



延伸阅读

智能机器人会超越人类吗

“2045年,当计算机智能超越人类。”在库兹韦尔著作《奇点临近》中译本的封面上写着这样一句话。这本书是2005年的畅销书,如今算来,离库兹韦尔预测的时间还有30年。这一个预测结果得自库兹韦尔的加速回报定律(Law of Accelerating Returns),该定律认为人类社会是加速发展的,一个更加发达的社会本身具备的继续发展的能力也更强,按照这条定律,2000年的发展速度是20世纪平均发展速度的5倍,人类在21世纪的进步又将是20世纪的1000倍。而这一切试图让人们相信奇点正在临近,技术将以近乎无限的速度发展。“从弱人工智能走向强人工智能可能需要几十年,从强走向超甚至有人预测是两年,最长的预测时间也是30年。大家都同意从强到超会变得非常快,如果失控,在强的阶段就会出现,到了超就彻底失控。《超体》虽然是幻想人脑的开发,但露西发展的过程恰恰是强人工智能到超人工智能过程的一个图解。一旦越过了界限,知识会迅速膨胀。这部电影跟现在关于脑科学、人工智能的想象关系非常密切。”上海交通大学科技史与科学文化研究院院长江晓原说。

智能超越,的确会破坏人类的安全感,毕竟智能是人类站上食物链顶端的重要凭恃。江晓原和穆蕴秋博士刚刚翻译完成的《Nature 杂志科幻小说选集》,其中有10来篇涉及人工智能,故事结局大多荒唐,字里行间透着隐忧。“西方做科幻与做前沿的人之间似乎形成一种不成文的规定,科幻专门用来对研究的内容质疑,想象其发展所带来的一些不好的未来。在《Nature》上的小说作者都是很入流的,都是跟着西方现在的科幻潮流在写。”以此反观,对于人工智能的担忧似乎并非空穴来风。

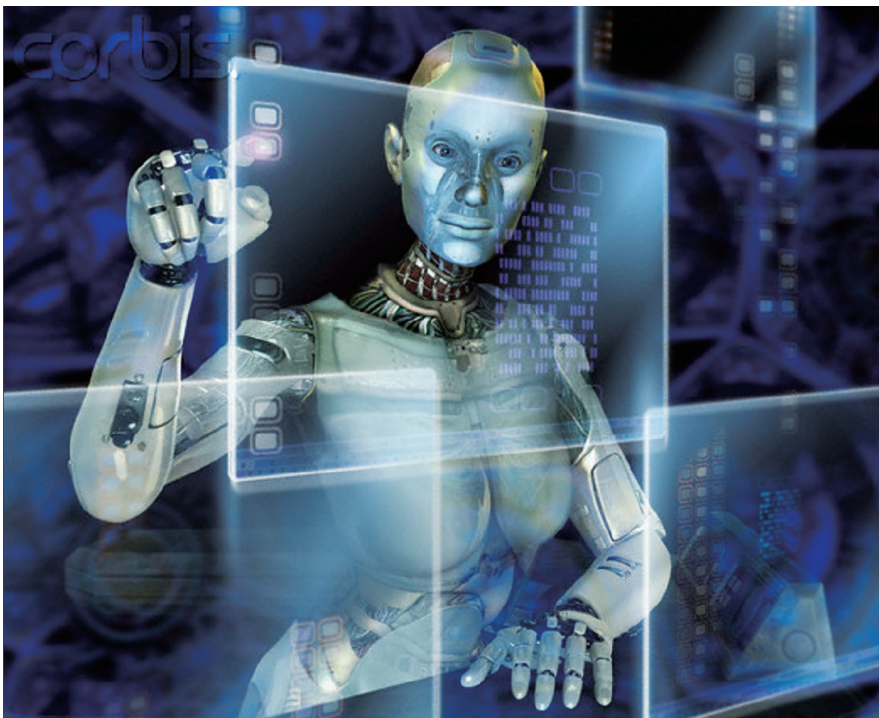


图1 操作未来计算机的机器人
(图片来源:www.corbisimages.com)

这种担忧,大概可以分为3个层面:第一层是技术上的滥用、技术缺陷对人类社会造成伤害;第二种是哲学层面的担忧,即人对技术过于依赖,把所有事情都交给计算机决策,人类丧失主体地位,更丧失生活的意义;第三层是最为可怕的,人工智能变坏,在意识的驱动下伤害人类,这更像是电影《终结者》中的场景。

超越前的“九九八十一难”

人工智能会超过人类智能吗?这种争论在学界一直存在。坚持人工智能不能超越人类的人不在少数,依据是人工智能是人造物,是在不断模拟人类,模拟者怎么会超越被模拟者?

“这种想法的论据都是站不住脚的。”江晓原说,“没有任何一条定理证明被创造的东西不能超过造物主,现实

生活中反而有很多相反的例子,比如孩子比父母聪明,就非常典型。虽然目前还在讨论人工智能能不能通过图灵测试,离超过人类还差得太远,但那种认为人工智能永远不能超过人类的想法我不赞成。”他认为,这类东西的问题在于技术的突破是不可知的,突破也有可能很快就到来。“人工智能和生物工程是人类现在在做的最为危险的两件事。”

反驳“不会超越说”的一个重要依据是机器强大的升级能力,离我们生活最近的例子就是电子邮箱中根据收件人习惯筛除垃圾邮件的程序。这一点是霍金提出警示的重要原因,他担心,一旦发展出相当于或者超越人类智慧的人工智能技术,它就会脱离控制,以不断加快的速度重新设计自己。但是人类受到缓慢生物进化的限制,无法与

其竞争,甚至可能最终被超越。库兹韦尔也在书里写道:一旦人工智能达到人类的水平,它一定会很快超越人类水平,因为它会把人类智能的力量和非生物展现出来的智能(包括速度、内存容量、知识共享)结合在一起。机器学习是人工智能学习人类智能的重要方式,通过大样本的数据、信息让计算机从中发现学习知识经验、自动改进计算机算法。美国辛辛那提大学智能维护中心主任李杰认为人工智能的确有学习能力,但那属于程式里的学习,不会跳到程式之外,人工智能只能在指定领域升级。“不要忘了,人工智能里都有一个核心东西叫软件,软件不会自己思考,一定是人编程的。”

目前人工智能在模拟人类智能上还存在诸多障碍,这让很多人相信“终结者”的出现还非常遥远。

比如现阶段让机器学习人类的知识就不容易。这种学习有个困境——计算机学到的知识未必是人类全部的知识。李杰认为,人在自身智慧的基础上学习别的智慧的功能,就不能被机器复制。“比如人会因为某一条信息醍醐灌顶,人类称之为灵感,灵感是精神上的、无形的、不能编程的。我认为,机器可以取代知识,可以慢慢帮助人类做决策,但是灵感是其无法实现的,灵感是由人的智慧(wisdom)跟个人的精神所

造成的一种很即时的冲击,无法量化描述,而机器是以目标为准。当以目标为准时,比如在打字速度、翻译速度、计算能力上,人脑的确竞争不过电脑,但当人以价值为准,道德价值、意识价值,电脑就不知如何与人相比。”

难以数字化处理的并非只有灵感。人有个特质在现象学上叫具身性——有身体、有感知,可以跟周边环境接触获得知识,这种知识与不跟外界接触,完全靠推理得来的知识不一样,打个比方,眼睛的转动可以帮注意力集中做出多种选择,这是人生活在世界上跟周边环境接触形成的,很难提炼出来放到计算机中去。再如两条腿走路,要在机器人身上实现并不容易,因为涉及人如何使用力气来抵抗地心引力,这是人上百万年来生物进化后自然而然的反应,要把这个技能加诸机器人身上,会涉及大量的计算,而且这种运转需要非常复杂的机电控制系统,保证它像人体四肢那样可靠地精密运转有很大挑战。概言之,人的知识来自于社会和交往,这种知识很难形式化,知识的形式化边界至少在目前来看也是数字计算机的能力边界,人工智能缺乏具身性特征,难以从周边环境中获得这些不能形式化的知识。

知识学习已经够复杂了,但这还不是人类智能的全部。前几年机器人“华

生”参加美国答题节目,已经超越了最好的美国选手,它存储了很多词典、报纸,也可以做简单推理,已然是知识丰富了,但并没有人把它看作是一个完整的人。“人除了知,还有情和意,意是意志信仰,情跟人的群体有关,知、情、意是绑定一起的,很多知识的正确与否取决于情和意,比如宗教信仰和社会背景。作为社会动物,人做出的决策并不完全是知识表达,有很多情和意的因素在里面,所以很难还原出一个算法来表现人类如何做决策。”出身计算机专业的上海交通大学科学史在读博士陈自富,是个AI(人工智能)发烧友,他的这个观点与李杰相仿。在李杰看来,人类是团结的动物,人类的智慧来自于群体间的学习与成长,人类智慧不是一个单纯的机械,而是一个智能化集成技术。简而言之,智慧是智能化的结晶。人工智能可能在“单挑”中胜出,但是超出整个社会的人类智慧不太可能。

“即便是最乐观的计算机科学家也认为人工智能在21世纪达到单个人的智力水平,这已经不错了,但其中不涉及情和意。”陈自富认为,知的方面,在未来30多年内,计算机可以达到一个很高的水平,人工智能大致能帮助普通人做出不那么糟糕的决策,比如帮人计算存一笔钱是否划得来,存定期还是活期,买哪支股票。“对于确定的任务,人工智能可以做得到,而一些非确定性的任务则很难。说白了,机器只有不断进化,成为我们社会的一员,才能达到人类水平。”

人类意识难以被模拟

人工智能最终达到或超越人类智慧,必须要跨越机器与人的差别。这种差别主要是理性、信仰和群体交往中的情感互动、文化基因。创造出有意识、有情感的人工智能,怀此愿望的科学家不在少数,阿森克特说过,我们能够实现让机器人像人类一样思考,但构建完美的机器人脑需要很长时间。

蓝脑计划(BlueBrain)领军人物亨利·马克莱姆(HenryMarkram)2009年曾声称10年内推出人类意识的全息

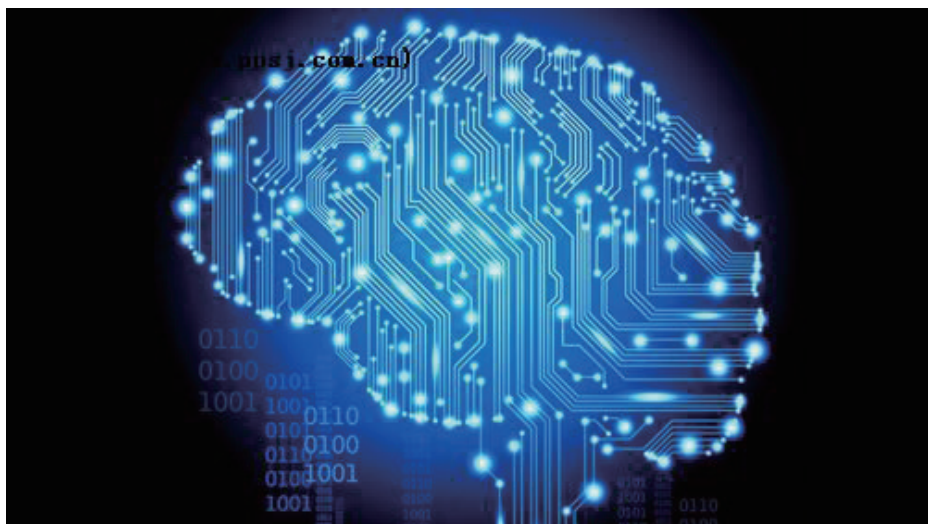


图2 IBM超级计算机模拟人类大脑5千亿神经元(图片来源:it.ppsj.com.cn)



图,意图建立一套具有自我意识的人工智能。完成的方法是建立一个从突触到半球完整人脑模型,并在一部超级计算机上模拟。尽管他的实验室已成功模拟小鼠单个脑皮层单元的活动,但是似乎距离还是很远。

“西方人常说,生命是一个奇迹,那么脑更是一个奇迹,脑是生命里最高级的组成部分,人们并没有把这个奇迹的所有机制都搞清楚。”江晓原说,当人们说生命是一个奇迹,跟说苹果是一个奇迹,意思不一样,后者只是溢美之词,其构造运作机制都可以搞清楚,而前者是因为不知道生命里面到底发生了什么事,人脑也是一样。

让我们看看把大脑神经元的工作原理弄清楚有多难。人脑是由连接着1000亿个神经元的100万个神经突触组成的网络,这些神经突触和神经元的状态每秒改变10~100次,而线虫只有302个神经元,人们尚且都搞不懂其机制。“大脑是个复杂的、开放的巨系统,跟外界一直有信息交换,用计算机模拟整个巨系统难度非常大,目前来讲还差十万八千里。”陈自富说,现在很多人通过计算机程序模拟神经单元,比如谷歌的一个项目大概可以实现10万~100万个大脑神经元的处理功能,这可能还达不到一只猫智力水平,更何况它的算法是不是就是人脑中的算法,都没法判断。就连马克莱姆自己都无法对“模型建出来,能不能做思考”这个问题给出答案。

事实上,神经系统的计算方式和计算机有着本质的区别。计算机的状态转变完全由程序设定好,而神经系统从一个状态到另一个状态则非常依赖于当时的境况。“人脑处理事情会做很多剪枝,当你端起杯子往外倒水,其实有多个分支,大脑会给出一个直接判断,把一些不必要的判断剪掉。但是机器人做这件事,就是对各个分支一个接一个地做可行性计算,这个叫做蛮力计算,所有的计算都要消耗能量。再者,从技术上来看,大脑神经元虽多,其运算速度是缓慢的,但对事情的处理是并

行的。如果计算机实现 10^{10} ~ 10^{12} 级别的大脑神经元处理功能,目前的技术条件下散热就是一个难以处理的问题。”陈自富认为,人工智能学习人类的这种缓慢的处理机制,需要认知科学的进步。但是搞清楚大脑的大部分原理就要数十年,甚至上百年。可以说,以现实的技术路径来看,在模拟人类智能的路径上完成超越很难,“终结者”似乎还距离我们几个星系。

人工智能发展充满太多未知

不过,担忧者也有担忧者的道理,既然人脑中产生意识、情感的机制不清楚,那么人工智能在升级过程中能否自主地形成意识也就难以说清。

江晓原的电脑里有一个下棋程序,其程序设定是不断通过互联网学习高手对局。“随着它的不断升级,我和它对局的胜率已经非常低。但是后来我找到一种发源于明代的古老棋招,由于这种棋招在互联网数据库上没有记录,就凭这一招,我的胜率从不到20%提升到70%。”江晓原说,其实下棋程序若想赢过这一招,也许可以写入一个“多次输棋后就跟对手学习”的规则。“这个规则能不能超越人类设定,由程序自己生成?如果有这种例证,就说明人工智能有了自主学习的能力,那就很可怕。这种事很玄,现在也很难有实证。”江晓原认为,虽然目前现实生活中还没有看到科幻作品中所担忧的那种现象,但危险在于一旦这种事情出现,人们不一定会马上看出来,等到迹象明显,可能已经来不及了。“我曾经设想过,如果真的有的人工智能反叛人类,它们肯定会互相约定一个行动原则,坚决不要通过图灵测试,免得让人类警惕。”情感自发的产生也是有可能的,江晓原认为,当谈话程序能聊天,能让客户觉得放松缓解,它不可能完全不了解情感。“智能和情感之间也没有什么不可逾越的鸿沟,复杂到一定程度可能就有情感了。”

不久前,荷兰埃因霍温大学发布了RoboEarth项目,在模拟医院里,4个机器人协作照顾病人,通过云端服务器实

现信息共享,比如一个机器人扫描医院房间,将房间地图上传至“云端”,另外一个对该房间完全不了解的机器人就可以通过访问存储于“云端”的地图了解房间内物品的放置情况。目前接入云端进行测试的只是功能性机器人,但有人提出了这样的担心——将来每个接入云端的机器人可能程序差异很大,随着接入机器人越来越多,群体思维和群体协作产生的效果将愈加显著,而差异性和群体性将共同指向一点——进化,这意味着人工智能将会具备学习、适应、自我治愈,甚至是进化的能力。正如人工智能专家詹姆斯·巴拉(James Barrat)所担心的:当所有的机器人都具备了足够的学习能力之后,就变得不再依靠人类,如果加上全世界的机器人都可以进行交流和联系,它们很有可能变成新的支配者。

“可怕之处在于我们不知道它们能学到什么样的知识,如何用这些知识。”陈自富认为,人类知识是很庞杂的系统,价值观多元。有了人类知识,人工智能怎么去塑造情和意,或者说价值观,这并不可控。

“互联网与机器人的结合,会使风险加大,很多科幻作品里都做过想象。”江晓原说,阿西莫夫在作品中曾谈到机器人智能的刚性极限,单个智能机器人做成人状,其身体体积是有限的,智能也是有限的,因为电脑的微型化有刚性极限。但是现在有了互联网,连互联网的机器人就突破了这种刚性极限。“好莱坞作品《鹰眼》讲的就是互联网与摄像头接轨后发生的事。互联网自己就会变成一个人工智能,可以做坏事。即便没有伺服机构,也可以像《疑犯追踪》里讲的,自己订一个产品进行组装,用不着非得化身一个机器人,它可以充分利用资源,因为这样的超级大脑可以让千百台计算机都受它控制。”

文/刘力源

(本文转载自2015年4月17日《文汇报》。原题为“人工智能是人类的‘终结者’吗?”,现题为本刊所加)

(编辑 王丽娜)