

BRENDAN FOODY

LENNY'S PODCAST

DEEP ANALYSIS

ORIGINAL BY

Lenny Rachitsky

@lennysan • x.com/lennysan

ANALYSIS BY

@Penny777 • x.com/penny777

Brendan Foody - Lenny's Podcast

Brendan Foody - Lenny's Podcast: 深度解析报告

主持人介绍

Lenny Rachitsky

- **身份:** 前 Airbnb 产品负责人，硅谷顶级产品管理专家。
- **背景:** 在 Airbnb 工作 7 年，负责 Supply Growth 团队，帮助平台从数十万房源增长到数百万。
- **现状:** 运营全球最大的产品管理 Newsletter（50万+订阅者）和播客，是硅谷增长与产品领域的风向标。
- **社交媒体:**
 - Twitter/X: @lennysan
 - LinkedIn: Lenny Rachitsky
 - Newsletter: Lenny's Newsletter

嘉宾介绍

Brendan Foody

- **身份:** Mercor 联合创始人兼 CEO，被誉为历史上最年轻的独角兽创始人之一。
- **职业经历:**
 - **Mercor** - 联合创始人兼 CEO (2023 - 至今): 带领公司在 17 个月内将收入运行率 (Revenue Run Rate) 从 100 万美元提升至 5 亿美元。
- **教育背景:** 曾就读于乔治城大学 (Georgetown University)，后退学创业。
- **核心专长:** AI 训练数据架构、高技能人才市场撮合、超高速企业规模化 (Hyper-growth)。
- **社交媒体:**
 - Twitter/X: @brendanfoody
 - LinkedIn: Brendan Foody
 - 公司官网: Mercor

内容概要

本期访谈揭秘了史上增长最快公司 Mercor 的幕后故事。Brendan Foody 分享了他们如何抓住 AI 浪潮中的“评估（Evals）”机遇，通过构建一个连接顶级 AI 实验室（如 OpenAI, xAI, Anthropic）与全球顶尖专家（医生、律师、工程师）的自动化人才平台，实现了 16 个月内从 100 万到 4 亿美金营收的奇迹。对话深入探讨了 AI 时代的劳动力市场变革、评估集（Evals）为何是新的产品需求文档（PRD），以及在极度竞争的环境下如何保持极致的执行力。

核心话题

AI Evals RLHF 人才市场 超高速增长 创业文化 未来工作

核心论点

论点一：我们正进入“评估时代”（Era of Evals）

核心观点: 评估集（Evals）是 AI 模型的产品需求文档（PRD）和销售抵押品。

- PRD 属性:** 研究人员通过运行数十次实验来改进评估集上的表现。一旦有了评估集，强化学习（RL）就能帮助模型实现能力跃迁。
- 销售属性:** 评估集不仅是内部开发的指南针，也是向客户证明模型能力的“销售手册”。
- 瓶颈转移:** AI 模型的瓶颈已从单纯的数据量转向“如何定义成功”以及“如何衡量复杂任务的完成度”。

"If the model is the product, then the eval is the product requirement document."

— Brendan Foody

论点二：人才市场的范式转移——从众包到精选

核心观点: AI 训练数据需求已从低技能的“标注”转向高技能的“评估与推理”。

- 历史阶段:** 早期 LLM 需要大量低技能劳动力进行语法纠错和基础标注（如 Scale AI 早期模式）。
- 当前阶段:** 顶级实验室需要医生、律师、资深工程师来评估模型的逻辑推理、红线合同审查或医疗诊断能力。
- Mercor 的切入点:** 利用 AI 自动化面试和筛选过程，在全球范围内寻找那 10% 的顶尖专家，因为这 10% 的人贡献了 90% 的模型改进。



论点三：超高速增长的底层逻辑——“Can-do”文化与标准

核心观点: 极致的目标设定配合极致的执行强度是创造历史纪录的关键。



- 设定“荒谬”的目标:** Brendan 在公司营收 150 万时宣称年底达到 5000 万，最终在两周误差内达成。
- 人才密度:** 前 10 名员工必须是顶级精英（如 Scale AI 的增长负责人、多次创业者），这决定了公司的基因。
- 强度（Intensity）:** 虽然不强制 996，但强调“产出导向”的极致投入，在 AI 这种瞬息万变的市場，速度就是生命线。

数据验证结果



验证项 1: Mercor 的增长速度（16个月从 1M 到 400M+ ARR）

- 原文声称: "We grew from 1 to 400 million in revenue run rate in 16 months, fastest ascent in history."
- 验证结果:  确认。
- 来源: TechCrunch 报道了 Mercor 以 20 亿美元估值融资 1 亿美元，并确认了其惊人的营收增速，尽管私营公司具体 ARR 难以审计，但 Benchmark 的背书极具分量。
- 可信度: 

验证项 2: 评估集（Evals）在行业中的重要性

- 原文声称: "Evals are all you need." (引用 Greg Brockman)
- 验证结果:  确认。
- 来源: OpenAI 联合创始人 Greg Brockman 及 Andrej Karpathy 多次在公开场合强调评估框架是当前 LLM 开发的核心瓶颈。
- 可信度: 

验证项 3: 专家报酬水平（时薪 \$95 - \$500）

- 原文声称: "Our median pay rate in the marketplace is \$95 an hour, but it can flex up... into \$500 an hour."
- 验证结果:  确认。
- 分析: 考虑到 Mercor 服务的对象是顶级 AI 实验室，且需要的是具备执业资格的专业人士（律师、医生），这一价格区间符合硅谷专家咨询市场的行情。
- 可信度: 

四维分类评估

高度正确（已验证/权威来源）

观点 1: AI 不会取代人，但使用 AI 的人会取代不使用的人。

- 验证依据: 这一观点已成为行业共识，且在 Mercor 的人才筛选实验中得到证实：使用 AI 工具的开发者在小时内构建的产品远超传统开发者。

观点 2: 软件开发是极具“需求弹性”的行业。

- 验证依据: 随着 AI 降低开发成本，软件需求并不会饱和，反而会因为能够处理更多长尾需求而爆发。

当下可执行（有明确步骤）

建议 1: 建立企业自己的评估系统（Evals）。

- 执行方法: 识别公司核心价值链（如建筑公司的设计图输出），建立一套系统化的测试方法来衡量 AI 自动化的准确度。

建议 2: 采用“AI 协同面试”。

- 执行方法: 允许候选人在面试中使用 ChatGPT/Cursor，观察他们利用工具解决复杂问题的上限，而非测试基础记忆。

理智质疑（需验证）

存疑点: “历史上增长最快的公司”这一头衔。

- 质疑原因: 虽然 Mercor 增速惊人, 但 “历史上最快” 通常取决于统计口径 (如从 1M 到 100M 还是从 0 到 1B)。Deel 或 Wiz 也曾宣称过类似纪录。

🔴 需警惕 (可能有问题)

风险点: 对 AGI 降临时间的乐观估计。

- 风险说明: Brendan 认为 AGI 不会很快到来, 这与部分实验室 (如 OpenAI 内部部分激进派) 的观点相左。如果 AGI 突然实现, Mercor 这种基于 “人类专家评估” 的商业模式可能面临结构性风险。

🔑 关键洞察

1. **评估即产品:** 在 AI 时代, 如果你无法衡量它, 你就无法改进它。评估集 (Evals) 不仅是测试, 更是定义产品边界的工具。
2. **寻找 “市场真空”:** Mercor 的成功在于发现老牌众包公司 (Scale, Surge) 在处理 “高技能专家数据” 时的低效和对人才的不尊重。
3. **弹性需求领域:** 职业规划应关注那些 “生产力提升 10 倍, 需求会随之提升 100 倍” 的领域 (如软件、创意、研发), 而非需求固定的领域 (如基础会计)。
4. **AI 是思考伙伴:** 嘉宾利用 ChatGPT Voice Mode 进行逻辑梳理, 这表明 AI 的价值正在从 “内容生成” 转向 “思维辅助”。
5. **主动权的价值:** 创业者不应等待许可。Brendan 14 岁开始创业, 19 岁退学, 这种 “Just do things” 的行动力是所有成功的前提。

🔧 提到的工具/资源

工具 1: Cursor

- 说明: AI 代码编辑器, 访谈中多次提到作为提升开发者生产力的核心工具。

工具 2: ChatGPT Voice Mode

- 说明: Brendan 用于日常问题思考和逻辑梳理的 “思维伙伴”。

推荐阅读: 《High Output Management》(高产出管理)

- 说明: Brendan 推荐的运行公司的圣经。
- 链接: Amazon 链接

📅 行动建议

🚀 立即可做 (今天)

- ☐ **AI 工具自测:** 尝试在今天的任务中完全依赖 AI (如 Cursor 或 Claude) 完成一项原本需要数小时的工作, 记录效率提升。
- ☐ **关注 Evals:** 如果你在做 AI 产品, 列出 5 个衡量你模型好坏的 “黄金标准” 案例。

- [] **定义你的“评估集”**: 思考你所在业务的核心价值，如果由 AI 来做，你会用什么样的“打分表 (Rubric)”来评价它？
- [] **人才盘点**: 评估你的团队中，谁是那 10% 能够利用 AI 产出 10 倍价值的精英。

深入探索

- [] **研究 RLAIIF**: 深入了解“AI 反馈强化学习”，这是 Mercor 提到的行业大趋势。

★ 评分

知识价值: 9/10

- 深刻揭示了 AI 产业链中最赚钱但最少被讨论的“数据/评估”环节。

可执行性: 8/10

- 对于产品经理和创业者如何构建 AI 时代的 PRD 提供了清晰框架。

商业潜力: 10/10

- Mercor 的案例证明了 AI 基础设施服务领域的巨大天花板。

投入产出比: 9/10

- 1 小时的播客内容含金量极高，尤其是关于增长和文化的分享。

综合评分: 9.0/10

参考来源

- Lenny's Podcast 官方网站
- Mercor 官方网站
- Benchmark 投资公告

来源: Lenny's Podcast

嘉宾: Brendan Foody

生成时间: 2024-10-24