

计算机科学家姚期智、历史学家葛兆光眼中的历史与未来——人工智能，能否跨越智慧的疆界

见习记者 涂佳煜

“机器能思考吗？”

70多年前，计算机科学之父艾伦·图灵在《计算机机械与智能》一文的开头提出这一问题，至今仍被追问不已。

人类的思维难以定义。图灵因而设计了一场“模仿游戏”，通过检验计算机替代人类后的游戏结果，倒推思维是否存在于机器中。后来，这一设想发展为我们熟知的“图灵测试”：计算机假扮人和裁判对谈，如果超过30%的裁判误以为和自己说话的是人而非计算机，那么就可以认为计算机“能思考”或“有智能”。

实际中，“图灵测试”并非判断人工智能的标准，但“有思维的机器”这一命题却为众多科幻作品提供了灵感。

不论是《2001太空漫游》中不具人形却有自我保全意识的计算机哈尔9000，还是《人工智能》中惹人怜惜的“类人”男孩戴维，拥有拟人类思维甚至情感的“荧幕机器”，无一不演绎出催人深省的故事，建构了许多人对“人工智能”的终极幻想。

然而，如果我们从科幻世界转向现实生活，看到的“人工智能”却是另一种截然不同的形态——语音助手、人脸识别、智能家居服务我们的日常生活，自动驾驶、智能制造、智能计算推动科技革命和产业变革。AlphaGo与李世石的世纪大战中，“机器战胜人类”这一科幻电影里的可怕结局，反被誉为人类技术胜利的又一次证明。

究竟，什么是“人工智能”？它将把人类的未来引向何方？当我们跨越“智慧的疆界”，又该以何种姿态拥抱新一轮的技术革命？近日，在西湖大学的湖心讲堂上，围绕这些话题，计算机科学家姚期智和历史学家葛兆光，展开了一场科技与人文的跨界对话。

寻找“智慧”源头

“怎样教一部计算机识别出手写的阿拉伯数字‘8’？”姚期智向观众抛出了个“脑筋急转弯”。

乍一看，“8”可以被分解为两部分，因此，可以告诉计算机，在图像中找到两个圈，判断它们的相对位置，这是传统的机器学习方法。姚期智说，“但‘深度学习’舍弃了这种人工分析曲线特征的思路，只给机器提供大量原始手写数据，通过优化算法结构，就能训练它最终辨别出‘8’来。”

深度学习赋予了人工智能强大的“自学”的能力，它的基础是被称作“人工神经网络”的算法模型。物如其名，该模型的灵感正来自人类自身的“智慧中枢”——大脑。

从某种意义上说，人脑可谓世界上最复杂精密的计算机。人脑大约有860亿神经元。在学习新知识时，外界信号刺激神经元间形成突触，信息便开始传递，同一个信号的反复刺激，则能强化突触传递该信息的能力。一个人的“知识网”，正是随着神经网络的生长越来越细密，越来越牢固。

作为人脑结构的数学抽象，“人工神经网络”中同样有一层层排布的“神经元”节点，它处理数据的过程也与突触传递信号颇为相似：输入层“神经元”接收初始信息后，上一层的数据经过加权函数运算激活下一层“神经元”，层层正向推进，直到在输出层得到结果。对比结果与预期，优化信息又“反向传播”，“训练”神经网络调整权重和偏差。这样一来，一旦有了海量输入数据的支撑，神经网络就能在循环往复的训练中渐趋完善。

“深度学习就好比换一种方式学语言。”姚期智又形象地打了个比方，“成年人学语言，要分析语音、语法、词汇，现在我们用小孩学语言的方法，让他浸润于该语言的母语环境中，自己琢磨，自然而然他就能说得很流利，效果还可能更好。”

早在上世纪末，图灵奖得主、深度学习领域的先驱人物杨立昆(Yann LeCun)就率先使用反向传播和卷积神经网络，开发出了能识别手写数字的系统LeNet，用于自动读取银行支票上的数字。这是深度学习在计算机视觉中的最早应用之一。但在当时，它还是默默无闻的小众算法，没有在学界业界掀起太大波澜。

直到2012年，深度学习才再度出现在大众视野。在这一年的ImageNet大规模视觉识别挑战赛上，基于深度学习卷积神经网络的AlexNet一战成名，图像识别错误率远低于第二名10%，开启了人工智能热的新时代。AlexNet也“开宗立派”，后来出现的许多新算法中无不有它的影子。到今天，深度学习在图像识别、语音识别、自然语言等领域大展身手，一度成为人工智能的代名词。

有人说，围棋是人类智慧的“最后一道防线”。真正让深度学习“出圈”的标志性事件非AlphaGo和围棋世界冠军李世石的人机大战莫属。倘若探究AlphaGo的制胜奥秘，除深度学习多层神经网络外，强化学习方法也功不可没。

姚期智以“老鼠吃奶酪”的经典案例解释当前在游戏和机器人领域应用广泛的“强化学习”：想象有一只老鼠在走迷宫，在出口它能吃到奶酪，每走错路它就会撞到墙上，久而久之，它再进入迷宫时就能根据记忆走出去。强化学习的关键，就在于让计算机自主与环境互动，在正负反馈中找到通路。

AlphaGo的训练就好比在棋局的迷宫中不断搜索胜利的“奶酪”——它并不钻研“棋艺”，而是通过吸收大量人类专家的棋谱，再进行大量自我博弈，根据结果评估得失，总结出走迷宫的“经验”。

人机对弈仅一年后，使用强化学习的AlphaGo Zero更是摆脱了除围棋规则外的任何棋谱，经过3天的训练就无师自通，以100:0的成绩击败了前辈AlphaGo。至此，人工智能棋手意义已不再是与人类冠军一决高下，它自身的迭代与突破，就足以证明人类智慧的胜利。

“深度学习和强化学习这两大算法引擎，极大推动了过去十几年我们目睹的人工智能‘新浪潮’。”姚期智总结道。



姚期智(线上)、葛兆光(中)、施一公(右)交流对话 西湖大学提供



2022世界人工智能大会在沪开幕，工作人员向参观者介绍一款智能文字识别AI系统对青铜鼎铭文的识别展示。

新华社发

学科疆界如何打破

“从今天起，确定几乎全部已知蛋白质的三维结构，都会像使用谷歌搜索一样简单。”《自然》杂志今年7月28日宣告，DeepMind公司研发的AlphaFold人工智能系统已经预测出约100万个物种中的2亿多种蛋白质结构，几乎涵盖了地球上所有已知蛋白质。整个生命科学界再度为之震颤。

蛋白质是生命不可或缺的元素，独特的折叠结构则决定了每种蛋白质的不同功能。五十年前，科学家就开始艰难攻克这个极富挑战性的生物现象，直到人工智能的出现带来了质的飞跃。

这场颠覆性革命几年前就已萌芽。2018年，AlphaFold首次在国际蛋白质结构预测大赛上亮相，预测准确率比得分最高的人类团队高出近20分。到了2020年，它的全局描述符表总体平均分已超90分，与人类的实验结果不相上下。换句话说，结构生物学家用复杂实验手段解析出的结构，人工智能通过计算就能得到。

“原子分辨率的结构常被看作是生物学问题的终极解释。以往生物学研究遵循着一定的模式，从发育到遗传，寻找到基因之后再回到细胞生物学探寻其功能，然后到生物化学研究对应的蛋白质如何起作用，最后再到结构来解释。”现场，西湖大学校长、结构生物学家施一公说，“人工智能的出现，打破了这一研究常规，把生命的结构基础瞬间拓展了几个数量级，结构走到了

生物化学、细胞生物学，甚至遗传学前面。”

如果说AlphaFold把我们对已知生命世界的认知推向了极致，那么当我们把目光转向物理学，自然界中不曾存在的未知物质，也在人工智能的帮助下被创造了出来。

姚期智说，“时间晶体”的实现就是一个很好的例子。十年前，“时间晶体”的概念被首次提出——这种奇妙的物质如同四维空间的万花筒，具有“时间平移对称性被破坏”的特征，即在时间上有规律地重复自身的状态，如“永动机”一般不消耗任何能量。

今年7月，浙江大学和清华大学团队在超导量子芯片上，结合人工智能算法和全数字化量子模拟，创造出了全新具有“拓扑性”的“时间晶体”——链状晶体边缘和内部特征不同，只有两端呈现出时间平移性的破坏。人工智能与量子计算强强联手，展现出在量子世界探索更多新奇物质和现象的巨大潜能。

不断涌现的成果表明，人工智能已成为兼具速度与准确性的科学发现工具。数百年来，数据驱动的开普勒范式和第一性原理驱动的牛顿范式是科学研究的两大支柱。如今机器学习，尤其是深度学习异军突起，让科学家看到了新范式诞生的曙光。那么，我们可否再往前走一步，大胆畅想，如果人工智能将彻底改变社会生活，科学与人文的疆界是否也将被突破？

科技向善是终极答案

“我所渴望探求的乃是天地之奥秘。我要弄清楚究竟是什么力量在支配着我？是构成事物的外部物质，还是大自然的本身意志和人的神秘的灵魂？”弗兰肯斯坦的自白是工业革命时代技术狂热的缩影，他的悲剧敲响科技伦理的警钟。

两个世纪前的隐忧始终未曾消散，百年间，每一种新技术的诞生，都或多或少在颠覆旧思想、旧方法的同时，带来争议、困惑和焦虑。而从第一个科学怪人的诞生一直到今天科幻电影大量渲染的“人造人”危机，我们对所谓人类“造物主情结”的警惕，似乎随着人工智能时代的到来愈演愈烈。

机器能思考吗？人工智能会发展出意识与情感吗？如果那一天真的到来，人类与人工智能的边界在哪里？现场，同样有观众抱有这样的疑问。

事实上，如果仅从技术层面看，科幻电影中描绘的那些拥有独立自我意识、自主思维能力的人工智能，属于“强人工智能”或“通用人工智能”的范畴，现在的知识技术水平对此还望尘莫及。

“从行为的视角，我们或许可以通过分析自我意识的‘表象’要素，通过计算机程序把它们表达出来。”姚期智回答，“但从本质上来说，什么是‘意识’，什么是‘情感’，

这又是另一个哲学问题。”正如现在，越来越多的机器能够通过图灵测试，让人在表面上无法分辨其和人类的不同之处，但背后是否有“自我意识”的作用，这很难定论。

既然如此，是否意味着科幻作品只是杞人忧天，人工智能的伦理问题离我们还是很遥远？

“我们人文学者，始终要给出一个偏向悲观的结论，这是为了提醒世人记住经验教训。但是科学家一定要传达乐观的精神，因为他们的工作面向未来，这两者其实并不矛盾。”葛兆光说。如玛丽·雪莱以超现实的故事试图打破对技术的绝对崇拜，虚构与想象的价值不在于预知未来，也并非为了激起恐惧，而是为了让人们保持对科技“双刃剑”效应的思考。

而随着人工智能研究向纵深迈进，两位学者都相信，“强人工智能”的实现不是“是与否”的问题，而是“还要多长时间”的问题。

在演讲结尾，姚期智特别引用了加州大学伯克利分校计算机科学家斯图尔特·罗素的名言：“想先设计出极度聪明的机器，然后再去研究如何控制它，别做梦了！”说到底，未雨绸缪永远不会为时过早。至于当下，人工智能也还面临着更现实的技术瓶颈与伦理考验。比如，有关人

工智能应用中隐私保护和算法公平的呼吁早已出现。又比如，社会亟待“可信AI”解除人工智能的信任危机。

姚期智指出：“现在的人工智能机器学习技术有鲁棒性(稳定性)差、可解释性不强之痛病。”试想，如果人工智能不能达到绝对稳定，那么在自动驾驶的应用中，即使千分之一的差错率都会造成巨大的灾难；如果由于深度学习的“黑箱”特征，人工智能医生无法向患者提供医疗诊断的依据，它的可信度又还剩几分？

引导科技向善，让人工智能更好地服务美好生活与社会进步，是科学家的使命，也取决于整个社会的立场与态度。

“成熟的文明像一艘古老的大船，很稳定、很安全，但是要转身时，可能就不如小船，小船在风雨飘摇中颠簸，但往往能灵活调转。”葛兆光向我们解释，古代中国根深蒂固、自成一体的文化传统与政治秩序，如何在它与科技进步间划出了一条本不存在的“疆界”，拉开了它与西方世界的距离。

如今，中国这艘现代化的大船正朝着科技强国的灯塔乘风破浪，在人工智能这条航道上，技术创新力与伦理价值的共同护航，方能助我们行稳致远。

延伸阅读

他们眼中的人工智能

人工智能正在成为引领本轮科技革命与产业变革的战略技术、驱动性力量。这是一个人工智能大显身手的时代。

——中国工程院院士 潘云鹤

对于预测人工智能将带来怎样一个世界，人类的想象力也许是贫瘠的。人工智能所造就的，可能是一个难以想象的世界和文明。

——《三体》作者 刘慈欣

人工智能能够帮助我们处理大量的数据，快速地计算。但这种便捷的方法是用人类的思维发展出来的。

——数学家 丘成桐

以前的所有工具都使人类更强大，因为人懂得工具，而工具不懂得人。一个农夫知道一把斧头能做什么，但斧头不明白农夫的需要和感觉。很快，人工智能就能比我们更加了解我们自己。它是会继续做我们手中的工具，还是我们会成为它的工具呢？

——《人类简史三部曲》作者 尤瓦尔·赫拉利



距离“机器人”(ROBOT)一词(出自1920年捷克斯洛伐克文学大师卡雷尔·恰佩克的科幻剧本《罗素通用机器人公司》)的诞生，已经百年有余。人工智能也成为科幻小说最热门的题材之一。

《人造神祇》收录了十二位中国实力派科幻作家的中短篇原创科幻小说。其中，《起风之城》《机器之道》《以前的黄昏》等作品曾获科幻银河奖。作家们思考的都是人类与人工智能之间的各种关系，以及由此演绎出的各种精彩故事。人工智能这个目前尚“机心莫测”的物种，在书中以各种迥然不同的面目跃然纸上，充分展示了人类忧心忡忡的心理下所迸发的丰富想象力。(本报记者 严粒粒 综合整理)

葛兆光谈科技与中国古代文明 西湖大学提供

