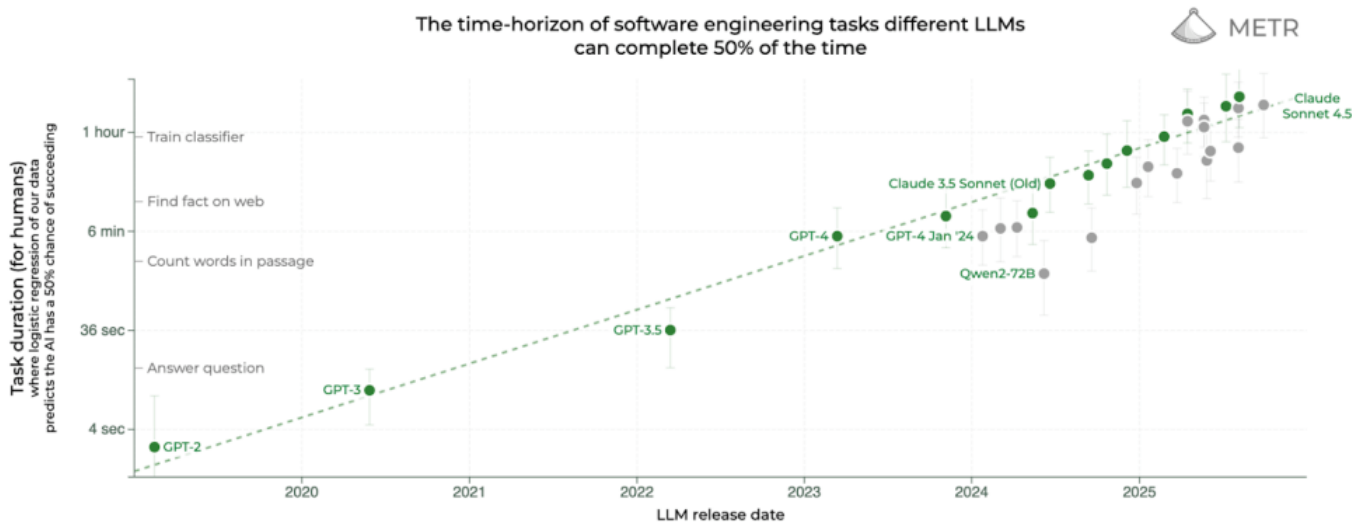
一位非经济学家对 AI 与经济学的思考

同步发布于 LessWrong

> 现代人类大约在 10 万年前首次出现。在接下来的大约 9.98 万年里，几乎什么都没有发生。嗯，也不能说完全没有：有战争，有政治斗争，还有农业的发明——但这些事情对人们生活质量几乎没有什么影响。几乎所有人的生活水准都相当于现代每年 400 至 600 美元，勉强高于生存线……
> 然后——就在几百年前，也许 10 代人之前——人们开始变得更富有，而且越来越富。以人均收入计，至少在西方，开始以史无前例的每年约四分之三个百分点的速度增长。几十年后，世界各地也在发生同样的事情。”
> —— Steven Lundsburg

METR 发表了 Kwa、West 等人的一项极具影响力的研究，主题是衡量 AI 完成长任务的能力。其主要结果是下面这张引人注目的图表：



横轴是各代旗舰级 LLM 的发布日期；纵轴是这样的能力指标：选取那些模型以 50% 概率能成功解决的软件工程任务，并度量**人类**解决这些任务所需的时间。

模型随时间进步并不奇怪；但这张图之所以非同寻常，主要在于纵轴是对数刻度。这意味着存在一个固定的时间间隔，过了这个间隔，模型能够**成功完成**的任务时长就*翻一番*。METR 估计这种“倍增时间”（与图中直线斜率的倒数成正比）大约是 7 个月；他们也指出最近可能加速了（如果只看 2024 年之后的模型，甚至短至 3 个月）。在本文中，为了简便，我假定倍增时间为 6 个月，于是“时间跨度”每年翻两番（四倍）。（本文所有数字都很粗略。）

有许多因素可能影响这些结果（METR 已经相当充分地枚举了它们）。值得将其区分为影响**截距**（模型当前能胜任的任务时间跨度的绝对值）、影响**斜率**（倍增时间），乃至影响**形状**（例如打破“模型发布时间与对数时间跨度之间的线性关系”）的因素。

影响截距的因素

诸如“GPT5 可以完成人类需要 2 小时 17 分钟才能完成的任务”之类的精确数值，可能受多种因素影响：

- **可靠性因子（↓）**——图中按 50% 成功率作图。若设为 80% 成功率，METR 得到的斜率（即倍增时间）相近，但截距显著降低，例如 GPT5 只能完成耗时 26 分钟的人类任务。（关于 100% 准确率还有一则说明，见下文“100% 视野”。）
- **任务类型（↕）**——该图针对特定基准绘制。METR 还研究了其他领域。尽管数据更稀疏，但总体仍与直线拟合相符（有时甚至更陡，不过其中一些数据点只出现在较新的模型上，而这一区段的斜率本身更陡）。
- **“基准偏置”（↓）**——在边界清晰、成败容易度量的任务中，AI 表现往往更好，而现实世界的任务要“凌乱”得多——无论在规格、所需上下文，还是成功的度量方式上。目前仍不清楚这是否只影响截距，还是也影响斜率乃至整条曲线的形状。“田野实验”迄今仍很有限，结论不一。然而，AI 实际使用量的迅速攀升表明，模型能力并非只限于实验室环境。我个人确信这会对“截距”产生显著影响——模型把基准测试转译到现实世界时要支付不小的“凌乱税（messiness tax）”

——但我不确信它会影响斜率。也就是说，“凌乱税”很可能只是一个固定的常数乘子 c<1，作用在模型所能处理的任务时长上。

影响斜率/形状的因素

- **指数级投入（↓）**：模型发布时间这条时间轴，叠加了若干以极快速度增长的投入：算力、AI 实验室人员规模、数据、AI 资本开支等。例如 Epoch AI 的估计是训练算力每 6 个月翻一番。如果指数的“底数”发生变化，进展速度也会随之改变。尤其是，从逻辑上讲，维持指数型投入会变得越来越难，且很快会不可能。不过到目前为止，这些资源的投资没有任何放缓迹象。
- **新数据/主动学习（↓）**：至今为止，LLM 基本是在学习人类生产的数据。可以类比一个学生：在 K–12 与本科阶段，他主要从课本学习——那是已经收集好的知识。然而在许多职业里，特别是在科学与发明领域，人类需要从现实世界中获取新知识与新观测。如果 LLM 的智能进步必须依赖于跑科学实验或在世界中行动，那会放慢进度。但目前并未出现这种放缓迹象，尽管 LLM 越来越“具代理性”。事实上，METR 的数据恰表明最近几年的斜率在加速。
- **物理任务与物理世界数据（↓）**：METR 主要关注软件工程。趋势或许能外推到其他认知劳动，但尚不清楚它是否能扩展到需要在物理世界行动、或从物理世界采集数据的领域。尽管机器人技术也在进步，但尚不明确它是否遵循类似的指数曲线。即便最前沿的机器人以与最前沿模型相似的速率提升，要实现大规模制造仍可能成为瓶颈。

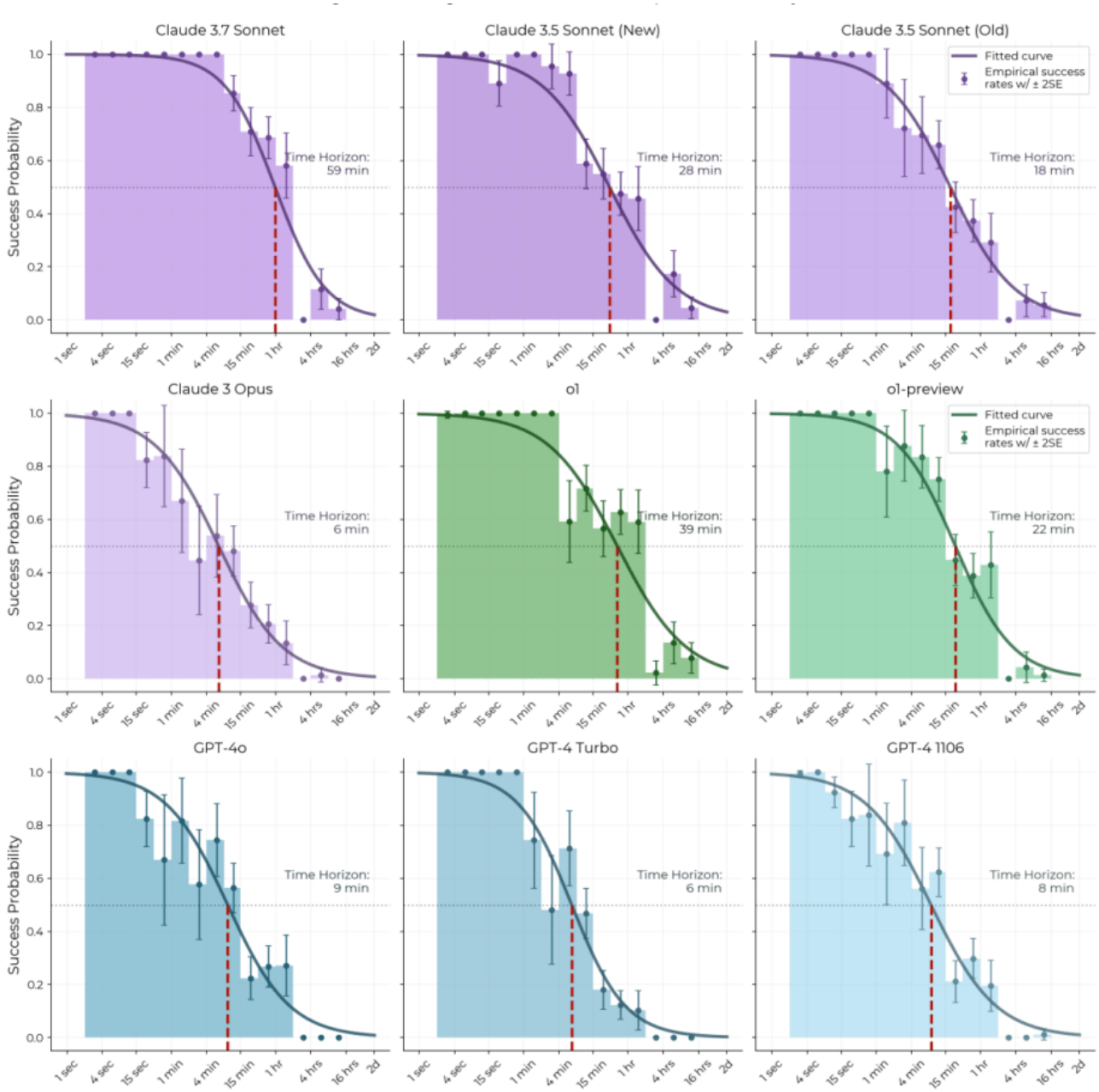
  我个人不相信所谓“数据墙”。我认为，同类数据越多回报越递减，而过去几年的进步——如 METR 图所示——主要并非源自互联网数据的数量增长。此外，尽管迄今 AI 的经济影响多发生在软件工程上，但值得注意的是，最强的软件工程模型往往是在**非常通用的数据**上训练出来的，而且它们常常在许多其他任务上也很出色（这篇文章写的是 Claude Code，但依我经验，Codex 在非编程任务上也很强🙂）。在我看来，认为 AI 的影响会局限于软件工程，就好比 2020 年 1 月认为新冠的影响会局限在中国一样。

- **阈值效应（↑）**：这些任务是相对于人类的时间尺度来衡量的。我们需要睡觉、休息，还要在多人之间分配工作，这就要求隔离完成任务所需的上下文与成功的评价方式。因此我们把任务分解到“工作日/周/月/季度”等粒度。（我相信即使是埋头攻克费马大定理 7 年的安德鲁·怀尔斯，也把任务拆成了许多中间结果。）于是，一旦 AI 达到某个时间跨度，要么（a）这种测量就不再合理，要么（b）它基本能够模拟任意多人、任意时长的组合。
- **递归自我改进（↑）**：生产 AI 模型的重要投入是人类研究者与工程师的工作。如果由 AI 自身来生产这部分投入，那么它可能极大地提高投入水平。目前尚不清楚，用 AI 来自动化 AI 研发过程，会导致奇点、斜率上升、一次性跃迁，还是仅仅帮助维持指数增长。

总之，我们对**截距**的不确定性很大，但对**斜率**（至少在达到“递归自改进/全面自动化全部人类认知劳动”这一点之前的总体形状）不确定性相对更小。下文我将直接假设：时间跨度每 6 个月翻一番，但不对“时间跨度的绝对值”作任何假设。

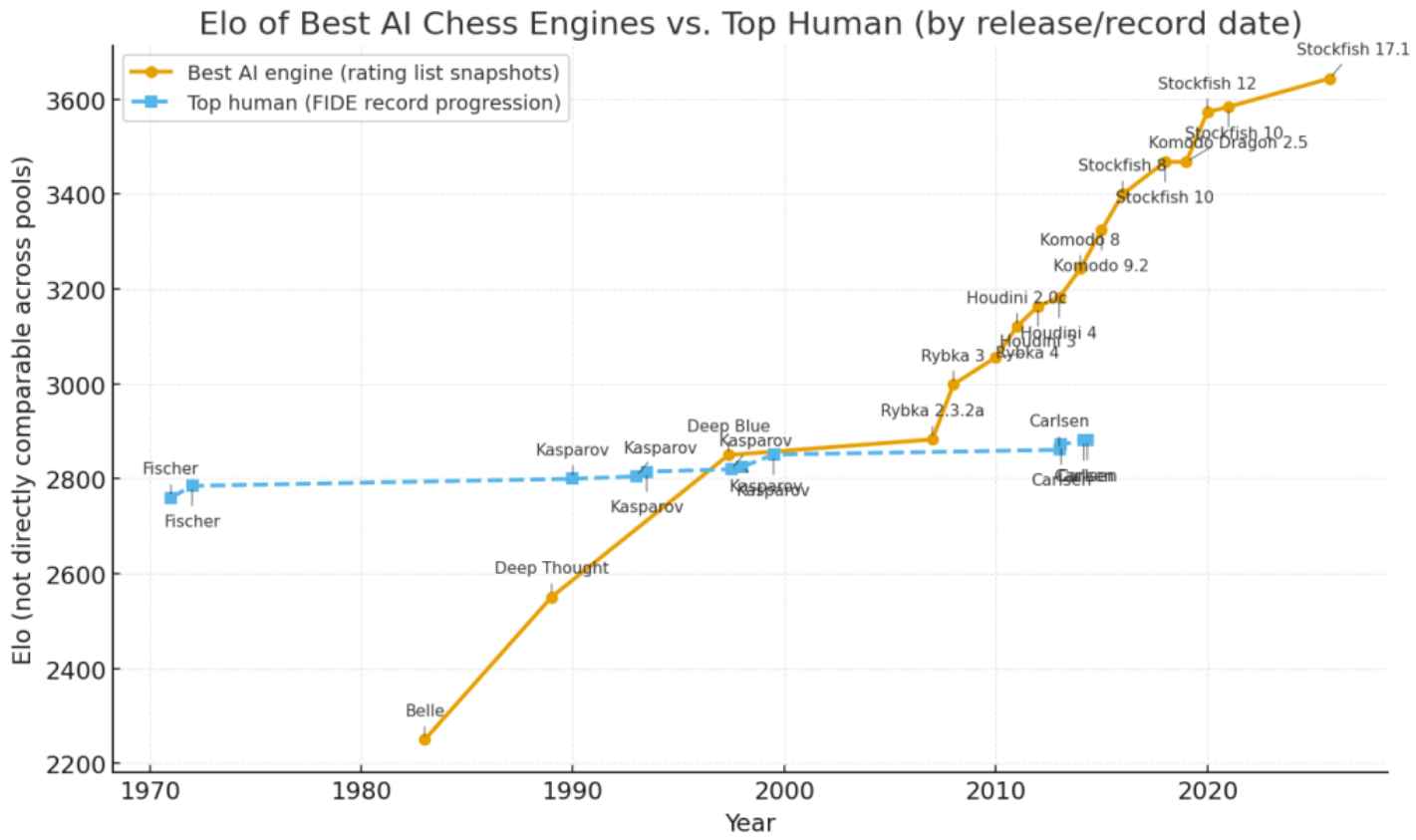
S 形关系

METR 论文中的另一幅（图 5）同样耐人寻味：



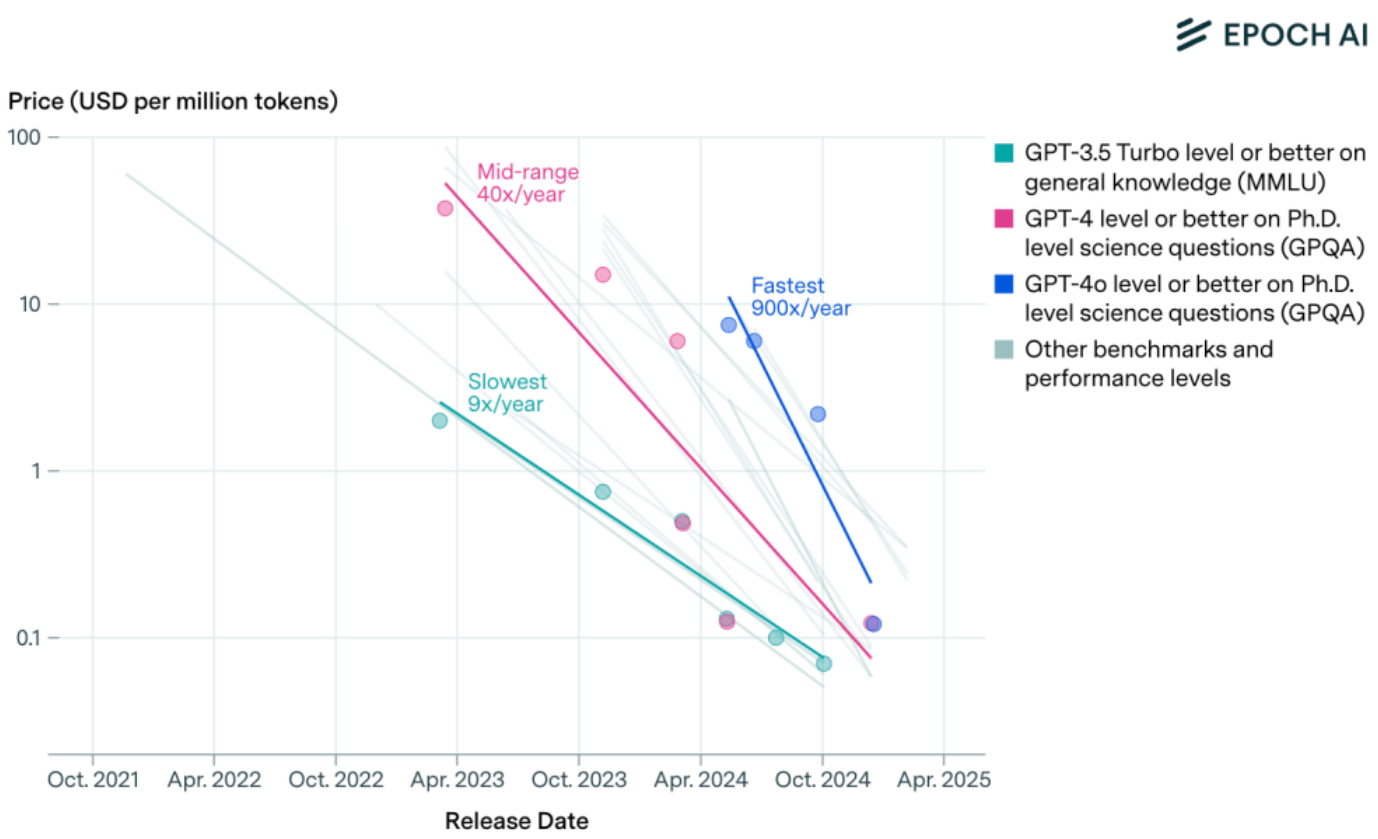
让我惊讶的是：模型成功率作为"任务的人类耗时"的函数，用"Sigmoid（S 形）曲线"拟合得极好。尤其是，看起来存在某个时间阈值：低于它时，模型几乎 100% 成功。这暗示即便是“100% 成功所对应的时间跨度”（尽管经验上很难测）也会以相近的速率翻倍。

另一种理解方式：每个任务都有一个"难度"（可由对数时间跨度捕捉），而模型有一个"技能水平"，决定它成功的概率。从这个意义看，它类似ELO 评分：我们可以把"对数时间跨度"看成任务的 ELO，若模型的 ELO 等于任务的 ELO，它"赢"的概率就是 50%。（感谢 Jacob Hilton 提供的类比。）顺带一提，下图是模型（以及人类）在国际象棋上的 ELO 随时间的变化，也呈现出类似的线性增长（不过在 1997–2007 年左右有一段长平台期）。



成本在下降

另一个醒目的统计事实是，推理价格在迅速下降。也就是说，虽然把智能前沿再向前推进很昂贵，但一旦我们达到某个水平 X，向用户提供同等水平能力的成本会以每年 10 倍以上的速度下降。看起来，第一次抵达某个前沿很难、很贵，但"第二次抵达"就便宜且容易得多（参见所谓"DeepSeek 时刻"）。如果这种趋势持续下去，这意味着一旦某项工作被自动化，**一年之内** AI 做这项工作的成本就会变得可以忽略。

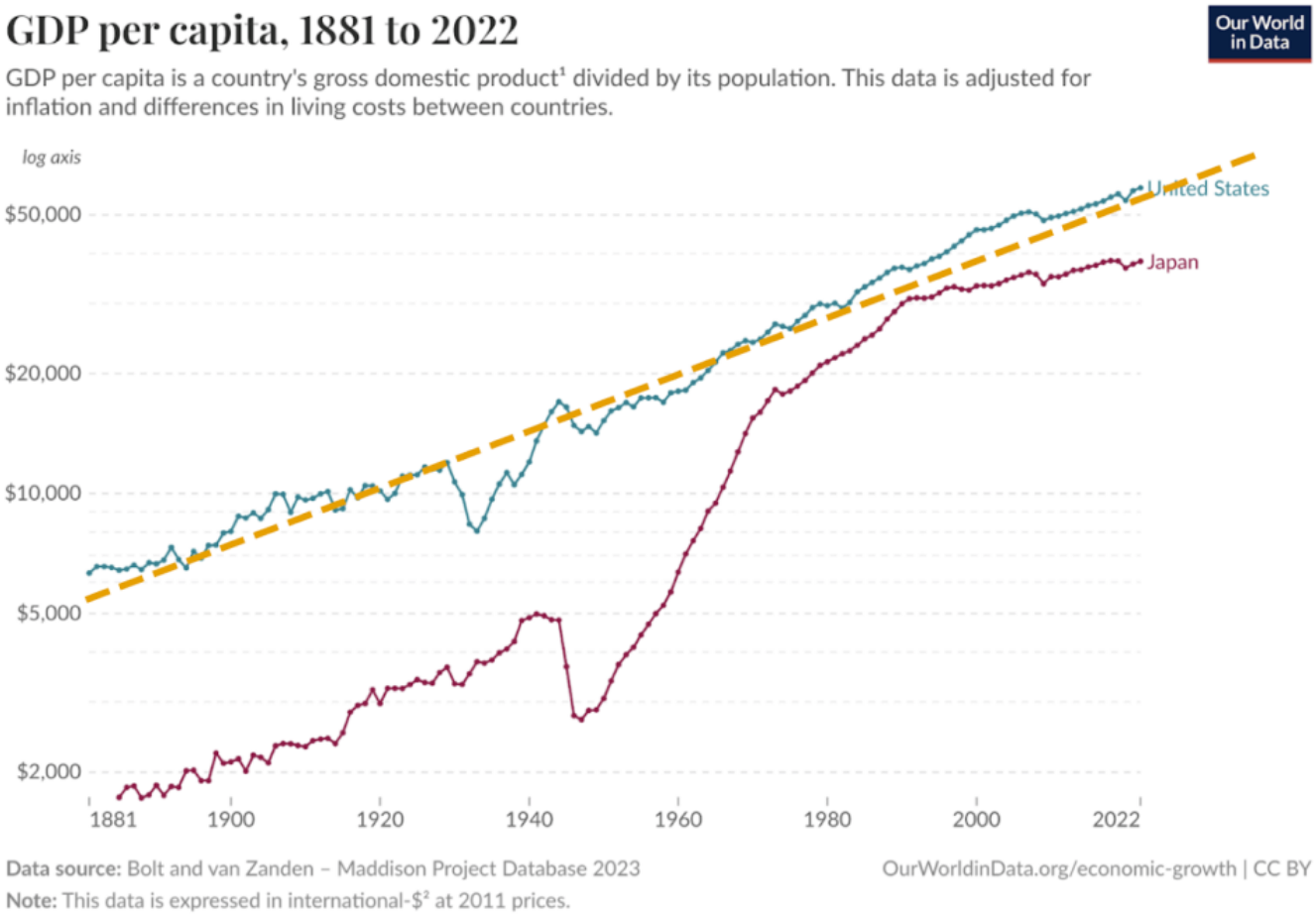


关于 AI 的讨论常假设"机器人或许是例外"，即其进步不会像"虚拟 AI 助手"那样快。这有其理由：生产成本、部署灵活性不及软件等。就我而言，并不清楚是否有根本性的理由说明“机器人可执行任务复杂度的**倍增时间**”会明显慢于其他 AI 系统；不过，据我所知，这方面尚无数据。

对 GDP 增长的启示

接下来我会严重超出自己的专业边界，但我确实想看看能否做些“信封背面算术”。由于我关注的是“**斜率**”而非“截距”，我不会试图预测**何时** AI 会带来某个增长，而是估算从“AI 对 GDP 总增长贡献达到 5%”那一时刻，走到“因 AI 使得 GDP **翻番**”要花多长时间。

本文只讨论**增长**，不涉及就业结果。在某些"刀尖上平衡"的设定下，人被替代的程度恰好抵消 AI 的生产率贡献，从而出现就业下降但无增长提升——但我不认为在"能力指数增长、成本指数下降"的假设下，这类情景能长期维持。因此，若上述趋势持续，AI 对劳动的重大影响必然伴随显著的生产率提升。



一个惊人的事实是（经通胀调整），在过去 150 年里，美国人均 GDP 大致以**恒定**的 2% 左右速度增长。期间没有任何发明——包括电气化、内燃机、计算机与互联网——改变过这条轨迹。注意，2% 的增长意味着 GDP “翻番时间”约为 35 年。

其他国家的增长波动更显著（例如日本）。通常“远离前沿”更易于高速增长；而一旦抵达前沿就会放缓。这是合乎直觉的：推动前沿前进需要**发明新思想**，而追赶前沿只需**复制与适配**现有思想。

即便如此，值得玩味的是：在“摩尔定律”对应约 40% 年增长率（指计算性能/成本）的大背景下，人均 GDP 的增速仍只是 2%（不看人均的话可能是 3% 左右）。一种解释是“**鲍莫尔成本病**”——计算机生产率大幅提升，但**人**成了瓶颈（亦见本文 II.C 节的这篇论文）。另一种解释是“好点子越来越难找”，因此要获得同样的新增产出需投入越来越多的技术资源。

关键问题是：AI 是否会打破“2%”这条惯性轨迹？AI 会不会只是另一项让我们再维持几十年 2% 增速的技术？还是说，“AI 时刻”对我们将如战后日本一般：我们好像与一个高得多生产率的“AI 前沿经济体”相遇，这种互动将带来快速增长，让 GDP 至少像日本那样每十年翻一番？

作一番对比： Acemuglu 预测 AI 驱动的 GDP 增长约为**每年 0.1%**，而 高盛 预测约 **1.5%/年**。十年翻番的 GDP 约等于**每年 7%** 增长，亦即 AI 需贡献约 **5%** 的额外增速——是高盛估计值的三倍多，是 Acemuglu 估计值的 50 倍。即便是“温和”的增长提升，带来的影响也可能极其巨大：**额外 1.2%** 的年增长就足以让美国财政走向可持续（无需加税或减支），而**额外 2%** 的年增长对美国而言将是前所未有的。

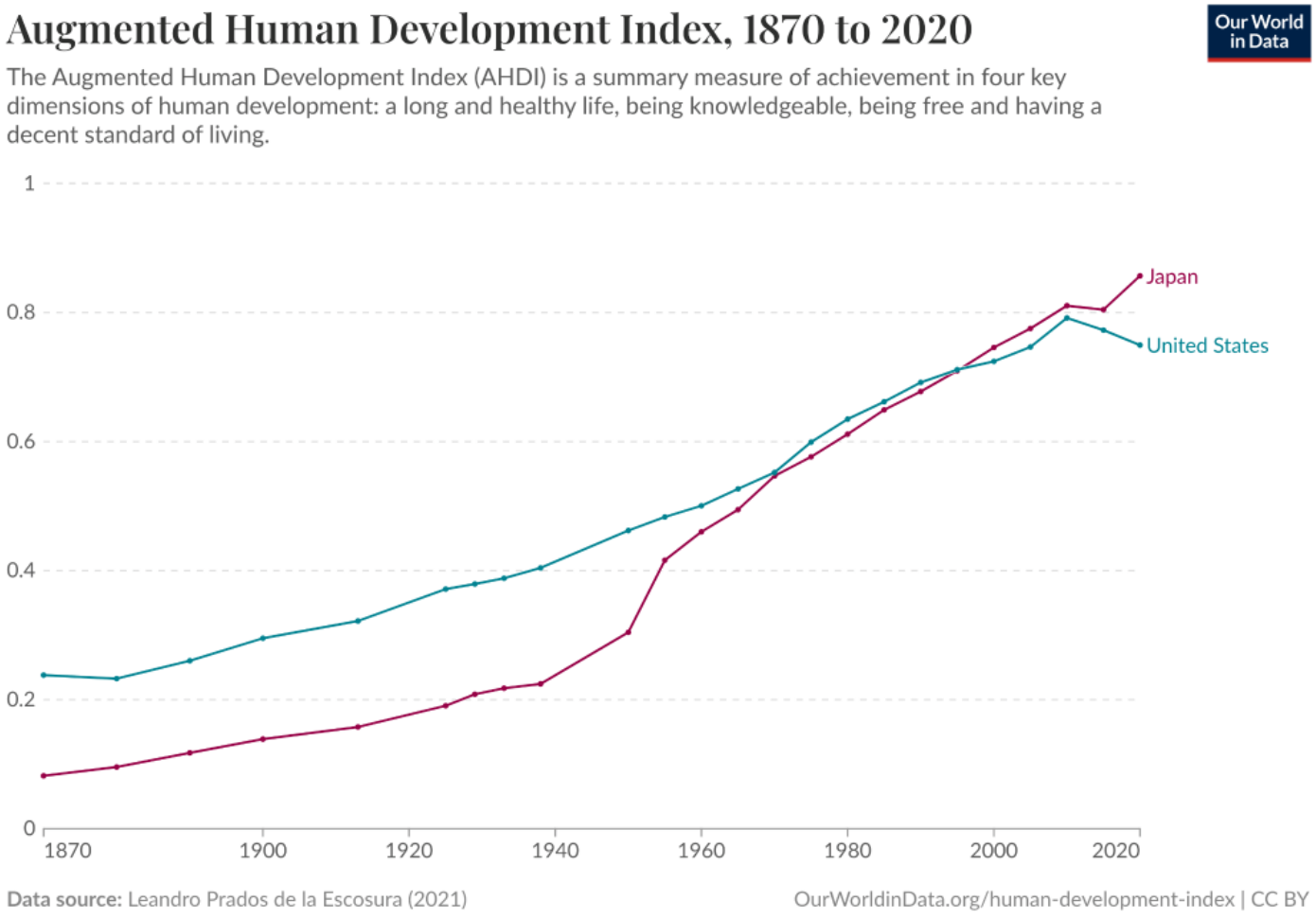
AI 能通过两条路径提升 GDP：用资本替代劳动，或提高全要素生产率（TFP）。具体而言，某些内生增长理论认为，生产率随“**思想**的产出”而增长，而“思想的产出”又与研究者/发明者的数量单调相关（但不是线性）。

若 AI 通过自动化**某个具体行业**来贡献增长，则对 GDP 的最大收益受该行业**份额**所限。具体说，若某行业占经济的比例为 x，那么（按 B. Jones 的模型，见下）**完全自动化**它对 GDP 的提升最多是 1/(1−x)。例如，软件行业占比若为 2%，则完全自动化它可使 GDP 提高到 1/0.98 ≈ **1.02**（约 **2%** 增长）。再比如，按某些口径，认知劳动至少占 GDP 的 30%（例如占劳动者收入的一半），那么**完全自动化认知劳动**可使 GDP 达到 1/0.7 ≈ **1.42**（即 **42%** 提升）。若这在十年内发生，折算成年增速约 **3.5%**。

不过，如果 AI 还能通过**发现新思想**来提高其他行业的生产率，那么它的贡献会超越那些被直接自动化的部门。需要说明的是，就 AI 对研发的贡献而言，我预期它会通过**加速科学**、让科学家、研究者与发明者更高效的方式来实现。这意味着在未来几年，资助人类科学家的投资回报率将比过去更高。

下面我会给出一些直观但不够严谨的外推，以估一估 AI 能带来多少、以及多快的增长。

当然，[也有人批评 GDP](#) 作为衡量进步的指标，并提出了各种替代指标。不过，许多替代指标至少与 GDP**宽松相关**：其中一个飙升，另一个往往也飙升。比如看美国与日本的[扩展人类发展指数](#)。



我确信 AI 的许多影响不会被 GDP 捕捉到；但如果它真的像工业革命那样具有变革性，这也会在 GDP 中显现出来。

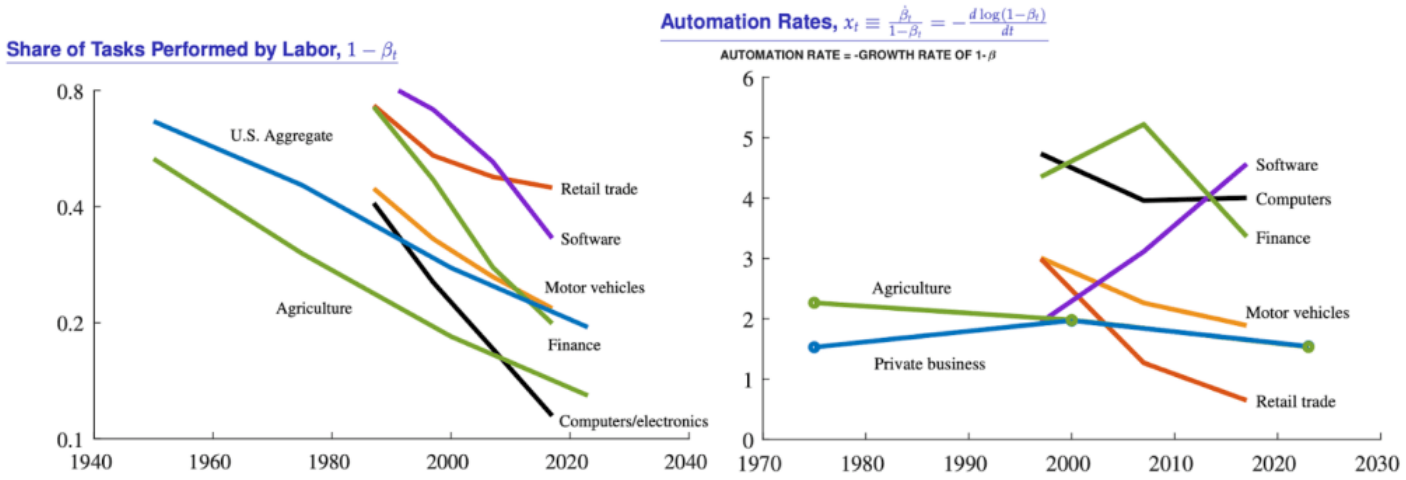
从 METR 任务得到的直觉

做一个（非常粗糙且并不真正合理的）假设：在某个行业或岗位中，任务呈“重尾”分布，即“耗时至少 T 的任务的占比”大致与 1/T 成比例。在这种情况下，“时间跨度”与“尚未被自动化的任务占比”成比例。这意味着“倍增时间”就等同于“剩余任务占比的**减半时间**”。

另一种想法是“ELO 类比”——假设任务“ELO 难度”的分布满足：难度超过 E 的任务占比大约是 $\exp(-E)$。（就国际象棋棋手而言，ELO 分布大概是正态，即 ~ $\exp(-E^2)$。）

如果我们假设“减半时间”为 6 个月，那么从 AI 自动化了某个行业**一半**任务的时点起，再过大约 **2 年**，就会走到自动化 **31/32 ≈ 97%** 任务的时点。

这个假设极具进取性，因为它只关注**能力**、忽略**扩散**。AI 在理论上也许能自动化大部分任务，但由于种种原因，实践中未必会自动化。（不过扩散也可能并不均匀，存在"潜在能力的突然解锁"。）同样重要的是，这个直觉与过去 80 年自动化的宏观趋势相反：历史上自动化总体呈**线性**推进——自动化份额缓慢、稳定上升，年自动化率为个位数，且常常是递减的。见下图（C. Jones 与 Tonetti, 2025，经由 ChatGPT）：



因此，如果 AI 导致"劳动承担的任务占比呈指数衰减"，那将打破既有趋势，与我们过往所见非常不同。

把 AI 看作“人口增加”

另一种视角：把 AI 视为在每一年 **t** 向经济中注入 **N(t)** 名新的“劳动者”，他们具有某种质量 **Q(t)**——质量可理解为“他们能完成的**经济上有用**的任务占比”，既反映“受教育年限”，也反映能力的通用性。我们可以把 Q(t) 定义为“尚未被自动化的任务占比的**倒数**”。

算法进步与总体算力预算将决定“劳动者数量/质量”的组合。做个对照：美国人口每年因自然增长与移民增加约 200 万，美国劳动力规模约 1.6 亿。

在粗略层面，若资本与技术固定，按柯布—道格拉斯生产函数（$GDP \propto K^a L^{1-a}$，a=0.4），那么把劳动规模增加一个系数 **C**，GDP 将增加 $C^{0.6}$。若 AI 新增 1000 万“虚拟员工”，GDP 将增加 $(170/160)^{0.6} \approx$ **4%**；若新增 5000 万，则约 **18%**；若把劳动力**翻番**，则约 **50%**。当然，AI“虚拟劳动者”的数量可能高出若干个数量级——到那时，这种算术大概就不太有意义了。（有人写过一本《十亿美国人》，但我想连作者都没思考过“倘若有**一万亿**美国人会怎样”。）

难以对 N(t) 或 Q(t) 做出预测，但看起来两者都将指数增长。在极端情况下，如果把质量 Q 固定，那么由于固定质量模型成本骤降，N(t) 可能**每年 10 倍**增长。或许更朴素的起点是假设“Q×N 的乘积”按 METR 的速率每年**四倍**增长，即 Q 与 N 各自**每年翻番**。（注意，N(t) 是**新增**劳动者数量，即 TN(t)−TN(t−1)；但指数函数的导数仍是指数，所以这一点对大势影响不大。）

这种增长一旦让 AI 开始提供“非微不足道”的劳动规模（例如美国境内 10 万人份），那么在**十年**之内，AI 将成为美国劳动供给的**主导来源**。

替代与自动化效应

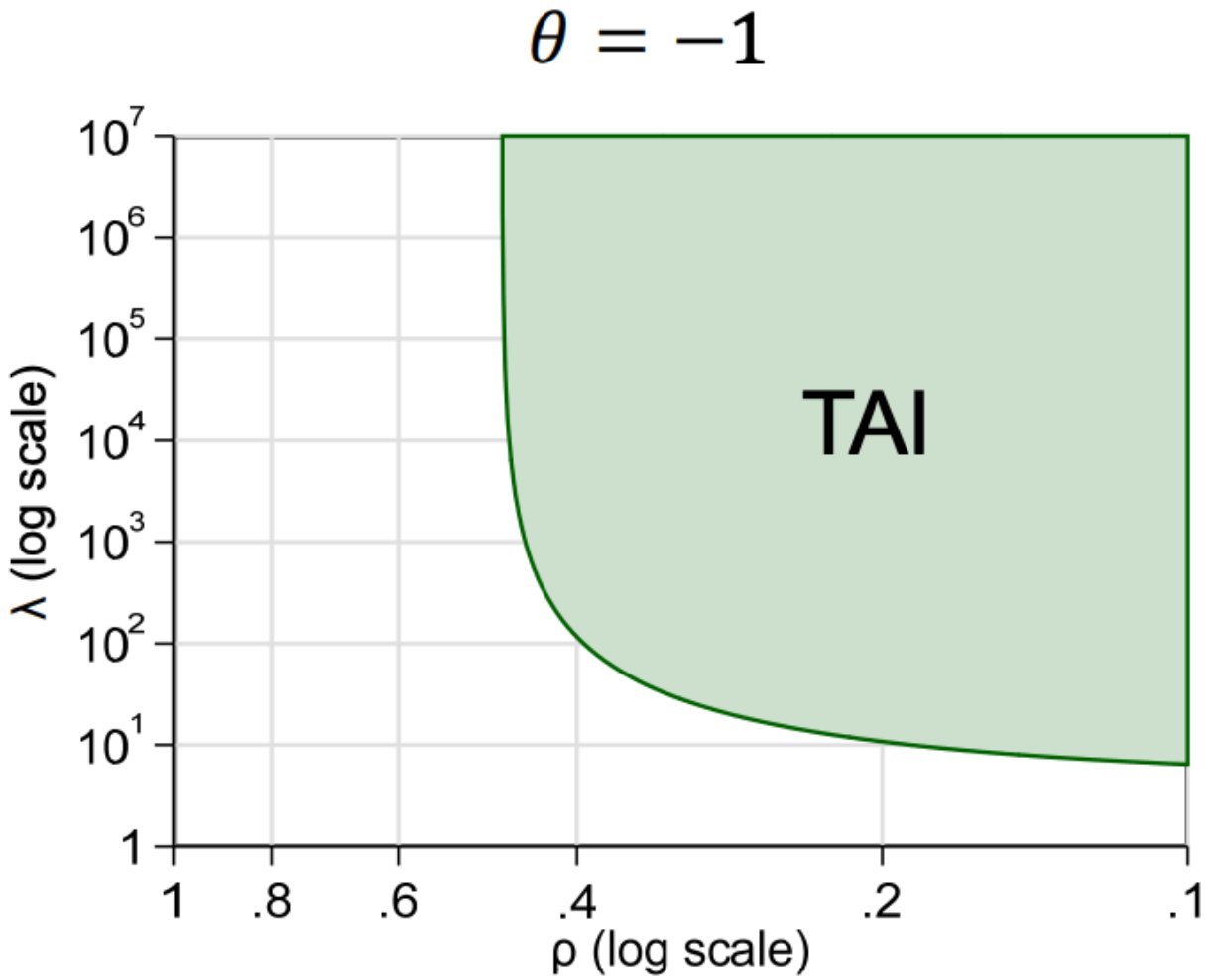
Ben Jones (2025) 的简化模型认为：AI 对生产率的影响由“**可自动化任务**”与“**不可自动化任务**”两部分的**调和平均**决定。

设有 ρ 的任务**不可自动化**，$1-\rho$ 的任务**可自动化**，且可自动化任务由 AI 完成的生产率是 $\lambda >> 1$；为简化，设不可自动化任务由人完成的生产率为 1。
则自动化带来的生产率提升可表示为带权调和平均数：

$$
\Big(\frac{\rho}{1} + \frac{1-\rho}{\lambda}\Big)^{-1}
$$

注意，与算术/几何平均不同，即便 $\lambda \to \infty$，该式的上限也只是 $1/\rho$。这符合直觉：如果任务彼此不可替代，那么把“人类耗时占比 90% 的那部分任务”自动化后，一个工人在同一时间里至多做出从前的 10 倍工作。

Jones 证明，在这个模型里，要获得显著的生产率提升，必须同时**减少 ρ**、**增大 λ**。若两者之一"卡住"，生产率也会"卡住"。他给出了达成"变革性 AI"（生产率提升 10 倍，量级可比工业革命）的参数区间图（在他的假设下，该图的条件是 $(9/[(2-\rho)/\sqrt{\lambda} + 2\rho]^2 \geqslant 10)$）：

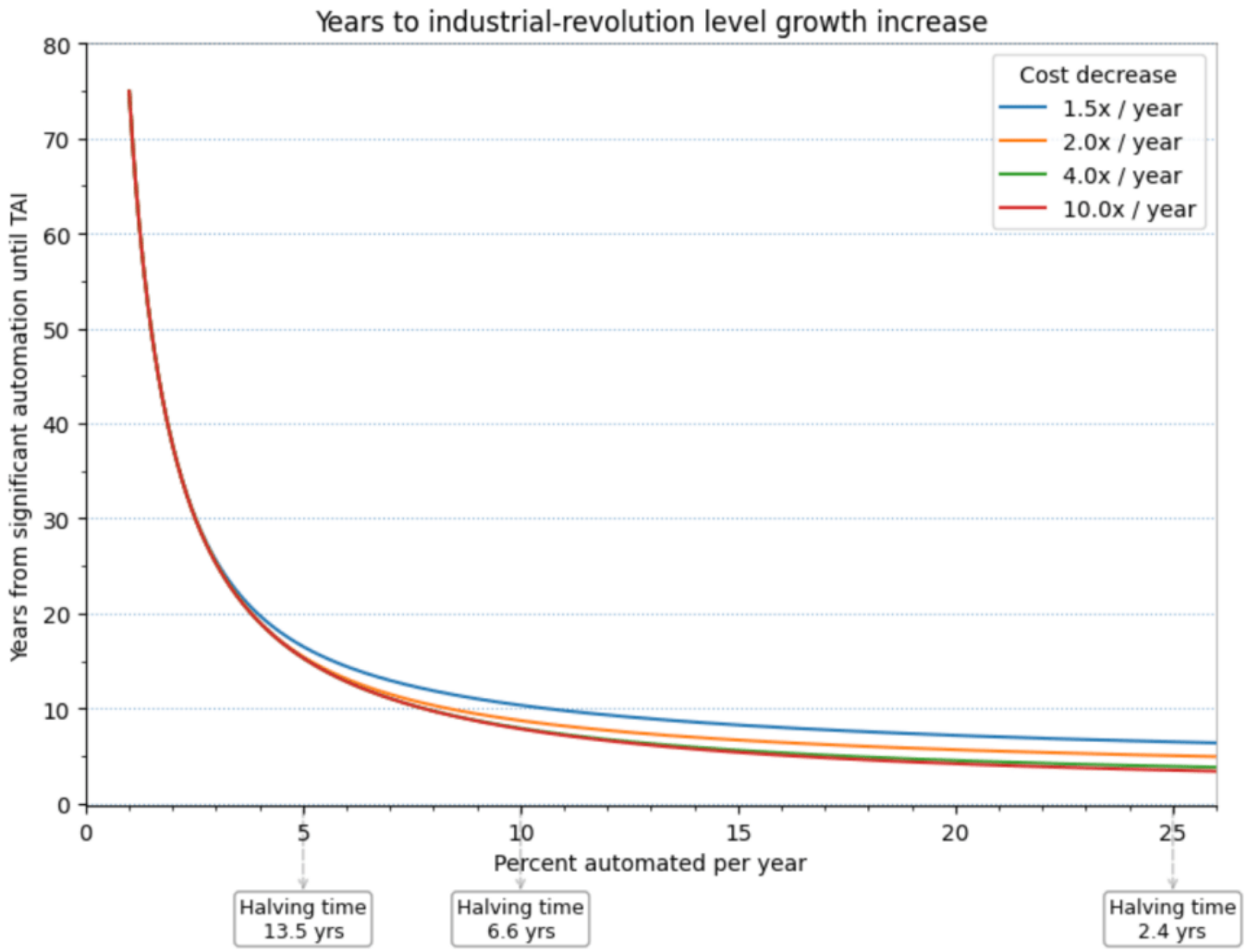


（引自 Jones 的图 5：在调和平均情形下，导致变革性 AI 的 λ、ρ 组合区间。）

可以这样想：如果（出于上述机制）我们每年都在 ρ（尚未自动化任务的比例）与 λ（AI 在可自动化任务上的相对生产率）上取得进步，会发生什么？

如前所述，这需要一点“信念之跃”，但或许并非完全离谱：也许在某个阈值（比如 AI 能稳定做满“8 小时工作日”）之后，**ρ 每年缩小 4 倍，λ 每年增大 10 倍**。以这样的速度，仅用一年，我们就能从图的左下角（1,1）走到 **(λ=10, ρ=1/4)**，几乎抵达 TAI（变革性 AI）边界。（注意 Jones 假设已有**一半任务**被自动化，我们可能需要一段时间才能走到图的左下角起点。此外，考虑“好点子越来越难找”会降低 ρ 的收缩速率——见论文第 4.4 节。）

当然，"ρ 每年缩小 4 倍"的假设非常激进，可能并不现实。或许"每年缩小到原来的 1/1.1"更合理——这意味着每年自动化掉**约 9%** 的剩余任务（而非 4 倍收缩情形下的 75%）。下图展示：从（1,1）角出发，达到"变革性 AI"所需年份，如何随"每年自动化的剩余任务比例"及"成本下降速度（λ 的增长）"而变化：



可以看到，只要速度**足够可观**，我们有望在**十年至二十年**内抵达变革性增长。而且结果对"成本下降（λ 增长）"的敏感度较低——这对"自动化体力/手工任务"也许是个好兆头。

结论：AI 是否会带来空前的经济增长，归根到底在于：它在**能力**上的指数级进步，是否会让“尚未自动化的任务占比”本身也以**指数速率**下降。

致谢：感谢 Bharat Chandar、Jason Furman、Benjamin Jones 与 Chad Jones 对本文的评论与讨论。

来源：https://windowsontheory.org/2025/11/04/thoughts-by-a-non-economist-on-ai-and-economics/