当你谈论人工智能时，到底在谈论什么？

原创 2015-08-07 王培 赛先生

研究者们对人工智能的研究目标有非常不同的理解，这在业内不是新闻，但外界往往搞不清其中的区别，所以会把计算机的一个新应用当成对思维规律的研究突破，并因此期待其它领域中的相关问题也会很快得到解决。这一问题部分解释了为什么公众对人工智能的期望，历来总是在乐观和悲观之间徘徊。

- 王培（美国天普大学计算机与信息科学系）

近来人工智能成了热门话题。一方面，它和自动驾驶汽车、机器人保姆等各种新发明紧密相关，以至于在各种产品前面加一个“智能”的修饰语成为时髦。另一方面，它又被联系于大规模失业、机器统治人类等各种灾难，一时间，各种呼吁和警告不绝于耳。

专家们对这个领域的看法也很不一致。有人认为人工智能已经在我们身边了，而未来的发展无非是出现更多更好的类似技术；另一些人则认为人工智能还没有真正到来，而其出现将导致革命性的后果。

简单在网上搜索一下就可以发现，“人工智能”这个词语好像不总是在同一个意思下被使用，很多争论因此基本变成了“鸡同鸭讲”。因为对人工智能的不同理解会导致对有关问题的不同回答，对这个领域的剖析需要从“正名”开始。

乍看起来这个问题并不复杂。人工智能无非就是“让计算机像人脑那样工作”。现在，计算机已经能够完成很多以前只有人脑才能完成的智力活动，这自然就会引出“计算机能否完成所有智力活动”的问题。如果答案是肯定的，后面就可以考虑“怎样做”和“应不应做”等问题。如果答案是否定的，后面就可以考虑“为什么”和“哪些能做，哪些不能做”等问题。这已经就是人工智能的基本问题了。

尽管上述说法没有错，但其对研究目标的刻画太过粗略，不足以指导研究工作的进行。同时，这种描述遗漏了一个关键点，因而使不同的理解成为可能。

人工智能概念中的“人工”指的是电子计算机及其相关系统，如机器人等机电设备。这就意味着，“人工智能”和“人类智能”不可能在任何方面都完全相同。比如说， 一个人工智能系统不

会有和人完全一样的躯体，也不会有和人完全一样的经验（尤其是社会经验），否则它就不是一个人工智能，而是个“人造人”了，那是一个完全不同的领域。这个前提是如此地显而易见，以至于通常会被略过不提，而下面我们恰要以此为分析的起点。

既然人工智能是在某些方面（而非在所有方面）和人类智能相同或相近，那么刻画这个概念的关键在于，指明在哪方面“人”和“机”应当是一样或差不多的，而在其它方面是无所谓的。正是对这个问题的不同回答，造成了人工智能领域中最根本的分歧。

本文把典型的回答划分为“五大派”：结构派、行为派、能力派、方法派、原则派，并依次加以介绍和分析，尤其是其被支持和反对的理由。

结构派

既然人类智能是人脑的产物，那么似乎“尽可能忠实地模拟人脑结构”就是实现人工智能的最可靠途径了。对神经网络模型的研究已经取得了显著的进展，尽管人脑的巨大复杂性导致其许多因素在模型中被忽略或简化了。各种脑模型试图不仅在细节上（如神经元及其相互联系）而且在大尺度上（如层次和区域）都忠实于人脑结构，这方面的例子包括《人工智能的未来》一书作者Jeff Hawkins提出模拟大脑新皮层的HTM模型等。

尽管结构派的理由看上去很有说服力，但在人工智能界只是个少数派。除了其复杂性令人望而却步之外，这一思路最常被人挑战的问题是其必要性。人脑作为进化的产物，其结构自然会反映某些历史的偶然性。作为一个生物器官，人脑结构的生物性特征对于机电设备来说也未必都有被模拟的必要。即使是在神经网络模型中，人脑也一般只是被用作灵感的源泉，而非设计的依据。

行为派

考虑到计算机和人脑在质料和结构等方面的本质差别，有理由认为人工智能只须在外部表现和行为上像人就足够了。这一派观点最流行的版本就是“图灵测试”。为了避免在“思维”等复杂概念上纠缠不清，Alan Turing建议用一个直截了当的办法来判定一台机器是否“能思维”：如果它的言语行为和人无法区分，那它就算是“能思维”了，因为思维的结果最终总是要体现在行为上的。

虽然图灵测试广为人知，但是现有的人工智能研发很少将通过图灵测试作为其目标。与脑模型的情况相似，“在外部行为上和人一样”大概有些过于“人类中心主义”了。现在的计算机系统已经在很多方面远超人类了，难道它们在展现智能时还硬要在这些地方成心装得笨些吗？

同理，如果有朝一日外星人真的来造访地球了，我们总不能用它们在行为上可否与人相区分来判断它们是否有智能吧?

能力派

人们通常是以解决难题的能力来衡量他人智力的。依此类推，一种普遍的看法是，把人工智能等同于能解决“以前只有人脑才能解决的问题”的计算机。一个典型的例子就是IBM公司开发的“深蓝”系统。下棋无疑是需要智力的。既然“深蓝”战胜了国际象棋世界冠军，我们大概必须认为它有很高的智能。按照这种“不论黑猫白猫，能抓住老鼠就是好猫”的观念，智能就是解决某些问题的能力，而下棋就是这样一个问题，至于计算机是不是“像人一样下棋”是无关紧要的，只要能赢就行。

这种标准现在仍常常在媒体上被采用，而一些计算机的新应用也因而被贴上“人工智能”的标签。但这种做法也同时导致了一个非常尴尬的后果：一旦某个问题能通过计算机来解决，它就不再是“以前只有人脑才能解决的问题”了。这样一来，“人工智能”岂不是只能被定义为那些尚未解决的问题，从而成了阿凡提吊在驴鼻子前面那根它永远也吃不到的胡萝卜了？如果我们放宽标准到“曾经只有人脑才能解决的问题”，那么四则运算也将成为人工智能问题了，如此，这个领域也未免太过宽泛。这可真是“宽严皆误”!

方法派

既然只靠问题很难为人工智能划界，把解决问题的方法也加进去怎么样?这可以说是目前主流学术界的态度。按照这种观点，“智能”是一组“认知功能”的总称，其中包括表征、搜索、推理、学习、规划、决策、感知、运动、通讯等等。每一种认知功能可以作为解决某一类问题的方法来独立研究，进而成为一个可以独立应用的人工智能技术，一般体现为某种算法。

尽管这一派目前在人工智能界占主流地位，但它也有自己的麻烦。最突出的问题就是人工智能的“碎片化”。由于各个研究领域都想先把自己的问题解决好，再去考虑与别人协作的问题，其结果就是各人自扫门前雪，而各种人工智能技术也因为它们在基本假设上的不相容而难以有效协作或整合。这就是目前人工智能并无一个统一理论或技术的根本原因。

原则派

有些研究者相信，人工智能和其它领域一样可以被建立在一个相对简单的统一基础之上，而“智能”代表着某种理性原则。在人工智能的早期，Allen Newell和Herbert Simon就曾试图把“智能”归结为“问题求解”，进而归结为“在状态空间中的路径搜索”。以目前的智能模型为

例，在Marcus Hutter的AIXI模型中，智能被定义成“在任意环境中找到最优解的能力”，但这种能力不是可计算的，各种智能系统只能设法逼近它。在笔者设计的NARS（纳思）模型中，智能被定义成“在知识和资源相对不足的条件下的适应能力”，据此设计的一个推理系统目前已在调试阶段。

这一派目前在人工智能界是极少数，并常被人称为患了“物理学嫉妒症”。大部分人不相信复杂如“智能”（以及与其密切相关的“认知”、“思维”、“意识”等）现象可以被归结为一组基本原则，就如在物理学中那样。

研究者们对人工智能的研究目标有非常不同的理解，这在业内不是新闻，但外界往往搞不清其中的区别，所以会把计算机的一个新应用当成对思维规律的研究突破，并因此期待其它领域中的相关问题也会很快得到解决。这一问题部分解释了为什么公众对人工智能的期望，历来总是在乐观和悲观之间徘徊。

谈及对人工智能的不同理解，人们通常的观点是把它们看作导向同一终点的不同途径。而本文所要指出的是：严格说来，上面列举的五派其实是五个不同的研究领域，各有其终极目标、基本假设、评价标准等等。由于历史原因，它们现在都被叫做“人工智能”，但实际上不是一回事。它们之间当然是有联系的，而且在某些时候的确是“同路人”，但最终仍将分道扬镳，而非殊途同归。

这个判断在很多人看来似乎是太极端了，毕竟大家还是“同根生”，都是以人类智能为师的。有人把各派看作著名的“摸象盲人”，用领导的口气总结说：“完整的人工智能将是各种观点的辩证统一”。不幸的是，这副万能良药在这里无济于事，因为在具体设计的问题上，各派的要求常常是相悖的：最“拟人”的方案未必是对计算机最合适的，而对某个具体问题最好的解，也未必是最有一般意义的解。当一盘菜无法同时取悦于几个口味完全不同的食客时，上一个“拼盘”往往不是好办法。

这种现象的根本原因是，上面的五派分别对应于对人类智能的不同抽象，而上面所列它们各自的次序，则对应于抽象程度：结构派最低，原则派最高。

当同一个对象可以在不同的抽象层次上被有效描述时，不同的描述层次会呈现不同的图像。低层描述保留了对象的更多内部细节，但模糊了其整体特征和外部联系，而高层描述则相反。不同的抽象层次就像是不同倍数的望远镜或显微镜镜头：不存在哪个镜头“更正确”的问题，但相对于一个特定问题而言，的确存在哪个镜头“更合适”的问题。当我们在谈论人类智能时，这五类描述的确存在内在联系，但当我们试图在计算机上再现这些描述所刻画的对象或过程时，那些联系就不再是必要的了，而由此得到的各种“人工智能”甚至可能风马牛不相及。

总而言之，在人工智能的旗帜下，不同的人实际上在干不同的事：有构建脑模型的，有模拟人类行为的，有开拓计算机应用领域的，有设计新算法的，有总结思维规律的。虽然这些研究都有价值且有联系，但它们不可以彼此替代，而把它们混为一谈则容易导致思想混乱。

参考资料：

[1]Stuart Russell and Peter Norvig， Artificial Intelligence: A Modern Approach， 3rd edition， 2010.

[2]Pei Wang， What do you mean by“AI”? In the Proceedings of the First Conference on Artificial General Intelligence， pp. 362–373， 2008.

欢迎个人转发分享，刊物和机构如需转载，请联系授权事宜：iscientists@126.com。

更多精彩文章：您可以回复“目录”，接收往期文章目录和每一篇的获取方式，也可以返回主页点击屏幕下方子菜单获取最新文章、往期文章或直达赛先生微博。谢谢！

关于我们

《赛先生》由百人传媒投资和主办，全球范围内邀请顶尖科学家轮值担任主编。关注请加微信号：iscientists 或扫描下方二维码。

微信扫一扫
关注该公众号