



Discussion Note | 论鉴系列

总 No.1 期/2026 年 1 月

2026 宏观技术展望：

代理自主性、物理融合与数字真理

执行摘要

随着 2026 拉开帷幕，全球科技产业正处于一个决定性的拐点。如果说过去两年的主旋律是生成式 AI 模型的单点突破与能力跃升，那么 a16z 发布的“Big Ideas 2026”系列报告则清晰地描绘了一个更为宏大的图景：我们正在从“工具时代”迈向“代理时代”。这一转变不仅关乎软件能力的提升，更是一场涉及基础设施、商业模式、工业生产以及社会契约的结构重组。

本报告综合了 a16z 在基础设施（Infrastructure）、增长（Growth）、生物与健康（Bio + Health）、游戏（Speedrun）、美国活力（American Dynamism）、应用（Apps）以及加密货币（Crypto）等多个领域的深度洞察。核心论点在于，2026 年将标志着数字世界与物理世界的深度融合，以及软件从被动响应人类指令转向主动执行复杂意图的历史性跨越。

三个元趋势相互交织：

1. **基础设施的重构：** 为了应对 AI 代理产生的“惊群效应”式工作负载，企业后端必须从为人类设计的低并发架构，转型为能够承受大规模递归任务的“原生智能体”（Agent-Native）基础设施。
2. **物理世界的软件化：** AI 不再局限于屏幕，而是深入工厂、电网和矿山。未来的工厂本身即产品，具备模块化和自主进化的能力。
3. **价值流动的加数字化：** 随着非人类经济主体（AI 代理）数量激增，传统的金融身份验证（KYC）和银行账户体系已无法适应需求。数字货币以及“了解你的代理”（KYA）协议将成为机器经济的金融基石，而“质押媒体”（Staked Media）则成为对抗 AI 生成内容泛滥的锚点。

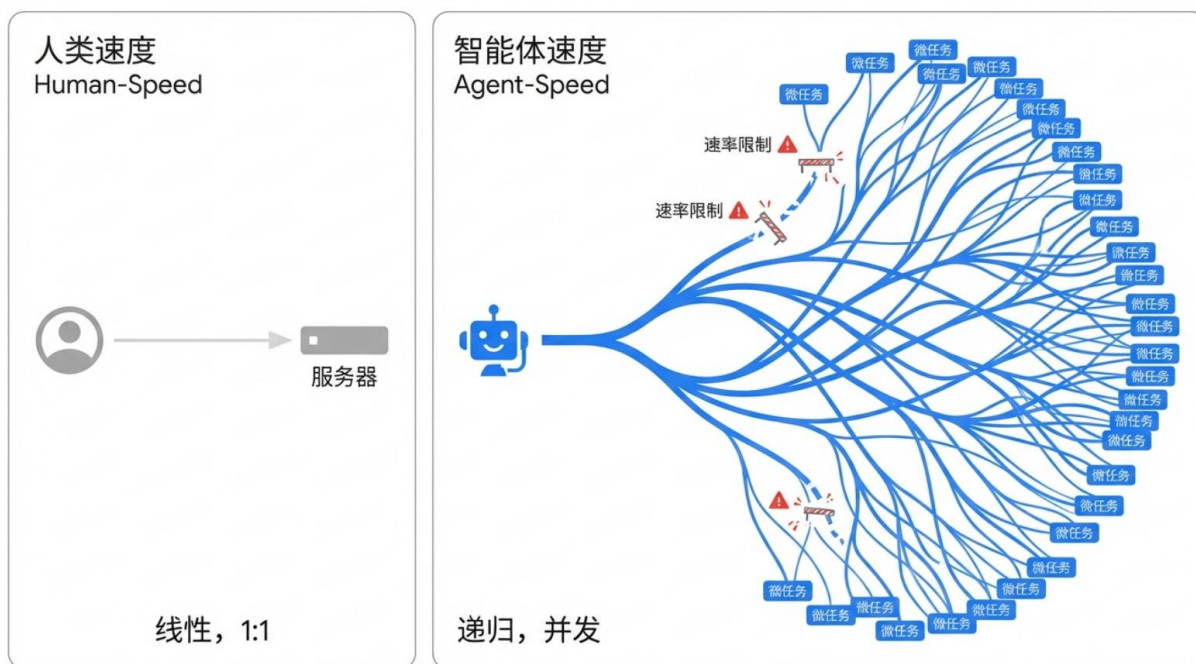
第一章：基础设施的范式转移——为自主性而生

1.1 从人类速度到代理速度：控制平面的彻底重构

在过去的二十年里，企业软件的后端架构是建立在一个基本假设之上的：人机交互的比率约为 1:1。即用户点击一次按钮，系统执行一次查询，页面加载一次。这种流量模式具有可预测性，并发量相对较低，且对延迟的容忍度是以人类感知的百毫秒级为基准的。然而，a16z AI 基础设施团队的 Malika Aubakirova 指出，2026 年我们将面临“原生智能体基础设施”成为标配的剧变。

AI 代理的行为模式与人类截然不同。当人类向代理下达一个高层次的模糊指令（例如“重构这个代码库”或“规划全球供应链路线”）时，这一单一意图会瞬间触发一个递归式的任务“扇出”（Fan-out）。代理会将目标拆解为成千上万个子任务，并在毫秒级的时间窗口内并发地发起数据库查询、API 调用和工具执行。对于传统的负载均衡器、速率限制器和连接池而言，这种突发性、大规模的流量模式在特征上与分布式拒绝服务攻击（DDoS）几乎无法区分。

The Thundering Herd: Human vs. Agent Workload Topology



While human interactions trigger linear, 1:1 system responses, a single agentic 'intent' triggers a recursive fan-out of thousands of concurrent processes. Legacy infrastructure interprets this bursty, high-concurrency pattern as a denial-of-service attack, necessitating new 'agent-native' control planes.

这种“惊群效应”（Thundering Herd）将成为系统运行的常态。为了生存，基础设施必须进行根本性的演进：

- **极度压缩的延迟方差：**代理通常以链式逻辑运行，任何一个环节的长尾延迟（Tail Latency）都会成倍放大，阻塞整个推理循环。因此，系统不仅要快，还要极其稳定。
- **数量级的并发提升：**系统必须支持比当前标准高出几个数量级的并发限制。
- **协调即瓶颈：**核心挑战从单纯的算力供给转移到了“协调”上——即如何在成千上万个并行执行线程之间进行路由、锁定、状态管理和策略执行。只有那些能够驾驭这种工具调用洪流的平台，才能定义下一代云基础设施。

1.2 驯服多模态世界的熵增

如果说计算层面临的是并发挑战，那么数据层则面临着“熵增”的危机。Jennifer Li 将非结构化的多模态数据视为 2026 年企业最大的瓶颈，同时也是最大的宝藏。目前，约 80% 的企业知识并不存在于整洁的数据库行中，而是沉淀在“半结构化的泥潭”里——PDF 合同、截屏、视频会议录像、系统日志和散乱的邮件往来中。

随着模型变得越来越智能，它们所摄入的数据输入却变得越来越混乱。这种“数据熵”（Data Entropy）——即企业知识库中真实性、结构性和新鲜度的持续衰减——直接导致了检索增强生成（RAG）系统的幻觉和代理执行的失败。试想，如果一个采购代理读取了一份过期的 PDF 合同扫描件，却忽略了后续在 Slack 上达成的最新修改协议，其执行结果将是灾难性的。

因此，2026 年的创业机会在于构建“多模态 ETL”（抽取、转换、加载）技术栈。这远超出了传统 OCR（光学字符识别）的范畴，它要求平台具备以下能力：

1. **结构提取**：将像素、音频和视频流转化为具备语义的 JSON 对象，使机器可读。
2. **冲突调解**：在相互矛盾的数据源之间（例如正式合同与临时邮件）判定“事实基准”（Ground Truth）。
3. **新鲜度治理**：确保提供给代理的上下文在时间维度上是准确的，防止因数据过时导致的决策错误。

这与 Jason Cui 提出的“AI 原生数据栈”演进趋势不谋而合。未来的数据架构必须解决“上下文问题”，通过构建统一的语义层，使代理能够在跨越多个孤立的记录系统时，依然保持对业务定义的理解一致性，从而实现真正的“与数据对话”。

1.3 视频世界模型：从观看介质到栖息之所

在创意工具领域，Yoko Li 描绘了一个关于视频技术的变革性愿景：到 2026 年，视频将不再是被动观看的流媒体，而将演变成我们可以“步入”的生活环境。目前的视频生成模型（如 Sora 早期的形态）大多只能生成短暂、不连贯的片段。而下一代“世界模型”（World Models）将赋予视频以物理和时间的一致性——模型将理解重力、物体恒常性和因果律。

这意味着视频从一种媒体格式转变为一种 *仿真基质*。机器人可以在这些视频生成的世界模型中进行强化学习训练，游戏设计师可以用它来快速原型化交互式现实。这被描述为“红药丸时刻”：我们不再是观看幻觉，而是进入一个遵循因果逻辑的生成式现实。这为物理世界的自动化提供了低成本、高保真的试验场，极大地加速了具身智能的发展。

第二章：代理经济与软件工具属性的终结

2.1 记录系统的衰落与行动系统的崛起

在企业软件市场，一场权力更迭正在发生。过去二十年，“记录系统”（System of Record, SoR）——如 Salesforce 的 CRM、SAP 的 ERP、ServiceNow 的 ITSM——占据了价值链的顶端。它们是企业数据的单一事实来源，也是员工必须登录操作的界面。然而，a16z 增长团队的 Sarah Wang 预测，在 2026 年，记录系统将失去其霸主地位。

AI 技术正在极度压缩“意图”与“执行”之间的距离。在过去，人类需要充当桥梁，通过繁琐的 UI 导航来将业务意图转化为数据库的更新。而现在，智能体能够直接读取、推理并写入数据层。这一变化导致传统的 SoR 正在被商品化为底层的“持久化层”（Persistence Tier）——一个仅仅负责存储状态的“哑数据库”，不再提供高价值的工作流服务。

价值正在向上迁移，汇聚于“行动系统”（System of Action）或“动态代理层”（Dynamic Agent Layer）。这是员工实际交互的智能界面，也是实际工作发生的地方。掌控了执行销售外联、自动报税或调度物流的代理层的公司，将掌控客户关系；而底层的记录系统将沦为管道。对于那些依赖“席位费”（Seat-based Pricing）的传统软件巨头来说，这是一个生存级威胁，因为代理不需要席位，而人类将不再访问它们的 UI。

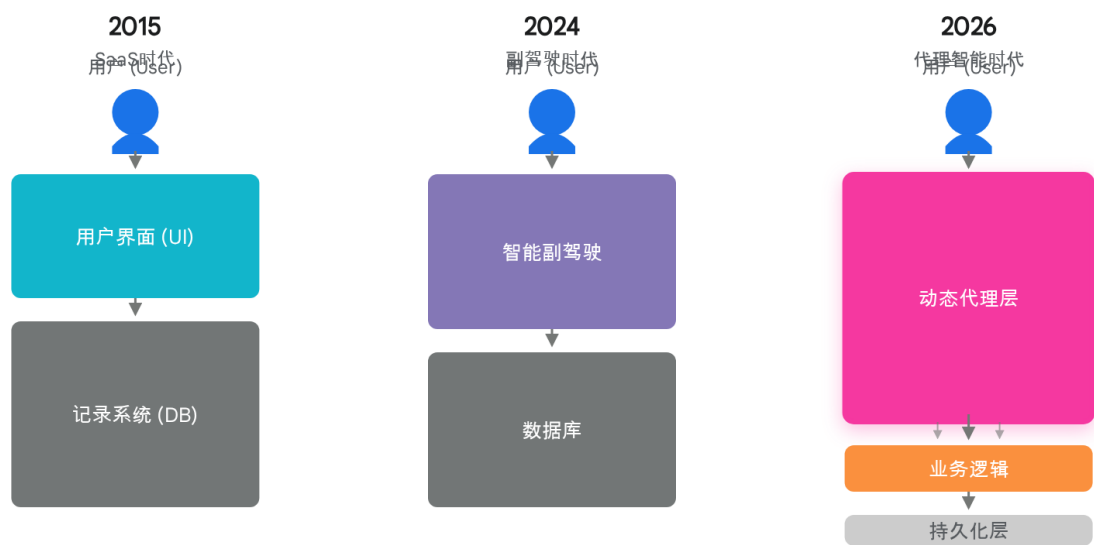
2.2 垂直领域的 AI 进化：多人模式

Alex Immerman 将这一逻辑延伸到了垂直行业软件（如法律、医疗、建筑）。垂直 AI 的发展经历了三个阶段：第一阶段是“信息检索”（找到相关条款）；第二阶段是“推理”（分析条款含义）；而 2026 年将开启第三阶段——“多人模式”（Multiplayer Mode）。

垂直行业的工作本质上是多方协作的生态系统：买方、卖方、监管者、供应商和顾问错综复杂地交织在一起。目前的 AI 工具大多是单机版的，导致用户在使用 AI 后，仍需通过传统的低效渠道（邮件、电话）进行跨组织沟通，形成了“没有授权的移交”。

价值迁移：从存储到执行

企业价值链演变图 (2015-2026)



历史上，价值主要积累在记录系统（数据库/UI）中。到2026年，“动态代理层”将截获用户意图，自主执行工作流。传统的记录系统将商品化为被动的“持久化层”。

数据来源: [a16z Big Ideas 2026](#), [a16z Crypto](#), [Productive Edge](#), [Firstsource](#)

未来的多人垂直 AI 将改变这一现状。在统一的软件基质中，代表买方的代理将直接与代表供应商的代理进行谈判。系统自动处理权限隔离、合规性检查和上下文同步。这种架构将网络效应引入了 AI 应用——这在过去是 AI 初创企业的软肋。此时，“协作层”本身构成了最强

大的护城河。

2.3 提示框的消亡与“我之年”

在消费者应用端，交互界面正在消失。Marc Andrusko 大胆预测，2026 年将是主流用户“提示框（Prompt Box）消亡”的一年。目前的聊天对话框被视为一种过渡性的“辅助轮”。下一代 AI 应用将具备零提示（Zero Prompt）的主动性：IDE 在你提问之前就建议代码重构，CRM 在通话结束的瞬间就起草好跟进邮件，设计工具在你工作的过程中实时生成变体。AI 将成为一种“隐形的脚手架”，由用户的意图而非显式的指令触发。

这与 Speedrun 团队 Joshua Lu 提出的“我之年”（The Year of Me）概念高度契合。工业时代的产品逻辑是“标准化”，而 AI 时代的产品逻辑是“超个性化”。产品将不再是为平均用户设计，而是为每一个具体的“我”实时生成。例如，Alphaschool 等初创公司正在构建的 AI 导师，能够根据每个学生的学习节奏和兴趣点动态调整课程；健康应用则根据用户的实时生物数据设计每日补剂方案。

随着交互方式的改变，衡量产品价值的指标也在发生根本性变化。Santiago Rodriguez 指出了“屏幕时间 KPI 的终结”。如果一个 AI 代理能在后台完美地替用户重新预订航班或完成报税，那么用户的屏幕停留时间可能趋近于零，但获取的价值却是无限的。因此，商业模式必须从基于注意力的经济（广告、时长）转向基于结果的经济（成功费、订阅费）。

第三章：经济活力与物理世界的复兴

3.1 AI 原生的工业基础

经济活力的核心论点在于，数字革命终于跨越了“气隙”（Air-gap），开始重塑物理世界。David Ulevitch 强调，美国正在重建其工业肌肉，但这就不是对旧系统的简单修补，而是建立一个“真正的 AI 原生和软件优先的工业基础”。

这不仅仅是在 1970 年代的设备上加装传感器。而是指那些从第 0 天开始就利用仿真、自动化设计和 AI 驱动运营的公司。AI 模型被用于设计更清洁的核反应堆，优化矿产开采流程，以及通过定向进化工程化出用于化工生产的酶。软件正在从“吞噬世界”转向“移动世界”，通过控制电池、电机和电力电子设备等“电工-工业技术栈”（Electro-Industrial Stack），重新定义工业和技术的未来。

3.2 工厂即产品

Erin Price-Wright 提出了“工厂复兴”这一概念，其核心在于采用“工厂思维”。2026 年的创始人将把工厂本身视为产品——它是模块化的、可复制的、软件定义的。通过将 AI 和自动化模块化部署，企业能够像亨利·福特的流水线一样，高效地处理复杂的定制化流程。这种思维方式将被应用于核反应堆的批量生产、数据中心的极速建设以及解决住房危机的房屋制造中。它可以显著缩短设计周期，并像处理代码一样处理复杂的监管和许可流程。

3.3 物理可观测性与数据十字军

正如 Datadog 和 Splunk 等软件可观测性工具让代码变得透明，Zabie Elmgren 指出的“物理可观测性”（Physical Observability）正在让现实世界变得可读。随着数十亿个联网传感器、无人机和摄像头的部署，我们第一次能够对电网、港口、城市进行实时的、连续的监

控。这种“可信的织物”是下一代机器人技术的前提，因为机器人需要一个可靠的、可观测的物理现实来安全运行。

然而，物理世界的数字化需要海量数据的支撑。Will Bitsky 描述了一场发生在关键行业中的“数据十字军东征”（Data Crusade）。能源、物流和制造业坐拥着巨大的、未开发的非结构化数据金矿（卡车行驶记录、仪表读数、维护日志）。在 2026 年，工业公司将不再仅仅依赖外部数据标注，而是开始积极捕获和利用自身运营产生的物理数据来训练专有模型，从而将其在物理运营上的比较优势转化为数据优势。

第四章：加密与 AI 的融合——构建信任的架构

在 a16z 的“Big Ideas 2026”报告中，加密货币（Crypto）不再是一个独立的金融赌场，而是被重新定义为 AI 和代理经济的“经济操作系统”。AI 提供了无限的数字劳动力，而加密技术则提供了这种劳动力所需的支付轨道、身份认证和真理验证机制。

4.1 “了解你的代理”（KYA）与代理经济的瓶颈

Sean Neville 敏锐地指出，代理经济发展的瓶颈正在从智能（Intelligence）转移到身份（Identity）。在金融服务领域，非人类身份（AI 代理）的数量已经达到了人类员工的 96 倍，但它们大多仍是“没有银行账户的幽灵”。目前的金融体系无法识别机器人，导致代理无法开设账户、签署合同或证明责任归属。

解决方案是“了解你的代理”（Know Your Agent, KYA）——这是一种新的加密原语。KYA 通过密码学签名凭证，将一个 AI 代理与其人类委托人（Principal）、约束条件（预算/权限）以及责任主体链接起来。只有通过 KYA，商家和银行才能信任并发起与机器人的交易，而无需担心欺诈或合规风险。没有 KYA，前文所述的“代理自主执行”愿景将因金融防火墙的阻拦而无法落地。

4.2 稳定币：机器商业的结算层

一旦代理拥有了身份（KYA），它们就需要一种货币。Christian Crowley 和 Jeremy Zhang 认为，稳定币正在成为“互联网的基础结算层”。以机器速度运行的代理无法忍受传统银行 ACH 结算的 T+2 延迟，也无法处理复杂的 SWIFT 电汇流程。它们需要一种可编程的、即时的、7x24 小时的价值传输工具。

稳定币去年的交易量估计已达到 46 万亿美元，超过了 PayPal 交易量的 20 倍，甚至接近 Visa。更重要的是，稳定币正在解锁“银行账本

的升级周期”。许多银行的核心账本仍运行在 1960 年代的 COBOL 大型机上，修改这些系统风险极高。通过稳定币和链上债券，金融机构可以绕过老旧的核心系统，在不重写底层代码的情况下，为客户提供实时的现代化金融产品。

4.3 质押媒体与真理危机

随着生成式 AI 模型以零边际成本制造出无限的内容（被戏称为“Slop”），互联网面临着前所未有的真理危机。Robert Hackett 提出了“质押媒体”（Staked Media）作为应对之策。在这一模型中，可信度不再来自于机构的背书，而是来自于“利益相关”（Skin in the Game）。

发布者和评论员需要通过密码学手段，为他们言论的真实性“质押”资产。如果他们的声明被证明是虚假的（这一过程可以通过预测市场或去中心化预言机来验证），他们将失去质押的资产。这为撒谎引入了直接的经济成本，从而在 AI 生成内容的洪流中重建了信噪比。为了构建一个功能完备的“智能体经济”（Agent Economy），必须建立三个核心支柱：身份（Identity）、结算（Settlement）和真理（Truth）。Know Your Agent (KYA) 协议为智能体提供了授权身份；稳定币（Stablecoins）提供了适应机器速度的结算轨道；而质押媒体（Staked Media）与隐私技术则确保了数据输入的真实性与策略的保密性。这三者共同构成了支撑自主代理运行的“加密原生操作系统”。

此外，Andrew Hall 指出，**预测市场**（Prediction Markets）将变得“更大、更广、更智能”。通过引入去中心化的治理机制和 LLM 预言机，预测市场将能够解决除选举之外更复杂、更细微的争议结果，实际上充当了互联网的“真理层”。

第五章：社会转型与新范式

5.1 健康 MAU 与 AI 原生大学

技术的融合最终将引发人类服务的深刻变革。Julie Yoo 在生物与健康领域提出了“健康 MAU”（Healthy MAU，月活跃用户）的概念，这将是医疗健康领域的新兴主导客群。传统医疗系统服务的是“患病 MAU”（病人），这是一种被动的、基于治疗的模式。

AI 驱动的持续监测和基于结果的定价，使得建立一种基于“预防”的商业模式成为可能——即通过持续与健康用户互动来保持其健康状况。这彻底翻转了保险的逻辑，从“风险分担”转向了“风险降低”。

同样，在教育领域，Emily Bennett 预测了第一所 **AI 原生大学** 的诞生。这不仅是学生使用 ChatGPT 辅助作业，而是一所从底层逻辑上就围绕智能系统构建的机构。在这里，课程设置会根据最新的科研进展实时重写（AI 可以瞬间阅读并综合新论文），日程安排会自动优化，评估体系也不再测试死记硬背，而是基于“AI 感知评估”（AI-aware evaluation），考察学生如何利用 AI 解决复杂问题的能力。

5.2 世界模型与故事讲述

在文化与娱乐领域，Jonathan Lai 预测 **世界模型**（World Models）将彻底改变故事讲述的方式。我们正在迈向一个“生成式 Minecraft”的时代。在这个时代，用户不仅是玩家，更是共同创造者，他们可以利用自然语言编程来修改游戏机制（例如“创建一个把所有东西变成粉红色的画笔”）。这种技术模糊了玩家与创作者的界限，将在这些生成式宇宙中催生出全新的数字经济体系。

结论

a16z 的“Big Ideas 2026”报告描绘了一个正在快速从静态的、人类中介的交互模式，转向动态的、代理中介的执行模式的世界。贯穿所有领域的统一主线是**自主性**（Autonomy）。

- **基础设施**正在重构，以承载这种自主性（原生智能体基建）。
- **加密货币**正在重构，以为这种自主性提供银行服务和身份认证（KYA/稳定币）。
- **工业**正在重构，以在物理世界中显化这种自主性（AI 工厂）。
- **社会**正在适应，以与这种自主性共存（质押媒体、AI 大学）。

对于投资者和构建者而言，信号是明确的：那个“为人制造工具”的时代正在结束，“为代理构建基础设施”的时代已经开始。未来的价值将不再属于那些仅仅提供效率工具的企业，而将属于那些构建了控制平面、信任层和物理接口，从而允许这一新的智能物种安全、有效运行的先行者。

Disclaimer: 本文由 AI 参与整理，若发现任何事实性问题，请联系我们：CEISD@Shanghaitech.edu.cn