用常数。

12.2 一个明确标为假设的可复算例子

以下是用于说明时间账本的分析性算例，不是企业实测。设传统流程与 AI 辅助流程都从冻结需求开始，验收口径相同；两组各探索四个候选，先在同一位工程师处做候选审查，之后有两台可并行运行的样件设备，最后只有一个性能试验台。假设需求冻结与首轮 CAD 整理需 2 个工作日；传统候选需要工程师逐个建模，每个用 3 日，AI 候选由系统生成后仍需每个 1.5 日人工审查；加工每件 2 日，性能试验每件 1 日。为明确算术口径，假定项目采用严格的批次阶段门：四个候选全部完成建模/审查后，才统一启动加工；四件样件全部完成制造后，才进入测试批次。再假设所有候选一次通过前置审查、设备无额外排队、无返工且测试按提交顺序执行。这些条件明显偏理想，仅用于展示这种批处理策略下的历时上界，不是允许候选完工即流转时的资源网络最短工期。

按上述严格批次阶段门，传统组最后一个候选在第 12 日完成建模，随后两台设备并行制造，四件样件分两批历时 4 日；制造批次全部结束后，单一测试台逐件测试需 4 日。自 CAD 整理结束后计算，批次结束时间为 12+4+4=20 日；加上前置 2 日，总计 22 个工作日。AI 组的四次审查需 6 日，之后制造 4 日、测试 4

SVSAC | sv-sac.org

日，批次完成时间为 2+6+4+4=16 个工作日。两个数都是“全部候选均需验证”且严格分批的假设上界，并非首个合格样机的平均历时或资源网络最短工期，也没有包含训练、接入或数据整理成本。若允许单个候选审查完成即进入加工、制造完成即进入测试，阶段间可流水重叠，工期会不同；在没有给定设备排程和具体依赖图时不能从此算例推导最短工期。算例只说明当节省发生在真实关键路径且未被下游队列抵消时，周期才有缩短机制。

现在改变一个条件：AI 组因为可以生成更多候选，审查量从 4 个增加到 12 个，而仍只有一位工程师、每件 1.5 日审查。审查需要 18 日，已长于传统组 12 日的建模时间；后续制造和测试资源仍有限，故扩大生成量不仅没有保证更快，还可能延长完成时间。再假设模型在审查后淘汰一半候选，能够减少设备负荷；但若误淘汰率未知，且最终通过样机可能落在被淘汰集合，候选减少也不能直接被记作收益。要在算例中体现成本，需要把开发的未通过候选、人员审查、制造失败及后续复核全部纳入。现实项目应以事件时间戳替代这些假设数值，并按资源节点验证节省是否发生在关键路径上。

12.3 试验预算是资金与机会成本的组合

实体样件不只是一次材料费用。每轮试验会消耗设计工程师、制造工程师、机台准备、材料、后处理、检验、试验台时间及缺陷诊断能力。共享设备上的机会成本也可能高于单件直接现金支出：一个不成熟候选占用了有限测试窗口，可能推迟其他项目的验证。反过来，早期失败试验如果及时排除不适用路线，也可能释放后续资源。因此“样件数减少”只有在质量、覆盖和放弃规则一致的条件下才可解释为效率提升。

并行化也不是无成本。多台设备同时加工会提高设备准备、操作人员、校准和物料管理需求；多个候选并行测试可能导致结果解释互相干扰，尤其在同一试验台需要改变夹具或重新校准时。若试验结果按批次到达，主动算法可能在批次完成后才更新下一轮，产生同步等待；异步调度则可以使用部分结果，但须确保中途更新不会破坏预设比较或安全检查。研究记录应区分并行开始、实际运行、等待资源和结果可用时间。

12.4 失败、返工与未完成候选的定义

失败不等于一个统一的 0 分。候选可能违反几何约束、求解不收敛、实体无法制造、样件损坏、测试设备故障、指标未过线或因预算截止未得到测试。不同原因对模型训练和项目速度的含义不同。把设备故障当作候选性能失败会污染标

SVSAC | sv-sac.org

签；把几何重建失败当作“没有样机”也可能漏掉 AI 表示层的可用性问题。记录失败时应保留阶段、原因、责任证据、是否可重试、实际耗用资源及最终处置。

返工次数同样需要计量口径。一个设计版本经过多个局部修改，不能既把每次修改计为独立候选，又把它们都计为独立样机。建议以冻结需求的项目或可追溯的设计分支为观察对象，再建立父版本、子版本和样件批次的关系。设计修改后，哪些既有仿真和测试仍然适用需要由技术负责人明确；重复利用结果能减少时间，但只有当变化未破坏其适用假设时才成立。

12.5 资源比较应同时报告工程工时、现金与机会成本

将时间结果做成货币收益估算，至少要分开直接现金支出、内部工时和共享设备占用。样件材料费、外部检测发票属于现金支出；工程师审查与失败诊断属于内部人时；测试台被一个项目占用而推迟其他项目，属于机会成本。三者可能重叠，例如外包检测费用已含技术人员服务时，不能再把同一段外部人时加到总成本。若组织没有可信的小时成本或设备分摊口径，宜并列报告数量和时数，而不应虚构一个精确 ROI。

一个透明的预算表可按设计版本列出 CAD/CAE 人时、计算费用、物料与加工费用、后处理、计量测试、返工、审查和等待时长，并在“可复用”栏记录已有夹具、模型或测试结果是否继续有效。对于尚未结束的候选，已经支出的资源与剩余估计应分开；对于被取消的路线，沉没成本不能靠最终成功版本平均摊薄。若 AI 降低单次仿真费用但增加版本数，其总成本方向取决于每个版本的资源账本和是否进入实物阶段。任何比较都应披露未计入项，并对成本范围而非单一点值做敏感性分析。

13 从“完成样机”到时间结果：删失、竞争事件和估计边界

13.1 观察期内未通过不等于历时为零

项目研究常在固定截止日结算。部分项目已验收，部分仍在加工或测试，也有项目因设计路线无效、预算取消或优先级变化而终止。若只对按期通过的项目计算平均天数，就会排除较慢或更复杂的任务，系统性地呈现更快结果。未验收项目至少包含两种不同状态：在观察截止时仍可能通过的右删失项目，以及已被明确停止且不再追求当前验收目标的竞争事件。把二者都当作成功或都从分析中删去，会混淆速度与项目选择。

一个可复现的时间结果定义应明确：起点是需求、验收测试和预算冻结日期；事件是第一个满足全部预定验收条件且由授权人签署的工程样机；删失日期

SVSAC | sv-sac.org

是研究观察结束或数据停止可用日期；竞争事件则包括正式取消、目标变更后不再具有可比性的项目、或资源转移导致原项目无法继续。若项目只是暂停等待供应商，需记录暂停区间并同时报告自然日和活动工作日；不能因暂停由外部造成就静默扣除，除非分析计划事前定义并能保持两组一致。

13.2 估计方法要服从分配与事件结构

对于样本较少、项目风险差异明显的研发研究，至少应报告按组的验收事件数、删失数、竞争事件数和观察期限分布。当验收是唯一事件、其他观测中止可在给定条件下视为独立删失时，Kaplan–Meier 曲线可估计尚未验收生存函数。若项目取消会阻止原验收事件发生，则标准 KM 把取消当删失所得曲线不等于现实世界的未验收比例或验收概率；取消与项目失败风险或 AI 流程相关时，删失还可能是信息性的。存在真实竞争终点时，应以累计发生函数描述验收等各终点的现实概率，或使用适当的多状态描述；若研究设计为团队分配，则统计推断也要反映团队层级，而不能把一个项目中的每个候选都当成独立样本。具体模型应由分配、事件数量和随访条件决定，不应为了输出一个精确比值而忽略识别限制。

平均历时和“在某期限内验收的概率”回答不同问题。完整案例中位数是仅对已观察到验收事件的项目计算中位数；它忽略尚未验收项目，通常会受完成案例选择影响。Kaplan–Meier 曲线所定义的中位事件时间则是估计的“尚未验收”生存函数首次降至 0.5 或以下的时点；若曲线在共同随访范围内未到达 0.5，该中位数不可估，不能外推一个数值。普通 KM 估计把竞争取消当作独立删失时，所对应的是在强假设下消除取消事件后的假想验收时间分布，不是现实项目在存在取消时的验收累计概率。现实世界中“先验收”与“先取消”互为竞争终点，应按实际验收概率估计累计发生函数；Kim 文献仅保留为竞争风险方法的待核线索，其相关原段未取得，不作为已核的具体方法归属。[12] 限制平均生存时间（RMST）是预先指定共同时间窗内生存曲线下面积，即该窗口内未发生定义事件的平均时间；它是依赖事件和删失定义的估计对象，不是无条件的平均开发工期。[13] 若取消为真实竞争终点，应明确要估计的是验收概率、取消概率还是无事件时间，并采用相应竞争风险或多状态方法，不能把标准 KM/RMST 悄然解释成实际验收机会。两组项目构成、截止期或任务难度不同，单一时间比很难解释为 AI 因果效果；应报告事件定义、随访期、删失机制、项目层级结构和不确定区间，并按研究设计允许的方式分层或调整。

SVSAC | sv-sac.org

13.3 加速与止损必须分开识别

主动筛选可能让团队更早发现某一设计族不可行，从而取消路线并转向替代方案。对当前候选而言，这不是成功样机，但对更大项目组合可能是有价值的止损。若研究只把最终目标达成视为唯一结果，会把有效排除风险的机制误判为失败；若把“迅速停止”与“迅速通过”合并为开发速度，则又会夸大成功交付。应将首次验收时间、停止决策时间、停止后的替代路线以及整个项目目标是否最终达成分别报告。

对照两种流程时，替代路线的可用性必须遵循相同标准。若 AI 组在某条路线失败后有权调用资深工程师手动补救，而基线组不能获得相当支援，比较反映的是流程组合而非单一工具。反之，若研究关注的是“AI 支持的完整工作流”，这些人工升级和外部服务也应纳入干预定义与资源成本，而不能把它们藏在系统边界之外。

13.4 后续识别需要联结完整历时数据

候选生成日志本身不足以重建样机开发时间。至少需要把冻结需求、建模版本、模型/代理版本、人工审查、分析提交与返回、制造开始与完工、后处理、测量、测试、变更和签署事件相连，并记录等待原因及人员/设备占用。基线与 AI 流程使用相同的需求范围、验收规则和观察期限，才可以判断周期差异是否来自工具、资源、任务选择或验收口径变化。由于实际项目无法总是随机分配，未来研究可采用团队层级随机分派的试点、匹配产品族的并行比较或预先设定的分阶段引入；每种设计都需要清楚写出假设，并报告何种结果会推翻“AI 缩短周期”的判断。

本章建立的只是从候选筛选、资源队列到时间结果的可测量框架。对于公开文献中已发表的计算、材料或机械设计案例，仍需逐一说明任务和测试边界；对于企业场景的因果判断，还要结合后续的现场研究协议、产业情境和治理条件。没有项目级历时记录时，合理结果是提出可复核的测量方案与机制预测，而不是给出一个看似精确的通用提速幅度。

13.5 将外部可比性作为结果解释的一部分

同一“需求冻结至首个通过样机”指标，在不同产品、制造方式和验收制度下并不代表同等难度。一个塑料概念件可能几天内通过装配检查；一个承载金属部件可能需要材料批次确认、热处理、疲劳测试和数周供应周期。外部比较至少应按材料/工艺、样机用途、关键性能要求、样件层级、资源可得性及验收方责任分层。若这些信息缺失，只能报告单一案例的观测历时，不能把它转成行业基准。

SVSAC | sv-sac.org

推广研究结论时还需辨别 AI 系统与组织能力的互补性。若一个团队已积累可调用的参数化 CAD 模板、校准过的代理模型、自动化制造接口和快速复核机制，另一个团队从零开始接入相同模型，二者的处理时间可能不同。此差异不意味着工具没有价值，却意味着效果估计针对的是“工具加其部署条件”的组合。将工作流成熟度或数据可用性作为预先定义的情境变量，可以检验哪些互补条件是作用机制的一部分，而不是事后用来解释不一致结果的借口。

本次分析框架尚未提供跨企业的历史数据或可推断的普遍效应。下一步可以先在单一产品族中建立事件数据和候选版本链，之后用另一种材料或验证任务重复协议，考察结果方向、误差类型和资源瓶颈是否迁移。若仅在一个产品族有正向结果，应把结论限制在该流程；若同一方法在不同情境中产生相反方向，则应检验候选质量、资源队列和验收制度是否解释异质性，而不是以平均值掩盖差异。

13.6 现场比较要处理学习、干预和边界移动

实际研发流程会随时间学习。工程师可能在第一轮 AI 使用后更熟悉提示和边界条件，因而第二轮任务更快；若只比较引入前后阶段，工具效果会与团队学习、人员变化、软件升级和产品成熟度混合。团队层级随机分派可以降低任务自选偏差，但研发团队之间仍可能共享设计模板和经验，造成组间干扰。分阶段推广应保存每个团队首次接触日期、培训时间、版本更新和跨组交流，分析时说明采用的是“按分派流程”的效应还是实际使用程度的关联。

更重要的是，干预边界必须在研究开始时明确。若 AI 只参与候选生成，研究估计的是这一环节；若它还改变设计审查、仿真调用、打样优先级或工程师分工，结果属于组合式工作流。中途增加专家审核或改变验收要求，可能是合理的安全响应，但会改变原先比较的问题。研究者不应为了保持统计整洁而压制必要的工程判断，而应记录决策、影响范围和时间，并分别报告原始方案与修订后的实际流程。这样才能判断观测到的提速或延迟来自模型、配套资源还是审批边界的变化。

14 把现场研究变成可执行比较：先冻结问题，再分派工作流

14.1 干预是什么，必须在第一张任务单上说清楚

“使用 AI”不是足以复现的干预定义。某团队可能只用生成模型提出几何初稿，另一团队则用代理模型筛选、脚本自动化 CAE、生成试验计划，并由工程师按模型推荐顺序加工。若把两者都编码为 AI 组，平均效果没有明确含义。现场

SVSAC | sv-sac.org

试点的方案应说明工具版本、允许输入、输出类型、使用环节、人工复核权限、可调用的传统工具、使用培训和任务分工。基线组也要写清楚可用 CAD/CAE、既有模板、专家咨询与外包资源，避免把“正常工作”当作不需定义的剩余类别。

建议将一个可执行的工作流作为干预包，而不是把模型名称当处理条件。例如，干预组可获得“受控生成候选–初筛–独立工程审查–计划内高保真分析–实体验证”的支持；对照组按项目原有的 CAD/CAE 与试制流程工作，两组都必须满足同一安全、合规和客户验收要求。研究结果届时估计的是这个完整工作流相对比较流程的差异，而不能拆成模型单独的因果效应。若需要分辨特定功能，可另行设计析因或多臂比较，但有限团队数可能无法同时识别每个组件的独立作用。

14.2 冻结需求、验收与任务资格

每个项目进入比较前，应形成版本受控的“任务基线包”：产品用途和系统边界、目标工况、适用材料/制造路线、设计自由度、不可越过的硬约束、客户或法规要求、首样验收测试及其判定规则。验收项目应尽量写成可观察结果，例如尺寸、接口配合、静载或疲劳工况、功能循环、表面质量与记录完整度；不能把“工程师认为可用”留作唯一标准。方案冻结后变更仍可能必要，但应保留变更前后版本、批准人、原因、影响的测试和已经完成证据的适用范围。

任务资格规则同样需要预先确定。可以按产品族、设计成熟度、材料/工艺、既有模板可复用程度、验收难度、外部供应周期和风险级别形成层次，再决定纳入、排除或单独报告的情形。正在等待客户需求、关键尺寸尚未确定、工艺资格未建立或供应商停产的项目，并不一定适合放入同一主比较；排除规则应在看到分组结果前应用。资格判定若依赖资深工程师意见，应有第二名评审者对一部分项目复核并记录分歧，而不是事后根据项目进展重新定义“可比任务”。

14.3 单位与分派：不能把同一团队的候选当作独立实验

由于设计模板、仿真经验、同事交流和操作技能会在团队内扩散，若任务在同一团队内逐个随机到 AI 或常规流程，团队可能把 AI 产生的经验用于对照任务，形成污染。较稳妥的试点可以以相对独立的项目团队或产品小组作为分派单位，将多个待开发项目先按产品族、成熟度和预期验证负荷分层，再在层内以团队为单位随机分派。任务是嵌套在团队中的观测，不会因为一组有多个候选就增加独立随机样本数。试点开始前要盘点团队之间共享人员、设计库、外包分析、工装和评审委员会的情况；若共享过强，团队随机化的独立性假设不成立，应改为按业务单元分阶段引入，或使用有明确边界的准实验设计。

SVSAC | sv-sac.org

分派时应由不参与项目交付的协调者在任务清单冻结后生成分配表，保存随机种子或执行记录、分层规则和任何人工覆盖。工程主管仍可因安全、客户或资源原因调换人员或改用另一工具，但这些偏离不能删除原分组信息。主要分析优先按最初分派比较工作流的意向处理效应；实际使用强度和越组情况另行描述，不能将“完成使用 AI”的项目与“没有使用 AI”的项目直接做无条件比较，因为使用程度通常与任务难度、工程师偏好和问题严重性有关。

14.4 学习、顺序效应与不能随机的工作

团队随机分派也无法自动消除随时间变化的工具熟练度。若干预组先集中培训，随后承担一批相对成熟任务，而对照组在同期处理复杂新结构，结果会同时受到任务选择与学习影响。设计上应预留共同磨合期，明确磨合任务是否计入主要效果；记录团队首次接触、培训和工具升级时间；对按日历时间实施的分阶段推广，比较尚未接触的同期团队，而非只比较一个团队的上线前后。

有些任务不能、也不应为了研究而随机分派：如涉及安全关键接口、客户明确指定方案、必须完成的认证测试或无法复用的唯一原型。可把这些任务放入外部可验证的机制观察或病例式记录，不与随机比较的主结果混合；若必须纳入，则应使用事前可审查的匹配规则和清楚限定的因果解释。安全和资格测试不可由分组决定是否执行。研究计划的目的，是分派可比的工作流和测试资源，而不是随机取消必要工程控制。

15 版本、盲评与独立参考：让验收决策建议不追随模型分数

15.1 设计到样件的实体链必须可追溯

可复核的试点记录应让每个受验样件回溯到需求基线、CAD 文件、模型建议、工程师修订、仿真网格、边界条件、材料批次、制造程序、后处理、测量设备和测试结果。设计版本号不应只依赖文件名；可使用受控仓库提交标识或文件哈希，并由版本链说明父子关系及变更原因。由 AI 生成后经人工修改的 CAD，属于新的实物候选，必须将修改前的模型输出和修改后实际送制的文件同时留存。模型训练/更新版本、检索资料快照、提示和参数也应留档，以确定生成工具在当时实际看到什么。

试验协议需限定“参考结果”由何种设备或独立工序给出。几何尺寸可以来自经校准的量具、坐标测量或全场光学测量；承载性能由预先选定的夹具、载荷和循环程序验证；装配通过则需定义插入力、配合状态或通电功能标准。一个高保真 CAE 模型不能同时作为候选选择器和唯一的最终真值。若独立实物测试受成本

SVSAC | sv-sac.org

限制，可在预定风险层随机抽取候选做盲态复核，并保留抽取概率；只有测试被选中候选将产生验证偏倚。

15.2 盲态评审如何在可行范围内实现

拓扑形状或文件格式可能暴露候选由何种工具生成，所以工程评审不一定能做到完全双盲。可行的程序是将样件、图纸和验证数据交由不知道分组的验收小组，使用去除 AI/对照标签的编号包；技术安全评审者可看到必要的技术细节，但不查看模型评分、提示内容或项目预期结果。评审者按预先设定的项目表记录每项判定、测量依据、偏差与复核建议，不在首轮判分前由研发团队解释“模型为什么认为它更好”。如果特征本身使分组不可隐藏，应记录评审者猜测的组别和信心程度，分析这种猜测是否与判定相关。

验收量表可以按问题类型设硬门槛和描述性维度。例如结构件先按尺寸、安装接口和指定载荷/循环条件判定通过或未通过；外观件将色差、平整度与装配配合分别记录；电气件将机械插合与连续性/功能分开。不同维度不可随意加权成一个“AI 质量分”。若设综合分，权重必须来自客户或工程要求并在分派前冻结。评审者分歧应先由独立裁决人按同一证据复核；若参考测量本身处于公差边界，应进行重复测量或不确定度评估，而非简单取有利的一次结果。

15.3 把结果从生成通量改成通过验收的路径

主结果可定义为从任务基线冻结到首个经盲态验收组签署的实体样机的自然日数，并同时记录工作日、关键路径活动日及停等供应时间。以样机是否通过验收作为事件，而非“设计文件已生成”“仿真已收敛”或“首次试模完成”。辅助结果包括首次样件合格比例、达到验收前的物理样件数、关键门槛返工、测量不确定度、工程人员时、设备/供应商占用、被取消路线及验收后强制改版。每一个辅助结果都要有分母：按项目、按实际制造样件、按独立设计分支的统计意义不同，需一并披露。

同一项目可能有多个并行样件或设计分支，因此“首次通过”只能定义为项目级时间结果；不能把一个项目中的十个候选当作十个相互独立的项目。所有候选可用于描述筛选过程，但因果推断要按分派的团队/项目层级进行。若样机从未通过而研究期结束，保留其观察和停止状态，不能把未通过编码成成功或给一个任意最大天数。竞争性取消与右删失的定义按前章确定，需在数据采集表中分开。

SVSAC | sv-sac.org

16 失败、偏离和边界改变：不要让现场不确定性变成删数规则

16.1 先把失败分解为可诊断事件

一项设计可能在几何修复、边界条件复核、材料采购、加工、后处理、尺寸检验、性能试验或客户验收中的不同阶段停下。事件表应至少记录发生时间、当前版本、阶段、失败类别、证据材料、实际耗用、是否重试、由谁决定下一步及其理由。模型越界、求解器发散、供应商交期延误、设备故障、样件破损、未达到验收条件和因优先级取消不能共用一个“失败”标签；它们对 AI 适用性和日历时间的解释不同。供应商或设备变化应被记录为情境变更，不因它“外生”就从主分析中抹掉。

16.2 偏离预案应允许工程师采取安全动作

随机分派并不要求团队机械服从不安全或不适用建议。协议应预先列出必须停止、人工接管或转回常规路线的触发条件，例如候选跨越硬安全边界、材料批次不可追溯、独立复核与模拟出现超出既定范围的差异、关键数据缺失或客户改变使用场景。触发后保存原方案、偏离时间、决策人、风险依据和后续路线；主分析仍按原分组计入，实际路线描述在次级分析中报告。这样既保护实际工程责任，也避免只统计“AI 成功且没有被接管”的精选子集。

16.3 缺失、退出与验收改变怎样处理

项目暂停可能源于供应、预算或客户；每一类原因均应在观察窗口内更新而非到结项时回忆。对研究数据不可得的单个节点，应记录缺失字段、责任主体和是否可恢复；不得用项目成功后的回忆值补出精确时间戳。需求变更若只是提高验收门槛，保留原目标和新目标两个版本；若产品功能已变到无法定义相同测试，记录为目标变更的竞争事件并报告数量。估计结果应按预先规定的共同目标口径分析，变更后的新目标另作描述。对关键字段缺失，可给出上下界或敏感性分析，而不选择最有利于工具的一种补法。

16.4 试点规模由可识别信息决定

小团队或少量唯一原型无法支持稳定的行业平均效果估计。没有基线项目历时、团队间差异、组内相关、预期事件率、观察期限和可接受区间时，不应编一个通用样本数或设定“至少快多少”门槛。实施前可用过去项目日志估计结果方差及团队内相关，模拟不同团队数和随访长度下区间可能有多宽；如果数据不足，

SVSAC | sv-sac.org

就把试点目标定位为流程可行性、事件链完整性和机制参数，而不是工作流效果确认。研究报告应明确哪些结果会被判作证据不足，哪些反向结果会推翻提速假设。

17 分析计划与因果解释：比较整个流程而非参与者的选择

分析计划在第一组任务启动前登记主要结果、分派层级、排除条件、分层变量、估计窗口、停止规则、缺失处理、竞争终点和中期观察权限。主分析按随机分派的团队或项目比较完整工作流，推断单位与随机化单位一致；项目数量多于团队数量时不能按独立项目误差计算而忽略组内相关。对于等待期差异明显的任务，可按预定产品族或验证负荷报告分层效果；若层内只有少数团队，应以不确定区间为重点，不给排名或强因果措辞。

如果团队随机试点不可行，可用匹配项目和分阶段引入，但匹配必须使用结果出现前已知的任务特征，并把剩余混杂列明。比如材料新颖度、验证周期、供应商成熟度、团队经验和工具部署前准备度均可能共同影响是否选择 AI 以及项目历时。简单把这些变量放进回归不一定修复自选偏差；研究者应说明设计识别依赖什么条件，以及哪些潜在因素仍无法观测。前后对照特别容易受到版本更新和日历时间影响，需报告同期业务变化，而非只给一个平均天数差。

结果应同时给出点估计、与分派单位一致的不确定区间、各组分母、接受事件、删失、取消、越组和安全接管数。若不能可靠估计组间差异，应明确证据不足，而不是把每个被试候选视为独立样本以制造窄区间。任何按实际使用、通过筛选或完成试验的子组分析，都应标为探索性，因为这些分组在分派后形成，可能受处理影响。机制分析可以追踪候选数、审查队列、物理试件、返工和资源等待，但这些中间结果不能替代主要终点；要说它们解释总周期，仍须满足额外的因果识别条件。

18 分析性情境一：注塑外观件，快算不等于快过试模

以下是用于比较机制的构造情境，不对应真实企业、产品或项目记录。设想一款带卡扣和可见表面的注塑外壳，需求冻结后，设计团队用 AI 辅助提出浇口、壁厚和局部加强筋候选；CAE 对充填、保压与翘曲进行筛选，随后由模具供应商加工试模。它的关键路径至少包含模具结构与制造、首轮成型、冷却和尺寸稳定、外观/装配验收，以及必要时的模具修正。生成多个参数方案可能缩短前段搜索，却不会消除钢料加工、试模窗口、产品调湿或重新修模所占的实体时间。

SVSAC | sv-sac.org

注塑件不能仅凭最大翘曲或仿真收敛判为通过；脱模和残余应力释放会改变外观面，卡扣及螺钉柱还须与对件配合。Dupuis 等的可读摘要支持数字图像相关用于板件全场收缩测量，但不能据此声称 AI 减少商业修模轮次。[14] Wang 等原文描述了九个局部加热区的聚丙烯平板模具、实测温度与翘曲及神经网络比较；其 2.5 节同时把训练、验证、测试比例写为 75%、15%、15%，合计 105%，存在报告内部不一致。[17] 本节不猜测改为哪一组比例。表 3 的分组数量也不能单独证明同件多点已按独立制件划分；这部分须原始分配记录或作者澄清。在此之前，该研究可作空间响应和测量边界的案例，不能以其拟合指标证明对独立新制件的泛化，更不能证明开发日历缩短。

因此，这个情境应将候选送入实体试模前的决策问题写为“在不违反卡扣强度、外观和装配门槛下，哪组可制造工艺/几何候选值得进入模具验证”，而非“哪一组预测翘曲最小”。若候选之间只差可在现有模具上调整的保压或冷却参数，快速搜索或许影响关键路径；若主要不确定性来自浇口位置、分型线、滑块、顶出结构或模具钢料加工，设计计算即使更快，瓶颈仍在工装制造。若产品已有稳定模具且问题只是一个局部参数窗口，AI 的新增价值也可能很小：传统小规模试验或工艺工程师的已验证窗口可能更省接入成本。

可检验的现场预测不是“AI 必然少试模”，而是：在任务确实受可搜索工艺变量限制、模具方案已冻结且供应周期可比的项目中，工作流若有效，应该先降低达到可行试模窗口前的设计/仿真迭代或无效试模；随后才可能降低从需求冻结到独立验收通过的时间。若计算迭代减少但模具变更、样件数、返工或签署时间未变，说明算法改善没有转化为实体关键路径收益。若外观验收随观察角度、照明或色差标准变化，项目间“通过”也不可比；应在试模前固定观察条件、抽样位置、配合件版本和判定者流程。对每轮模具修正，都需将原因映射到实际误差来源：材料批次、模温控制、测量误差、模具尺寸或结构假设，不应一律归因为 AI 预测失败。

这一情境还揭示了比较中容易忽略的选择问题：供应商通常会优先安排较成熟或更重要的模具项目，团队也可能只把需求清楚、历史数据多的零件交给新工具。如果 AI 项目因此更易成功，按“已完成项目平均试模次数”比较会偏向工具；反过来，管理者也可能把最棘手的翘曲问题交给 AI，造成偏向无效的表面差异。纳入规则和任务基线应在分派之前确定，并把模具复杂度、材料新颖度、已有工艺窗口和供应商排期作为任务特征记录。小样本下这些调整不能自动消除选择偏差，但能使不确定性显性化。

SVSAC | sv-sac.org

19 分析性情境二：金属承载件，拓扑候选必须服从代表性疲劳验证

第二个构造情境是一件减重承载支架：AI 或拓扑优化提出几何，再由工程师重建可制造 CAD，选择增材或切削路线，完成后处理、尺寸检查、静载和疲劳测试。这里的关键路径不是单一“求解时间”，而是需求载荷谱能否确定、几何能否变为可制造和可检验实体、材料与制造方向是否代表目标产品，以及疲劳测试是否覆盖真实服役条件。优化器可能在几个小时内输出轻量候选，但每次改变支撑、圆角、打印方向、热处理或表面处理，都会改变验证对象。

既有稿引用 Chen 等航空支架研究，列有约 37%减重、Ti-6Al-4V 和未热处理状态，设计载荷 50/5 kN、试验载荷 100/10 kN 及 6,201 次循环断裂等具体信息。[15] 该文原文未能重新读取，故这些数值、条件及目标载荷寿命推算均保留为待核引述，不能当作已确认的试验结论。需要逐项核对载荷上下限、制造及热处理状态、样本数、失效定义和曲线外推依据；任一关键条件不同都影响桥接。可保留的一般方法论点是：几何或减重优化、模型预测与目标工况实体验收是不同结果，不能用一个结果替代另一个；该论点不依赖暂未证实的具体循环数。

若项目目标是证明支架满足一个明确载荷谱，测试报告需要注明夹具、加载比、频率、环境、表面状态、打印方向、后处理、样件批次和失效判据。若因试验资源而采用较高载荷加速，必须事先说明模型怎样将加速条件映射到目标条件、哪些材料和缺陷假设成立，以及何种不确定性仍需保留；不能在试件未达目标条件时把模型推算写成“已通过”。疲劳寿命又常受微观缺陷、残余应力和表面粗糙度影响，单个漂亮样件不能代表过程分布；而完整疲劳试验可能成为整个项目的时间下界，算法压缩初始设计搜索未必改变它。

在这里，AI 可能产生两类相反影响。其一，筛掉低刚度或明显不可制造候选，减少进入昂贵试件阶段的设计分支；其二，扩大候选集，使工程评审和实体疲劳试件需求上升。即使最终候选更轻，若筛选规则更依赖未经验证的疲劳代理，过早淘汰有稳健裕量的常规方案，也可能增加后续返工。可检验的机制指标应因此包括：候选进入独立验证的数量、被淘汰方案抽样复核中的漏选、几何重建偏差、实际制造/后处理路线变更、达到验收的样件数和总历时。仅比较生成数量、有限元静态安全系数或质量比，无法判断项目是否更快获得可签署的合格实体。

反例也很关键。若设计空间边界已由成熟规范限定，既有部件具有相似载荷和可直接复用的疲劳数据，AI 生成的只是几何候选排序，那么少量增量试件也许足以完成工程确认；反之，对于新材料、复杂增材方向或失效后果高的唯一部

SVSAC | sv-sac.org

件，独立试验和资格流程不可能因为仿真更快而消失。项目间应按证据要求和失效后果定义“样机完成”，而非把不同资格层级都统称首样成功。若一组允许设计计算替代某些实测，而另一组仍执行相同试验，两组时间差比较的是不同验收任务，不是同一标准下的开发效率。

20 分析性情境三：机电装配件，CAD 接口不等于功能配合

第三个构造情境是一种带线束、插接器、壳体和控制板的小型机电组件。设计生成或参数优化可能压缩支架、外壳和布线路径的迭代，但可交付样机还需要将特定供应商部件装配、完成插合与通断检查，并在振动、温升或寿命等要求下验证。产品结构中的接口依赖供应件的真实尺寸、柔性、插拔力、线束回弹、锁止特征和操作空间；数字装配可显示零件不相交，却不必然预测插入力、偏斜卡滞或重复连接后的电连续性。

Yumbla 等的出版摘要可确认研究了 70 类插接式连接器，使用 Robotiq 两指夹爪和 ABB 六自由度机器人测量配合公差，并认为所测公差大于工业机器人重复定位尺度，位置控制可用于其研究设置。[16] 完整方法未能取得，既有稿所写的力/连续性传感程序或 CAD 容差与实测对照无法确认，故不再把这些细节归为已核事实。它是特定装配测量研究，不是 AI 开发提速试验，也不代表所有连接器或供应商。若某产品的验收还要求插拔力或电气连续性，应另找相应测试依据；这是我们提出的验收边界，不是该摘要已经证明的结果。

在构造情境中，团队应在任务开始时冻结连接器料号、供应商和批次范围、配对件、线缆规格、插入方向、可接近空间及功能测试。随后把验收拆成机械和电气两个部分：配合成功不能只用“可以插上”描述，还需定义插入力/行程的可测要求、锁止、拔出或循环条件；电气侧定义连续性、接触稳定性和功能状态。若供应商在中途换型或缺货导致替代料，必须把接口规格和样件版本作为需求变更，而不是把新部件测得的结果覆盖原结果。对某些件而言，供应波动比设计搜索更可能决定日历周期；另一些任务则因布线可达和公差链反复返工而受设计影响，不能先验认定统一瓶颈。

该情境的 AI 收益边界与前两个不同。若工具擅长生成可编辑的装配布局，并且连接器的实际包络与配合规则已稳定测得，它可减少空间冲突和人工改版；若样件验收主要失败在供应批次、线束工艺或人工装配一致性，设计生成提速可能几乎不影响功能确认时间。若模型对单个连接器静态 CAD 包络预测很好，但装配机器人/人员的插入角度或柔性线束导致失败，原因不是简单的“模型准确率

SVSAC | sv-sac.org

低”，而是模型所用任务定义遗漏了动态配合变量。可用失败编码区分几何干涉、可达性、插入力、锁止、电连续性、热/振动后功能及物料等待，再观察 AI 组减少的是哪类事件。

一个有信息量的现场比较应报告候选版本之间的接口改动和已冻结部件清单，记录每次实体装配的对象和执行者，独立测量插合力/行程并做电气测试。若只能抽测部分候选，抽样应覆盖被模型排序靠前和靠后的设计，而不是只确认胜出件；否则无法知道筛选过程是否遗漏了稳定的常规方案。量具、力传感器或通断测试程序需按实际测量重复性安排复测，单次插合成功不足以推断产品族的装配能力。

20.1 三种情境为何不能共用“样机提速”解释

三个构造情境都能使用生成式设计、代理模型或自动化搜索，但物理约束和验收关键路径不同。注塑件的主要不确定性可能在模具/热过程的空间响应与反复修正；承载金属件的主要约束可能是载荷条件、制造方向和疲劳证据；机电组件则可能受接口动态配合、供应件和功能闭环影响。若三者只用“首版 CAD 到达时间”比较，所得指标会共同忽略试模、疲劳和装配测试；若用“从需求冻结到签署验收”比较，仍需确保各自验收标准足以代表相同产品阶段，而不是三个不同严格程度的门槛。

可以将跨情境机制写成可反驳的因果链：工具先改变候选生成/筛选和工程审查负荷；这影响进入工装、打印或装配的实体版本数；实体版本数及其信息质量决定测试资源、返工和供应等待；最终这些节点改变首个满足冻结要求的样机时间。链条中任何一环未发生预期变化，都可能使计算提速不转化为产品进度。更直接的解释也须比较：团队经验、既有模板、供应商成熟度、项目优先级、材料/部件可得性、验收口径变化，都可能同时影响 AI 使用和周期。

一个项目级比较表可为每类任务记录：工具能够改变的决策节点；它无法改变的物理/认证节点；实体验收的独立参考；最可能的瓶颈迁移；以及能反驳“提速”解释的观察。注塑件可把模具修改轮次减少但签署时间不变作为反证；金属件可把轻量化提高但目标载荷疲劳未验证作为验收失败；机电件可把 CAD 间隙问题减少但插合/电气试验不变作为局部模型作用的边界。该表不是通用的结果评分量表，真正比较时要在每个产品族内部定义可比较任务，跨族汇总仅可作为异质性描述。

比较还存在层级问题。算法可能在零件层提升局部候选表现，却在系统层增加接口数量、部件版本和评审负担。对三个产品族而言，单个零件的最优设计未

SVSAC | sv-sac.org

必减少整机装配时间：注塑壳体可能更轻但卡扣难装，支架可能更薄却遮挡工具路径，机电外壳可能节省体积但线束弯曲半径不足。是否把系统装配及维护纳入验收，要取决于冻结的产品需求；若纳入，则样机单位必须是规定系统级对象，不可把零件级数字与整机级验收混为一个终点。

最终能得到的结论应限于被观察的任务、团队、工具版本和验收条件。某个构造情境中的因果链只是研究设计的逻辑，不是对任何企业已经发生的证据。后续现场研究如果没有独立参考、版本记录、任务可比性和完整失败记录，即使报告样机周期缩短，也很难分辨是模型效率、项目选择、验收放宽还是供应条件变化所致。

21 验收边界附近的测量与模型不确定性：验证资源应投向会改变决定的信息

21.1 “预测不确定”不是一种误差

一项工程预测至少可能受到四种不同来源影响：模型输入参数未知、代理/求解器近似误差、物理模型与实际对象之间的结构差异，以及实物测量本身的不确定性。它们不能都压缩成一个置信区间。增加同类仿真点或重复神经网络训练，可能降低近似器的不稳定，却不会自动修复遗漏的接触、材料非线性或制造残余应力；增加测量重复次数能减少部分随机测量分量，却不会校正量具偏差或不代表目标载荷的试验条件。Kennedy 与 O’Hagan 提出的计算模型校准框架明确把参数不确定性和模型与观测之间的差异纳入预测；其论点是，拟合参数后仍需对剩余不确定性与模型不足留有空间，而不是把最佳拟合当成真实系统。[19]本篇只依据作者摘要和出版元数据描述这一方法方向，未把论文示例当作制造现场验证证据。

测量应明确被测量、测量模型、校准或修正、重复性及覆盖区间。本节引用 JCGM 106:2012 第 7.1、7.5 节及第 8.3.2 节的相关内容：符合性概率取决于给定测量信息下量值的分布，覆盖区间与公差的关系影响判定，护带可改变错误接受风险。[18] 覆盖区间跨界并不能在缺少分布和决策规则时确定一个通过概率。该指南是计量框架，不是本次分析产品的法定认证规则，不能替代产品安全决策。该指南的目录状态及 2026 年更新未及核对，故不据此作现行法规或最新版本声明。

在样机验收附近，三类量常被混在一起：预测结果离目标的距离、预测误差大小、以及测量系统能否区分合格与不合格。即使一个代理模型在验证集上的平均误差很小，若候选的预计性能离硬门槛也很近，误差尾部仍可能改变选型；反

SVSAC | sv-sac.org

之，预测不确定度较大但所有合理结果都远在可行区内时，额外高成本测试也未必改变决策。测量系统的不确定度与模型不确定度还可能相关，例如测试结果用于校准模型，而量具偏差和夹具变形共同影响比较。此时简单把两项标准差平方相加会暗含独立、近似线性等假设，不能由一个公式自动保证。

21.2 错误接受、错拒和多指标决策

工程验收至少涉及两种方向相反的错误。错误接受是实际不满足要求的对象被判断为通过，可能导致安全、可靠性、返修或客户后果；错误拒绝是合格对象被判不通过，可能增加样件、材料、等待和机会成本。哪个错误更严重取决于用途、失效后果、是否可逆、测试可重复性和后续屏障。安全关键指标上，模型给出的“通过概率高”不能自动授权放行；模型概率只描述在既定建模假设和数据条件下的不确定性，不是工程许可或法规签署。工程责任人和适用规范仍决定是否需要独立验证、保守边界或停止。

多指标更难用一个概率数字概括。某承载件可能同时有质量、刚度、屈服裕度、疲劳寿命、装配尺寸和材料供应约束；注塑件则可能有翘曲、外观、卡扣力和周期；这些结果既有不同单位，也有不可互换的硬约束与偏好目标。若优化器把所有结果加权为一个综合分，权重变化便可能改变排名，而该加权并非模型从数据中“发现”的客观优先级。验证方案应对每个不可交换的门槛单独报告测量对象、判定规则和不确定性，再将允许折衷的目标放到明确的决策层级。若任一硬门槛未被独立测量，综合分不得掩盖证据缺口。

资源有限时，额外验证的分配应由“它可能改变哪个决定”以及错判后果共同决定，而不是只挑模型方差最大的候选。可把每个待验候选放入一个决策矩阵：预计值相对门槛的位置、测量与模型不确定性、潜在错判损失、验证成本、结果能否影响其他设计分支。靠近门槛且错误接受损失高的候选，通常需要更独立或更严格的证据；远离门槛但验证结果将用于校准一整类后续设计的候选，也可能有较高信息价值；预测方差高但无论结果如何都不会改变设计路线的测试，则未必值得占用稀缺设备。这里的“通常”是资源配置假设，需要在不同工程体系中用决策日志检验，不是统一排序定律。

21.3 建立分阶段的验证账本

一项可操作的验证账本可将样机项目按以下层次记录。第一层是输入证据：材料批次、工况、边界条件、传感器/量具、CAD 和制造版本；第二层是模型证据：训练覆盖、求解器和代理版本、参数不确定性、已知模型差异、外推距离；第三层是验收证据：实测指标、重复测量、覆盖区间、独立盲评、判定人及任何

SVSAC | sv-sac.org

未执行测试；第四层是决定：通过、拒绝、补测、限制用途或工程接管。每次从较便宜证据升级到较贵证据时，应记录升级触发条件与预期改变的决策。这样可以区分“高风险候选获得更多测试”与“模型组受到更多资源倾斜”两种解释。

一种分阶段规则可以先用低成本几何/解析检查排除明显不满足项，再用高保真仿真和独立审查检验边界候选，最后把风险高、接近验收边界或具有代表性信息价值的候选送实体试验。但阶段门不是减少验证的许可证。筛查器的漏排率未知时，不应以其筛掉低分候选作为已证明有效；验证样本最好包含一定比例的被淘汰对象，抽样概率和纳入方式预先记录。若测试结果影响产品安全或资格，独立工程批准优先于实验效率。对每一层都需明确哪些情况必然升级，哪些情况可在充分证据下停止，以及谁有权限批准该决定。

21.4 一项明确标为假设的量测例：读数相同，行动仍可不同

设某机加工接口的单边尺寸上限是假设的 10.00 mm，项目已经事先规定：只有在测量模型给出的覆盖区间上端不超过该上限时，才可按这项尺寸判定通过；区间跨过上限时先暂缓结论并复测，不直接放行或报废。第一次方法的读数假设为 9.98 mm，扩展不确定度为 0.04 mm（假设覆盖因子和模型条件使其对应一个约 95%的覆盖区间），则区间约为[9.94, 10.02] mm，跨越上限。此结果并不是“有 95%概率合格”或“有 5%概率不合格”；覆盖概率描述的是在既定测量模型及其输入假设下，测量所表征被测量值的区间覆盖性质，不等同于符合概率、错误接受概率或整件安全概率。JCGM 106 对错误判定风险的定义还需结合公差、验收规则和有关被测量可能值的分布来计算，不能从覆盖区间读出一个数值。

再假设项目改用校准和重复性更好的方法，得到同一中心读数 9.98 mm、扩展不确定度 0.01 mm；在相同且预先声明的上端保护规则下，区间约为[9.97, 9.99] mm，这项单边尺寸证据可以通过该规则。第二次结果改变的是对测量量的证据，不是产品真实尺寸被“改善”；也不证明其他接口、强度、疲劳或系统级要求合格。若保护规则改为直接比较单次读数，或上限来源、校准偏差、相关误差模型不同，决策可能变化。例中数值全为教学假设，没有估计实际假接受概率，也不规定任何工业公差。资源账本应记录为提高测量能力付出的重新装夹、校准、重复测量和设备占用；若边界候选的测量常常跨界，降低量测不确定度可能比增加模型采样更能改变验收行动。

21.5 测量能力和模型可信度可以共同限制决策建议

在极限附近，重复测量不是形式主义。例如某特征测量结果落在公差内侧，但测量区间跨过边界；一次通过不能消除不确定性。重新装夹、换操作员或换量

SVSAC | sv-sac.org

具可能揭示重复性与重现性差异，但只有在测量程序和被测样件稳定时，重复测量才估计同一对象的量具波动。若制造批次或样件版本同时变化，所谓“重复”已包含产品差异。试验协议需要把重复量测、重复制造和重复工况区分开，并说明每一种增加了什么信息。

模型校准也并非总能区分输入参数错误和模型形式错误。若可调参数与差异项能相互补偿，一组校准数据可能使预测在已测工况很好匹配，却在新材料或边界下失准。因此，模型复核应留出与校准不同的工况、方向或版本作外部检验；若只能使用同一批数据反复选择模型，报告需承认验证与选择未独立。可迁移性更需在变更边界附近检验，而不是把全数据集平均拟合误差直接搬到一个关键候选上。

本章的增量在于把“模型可信度”落到验收决定和稀缺验证资源分配上：不确定性量化本身不等于安全结论，测试价值取决于结果会否改变分派、制造或放行决定，错接受与错拒的损失也不对称。具体门槛和可接受风险只能由产品任务、责任制度与有效要求确定；本篇没有给企业或行业设定通用概率阈值。

22 设计变更后的证据再利用边界：版本可追溯不等于验证仍有效

22.1 先问原证据验证了什么命题

版本追踪回答“这个结果来自哪个样件、材料和模型”，证据再利用则回答“变更后的对象是否仍满足原验证命题”。二者不能混为一谈。设计文件父子关系完整，并不意味着子版本可自动继承父版本的试验结论；相反，即便几何文件变化不大，材料批次、热处理、载荷方向、装配预紧、供应件或测试夹具若改变了主要失效机制，原证据也可能失去适用性。每一项仿真或实体测试应登记一个明确的验证对象、适用工况、假设与失效模式，而不是只留下“通过”标签。

可先把已有证据分成四类：几何/接口测量证据、特定材料和制造过程的属性证据、指定边界条件下的性能证据，以及整机或系统任务证据。几何尺寸证据可能支持接口名义尺寸仍相同的变更，但不自动支持疲劳寿命；同一材料牌号的材料数据也不证明更换打印方向或后处理后仍保持相同缺陷分布；单一静载试验更不能替代不同均值应力、载荷比或循环谱的疲劳证据。证据适用范围应逐项写，不因它们存于同一个项目数据库就被合并为“产品验证包”。

SVSAC | sv-sac.org

22.2 变更影响矩阵：比较失效机制，而非仅比较几何距离

在每次变更审查中，可用一张“影响–证据”矩阵记录变更项、可能被改变的物理机制、已有证据覆盖、缺口和补验证建议。几何尺寸变化需说明其相对关键应力集中区、接触面、流道、配合界面或测量基准的位置；材料/供应商变化需检查成分、批次、热处理与缺陷形态；制造路线变化需检查层间方向、刀具痕迹、残余应力、支撑移除和检验可达性；载荷变化需比较最大/最小载荷、循环、频率、温度和固定方式。相似性不是“整体 CAD 差异小于某百分比”，而是哪些控制失效的因素保持不变、变化是否远离关键响应区，以及已有模型和试验是否已在该变化范围内验证。

矩阵的决定结果不必只有“重复全部测试”或“全部继承”。可能的选择包括：仅重做尺寸与配合确认；在已验证范围内复算并抽测；用边界工况或最坏组合进行补充验证；重做独立性能测试；或暂停并重新定义需求。任何“无需重测”都是一个技术主张，应说明证据、界限与责任人，而非默认选项。若变化超出校准域、触发新失效模式或导致原参考不再代表产品使用，统计上的相似性不能补足实体证据。

22.3 一个具体变化链与一个反例

第 19 章承载支架案例的具体载荷、材料状态和循环数仍有待原文核对。[15] 以下仅讨论条件化的证据桥接：若团队只改不影响应力场的文字标注，结构疲劳结果可能仍相关，但不必靠疲劳试验验证标注本身；若改材料批次、构建方向、热处理或关键过渡圆角，则残余应力、缺陷方向和应力集中可能变化，需说明原试验为何仍适用，或补充目标工况试验。原工况与目标工况尚未核清时，不能把一条模型寿命曲线当成充分桥接证据，更不能由此宣布新品已验收。

反例是一个通过成熟工艺制造的标准化零件，在规范允许的设计空间内只改变远离受力区域的非承载文字刻印。若可证明刻印不改变壁厚、表面缺陷、质量或检验目标，已有主承载试验也许仍覆盖其核心失效机制；完全重测反而浪费资源。但即使如此，证明刻印不影响局部应力或腐蚀敏感性仍需适用于具体材料和位置，而不是从“尺寸改动很小”直接推断。相同体积差异对一个疲劳裂纹起始区可能重大，对远离应力的标识可能无关，因此证据再利用应该按失效机理和决策后果区分。

22.4 证据迁移需要明确适用域和反驳条件

项目变更中常把“看起来相似”当成证据继承理由。更可复核的做法是为每个已有结果标记适用域：材料及供应来源范围、几何/参数区间、工况、温度、制造

SVSAC | sv-sac.org

和后处理、传感/夹具、模型版本与测量不确定度。新版本若在这些维度内变化，团队可论证其是否属插值；超出范围则属于外推，需要新的验证或保守限制。该界定也应包含模型误差和测量能力：即便新版本处于数值区间内，如果目标响应依赖此前未观测的接触/失效机制，覆盖范围仍是表面的。

“保留旧测试结果”与“以旧结果证明新设计合格”也要分开。旧证据不应删除，因为它能显示设计演进和失败原因；但被引用用于新设计时，应标出版本差异、共享的机制假设、变化对结果的方向性影响及尚未验证部分。可设定可反驳预测，例如“变更后的模型应在独立复算中维持原候选排序”“新增样件的应变热点仍出现在同一位置”“换批次后接口尺寸分布不发生方向性偏移”。若观察到相反结果，停止继承或更新适用域，而非以调参消除反例。

再利用边界对周期研究尤其重要。AI 组如果因自动比对而减少重复测试，可能确实节约时间；但若它只是少做了必要试验，端点“更快验收”并不等于同一证据强度下的提速。因此主研究需把补测决策、被复用证据、批准人、测试覆盖和漏检后果列入工时/事件链。未来研究可比较严格相同证据要求下的开发时间，并把合理的风险分级再利用作为流程处理的一部分；两种问题不应混成一个结果。

23 一个可复算的有限样例：固定分母、观察窗口和资源账本

23.1 构造目的与设计边界

以下数据是明确假设的教学样例，不是任何真实企业项目、实验或抽样结果。它只说明如何同时呈现固定分母下的窗口结果、已通过者时间和资源消耗；没有进行随机试验、显著性检验或因果估计。设有四个相互隔离的项目团队，每队各有两个预先符合条件的样机任务。为降低知识泄漏，团队内共享人员与 CAD 库但团队间原则上不共享人员、候选模型或评审材料；供应商和测试设备按同一规则排期。若现实中仍共享评审委员会或设备队列，则团队分派不能视为完全无干扰，本算例不处理该复杂性。

四个团队以团队为分派单位，两个团队使用 AI 辅助的完整工作流，另外两个按常规流程。任务清单、产品族、需求冻结口径和 40 日观察窗口在分派前确定。每个团队处理两个任务，所以每组分母都是四个任务，但独立分派单位只有两个团队；八个项目绝不应被当成八次独立随机分派。每个项目均从需求与验收冻结日记时，以首个经独立签署、满足原要求的实体样机为通过事件；第 40 日仍未通过则计为观察期内未通过，正式取消记录为竞争事件。日历日连续计数，外部供应等待不扣除。

SVSAC | sv-sac.org

23.2 假设事件表

分组与项目	到第 40 日状态	若通过则验收日	工程师总工时	实际设备/试验台占用	事件说明
AI-A	通过	18	52 小时	6 设备日	首个版本经盲评通过
AI-B	通过	29	76 小时	8 设备日	一次实体返工后通过
AI-C	仍在进行	—	68 小时	5 设备日	待供应商件，尚未完成测试
AI-D	取消	—	41 小时	3 设备日	第 25 日客户目标变更，原目标停止
常规-C	通过	20	61 小时	7 设备日	首个版本通过
常规-D	通过	38	88 小时	10 设备日	两次测试/改版后通过
常规-E	仍在进行	—	84 小时	9 设备日	测试排队，观察窗内未签署
常规-F	取消	—	50 小时	4 设备日	第 30 日预算转移，项目停止

表内的“工程师总工时”假设已包含需求解释、模型接入/培训摊销、数据准备、设计、审查、协调、失败分析和测试报告；“设备/试验台占用”只记录实体设备实际被项目占用的日数，不等于等待排队的日历时间。若项目采用新工具的一次性接入投入未分摊入这四个任务，表中会低估 AI 组成本；在示例里假设团队接入和训练已分摊计入各 AI 项目。样件材料费、第三方检测费和共享设备机会成本未换算成金额，避免虚构费率或把内外部人时重复计价。取消为竞争事件而

SVSAC | sv-sac.org

非普通未完成；常规-E 与 AI-C 截至第 40 日仍可能通过，属于行政观察终止状态。

23.3 固定分母与完整案例会给出不同摘要

按预定固定分母，AI 组 40 日内通过 2/4，常规组也通过 2/4；观察窗口内的粗通过比例均为 50%。这个描述保留了未完成和取消任务，因此没有只挑已通过项目。若只算已通过者，AI 组验收日中位数为(18+29)/2=23.5 日，常规组为(20+38)/2=29 日，看似 AI 组更快。两项中位数只是每组两名通过者的完整案例描述，不能替代所有分派项目的时间估计；组内尚未通过与已取消者被排除后，选择对象已经依赖项目进展，因果解释不成立。

40 日固定窗口通过比例同样不等于“实际完成时间的平均差”。样例中两组通过比例一样，但完成者的历时不同，另一半项目则仍在运行或被取消。若把取消当成一般独立删失后套用普通 KM，所得曲线不会直接给出现实条件下的验收概率；若研究目标是“验收先于取消”的概率，应预先选择竞争风险估计对象。样例只有四个项目/组且两支团队/组，任何模型区间都会极不稳定；在这里计算推断区间或 P 值会制造虚假精度。真实研究须按团队层随机分派，将组内任务相关性纳入推断，并报告足够团队数、随访和事件构成。

23.4 固定分母结果与受干预影响的资源比率

假设工时合计，AI 组为 237 小时，常规组为 283 小时；设备占用分别为 22 日和 30 日。它们是本构造样例的描述性总和，不是“每个项目节省 46 小时”或“因 AI 减少 8 个设备日”的因果估计。首先，两组只各有四个项目且团队数很少；其次，项目复杂度、取消和通过状态仍会决定成本累积；第三，总工时包括了分摊的接入/培训，但不含货币化第三方和机会成本。按每个通过样机的工时去除未通过或取消项目，会形成受处理影响的分母，可能把失败路线投入从结果中消去。因此主展示应同时保留固定的全部分派任务分母、窗口通过数、取消和未完成数，以及总资源账本。

如果管理者确实需要“每个验收通过样机成本”，可另作明确命名的描述性比率：全组投入工时/该组已通过数量。此示例为 AI 组 237/2=118.5 小时、常规组 283/2=141.5 小时。但它不是因果效率率，因为分母“通过数”可能受工具影响且四个样机不是固定产出量；每组分母仅为 2，也使微小事件变化造成比率大幅波动。更稳健的决策呈现是分开显示验收概率、完成时间、工程资源、设备占用和失效后果，再根据事先说明的业务目标权衡，而不是用一个商数掩盖目标冲突。

SVSAC | sv-sac.org

23.5 如何从示意账本走向真实可审计数据

现场数据应逐任务包含：团队分派、任务资格、版本/供应件、模型和人工活动、仿真调用、等待区间、设计分支、制造/测试事件、通过签署、取消原因、完整工时及设备日志。研究者在看到分组结果前定义 40 日或其他观察窗口、起点、事件、竞争事件、缺失规则和固定分母。工时记录应覆盖访问权限/培训、数据清理、求解配置、重建、评审、供应商沟通、试验准备、返工、报告和质量签署；若把任何一次性成本摊销，需说明估算任务寿命和分摊方式并作敏感性分析。

每组的工程、制造、质量和项目负责人可以在组别可见时登记活动，但验收判定仍由不知分派的评审人员按冻结标准执行；无法盲态的实体工艺评审，应记录其可见信息和可能的期望偏差。所有候选、淘汰、失联或停做项目均保留在任务清单。若竞争事件发生，固定窗口内的“是否通过”作为描述性目标依然可读，但其解释是“分派任务中在窗口内签署通过的比例”，并不等同于消除取消后的潜在通关机会。此类区别使算例可复算，也让未来项目不会用完成者的快慢来代替整个研发工作流的结果。

24 证据综合：计算搜索、流程标准化与实体验证回答不同问题

24.1 已发表证据支持到哪一层

已读 DeepCAD 相关原段涉及 CAD 命令生成，Yoo 等摘要与引言涉及概念车轮的自动 CAD/CAE 衔接；[2,3]这些范围支持数字探索和接口问题，不证明量产、实物验收或项目历时。PINN 首面与引言可提示软约束训练的失败边界，[6]不替所有代理模型定论。Folch 多保真文献未读，[11]不作为电池数据或已验证节时的依据；本篇只提出低保真能否传到高保真、候选排序是否稳定及资源是否可用等条件问题。

实体实验把证据推进一步，但仍没有回答普遍周期效应。Gongora 等报告将贝叶斯优化与高通量机械实验结合，并在作者摘要中称识别高性能结构所需试验数相较网格搜索减少近 60 倍；这属于作者机构摘要范围的实验数比较，不是完整现场项目历时，也不是随机并行流程比较。[7]Fadhel 等在特定铜合金和定向能量沉积参数空间中以物理试验寻找可行打印配置，报告相对过往手动实验过程的实验发现效率；其材料、设备和质量终点限定了外推范围。[5]这些研究表明自适应选择可以影响“用哪些试验发现可行窗口”，但不等于从需求冻结、设计、采购、加工、返工到签署验收的全周期变短。

SVSAC | sv-sac.org

PLA 支架与金属疲劳文献仍为待核线索，[10,15]不能在综合段恢复成已确认的制造、减重或断裂结果。已读 DIC 摘要及局部热翘曲相关段，分别涉及空间收缩测量与温度/翘曲关系，[14,17]不等于商业修模提速；连接器摘要只支持其 70 类型和机器人配合测量范围，[16]不支持本篇未读到的 CAD 比较。它们在各自边界内提示数字模型与实体目标之间的验证要求，不能拼成已完成的 AI 提速试验。

由此可以分为三个相互关联但不可替代的终点：算法是否减少候选生成/计算成本；在同等过程控制下，是否减少达到决策所需的独立分析或实体试验；是否更早获得按冻结标准通过的工程样机，并在可比资源口径下不增加风险与负担。本篇审阅到的证据对第一层支持最多，对部分研究场景的第二层有实例，对跨企业、同一验收标准下的第三层尚不足以估计普遍效应。任何直接把 GPU 时间、候选数或某次试验数变化写成通用样机周期收益的说法都越过了证据链。

24.2 反向结果和替代解释

AI 辅助组即使表现更快，也有至少四类非模型解释。第一，管理者可能把更清楚、数据更齐或战略优先级更高的任务交给工具，任务选择会造成表面差异。第二，AI 上线往往伴随需求冻结、模板清理、脚本自动化、培训和项目经理加强跟踪；这些一般流程改变本身可能降低返工。第三，AI 组可能获得更多工程审查、算力、供应商窗口或试验设备，提速来自资源倾斜而非推理。第四，验收门槛可能随项目进展而放宽，例如改用未代表服役载荷的模拟、减少独立测量、提前签署或把失败候选从分母中移除。若未追踪标准和全部任务，这些解释都难区分。

相反，观察到 AI 组较慢也不一定证明算法无效。新工具早期需要接入、数据清理和工程师学习；如果模型生成候选数上升，审查/制造队列可能暂时拥堵；如果最难任务被有意分派给 AI，简单前后均值会偏向负面。更关键的问题是这些成本是否一次性，未来能否摊销、是否可通过容量规划解决，和提速是否仍在同一验收定义下存在。只有把准备成本、使用阶段、后续维护和失败路径都记入，才可区分部署期成本与稳定运行收益。

流程标准化与 AI 的作用可以作可辨别预测。若只实施任务基线、CAD/CAE 版本和验收记录标准化，理论上应主要减少信息缺口、跨职能退回和重复确认；若 AI 候选排序有效，在相同冻结流程之上还应降低到达可行设计所需的候选/试验，且独立验收不恶化；若主要是资源倾斜，模型组可能缩短队列等待，但工程工时或设备占用不降。一个可检验的后续试验可采用团队为单位的 2×2 设计：常

SVSAC | sv-sac.org

规/标准化工作流与 AI/非 AI 功能交叉分配，并保持安全控制和验收规则一致。主结果按原始团队分派分析，记录团队共享人员、CAD 库、供应商和评审资源；若四个条件下团队数不足以识别交互，则只报告可行性与区间，不宣称机制分离成功。

此外，单个中间指标容易产生错误解释：候选数下降可能是算法更会筛选，也可能是团队少探索；实体样件减少可能是信息价值提高，也可能是试验预算缩紧；首样更快可能是验收降级，而最终再设计更多。因而应把关键链条两侧都观察：候选产生/淘汰、独立复核样本、试件数量与失败原因、资源队列、首次签署验收时间、后续改版和验收后质量事件。任何一项变化都应在固定分母与同等观察期内解释，不能只选择支持工具的过程指标。

24.3 证据升级矩阵：把“有效”限定在被实际观测的对象

已观察的结果	可以支持的解释	仍未证明的命题	下一步的关键观察
生成/重构基准表现改善	特定表示和数据域内的候选表达能力提高	工程师可直接制造、独立验收或周期缩短	几何修复工时、接口错误、制造/测试结果
代理/多保真搜索减少求解调用	在该目标函数或研究任务中搜索成本可能下降	高保真排序维持、实体误差受控	边界候选独立复算、被淘汰集合抽样验证
自适应物理实验发现可行参数	试验选择可能减少无信息探索	跨材料/机器的普遍试验节省或样机历时收益	同条件同期对照、包含培训/排队/复测的完整历时
单件样机通过某项实体试验	该样件在指定工况下达到该项测量判据	产品族过程能力、目标外工况和量产资格	代表批次、规定载荷/环境、适用域与后续质量
企业报告首样时间下降	在记录口径下项目事件发生较早	差异由 AI 造成且验收强度未变	分派设计、固定分母、盲评、版本和资源日志

这张矩阵并非给研究质量打一个总分，而是防止跨层级推论。尤其是“实体验证”也有不同强度：实验室单件、与需求一致的部件试验、系统级验收和生产代表性批次不是同一结果。研究综合应先注明每个来源的物理对象、比较基线、观察

SVSAC | sv-sac.org

时间、筛选规则和验收者，再说明其对端到端命题增加了哪一段证据。如果某一层尚无研究，就直接保留为空白，而非以相邻层级的更多论文数量补齐。

24.4 何种反向结果会改变总体判断

不同证据层级上的负结果也要准确解释。若 CAD 生成在数据集基准上失败，它主要质疑特定表示或任务范围，不直接排除物理实验选择工具；若候选在目标约束附近排序不稳，说明不能把该模型用于无人复核的筛选，但它仍可能用于低成本预筛或解释设计空间；若实验室中出现可行参数而企业项目周期不变，说明可行性发现不等于组织层面时间收益，或下游验证依然占关键路径。相反，如果同标准对照中 AI 组生成候选更少、实体验证结果相当、失败未转移到验收后且固定分母下通过时间更早，这才为端到端机制增加较强支持，但依然只适用于研究覆盖的任务和部署条件。

由此，研究者不应把“没有观察到周期改善”简化成模型性能无效，也不应把“模型在实验室找到可行解”升级成产品开发有效。前者可能是正确的局部作用未落在关键路径，后者可能是实验设备、工程权限或需求审批成为限制。比较应把模型对候选的影响、工作流对任务交接的影响和组织对资源的配置分开描述；若数据不能分辨它们，就把机制结论留为待识别，而不是用一个综合叙述填满因果链。

25 优先验证的关键问题：从单一产品族的流程识别到跨情境重复

25.1 先证明被测流程和终点稳定

下一步不必从“大规模 AI 投资回报”调查开始。较优先的工作是在单一产品族内抽取连续、符合资格的任务，建立从冻结需求到独立验收的版本/事件联结，并回溯足够的既有项目以估计团队内相关、事件比例、供货等待和验收历时分布。过去日志可用于规划可能的区间宽度及合理观察期，却不能替代未来随机化的结果。须在分派前定义项目资格、阶段门、工程样机单位、通过/取消/未完成及资源计量；不能只纳入最终有完整试验记录的成功项目。

若不同团队之间确有稳定边界，可按团队分派完整工作流，并在产品族、成熟度、验证负荷等层内随机；分派前检查共享工程师、设计库、供应商和评审机制。若此类共享使组间污染不可控，可采用按团队分阶段引入或限定业务单元的比较，同时明确日历趋势、人员经验和供应环境仍会混杂。第一阶段的研究成功

SVSAC | sv-sac.org

标准应是任务登记、版本追溯、独立验收、资源采集和失效记录确实能完整执行；如果这些数据链断裂，就不应跳到一个“提速率”。

25.2 预先区分能改变机制解释的观察

一项有辨识力的现场协议应列出可能反驳每种解释的观察。模型筛选机制预期：在相同需求、阶段门与实体验收下，进入制造/试验的候选更少或单位决策所需高成本测试更少；对被筛除集合的抽样复核若发现大量可行方案，则这一机制被削弱。流程标准化机制预期：版本错配、信息退回和跨职能等待减少，即使工具组与常规组候选数接近；如果返工原因与记录完整度无关，标准化解释就需修订。资源倾斜机制预期：被试点组设备/工程工时增加且等待下降；如果资源相同但队列时长也未变，则资源倾斜不是主要解释。验收口径改变的反证则来自盲评差异、测试缺失或需求门槛版本不一致。

这些预测需要同时检查时间和风险。假如首个通过样机提前，但独立盲评下通过率降低、后续改版增加或测试不再覆盖目标工况，不能把它报告为等效质量下的提速。相反，工具减少低价值试件但导致验证等待日数不变，说明其节省可能落在预算而非日历时间。对小样本，不可因点估计方向符合预期就声称机制成立；要展示区间、项目轨迹和发生在何处的差异，并预先定义在什么数据质量或精度下结论只能是“不确定”。

25.3 再检验外部效度，而非把异质性平均掉

只有当单一族群中的评估协议稳定，才有意义迁移到第二种制造路线或验收结构。应有意选取对机制有对照性的任务：一类是软件搜索可直接改变关键路径、而物理验证较短的项目；另一类是试验、后处理或外部供应占据较长时间的项目；第三类是安全/资格门槛高且候选不能少做验证的项目。若 AI 只在第一类改变日历时间，结论便应限定在其关键路径条件，不应把另外两类的“工具使用”合并平均成行业效应。

外推依赖数据域、表示和部署成熟度。已读范围涉及 CAD 命令生成、概念车轮及铜合金沉积中的不同变量和终点；[2,3,5]金属承载件疲劳文献尚待核，不作为本篇已取得试验。[15] 跨产品族、材料、供应商或求解器版本时，需区分插值与外推，并检验排序、残差和错判是否变化。后续系统综述可依透明检索、双人筛选和预定规则补充范围，但综述本身仍不产生企业级因果效果。

SVSAC | sv-sac.org

25.4 把反证、缺失和资源停止规则写进方案

协议应在首个任务分派前登记主要终点、主要比较、允许的次级分析、观察窗、停止条件和解释规则。主要时间结果可按团队随机分派估计从需求冻结到独立签署的历时；窗口内未通过者不能因研究人员偏好而删去。取消、目标不可比变更、资源暂停和仍在等待应分别编码。若出现测量边界争议，先按预设流程复测或交裁决；不得让 AI 组和对照组在相同争议下采用不同的有利规则。若发生安全边界越界、参考结果与模型严重不一致或关键数据不可追溯，应停止该候选或升级审核，同时保留发生率用于比较。

主分析应遵循团队层原始分派；在实际使用强度、工具版本或“被筛选/进入制造”子组中的估计只作探索性描述，因为这些状态均可能受分派和项目进展影响。若团队间共享知识、候选库、评审资源或供应商窗口，应记录共享方向与时点；必要时缩小随机化单元之外的因果主张。若某一处理组在追踪中获得额外试验台日数，应报告此资源差异，并将“AI 工作流加额外容量”与“相同容量下 AI 工作流”视为不同核心问题。

停止不应由期望的加速幅度决定，而应由安全事件、数据质量失效、资源预算耗尽或预先规定的可行性条件决定。若由于区间过宽而无法判断，报告不确定而非宣布无效；若期望方向与实际方向相反，应如实给出项目轨迹、失败类型和不确定范围。对于小样本次分析，预先写出可能导致“不可解释”的情形，本身比事后寻找显著性更有研究价值。

25.5 研究局限

本篇是定向综述与机制分析，不声称穷尽工程设计 AI 研究。DeepCAD、BEAM、桁架逆设计、翘曲、NIST 项目和 JCGM 的依据限于实际读到的相关原段，若干文献仅有摘要或首面，5 项原文不可读；来源记录保留具体深度及未核数字，元数据不代替主张支持。不同材料小样、算法基准、计量与实体试验不能汇总为通用提速率。本篇没有企业访谈、制造日志、合同或真实项目对照，数值与情境属于构造分析。原文报告比例不一致和关键试验细节缺口仍须保留，不由方法合理性自动消除。

另一个限制是“样机验收”会随产品阶段改变。概念件、工程验证件、型式/资格件和生产代表件所需证据不同，单一时间变量可能掩盖目标差别。我们提出的团队级随机化、版本追溯和不确定性处理是可执行的研究方案，不等于已经在某个企业被验证可行。现实中任务数少、泄漏风险高和资源不可随机都可能限制设

SVSAC | sv-sac.org

计；遇到这些条件，应缩小因果主张而不是用复杂模型弥补设计无法识别的问题。

26 决策建议：AI 辅助设计何时可能缩短获得可验收样机的时间

本篇的核心问题是，AI 辅助设计、仿真与打样在什么条件下能够缩短从需求冻结到首个可验收工程样机的历时。现有证据支持一种有条件的答案：AI 能够加速候选生成、部分预测和设计空间搜索，并在一些研究场景中帮助安排更有信息价值的计算或物理试验；但这类局部性能尚不能单独证明工程样机更早通过同一验收标准。要形成端到端提速，工具须直接减少关键路径处理时间，或经候选/排序改进减少低价值迭代或错误返工，且节省未被下游等待抵消，设计表达须能进入 CAD/CAE 和实际制造，模型/测量不确定性须经相称而独立的证据约束，且加工、审查、供应和测试资源不能成为被新增候选挤压的瓶颈。

这项判断也回应了以生成量或单任务加速作为主指标的不足。多生成方案、求解更快、或少做几次试验都不是最终研发结果；固定分母的通过事件、全体资源账本和后续修改共同决定“更早而且仍合格”是否成立。错误接受和错误拒绝的代价不同，证据再利用也取决于变更是否保留了原有失效机制和验证适用域。对高风险或接近验收边界的样机，模型概率不能代替工程责任和规定的独立检验。

因此，实践中的有限启示是先明确任务终点和工作流，再试点改变其中确有数据或分析瓶颈的环节；同等保留需求、验收、版本和失败证据，并预先规定哪些结果会反驳提速判断。若工具只减少计算时长而队列、制造或验收不变，收益可能体现为成本或探索能力，未必是日历时间；若候选增加导致审核和实物验证拥堵，项目周期甚至可能延长。本篇没有证据给出通用提速百分比、投资回报率或行业门槛，也不保证某种模型对任何企业有效。

结论应限于本篇可见证据和待检验机制。未来只有在可比较任务、独立实体验收、完整过程日志和与分派单位一致的推断下，才能回答某个 AI 工作流是否更快地产生了同等或更强证据的样机，以及这种作用是否会迁移到其他产品族。若这些条件尚未满足，审慎的结论不是“AI 无效”，而是端到端周期效果仍未识别。

SVSAC | sv-sac.org

给企业的下一步建议

AI 打样提速的竞争价值，在于更快取得能被客户和工程端共同验收的样机，而不是更快生成图纸。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：拆分需求澄清、设计、仿真、加工与验证的等待时间，先处理关键瓶颈，再引入适配的 AI 工具。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，进一步形成开发流程诊断、仿真与实物验证安排、样机提速路线图。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。

工程开发中“更快”至少有四种不同含义：设计候选生成耗时更短、单次仿真运行更快、通过验证的候选更早出现，以及从需求冻结到工程样机验收的总历时缩短。前两者是局部计算结果，后两者才直接对应研发周期；它们并不自动相等。本篇研究的问题是：AI 辅助设计、代理模型与自适应试验在什么条件下能够缩短“获得可验收工程样机”的历时，又在什么条件下只是把排队、返工和不确定性迁移到网格收敛、模型复核、加工或物理测试环节？本篇采取定向范围证据梳理与工程机制分析，不是系统综述、企业田野调查或随机对照试验。我们比较生成式 CAD 与逆向设计、机器学习代理、仿真驱动优化和物理试验闭环四类任务，分开讨论几何表达、预测误差、目标附近的约束可行性、制造误差和验收等待。已核阅的研究显示，模型可在训练后快速生成候选，或以学习所得近似代替部分高成本计算；但这类证据主要支持搜索或评估子任务的计算效率，不能单凭候选数或推理时间推断实体样机已经通过强度、疲劳、尺寸、表面及装配要求。本篇由此提出一条可检验的中间链：只有当工具减少关键路径处理时间、无信息实物迭代或错误返工，这种节省未被下游等待抵消，且自动生成的几何可进入可追溯的 CAD/CAE 及制造流程、模型不确定性经独立验证、加工与测试队列不成为新瓶颈时，局部提速才有机会转化为端到端历时缩短。该机制及边界均待现场数据检验。

SVSAC | sv-sac.org

参考文献

[1] Lyle Regenwetter; Amin Heyrani Nobari; Faez Ahmed. Deep Generative Models in Engineering Design: A Review. Journal of Mechanical Design, 144(7), 071704 (2022). https://dspace.mit.edu/bitstream/handle/1721.1/150666/2110.10863.pdf?isAllowed=y&sequence=2

[2] Rundi Wu; Chang Xiao; Changxi Zheng. DeepCAD: A Deep Generative Network for Computer-Aided Design Models. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6752–6762 (2021). https://arxiv.org/html/2105.09492v2

[3] Soyoung Yoo; Sunghee Lee; Seongsin Kim; Kwang Hyeon Hwang; Jong Ho Park; Namwoo Kang. Integrating Deep Learning into CAD/CAE System: Generative Design and Evaluation of 3D Conceptual Wheel. Structural and Multidisciplinary Optimization, 64, 2725–2747 (2021). https://arxiv.org/pdf/2006.02138

[4] Jan-Hendrik Bastek; Siddhant Kumar; Bastian Telgen; Raphaël N. Glaesener; Dennis M. Kochmann. Inverting the structure–property map of truss metamaterials by deep learning. Proceedings of the National Academy of Sciences, 119(1), e2111505119 (2022). https://doi.org/10.1073/pnas.2111505119

[5] Azza Fadhel; Nathaniel W. Zuckschwerdt; Aryan Deshwal; Susmita Bose; Amit Bandyopadhyay; Jana Doppa. Discovery of Feasible 3D Printing Configurations for Metal Alloys via AI-Driven Adaptive Experimental Design. Proceedings of the AAAI Conference on Artificial Intelligence, 40(47), 39939–39948 (2026). https://arxiv.org/html/2601.17587v1

[6] Aditi S. Krishnapriyan; Amir Gholami; Shandian Zhe; Robert M. Kirby; Michael W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. Advances in Neural Information Processing Systems, 34 (2021). https://arxiv.org/pdf/2109.01050

[7] Aldair E. Gongora; Bowen Xu; Wyatt Perry; Chika Okoye; Patrick Riley; Kristofer G. Reyes; Elise F. Morgan; Keith A. Brown. A Bayesian experimental autonomous researcher for mechanical design. Science Advances, 6(15), eaaz1708 (2020). https://research.google/pubs/a-bayesian-experimental-autonomous-researcher-for-mechanical-design/

[8] National Institute of Standards and Technology. Metrology for AM Model Validation. NIST research project/program page, updated 29 April 2026 (2026). https://www.nist.gov/programs-projects/metrology-am-model-validation

[9] Ari Seff; Yaniv Ovadia; Wenda Zhou; Ryan P. Adams. SketchGraphs: A Large-Scale Dataset for Modeling Relational Geometry in Computer-Aided Design. arXiv preprint, arXiv:2007.08506 (2020). https://arxiv.org/abs/2007.08506

[10] Onyekachi Okorie; Asma Perveen; Didier Talamona; Konstantinos Kostas. Topology Optimization of an Aerospace Bracket: Numerical and Experimental Investigation. Applied Sciences, 13(24), article 13218 (2023). https://www.mdpi.com/2076-3417/13/24/13218

[11] Jose Pablo Folch; Robert M. Lee; Behrang Shafei; David Walz; Calvin Tsay; Mark van der Wilk; Ruth Misener. Combining multi-fidelity modelling and asynchronous batch Bayesian

SVSAC | sv-sac.org

Optimization. Computers & Chemical Engineering, 172, article 108194 (2023). https://www.sciencedirect.com/science/article/pii/S0098135423000637

[12] Haesook T. Kim. Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis. Clinical Cancer Research, 13(2 Pt 1), 559–565 (2007). https://aacrjournals.org/clincancerres/article/13/2/559/178178/Cumulative-Incidence-in-Competing-Risks-Data-and

[13] Patrick Royston; Mahesh K. B. Parmar. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Medical Research Methodology, 13, article 152 (2013). https://link.springer.com/article/10.1186/1471-2288-13-152

[14] Antoine Dupuis; Jean-Jacques Pesce; Jean-Baptiste Marijon; Stéphane Roux; Gilles Régnier. Use of a digital image correlation method for full-field shrinkage measurement in injection molding. The Journal of Strain Analysis for Engineering Design, 57(8), 702–713 (first published online 2 December 2021) (2022). https://journals.sagepub.com/doi/10.1177/03093247211060215

[15] Yisheng Chen; Qianglong Wang; Chong Wang; Peng Gong; Yincheng Shi; Yi Yu; Zhenyu Liu. Topology Optimization Design and Experimental Research of a 3D-Printed Metal Aerospace Bracket Considering Fatigue Performance. Applied Sciences, 11(15), article 6671 (2021). https://www.mdpi.com/2076-3417/11/15/6671

[16] Francisco Yumbla; June-Sup Yi; Meseret Abayebas; Mahir Shafiyev; Hyungpil Moon. Tolerance dataset: mating process of plug-in cable connectors for wire harness assembly tasks. Intelligent Service Robotics, 13(2) (2020). https://link.springer.com/article/10.1007/s11370-019-00307-5

[17] Jian Wang; Tao Liu; Kaihuang Zheng; Hao Liu; Hongdao Cui; Hang Li. Local thermal warpage deformation of polypropylene injection molded flat part and neural network prediction model. Frontiers in Materials, 11, article 1421546 (2024). https://www.frontiersin.org/journals/materials/articles/10.3389/fmats.2024.1421546/full

[18] Joint Committee for Guides in Metrology (JCGM). Evaluation of measurement data — The role of measurement uncertainty in conformity assessment. JCGM 106:2012 (2012). https://www.bipm.org/documents/20126/2071204/JCGM_106_2012_E.pdf

[19] Marc C. Kennedy; Anthony O'Hagan. Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), 425–464; first published online 6 January 2002 (2001). https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00294

SVSAC | sv-sac.org