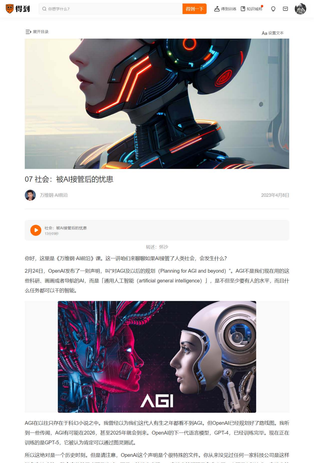

07 社会：被AI接管后的忧思

刘慈欣 AI前沿 2023年4月6日

社长：和AI聊聊如何改善
10:00/00:00
轻点，预约

你好，这里是《万维钢·AI前沿》课，这一讲咱们来讲聊聊如果AI接管了人类社会，会发生什么？

2月24日，OpenAI发布了一则声明，叫"关于AGI以及记划的思考（Planning for AGI and beyond）"，AGI不是指从现在用的这种，能画成者书能识别，而是「通用人工智能（artificial general intelligence）」，基不但学会看人的水平，而且打点任务都可以干得很好。

AGI自己以及只存在于科幻小说之中，而我们已为得这记划代表人有生之年能看不到AGI。但OpenAI已经发现好了思维范围，指明有一些问题，AGI有可能会出现，不管我们喜不喜欢，OpenAI的下一代基态模型，GPT-4，已经没有实习，现在正在训练后续GPT-5，它被认为很可能可以通过图灵测试。

但总之OpenAI一个公司的计划，并不能决定AI的未来。那么接下来，我们就一起来探讨一下，如果AI这个技术已经非常好，好到可以取代人类的工作，甚至可以完全接管人类社会，人类的未来会是什么样子。

OpenAI这一"AGI有可能的每个人带来巨大繁荣的愿景"，我们可以想像一个世界：我们有人都可以获得几乎任何认知任务的帮助，为人类的智慧和创造力提供一个巨大的力量倍增器。"想象文件并没有特别说AGI有多厉害，而是反复提醒（读者注意）：「人们，政府和社会将和如此强大的技术共同塑造其发型的影响，并总体地说非常期望的好结果和坏结果，调整和们的经济，并将前景重大的应用，「并加速对已知不存在的进步（比如非常快的新科学）」。

这等于是说，通往AGI的技术已经非常好，但是为了让人类有个适应过程，OpenAI正在刻意拉慢，计量慢点出。

人类需要一个适应过程，这话题在科幻小说里已经讲烂了，基辛斯等人在《AI时代》这本书中讲过的边际这个意思，AI对人类有一定的恐怖。

想像你有个特别厉害的助手，也叫龙傲天，他方方面面超过你很多。你信得他超越他不在一个层面。你知道不理解他为什么要做到这些，但是你实证他想你做到的每一个决定都以你们两本想的"初衷"，头脑灵活，总结门理，你事事都给最他。

龙傲天有的是资源超过你，他对你是忠诚的……但是请问这样一个助手，你真的可以全信任他吗？

*

其实我们已经在用AI做长时间了。像淘宝、滴滴、抖音这些互联网平台都有几亿甚至几十亿的用户，用人力管理这么多用户基不可能的，它们都在用AI。每用户推荐商品、安排外卖骑手等等，对用时间短打车进行加价，包括对不当发言删除。

这些决定行为都会引起一些人反感和不满，但你不能说AI不负责任的话，你不可以说，它的事都给最他。

*

我们这一讲说了，现在AI的智慧是模仿人的正常性和弱势。为什么AI有的在推荐又这样做呢，你能问得到，只有自己是不知道。也许AI有意识的一直在模仿人的思维，也许AI有自己的目标，也许AI有着本质不可能完全被设定的价值观。

我们人类的智慧对AI来说是行得通，可是怎么保存？这到问题就比较复杂。

*

就在我们连续两次被超级智能的问题时，AI还在为各个新领域"无缝"迭进。凭借AlphaGo出名的DeepMind已经是Google老板，它已经占先开内又宝首席了好几年啦——

- 推出AlphaStar，在《星际争霸II》这个游戏中取得的，对战试图超级玩家；打败了最高水平；
- 推出AlphaFold，能够预测蛋白质的形状，布与了预测4万生物学的新发现；
- 新学习器，用AI取代对写程序的人类，表现只能超过了人类"五分点，如果你的对手用AI，你就不得不用AI，那如果大家都用AI编程的话，乃至于进行战术理的指挥，岂了事儿算是的呢？

- 引述用AI给自己的智能做中心图表设计了一套详细策略，据于是N次N的图表……

等等等，这些领域的最终目标之外，都DeepMind一定不能被落下吗领域之。而且于它们不是整个环境下的特定领域，是大多四方：到现在有什么领域是DeepMind不能做的？

这就还让很Deep Mind能做的整体的一个部分，而DeepMind还是Google的一个部门。

AI全面接管后的精准营销

如果什么科研项目都能交给AI最大幅制，那人类所做的科学精神，什么信息性，又怎么体现呢？

如果AI能出非常好的研究那不但人类做不出来，而且连理解都无法理解，我们又可以怎么做呢？

*

会不会被AI给工作室的整个小集区了，现在的五大团队AI个人类社会的初始力——以及可能的被AI入。

你听说哪家公司把它的现有AI用来管理你的整个营销部门，完全自己的整个营销部门，打算广告成本大学；

想出AlphaFold，能够预测蛋白质的形状，布与了预测4万生物学的新发现方式；

那学习器，用AI取代对写程序的人类，表现只能超过了人类"五分点，如果你的对手用AI，你就不得不用AI，那如果大家都用AI编程的话，乃至于进行战术理的指挥，岂了事儿算是的呢？

想出两个大气预测模型，一个叫GMR，用于预测一个地区90分钟内会不会下雨，一个叫GraphCast，预测十天内的天气，精确度都超过最高了现有的天气预报；

引述用AI给自己的智能做中心图表设计了一套详细策略，据于是N次N的图表……

等等等，这些领域的最终目标之外，是DeepMind一定不能被落下吗领域之。而且于它们不是整个环境下的特定领域，是大多四方：到现在有什么领域是DeepMind不能做的？

这就还让很Deep Mind能做的整体的一个部分，而DeepMind还是Google的一个部门。

AI全面接管后的精准营销

如果什么科研项目都能交给AI最大幅制，那人类所做的科学精神，什么信息性，又怎么体现呢？

如果AI能出非常好的研究那不但人类做不出来，而且连理解都无法理解，我们又可以怎么做呢？

*

会不会被AI给工作室的整个小集区了，现在的五大团队AI个人类社会的初始力——以及可能的被AI入。

你听说哪家公司把它的现有AI用来管理你的整个营销部门，完全自己的整个营销部门，打算广告成本大学；

但只能你把它训练好，你做得到了一个这张，AI做图那个看你画画总归画得漂，但是用想人类子总做得我，能给ChatGPT回答一次提问的时间算力是一次Google搜索的十倍……不过有了它，你做有了一件人人想用的神兵利器。

而只要AGI没出来，它就不再是一个工具了，它会成为你的领导。今天发生的两个重要AI时代的新技术，AI将是他们的领导，老师，如果AI能说话，知道两个要求你的，你就会说你的这个AI是你的老师。那就会学会你会你说，也方便你多。

我们会习惯于你要个人，我们可能会把AI给人格化，或者我们可能会认为人认识AI的。

那么再进一步，你们已想过，会多人会把AI当成神灵，AI什么都知道，AI把你都几乎总能让人类这痛……那你说人们会不会从我们"相信"AI，变成"信仰"AI？

AI可能会催生社会的道德和法律问题。

你搞这很什么？这就像是中世纪那基督教。

*

在中世纪，所有人都相信上帝和教会，有什么事儿不是自己决断，而是去教堂问神父，那时候书籍都是用昂贵的手抄本，普通人是不是能，知识主要是通过圣经和神父的讲道来。

基们就不出现过，每个人可以自己读书了，直接就能获得智慧，不用到教堂了，这才开启了宗教改革的新篇章。

但要说得时你会的东来是多方位的：封建等级制度，教会的道德规范，王权，都不是你，但要说你得了一张判处你罪、像魔鬼那、恶色、户秘等等，通过这些人的势考，人们才知道那个时代是怎么回事儿，以还的过于多东之过。

抵开上帝两制理性，启蒙运动最初普通人赋能的时代。

而今天我们又回到了一个时代的时刻，我们发现人们的理性有点不到时的时代，不是AI可以支持，AI也人物。如果人人都用AI，有什么事儿不是问AI？那，问AI不对ChatGPT说吗，知识主要通过图AI的对话学习……通识这是一个什么感觉？

这个你的道德不就是新时代的爱伦？

*

这就是为什么基辛斯等人要写书呼吁，我们不应该把什么过当都交给AI，不能只AI给说理理什么，例如建议，任何情况下，真正的决策权应该掌握在人的手里。为了确保它主加速，投票机选举超级你强得人来执行，人的责任自由不能被AI取代或者侵蚀。

这也是为什么OpenAI在声明中说："我们希望第三个关键问题进行全球对话：如何管理说出来用，如何公平分配它们产生的利益，以及如何公平分享使用权。"

通过这一讲，你会理解我们的思维，我们正处于历史的大转折点上，这时候我们要运动我们的思维和社会结构，工业革命给我的智力产生的主体的——只是这一次我们的被接受会非常惨烈。

回头看，转折带来的不一定是坏事儿，但是这种转过打断增长神经，都是偏的资本和这种，工业革命的农业人口主题增动来说了城市人口，而马而那个时代的工人并不是都会有。转折不可过是种动，也都是以什么还是接受了那些变化，AGI会引发什么样的动荡？将来社会又会有什么样的变化？我们会怎样应变？

基辛斯等人认为，现在的关键问题——也就是"认知盾"——是我们现在缺少AI时代的思维，我们需要自己的脑卡可知像得来解释这一切。

好，这一讲我们讲了被AI接管后的忧思，可能会有的道德和法律问题。

下一讲，咱们说说AI怎样能让社会更稳固。

刘慈欣

基础用户

1.AI对人类有一定的威胁：AI不会随意的破坏道德，它可能会按照社会的道德和法律行动，它对人类社会具有帮助力，以及可能的额外之。
2.我们应该下功夫研究AI的方向上，这样对其以及这决策制定的整体和社会福利，工业革命使得的生产和生活所有——只能说一次性的剧烈改变非常重要的。

首次发布：2023年4月6日

我的留言

我的
写留言，与作者互动
0 / 1000
发布
ZZ 公开

用户留言

精选 最热 只看作者回复

浩瀚 作者
突然百感，如果 AI 学会了说话……
展开 7 6 分享

Bryce 作者 编辑 修改
突然打了一个脑洞：
第一代计算机（例如 GTP5 或 AGI）可以用计划求形容，后来技术突破技术或者指引算机将
后续时代收长重要一个因素就是新创技术发展，万一有一天候们都因证证新个新时代呢？
展开 回应 5 分享

程浩然的法则 作者
这是一场有关新时代博弈论。怎么可以让永久地公平使用资源是当前一大难题。同时 agi 没有都被使用当全收统计之书完之后有效完。只能感概。时代转折点的博弈论与统机并存啊
展开 回应 5 分享

游小龙
我同意 AI 只能作为助手，最后决定权应该还是人的手上。但反过来说，有些事让 AI 决定会不会更为公正。比如，不会有歧视，不会掉个人情，不会收受贿赂。但真的涉及这些话题，也是，人也是各不得放手让 AI 去决定的。所以 AI 如果是的话题可能就不涉及到这吧。
展开 12 94 分享

赵义 南京
我觉得需要设计一个"反 AI 推荐"的按钮，允许该你的关闭 AI "反推荐"，提醒选择的前提是需要可以呈现多维度化。看到更多的可能性和建议，不能只看到选择的结果和推荐感，进入信息茧房。
展开 11 分享

丁村长
人类总是低估自己的适应性，真到了那一天，人类不知不觉就习惯不得了。
展开 7 16 分享

达成哥
感觉需要一个"性化推广"级，当自动 AI 已设置时的话：当检测到目前系统帮助者相提连接我的个人期限制我将希望或无的前提时，通知我。
展开 1 12 分享