01 大变局：一个新智慧形态的产生

万维钢·AI前沿
2023年4月6日

进入课程 ›

时长：13分15秒 大小：9.09MB

00:00 / 13:15

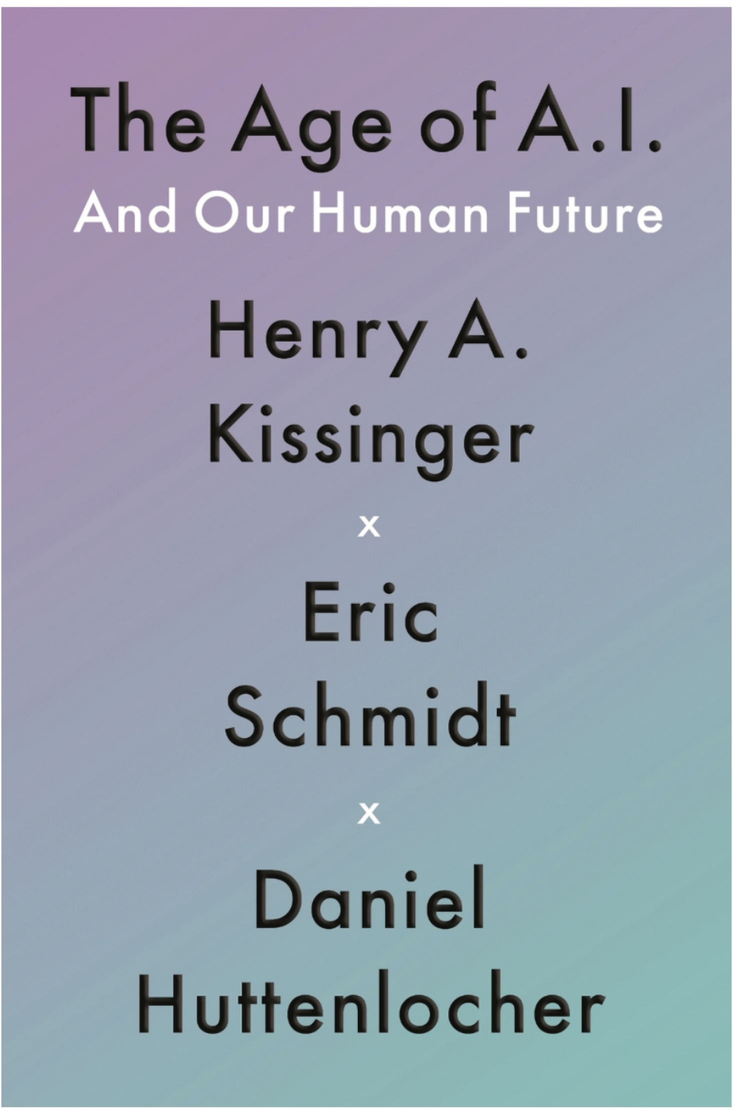
转述：怀沙

你好，这里是《万维钢·AI前沿》课。这一讲，我们来讲一个新智慧形态的产生。

ChatGPT让2023年成了AI之年，正如iPhone在2007年开启了智能手机时代，我们现在正式进入了人工智能时代。

咱们的这门课程就会来跟你聊聊对这个时代的最新理解和思考，以及AI的应用场景和应对方法。

这一讲我们先从2021年出的一本书讲起，书名叫《AI时代：以及人类的未来》（*The Age of AI: And Our Human Future*）。



这本书的三个作者很厉害，分别是无需介绍的亨利·基辛格，Google前CEO埃里克·施密特，麻省理工学院苏世民计算机学院的院长

丹尼尔·胡滕洛赫尔（Daniel Huttenlocher）。

作者咖位这么高，但这并不是一本献礼式、应景式的书，这书里有真思想，有关于这个时代的高观点。

* * *

2020年，麻省理工学院宣布发现了一种新的抗生素，叫Halicin。这是一种广谱抗生素，能杀死那些对市面上现有的抗生素已经产生耐药性的细菌，而且它自己还不会让细菌产生耐药性。

Halicin
Discovered by AI

这个幸运的发现，是用AI完成的。研究者先搞了一个由两千个性能已知的分子组成的训练集，这些分子都被标记好了是不是可以抑制细菌生长，用它们去训练AI。AI自己学习这些分子都有什么特点，总结了一套“什么样的分子能抗菌”的规律。

AI模型训练好之后，研究者用它一个个考察美国FDA已经批准的药物和天然产品库中的61000个分子，要求AI按照三个标准从中选择一种抗生素：1）它具备抗菌效果；2）它看起来不像已知的抗生素；3）它必须是无毒的。

结果AI最后只找到一个符合所有要求的分子，这就是Halicin。然后研究者做实验证明，它真的非常好使。它大概很快就会用于临床，造福人类。

用传统的研究方法，这件事是绝对做不成的：你不可能测试61000个分子，那成本太高了。这只是当代AI众多的应用案例中的一个，它很幸运但是它并不特殊。

我们之所以先讲这个例子，是因为它带给我们一个清晰的认知震撼——

Halicin可以作为抗生素的化学特征，是人类科学家所不理解的。

关于什么样的分子可以做抗生素，科学家以前是有些说法的，比如原子量和化学键应该具有某些特征——可是AI这个发现用的不是那些特征。AI在用那两千个分子训练的过程

中，找到了一些不为科学家所知的特征，然后用那些特征发现了新的抗生素。

那些是什么特征呢？不知道。整个训练模型只是一大堆——也许几万到几十万个——参数，人类无法从那些参数中读出理论。

这可不是特例。AlphaZero完全不用人类棋手的棋谱，通过自己跟自己对弈学会了下国际象棋和围棋，然后轻松就能打败人类。然后你看看它的走法，它经常会走一些人类棋手匪夷所思、没有考虑过的走法。比如在国际象棋里它看似很随便就可以放弃皇后这样的重要棋子……有时候你事后能想明白它为啥那样走，有时候你想不明白。

这个关键在于，AI的思路，不同于人类的理性套路。

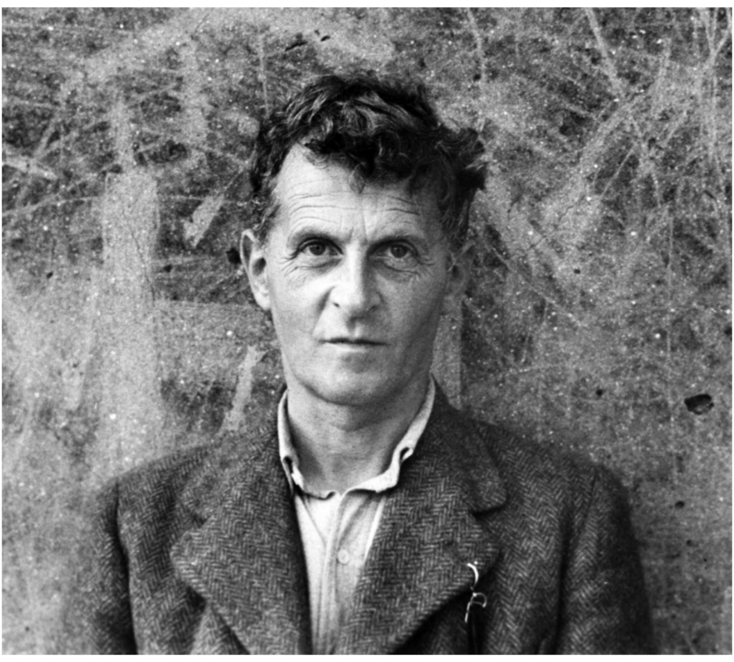
也就是说，当代AI最厉害之处并不在于自动化，更不在于它像人，而在于它*不像*人：它能找到人类理解范围之外的解决方案。

⁂

这不是汽车取代马的发明，也不仅仅是时代的进步。这是哲学上的跨越。

人类从古希腊、古罗马时代就在追求「理性」。到了启蒙运动，人们更是设想世界应该是由一些像牛顿定律这样的明确规则确定的，康德以后人们甚至想把道德也给规则化。我们设想世界的规律应该像法律条文一样可以一条条写下来。科学家一直都在把万事万物分门别类，划分成各个学科，各自总结自己的规律，打算最好能把所有知识编写进一本百科全书。

然而进入20世纪，哲学家维特根斯坦提出了一个新的观点。他说你们这种按学科分类写条文的做法根本不可能穷尽所有的知识。事物之间总有些相似性是模糊的、不明确的、难以用语言说明的。想要丁是丁卯是卯全都理性化，根本做不到。



[Ludwig Wittgenstein, 1889-1951]

现在AI找到的，恰恰就是一些难以被人所理解，不能用明确的规则定义的智慧。这是柏拉图理性的失败，是维特根斯坦的胜利。

其实不用AI你也能想明白这个道理。比如说，什么是“猫”？你很难精确定义猫到底是什么东西，但是当你看到一只猫的时候，你知道那是猫。这种认知不同于启蒙运动以来人们说的规则式的理性，但是你可以说这是一种「感觉」。

一种难以明说、无法告诉另一个人的感觉。我们对猫的认识很大程度上是感性的。

而现在AI有这种感觉。当然，人一直都有这种感觉，这本来没什么，康德也承认感性认知是不可缺的。问题是，AI通过这样的感觉，已经认识到了一些人类无法理解的规律。康德原本认为只有理性认知才能掌握世界的普遍规律。

AI感受到了人类既不能用理性认知，也感受不到的规律。而且它可以用这个规律做事。

人类已经不是世界规律唯一的发现者和感知者。

你说这是不是启蒙运动以来未有之大变局。

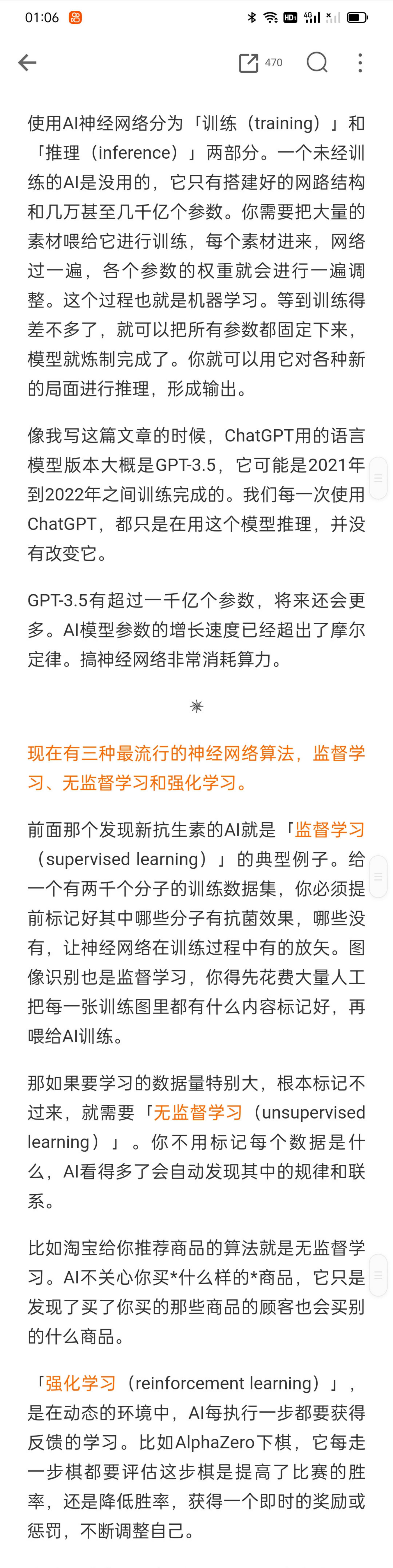
* * *

我们简单介绍一下当代AI的原理。现在有些人谈论AI是把AI当做了一种“超级智能”，仿佛神灵一般，说能把人类如何如何——那种讨论没什么意义。如果神灵都已经降临人间了我们还在这聊什么？不要高推圣境。

现在的AI不是什么通用人工智能（AGI），而是一种非常特殊的智能，也就是通过机器学习训练的神经网络系统。

上世纪八十年代以前，科学家还在尝试用启蒙运动理性的思路，把解决问题的规则输入给计算机执行。后来发现那条路走不通，因为规则太多了，根本弄不过来。这才有了神经网络。现在是我们根本不用告诉AI任何规则，也就是把学习世界的过程都委托给机器，有什么规则你自己学去吧。

这个思路受到了人脑神经网络的启发，但是并不完全一样。我在精英日课第三季讲过，AI神经网络的基本概念，它分为输入层、很多中间层和输出层，一般的深度学习网络大概要有10层。



使用AI神经网络分为「训练（training）」和「推理（inference）」两部分。一个未经训练的AI是没用的，它只有搭建好的网路结构和几万甚至几千亿个参数。你需要把大量的素材喂给它进行训练，每个素材进来，网络过一遍，各个参数的权重就会进行一遍调整。这个过程也就是机器学习。等到训练得差不多了，就可以把所有参数都固定下来，模型就炼制完成了。你就可以用它对各种新的局面进行推理，形成输出。

像我写这篇文章的时候，ChatGPT用的语言模型版本大概是GPT-3.5，它可能是2021年到2022年之间训练完成的。我们每一次使用ChatGPT，都只是在用这个模型推理，并没有改变它。

GPT-3.5有超过一千亿个参数，将来还会更多。AI模型参数的增长速度已经超出了摩尔定律。搞神经网络非常消耗算力。

* * *

现在有三种最流行的神经网络算法，监督学习、无监督学习和强化学习。

前面那个发现新抗生素的AI就是「监督学习（supervised learning）」的典型例子。给一个有两千个分子的训练数据集，你必须提前标记好其中哪些分子有抗菌效果，哪些没有，让神经网络在训练过程中有的放矢。图像识别也是监督学习，你得先花费大量人工把每一张训练图里都有什么内容标记好，再喂给AI训练。

那如果要学习的数据量特别大，根本标记不过来，就需要「无监督学习（unsupervised learning）」。你不用标记每个数据是什么，AI看得多了会自动发现其中的规律和联系。

比如淘宝给你推荐商品的算法就是无监督学习。AI不关心你买*什么样的*商品，它只是发现了买了你买的那些商品的顾客也会买别的什么商品。

「强化学习（reinforcement learning）」，是在动态的环境中，AI每执行一步都要获得反馈的学习。比如AlphaZero下棋，它每走一步棋都要评估这步棋是提高了比赛的胜率，还是降低胜率，获得一个即时的奖励或惩罚，不断调整自己。

自动驾驶也是强化学习。AI不是静态地看很多汽车驾驶录像，它是直接上手，在实时环境中自己做动作，直接考察自己每个动作导致什么结果，获得及时的反馈。

我打个简单的比方——

* 监督学习就好像是学校里老师对学生的教学，对错分明有标准答案，但是可以不给讲是什么原理；

* 无监督学习就好像一个学者，自己调研了大量的内容，看多了就会了；

* 强化学习则是训练运动员，哪个动作出错了立即给你纠正。

* * *

机器翻译本来是典型的监督学习。比如你要做英译中，就把英文的原文和中文翻译一起输入给神经网络，让它学习其中的对应关系。但是这种学法太慢了，毕竟很多英文作品没有翻译版……后来有人发明了一个特别高级的办法，叫「平行语料库（parallel corpora）」。

先用对照翻译版来一段时间的监督学习，作为「预训练」。模型差不多找到感觉之后，你就可以把一大堆同一个主题的英文也好、中文也好，别管是文章还是书籍，不需要互相是翻译关系，各种材料都直接扔给机器，让它自学。这一步就是无监督学习了，AI进行一段沉浸式的学习，就能猜出来哪段英文应该对应哪段中文。这样训练不是那么精确，但是因为可用的数据量要大得多，训练效果好得多。

购买课程微信：15612277539

像这种处理自然语言的AI现在都用上了一个新技术叫「transformer」。它能更好地发现词语跟词语之间的关系，而且允许改变前后顺序。比如“猫”和“喜欢”是主语跟谓语的关系，“猫”和“玩具”则是两个名词之间的“使用”关系，它都可以自行发现。

还有一种流行技术叫「生成性神经网络（generative neural networks）」，特点是能根据你的输入生成一个什么东西，比如一幅画、一篇文章或者一首诗。生成性神经网络的训练方法是用两个具有互补学习目标的网络相互对抗：一个叫生成器，负责生成内容，一个叫判别器，负责判断内容的质量，二者随着训练互相提高。

GPT的全称是「生成式预训练变换器（Generative Pre-trained Transformer）」，就是基于transformer架

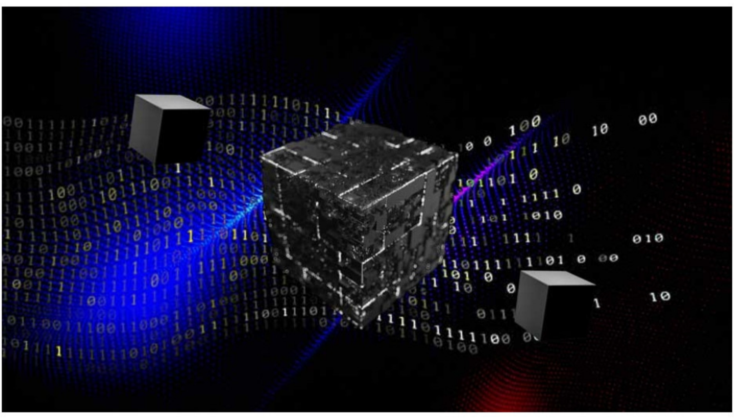
构的、经过预训练的、生成性的模型。

* * *

当前所有AI都是大数据训练的结果，它们的知识原则上取决于训练素材的质量和数量。但是，因为现在有各种高级的算法，AI已经非常智能了，不仅能预测一个词汇出现的频率，更能理解词与词之间的关系，有相当不错的判断力。

但是AI最不可思议的优势，是它能发现人的理性无法理解的规律，并且据此做出判断。

AI基本上就是一个黑盒子，吞食一大堆材料之后突然说，“我会了”。你一测试发现它真的很会，可是你不知道它会的究竟是什么。



因为神经网络本质上只是一大堆参数，不可理解性可以说是AI的本质特征。事实是连OpenAI的研究者也搞不清ChatGPT为什么这么好用。

要这么说的话，可以说我们正在目睹一个新智慧形态的觉醒。

好，以上就是这一讲的全部内容。

下面的链接是我让ChatGPT给你出了五道题，分别是五个AI应用场景，考一考你记不记得它们用的是什么学习方法。我们下一讲见。

➡️点击这里，试试能得多少分？

划重点　添加到笔记

1.AI感受到了人类既不能用理性认知，也感受不到的规律。而且它可以用这个规律做事。人类已经不是世界规律唯一的发现者和感知者。
2.现在有三种最流行的神经网络算法，监督学习、无监督学习和强化学习。