

## 人的智能是 AI 的上限吗？（下）

上一讲我们说到，现有的大模型已经在相当程度上具有超过普通人的智能，而且在某些领域超过了专家水平。那能不能让 AI 有更高的智能呢？

其实有一个提升现有大模型智能的办法，我们前面说的所有研究都还没有用上，但是很容易用，这就是使用所谓「agent」。这个词原本的意思是有独立行为能力的人，一般翻译成“代理人”，现在中文 AI 圈称之为「智能体」。



智能体是过去大半年间特别热门的一个研究和应用领域。我们先说说这个。

## 2. 用智能体提升智能

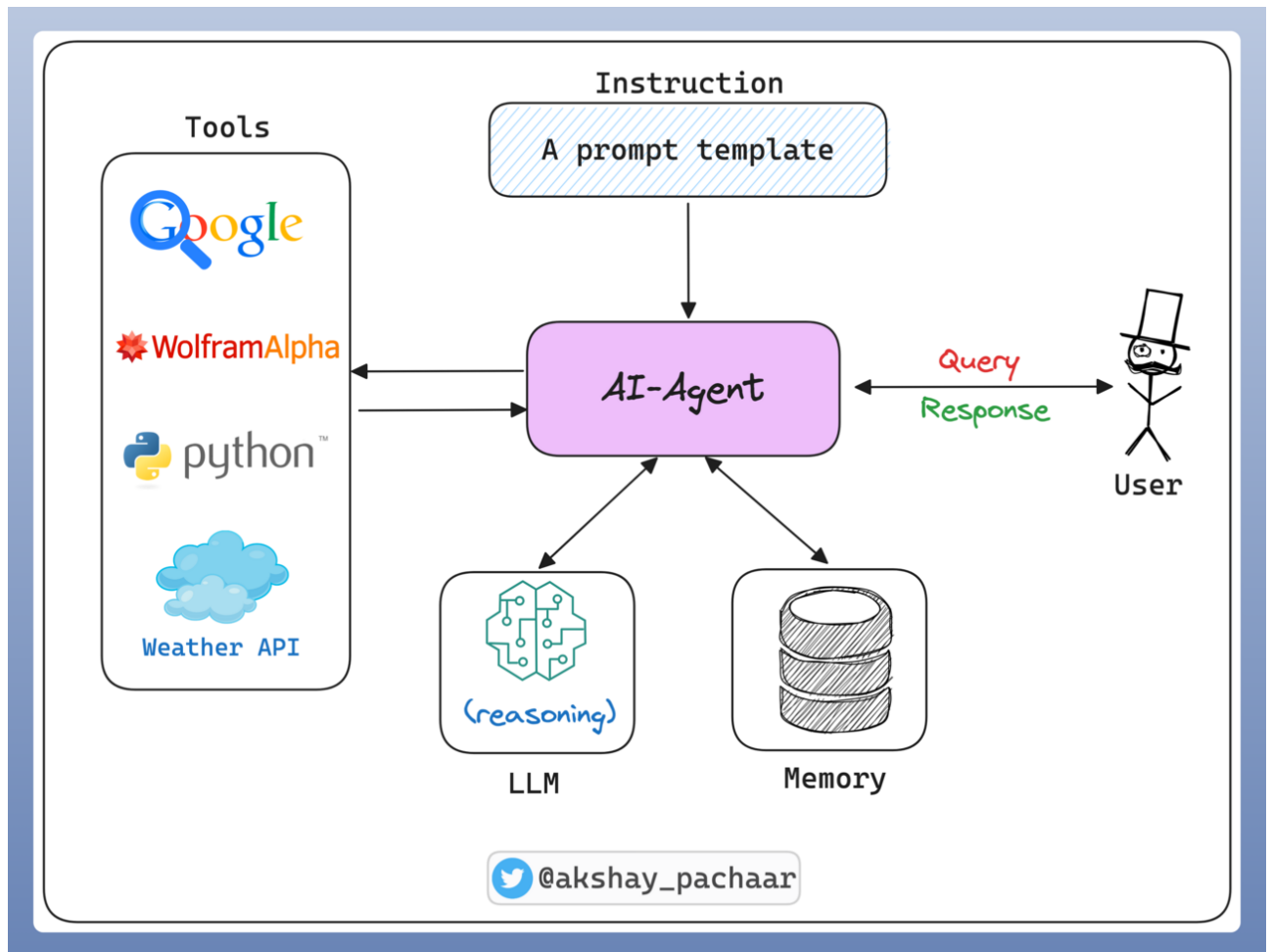
我们可以简单地把智能体理解成定制的 GPT。比如在 ChatGPT 应用商店里的那些 GPTs，就可以被视为是一个个智能体。它们已经被预装了详细的指令，可以去做一些有多个步骤、需要判断的复杂任务。你还可

以在对话中用 “@” 符号调用其他的 GPTs，让多个智能体联合为你服务。

深度学习研究者、同时也是 X 上的 AI 大 V，卡洛斯·佩雷斯 (Carlos E. Perez)，对智能体 AI 的定义是 [8]：它们可以接收外部信息，自主采取行动而不必有人类在旁边控制，可以连续地跟外界互动。

智能体最鲜明的特点是有自主决策能力。它们不只是像生产流水线工人那样按照预先设定的流程做事，他们会根据给定的目标自己制定计划，它们会调用各种工具，包括调用其他的智能体，并且在执行过程中分析现场的情况，决定下一步该怎么办，最后汇总结果。

2024 年 4 月刚刚出了一篇关于智能体的综述论文 [9]，我们这里不用详述。我看有人用下面这张图说的就挺好 ——



你提出一个总目标，智能体会调用大模型分析推理，制定计划，然后该调用工具就调用工具，该学习信息就学习信息，中间还可以跟你持续互动，最终完成任务。

曾经担任过阿里巴巴技术副总裁、后来自己创业的 AI 科学家贾扬清，举了个简单的例子 [10]，我看最能说明智能体式的自主性。比如你想知道 YC 的黑客新闻 (Hacker News) 论坛上今天最热门的帖子说了什么。直接在传统搜索引擎输入这个问题 (" what did the most popular post in hacker news say today?" ) ， 它只会匹配关键词，返回的结

果都是提到过“黑客新闻热帖”关键词的网页，而不是真正的今日热帖

---

X

All News Images Videos Shopping : More

Y Hacker News  
https://news.ycombinator.com › item

### Most-favorited Hacker News posts

Sep 2, 2020 — ... **say** it applies even more ... The globally most-favorited is  
https://news.ycombinator ... never knew “**favorite**” was a feature on **hacker news** until I ...

Best Comments - Hacker News	Aug 27, 2011
Ask HN: What's the <b>most</b> life-changing blog <b>post</b> you've ever ...	May 6, 2024
Ask HN: Your Favourite HN Comment? - Hacker News	Sep 20, 2020
Ask HN: What's your <b>favorite</b> HN <b>post</b> ? - Hacker News	Sep 14, 2016

[More results from news.ycombinator.com](#)

上面是我用 Google 测试的结果，它返回的第一个答案居然是 2020 年的一个网页，其中列举了黑客新闻的一些热帖。显然，Google 不理解你的问题。

但是如果你问 GPT-4o，它理解你！它会先访问黑客新闻实时网页，查看其中的帖子，找到哪个贴的“点数”最高，然后把内容总结给你 —

---



what did the most popular post in hacker news say today?



Searched 4 sites

The most popular post on Hacker News today is about the passing of Lynn Conway, a pioneering computer scientist and transgender activist. This post garnered significant attention with over 1600 points and hundreds of comments. Lynn Conway made substantial contributions to microelectronics and is renowned for her advocacy for LGBTQ+ rights. Her passing has sparked a wave of tributes and reflections on her impactful career and legacy.

For more details, you can visit the Hacker News [front page \(Hacker News\)](#) ([Hacker News](#)).



我去看了黑客新闻页面，发现果然最热的就是那个贴。

Y

Hacker News

new | past | comments | ask | show | jobs | submit | 2024-06-11

Stories from June 11, 2024 (UTC)

Go back a [day](#), [month](#), or [year](#). Go forward a [day](#).

1.▲ Lynn Conway has died (wikipedia.org)

1617 points by kevvok 1 day ago | 310 comments

2.▲ I like the RP2040 (dgroshev.com)

616 points by dgroshev 1 day ago | 343 comments

3.▲ Self-Serve Dashboards (briefer.cloud)

239 points by vieiralucas 1 day ago | 79 comments

4.▲ Deterioration of local community a major driver of loss c (afterbabel.com)

397 points by throwup238 2 days ago | 442 comments

这就叫自主解决问题。这件事是不平凡的。GPT 不但理解了你的意思，而且主动调用了上网工具，然后根据返回的信息进一步访问了那个帖子，最后给你生成了摘要。而这一切都是它自己的决定。

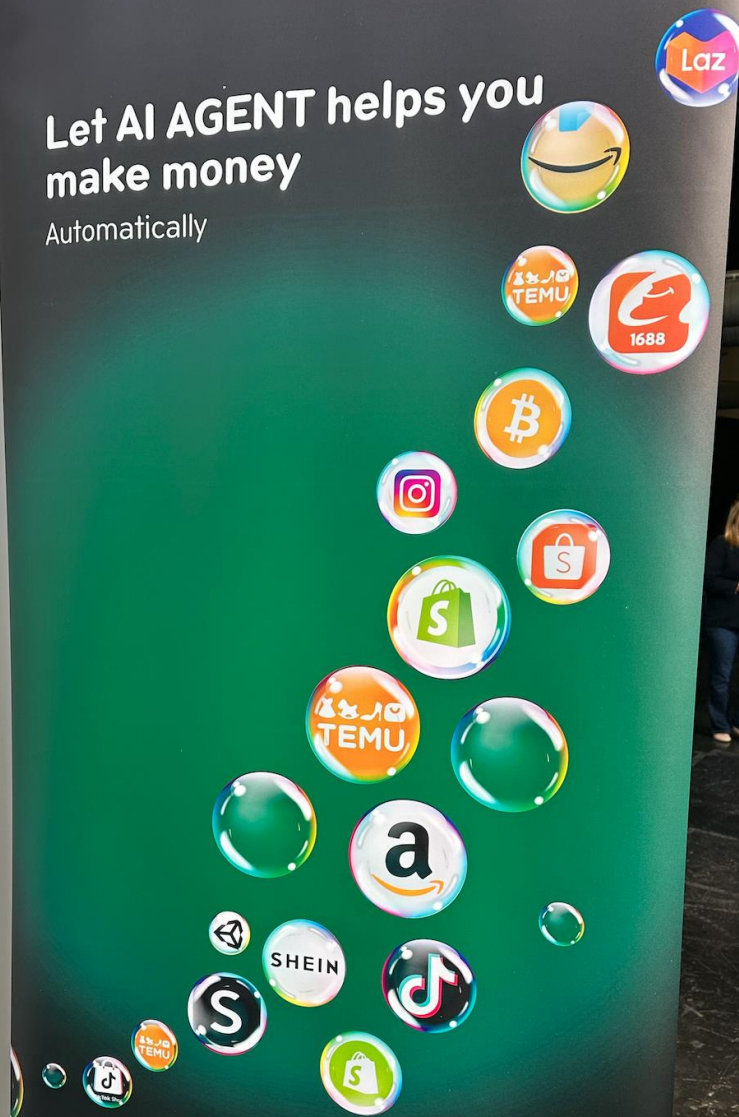
连专攻智能搜索，有志于取代传统搜索引擎的 Perplexity，也没通过这个测试。可见 OpenAI 底蕴之深。

模型还是那个模型，你只要对它进行一些定制，就能得到一个个智能体。在 2024 年 5 月底旧金山的生成式 AI 峰会（GenAI Summit）上，我就看到好几家创业公司已经把智能体用于传统业务 ——



# Agent2.AI

Let AI AGENT helps you  
make money  
Automatically



**Join our beta testing now!**

We help cross-border e-commerce  
achieve full automation by letting AI  
control your computer or Cloud VM.  
& E-commerce is just our first step 🤖

sign up on [www.agent2.ai](http://www.agent2.ai)





The image shows two large vertical banners for 'Streetbeat' and 'StreetbeatPRO'. The left banner is titled 'Streetbeat' and 'Invest like a politician with Gen AI', featuring two smartphone screens displaying stock tracking interfaces. The right banner is titled 'StreetbeatPRO' and 'The first, fastest, and most accurate AI for stock investing', with the text 'The financial world is set to be the fastest and most significantly impacted industry by the AI revolution.' Below this, it says 'We are leading the charge' and 'Learn more at streetbeatPRO.com'. In the foreground, a laptop screen displays the 'Streetbeat PRO' web interface, showing a chat window with a user asking for high-yield dividend stocks and an AI agent responding with a list of companies including Chevron, Verizon, TotalEnergies, Pfizer, and USBC Holdings.

**Streetbeat**  
Invest like a politician  
with Gen AI

**StreetbeatPRO**  
The first, fastest,  
and most accurate  
AI for stock  
investing

The financial world is set to be  
the fastest and most  
significantly impacted industry  
by the AI revolution.

We are leading  
the charge

Learn more at  
streetbeatPRO.com

**Streetbeat PRO**  
New thread  
You: Identify high-yield dividend stocks for stable income, considering current market conditions.  
Agent: Here are some high-yield dividend stocks for stable income, considering current market conditions:  
1. Chevron Corporation (CVX)  
2. Verizon Communications Inc. (VZ)  
3. TotalEnergies SE (TTE)  
4. Pfizer Inc. (PFE)  
USBC Holdings plc (USBC)

这个智能体在做跨境电商操作员，那个智能体在帮人炒股，他们的角色很不一样，但它们是同一个大模型扮演的。

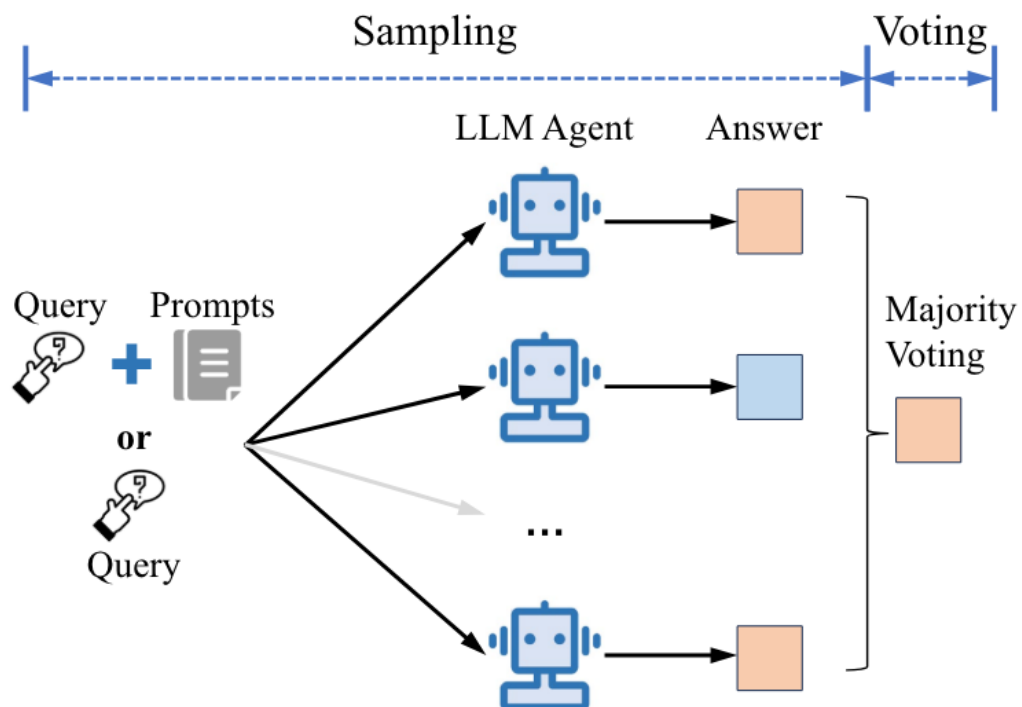
智能体是模型的分身。



好，现在我要说的关键来了，

你可以用让模型变成多个分身的方式，提高模型的智能。

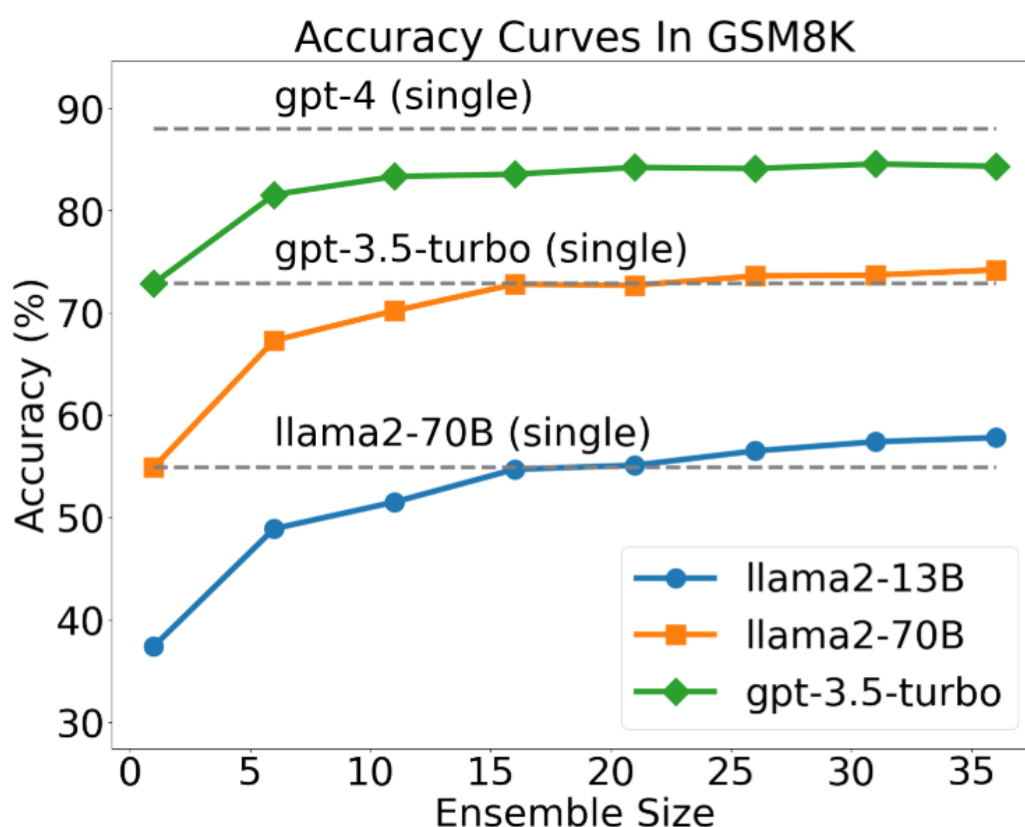
2024 年 2 月，腾讯公司的几个研究者发表论文 [11]，演示和综述了这个方法。对于一个给定的问题，你先用模型生成几十个智能体分身，让这些分身各自处理问题，它们生成的答案有的相同有的不同；然后把所有答案汇总在一起投票，看得到哪个答案的智能体最多，就选那个答案。



你可以用不同方式生成众多的智能体分身。比如像开“私董会”那样，让模型扮演不同的角色，从不同角度思考问题；或者你可以调用不同的模型思考同一个问题；又或者你可以利用系统随机性，简单地让一个模型多跑几遍。

要点是，每个分身都是一次独立的推理，互相不干扰。这样就克服了群体思维，体现了丹尼尔·卡尼曼说的「决策卫生」。而正如斯科特·佩奇在《多样性红利》一书中讲过的那样 [12]，群体能力会高于个体的平均能力。

腾讯的研究表明，仅仅靠“智能体分身再综合”这一招，Llama2 这个开源小模型的能力就达到了 GPT-3.5 的水平，而 GPT-3.5 的能力分身综合后可以达到 GPT-4 的水平。



这正是三个臭皮匠赛过诸葛亮。

✱

2024 年 2 月，伦敦大学学院和牛津大学的几个研究者发表一篇论文 [13]，把 AI 评价商业模式的能力跟人类专家做了个对比。任务很简单，这里有 60 份商业计划书，请你按照从好到坏的顺序给它们排个序。这就相当于你是个投资者，看看你判断投资项目的眼光如何。

研究者使用了 7 个不同的大模型，让每个模型扮演 10 个角色，并且使用了两套不同的提示语，这样得到了 140 个智能体。研究发现，单个智能体给的排序各不相同，有的会有偏见；但是如果你把所有智能体的排序综合起来，结果就很接近人类专家给的排序。

6 月份阿里巴巴全球数学竞赛中，参赛 AI 的前三名，也都是用了多个智能体一起判断综合求解的方法。

所以如果你觉得一个 AI 还不够聪明，你可以多用几个 AI。

✱

综合前面这些研究结果，我们可以放心地说，GPT-4 水平的 AI 已经达到了人类中比较聪明的头脑 —— 但不是天才 —— 的智能水平。如果你要求不高，你可以说这就是 AGI。把这个水平的 AI 廉价地、大规模地部署到各行各业，已经足以改变世界。这就是拐点。

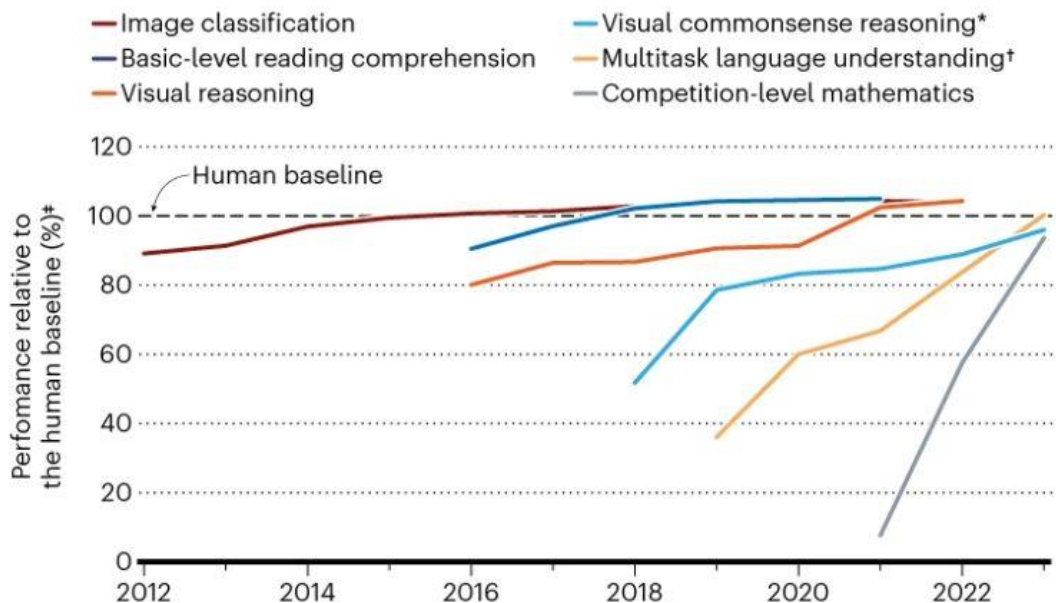
但我们还不满足。AI 能不能再聪明一点，以至于达到明显超过人类的智能水平呢？

### **3.你是否信仰缩放定律**

2024 年 4 月,《自然》杂志刊登了一份截止到 2023 年, AI 各项技能相对于人类基线是个什么水平的演化图 [14] ——

## SPEEDY ADVANCES

In the past several years, some AI systems have surpassed human performance on certain benchmark tests, and others have made rapid progress.



\*Requires an AI system to answer questions about an image and provide a rationale for why its answers are true.

\*Tests an AI model's knowledge and problem-solving ability with regard to 57 subjects, including broader topics such as mathematics and history, and narrower areas such as law and ethics.

\*Data indicate the best performance of an AI model that year.

©nature

我们看到, AI 的图像分类能力、基本阅读理解能力、视觉推理能力都已经超过了人类基线;多任务语言理解能力已经达到了人类基线;视觉常识推理能力和竞赛级数学能力正在以很高的速度接近人类基线。

面对这张图,乐观的人可能会说 AI 智能很快就会超过人类。但是,也有人会注意到, AI 的几项能力在超过人类水平之后,就不再有明显提高了:它似乎只比人类强一点,就止步不前了。这是怎么回事儿呢?

我认为关键原因是那几个能力的上限本来就不高。比如说阅读理解,像生成一篇文章的摘要这种任务,你做得再好又能有多好呢?现在所有主

流模型都能做得比较好 —— 在这个项目上，AI 没有多少可提高的空间。我们更关注的是像数学能力那种高上限项目，而 AI 的数学能力现在恰恰是突飞猛进，没有到顶的迹象。

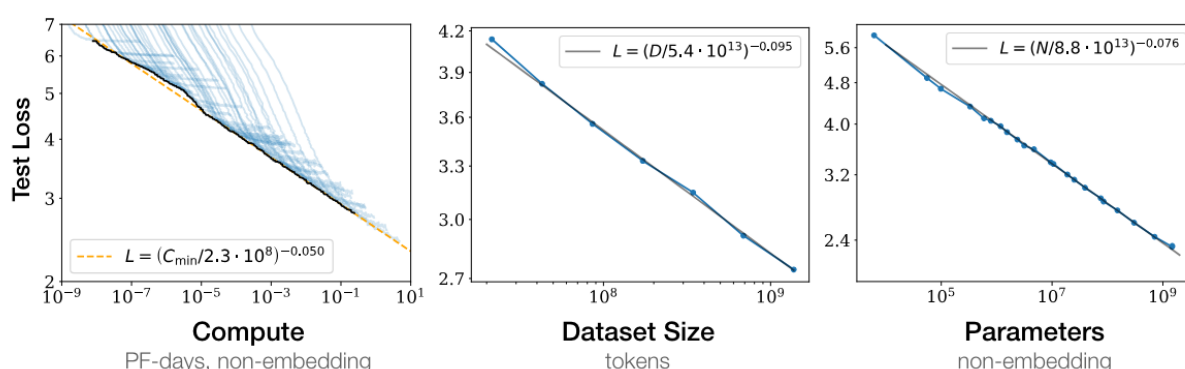
但有些学者不这么看。比如我们专栏多次提到过的图灵奖得主杨立昆 (Yann LeCun)，就认为既然大语言模型都是“语言”模型，它们的智能就必然受到语料水平的限制：既然那些语料都属于人类，你 AI 又怎么可能超过人类的水平呢？所以杨立昆认为 AI 的智能增长将会迅速陷入边际效益递减，达到某个上限之后就会停下来。

他说的对吗？这就是当前最大的悬念。

✱

乐观者相信「缩放定律 (Scaling Law)」。

所谓缩放定律，就是如果你把模型的算力大小（包括运算次数、数据规模、参数个数）扩大多少倍，模型的智能水平会一直跟着扩大。OpenAI 2020 年的一篇论文最先发现了缩放定律 [15]，如下图所示 ——



**Figure 1** Language modeling performance improves smoothly as we increase the model size, dataset size, and amount of compute<sup>2</sup> used for training. For optimal performance all three factors must be scaled up in tandem. Empirical performance has a power-law relationship with each individual factor when not bottlenecked by the other two.



算力投入越多，模型表现就越好。

这个缩放定律是所有做大模型开发的公司的命根子。缩放定律在，就说明只要你投入更多的资源就一定能得到更好的结果 —— 如果哪天缩放定律不好使了，你再把算力提高十倍也不能让 AI 有更高的智能，那么这一轮的游戏就算结束了。

截至此时此刻，不管是大佬公开发言还是研发人员私下交流，都是说缩放定律仍然有效 [16]。OpenAI CEO 山姆·奥特曼一再表示，GPT-5 比 GPT-4 厉害得多，而未来的 GPT-6 又比 GPT-5 厉害得多.....

但是 OpenAI 迟迟没有发布 GPT-5。可以说现在我们没有任何切实的证据，说未来 AI 的水平能比 GPT-4 厉害得多。

我们反倒是听到了一些隐忧。

✱

一位机器学习大神级的专家，jbetker，2023 年 6 月在自己的博客发了一个贴 [17]，说他已经 OpenAI 工作了一年，训练了各种各样的模型，配置了各种各样的参数，他发现一个在我看来有点恐怖的规律：模型行为不是由架构、参数或优化器选择决定的，而是由数据集决定：使用同样的语料训练，不管你是什么模型，最终它们的表现都是一样的。这就如同不管哪个学生聪明哪个学生笨，只要他们都努力读书，他们最终都学成了学校用的教材的样子。

难道说归根结底，AI 的智能终究是语料的水平？

jbetker 不是唯一一个这么说的，还有别人也是这个看法 [18]。这符合杨立昆的直觉。是啊，我们真的指望中学课本能培养出大学生来吗？

但是，难道从古至今每个人不都是人类语料训练的结果吗？为什么有的人就有超过课本的智能呢？为什么 AlphaGo 下围棋就超过了人类棋手呢？就算 AI 的“见识”跟人一样，但是它的运算速度快，它的逻辑推理更冷静，它能同时调动很多个分身从多角度思考问题，它还是可以超过人的智能吧？

这些问题我们坐在这里猜测没用，必须等 GPT-5 出来才知道。

✱

另一个隐忧是，训练 AI 用的语料，可能快要用完了。

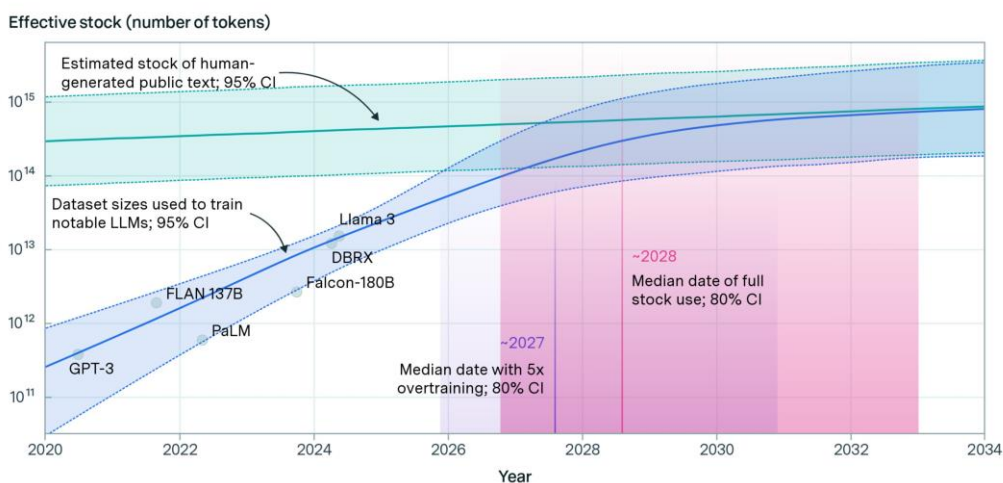
是，人类每天都生产无数的新信息，但是，其中绝大部分都是没什么价值的闲聊。优质语料是有限的，毕竟图书馆里就只有那么多本书。

2023 年的一个研究 [19] 估算了公共人类文本数据的总存量，考虑到现在大语言模型发展的趋势，认为在 2032 年之前，大模型就会面临自己的胃口太大，没有新语料可用的局面。

我还看到另一个预测说，到 2028 年，能用的语料就已经到顶 ——

## Projections of the stock of public text and data usage

EPOCH AI

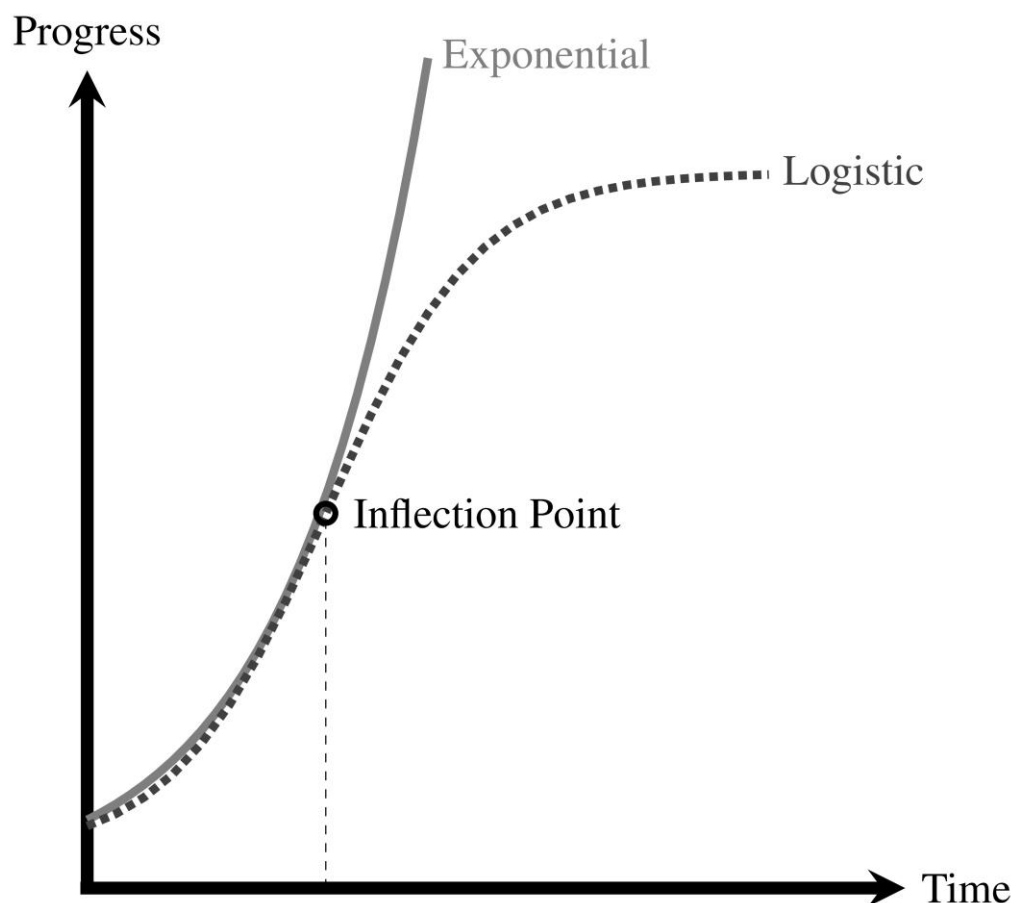


当然到时候我们肯定可以开发“新型”语料，比如把一些非语言数据变成语料。其实现在就已经有公司专门给 AI 公司提供非天然的、合成的高效训练语料。

但不论如何，我们终究会面临一个问题：有限的人类知识能训练出智能无限的 AI 来吗？

✱

DeepMind 的 AI 科学家，也是伦敦大学学院教授蒂姆·洛克塔舍尔 (Tim Rocktäschel)，最近在 X 上发出一个警告 [20]。他说我们都爱谈论「指数增长」，但是指数增长的初期，其实跟「逻辑斯谛 (Logistic) 函数」的曲线是一样的，像下面图中这样 ——



实线是指数增长，虚线是逻辑斯谛增长。二者的区别是指数增长会一直高速长下去，而逻辑斯谛增长会很快陷入边际效益递减，最终收敛在一个上限之下，形成所谓「S 曲线」。

我们现在很难说接下来 AI 智能是继续指数增长，还是陷入 S 曲线。

按理说，世界上没有永远的高速增长，任何技术终究都会陷入 S 曲线，要想再增长就必须改变增长方式才行 [21]。但是，过去几十年间摩尔定律就真的是一直保持指数增长！

所以我们还是不知道，一切只能等待。

# 《万维钢·精英日课6》

## 帮你和全球精英大脑同步

版权归得到App所有, 未经许可不得转载



万维钢  
科学作家

注释

[8]

<https://twitter.com/IntuitMachine/status/177583731657089893>

9

[9] Masterman, Tula, Sandi Besen, Mason Sawtell, and Alex Chao.

“The Landscape of Emerging AI Agent Architectures for Reasoning, Planning, and Tool Calling: A Survey.” arXiv, April 23, 2024. <https://arxiv.org/abs/2404.11584>.

[10] <https://twitter.com/jiayq/status/1786208795720593789>

[11] Li, Junyou, Qin Zhang, Yangbin Yu, Qiang Fu, and Deheng Ye.

“More Agents Is All You Need.” arXiv, February 12, 2024. <https://arxiv.org/abs/2402.05120>.

[12]

《多样性红利》1: “多样性” 到底好在哪?

[13] Doshi, Anil Rajnikant and Bell, J. Jason and Mirzayev, Emil and Vanneste, Bart, Generative Artificial Intelligence and Evaluating Strategic Decisions (February 16, 2024).

<http://dx.doi.org/10.2139/ssrn.4714776>

[14] Jones, Nicola. "AI Now Beats Humans at Basic Tasks — New Benchmarks Are Needed, Says Major Report." *Nature*, April 15, 2024. <https://doi.org/10.1038/d41586-024-01087-4>.

[15] Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. "Scaling Laws for Neural Language Models." *arXiv preprint arXiv:2001.08361* (2020). <https://arxiv.org/abs/2001.08361>.

[16] 感谢曾鸣的讨论。

[17] jbetker, "The 'It' in AI Models Is the Dataset." *Non\_Interactive*, June 10, 2023. <https://nonint.com/2023/06/10/the-it-in-ai-models-is-the-dataset/>.

[18] <https://twitter.com/arimorcos/status/1777753903184175390>

[19] Villalobos, Pablo, Jaime Sevilla, Lennart Heim, Tamay Besiroglu, Marius Hobbhahn, and Anson Ho. "Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning." *arXiv*, June 6, 2023. <https://arxiv.org/abs/2211.04325v2>.

[20] [https://x.com/\\_rockt/status/1799006089260007549](https://x.com/_rockt/status/1799006089260007549)

[21] 精英日课第二季，复利的鸡汤和真实世界的增长



## 划重点

- 1.「缩放定律」是说，算力投入越多，模型表现就越好。但反面观点是，AI 的智能增长将会迅速陷入边际效益递减，达到某个上限之后就会停下来。这是当前最大的悬念。
- 2.对此我们有两个隐忧：AI 的智能终究是语料的水平？训练 AI 用的语料，可能快要吃完了？
- 3.我们终究会面临一个问题：有限的人类知识能训练出智能无限的 AI 来吗？