【知识福利社】你的私人书院，微信：312005650 全网热门课程众筹共享！

07 社会：被AI接管后的忧患

万维钢·AI前沿
2023年4月8日

进入课程 ›

时长：13分09秒 大小：9.02MB
00:00 / 13:09

转述：怀沙

你好，这里是《万维钢·AI前沿》课。这一讲咱们来聊聊如果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“对AGI及以后的规划（Planning for AGI and beyond）”。AGI不是我们现在用的这些科研、画画或者导航的AI，而是「通用人工智能（artificial general intelligence）」，是不但至少要有人的水平，而且什么任务都可以干的智能。

AGI在以往只存在于科幻小说之中。我曾经以为我们这代人有生之年都看不到AGI。但OpenAI已经规划好了路线图。我听到一些传闻，AGI有可能在2026，甚至2025年就会到来。OpenAI的下一代语言模型，GPT-4，已经训练完毕。现在正在训练的是GPT-5，它被认为肯定可以通过图灵测试。

所以这绝对是一个历史时刻。但是请注意，OpenAI这个声明是个很特殊的文件。你从来没见过任何一家科技公司是这样说自家技术的：整个文件的重点不是吹嘘，而是一种忧患意识——它忧患的既不是自家公司，也不是AGI技术，它忧患的是人类怎么接受AGI。

OpenAI说：“AGI有可能给每个人带来令人难以置信的新能力；我们可以想象一个世界，我们所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。”然后文件并没有继续说AGI有多厉害，而是反复强调要「逐渐过渡」：“让人们、政策制定者和机构有时间了解正在发生的事情，亲自体验这些系统的好处和坏处，调整我们的经济，并将监管落实到位。”并且说它部署新模型会「比许多用户希望的更谨慎」。

这等于是说，通往AGI的技术已经具备，但是为了让人类有个适应过程，OpenAI正在刻意压着，尽量慢点出。

人类需要一个适应过程，这也是我们这一讲的主题。基辛格等人在《AI时代》这本书中表达的也是这个意思，AI对人类有一定的危险。

想象你有个特别厉害的助手，名叫龙傲天。他方方面面都比你强，你连思维都跟他不在一个层面。你常常不理解他为什么要那样决定，但是事实证明他替你做的每一个决定都比你自己原本想的更好。久而久之，你就习惯了，你事事都依赖他。

龙傲天所有的表现都证明，他对你是忠诚的……但是请问刘波，你真的完全信任他吗？

✳

其实我们已经在用AI很长时间了。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是不可能的，它们都在用AI。给用户推荐商品、安排外卖骑手接单、对拥挤时段打车进行加价，包括对不当发言删帖，这些决定已经要么全部、要么主要是AI做的。

而问题随之而来。如果是某公司的某个员工的操作伤害了你的利益，你大可抗议，要求他负责；可如果你感到受到了伤害，公司却说那是AI做的，连我们自己都不理解，你同意吗？

我们上一讲说了，现在AI的智慧是难以用人的理性解释的。为什么抖音向你推荐了这条视频，你质问抖音，抖音自己都不知道。也许抖音设定的价值观影响了AI的算法，也许抖音根本就不可能完全设定AI的价值观。

政府和人民都要求对AI算法进行审查，可是怎么审查？这些问题都在探索之中。

✳

就在我们连简单应用都没想明白的同时，AI正在各个新领域突飞猛进。凭借AlphaGo出名的DeepMind已经被Google收购，它在过去几年内又取得了如下成就——

- 推出AlphaStar，在《星际争霸II》这样一个规则复杂的、开放式的游戏环境中，打到了最高水平；
- 推出AlphaFold，能够预测蛋白质的形状，改写了领域内生物学的研究方式；
- 医学方面，用AI识别X射线图片帮助诊断乳腺癌，把对急性肾脏损伤的诊断比主流方法提前了48小时，对老年人眼睛里的老年性黄斑变异做出了提前好几个月的预测；
- 推出两个天气预报模型，一个叫DGMR，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，精确度都显著高于现有的天气预报；
- 它还用AI给谷歌的数据中心重新设计了一套冷却系统，能节省30%的能源……

等等等。这些成就的最可怕之处不是DeepMind一出手就颠覆了传统做法，而在于它们不是集中在某个特定领域，是大杀四方：到底还有什么领域是DeepMind不能颠覆的？

这些还只是DeepMind能做的事情中的一小部分，而DeepMind只是Google的一个部门。

AI全面接管科研就在眼前。

如果什么科研项目都能交给AI暴力破解，那人类所谓的科学精神、什么创造性，又怎么体现呢？

如果AI做出来的科研结果不但人类做不出来，而且连理解都无法理解，我们又何以自处呢？

✳

会不会被AI抢工作那些都是小事儿了。现在的大问题是AI对人类社会的统治力——以及可能的破坏力。

华尔街搞量化交易的公司已经在用AI直接做股票交易，效果很好。可是AI交易是以高频进行，在没有任何人意识到之前，就有可能形成一个湍流，乃至于引发市场崩溃。这是人类交易员犯不出来的错误。

美军在测试中用AI操控战斗机，表现已经超过了人类飞行员。如果你的对手用AI，你就不得不用AI。那如果大家都用AI操控武器，乃至于进行战术级的指挥，出了事儿算谁的呢？

再进一步，根据现有的研究案例，我完全相信，如果我们把司法判决权完全交给AI，社会绝对会比现在公正。大多数人会服气，但是有些人输了官司会要求一个解释。如果AI说只是我的算法判断你再次犯罪的概率有点高，我也说不清具体因为啥高，你能接受吗？

理性人需要解释。有解释才有意义，有说法才有正义。如果没有解释，也许……以后我们都习惯于不再要求解释。

我们可能会把AI的决定当做命运的安排。

小李说：我没被大学录取。我的高考成绩比小王高，可是小王被录取了。一定是AI认为我的综合素质不够高……我不抱怨，因为AI自有安排！

老李说：是的孩子，继续努力！我听人说了，AI爱笨小孩！

你能接受这样的社会吗？

✳

AI到底是个什么东西？现阶段，它已经不是一个普通工具，而是一个法宝。你需要像修仙小说里一样，耗费巨量的资源去炼制它。

据摩根斯坦利分析，正在训练之中的GPT-5，用了25000块英伟达最新的GPU。这种GPU每块价值一万美元，这就是2.5亿了。再考虑研发、电费、喂语料的费用，这不是每家公司都玩得起的游戏。那如果将来训练AGI，又要投入多少？

但只要你把它训练好，你就得到了一个法宝。AI做推理不像训练那么消耗资源，但是用的人多了也很费钱，据说ChatGPT回答一次提问消耗的算力是一次Google搜索的十倍……不过有了它，你就有了一件人人想用的神兵利器。

而只要AGI出来，它就不再是一个工具了。它会成为你的助理。今天出生的孩子都是AI时代的原住民，AI将是他们的保姆、老师、顾问和朋友。比如孩子要学语言，直接跟AI互动交流比跟老师、跟家长学都会快得多、也方便得多。

我们会习惯于依赖于AI。我们可能会把AI给人格化，或者我们可能会认为人没有AI好。

那么再进一步，你可以想见，很多人会把AI当成神灵。AI什么都知道，AI的判断几乎总是比人类正确……那你说人们会不会从强烈*相信*AI，变成*信仰*AI？

AI可能会接管社会的道德和法律问题。

你猜这像什么？这就像是中世纪的基督教。

✳

在中世纪，所有人都相信上帝和教会，有什么事儿不是自己判断，而是去教堂问神父。那时候书籍都是昂贵的手抄本，普通人是不读书的，知识主要是通过跟神父的对话传承。

是印刷术出现以后，每个人可以自己读书了，直接就能获得智慧，不用迷信教会了，这才开启了讲究理性的启蒙运动。

启蒙运动对社会的改变是全方位的：封建等级制度、教会的崇高地位、王权，都不复存在。启蒙运动孕育了一系列政治哲学家，像霍布斯、洛克、卢梭等等，通过这些人的思考，人们才知道那个时代是怎么回事儿，以后的日子该怎么过。

抛开上帝拥抱理性，启蒙运动是给普通人赋能的时代。

而今天我们又开启了一个新的时代。我们发现人的理性有达不到的地方，可是AI可以达到，AI比人强。如果人人都相信AI，有什么事儿不是自己判断，而是打开ChatGPT问AI，知识主要是通过跟AI的对话学习……

再考虑到AI还可以轻易地向你推荐一些最适合你吸收的内容，对你进行定点宣传，你舒舒服服地接受了……

这不就是神又回来了吗？

再想一步。假设很快就有公司炼制成功了AGI，而AGI的技术特别难、炼制特别昂贵，以至于是其他人难以模仿的。那如果这些掌握AGI的公司成立一个组织，这个组织因为可以用AGI自行编码设计新的AGI，迭代越来越快，水平越来越高，领先优势越来越大，以至于任何人想要接触最高智能都必须通过他们……请问这是一个什么组织？

这个组织难道不就是新时代的教会吗？

✳

这就是为什么基辛格等人要写书呼吁，我们不应该把什么任务都交给AI，不能让AI自动管理社会。他们建议，任何情况下，真正的决策权应该掌握在人的手里。为了确保民主制度，投票和选举都必须得由人来执行，人的言论自由不能被AI取代或者歪曲。

这也是为什么OpenAI在声明中说：“我们希望就三个关键问题进行全球对话：如何管理这些系统，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，你会理解他们的忧患。我们正处于历史的大转折点上，这绝对是启蒙运动级别的思想和社会转折，工业革命级别的生产和生活转折——只是这一次转折的速度会非常非常快。

回头看，转折带来的不一定都是好事儿。启蒙运动导致过打着理性旗号的法国大革命和战争；工业革命导致了城市人口大规模增加、城市人口，而马克思那个时代的工人并不是很幸福。转折引发过各种动乱，但是最后社会还是接受了那些变化。AI又会引发什么样的动乱？将来社会又会有什么样的变化？我们会怎样接受？

基辛格等人认为，现在的关键问题——也就是“元问题”——是我们现在缺少AI时代的哲学。我们需要自己的笛卡尔和康德来解释这一切……

好，这一讲我们讲了被AI接管后的社会，可能会有的道德和法律问题。

下一讲，咱们说说AI怎样能让社会更聪明。

内容禁止商业 仅供社群内部免费交流使用
【知识福利社】微信:312005650

划重点　添加到笔记

1.AI对人类有一定的危险：AI全面接管科研就在眼前，它可能会接管社会的道德和法律问题，它对人类社会具有统治力，以及可能的破坏力。
2.我们正处于历史的大转折点上，这绝对是启蒙运动级别的思想和社会转折，工业革命级别的生产和生活转折——只是这一次转折的速度会非常非常快。

本课取材自：
万维钢·精英日课5
帮你和全球精英大脑同步
更多前沿认知 戳此了解
万维钢 科学作家

首次发布：2023年4月8日

上一篇 06 决策：AI的预测 + 人的判断
下一篇 08 经济：AI让资源调配更有效

可免费赠送给20位朋友
微信　朋友圈

用户留言

最新　最热　只看作者回复

优格 4月8日　+ 关注
突然在想，如果AI学会了说谎...
转发　4　4　分享

Skye 4月8日　+ 关注
突然开了一个脑洞：
第一代计算机（类比GTP5或AGI）可以用巨型来形容，后来技术突很快变成微型计算机等
历史时代变化重要一个因素就是新的技术变革，万一有一天我们真的能见证那个新时代呢？
转发　评论　3　分享

柯洁的阿法狗 4月8日　+ 关注
这是一场有关新时代博弈论。怎么可以让各大场公平使用资源是当前一大难题，同时agi没有被使用得当会变成什么？听完之后脊骨发凉。只能感慨，时代转折点机遇与危机并存啊
转发　评论　3　分享

甜小姐 4月8日　+ 关注
我同意AI只能作为助手，最后决定权应该在人的手上。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，不会碍于人情，不会收受贿赂。但真的涉及这些领域，怕是，人也是舍不得放手让AI去决定的。所以，AI 能决定的领域可能是不涉及利益的。
1　12　71　分享

亚义 南京 4月8日　+ 关注
我觉得需要设计一个“反AI推荐”的按钮，允许适时的关闭AI”反效率”。提醒选择的底层逻辑可以呈现多维度化，看到更多的可能性和疆域，不能只看到选择的效率和愉悦感，进入信息茧房。
转发　评论　7　分享

丁梓怙 4月8日　+ 关注
人类总是低估自己的适应性。真到了那一天，人类不知不觉就习而不察了。
转发　2　11　分享

达成骞 4月8日　+ 关注
感觉需要一个个性化客户端，当启动AI后设置时约定：当给我的回答和帮助有明显违背我的个人原则和我希望成为的形象时，请闭嘴。
转发　1　9　分享

以上留言由作者筛选显示

【知识福利社】你的私人书院，微信：312005650 全网热门课程众筹共享！