# 认知交汇

Jul 7, 2025

图片来源：“the miracle” ~ @themeloart

*关于扩展性、普适性与AI能力的思考*

上周，我和我的MBA学生们在课堂上——受一些思考启发——讨论了一个问题：不断扩大的AI模型是否会持续改进，或者我们是否正走向又一次“AI寒冬”。

有人半开玩笑地称我为“加速主义者”。我不觉得这个标签合适，但我确实相信，当前进展背后的势头是真实的、迅速的，并且不太可能停滞。以下是我向班级陈述的观点……

## 为什么规模仍然奏效

OpenAI的《神经语言模型的扩展法则》早期提供了实证证据：只要模型规模、数据量和算力同时增加，AI模型的错误率就会下降，也就是说AI可以继续“学习更多”。这个趋势在七个数量级上呈现出幂律关系。目前还未出现饱和迹象——每增加一个数量级，仍然带来改进，尽管回报在逐步递减。

“痛苦的教训”遇上“普适逼近定理” 强化学习先驱Richard Sutton将通过蛮力学习取得的进展称为“痛苦的教训”：从长期来看，利用日益廉价的计算资源的一般性方法，会胜过那些由人类设计的精巧学习技巧。

数学家们在原理上证明了类似的结论：普适逼近定理（UAT）保证，只要参数足够多，一个模型可以以任意精度逼近任何函数。虽然这只是存在性证明（并不保证你真能造出这样的AI），但它确立了一个理论上限，而我们的经验性扩展曲线正逐步逼近。

将这两个思想结合起来，就能形成一个引人注目的叙事：

现实的下限：扩展法则：更大的模型 + 更多数据 ⇒ 持续降低的错误率。

理论的上限：UAT：在无限规模的前提下，模型可以将误差降至零 —— 它可以“学会任何事”。

# 为什么我认为进展不会停滞

硬件和资本仍在流入。 传统计算范式下的算力可能开始趋于平台期，但数十亿美元级的数据中心建设计划表明，算力供应远未枯竭。

算法还在进步。 Deepseek就是一个例证，说明我们在优化软件堆栈方面还有很大空间——每一次浮点运算（FLOP）都能榨出更多能力。Alpha Tensor和Alpha Search是新算法由AI设计的早期实例。

数据供应在演化，而非枯竭。 也许Common Crawl的数据已接近极限（我并不这么认为），但我们已经看到合成数据、模拟轨迹、多模态流和具身AI的出现，用以满足模型那近乎无止境的“胃口”。

使用价值的门槛已被突破。 ChatGPT、Gemini 和 Claude 等工具已经在现实中发挥作用（至少对个人而言）；这种商业反馈回路为持续的研发提供了资金保障，也防止了我们再次进入AI寒冬。

这些都不意味着能得出具体的时间表，但模型能力的提升斜率之陡峭出人意料——这也是为何近期“AGI 2027”预测层出不穷。OpenAI前研究员Leopold Aschenbrenner在《情势感知：未来十年》中提出，未来再提升2~3个数量级的“有效算力”可能就能构建出能自动化大部分AI研究的系统。

你是否认同这个具体日期并不重要，较为安全的预测是：模型能力会继续提高——而且比大多数人预计的更快。

# 当机器超越人类思维

如果模型在未来十年内越过人类认知的界限，争论的焦点将从“它们能不能？”转向“那现在该怎么办？”

技能替代：知识型工作任务（编码、法律研究、设计）可能面临类似20世纪工厂所经历的剧变。

决策速度提升： 董事会战略、并购尽职调查或新药发现周期可能从几个月压缩到几分钟——组织若能学会将判断委托给机器，同时保持人类“在回路中”，将会获得优势。

权力集中： 对算力依赖极强的前沿模型将有利于能够负担多亿美元集群的企业或国家，这将加剧反垄断争议和地缘政治风险。

对齐与责任： 当一个大语言模型比任何医生都更准确地生成医学建议时，谁来签署处方？法律框架远远落后于技术现实。

意义与身份： 如果推理、专业知识、创造力和洞察力都变成可商品化的公共服务，那人类工作中真正独特而有价值的部分还剩下什么？

这些不再是科幻假设，而是2025–2035年间的战略规划议题。

# MBA班级的一点结语

“加速主义”有时意味着“越快越好，无视后果”。那不是我的立场。我只是观察到：• 曲线依然陡峭；• 看不到明显的停滞迹象；• 资本、人才和算力依旧不断流入，因为这项技术已经有实际用途。

相信进展会持续，并不等于鼓吹鲁莽的速度；治理和安全同样重要。但现在下注“模型能力不会继续提升”——感觉是会输的一方！

扩展法则（尽管它们不是严格意义上的“法则”，而是一种经验观察）为我们提供了一个不断提升的下限轨迹；UAT则表明理论上的上限是无限的。

从我所看到的数据来看，没有任何迹象表明我们即将撞墙——真正重要的工作，是为一个机器将常规地超越人类思维的世界做准备，并负责任地塑造那个未来。

对于下一届MBA领导者来说，这是一项令人兴奋、有时也让人畏惧的任务：负责任地驾驭这种势能，将不断进化的模型转化为现实世界的价值，同时不忘伦理和治理。

欢迎质疑——毕竟，好的反驳是检验信念的方法。但如果你今天问我，模型性能是否会持续上升并最终超越人类基准，我的答案是：会的。（许多破折号是人类自己添加的。）

# 推荐阅读与收听

主题 - 标题与作者 - 推荐理由

扩展证据 - “Scaling Laws for Neural Language Models” – Kaplan 等 (2020) 最初提出幂律关系的论文。 https://arxiv.org/abs/2001.08361

计算-数据平衡 - “Training Compute-Optimal LLMs” – Hoffmann 等 (2022) - 解释了“Chinchilla比率”。 https://arxiv.org/abs/2203.15556

蛮力哲学 - “The Bitter Lesson” – Richard Sutton (2019) - 为什么规模胜过人工技巧。 http://www.incompleteideas.net/IncIdeas/BitterLesson.html

理论上限 - 普适逼近定理 – Cybenko、Hornik 等 - 显示任何函数在理论上都可以用足够大的模型逼近。 https://en.wikipedia.org/wiki/Universal_approximation_theorem

近期AGI情景 - “Situational Awareness: The Decade Ahead” – Leopold Aschenbrenner (2024) - 详述AGI将在2027年实现的预测。 https://situational-awareness.ai/

批判评论 - “The AGI-in-2027 Thesis” – Evan Armstrong, Every (2024) - 对2027预测的平衡性评估。 https://every.to/napkin-math/the-agi-in-2027-thesis

宏观政策视角 - Nick Bostrom，《超级智能》(2014) - 探讨控制与风险的经典著作（依然相关——我第一次读这本书时也非常怀疑）。