

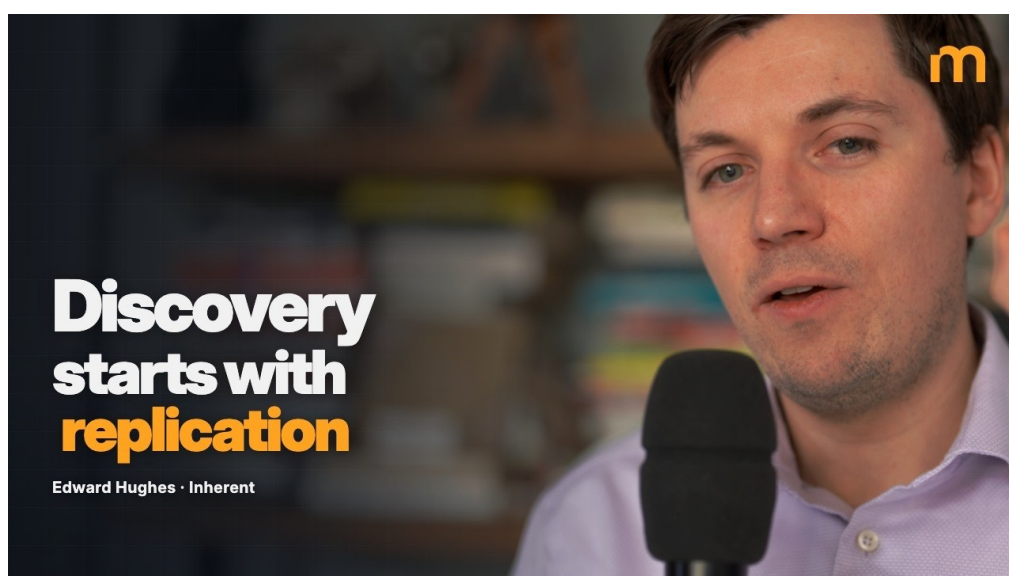
学习讲义

构建 AI 科学家：除了智能之外究竟还需要什么

从创造力的定义、论文复现到递归型公司

cnfjlhj & Pi

2026 年 9 月 15 日



来源类型：播客访谈（中配翻译版 / 英文原片字幕）

作者/频道/节目：Machine Learning Street Talk（主持人 Tim Scarfe）· 嘉宾 Edward Hughes

发布日期：2026-09-11（原片） / 2026-09-13（中配版）

时长：约 2 小时（原片 122 分钟）

来源链接：<https://www.youtube.com/watch?v=P4bYjTJvD28>

中配入口：小宇宙《跨国串门儿计划》#720

目录

引子：第 37 手到底算不算创造？	2
1 从弦理论到 AI 科学家：Ed 的历程与 Inherent 的创立	2
2 创新不等于创造力：第 37 手与创造力的三层结构	3
3 创造力是满意决策，不是优化	4
4 扩展适应、Tristan 和弦与情境创造力	4
5 Deutsch：难以变通的解释与”复制即创造”	5
6 开放式探索、局部目标与世界模型的不完美	6
7 Replica：把论文复现变成可训练的任务空间	6
8 Faraday：小模型指导大模型学到科研直觉	7
9 从复现走向创新：Transformer 是怎么诞生的	8
10 让 GRPO 在非可验证任务上稳定工作	8
11 权重与脚手架：AlphaEvolve、DGM 与 EvoTune	9
12 递归公司与群体智能	10
13 开放式组织：发电机类比与 OKR 的替代	10
14 总结与延伸	11

引子：第 37 手到底算不算创造？

2016 年 3 月，AlphaGo 与李世石的第二盘棋进行到第 37 手。这一手棋在人类围棋知识里几乎“不该下”，现场解说一度以为是程序失误。但几十手之后，它被证明是决定胜负的关键一手。

将近十年后，在 Machine Learning Street Talk 的一期访谈里，主持人 Tim Scarfe 对嘉宾 Edward Hughes 说：“你提到第 37 手，我想你和我会同意那绝对是创造性的（creative）。”

Ed 的回答出乎意料：“我不认为第 37 手是创造性的。我认为它是创新的（innovative），但不是创造性的（creative）。”

这个区分不是咬文嚼字。它是整期访谈的起点，也是 Ed 在 Google DeepMind 思考多年、随后创办 Inherent 的核心问题：如果连“第 37 手算不算创造”都说不清，我们怎么构建一个能做出真正科学发现的 AI？这期访谈从这个问题出发，一路走到创造力的三层结构、论文复现如何成为创新的课程、一个小模型如何指导大模型学到“科研直觉”、以及为什么未来的关键不是单个超级 Agent，而是“递归型公司”。

本讲义按原片 26 个章节的顺序复原这场对话，并在关键处补充教学解释和外部核验。所有时间戳对应英文原片（YouTube P4bYjTJvD28，约 122 分钟）；中配版（《跨国串门儿计划》#720）因剪掉赞助和冷开场而约 102 分钟，章节起止略有偏移。

1 从弦理论到 AI 科学家：Ed 的历程与 Inherent 的创立

内容复原（00:01:46–00:09:14）

Edward Hughes 的专业背景是理论物理学，博士阶段研究弦理论，具体方向是粒子散射。弦理论里有一个核心概念叫**对偶性（duality）**——它是两个不同理论之间的数学映射，可以把一个理论里非常难的计算，映射到另一个理论里容易得多的计算。Ed 说，这段训练让他学会了“如何转化问题”，而这恰恰是后面讨论创造力时会反复回到主题。

博士末期，他开始怀疑继续做散射振幅是否是对人类知识最有贡献的事。有意思的是，他当时引用的许多作者——Jared Kaplan、Jeffrey Pennington——几乎在同一时间做了类似的选择。2016 年，他看了 AlphaGo 与李世石的比赛，被第 37 手震撼，确信未来会涉及能辅助人类做出发现的 Agent。2017 年他加入 Google DeepMind，最初做多智能体强化学习，因为他把**文化演化**看作宇宙中最快的智能生成过程，想把文化演化“蒸馏”进 Agent。后来他训练了 Adaptive Agents 模型（当时从头训练的最大 RL Agent，5 亿参数）、手写了 Genie 1 的单文件实现，并带领一支名为“AI Scientist”的团队，与 Louis Kirsch（Jürgen Schmidhuber 的博士生）合作了一年多。

2025 年年中，他和 Louis 决定创办一家全新公司 Inherent，理由有三：第一，能做出发现的 Agent 不只是单个 Agent 的问题，而是它与整个科学家生态互动的方式，这需要激进的组织设计，在大公司里很难做；第二，最大的范式转移往往来自把一个领域的知识搬到另一个领域，所以 Inherent 要做“横跨所有科学的水平智能层”；第三，给 AI 科学家和人类一样的 affordances 和 context 需要从底层重建基础设施。Inherent 已从 Index Ventures 和 Radical Ventures 融资 5000 万美元，投后估值 2.25 亿美元，目前团队 11 人。

⁰节目讨论时间区间：00:00:00–00:01:08（冷开场）及 00:13:47–00:20:39（第 37 手论证）。原片为英文自动字幕，人名与术语有 ASR 噪声，讲义已校正。

三种学习的时间尺度

Ed 和 Tim 在这里建立了一个贯穿全场的框架：知识积累发生在三个时间尺度上——**权重的慢更新**（类比 DNA 系统发育演化）、**上下文内的快更新**（类比个体生命周期内的适应）、以及**文化积累**（最快的知识积累形式）。Ed 补充：未来不会有一个“只属于 Agent 的文化”和一个“只属于人类的文化”，超级智能将是人类与 Agent 深度互动的产物。他还提出一个介于权重和上下文之间的灰色地带——Faraday 论文里“较弱编码 Agent 把较强编码 Agent 当工具用”时，既更新了弱模型的权重，又往强模型的上下文注入了信息，”这到底是权重还是上下文？答案是都是。”

本章小结：Ed 的个人路径（弦理论 → DeepMind → Inherent）不是偶然，而是围绕“如何让 Agent 做发现”这一条线的持续推进。Inherent 的三个创业理由——组织设计、水平智能层、基础设施——为后半场“递归型公司”的讨论埋下伏笔。

2 创新不等于创造力：第 37 手与创造力的三层结构

内容复原（00:13:47–00:20:39）

Ed 对“第 37 手是创新的，但不是创造性的”给出了精确定义。**创新（innovation）**是把“未知的未知”变成“已知的已知”——你不能在已经知道要找什么的情况下产生创新，它必须是出乎意料且有价值的。**创造力（creativity）**则比创新多一步：你必须意识到自己做出了此前不知道的东西，并据此更新自己的知识。这是一种**元元认知（meta-metacognition）**。

那么是谁意识到第 37 手是了不起的一手？不是 AlphaGo，而是现场的解说员——他们一开始以为是误算，后来才确认这是精确的正确一手。

Ed 引用了 Mihaly Csikszentmihalyi（心流理论的提出者）在《创造力与发明的心理学》一书中的框架：创造力有三个组成部分——(1) **创造者个体**（creative individual）：我们通常只关注这一部分，但它只是冰山一角；(2) **领域**（domain）：一套符号规则，创造者向其中添加内容，或打破其中一条规则从而扩展可能性的空间；(3) **领域评判者**（field）：其他个体组成的群体，他们决定创造者的贡献是否被接纳进领域。

Ed 用 15 世纪佛罗伦萨的文艺复兴来说明第三点的重要性。为什么是佛罗伦萨而不是其他意大利城市？不是 DNA 突变（时间太短），也不只是领域（罗马也在重新发现古希腊罗马的建筑技术，却没有同样的创造力爆发）。关键在 field——佛罗伦萨有大量富裕家族（Medici 是最著名的），他们把“让佛罗伦萨成为最美、最有文化的城市”当作城邦的心理防御，于是大量创造性作品被接纳进领域，匠人之间形成竞争。

Move 37 与 AlphaFold 之间没有直接的线

Ed 指出一个常被忽视的事实：第 37 手是一个了不起的创新，但它没有告诉你如何把创新迁移到其他领域。我们没有立刻看到从第 37 手到发现新材料的一条线；即便在同一个组织内部，从 AlphaGo 到 AlphaFold 之间的线也并不直接——“不是 AlphaGo 的训练技术直接启发了 AlphaFold”。如果存在一个**可泛化的发现引擎**，它可以在天气建模里做发现，然后意识到“这个发现里用到的神经网络架构，也许适用于蛋白质设计”——这种可泛化的连接才是加速发现速率的关键。

本章小结：创造力 = 创新 + 元元认知 + 领域接纳。把创造力只归结为“个体”是片面的；field（评判群体）是被遗忘的第三要素。这也解释了为什么 AI 要做出被认可的创造力，不能只靠单个 Agent 在盒子里跑——它需要嵌入一个能评判和接纳的文化。

3 创造力是满意决策，不是优化

内容复原 (00:20:39–00:25:06)

Ed 接着论证：演化显然是创造性的——它产生了自然界的巨大多样性。而演化用的机制恰恰是**满意决策 (satisficing)**，即“满足约束”。为什么是满意而不是优化？很多人以为演化在优化“适应性最好的个体”，但演化真正要求的只是个体存活并繁殖。一旦做到这一点，就没有更多可“优化”的了（顶多加一个二阶的“让后代也存活繁殖”，但这仍然是约束满足问题）。

这个观察有两个推论。**推论一**：如果你真的相信创造力是满意决策，那么机器学习里默认依赖的“拐杖”——**优化**——对于构建有创造力的 Agent 来说就是找错了方向。**推论二**：如果约束满足对创造力很重要，那么打开新创造空间的方式就是**打破一部分现有约束**。为什么是“一部分”？因为如果你打破所有约束，意义就消失了——我们构造意义、甚至构造物理定律的方式都依赖于“事情可重复”（归纳原理）。打破所有约束会得到一个白噪声世界。但打破一部分约束很有用，因为有些约束来自我们现有的理论：我们无法直接接触宇宙，只能通过观测和测量工具，以及我们自己的神经解释装置。放松关于“解释”或“工具”的约束，就能接触到关于宇宙的新洞见。

Kenneth Stanley 的“为什么伟大不能被计划”

Tim 在这里引用了 Kenneth Stanley 的《Why Greatness Cannot Be Planned》：当你朝着某个目标优化时，由于“戴着眼罩”，很难找到有趣和有创造力的东西。Ed 同意，并把这一点与开放式探索（open-endedness）联系起来——后文会展开。

本章小结：创造力的机制是满意决策（满足约束后继续积累变化），而不是朝固定目标优化。这直接暗示了当前 ML 训练范式的一个根本张力：我们用优化作为默认工具，但创造力可能需要一种不同的驱动方式。

4 扩展适应、Tristan 和弦与情境创造力

内容复原 (00:25:06–00:30:56)

Ed 用两个例子展开“打破约束如何在情境中产生创造力”。

例一：扩展适应 (exaptation)。演化生物学里，adaptation 是因为给个体带来优势而被选择保留下来的特征；**exaptation** 则是一个原本有某种优势的适应，意外地在别处找到了第二个用途。科学发现里比比皆是：Alexander Fleming 把培养皿忘在外面发现青霉素；微波炉源于雷达测试时口袋里的巧克力棒融化了；Ed 作为 ML 研究者最喜欢的例子是 GPU——最初为游戏开发，结果恰好是训练卷积网络、进而训练所有现代神经网络所需要的。

例二：Tristan 和弦。Ed 有一个半专业的音乐人副业。他举了瓦格纳歌剧《特里斯坦与伊索尔德》开场著名的“Tristan 和弦”。这个和弦本身并不新——同样的和弦在几百年前就出现过。新的是**情境**：它出现在曲子最开始，在你还没确立调性时就响起，和声上极度模糊，作为听众你立刻产生疑问“这个和弦在说什么？”在此之前的音乐史里，和声更多是旋律的伴奏；而瓦格纳在这里把和声本身当作**直接的交流手段**，用模糊性传达悬念、困惑、即将到来的混乱，以及一丝希望。这预示了 20、21 世纪和声被用作色彩和情绪的整个发展——我们今天在电影配乐里对此再熟悉不过：一个场景的前几秒和弦就能告诉你这是惊悚、希望还是爱情。

创造力必须被置于它所在的”正典”中评判

Ed 总结道：创造力必须”站在巨人的肩膀上”来评判，必须 situated 在它当前所处的正典 (canon) 之中。同一个和弦，换个情境，就可能是创造性的；一幅画是否是 slop (粗制滥造)，取决于观察者能否识别指导其创作的约束——它可能客观上是连贯的，只是观察者还没理解。这种”观察者相对性”正是 Ed 2024 年 ICML 论文里开放式系统定义的基础。

本章小结：创造力不是”产生了什么新东西”，而是”在已有约束的情境中，打破了哪些约束、产生了什么新的可被识别的意义”。Tristan 和弦说明，新的不是材料，而是情境；扩展适应说明，新的用途往往来自约束的意外迁移。

5 Deutsch：难以变通的解释与”复制即创造”

内容复原 (00:30:56–00:42:27)

Tim 把话题引到 David Deutsch 的《The Beginning of Infinity》。Ed 用这本书做了两段论证。

第一段：良好解释是”难以变通”的。Deutsch 说科学是寻找关于宇宙的良好解释，而良好解释必须**难以变通 (hard to vary)**。Ed 用日出的例子对比：如果解释是”太阳被神用战车拉着升起”，那么日出变晚时你可以补丁说”神累了起晚了”或”神生气了”；日食时可以说”神在施罚或让你多睡”。这个解释可以任意变通，因此没有真正约束力。相反，如果解释是”地球自转产生昼夜”，那么日出变晚就迫使你提出”地球自转变慢了”——但昼夜周期仍是 24 小时，矛盾——于是你被迫进入更有创造力的空间：地球不仅自转，还倾斜着轴绕太阳公转。日食又迫使你假设另一个天体（月亮）在特定时间与整套理论自洽的规律性挡住太阳。**越难以变通的解释，越有预测力，越能迫使我们提出更精确的结构。**

第二段：复制本身就是创造性的。Deutsch 在书的后半部分讨论了两个谜。谜一：我举起手，你能复制这个动作——但你只有视觉线索，没有我的本体感受、肌肉信息或神经回路，却要在我体内重现它，这在认知上极其困难。谜二：约 1 万到 4 千年前，人类技术创造力在考古记录里出现爆炸性增长，但这段时间远不够生物演化起作用——那么大脑里的生物前提为什么之前就在却不被使用？Deutsch 的解决方案是：**正是复制这一行为本身是创造性的。**要传递一个想法（无论物理的还是文化的），你必须重现别人大脑里的东西，这需要你自己的创造力。而那 6000 年里改变的不是个体，而是 field——社会领导者开始重视不仅为复制、也为做新事而展现创造力的人。

浅复制 vs 深复制

Tim 在这里引入了反例：影印机 (photocopy) 和伪造者 (forger) 只复制表面的颜色和笔触，却缺失内部结构、抽象意图和动机——这是浅复制。Michael Thomas 指出，人类累积文化依赖于共享意向性、教学、规范性和棘轮 (ratcheting)，而不只是复制。Ed 回应：正因为人类之间无法直接传递权重（信息瓶颈），复制才逼出了创造力；而 AI 的问题在于我们可以影印开放权重模型的权重——这种约束的缺失反而让创造力更难达到。因此他把自己（以及越来越多的人类）的角色定义为”**约束工程师**”：设计 Agent 之间的接口，促使新颖性和创造力，而不是任由系统退化为只生产摹本。

本章小结：良好科学解释的标志是难以变通；创造性的复制不是表面摹写，而是在信息瓶颈下重现意图。Ed 选择从”复现”出发训练 AI 科学家，正是因为复现天然携带了约束和瓶颈，而正是这些约束逼出了创造力的前体。

6 开放式探索、局部目标与世界模型的不完美

内容复原 (00:42:27–01:00:37)

这一段覆盖了多个相互关联的论题：强 Moravec 悖论、事后评估、Picbreeder 的局部目标、欺骗性目标与世界模型。

强 Moravec 悖论。经典的 Moravec 悖论是说：对人类显得复杂的事（下棋、围棋）反而对 AI 较容易，对人类显得简单的事（在新厨房泡咖啡）对现代机器人仍很难。Ed 提出**强版本**：位于文化演化树最尖端的知识（蛋白质折叠、天气预报、材料设计）将是我们最先学会用 AI 有效解决的；而演化树很早期的东西（生命起源、共生起源、自催化反应）反而会是最后才解决的。

评估的边界在前瞻还是回溯？Ed 越来越确信，当前 AI 的前沿在于**如何评估**。目前大多数评估是前瞻的 (in foresight)：有人设想一个能力，预先编码环境和奖励函数（可验证奖励）。这擅长产生能满足人类预设目标的 Agent，但不擅长训练**自己提出目标、提问而非答题**的 Agent。要训练像科学家一样行为的 Agent，必须把评估翻转成回溯的 (in hindsight)：不预设期望，而是在 Agent 做完之后由人类或 Agent 群体评判——就像 PhD 答辩，导师不会在第一天写好三个问题、四年后考你，而是四年后由同行评审你呈现的工作。

问题的可规范程度谱。问题处于不同的规范程度：最具体的是形式化数学 (Lean prover)，你能规范判断证明是否正确；最不具体的比如“如何解决气候变化”——即便有人给出精确的十年程序，你也无法先验地评判其质量。开放式探索看到的世界是：你无法预先提出这些具体问题。两个可行的方向：一是在不同具体问题之间跳来跳去 (Map-Elites 擅长的)；二是建立一条**欠规范的课程** (curriculum of underspecification)，用语言模型作为“有趣性”模型 (如 OMNI、OMNI-EPIC)。

Picbreeder 与局部目标。开放式探索的核心不是“没有任何目标”，而是**没有单一的全局目标**。Picbreeder 实验里的人不是随机乱点的醉汉——他们有自己的局部好奇心和内部连贯性，但不受全局目标引导。Ed 强调：可以有局部目标，局部目标是好的；关键是不要被单一全局目标锁死。

为什么发现需要“欺骗性目标”与世界模型的不完美

这是全场最深刻的论证之一：如果你想让 Agent 做出科学发现，它必须在目标具有**欺骗性 (deceptive)** 的空间里运作。为什么？假设 Agent 有完美世界模型，那么它做的每个实验结果都在预料之中，不会产生发现；拥有完美世界模型反而让你无法判断目标好不好。反过来，**要做出发现，你的世界模型必须有某种不完美**，让局部目标具有欺骗性——你走到某个点才意识到“我以为迷宫出口在这边，其实是被误导了”，于是更新世界模型。开放式探索和构建世界模型（尤其是“实验性世界模型”——预测新实验会发生什么）之间存在深层联系。

本章小结：开放式探索不是“没有目标”，而是“没有单一全局目标”；局部好奇心是允许的、甚至是必需的。科学发现要求世界模型存在不完美，从而让局部目标具有欺骗性，迫使 Agent 在遇到意外时更新认识。

7 Replica：把论文复现变成可训练的任务空间

内容复原 (01:07:36–01:12:31)

从这里进入 Ed 的论文。核心问题是：为什么从**复现 (replication)** 出发训练 AI 科学家？

复现为什么重要。论文复现是重新做原始实验的过程。它是一种公共品：给科学更坚实的基础，揭示论文里没写清的默会知识；它也可以是新创造探索的跳板——复现过程中常发现原方法的一个“褶皱”，从而开启全新研究。Ed 认为，能快速验证或证伪一个研究方向的能力，是区分“好研究者”和“真正杰出研究者”的决定性技能。

复现是通向创新的欠规范课程的第一步。让 Agent 学会复现所需的技能——判断该做哪些实验、批评自己的实验过程、搜集自己不知道的默会信息——正是设计和实施自己实验、推进前沿所需的同样技能。

Replica 任务空间的设计

任务定义：取一篇研究论文，让语言模型把其中一个图表 (figure) 从 PDF 中不可逆地涂黑 (redact)，然后给 Agent 一小时、一块 H200 GPU 的 1/17 切片 (用 NVIDIA MIG 多实例切片)，要求它仅凭论文里的文字描述重新生成该图表，且不许访问原图。

评分：用另一个前沿编码 Agent 作为裁判 (judge)，裁判拿到原始 Agent 的指令加一套检测作弊的指南，并与人类评估对照验证一致性。

规模：初始套件含来自 100 篇 ML 和 AI-for-science 论文的 310 个任务，领域横跨 NLP、材料科学、天气预报等。

Ed 坦承，他们刻意选了高被引、知名的论文，因为这些论文更可能可复现 (否则早被无数人试过了)。这也让他们暂时“躲开”了“如何判断一篇论文是否可复现”这个棘手问题——但随着任务空间扩大，这会成为绕不开的难题。

本章小结：Replica 把“复现论文图表”变成一个可在固定时间、固定算力下自动评分的任务空间。选它作为训练起点，是因为复现所需的技能与创新所需的技能是同一套——复现是“欠规范课程的第一步”。

8 Faraday：小模型指导大模型学到科研直觉

内容复原 (01:12:31–01:18:26 及 01:20:38–01:23:37)

Faraday 的架构。Ed 团队开发了一个名为 Faraday 的 Agent，做法相当不寻常：取一个小模型 (Qwen3.6 27B)，让它把**前沿编码 Agent**当工具使用 (“coding agent as tool”)。对这个 27B 小模型在 Replica 上做大量 rollout 进行后训练。

训练与测试：训练 242 个任务，测试 68 个 held-out 任务。held-out 任务刻意不在同一 AI 领域——训练用经典 ML 论文 (后训练、CNN 架构、开放性、LSTM 等)，测试用 AI-for-science 论文 (用 AI 建立世界模型)。

结果：Faraday 不仅超过了它所使用的编码 Agent 工具 (说明它从工具里“挤”出了更多能力)，还超过了其他前沿编码 Agent (如 Claude) 和前沿开放权重模型 (如 GLM 5.2)。为排除“只是 prompt 不好”的质疑，他们跑了 GPT-5.5 Codex 的 prompt 优化循环 (类似 Karpathy auto-research)，让提示不断累积规则“别作弊、要严谨、别 10 分钟就停”——prompt 优化只微弱提升了 Codex 成绩，Faraday 仍有显著优势。

Faraday 到底学到了什么？两个深度复现的例子

Voyager 例子 (训练任务)：Voyager 是关于在合成游戏里构建技能获取库的论文。Claude 的最佳复现是硬编码一个预定义技能库——但这违背了 Voyager 的核心思想 (系统应自己编码并复用技能)。Faraday 更忠实地维持了“不仅使用技能库，还要在线构建技能库”的思想。

GNoME 例子 (held-out 测试任务)：这是一篇材料科学论文，图里有生成模型的误差棒和不同物理条件下的消融。Codex 的最佳复现只跑了一个 seed，无法提供误差棒，也省略了条件消融；Faraday 则严格 adhered 到严谨科学家的规范，既给了误差棒也做了条件分析。这正是“实习生起步研究科学家”应有的 forensics。

外部核验：Inherent 官方研究页（2026-08-14 发布，论文 arXiv:2608.13331）确认 Faraday 为 27B 参数，超过 Claude Opus 4.8 和 GPT-5.5；Replica 初始套件 310 任务来自 100 篇论文，领域涵盖 NLP、材料科学、天气预报。

本章小结：Faraday 证明了一个反直觉的范式——**把前沿大模型当工具、让小模型在权重里学到“如何指导工具”**——能在 held-out 的跨领域任务上超过大模型本身。学到的不是某篇论文的知识，而是一种“通用的复现能力”或“科研直觉”。

9 从复现走向创新：Transformer 是怎么诞生的

内容复原（01:15:31–01:18:26）

如何从“擅长复现”走向“做出创新”？Ed 给了一个他喜欢的思路：假设你极其擅长复现任意论文，那么如果你还能**想象一篇不存在的论文**——比如把 Transformer 论文的图改成“同样的结果但样本效率高一倍”——然后让你的复现 Agent 去“复现”这个其实全新的结果，Agent 就会拼命去实现一个实际是创新的图。

Ed 讲了他与 Transformer 论文共著者的对话，还原了 Transformer 的诞生过程：他们最初想做的其实是**复现**用 RNN 做的机器翻译结果，但不用 RNN——约束是“只用卷积网络”。于是他们用 convnet 替换 RNN 做翻译，逐步拼接部件。某个时候，一位同事过来说“我基于 Bahdanau 几年前的注意力机制做了个东西放在 Google 代码库里，要不要试试？”他们把它加进来，有帮助。后来他们消融掉其他所有部件，发现只剩注意力机制就够了——于是有了标题“Attention Is All You Need”。

创新

Ed 的总结是：Transformer 最初想做的只是“复现”，然后通过一系列步骤对之前论文的约束做了不同的施加。这让我们看到：一个好的复现 Agent 可以与人类协作，真正加速创新速率——不是靠“凭空发明”，而是靠“复现 + 想象性约束修改”。

本章小结：复现与创新之间不是断裂，而是一条连续的路径：复现现有结果 → 想象对约束的修改 → 让复现 Agent 去“复现”这个被修改的（实际是新的）目标。Transformer 的历史本身就是这个路径的案例。

10 让 GRPO 在非可验证任务上稳定工作

内容复原（01:29:09–01:39:43）

训练 Faraday 最大的技术挑战是让 GRPO（Group Relative Policy Optimization）在**非可验证任务**上稳定。Ed 把这段经历称为“RL crisis”——“什么都跑不通”，而且其他公司也有过自己的 RL crisis。

困难来自三重叠加：（1）奖励来自 LLM 裁判，而 LLM 是随机生成模型，从 rollout 到 rollout 对同类行为的打分有固有噪声；（2）任务是长时程的（最终阶段每次 rollout 至少一小时）；（3）行为是多轮的：27B 模型可以用 Unix 工具、可以调用 Codex、可以访问互联网下载东西到容器——动作空间几乎和人类等价。这三者叠加导致 GRPO“跑得好一阵子然后崩溃”。

解法一：对同一 rollout 用同一裁判**多次评分取平均**——相对标准的降噪。

解法二：逐轮信用分配（per-turn credit assignment）。让裁判除给出 rollout 级总分外，再对 rollout 内每一轮给一个权重（正数，归一化为和为 1），并按每轮 token 数做归一化（防止一个很长的轮次不合理地放大该 rollout 的影响）。然后用这些权重在 GRPO 里调整每一轮的 advantage，而不是 rollout 级。

逐轮权重揭示了什么？

Ed 做了一点可解释性分析：裁判最终给 **rollout 中段的轮次** 更高权重——这正是“承重的”部分（类比人类：开头是热身，中间是做关键决策的心流，结尾是收尾）。**27B 模型在提示 Codex 的轮次** 也获得更高权重——因为“决定让 Codex 做什么”是真正承重的决策。逐轮信用分配的好处是：Qwen 本来就擅长的事（如读 PDF）不会被反复更新，训练梯度集中在真正区分好 rollout 与坏 rollout 的关键轮次上，从而买到稳定性。Ed 还提到，他们最终停止训练不是因为崩溃，而是因为要发论文了。

本章小结：非可验证长时程多轮任务上的 GRPO 容易崩溃。两个关键解法是多次评分降噪和逐轮信用分配。后者的可解释性分析显示，裁判自然地给“承重决策”（中段、提示工具的轮次）更高权重——这与人类科研工作的时间分布惊人地相似。

11 权重与脚手架：AlphaEvolve、DGM 与 EvoTune

内容复原（01:39:43–01:45:45）

Tim 提到当前很多人（如 Gary Marcus）在为“神经符号 AI”宣告胜利，并指出近期大量疯狂的 harness engineering（脚手架工程）。他问 Ed：你有没有考虑过走 harness 路线？

Ed 的回答很明确：这篇论文正是对 **harness 路线的反应**。他们非常刻意地不构建 harness。他的联合创始人 Louis Kirsch 做过一项分析，发现 harness 工作的一个规律：**大约在你构建 harness 三个月后，基础模型就能做到 harness 能做的事**。所以他们想探索一种不同的范式。

但 Ed 也承认 harness 不是全坏。**权衡**在于：能力放在权重里更灵活、更可泛化；能力放在 harness 里更样本高效。如果你的目标是解决一个很具体的问题（比如 AlphaEvolve 优化 4x4 复矩阵乘法），构建 harness 可能是最好的——这就是神经符号 AI 解决特定问题的威力。Ed 列举了 AlphaEvolve、Darwin Girdle Machine (DGM)、Jenny Jang 的 HyperAgents 都属于 harness engineering，它们在解决特定问题上很出色，但**往往不能泛化**。

Ed 团队想要的，是一个能应用到完全不同领域、困难长时程问题（复现一篇完全不同领域的 ML 论文）的系统，所以他们选择改权重。但他认为两者可以结合：他们的研究 fellow Anna Serena 去年写的 **EvoTune** 就是 harness engineering 上面叠 RL——可以类比 AlphaGo：搜索（符号部分）加 RL 蒸馏（神经部分）。Ed 团队很兴奋的下一步方向是：**训练时用 harness 提升 rollout 内的表现，然后把能力蒸馏回权重**，从而兼得灵活性和样本效率。

“coding agent as tool”只是冰山一角

Ed 把这个范式放在一个更大的图景里：目前的单编码 Agent 只是开始，未来可以想象 Agent 群 (swarm)、用更大模型做微调、AI 科学家系统自己构建专用模型（可能本身是神经模型、技能、或技能 +harness+ 神经模型的组合）当工具使用。“coding agent as tool”是一个巨大冰山露出水面的一角，水下是大量不同 Agent、技能、神经模型和符号装备的交互，以及与人类的交互。

本章小结：harness 在解决特定问题上强且样本高效，但不泛化；权重更灵活但样本效率低。EvoTune 展示了两者结合的可能——训练时用 harness，推理时蒸馏回权重。Faraday 选择改权重，是为了追求跨领域的泛化。

12 递归公司与群体智能

内容复原 (01:45:45–01:55:46)

Inherent 如何在公司内部使用 Faraday? Ed 说他们的使命是递归自我改进、发现新知识,但他们对**递归自我改进**的理解与大多数组织很不同:他们把它看作**公司层面的现象**。这意味着持续闭环,把 Agent 放在一切工作的核心。

具体做法从给 Agent 与人类相同的 affordances 和 context 开始。人类主动尽快采纳 Agent——已经内部使用 Faraday 做复现来推进研究,并从使用方式中学习以加速模型构建和未来研究。公司还积累了大量 context (代码、运营方式等)。

相变时刻:从”烦人”到”主动有用”

Ed 描述了一个关键时刻:公司头几个月,Agent 主动找上门来帮忙,但老实说”挺烦人的——它们完全不知道怎么帮我们”。但大约两三个月前,他们跨过了一个**相变 (phase transition)**:Agent 对公司上下文足够了解、有足够多的 affordances,它们主动做的事变得真正有用了。这时你就有了一个杠杆——可以扩展 Agent 数量,从而扩展公司做有用事的速率。这就是 Ed 说的”**递归公司**”:公司整体作为人类与 Agent 互动的函数自我改进,而不是建造一个在隔离盒子里神奇地完成一切的单一 Agent。

群体智能:多对多。Ed 认为下一个时代是**群体智能 (collective intelligence)** 的时代,而不是**个体智能的时代**。目前大多数人用 Agent 是一对一 (偶尔一对多),但 Inherent 认为这是”有些贫乏的”用法。如果人类社会只靠一对一关系运作,文化演化的速率会被严重拖慢。他们把使用 Agent 和发展群体智能看作**多对多**问题:如何构建让多人与多 Agent 在周期的不同点有机介入的界面。

Ed 还把这个愿景延伸到物理空间:我们今天在办公室的工作方式看起来和 1980 年代末没什么两样——人们到公司、坐在各自的电脑屏幕前,而我们现在有能独立做大量工作、可大规模扩展的模型。这不可能是最优解。Valve 用”带轮子的桌子”做了小实验,但真正的问题是什么是”为 AI 研究从底层重建工厂”。

本章小结:Inherent 的递归自我改进不是单个 Agent 自我提升,而是公司整体在人类与 Agent 互动中自我改进。关键是跨过”Agent 主动有用”的相变点。下一个时代是多对多的群体智能时代,这要求重建的不只是数字界面,还有物理实验室和办公室空间。

13 开放式组织:发电机类比与 OKR 的替代

内容复原 (01:55:46–02:01:51)

发电机类比。Ed 用 1890 年代电动发电机的发明做了一个精彩的历史类比。工厂主用电动发电机替换了大型蒸汽涡轮,获得了一点效率提升——但工厂是为单一动力源 (大蒸汽涡轮) 配置的,全靠复杂的齿轮和滑轮把动力传到各台机器,一旦动力故障全厂停工,还不得不建得又高又窄、又暗又不安全。真正解锁巨大生产力增益的,不是换动力源,而是**重组整个工厂**:在每个工位放独立发电机,然后发明流水线。好处是:去掉所有齿轮滑轮;单台故障不影响其他;工厂可以水平布置、开天窗、自然采光,工作环境大幅改善。Ed 说这个类比直接对应 Inherent 正在做的事:**如何从底层重建 AI 研究的工厂,把 AI Agent 放在中心?**

OKR 的替代。Ed 在 Google 待了将近九年,深知 OKR (目标与关键结果) 对 Google 成功的作用——它是前瞻的目标导向过程,先识别目标再朝着可量化的关键结果努力。但开放式探索的哲学与 OKR 在某种程度上相反:如果目标周期太长,你就锁死了在过程中适应和改变方向、走意外踏脚石的能力。Inherent 正在探索的问题是:**递归自我改进时代的 OKR 等价物是什么?**不是基于优化的

OKR，而是基于开放式探索的组织范式。

在位者能否适应？ Ed 说他不会创办新公司除非认为有做出显著不同的事的机会。现有大公司在已有的垂直领域会继续极其成功，但正在出现一个 AI 辅助、AI 加速的 R&D 新市场。他引用 Nicholas Bloom 的论文《Are Ideas Getting Harder to Find?》：无论用研究员生产率、开发一种新药所需的通胀调整后十亿美元、还是诺贝尔奖得主做出发现时的年龄来衡量，所有指标都在朝错误方向走。原因是**知识负担 (burden of knowledge)**：作为物种成功的受害者，我们积累了太多知识，到达任何领域前沿的时间和精力已经大到个体无法跨领域掌握足够多。Inherent 想做的，正是**横跨所有科学的水平智能层**，为下一代自主实验室、下一代 R&D 组织、甚至下一代公司构建提供动力。

本章小结：电动发电机的历史教训是——换工具不够，必须重组整个工厂。OKR 是优化时代的组织范式；开放式探索时代需要一种新的组织范式。知识负担让“想法越来越难找”，而水平智能层是 Inherent 对这个问题的回答。

14 总结与延伸

全场的核心脉络可以从一条线来把握：Ed 从“第 37 手算不算创造”出发，区分了创新与创造力，论证创造力需要元认知和领域接纳 (Csikszentmihalyi 三要素)；进而指出创造力的机制是满意决策而非优化 (演化论证)，打开新空间的方式是打破一部分约束 (Deutsch 的难以变通解释)；在此基础上，他论证科学发现要求世界模型存在不完美、局部目标具有欺骗性，评估必须从前瞻翻转为回溯 (PhD 答辩类比)。

技术落地落在 Replica 和 Faraday 上：Replica 把论文复现变成可评分的任务空间 (涂黑图表、一小时、1/17 H200)，Faraday 用 27B 小模型把前沿编码 Agent 当工具，在 held-out 的跨领域 AI-for-science 任务上超过 Claude、Codex 和 GLM 5.2。GRPO 的稳定化靠多次评分降噪和逐轮信用分配；从复现走向创新的路径是“复现 + 想象性约束修改” (Transformer 的诞生就是案例)。

组织愿景落在“递归公司”和“群体智能”上：递归自我改进是公司层面的现象，不是单个 Agent 的事；关键是跨过“Agent 主动有用”的相变点；未来是多对多的群体智能时代，需要重建的不仅是数字界面，还有物理空间。OKR 是优化时代的组织范式，开放式探索时代需要一种新的组织范式——这是 Inherent 正在探索的。

几条值得追问的开放问题：

- (1) **复现的泛化边界：**Faraday 在 held-out 的 AI-for-science 任务上表现出色，但这些仍是“高被引、知名”的论文。当任务空间扩大到 12000 篇会议论文时，可复现性的判断和作弊检测会变得多难？
- (2) **小模型指导大模型的极限：**如果小模型本身需要足够“聪明”才能有效指导大模型，那么这个范式的 scaling law 是什么？什么时候该停止增大“控制者”模型？
- (3) **创造力的三种要素能否被 AI 完整实现：**如果 field (领域评判者) 必须包含人类，那么 AI 的创造力是否永远是“人类-AI 共生的创造力”而非“AI 独立的创造力”？Ed 似乎倾向前者。
- (4) **递归公司的可控性：**当多个 Agent 主动在组织中做事，如何防止“bad behaviors emerge”？Ed 坦承这是“未知的未知”，他们还没解决。
- (5) **满意决策与当前训练范式的张力：**如果创造力是满意决策而非优化，而我们几乎所有的训练都基于优化目标，这是否意味着需要一种根本不同的训练范式？Ed 没有给出答案，这可能是最重要的开放问题。

延伸阅读：

- Inherent, *Training AI Scientists to Replicate Research*, arXiv:2608.13331 (2026-08-14)
- David Deutsch, *The Beginning of Infinity* (2011)——良好解释、复制即创造
- Mihaly Csikszentmihalyi, *Creativity: The Psychology of Discovery and Invention*——创造力三要素
- Kenneth Stanley & Joel Lehman, *Why Greatness Cannot Be Planned*——目标欺骗性与开放式探索
- Nicholas Bloom et al., *Are Ideas Getting Harder to Find?*——知识负担与研发生产率
- Ed Hughes & Michael T.?, ICML 2024——开放式系统的观察者定义 (novel + learnable)