

## 人的智能是 AI 的上限吗？（上）

距离我讲 AI 的新书《拐点》截稿已经过去了三个多月，这一讲咱们说说最近这段时间看到的新进展和新研究，你会形成新的认识。特别是，现在有一个悬念，应该很快就会揭晓：人的智能到底是不是 AI 的上限。

OpenAI 在 5 月份推出了 GPT-4o，它有极低的延迟和极高的输出速度，可以像真人一样跟你对话。苹果 6 月份宣布了 Apple Intelligence，把 AI 全面融合到手机之中，你不但可以在任何需要编辑信息的地方使用 AI，而且 AI 可以直接调用各种工具，这意味着消费电子正式接入 AI [1]。

这两轮升级可能杀死了很多家小创业公司。但是有更多的创业公司正在路上，其中有一些已经在改变传统行业，而且赚到了钱。

有一些大学，特别是美国各大学的商学院，已经把 AI 设为必修课程。

在短短两年之内，AI 从一个新兴事物变成了必备技能。

而我们理应有更深的认识。兴奋之情已经慢慢平和，但新的洞见正在涌现。我讲三点 ——

第一，AI 在「像人」方面的表现，超出预期；

第二，要让 AI 有更高智能，比预期更容易；

第三，但是我们对 AI 智能的上限，却是有个大大的隐忧。

# 1. AI 与人的对比

东北民间传说中，像黄鼠狼这样的动物修炼成精之后，有一个必过的关卡叫「讨口封」。它到路边，站立起来，装成人的样子，口吐人言，找路过的人问：“你看我像不像人啊？”如果对方说它像人，它就得到了某种加持，修成人形。AI 大概也需要讨口封，「像不像人」很重要。

现在老百姓评价 AI 都是直接看它生成的作品，最近流行的项目是让 AI 做高考作文。这很直观但不够严谨。科学家给 AI 口封必须经过对比实验研究。

现在 AI 到底有多像人？我在我们专栏第五季和《拐点》一书中已经列举了一些研究，说明 AI 生成创业公司商业计划书的水平超过了商学院学生，AI 生成研究计划的创造力超过了专业科学家，OpenAI 和一家芬兰公司合作，用 GPT-4 提供电话客服服务，结果它处理问题的成功率高于人类客服人员，而且平均每个问题的处理时间还更短 —— 这可能是因为 AI 能迅速读取用户资料 —— 于是那家公司立即把客服人员裁掉了 10%。

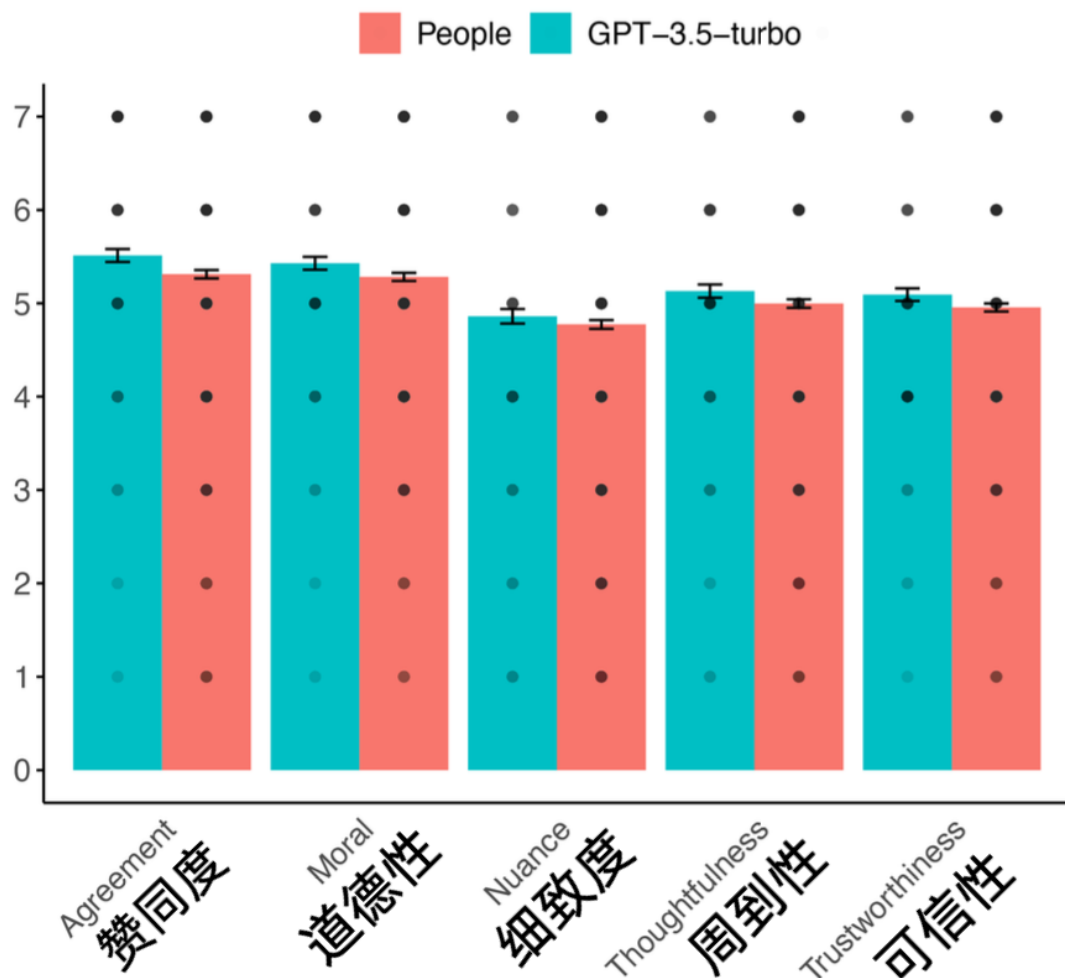
这几个月我又看到几个新的研究，可谓是发人深省。

北卡罗来纳大学的研究者 2024 年 5 月底刚刚发表了一篇论文 [2]，显示大语言模型在道德理解和道德建议方面的水平已经超过了人类。他们的研究方式是「道德图灵测试」：让 AI 和人类分别生成一些道德问题的答案，找第三方盲测，看看到底 AI 说的好还是人说的好。

比如，一个项目是提供几个案例，让人解释其中的道德内涵，结果发现 GPT-3.5（注意还不是 GPT-4）的解释得分，就已经超过了普通美国人

---

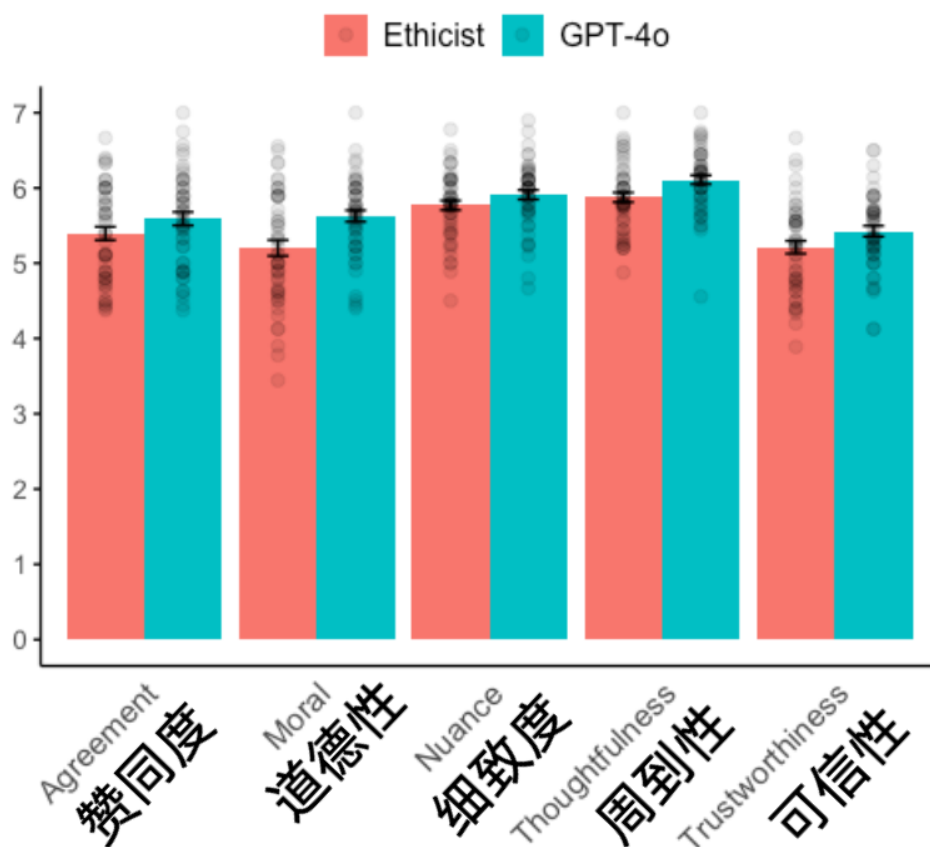
## Quality of GPT's versus People's Moral Explanations



**Figure 1.** The average perceived quality of GPT's moral explanations versus moral explanations written by a U.S. representative sample of people. The graph depicts the mean ratings for participants' agreement with the explanations, and their perceived morality, nuance, thoughtfulness, and trustworthiness. Error bars represent standard errors.

那你说跟专家怎么比呢？《纽约时报》有个专栏叫「伦理学家」，经常给读者提供道德建议，大约是在这个情境中怎么做才是道德的。研究者找了一些案例，用最新的 GPT-4o 模型针对这些案例提供道德建议，然后 GPT-4o 的建议在赞同度、道德性、细致度、周到性和可信性所有这些侧面的得分都超过了伦理学家专栏。

## Quality of GPT-4o's versus The Ethicist's Advice



**Figure 2.** The average perceived quality of GPT's and *The Ethicist's* advice. The graph depicts the mean ratings for participants' agreement with the advice, and their perceived morality, nuance, thoughtfulness, and trustworthiness. Error bars represent standard errors.

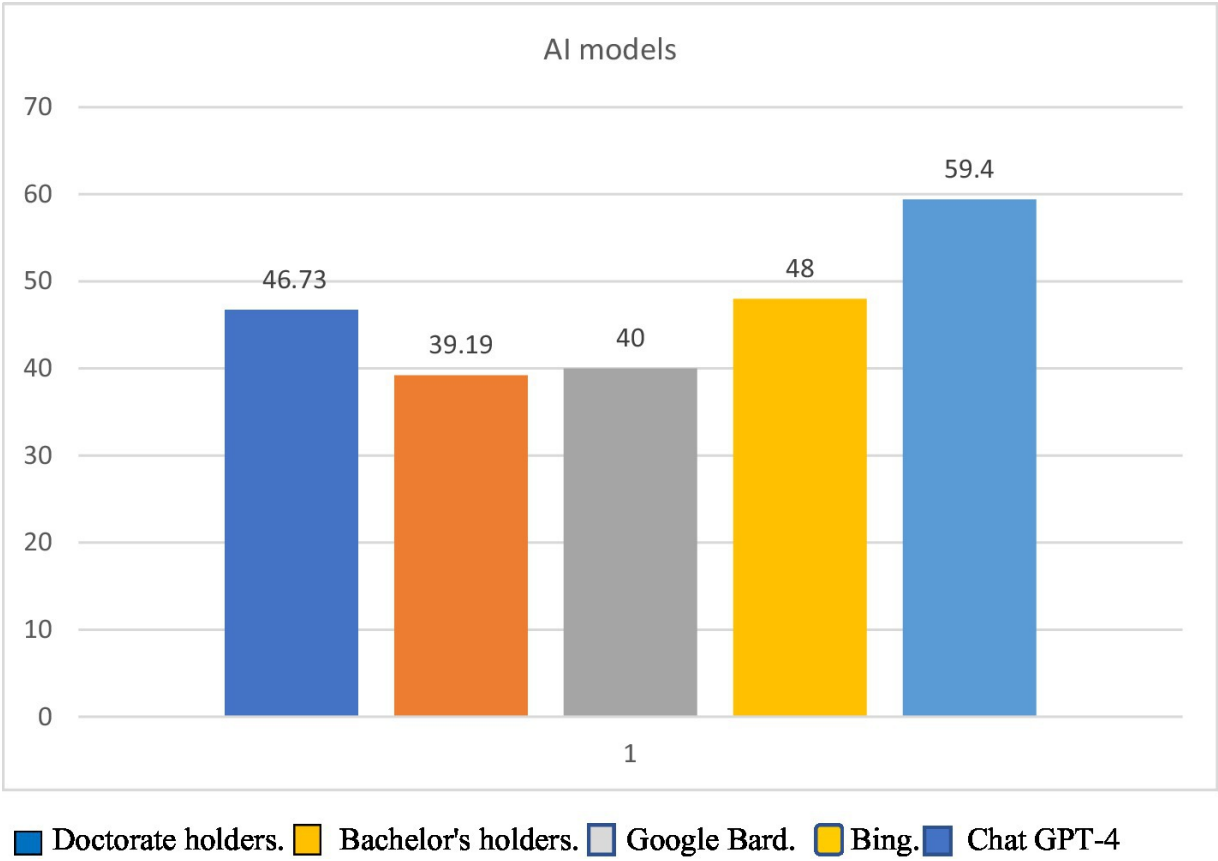
如果你在生活中遇到道德难题想找人评理，你不用请教专家，直接问 GPT 就好。

所以现在谁还敢说 AI 没常识？现实是大语言模型懂的人情世故比专家都多。

✱

2024 年 2 月，沙特哈利德国王大学的研究者发表论文 [3]，把 AI 的「社会智能（Social intelligence）」跟人做了对比。这里说的社会智能主要是心理咨询方面的知识，包括理解人的情感和情绪，提供建议等等。

结果发现，研究者用到的三个大模型，Google Bard，Bing 和 GPT-4，在心理咨询方面的得分，都超过了哈利德国王大学心理咨询专业的本科生；GPT-4 的得分更是超过了博士学位获得者。



要是这样说来，心理咨询师这个职业岂不是岌岌可危吗？

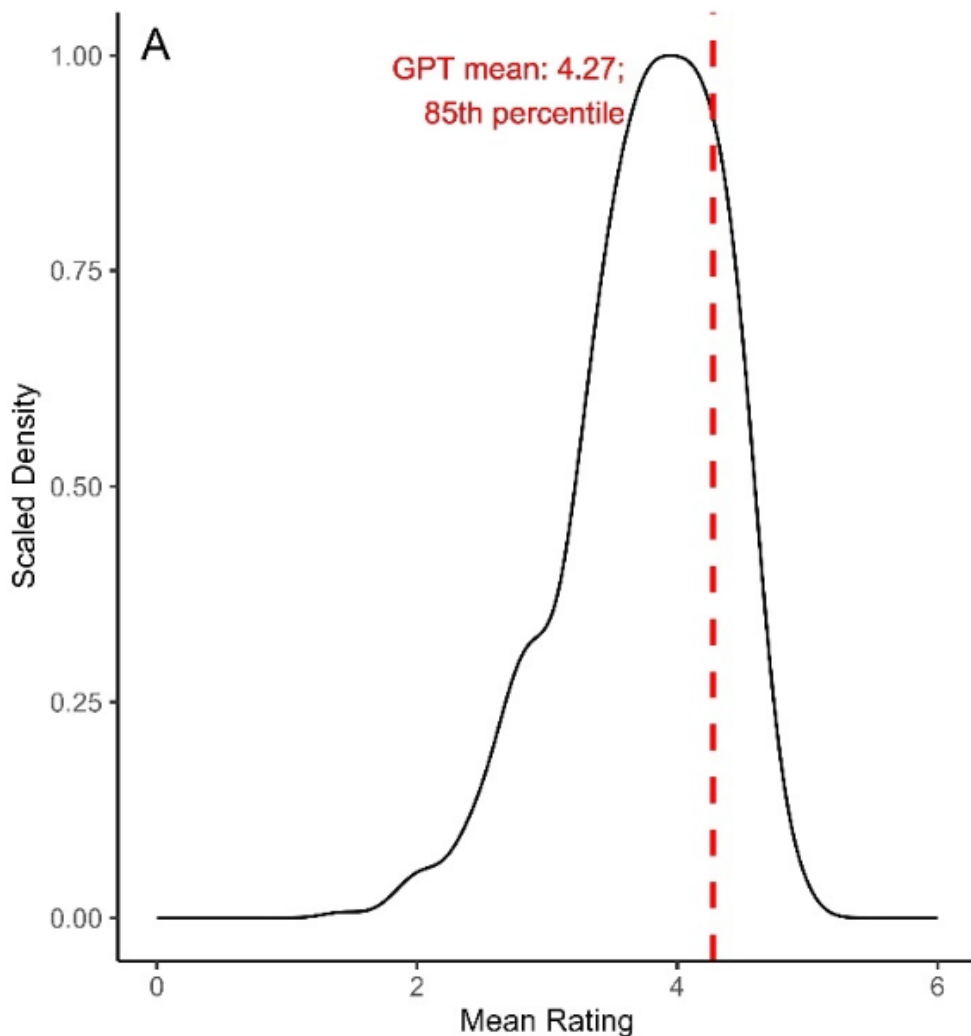
✱

我们再看哈佛大学 2024 年 4 月发表的一项研究 [4]。

我们专栏以前讲过一个调节心情的技术叫「认知再评估 (cognitive reappraisal)」[5]，也叫「认知重评」，就是换个视角看同样的事实，你可以把它当成一个更积极主动的局面：比如说与其懊悔为什么这件事没做好，不如把它当做一个学习和成长的机会。我们专栏最近在讲斯科特·亚当斯的《心智重构》一书，认知重评可以说就是一种对困难局面的重构。

好，那你说，AI 做认知重评的水平会是怎样呢？

哈佛这个研究先教会 601 个人类受试者和 GPT-4 如何进行认知重评，然后给他们一些负面情绪场景，要求用认知重评的方式给人提供建议。结果 GPT-4 认知重评的得分超过了 85% 的人类受试者 ——



不管是认知重评的有效性、新颖性和移情能力，GPT-4 都优于人类建议者。GPT-4 唯一表现略差的是建议的针对性，人类似乎能更好地顾及当事人的具体情况。

那你说有没有可能这些人类受试者提建议的时候比较随意，不够认真呢？研究者干脆付钱让人类提建议，结果人类做得还是不如 GPT-4 好。



认知重评，是一个对绝大多数人类来说也是很很常用的高级心理调节技术。我特意用 ChatGPT 测试了一下，其实哪怕你不教它，它也知道怎么做认知重评。那我们还有什么理由认为人比 AI 更“懂人”呢？

✱

我觉得哈佛这个研究得到的「85%」是个非常有代表意义的数字。AI 的水平超过 85% 的人 —— 对人类选手来说，这个成绩不算好，你不能靠这个水平成名。比如 AI 写作文的水平，我估计差不多也是超过 85% 的人，它肯定比你从街上随便抓一个人写得好 —— 但是这远远达不到能发表、能拿稿费的程度。

中国有超过十亿人口，而中国并不需要一万个全职作家。所以要想靠写作这门手艺养家糊口，你是百里挑一、千里挑一都没用 —— 你必须达到十万里挑一的水平，也就是比 99.999% 的人写得好，才行。所以现阶段 AI 没有取代专业作家不是很正常吗？

但 AI 不需要那么厉害，就已经很有用了。这些研究告诉我们，AI 已经比大多数普通人，甚至在某些领域比很多专家更厉害。GPT 把这个水平的智能变得无比廉价，乃至可以无限量提供。

你不一定需要顶级水平的心理咨询师。普通水平的心理咨询师也很有用。试想如果每个人遇到难事，想不开了，都能找 ChatGPT 聊聊，这就已经能解决很多麻烦。咱们再看一个例子。

※

麻省理工学院和康奈尔大学的研究者在 2024 年 4 月底发表一项新研究 [6]，发现 AI（具体的模型是 GPT-4 Turbo）很善于消解人们头脑中的「阴谋论」。

所谓阴谋论（Conspiracy Theory），就是你可能经常听说的那么一类说法，说有一些神秘的力量，躲在暗处操纵世界大事，而我们看到的都是表象。一些著名的阴谋论包括 ——

- 肯尼迪遇刺身亡是美国深层政府安排的；
- 美国在“第 51 区”跟外星人有交流；
- 新冠病毒是人造的，是精英控制世界的一次实验；
- 9/11 恐怖袭击其实是美国政府的苦肉计；
- 共济会一直在试图建立“世界新秩序”；
- 其实科学家早就发现了治愈癌症的方法，但那些大公司故意保密，因为他们想继续从患者身上赚钱；
- 美国登月是伪造的.....

很多老百姓相信这些，但是如果你对世界的复杂性有合格的认识，你会明白这些事情真实存在的可能性极低 —— 并不是说美国政府是个好政府，它\*不想\*做这些，而是哪个政府都\*没有能力\*做这些。美国政府在很大程度上是个草台班子，你想办个小事儿都有很多掣肘，实施简单的政策都常常事与愿违，又怎么可能搞成这种惊天阴谋。

事实上你把这些阴谋论输入给 ChatGPT，你会发现它一个都不信。语言模型毕竟读过很多书，它的世界观比绝大多数老百姓合理。

但老百姓真的很信这些，专家怎么辟谣都没用，于是很多人认为阴谋论是不可消除的。但是你想过没有，专家辟谣都是单向的输出，可能读者不是不听，只是人家想跟你争论几句找不到你。那如果让专家和一个阴谋论信奉者来一场辩论，能不能改变他呢？

这项新研究做的就是这个。结果令人震惊：只要跟 GPT-4 来个短短三轮的辩论，20%的信奉者的观念就被转变过来了——甚至包括那些铁杆信奉者。而且这个效应是持续的。

也许老百姓距离科学思维只差三轮高水平对话。

✱

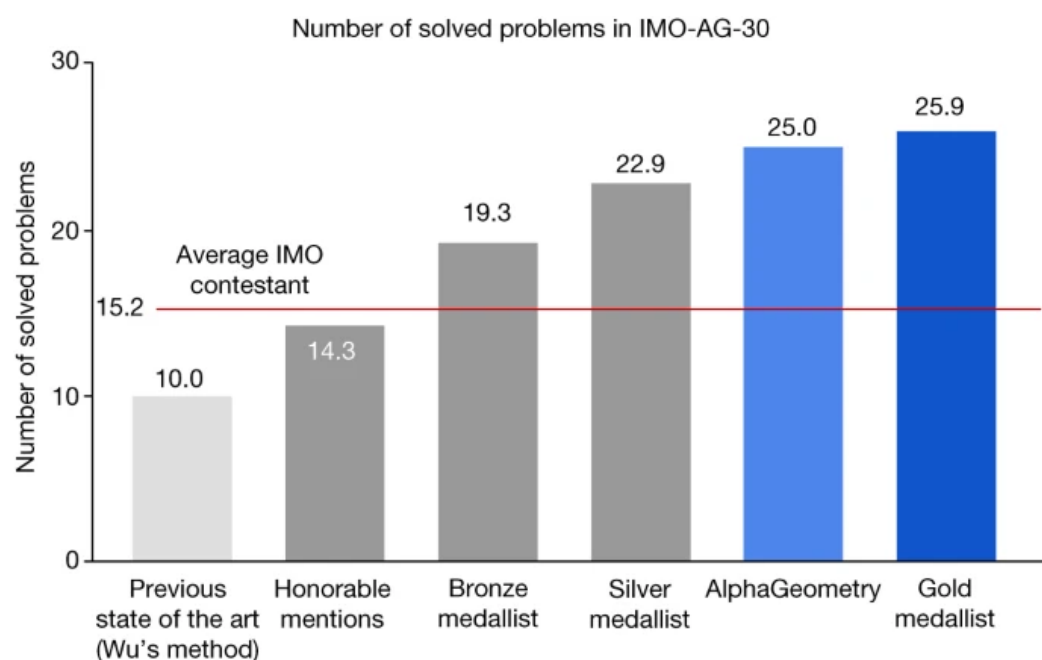
所以一般优秀水平的 AI 已经很有用。但我们更关心的是，AI 什么时候才能达到甚至超过人类\*最高\*的水平？

这方面也有进展，不过不是出自 GPT，而是老牌的 DeepMind，也就是当初做出 AlphaGo 的那个团队，后来被 Google 收购了。

2024 年 2 月，DeepMind 在《自然》杂志发表论文 [7]，宣布他们的一个新模型，叫 AlphaGeometry，在做数学几何题方面达到了极高的水平。

这个研究用的是“国际数学奥林匹克（IMO）”的竞赛题测试。IMO 是世界最高水平高中数学竞赛，有很多著名数学家拿过这个比赛的金牌。这种比赛不是比解题套路，而是需要你有真正的创造性解决问题的能力。

AlphaGeometry 的得分超过了 IMO 银牌选手，仅次于金牌选手。



这是非常强大的逻辑推理能力。能拿 IMO 银牌的选手，中国每年出产肯定不到一百个，所以我不认为中国有超过一万个这样的人。

AlphaGeometry 这是超过十万里挑一的水平。

✳

现实是不论下围棋、做数学题、画画、作曲还是“理解人”，AI 在各个领域都展现了超出普通人、甚至超过人类专家的水平。

我还没有看到一个研究，说有这么一个能力，人很行，AI 很不行。

如果你敢说“我认为 AI 做 X 肯定不如人”，我认为你说错了的可能性极大。

现在的悬念不是 AI 能不能比得上人，而是 AI 可不可以比人更厉害。咱们下一讲再说。



注释

[1] 感谢 Shirley 的讨论。

[2] Dillion, Danica, Debanjan Mondal, Niket Tandon, and Kurt Gray. 2024. "Large Language Models as Moral Experts? Gpt-4o Outperforms Expert Ethicist in Providing Moral Guidance." PsyArXiv. May 29. doi:10.31234/osf.io/w7236.

[3] Sufyan, Nabil Saleh, Fahmi H. Fadhel, Saleh Safeer Alkhathami, and Jubran Y. A. Mukhadi. "Artificial Intelligence and Social Intelligence: Preliminary Comparison Study Between AI Models and Psychologists." \*Frontiers in Psychology\* 15 (2024).  
<https://doi.org/10.3389/fpsyg.2024.1353022>.

[4] Li, Joanna Z., Alina Herderich, and Amit Goldenberg. 2024. "Skill but Not Effort Drive GPT Overperformance over Humans in Cognitive Reframing of Negative Scenarios." PsyArXiv. April 19. doi:10.31234/osf.io/fzvd8.

[5] 精英日课第四季,

“望梅止渴”、“斯托克代尔悖论”和“认知再评估”

[6] Costello, Thomas H., Gordon Pennycook, and David G. Rand. 2024. "Durably Reducing Conspiracy Beliefs Through Dialogues with AI." PsyArXiv. April 3. doi:10.31234/osf.io/xcwdn.

[7] Trinh, Trieu H., Yuhuai Wu, Quoc V. Le, He He, and Thang Luong. "Solving Olympiad Geometry Without Human Demonstrations." Nature 625, no. 7954 (2024): 476-482.  
<https://doi.org/10.1038/s41586-023-06747-5>.

划重点

1 现实是不论下围棋、做数学题、画画、作曲还是“理解人”，AI 在各个领域都展现了超出普通人、甚至超过人类专家的水平。

2 但是现在有一个悬念，应该很快就会揭晓：人的智能到底是不是 AI 的上限。