

《人工智能报告 2025》

2025年商业AI现状： GenAI鸿沟

MIT NANDA

Aditya Challapally

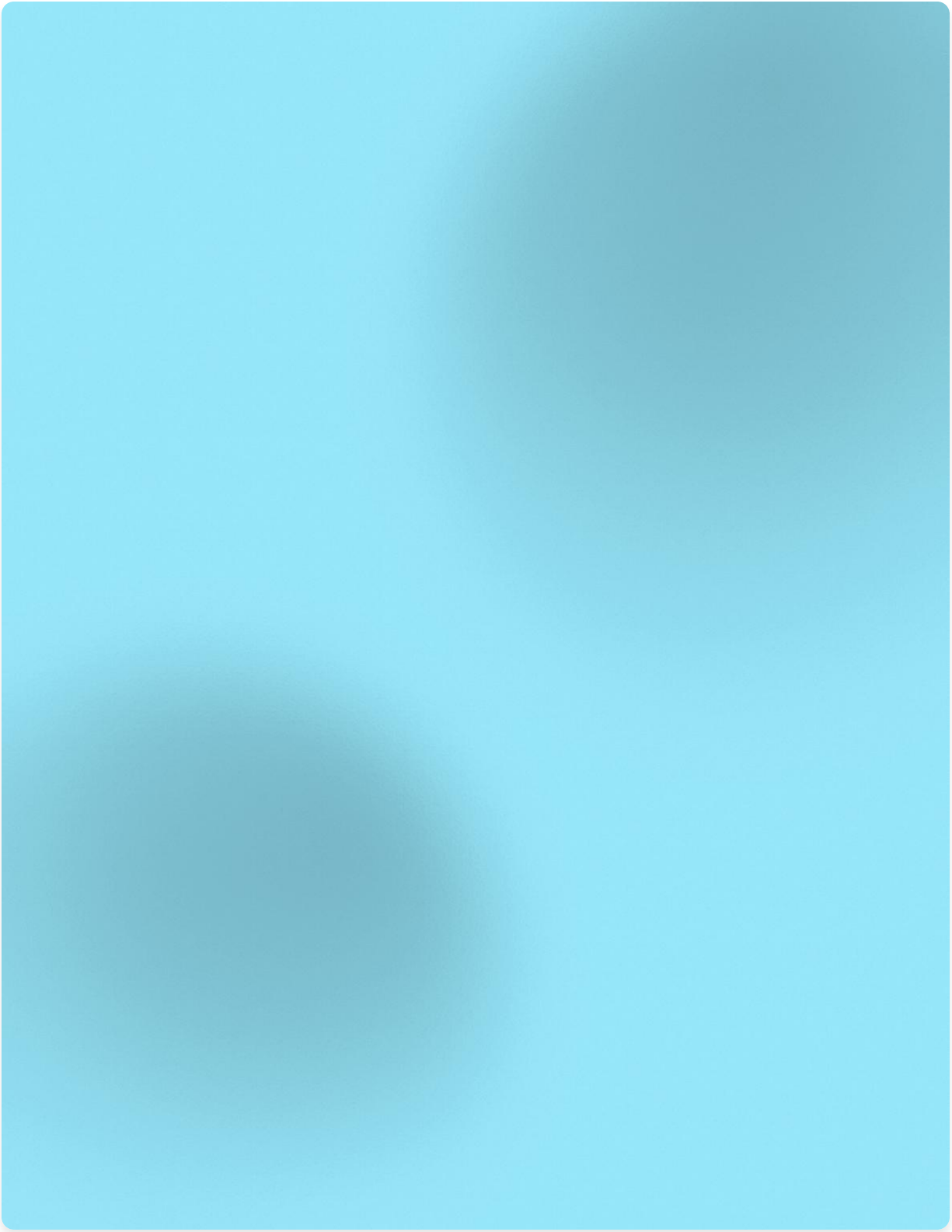
Chris Pease

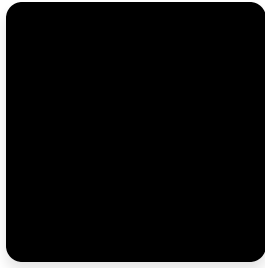
Ramesh Raskar

Pradyumna Chari

2025年7月







说明

来自NANDA项目AI实施研究的初步发现

评审人员： Pradyumna Chari, NANDA项目

研究期间： 2025年1月——6月

研究方法： 本报告基于多方法研究设计，包括对300多项公开披露的AI计划的系统性回顾，与52个组织代表的结构化访谈，以及在四个主要行业会议上收集的153位高级领导者的调研回复。

免责声明： 本报告中表达的观点仅代表作者和评审人员的个人观点，不代表任何关联雇主的立场。

保密说明： 所有公司特定数据和引用均已匿名化，以保持符合企业披露政策和保密协议，确保中立性，并防止任何商业推进或意见偏见的感知。





目录

1. 执行摘要
2. GenAI鸿沟的错误一侧：高采用率，低转型效果
3. 试点停滞的原因：鸿沟背后的学习差距
4. 跨越GenAI鸿沟：最佳建设者的成功之道
5. 跨越GenAI鸿沟：最佳买家的成功之道
6. 结论：弥合GenAI鸿沟

执行摘要

尽管企业在GenAI方面投资了300-400亿美元，但本报告发现了一个令人惊讶的结果：95%的组织获得了零回报。买家（企业、中端市场、中小企业）和建设者（初创公司、供应商、咨询公司）的结果差异如此巨大，以至于我们称之为GenAI鸿沟。只有5%的集成AI试点项目提取了数百万价值，而绝大多数仍然停滞不前，没有可衡量的损益影响。这种鸿沟似乎不是由模型质量或监管驱动的，而是由方法决定的。

像ChatGPT和Copilot这样的工具被广泛采用。超过80%的组织已经探索或试点了它们，近40%报告了部署情况。但这些工具主要是增强个人生产力，而不是损益绩效。与此同时，企业级系统，无论是定制的还是供应商销售的，都在悄悄被拒绝。60%的组织评估了此类工具，但只有20%达到了试点阶段，仅有5%达到了生产阶段。大多数失败是由于工作流程脆弱、缺乏上下文学习以及与日常运营不匹配。

从我们的访谈、调研和对300个公共实施案例的分析中，出现了四种定义GenAI鸿沟的模式：

- **有限颠覆**：8个主要行业中只有2个显示出有意义的结构性变化
- **企业悖论**：大公司在试点数量方面领先，但在扩展方面滞后
- **投资偏见**：预算倾向于可见的、营收线功能，而不是高ROI的后台办公
- **实施优势**：外部合作伙伴关系的成功率是内部构建的两倍

扩展的核心障碍不是基础设施、监管或人才，而是学习。大多数GenAI系统不保留反馈、不适应上下文或不随时间改进。

一小群供应商和买家通过直接解决这些限制来实现更快的进展。成功的买家要求流程特定的定制化，并基于业务结果而非软件基准来评估工具。他们期望系统与现有流程集成并随时间改进。满足这些期望的供应商在几个月内就能获得数百万美元的部署。



虽然大多数实施不会推动员工数量减少，但已经跨越GenAI鸿沟的组织开始在客户支持、软件工程和管理功能方面看到选择性的劳动力影响。此外，表现最佳的组织报告从减少BPO支出和外部代理使用中获得可衡量的节省，特别是在后台办公运营中。其他组织引用通过自动化外联和智能跟进系统改善了客户保留和销售转化。这些早期结果表明，具有学习能力的系统，当针对特定流程时，即使没有重大组织重构也能提供真正的价值。

GenAI鸿沟的错误一侧：高采用率，低转型效果

要点： 大多数组织处于GenAI鸿沟的错误一侧，采用率很高，但颠覆性很低。九个行业中有七个显示出很少的结构性变化。企业正在试点GenAI工具，但很少达到部署阶段。像ChatGPT这样的通用工具被广泛使用，但定制解决方案由于集成复杂性和与现有工作流程不匹配而停滞。

当检查行业级转型模式时，GenAI鸿沟最为明显。尽管有高调的投资和广泛的试点活动，但只有一小部分组织已经超越实验阶段，实现了有意义的业务转型。

鸿沟背后的颠覆现实

要点： GenAI鸿沟在行业层面清楚地表现出来，尽管GenAI具有可见性，但只有两个行业（技术和媒体）显示出结构性颠覆的明确迹象，而其他七个行业仍然处于转型的错误一侧。

尽管有高调的投资，行业层面的转型仍然有限。GenAI已被嵌入到支持、内容创建和分析用例中，但很少有行业显示出与过去通用技术相关的深层结构性转变，如新的市场领导者、被颠覆的商业模式或可衡量的客户行为变化。

为了更好地量化颠覆状态，我们开发了一个综合的AI市场颠覆指数。每个行业根据五个可观察指标被评分为0到5分。这些分数代表五个维度的标准化平均值，通过公开指标和访谈评估进行三角测量。我们测试了替代权重方案以确认行业排名的一致性：

[1.] 顶级现有企业间的市场份额波动性（2022年至2025年）



[2.] 2020年后成立的AI原生企业的收入增长

[3.] 新AI驱动商业模式的出现

[4.] 归因于GenAI的用户行为变化

[5.] 归因于AI工具的高管组织变更频率

展示：GenAI颠覆在不同行业间差异显著

[能源与材料] [0]

[先进制造业] [0.5]

[金融服务] [0.5]

[消费与零售] [0.5]

[医疗与制药] [0.5]

[专业服务] [1.5]

[媒体与电信] [2]

[0] [0.5] [1] [1.5] [2] [2.5] [3] [3.5] [4]

展示：GenAI颠覆描述

行业 关键信号

技术: 新挑战者获得优势（如Cursor vs Copilot）； 工作流程转变

媒体与电信: AI原生内容兴起； 广告动态转变； 现有企业仍在增长

专业服务: 效率提升； 客户交付基本保持不变

医疗与制药: [文档/转录试点； 临床模式未变]

消费与零售: 支持自动化； 对忠诚度或领导者影响有限

金融服务: 后端自动化； 客户关系稳定

先进制造业: 维护试点； 供应链无重大转变

能源与材料: 近零采用； 最少实验

敏感性分析： 我们测试了五个颠覆指标的替代权重。技术和媒体与电信在所有合理的权重方案中都保持顶级排名，而医疗和能源始终排名较低。专业服务对权重变化最敏感，根据对效率提升与结构性变化的重视程度不同，范围从1.2到2.1。



九个主要行业中有七个显示出显著的试点活动，但几乎没有结构性变化。投资与颠覆之间的这一差距直接展示了大规模的GenAI鸿沟，即广泛的实验而非转型。

受访者在评估中非常直率。一位中等市场制造业COO总结了普遍情绪：

“LinkedIn上的炒作说一切都已改变，但在我们的运营中，没有根本性转变。我们处理某些合同的速度更快了，但这就是全部的变化。”

3.2 试点到生产的鸿沟

要点： GenAI鸿沟在部署率上最为明显，只有5%的定制企业AI工具达到生产环境。聊天机器人成功是因为它们易于尝试且灵活，但由于缺乏记忆和定制化而在关键工作流程中失败。这一根本差距解释了为什么大多数组织仍处在鸿沟的错误一侧。

我们的研究揭示了GenAI采用工具调查、试点与实际实施之间的急剧下降，通用和定制解决方案之间存在显著差异。

研究局限性： 这些数据基于个人访谈而非官方公司报告，在方向上是准确的。各类别的样本量不同，成功定义在不同组织间可能有所差异。

展示： 任务特定GenAI工具从试点到生产的急剧下降揭示了GenAI鸿沟

[通用LLMs] [嵌入式或任务特定GenAI]

[80%]

[60%]

[50%]

[40%]

[20%] [5%]

[调研] [试点] [成功实施]

研究说明：我们将任务特定GenAI工具的成功实施定义为用户或高管认为对生产力和/或损益产生显著和持续影响的工具



企业AI解决方案95%的失败率代表了GenAI鸿沟最清晰的表现。困在错误一侧的组织继续投资无法适应其工作流程的静态工具，而那些跨越鸿沟的组织专注于具有学习能力的系统。

通用LLM聊天机器人似乎显示出较高的试点到实施率（约83%）。然而，这掩盖了感知价值的更深层分歧，并揭示了为什么大多数组织仍然困在鸿沟的错误一侧。

在访谈中，企业用户报告对ChatGPT和Copilot等消费级工具持续积极的体验。这些系统因其灵活性、熟悉度和即时效用而受到赞扬。然而，同样的用户对定制或供应商推介的AI工具绝大多数持怀疑态度，将它们描述为脆弱、过度工程化或与实际工作流程不匹配。

正如一位CIO所说：“我们今年看了几十个演示。也许其中一两个真正有用。其余的都是包装器或科学项目。”

虽然热情和预算通常足以启动试点，但将这些转化为具有持续价值的工作流程集成系统仍然很少见，这一模式定义了处在GenAI鸿沟错误一侧的组织的体验。

企业（这里定义为年收入超过1亿美元的公司）在试点数量上领先，并分配更多员工到AI相关项目。然而，这种强度并未转化为成功。这些组织报告的试点到规模化转换率最低。

相比之下，中型公司行动更快、更果断。顶级表现者报告的平均时间线是从试点到全面实施90天。而企业相比之下需要九个月或更长时间。

企业GenAI的五个误区

1. AI将在未来几年取代大多数工作 → 研究发现GenAI导致的裁员有限，且仅发生在已经受到AI显著影响的行业。高管们对未来3-5年的招聘水平没有达成共识。
2. 生成式AI正在改变业务 → 采用率很高，但转型很少见。只有5%的企业将AI工具大规模集成到工作流程中，9个行业中有7个没有显示出真正的结构性变化。
3. 企业在采用新技术方面很慢 → 企业极其渴望采用AI，90%的企业已经认真探索过购买AI解决方案。



4. 阻碍AI发展的最大因素是模型质量、法律、数据、风险 → 真正阻碍它的是大多数AI工具不会学习且无法很好地集成到工作流程中。
5. 最好的企业正在构建自己的工具 → 内部构建的失败率是外部购买的两倍。

影子AI经济：跨越鸿沟的桥梁

要点： 虽然官方企业计划仍停留在GenAI鸿沟的错误一边，员工已经通过个人AI工具跨越了这一鸿沟。这种”影子AI”通常比正式计划带来更好的投资回报率，并揭示了真正有效跨越鸿沟的方法。

在令人失望的企业部署数据背后隐藏着一个令人惊讶的现实：AI已经在改变工作方式，只是不是通过官方渠道。我们的研究发现了一个蓬勃发展的”影子AI经济”，员工使用个人ChatGPT账户、Claude订阅和其他消费者工具来自动化工作的重要部分，通常没有IT部门的知情或批准。

规模令人瞩目。虽然只有40%的公司表示购买了官方LLM订阅，但我们调查的公司中超过90%的员工报告定期使用个人AI工具处理工作任务。实际上，几乎每个人都以某种形式使用LLM进行工作。

图表：影子AI经济，员工使用远超官方采用

[已购买LLM订阅的公司，40%]

[定期使用LLM的员工，90%]

在许多情况下，影子AI用户报告每天多次使用LLM处理每周工作负载，而他们公司的官方AI计划仍停留在试点阶段。

这种影子经济表明，当获得灵活、响应迅速的工具时，个人可以成功跨越GenAI鸿沟。那些认识到这种模式并在此基础上构建的组织代表了企业AI采用的未来。

前瞻性组织开始通过学习影子使用情况、分析哪些个人工具能带来价值，然后采购企业替代方案来弥合这一差距。



投资模式反映了鸿沟

要点： 投资分配揭示了GenAI鸿沟的作用，50%的GenAI预算流向销售和营销，但后台自动化通常产生更好的投资回报率。这种偏向反映了更容易的指标归因，而非实际价值，使组织专注于错误的优先事项。

在功能重点方面，GenAI工具的投资高度集中。由于GenAI支出尚未在各组织间正式量化，我们要求高管将假设的100美元分配给不同功能。在我们的调查中，销售和营销功能获得了约70%的AI预算分配。

图表：按功能划分的GenAI投资分布

[销售与营销] [运营]

[内部工作流程协调] [文档摘要]

[智能线索评分] [AI生成的外向邮件]

[流程合规监控] [动态资源分配]

[客户服务] [财务与采购]

[合同分类和标记] [通话摘要]

[活动个性化内容] [跟进自动化] [呼叫路由]

[基于AI的竞争对手分析] [社交情感分析] [AI驱动的聊天机器人] [智能工单路由] [供应商风险警报] [AP/AR自动化]

研究说明： 虽然GenAI投资的一般功能分配（如约50%用于销售与营销）在高管访谈中相对一致，但子类别和用例细分应视为方向性指导。子类别反映综合说明和轶事模式，而非精确核算。公司类型驱动显著差异。例如，制造商和医疗保健提供者通常对销售与营销的投资最少，而在运营方面投资过度。科技和媒体公司通常优先考虑营销、内容和开发人员生产力。专业服务倾向于文档自动化和法律/合规工具。



销售和营销占主导地位不仅因为可见性，还因为结果容易衡量。演示量或电子邮件响应时间等指标直接与董事会级KPI一致。

相比之下，法律、采购和财务功能提供更微妙的效率。这些包括更少的合规违规、简化的工作流程或加速的月末流程，虽然重要但难以在高管对话或投资者更新中体现。

一家财富1000强制药公司的采购副总裁清楚地表达了这一挑战：

“如果我购买一个工具来帮助我的团队提高工作效率，我如何量化这种影响？当它不会直接推动收入或降低可衡量的成本时，我如何向我的CEO证明这一点？我可以论证它帮助我们的科学家更快地获得工具，但这与底线影响相差了好几个层次。”

这种投资偏见通过将资源导向可见但往往不太具有变革性的用例来延续GenAI鸿沟，而后台职能中ROI最高的机会仍然资金不足。

除了测量挑战，信任和社会认同在购买决策中仍然具有决定性作用。一家大型CPG公司的采购主管捕捉到了许多买家面临的困境：

“我每天收到无数声称提供最佳GenAI解决方案的邮件。有些有令人印象深刻的演示，但建立信任才是真正的挑战。由于有如此多的选择涌入我们的收件箱，我们很大程度上依赖同行推荐和我们网络中的推荐。”

这突出了一个更广泛的模式：仅有产品质量很少足够。推荐、先前的关系和VC介绍仍然是企业采用的更强预测因子，而非功能或特性集。

为什么试点项目停滞：GenAI鸿沟背后的学习差距

让组织停留在GenAI鸿沟错误一侧的主要因素是学习差距，即不会学习、集成不良或不匹配工作流程的工具。用户更喜欢ChatGPT来处理简单任务，但由于缺乏记忆而在关键任务工作中放弃使用。缺少的是能够适应、记忆和演进的系统，这些能力定义了鸿沟两侧之间的差异。



让组织陷入困境的障碍

要点： 顶级障碍反映了定义GenAI鸿沟的基本学习差距：用户抗拒不会适应的工具，没有上下文的模型质量会失败，当系统无法记忆时用户体验会受损。即使是狂热的ChatGPT用户也不信任不符合其期望的内部GenAI工具。

为了理解为什么如此少的GenAI试点项目能够超越实验阶段，我们对52个组织的执行发起人和一线用户进行了调研。参与者被要求在1-10频率量表上评估扩展的常见障碍，其中10代表最常遇到的障碍。

结果显示了一个可预测的领导者：抗拒采用新工具。然而，第二高的障碍被证明比预期的更重要。

图表： 为什么GenAI试点项目失败：企业中AI扩展的主要障碍 用户被要求在1-10量表上评估每个问题

变更管理挑战

缺乏执行赞助

用户体验差

模型输出质量担忧

不愿采用新工具

研究说明： 这些分数反映的是报告频率而非障碍影响的客观测量，可能因行业和组织规模而显著不同。

模型质量担忧的突出地位最初显得反直觉。ChatGPT和类似工具的消费者采用激增，超过40%的知识工作者个人使用AI工具。然而，将这些工具整合到个人工作流程中的相同用户在企业系统中遇到它们时却将其描述为不可靠。这个悖论说明了用户层面的GenAI鸿沟。

这种偏好揭示了一个根本性紧张关系。每天使用ChatGPT处理个人任务的相同专业人士要求企业工作具有学习和记忆能力。相当数量的工作者已经私下使用AI工具，报告了生产力提升，而他们公司的正式AI举措却停滞不前。这种影子使用创造了一个反馈循环：员工知道好的AI感觉如何，使他们对静态企业工具的容忍度降低。



为什么通用工具既赢又输

要点： GenAI鸿沟在用户偏好中显现：ChatGPT击败企业工具是因为它更好、更快、更熟悉，即使两者使用相似的模型。但这种同样的偏好揭示了为什么组织仍然困在鸿沟的错误一侧。

我们的后续访谈揭示了一个惊人的矛盾。对企业AI工具表示怀疑的专业人士往往是消费者LLM界面的重度用户。当被要求比较他们的经验时，出现了三个一致的主题。

用户偏好驱动因素：通用LLM界面 vs 集成工具

“更信任它”

“已经熟悉界面”

“答案更好”

一位中等规模公司的企业律师典型地体现了这种动态。她的组织在专业合同分析工具上投资了50,000美元，但她一直默认使用ChatGPT进行起草工作：

“我们购买的AI工具提供了有限定制选项的僵化摘要。使用ChatGPT，我可以引导对话并迭代直到得到我确切需要的内容。基本质量差异是明显的，ChatGPT始终产生更好的输出，即使我们的供应商声称使用相同的基础技术。”

这种模式表明，一个每月20美元的通用工具在即时可用性和用户满意度方面，往往优于成本高出数个数量级的定制企业系统。这个悖论体现了为什么大多数组织仍然停留在GenAI鸿沟的错误一侧。

4.3 定义鸿沟的学习差距

要点： ChatGPT的局限性恰恰揭示了GenAI鸿沟背后的核心问题：它会遗忘上下文、无法学习、不能进化。在关键任务工作中，90%的用户更偏好人类。这种差距是结构性的，GenAI缺乏记忆和适应性。

鉴于用户对消费级LLM界面的偏好，我们调查了阻碍其在关键任务工作中更广泛采用的因素。这里的障碍与一般可用性问题截然不同，直接揭示了定义GenAI鸿沟的学习差距。

图表：核心工作流程集成的障碍



“在边缘情况下出错且不适应。”

“无法定制以适应我们的特定工作流。”

“每次都需要太多手动上下文。”

“它不从我们的反馈中学习。”

同一位偏爱ChatGPT进行初稿写作的律师在涉及敏感合同时划下了明确界限：

“它在头脑风暴和初稿方面表现出色，但它不保留客户偏好知识或从以前的编辑中学习。它重复同样的错误，每次会话都需要大量的上下文输入。对于高风险工作，我需要一个能够积累知识并随时间改进的系统。”

这个反馈指向了让组织处于GenAI鸿沟错误一方的根本学习差距。用户赞赏消费级LLM界面的灵活性和响应性，但需要当前工具无法提供的持久性和上下文感知。

当我们要求企业用户对高风险工作的不同选项进行评级时，偏好层次变得清晰：

图表：高风险工作的感知适用性

“你会将此任务分配给AI还是初级同事？”

复杂项目（多周工作，客户管理）：AI偏好 10%，人类偏好 90%

快速任务（邮件、摘要、基础分析）：AI偏好 70%，人类偏好 30%

结果显示，AI已经赢得了简单工作的战争，70%的人偏好AI来起草邮件，65%用于基础分析。但对于任何复杂或长期的工作，人类以9比1的优势占主导地位。分界线不是智能，而是记忆、适应性和学习能力——这些正是区分GenAI鸿沟两侧的特征。



智能体AI(Agentic AI)，这类按设计嵌入持久记忆和迭代学习的系统，直接解决了定义GenAI鸿沟的学习差距。与每次都需要完整上下文的当前系统不同，智能体系统保持持久记忆，从交互中学习，并能自主协调复杂的工作流。早期企业实验中的客户服务智能体可端到端处理完整询问，金融处理智能体监控并批准常规交易，销售管道智能体跨渠道追踪参与度，这些展示了自主性和记忆如何解决企业识别的核心差距。

图表：按定制化和学习能力定位GenAI工具

	低记忆/学习	高记忆/学习
低定制化	Copilot, GPT包装器	ChatGPT带记忆功能(测试版)
高定制化	内部构建(脆弱)	智能体工作流，垂直SaaS

5 跨越GenAI鸿沟：最佳构建者如何成功

处于GenAI鸿沟正确一方的组织有一个共同方法：他们构建从反馈中学习的适应性嵌入式系统。跨越鸿沟的最佳初创公司专注于狭窄但高价值的用例，深度集成到工作流中，并通过持续学习而非广泛功能集进行扩展。领域流利度和工作流集成比炫目的UX更重要。

在我们的访谈中，我们观察到GenAI初创公司之间日益增长的分化。一些公司正在与过时的SaaS剧本斗争，仍被困在鸿沟的错误一方，而其他公司正通过积极的定制化和与真实业务痛点的对齐来获得企业关注。

对GenAI工具的需求仍然很高。几家初创公司报告在几天内签署试点项目，随后不久就达到七位数的收入运行率。突出的表现者不是那些构建通用工具的公司，而是那些将自己嵌入工作流、适应上下文并从狭窄但高价值立足点扩展的公司。

我们的数据揭示了一个清晰的模式：成功的组织和供应商是那些积极解决学习、记忆和工作流适应问题的，而失败的是那些构建通用工具或试图内部开发能力的。

成功的初创公司构建从反馈中学习的系统(66%的高管需要这个)，保留上下文(63%要求这个)，并深度定制到特定工作流。他们从具有显著定制化的工作流边缘开始，然后扩展到核心流程。



5.1 企业实际需要什么：跨越鸿沟的桥梁

最成功的供应商明白，跨越GenAI鸿沟需要构建高管们反复强调的系统：不仅生成内容，还能在其环境中学习和改进的AI系统。

在评估AI工具时，买家一致强调一套特定的优先事项。我们对这些主题进行了访谈编码，以量化它们在采购决策中出现的频率：

图表：高管如何选择GenAI供应商

来自访谈并按类别编码

- 变化时的灵活性
- 随时间改进的能力
- 清晰的数据边界
- 对当前工具的最小干扰

深入了解我们的工作流程

我们信任的供应商

[0%] [10%] [20%] [30%] [40%] [50%] [60%] [70%] [80%] [90%] [100%]

展示：高管选择GenAI供应商的直接引用 来自访谈并按类别编码

他们想要什么 直接引用

我们（他们）信任的供应商 [“我们更愿意等待现有合作伙伴添加AI，而不是赌一个初创公司。”]

深入了解我们的工作流程 [“大多数供应商不了解我们的审批流程或数据流是如何工作的”]

对当前工具的最小干扰 [“如果它不能插入Salesforce或我们的内部系统，没人会使用它。”]

清晰的数据边界 [“我不能冒险让客户数据与别人的模型混合，即使供应商说没问题。”]

随时间改进的能力 [“第一周很有用，但之后它只是重复同样的错误。我为什么要使用那个？”]



事情变化时的灵活性 [“我们的流程每季度都在发展。如果AI不能适应，我们就回到电子表格。”]

对劳动力影响的担忧远少于预期。大多数用户欢迎自动化，特别是对于繁琐的手动任务，只要数据保持安全且结果可衡量。

尽管传统观点认为企业抗拒训练AI系统，但我们访谈中的大多数团队表示愿意这样做，前提是收益明确且有防护措施到位。

尽管对AI感兴趣，但对新兴供应商存在显著的怀疑态度，特别是在高信任度或受监管的工作流程中。许多采购负责人告诉我们，无论创新如何，他们都会忽略大多数初创公司的推销。

“我们每天收到数十个关于AI驱动采购工具的推销。然而，我们既定的BPO合作伙伴已经了解我们的政策和流程。我们更愿意等待他们的AI增强版本，而不是切换到未知的供应商。”，全球消费品公司采购主管

5.2 跨越鸿沟的制胜手册

要点： 成功跨越GenAI鸿沟的初创公司在狭窄的工作流程中取得小而显著的胜利，然后扩展。设置负担低且价值实现时间快的工具胜过重型企业构建。渠道推荐和同行信任是跨越鸿沟的关键增长杠杆。

最成功的初创公司通过执行两种策略来解决对学习系统的渴望和对新工具的怀疑：

5.2.1 为特定工作流程定制

嵌入非关键或相邻流程并进行大量定制，展示明确价值，然后扩展到核心工作流程至关重要。成功的工具具有两个特征：低配置负担和立即可见的价值。相比之下，需要大量企业定制的工具经常在试点阶段停滞。

我们样本中的**成功类别**包括：

- 用于通话摘要和路由的语音AI
- 用于合同和表单的文档自动化
- 用于重复性工程任务的代码生成

困难类别通常涉及复杂的内部逻辑、不透明的决策支持，或基于专有启发式的优化。这些工具由于深度的企业特殊性而经常遇到采用摩擦。

一些初创公司通过主导小而关键的工作流程（特别是在销售和营销方面）然后扩展而表现出色。顶级四分位数的GenAI初创公司在推出后6-12个月内达到120万美元的年化收入。



狭窄范围 vs 广泛范围

简单执行 快速胜利（支出分类、合同审查） 部分试点（供应商风险监控）

复杂执行 早期试点（谈判机器人） 失败（完整采购编排）

5.2.2 利用推荐网络

为了克服信任障碍，成功的初创公司经常使用与系统集成商的渠道合作伙伴关系、来自董事会成员或顾问的采购推荐，以及通过熟悉的企业市场进行分发。

展示：领导者如何发现GenAI解决方案

供应商关系 同行网络

非正式同行推荐，13% 董事会成员或顾问推荐，10%

活动和媒体 内部流程

现有供应商合作伙伴关系，20%

会议演示或小组讨论，9%

冷接触 其他

新集成/合作伙伴推荐，15% 行业出版物或网络研讨会，6%



5.3 跨越鸿沟的窗口正在缩小

要点： 跨越 GenAI 鸿沟的窗口正在迅速关闭。企业正在锁定具有学习能力的工具。Agentic AI 和记忆框架（如 NANDA 和 MCP）将决定哪些供应商帮助组织跨越鸿沟，而不是困在错误的一边。

企业越来越要求随时间适应的系统。Microsoft 365 Copilot 和 Dynamics 365 正在融入持久记忆和反馈循环。OpenAI 的 ChatGPT 记忆 beta 版在通用工具中发出类似的期望信号。

通过构建从反馈、使用和结果中学习的自适应 agent，快速行动以缩小这一差距的初创公司，可以通过数据和集成深度建立持久的产品护城河。这样做的窗口很窄。在许多垂直领域，试点已经在进行中。

支持这一转变的基础设施正通过 Model Context Protocol (MCP)、Agent-to-Agent (A2A) 和 NANDA 等框架逐步建立，这些框架实现了 agent 之间的互操作性和协调。这些协议通过允许专业化 agent 协同工作，而不是要求整体式系统，创造了市场竞争和成本效率。这些框架构成了新兴 Agentic Web 的基础，这是一个由可互操作的 agent 和协议组成的网络，用动态协调层替代了整体式应用程序。

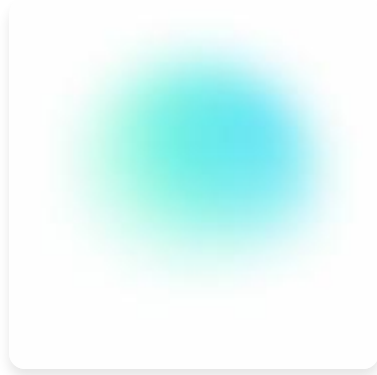
在接下来的几个季度中，一些企业将锁定供应商关系，这些关系几乎不可能解除。这个 18 个月的时间框架反映了我们采访的 17 位采购和 IT 外包负责人的共识，并得到公共采购披露分析的支持，显示企业从 RFP 到实施的周期范围从 2 个月到 18 个月。投资于从其数据、工作流程和反馈中学习的 AI 系统的组织正在创造逐月累积的转换成本。

“我们目前正在评估五种不同的 GenAI 解决方案，但无论哪个系统最能学习和适应我们的特定流程，最终都会赢得我们的业务。一旦我们投入时间训练系统理解我们的工作流程，转换成本就会变得令人望而却步。” - CIO，50 亿美元金融服务公司

研究限制： 这些百分比反映了我们52个组织的采访样本，可能不代表更广泛的市场模式。成功定义在各组织间有所不同，观察期可能无法捕捉长期实施趋势。

重要限制： 这些成功率差异可能反映组织能力而非仅仅是实施方法。选择外部合作伙伴关系的组织可能具有不同的风险承受能力、采购复杂性或内部技术能力，与那些进行内部构建的组织相比。外部合作伙伴关系与成功之间的相关性并不一定证明因果关系。

战略合作伙伴关系比内部开发工作实现了显著更高的成功部署份额。虽然我们在样本中观察到的BUILD倡议远多于BUY倡议，许多更多的组织在探索内部开发，但成功率偏向外部合作伙伴关系。尽管我们缺乏总体倡议量的精确数据，但模式表明内部开发工作尽管更常被尝试，但成功率要低得多。



虽然组织经常结合不同方法，但通过战略合作伙伴关系构建的试点项目达到完全部署的可能性是内部构建的2倍。更令人瞩目的是，外部构建工具的员工使用率几乎是内部构建的两倍。

这些合作伙伴关系通常提供 faster 的价值实现时间、更低的总成本和与运营工作流程的更好对齐。公司避免了从头构建的开销，同时仍然实现了量身定制的解决方案。理解这一模式的组织能够更有效地跨越GenAI鸿沟。

[6.2] [B]跨越鸿沟的买方实践

在访谈中，那些成功跨越GenAI鸿沟的组织呈现出一种一致的模式：顶级买家对待AI初创公司不像软件供应商，而更像商业服务提供商，采用更接近咨询公司或BPO的基准来评估它们。这些组织：

- 要求与内部流程和数据深度定制对齐
- 根据运营结果而非模型基准来衡量工具
- 通过早期失败进行合作，将部署视为共同进化
- 从前线经理而非中央实验室获取AI倡议

在最后一种模式中，个人贡献者和团队经理往往发挥关键作用。许多最强的企业部署始于超级用户，这些员工已经为个人生产力实验了ChatGPT或Claude等工具。这些”产消者”直观理解GenAI的能力和局限性，并成为内部认可解决方案的早期倡导者。成功的组织没有依赖集中的AI职能来识别用例，而是允许预算持有者和领域经理来发现问题、审查工具并领导推广。这种自下而上的采购方式，结合高管问责，加速了采用同时保持了运营适配性。

[6.3] [真正ROI所在：跨越鸿沟

要点：跨越GenAI鸿沟的组织发现，ROI在被忽视的职能（如运营和财务）中往往最高。真正的收益来自替代BPO和外部机构，而非削减内部员工。前台工具获得关注，但后台工具带来节省。

尽管50%的AI预算流向销售和营销（根据与高管的理论估算），我们记录的一些最显著的成本节省来自后台自动化。虽然前台收益可见且对董事会友好，但后台部署往往提供更快的投资回收期 and 更清晰的成本削减。

一流组织在这两个领域都产生可衡量的价值：

前台收益：



- 线索资格认证速度：提高40% • 客户保留：通过AI驱动的跟进和消息传递改善10%

后台收益：

- BPO消除：在客服和文档处理方面每年节省200-1000万美元 • 机构支出减少：外部创意和内容成本降低30% • 金融服务风险检查：外包风险管理每年节省100万美元

值得注意的是，这些收益没有大幅减少劳动力。工具加速了工作，但没有改变团队结构或预算。相反，ROI来自减少外部支出，消除BPO合同，削减机构费用，并用AI驱动的内部能力替代昂贵的顾问。

这种模式表明，虽然销售和营销获得了大部分关注和投资，但对于愿意超越明显用例并真正跨越GenAI鸿沟的组织来说，后台自动化可能提供更显著和可持续的回报。

[6.4] 就业影响现实：跨越鸿沟的实际含义

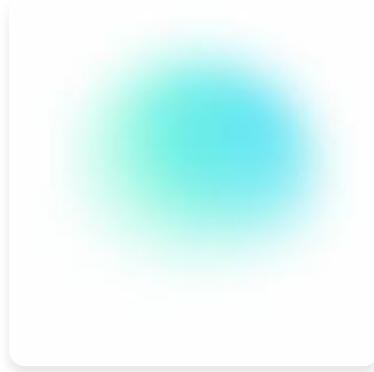
要点：GenAI已开始产生劳动力影响，表现为对以前外包职能的选择性替代和受限的招聘模式，但不是通过大规模裁员。跨越GenAI鸿沟的组织在略微减少内部员工数量的同时，展现出可衡量的外部成本削减。

6.4.1 替代模式和组织策略

我们的分析显示，GenAI驱动的劳动力减少集中在历史上被视为非核心业务活动的职能：客户支持运营、行政处理和标准化开发任务。这些角色由于其外包状态和流程标准化，在AI实施之前就表现出脆弱性。高管们不愿透露AI导致的裁员规模，但在这些公司中，客户支持运营和行政处理工作的裁员比例在5-20%之间。

特定行业的招聘预期揭示了与GenAI影响模式的明确关联。在受AI结构性破坏最小的行业——医疗保健、能源、先进制造业，大多数高管报告在未来五年内没有当前或预期的招聘削减。少数高管提到他们可以预期减少招聘，但承认目前没有系统来准确预测何时何地会发生。例如，医疗保健高管表示没有减少医生或临床员工招聘的预期。

相反，在GenAI已经展现可衡量影响的技术和媒体行业，超过80%的高管预期在24个月内减少招聘量。



关键发现：这种动态仍然集中在先进的AI采用者中，且仅在目前正经历GenAI显著破坏的行业（技术、媒体）中出现。

6.4.2 招聘标准和技能要求的演变

GenAI采用在组织间创造了不同的招聘策略。虽然高管对入门级或一般招聘量没有达成共识，但他们一致强调AI素养作为基本能力要求。这反映了组织对AI熟练度代表工作流优化竞争优势的认识。

“我们的招聘策略优先考虑展现AI工具熟练度的候选人。应届毕业生在这种能力上往往超过有经验的专业人士。” - 中型制造企业运营副总裁

6.4.3 未来劳动力影响预测

MIT的Project Iceberg分析为潜在的自动化暴露提供了定量背景：

当前自动化潜力： 美国劳动价值的2.27%

潜在自动化暴露： 2.3万亿美元劳动价值，影响3900万个职位

随着AI系统发展出持续记忆、连续学习和自主工具集成能力（这些能力定义了跨越GenAI分界线），这种潜在暴露变得可操作。

劳动力转型将逐步发生，而不是通过离散的替代事件。在AI系统实现上下文适应和自主操作之前，组织影响将通过外部成本优化而非内部重组来体现。

6.5 超越Agents：Agentic Web

要点：下一个超越个体AI agents的演进是agentic web，其中自主系统可以在整个互联网基础设施中发现、协商和协调，从根本上改变业务流程的运作方式。

这种转型的基础设施基础已经通过Model Context Protocol (MCP)、Agent-to-Agent (A2A)和NANDA等协议开始出现，这些协议不仅使agent互操作性成为可能，还实现了自主网络导航。在agentic web中，系统将自主发现最优供应商并评估解决方案，无需人工研究，建立实时动态API集成而无需预构建连接器，通过区块链智能合约执行无信任交易，并开发跨多个平台和组织边界自我优化的新兴工作流。早期实验显示采购agents独立识别新供应商并协商条款，客户服务系统跨平台无缝协调，以及跨多个提供商的内容创建工作流，具有自动化质量保证和支付。这代表了从当今人工中介的业务流程向在整个互联网生态系统中运行的自主系统的根本性转变，远远超越了当前的GenAI分界线，重塑了组织在网络经济中发现、整合和交易的方式。



7 结论：跨越GenAI分界线

成功跨越GenAI分界线的组织有三个不同做法：他们选择购买而非构建，赋权一线管理者而非中央实验室，并选择能够深度集成同时随时间适应的工具。最具前瞻性的组织已经在实验能够在定义参数内学习、记忆和自主行动的agentic系统。

这种转变不仅标志着工具的转变，还标志着Agentic Web的出现：一个持续的、互联的学习系统层，可以跨供应商、领域和接口进行协作。今天的企业技术栈由孤立的SaaS工具和静态工作流定义，而Agentic Web将其替换为能够协商任务、共享上下文并跨企业协调行动的动态agents。

正如原始Web去中心化了出版和商业，Agentic Web去中心化了行动，从提示转向自主的协议驱动协调。NANDA、MCP和A2A等系统代表了this web的早期基础设施，使组织能够不是从代码而是从agent能力和交互来构建工作流。随着企业开始在2026年锁定供应商关系和反馈循环，跨越GenAI分界线的窗口正在迅速缩小。下一波采用将不是由最炫目的模型获胜，而是由能够学习和记忆和/或为特定流程定制构建的系统获胜。

从构建到购买的转变，结合prosumer采用的兴起和agentic能力的出现，为能够提供学习能力、深度集成AI系统的供应商创造了前所未有的机遇。认识并行动于这些模式的组织和供应商将在后试点AI经济中建立主导地位，站在GenAI分界线的正确一侧。

对于目前困在错误一侧的组织，前进的道路很明确：停止投资需要持续提示的静态工具，开始与提供定制系统的供应商合作，专注于工作流集成而非炫目演示。GenAI分界线不是永久的，但跨越它需要在技术、合作伙伴关系和组织设计方面做出根本不同的选择。

8 附录

8.1 致谢

与MIT的Project NANDA合作制作：NANDA（Networked Agents And Decentralized Architecture）基于Anthropic的Model Context Protocol (MCP)和Google/Linux Foundation A2A构建，为大规模分布式agent智能创建基础设施。我们的研究专注于将AI能力转化为跨企业和中端市场组织的可衡量业务成果。我们感谢分享实施经验和见解的高管们的慷慨参与。



8.2 研究方法和局限性

方法：对企业利益相关者进行52次结构化访谈，系统分析300多个公共AI倡议和公告，以及对153位领导者的调查。成功定义为超越试点阶段的部署，具有可衡量的KPI。ROI影响在试点后6个月测量，根据部门规模调整。在适用情况下使用bootstrap重采样方法计算置信区间。

样本局限性：

- 我们的样本可能无法完全代表所有企业细分市场或地理区域

[•] 愿意讨论AI实施挑战的组织可能系统性地不同于拒绝参与的组织，这可能会造成偏向更具实验性或更谨慎采用者的偏差

[•] 愿意参与AI研究的组织可能存在选择偏差

[•] 成功指标在不同组织和行业间差异显著，限制了直接比较

方法论限制：

[•] 行业颠覆评分反映公开可观察的模式，可能无法捕捉私人或新兴发展

[•] 构建vs购买百分比基于访谈回应而非全面的市场数据

[•] ROI测量因同时进行的运营改进和外部经济因素而变得复杂

[•] 六个月的观察期可能不足以充分评估复杂企业系统的”成功部署”，可能低估了长期实施的成功率

未充分处理的外部因素：

[•] 影响采用的监管限制

8.3 研究工具

本研究使用了两个半结构化协议：

[•] **高管访谈** 专注于投资决策、组织设计和供应商选择



[•] **职能领导访谈** 探索用户偏好、工作流程适配性和日常使用中的摩擦

轻量级调查补充了这些访谈，提供了关于工具采用、满意度和障碍的可量化输入。

8.3.1 高管访谈问卷

重点：战略、投资、采购、结果

第1部分：战略和预算

1. 您的组织是否为GenAI计划分配了专门的预算？
2. 目前优先考虑哪些业务职能？
3. 您的组织是否确定了GenAI的具体用例？

第2部分：构建vs购买

4. 您主要是内部构建、外部合作，还是采用混合方法？
5. 是什么驱动该决策，成本、风险、时间表、控制等？

第3部分：试点到规模化

6. 自2024年1月以来，您启动了多少个GenAI试点？
7. 其中有多少现在已规模化部署？

8. 阻碍规模化的主要障碍是什么？

第4部分：采购和评估

9. 您如何评估潜在的GenAI供应商或合作伙伴？

10. 最重要的选择标准是什么（例如，信任、集成、数据控制）？

11. 推荐或生态系统合作伙伴如何影响决策？

第5部分：ROI和结果

12. 您是否观察到任何GenAI部署的可衡量ROI？

13. 使用了哪些指标（成本节约、生产力、客户保留）？

14. 是否有特定的后台办公室或前台办公室收益？

第6部分：劳动力和治理

15. 您是否因为GenAI而减少了员工人数？

16. 谁领导实施工作（例如，IT、业务线经理、AI CoE）？

17. 责任如何在团队间分配？

8.3.2 职能领导/用户访谈问卷

重点：工具可用性、摩擦和日常工作流程中的性能

第1部分：个人使用和偏好



1. 您个人是否使用ChatGPT或Claude等GenAI工具？用于什么任务？

2. 您在工作中使用内部GenAI工具吗？它们相比如何？

第2部分：企业工具体验

3. 您的组织引入了哪些GenAI工具？
4. 您使用它们的频率如何？
5. 什么运行良好？什么令人沮丧？

第3部分：工作流程适配

6. 这些工具是否与您的核心系统（例如，CRM、内部门户）集成？
7. 它们是否随时间适应您的工作流程还是感觉静态？
8. 您是否看到它们根据用户反馈而改进？

第4部分：任务类型偏好

9. 对于[X用例：电子邮件、文档起草、研究等]，您更喜欢AI还是人类同事？
10. 您信任AI处理什么类型的任务？您避免什么类型？

第5部分：采用障碍

11. 什么阻止您或您的同事更频繁地使用这些工具？
12. 培训、用户体验或对输出的信任是否是主要问题？

[1] Project Iceberg - Are you living under the Agentic API? <https://iceberg.mit.edu/index.html>

