

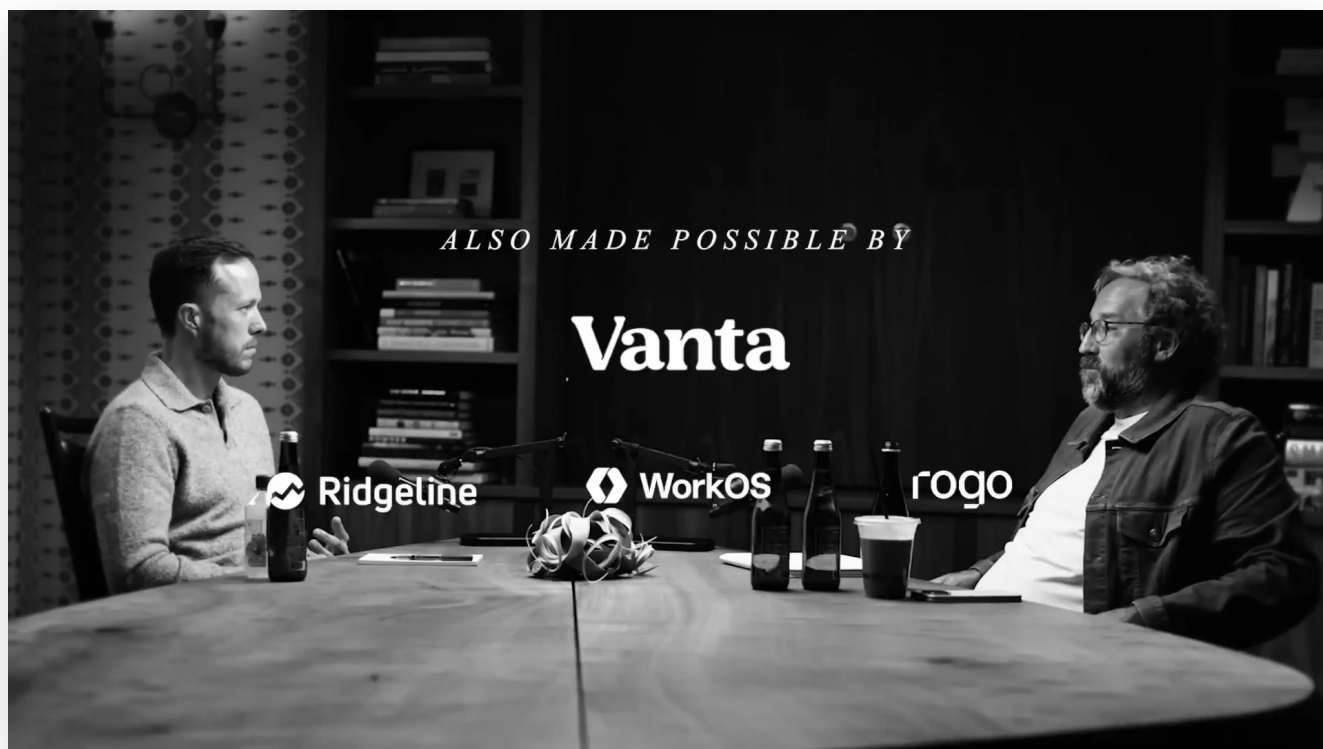
Why the Markets Are Pricing AI Wrong | Gavin Baker – 图文解说

一张关键画面对应一段完整解说，按原视频时间推进。

文章摘要

本期围绕《Why the Markets Are Pricing AI Wrong | [Gavin Baker](#)》展开，按时间顺序讨论2022 压缩进一个月：股价下跌与基本面加速、开源 token 与 GPU 价格：被忽略的需求加速、信用周期与算力融资：真正的压力测试、现货价格与需求：寻找真正的负面数据点。以下内容保留完整问答、例子、数字、画面信息和限定条件，适合先读这段摘要建立框架，再结合每个关键帧逐段阅读。

01. 2022 压缩进一个月：股价下跌与基本面加速（00:00–01:26）



场景 1

本节主旨 2022 压缩进一个月：股价下跌与基本面加速

完整内容

Gavin Baker: 我这次出来就是想被吓到。我不想像个疯子一样，看着这些股票越来越便宜，却觉得预期远期回报在上升。我的任务是做压力测试：请告诉我一些负面消息。但我还没找到一个量化指标能证明底层基本面在恶化；基本面在改善，股票却在下跌。以我们录制时的价格计算，NVIDIA 的远期市盈率是十年来最低，市场百分之百认为它们赚得太多了。

Patrick O'Shaughnessy: Gavin，才过了两个月。模型发布周期越来越短，我们俩几乎已经按模型发布节奏录播客了。

Gavin Baker： 有人指出，我们的播客过去总和局部市场高点重合；这次没人能这么说了。这个月太疯狂了。我会把七月形容为“一个月里经历了 **2022 年**”。确实有一些基本面上的负面因素，我们应该谈谈。

画面说明

双人棚内访谈中景，**Gavin Baker** 与 **Patrick O'Shaughnessy** 面向麦克风交替发言；画面显示节目品牌元素。

02. 开源 token 与 GPU 价格：被忽略的需求加速（01:28–10:20）



场景 2

本节主旨 开源 token 与 GPU 价格：被忽略的需求加速

完整内容

Gavin Baker: 总体来说，基本面在显著改善。很多 AI 股票从高点直线下跌了百分之四十到六十，但你在这里待了整个夏天，听到过哪怕一个 AI 负面的量化指标、哪怕一次增速放缓吗？没有；每个指标都在加速。无论看 GPU 可得性、租赁价格、这个月的 DRAM 现货价，还是 token 增长，结论都一样。

市场看不到 Anthropic、OpenAI，以及美国本土把开源推理变现的 inference cloud（Fireworks、Together 等），所以图景被扭曲了。开源增长非常快，GLM-5.2、Kimi K3 等模型带来能力跃迁；NVIDIA 的开源布局也在继续，OpenAI 加速，Anthropic 仍然强劲，并且很可能已经产生大量自由现金流。只看半导体和超大规模云厂商的现金流图，会漏掉这些私营公司。

大家在 2024、**2025 年**都以为 GPU 价格会慢慢下跌，悲观者甚至认为会断崖式下跌；没人想到到 **2026 年**老 GPU 价格还会垂直上升。新云厂商为了融资被迫签长期合约，于是已安装算力按合约价交易，远低于现在的现货价。合约到期、算力重定价后，会出现明显加速，并回答 ROI 的问题。

本季度已经看到这一点：Microsoft、Meta 和 Amazon 报告的经营现金流从 28 加速到 32；剔除一次性法律费用等项目，可能从 28 到 35。**这个规模上的加速很有意义，而且还没算上以溢价交付的 Rubin（ASR 中称为 Rubicon）以及合约重定价。**

这个月市场还误读了 Meta 出租算力、SpaceX 出售交易优化集群、OpenAI/开源模型发布以及中国 DUV 设备的消息。Meta 并没有削减资本开支，市场只是把行动解读成灾难。GLM-5.2 与 Kimi K3 让 token 组合从高毛利前沿模型转向开源模型，硅谷数据 token 指数因此暂时走平，但一个 token 仍需要同样的 FLOPs、**内存**和功耗；开源只是在拿走前沿模型层的利润，并把更多利润和需求推入 **AI 基础设施层**。NVIDIA 支持开源，也说明开源不必然损害其业务；而只有一两个收取约 **90%** 毛利的前沿模型，对人类和社会都不是好事。

画面说明

Gavin 单人近景为主，镜头在双人机位间切换；本段画面以演讲者为主，没有可读的外部数据图表。

03. 信用周期与算力融资：真正的压力测试（10:20-18:29）



场景 3

本节主旨 信用周期与算力融资：真正的压力测试

完整内容

Gavin Baker: 真实收益率上升，利差扩大。Meta 上周发行债券，定价不像大家想象中那样好；所有公司的 CDS 都在走阔。聪明的私募资本会说这只是对冲承诺，但事实仍然不好看：CDS 上升、利差扩大、实际利率上行。如果这轮建设必须依赖信用融资，那会非常可怕；正因如此，已安装算力的合约价与现货价差异很重要。

赞助口播: Ramp 帮助财务团队更精简、更快、更高效，客户平均每年节省 5%，使用 Ramp 的企业包括 Visa、Cursor、Stripe、Notion、11x、Shopify 等七万多家公司，详情见 ramp.com/invest。OpenAI、Cursor、Perplexity 等 AI 和软件公司使用 WorkOS 在一夜之间完成

企业就绪，详情见 workos.com。Rogo 的 Felix 是面向投资机构的个人财务 agent，可以用机构自己的模板、上下文和标准，把邮件要求变成 PowerPoint、Excel 模型和带来源的研究，详情见 rogo.ai/felix。

Patrick O’Shaughnessy：未来六个月的融资情况很关键。这轮建设需要多少信用？

Gavin Baker：这就是经典的资本周期：债务驱动的建设要求即时偿还，供需稍微错位就会迅速失衡，互联网周期已经发生过。假设超大规模云厂商的 Blackwell 和下一代 Rubin 按落后一代芯片的速度变现，经营现金流大约是 1.3 至 **1.4 万亿美元**；如果它们的变现能力更强，可能接近 **2 万亿美元**，相当于减少约 **7000 亿美元**信用需求。算力重定价后，信用指标会改善，可能大部分甚至全部建设都能由经营现金流承担。

我曾担心 Blackwell 出现“空气缺口”：先花数千亿美元买来、主要用于训练的算力，却还没有回报。但 Anthropic 的需求让我在四月、五月、六月暂时忽略了这个风险；七月多重因素叠加，市场又不再忽略它。核心问题是：硅谷私营公司的量化需求信号能否继续，让已安装算力在合约到期时涨价？如果能，现金流上升就足以支持未来几年的投入。市场像 **2022 年**一样下跌，但除信用以外的理由都很牵强；我仍坚持投资中最重要的三个词是“安全边际”。

画面说明

画面保持访谈机位，间或出现 Ramp、WorkOS、Rogo 赞助卡与读稿画面；未见可核验的现场数据图表。

04. 现货价格与需求：寻找真正的负面数据点（18:29–24:24）



场景 4

本节主旨 现货价格与需求：寻找真正的负面数据点

完整内容

Gavin Baker: 我今天刚和一家大家都想合作的热门初创公司谈过。他们七个月前租了一组几千张 Blackwell，GPU 小时价格在 **2 美元** 中段；现在租同样规模、同样 B200 配置的集群，七个月后希望价格接近 **4 美元**。按起点不同，六七个月上涨了 **50% 到 60%**。另一个**推理云**也说，合约到期时准备为 Blackwell 多付 **100%**。这意味着超大规模云厂商一直在低估算力价值。

Patrick O'Shaughnessy: 你听到的唯一负面量化指标是什么？

Gavin Baker: 第三方数据似乎显示 Anthropic 的曲线稍微偏离原来轨迹；这可能是真的，但 OpenAI 和开源在大幅加速，三者合计仍在净加速。开源像公募市场看不见的“暗物质”：**推理云**在播客或会议里说需求很明显加速，虽然不是审计财报，但与 GLM-5.2、Kimi K3 的能力跃迁相符，

NVIDIA 也会把自家模型逐步推进前沿。

我已经把每个假设都做了压力测试，底层基本面在改善，而 NVIDIA 的远期市盈率是十年来最低；只有“解放日”和 DeepSeek 时更便宜，而且那两次都很快 V 型反转。市场认为它们赚得太多了，我们必须谦逊，也许市场是对的，但我来硅谷就是尽量寻找负面数据点，除了 Anthropic 那条有争议的数据，几乎没有找到。OpenAI、开源、Grok、Cursor 的第三方数据都很强。

我有位 Fidelity 的朋友说，过去三年最简单的做法是尽快做最笨、最表面的事，在叙事之间轮换、不断减仓。可是我们做的工作显示，除信用以外，那些让市场下跌的叙事在事实层面并不成立。即使信用不足以建设全部需要的 FLOPs，现有 FLOPs 也会更有价值；我还要继续讲今天看到的具体基本面改善。

画面说明

Gavin 在单人机位中持续回答，画面偶尔切回主持人；本段没有可读的价格图表，信息主要来自口述数字。

05. Claude 解释一切：市场如何被同一种解读驱动（24:25–29:10）



场景 5

本节主旨 Claude 解释一切：市场如何被同一种解读驱动

完整内容

Gavin Baker: 不论散户还是机构，每条新闻都会立刻被喂给 Claude，有时是 Claude Code 或 Claude agent。它是概率性的，但解读方式可能没有太大差异。媒体从 Walter Cronkite 这个唯一真相来源，变成碎片化媒体；Claude 几乎成了股市的 Walter Cronkite，大家相信它说的每件事。它非常聪明，但不总是正确，而股市本质上是在对未来做概率性的贝叶斯解读。

于是新闻先进入 Claude，Claude 这样解释，很多人就按 Claude 的观点交易。半导体圈有人贴出日本电容器股票图，说六周走完了一个完整的电容器周期；股票可能翻倍、三倍或四倍，随后又迅速回落，实际基本面甚至还没发生，原本三年的周期被压缩到了六周。

Patrick O'Shaughnessy: 这里正在发生的创新，会提高训练和推理服务的每个环节效率，长期会怎样影响公开市场？有没有哪种长周期创新让你尤其兴奋或好奇？

Gavin Baker: 很多人似乎已经接近解决持续学习和样本高效学习。如果模型不必用 300 万亿 token 训练，而是用约 10 万亿 token 就能高效学习再投入世界，短期内可能造成需求断点。训练占半导体需求和计算的比例最终会趋近一个很小但不为零的数。SSI 说八月要发布模型，新一代实验室都在做这件事；这会对世界非常好，但我很难相信它会真的打击 AI 基础设施需求。NVIDIA 深度参与这些初创公司，Jensen 能看到每个实验室的进展，他所看到的东西显然让他保持乐观。

Patrick O'Shaughnessy: 如果你必须列出让自己真正害怕、立场反转的条件，会是经营现金流不再继续增长、整个建设转而依赖债务吗？

Gavin Baker: 是的；如果 OpenAI、Anthropic 等实验室的总量平台期或下降，会很负面，除非是开源 token 净增长、扩大了整个蛋糕。长期会是多模型世界：AI 原生公司把自己的数据用于微调开源模型，放在路由器后面，先调用自家模型，必要时再由 Claude 或其他前沿模型检查，很多场景可以用一半价格得到略好结果。成本下降并不等于算力需求下降，只是把高毛利 token 换成低毛利 token；token 仍然需要同样的计算。企业用路由器会稳定甚至降低支出，却可能增加 token 数和 GPU 小时。现在只有约五十万到二十五万人在使用 agentic AI；如果扩大到全球人口的 1%、一亿、五亿，会发生什么？

画面说明

Gavin 单人机位回答，主持人从双人机位追问；画面未显示 Claude 或股票图表，主要可见的是棚内麦克风与节目布景。

06. 多模型路由与劳动替代：token 需求从哪里来（29:11–37:05）



场景 6

本节主旨 多模型路由与劳动替代：token 需求从哪里来

完整内容

Gavin Baker: 如果 GPU 价格出现持续、剧烈的收缩，市场会立刻反应；但我没听过任何人说 GPU 太多，大家反而像在毒品市场里抢货。若 Anthropic、OpenAI 等实验室总量停滞或下降，除非开源 token 把蛋糕做大，否则确实是负面。未来会是多模型：AI 原生公司用云平台的监督微调和强化学习定制开源模型，再通过路由器把请求分到自家模型或 Claude、Grok 等前沿模型，很多时候成本减半、结果略好。

一半成本来自毛利率从前沿模型的约 90% 换到开源模型的约 30%，并不是少用了计算。大公司发现 AI 预算三个月就烧完，会用路由器降低账面支出，却可能把 token 转移到更便宜的模型，实际 token 和 GPU 计算小时继续增加。AI 原生公司把更多投入直接放进 token，不减速；美国东海岸、

内陆公司和欧洲则处于不同的采用阶段，甚至还在研究如何监管。

关键问题是客户在哪里。**超大规模云厂商的经营现金流必须来自更快的生产率增长或劳动替代。**许多 AI 原生公司不是大规模裁员，而是少招人；每位员工对应的毛利美元比过去高。知识工作总额约 **25 万亿美元**，即使按 **20%** 计算，也相当于 **5 万亿美元**，来源可以是劳动替代或更快的经济增长，我们更希望是后者。

Patrick O'Shaughnessy：我听到一位连续创办多家公司的科技 CEO 说，创始人控制的公司并没有大规模裁员，即使是最可能用 AI 提效的人也是如此；这说明他们仍然相信“人加 token”会带来机会。

Gavin Baker：这就是增长的多头情景。Cognition、Ray、Stripe 等公司的数据表明，AI 花费最多的公司增长明显更快。Cognition Index 很惊人，虽然没有完全控制行业差异；连水管工或 HVAC 承包商都能因 AI 做得更好。现在大家都在引用长期供应协议（LTA），因为所有东西都短缺；如果有弱点，可能只是我们还来不及把足够的千兆瓦接入电网。

画面说明

访谈近景持续展示 Gavin 与 Patrick 的对话；未出现可读的统计图或公司界面，视觉证据限于说话人和棚内机位。

07. 长期供应协议（LTA）：内存供给的博弈（37:05–42:06）



场景 7

本节主旨 长期供应协议（LTA）：内存供给的博弈

完整内容

Gavin Baker: 千兆瓦最终会接入电网，监管政策也在改善；涡轮机和柴油发电机厂商正在响应，甚至从旧飞机上拆涡轮机、翻新后再利用。资本主义很擅长解决这类问题。但市场的一个重要转变是，尤其在**内存**领域，大家从追逐短期业绩上行，转向签订供应链长期协议（LTA）：客户预付款，协议设定价格下限和上限。

想想打破 LTA 的博弈。规模上真正重要的买家是 Amazon 的 Trainium、Google 的 TPU、AMD，以及远大于其他几家的 NVIDIA。对给定计算量，**内存**越多，产出的 token 越多；这是提高单位计算 token 产出的最重要手段，也会直接降低成本，所以需求没有出现弹性下降。

假设 2027 或 **2028 年**某家公司为了低价打破协议，而我们并未处于严重过剩。接下来两三年，一旦杠杆重新回到内存厂商手里，打破协议的公司可能直接失去生意。Google 如果在过剩时打破 LTA，价格下跌后产能收缩，下一轮短缺时它还能拿到什么配额？这是周期行业，过剩之后会紧缺；供给分配和预购量将在未来几年决定市场份额。

过去 Apple 是压倒性的大买家，几乎可以随意压价，因为供应商没有别的选择。**现在至少有四个玩家，还有很多初创公司；如果一家打破价格协议，供应商可以说“那我们也打破数量协议，把产量给竞争对手”，结果就是丢掉份额。**围绕 LTA 的博弈，可能摧毁一家公司的业务和品牌，这与过去不同。

画面说明

画面为 Gavin 单人讲解，插入双人机位；选定画面显示棚内访谈构图，没有实际 LTA 合同或供应链图表。

08. NVIDIA 的信用包装与算力竞赛 (42:08–50:16)



场景 8

本节主旨 NVIDIA 的信用包装与算力竞赛

完整内容

Gavin Baker: 当前环境偏向 NVIDIA，我不理解它为什么交易在这么低的倍数。没有什么比 NVIDIA GPU 更容易融资；如果需要土地和电力，他们还在做撮合。NVIDIA 推出类似“信用包装”的商业模式：当 GPU 价格高于某个底价时分享收入，可能很快通过特许权形成巨大的云业务，也能缓解现金流错配。

这不完全是风险投资或供应商融资：别人在给 GPU 买家融资，NVIDIA 仍然做股权投资，并在协议里写明资金不能买 NVIDIA 芯片；但钱是可替代的。内存公司如果有现金，也会向 GPU 买家、Trainium 等客户提供类似安排，先拿出一部分现金、再分享未来收入，把长期协议延伸为经常性收入的 royalty。NVIDIA 正在做的正是这个，而且市场严重低估了它。

Jensen 看到所有实验室的进展，持续看好 AI；既然有现金流、又看好 AI，就同时要股权上行和收入分成是合理的。它提高每千瓦收入，也强化竞争地位：初创公司使用谁的芯片，价格、HBM DRAM 供给以及能否以同样利率融资，都会影响选择。

Anthropic 如果像 OpenAI 一样积极买算力，可能已经领先；现在 OpenAI、Grok 都重新进入竞争，SpaceX 也通过 Falcon 和 Crew Dragon 大举进入。看到竞争后，谁还会因为“买太多算力可能破产、不买又会输”而松油门？只要能由经营现金流支持，就很难后退。

赞助口播： Vanta 为包括 Ramp、Cursor、Harvey 在内的 1.6 万多家快速成长企业自动化安全与合规，Vanta agent 像全天候 GRC 工程师，最多可把供应商评估时间减少 50%，详情见 vanta.com/invest。RidgeLine 为投资机构提供带 AI 的统一记账、对账、报告、交易和合规系统，详情见 RidgeLine AI。

画面说明

Gavin 在访谈近景中分析 NVIDIA 的融资模式，段末切至 Vanta 与 RidgeLine 赞助读稿卡；未出现可读的财务模型。

09. 中国 DUV 与开源软件的第二条增长曲线（50:17–58:28）



场景 9

本节主旨 中国 DUV 与开源软件的第二条增长曲线

完整内容

Patrick O’Shaughnessy: 你在旅途中遇到过比自己更看多的人吗？他们相信什么，是你没有纳入预期的？中国传出 DUV 设备的消息后，市场反应从“这相当于 ASML 在 2001 年的阶段”到“这是全球尖端算力供应新故事的开端”，你怎么看？

Gavin Baker: 基本上每个人都比我乐观。有人估算租一台 H100 一年要 25 万美元，大约是现在现货价的 15 倍；这完全不在我的贝叶斯预期空间里。若把计算价格、计算量和推理利润同时乘起来，就能解释实验室加开源的加速。我在股票市场上像个愚蠢的乐观主义者，但和实验室或生态里的每个人交谈时，我相对所有人都更悲观。

DUV 就像螺旋桨飞机，EUV 像喷气发动机或跃迁到另一阶段。中国以前没有 DUV，现在据称有了，这是重要的相变；即使仍落后 **25 年**，也不能忽视。市场可能过度反应，因为如果真的影响 ASML 订单，也许要五年后才发生，到那时市场已经反复担忧、遗忘。中国非常聪明、勤奋，也知道这对国家很重要，但不能跳过“学习—实践”的周期，不能把做的过程传送到未来。我们正在脱钩，这是不幸但会自我强化的过程。

Gavin Baker：开源模型更接近前沿、Fireworks 让模型定制变得简单，对软件行业和 AI 原生公司是好消息。两年前，很多从成立到年收入 **5000 万美元**、九个月内产生现金流的公司被叫作 ChatGPT wrapper；现在它们用自己的垂直数据微调开源模型，获得独有数据，防御性更强。Fireworks 的 DEXIS 可以接收 Cloud Code、OpenAI Codex、Grok Build 的数据，做强化学习，再用路由器分发查询；只需几行代码，便能让 AI 原生公司把前沿模型只用于 **30% 到 60% 的 token**，其余由自己的模型处理。Cursor 的“AI speed running”也是先让前沿模型规划，再把任务交给较弱模型，效率可提高约 **15 倍**。

画面说明

访谈双人/单人机位交替，画面未展示 DUV/EUV 设备照片或 DEXIS 界面；重要信息来自口述类比与数字。

10. 前沿模型与开源 token：价值如何重新分配（00:58:29–01:05:03）



场景 10

本节主旨 前沿模型与开源 token：价值如何重新分配

完整内容

Gavin Baker: 另一种观点是，前沿模型会蒸馏能力，在每个智能层级提供低成本服务，开源就没有位置；但我们应该保持谦逊。许多 AI 原生公司已经生成了相当多的领域专有数据。在推理云和路由器成熟前，它们只能接受服务条款；现在可以脱离这条轨道，获得独立性、持久性和安全性。

便宜的 token 反而可能大幅提升最尖端 token 的价值：如果今天有很便宜的“120 IQ”开源模型，能够编排它们的“160 IQ”模型会更有价值。过去绝大多数经济收益都流向前沿模型，但推理云让开源 token 可能占据大多数处理量。Together 等基于模型的推理云增长几乎和早期前沿实验室一样快，却几乎不烧现金；按 SaaS 的 Rule of 40 看也很惊人。

Patrick O'Shaughnessy: 组织里 CEO 的薪酬远高于中位数员工，是否可以类比为前沿 token 与开源 token 的分配？

Gavin Baker: 可能是。前沿 token 仍会占据大部分经济价值，但不再是全部；开源 token 可能成为处理量的多数。对基础设施来说，token 就是 token：制造它需要相同的 FLOPs、瓦特、空间和冷却。

AI 最大的风险是监管。纽约已经在讨论数据中心暂停建设，而我们生活在一个后事实、后逻辑的政治世界。AI 行业公关做得很差，普通美国人的叙事是：数据中心会推高电价、拿走水、抢走工作。现实是，数据中心与电表后方签约，周围居民的电价通常会下降；开发商会建设医院、学校、警察和消防设施，并带来水管工、电工、HVAC 承包商等持续的高薪蓝领工作。数据中心可能是我这一生中对蓝领工资最有利的事情之一。

有人在书中把数据中心用水量高估了一万倍，作者已经多次承认错误，但错误仍像“菠菜含铁量”的都市传说一样传播。行业需要自己讲清楚电价、社区收益、就业和 AI 治病救人的故事，而不能假定公众已经知道。

画面说明

Gavin 近景回答监管与数据中心议题，画面保持固定棚内构图；未出现数据中心或菠菜的配图。

11. 讲清数据中心价值与 SRAM 加速器 (01:05:03–01:11:41)



场景 11

本节主旨 讲清数据中心价值与 SRAM 加速器

完整内容

Patrick O'Shaughnessy: 菠菜含铁量的传言也是因为学术书里小数点放错位置，后来一路传播；我自己也一直以为菠菜铁很多。

Gavin Baker: 同样的事情在 AI 叙事中发生。行业需要有人说真话，甚至成立基金会或政治行动委员会，在四强赛、NFL、大学橄榄球和世界大赛期间投广告：签署数据中心承诺的社区，电价会下降，会得到实质性社区投入和大量持续的高薪蓝领工作。建设初期有一波岗位，之后还有运维、升级和技术迭代；电价更低、社区更富裕，水和环境没有额外影响，数据中心也可以建在城外十英里。

AI 也在挽救生命、治疗罕见疾病。ASCO 今年的氛围是一次会议里出现了最多科学突破，其中一部分归功于 AI。行业必须把这些故事讲给有病孩子、父母和亲人的普通人听；硅谷的人觉得一切显而易见，却没意识到这与美国大多数人的看法差异极大。

Gavin Baker： 另一个缺失点是 SRAM 加速器。它们不受 HBM DRAM 限制，常在较老制程上制造，不与最新 GPU 争夺同一资源。将 prefill、attention 和 feed-forward network 解耦后，可能在一块没有 HBM 的芯片上做 prefill，在高功率 HBM 芯片上做 attention，再把 feed-forward network 放到 SRAM 芯片上。无论怎样调整芯片上计算、HBM DRAM 与 SRAM 的比例，工作负载都会变化；把三部分拆开，能提升现有和新增算力的 ROI。

Patrick O'Shaughnessy： 除了 Anthropic、OpenAI、Microsoft、Amazon、NVIDIA，是否会有黑马达到“权力的游戏”级别，比如 Micron？

Gavin Baker： Leapfrog 可能是黑马；Fireworks 的 Lisa 是杀手级人物；Scott Wu 的 Cognition 也很值得关注。接下来聊 SpaceX：它可能是最重要的新上市公司，但市场目前还没真正理解它。

画面说明

画面从双人对谈切到 Gavin 单人近景；赞助结束后仍是棚内访谈，没有显示 SRAM 架构图或公司标识以外的图表。

12. SpaceX、轨道算力与未来的概率（01:11:41–01:18:43）



场景 12

本节主旨 SpaceX、轨道算力与未来的概率

完整内容

Gavin Baker: SpaceX 的基本面自 IPO 以来一直变好：Grok 4.5、收购、核心业务都在加速；过去三年它比任何人都快、以更低价格接入更多算力。它把巨量算力一夜之间带入市场，市场几乎无感地吸收了。有人估计 SpaceX 想带来 8 吉瓦算力；我不会做空 Elon，但这会是惊人的成就。它签约后利率是上升而不是下降，按每吉瓦约 500 亿美元变现，明年共识约 730 亿美元。即使暂时忘掉 Starlink V3、直连手机、Grok 4.5 和 Cursor，只要算力计划兑现一部分，收入就可能迅速达到百亿美元级别。

到目前为止，一年内接入超过 500 兆瓦电力的只有超大规模云厂商、CoreWeave、Crusoe 和 SpaceX；SpaceX 接入得最快、成本最低，集群口碑也很好。市场需要看到这些事实，因为这不是今天市场对 SpaceX 的主流解读。纽约某大型对冲基金的做空逻辑是算力现货价会下降 90%，新增

算力赚不到预期收入；这可能发生，但 Elon 的公司过去反复做成看似不可能的事。他的一句话是“我们专门把不可能的事做晚”。

我在 Starbase 待过，轨道算力每天都显得更真实。StarCloud 是一家轨道算力公司，SpaceX 似乎会让它使用 Starlink 激光技术；PitchMark 的投资人并不属于 Elon 生态，却仍以合理估值投资了它。这让我检查自己是否疯了：也许我、Elon、Bitmark 和 SpaceX 工程师都疯了，但这种共同判断不太可能全是疯狂。

Patrick O'Shaughnessy：我们现在坐在 Big Spark 办公室那张著名的晚宴桌旁。

Gavin Baker：谢谢 Bitmark，也谢谢 Eric 和所有合作伙伴。未来一年，时间会告诉我们这些股票谁对谁错；未来是概率性的，但现在确实是令人兴奋的时刻。如果继续按模型发布周期录制，几周后再见。

赞助口播：Ramp 让企业把支出系统变成资本配置策略，详情见 ramp.com/invest。Vanta 随企业成长自动化合规并提供安全与风险的单一事实源。RidgeLine 帮助资产管理公司扩张五倍，详情见 RidgeLine Apps。OpenAI、Cursor、Perplexity 等使用 WorkOS 一夜之间完成企业就绪，详情见 workos.com。

画面说明

画面显示双人访谈及收尾赞助卡；结尾回到主持人与嘉宾同框，未出现轨道算力示意图。