

张钹院士：人工智能技术已进入第三代

AI 报道9月17日

来源：经济观察报



近日，中科院院士、清华大学人工智能研究院院长张钹教授接受记者采访时认为，目前基于深度学习的人工智能在技术上已经触及天花板。从长远来看，必须得走人类智能这条路，最终要发展人机协同，人类和机器和谐共处的世界。未来需要建立可解释、鲁棒性的人工智能理论和方法，发展安全、可靠和可信的人工智能技术。



张钹院士：AI 奇迹短期难再现 深度学习技术潜力已近天花板

在 **AlphaGo** 与韩国围棋选手李世石对战获胜三年过后，一些迹象逐渐显现，张钹院士认为到了一个合适的时点，并接受了此次的专访。

深度学习目前人工智能最受关注的领域，但并不是人工智能研究的全部。张钹认为尽管产业层面还有空间，但目前基于深度学习的人工智能在技术上已经触及天花板，此前由这一技术路线带来的“奇迹”在 **AlphaGo** 获胜后未再出现，而且估计未来也很难继续大量出现。技术改良很难彻底解决目前阶段人工智能的根本性缺陷，而这些缺陷决定了其应用的空间被局限在特定的领域——大部分都集中在图像识别、语音识别两方面。

同时，在张钹看来，目前全世界的企业界和部分学界对于深度学习技术的判断过于乐观，人工智能迫切需要推动到新的阶段，而这注定将会是一个漫长的过程，有赖于与数学、脑科学等结合实现底层理论的突破。

作为中国少有的经历了两个人工智能技术阶段的研究者，张钹在过去数年鲜少接受采访，其中一个原因在于他对目前人工智能技术发展现状的估计持有部分不同看法，在时机未到之时，张钹谨慎的认为这些看法并不方便通过大众媒体进行传播，即使传播也很难获得认同。

一、“奇迹并没有发生，按照我的估计，也不会继续大量发生”

经济观察报：您是如何估计和评价目前人工智能发展的现状？

张钹：这一轮人工智能热潮是本世纪初兴起的。首先是出现在学术界。学术界过去对人工智能是冷遇的，但是多层神经网络的出现带来了一些改变，神经网络的理论在上世纪 50 年代就有了，但是一直处于浅层的应用状态，人们没有想到多层会带来什么新的变化。

真正引起大家注意的就是 2012 年斯坦福的实验（注：2012 年谷歌和斯坦福利用多层神经网络和大量数据进行图像识别的实验），过去实验的图像样本数最多是“万”这个级别，斯坦福用了 1000 万，用多层神经网络来做，结果发现在人脸、人体、猫脸三个图像类别中，这个模型的识别率大概有 7%-10% 的提高。

这给大家非常大的震动，因为通常识别率要提高 1% 要做好多努力，现在只是把层数增加了，竟然发生两大变化，一个是识别率提高这么多；第二个是能处理这么大数据。这两个变化给大家非常大的鼓舞，何况在 2012 年之前，人工智能没有解决过实际问题。

经济观察报：这种突破的原因是什么？

张钹：现在分析下来是三个原因，大家也都非常清楚了，一个大数据、一个是计算能力、一个是算法。认识到之后，一夜之间业内业外对深度学习都非常震动，然后就发生了三件历史性的事件。

第一件事是 2015 年 12 月，微软通过 152 层的深度网络，将图像识别错误率降至 3.57%，低于人类的误识率 5.1%；第二件事，2016 年微软做的语音识别，其词错率 5.9%，和专业速记员水平一样；第三件事：AlphaGo 打败韩国围棋选手李世石。

通过人工智能，利用深度学习、大数据这两个工具，在一定条件下、一定领域内竟然能够超过人类，这三件事情给大家极大的鼓舞。

特别是对于业外的人，都认为我只要掌握了大数据，利用深度学习说不定还能搞出奇迹来，于是大家做了很多很多预测，比如在多短时间内计算机会在什么事情上能超过人。

但实际上，在这个之后，奇迹并没有发生，按照我的估计，今后也不会大量发生。准确一点说，今后或许会在个别领域取得进展，但是不会像之前预计的那样全面开花。特别是中国市场乐观的认为“中国市场大、数据多，运用又不受限制，所以将来奇迹一定会发生在中国”。

结果很多企业在做的时候发现，不是那么回事。**从目前的情况来看效果最好的事情还是这两件：图像识别、语音识别。**我看了一下，中国人工智能领域 20 个独角兽 30 个准独角兽企业，近 80% 都跟图像识别或者语音识别有关系。

经济观察报：为什么会出现这样的情况？或者说在这么长时间后，我们对人工智能目前能做什么有一个清晰的认识了吗？

张钹：人工智能在围棋上战胜人类后产生了这种恐慌，“大师才能做的事，人工智能居然能做，我的工作这么平凡，肯定会被机器所替代”。这里需要考虑一下它的局限性，我一直在各种各样的会上谈到不要过于乐观。

人工智能能做的那三件事（语音识别、图像识别、围棋）是因为它满足了五个条件，就是说只要满足了这五个条件，计算机就能做好，只要有任何一个或者多个条件不满足，计算机做起来就困难了。

第一个是必须具备充足的数据，充足不仅仅是说数量大，还要多样性，不能残缺等。

第二个是确定性。

第三个是最重要的，需要完全的信息，围棋就是完全信息博弈，牌类是不完全信息博弈，围棋虽然复杂，但本质上只需要计算速度快，不要靠什么智能，可是在日常生活中，我们所有的决策都是在不完全信息下做的。

第四个是静态，包括按确定性的规律演化，就是可预测性问题，在复杂路况下的自动驾驶就不满足这一条；实际上它既不满足确定性，也不满足完全信息。

第五个就是特定领域，如果领域太宽他做不了。单任务，即下棋的人工智能软件就是下棋，做不了别的。

经济观察报：就是说在满足这五个条件的前提下，目前的人工智能是胜任部分工作的？

张钹：如果你的工作符合这五个条件，绝对会被计算机替代，符合这五个条件的工作特点很明显，就是四个字“照章办事”，不需要灵活性，比如出纳员、收银员。如果你的工作富有灵活性和创造性，计算机绝对不可能完全代替，当然部分代替是可能的，因为其中肯定也有一些简单和重复性的内容。如果认识到这一条就会认识到人工智能仍处于发展阶段的初期。不是像有些人估计的那样“人工智能技术已经完全成熟，而进入发展应用的阶段”。

二、“深度学习技术，从应用角度已经接近天花板了”

经济观察报：我们应该怎么去定义目前的深度学习技术路线，它是基于概率学的一个事物吗？

张钹：现在的深度学习本质是基于概率统计，什么叫做概率统计？没有那么玄，深度学习是寻找那些重复出现的模式，因此重复多了就被认为是规律（真理），因此谎言重复一千遍就被认为真理，所以为什么大数据有时会做出非常荒唐的结果，因为不管对不对，只要重复多了它就会按照这个规律走，就是谁说多了就是谁。

我常常讲我们现在还没有进入人工智能的核心问题，其实人工智能的核心是知识表示、不确定性推理这些，因为人类智慧的源泉在哪？在知识、经验、推理能力，这是人类理性的根本。现在形成的人工智能系统都非常脆弱容易受攻击或者欺骗，需要大量的数据，而且不可解释，存在非常严重的缺陷，这个缺陷是本质的，由其方法本身引起的。

经济观察报：就是说通过改良的方式无法彻底解决？比如我们再增加神经网络层数和复杂性或者再提升数据的量级，会解决它的缺陷吗？

张钹：改良是不行的，深度学习的本质就是利用没有加工处理过的数据用概率学习的“黑箱”处理方法来寻找它的规律，这个方法本身通常无法找到“有意义”的规律，它只能找到重复出现的模式，也就是说，你光靠数据，是无法达到真正的智能。

此外，深度学习只是目前人工智能技术的一部分，人工智能还有更大更宽的领域需要去研究，知识表示、不确定性处理、人机交互，等等一大片地方，不能说深度学习就是人工智能，深度学习只是人工智能的一部分。一直到去年人工智能大会交流的论文还是三分之一是机器学习方面，三分之二是其他方面。

经济观察报：学界在这上面还是有一个比较清晰的认识？

张钹：我可以这么说，全世界的学界大多数有清晰的认识；全世界的企业界大多持过于乐观的估计。

为什么出现这样的情况呢？因为从事过早期人工智能研究的人，大多已经故去或者年老，已经没有话语权。现在活跃在人工智能研究第一线的都是深度学习、大数据兴起以后加入的，他们对人工智能的了解不够全面。

经济观察报：如果说每一个技术路线都有一个“技术潜力”，那么在深度学习方面，我们已经把这个潜力用了多少？

张钹：科学研究是很难精确估计的，但是深度学习如果从应用角度，不去改变它，我觉得已经接近天花板了，就是说你要想再出现奇迹的可能性比较小了。

经济观察报：那基于此，目前商业公司在底层技术和产业应用上还是有很大的空间吗？

张钹：只要选好合适的应用场景，利用成熟的人工智能技术去做应用，还有较大的空间。目前在学术界围绕克服深度学习存在的问题，正展开深入的研究工作，希望企业界，特别是中小企业要密切注视研究工作的进展，及时地将新技术应用到自己的产品中。当然像谷歌、BAT 这样规模的企业，他们都会去从事相关的研究工作，他们会把研究、开发与应用结合起来。

经济观察报：有一种观点认为我们强调的“白盒”（可理解性）它实际上是从人的思维来强调的，但是通过大数据、概率统计工具离散到连续的投射，它实际上是机器的思维，你不一定需要它给你一个解释，只要正确的答案就可以了？

张钹：目前有两种意见，一种观点认为智能化的道路是多条的，不是只有一条路能通向智能，我们通过自然进化产生了自然智能，那么我们为什么不能通过机器产生机器智能？这个智能和自然智能不会是完全一样的，条条大路通罗马，我们通过自然进化获得的智能也不见得是最佳的。这个观点我赞成，机器智能与人类不相同，其实是有好处的，恰恰可以互补，发挥各自的长处。

但是从长远来看，必须得走人类智能这条路，为什么？因为我们最终是要发展人机协同，人类和机器和谐共处的世界。我们不是说将来什么事情都让机器去管去做，人类在一边享受。我们要走人机共生这条路，这样机器的智能就必须和人类一样，不然没法共处，机器做出来的事情，我们不能理解，我们的意图机器也不知道，二者怎么能合作？

经济观察报：就是必须具有可解释性？

张钹：是，就是可解释性，你要它做决策，你不理解它，飞机就让它开，谁敢坐这架飞机？所以目前的阶段，车和飞机还是不能完全让机器开的。为什么司机坐在上面我们放心？因为我们和他同命运，要撞死一块撞死，机器和你可不是同一命运，它撞不死，你撞死了。

有的人非常脱离实际的去想这个问题，这是不对头的，人类怎么会去那样发展机器呢（注：指把人类的命运全部交给机器）？人类不会去那么发展的，有些人在那边担忧什么机器人统治人类，我说这最多只能算远虑。

经济观察报：所以图灵的论文中也说这种观点“不值一驳”。

张钹：是，那是远虑，我们目前还有很多近忧，发展人工智能必须要考虑安全问题，这已是现实问题。

你看语音合成，利用现有的技术可以做到以假乱真，和真人基本没有差别。现在看来这种技术不能推广应用，因为一旦推广就全乱套了，只要搞一段用语音合成技术做成的假录音，就可以让任何一位名人身败名裂。这些都是非常危险的技术。人工智能的治理已经提到日程上了。

三、“我们培养不出爱因斯坦、培养不出图灵”

经济观察报：一种观点认为中国有更多的数据和更多的工程师，这种规模能倒推带来基础研究层面的突破或者决定技术的路线？

张钹：这里混淆了好多概念，科学、技术、工程。**科技水平需要三个标准来衡量，一个是科研水平、一个是技术水平、一个是工程实践能力，或者产业化能力。**

我们中国什么情况？从工程角度来看，在一些领域我们“接近世界水平”；技术水平我用的词是“较大差距”，因为不少东西还是外国会做我们不会做；科学研究领域我用的词是“很大差距”，科学研究就是原创，实际上，所有人工智能领域的原创成果都是美国人做出来的，人工智能领域图灵奖得主共十一人，十个美国人，一个加拿大人。

经济观察报：数据显示中国在人工智能领域的论文发表量和被引用次数都已经进入前列位置，这是否说明中国人工智能科学研究领域的突破？

张钹：如果单从论文来看研究水平，基本反映在三个指标上：数量、平均引用率、单篇最高引用率。拿人工智能来讲，中国研究者论文的数量和平均引用率都还不错，但是单篇最高引用率和世界差距就很大，而这个指标恰恰是反映你的原创能力。

也就是说深度学习这个领域，我们的平均水平达到世界水平了，但是最高水平和世界差距还是很大的。不过还是要肯定的，我们应用上发展比较快。

经济观察报：清华在这方面有什么优势吗？

张钹：在人工智能重要的会议杂志上，这十年期间论文数量、平均质量 CMU（美国卡耐基梅隆大学）排第一，清华大学排第二。我们培养的人，在计算机这个领域，清华的本科、博士生都是世界一流的。

目前我们的跟踪能力是比较强的，一旦有人起个头，我们能迅速跟上去。但是很可惜，我们缺乏顶尖人物，也培养不出顶尖的人才，如爱因斯坦、图灵等。

我个人认为原因之一，可能与中国的文化有点关系，我们的从众心理很严重，比如在人工智能领域，深度学习很热，发表的论文作者中几乎 70% 是华人，但是其他非热门领域，包括不确定性推理、知识表示等几乎没有华人作者。这就是**从众扎堆，不愿意去探索“无人区”**。

当然也不要着急，科学研究本来就是富人干的事情，是富国干的事情，我们还是发展中国家，科学研究起点比较低，暂时落后是难免的，我们会迎头赶上。

四、“低潮会发生，但不会像过去那样”

经济观察报：如果说深度学习已进天花板，那么人工智能未来的前进方向将会在哪？

张钹：最近我们准备提出一个新的概念，就是第三代人工智能的概念，人工智能实际上经历过两代，第一代就是符号推理，第二代就是目前的概率学习（或深度学习），我们认为现在正在进入人工智能的第三代。原因很明显，第一代、第二代都有很大的局限性。

经济观察报：你所说的第三代人工智能技术是有明确的实现方向或者特点吗？

张钹：我们现在提出的是要**建立可解释、鲁棒性（注：可以理解为稳健性）的人工智能理论和方法，发展安全、可靠和可信的人工智能技术**。

经济观察报：这样的技术可能要等很久？

张钹：是啊，很难预计，我们也很着急。

经济观察报：是不是还得回归到数学等理论层面里再去找新的方法？

张钹：这个目前我们**有两条路，一个是和数学结合，一个是和脑科学结合**。你想想如果没有新的数学工具，没有来自于脑科学启发下的新思路，哪来的新理论？另一方面是要把**数据驱动和知识驱动结合起来**，因为通过数学、脑科学上寻求突破是比较艰难的，前面这件事现在则完全能够做。

经济观察报：这个结合是指之前几十年人工智能的经验统合到一块？

张钹：是的，至少有一个方向就是要把第一代和第二代结合，利用各自的优势。但是这两个结合很困难，因为他们在不同空间中操作，一个是向量空间，一个是符号空间，也需要有新的数学工具的加入。

经济观察报：看人工智能历史，每一代技术之间有很长的间隔期，第三代人工智能技术也会这样吗？

张钹：我认为会更长，因为需要攻坚，因为遇到的问题更困难。

经济观察报：会不会再过 10 年、20 年，人工智能在学界或者公众心中，又变成一个“隐学”，就像 70、80 年代那样，大众又不会再经常提起来这个词？

张钹：低潮会发生，但不会像过去那样，原因在哪？因为有大数据、互联网和强大的计算资源，这些都会支撑人工智能继续走下去，尽管有的时候还只是表面上的繁荣。

附：在 2018 全球人工智能与机器人峰会上，清华大学人工智能研究院院长张钹院士做题为“走向真正的人工智能”（Towards A Real Artificial Intelligence）的大会报告。以下为报告全文，供大家学习交流。

张钹院士：走向真正的人工智能

我今天要讲的中心思想就是：我们现在离真正的人工智能还有一段很长的路。为了讲清这个思想，我必须回答下面三个问题：

第一，什么叫做真正的人工智能？我们的目标是什么？第二，为什么我们需要真正的人工智能？

第三，我们如何走向真正的人工智能？

我现在回答这三个问题。

首先我们如何评价目前人工智能取得的成果，我们的评价很简单，针对这 5 件事：

第一是深蓝打败人类国际象棋冠军；第二是 IBM 在电视知识竞赛中打败了美国的前两个冠军，这两件事是一种类型，后面的三件事是另外一种类型；即 2015 年微软在 ImageNet 上做图象识别，它的误识率略低于人类。还有百度、讯飞也都宣布在单句的中文语音识别上，它的误识率也略低于人类。还有一个是大家非常熟悉的 AlphaGo 打败了李世石。这 5 件事情都是机器在一定的范围内超过了人类，我们如何来评价这 5 件事？

大家一致认为这 5 件事之所以成功，是由于前面三个因素，一是**大数据**，二是**计算能力提高**，第三是有**非常好的人工智能算法**。这三个因素大家都讨论得非常多了，没必要我再来说，我现在要说的最后一个因素是被大家所忽略的，这个因素是说，**这所有的成果必须建立在一个合适的应用场景下**。这 5 件事虽然领域很不一样，但是它们都满足完全一样的条件，或满足下面的 5 个限制，首先你必须有丰富的数据或者丰富的知识，如果这两件东西没有，或者很少，你不用来谈人工智能，因为你无法实现无米之炊。人工智能唯一的两个资源，一个是数据，一个是知识。还有确定性信息、完全信息、静态的、单任务和有限领域。这 5 个条件里面任何一个条件不满足，现在的人工智能做起来就非常困难了。

大家想想这 5 个限制条件下的应用场景是什么样的应用场景？就是照章办事，不需要任何灵活性，这显然不是智能的核心。

我们现在分析一下上述 5 个场景。下象棋是完全信息博弈，信息完全和确定，没有问题。其次，它遵循着完全确定的游戏规则演化，我们把这种情况也叫做静态。Watson 机器人也是这样，Watson 是什么样的对话问题呢？它为什么选择知识竞赛呢？我们知道知识竞赛提的问题都没有二义性，都是明确的，它的答案总是唯一性的。所以这样的问答对机器人来讲是非常容易的。它涉及的领域虽然比较宽，但也是有限的，包括大家觉得很玄乎的围棋，也完全符合上面 5 个条件，所以对计算机来说也是很容易的。目前计算机打麻将就不行，因为牌类是不完全信息博弈，所以比棋类要难。总之，我们对目前人工智能取得的成果要有一个正确的评价。

目前的人工智能技术在以下领域都可以找到它的应用，它们是交通、服务、教育、娱乐等等，但我要强调是这些领域里面只有满足上述 5 个条件的事情，计算机做起来才会容易，如果不满足这些条件，计算机就做起来就困难了。大家常常关心什么样的工作会被机器所替代，我可以明确告诉大家，**满足这 5 个条件的工作，总有一天会被计算机取代，就是那些照章办事，不需要任何灵活性的工作，比如说出纳员、收银员等等**。在座的所有工作都不可能被计算机完全代替，但不排斥你的工作中有一部分会被计算机取代，老师、企业家等的工作不可能被计算机完全代替。

为什么有这 5 个限制？原因在于我们现在的人工智能是没有理解的人工智能。

我们先看**符号模型，理性行为的模型**，举 Watson 的例子，它是个对话系统，我们现在所有做的对话系统都跟这个差不多，但是 Watson 做得更好些，它里面有知识库，有推理机制。沃森除了专家知识之外，还有大量互联网上大众的知识，还运用了多推理机制。请看，这就是 Watson 系统的体系结构。它里面有哪些知识呢？有很多，包括百科全书、有线新闻、文学作品等等。所有的知识用纸质来表示有 2 亿页，用存储量表示达到了 4TB。它能回答什么问题呢？用它的例子来说明。第一个问题，1974 年 9 月 8 日谁被总统赦免？这对美国人来讲很好回答，同样对计算机来讲也很好回答，你用这几个关键字「1974 年 9 月 8 日」、「被总统赦免」，就能在文献里头查出来是谁，他就是尼克松。也就是说根据问题中的关键字，可以在已有的文献里头直接找到答案，这就是一般的网络检索方法。

第二个问题，荧光粉受到电子撞击以后，它的电磁能以什么方式释放出来？我们用「荧光粉」、「电子撞击」、「释放电磁能」等关键词，也可以找到答案：「光或者光子」。这种方法就是平时网络搜索的原理，应该说没有什么智能。

回答下面的问题就需要「智能」了，跟智利陆地边界最长的是哪个国家？跟智利有陆地边界的国家可以检索到，它们是阿根廷和玻利维亚，但是谁的边境长？通常查不到。Watson 具备一定的推理能力，它从边界间发生的事件、边界的地理位置等等，经过分析推理以后就可以找出答案，它就是阿根廷。下一个问题也属于这种性质，跟美国没有外交关系的国家中哪个最靠北，跟美国没有外交关系的国家有 4 个，只要检索就行了，但是哪个国家最靠北，没有直接答案，但可以从其它信息中推导出来，比如各个国家所处的纬度、气候寒冷的程度等等分析出来，答案是北朝鲜。

智能体现在推理能力上。但是很不幸，现在的对话系统推理能力都很差。Watson 系统好一些，但也很有限。换句话说，**我们现在的对话系统离真正的智能还很远。**

我们通过索菲亚机器人就可以看出来，索菲亚的对话是面向开放领域，你可以随便提问，问题就暴露出来了。大家在电视上看到索菲亚侃侃而谈，问什么问题都能答得很好，这里面有玄机，如果你的问题是预先提出来的，因为里头有答案，因此回答得非常好，在电视上给大家演示的都是这种情况。

如果我们临时提问题，问题就出来了。这是一个中国记者给索菲亚提的 4 个问题，它只答对了「你几岁了」，这个问题很简单，它答不上来，它的回答是「你好，你看起来不错」，答非所问，因为它不理解你所问的问题。只有第二个问题它是有准备的，里面有答案，所以答得很好。「你的老板是谁」，这个肯定它有准备。第三个问题，「你能回答多少问题呢」？它说「请继续」，没听懂！。再问第四个问题，「你希望我问你什么问题呢」？它说「你经常在北京做户外活动吗」？这就告诉我们说，现代的问答系统基本上没有理解，只有少数有少量的理解，像 Watson 这样算是比较好的。

为什么会这样？也就是说**我们现在的人工智能基本方法有缺陷，我们必须走向具有理解的 AI，这才是真正的人工智能。**我这里提出的概念跟强人工智能有什么区别？首先我们说它在这点上相同的，我们都试图去准确地描述人类的智能行为，希望人工智能跟人类的智能相近，这也是强人工智能的一个目标，但是强人工智能只是从概念上提出来，并没有从方法上提出怎么解决。大家知道强人工智能提出了一个最主要的概念，就是通用人工智能。怎么个通用法？它没有回答。我们现在提出来的有理解的人工智能是可操作的，不只是概念，这是我们跟强人工智能的区别。

人机对话的时候，机器为什么不能理解人们提的问题。我们看一个例子就知道了，我们在知识库把「特朗普是美国总统」这个事实，用「特朗普-总统-美国」这三元组存在计算机里面，如果你提的问题是「谁是美国总统」？机器马上回答出来：「特朗普」。但是你如果问其它有关的问题，如「特朗普是一个人吗」？「特朗普是一个美国人吗」？「美国有没有总统」？它都回答不了。它太傻了，任何一个小学生，你只要告诉他特朗普是美国总统，后面这几个问题他们绝对回答得出来。机器为什么回答不了后面的三个问题呢？就是这个系统太笨了，没有常识，也没有常识推理。既然特朗普是美国的总统，美国当然有总统，但是它连这一点常识的推理能力都没有。所以要解决这个问题，必须

在系统中加上常识库、常识推理，没有做到这一步，人机对话系统中机器不可能具有理解能力。但是大家知道，建立常识库是一项「AI 的曼哈顿工程」。大家想想常识库多么不好建，怎么告诉计算机，什么叫吃饭，怎么告诉计算机，什么叫睡觉，什么叫睡不着觉，什么叫做梦，这些对人工智能来说都非常难，美国在 1984 年就搞了这样一个常识库的工程，做到现在还没完全做出来。可见，要走向真正的人工智能，有理解的人工智能，是一条很漫长的路。

这里介绍一点我们现在做的工作，加入常识以后，对话的性能会不会有所改善。**我们的基本做法是建立一个常识图谱，用这个图谱帮助理解提出的「问题」，同时利用常识图谱帮助产生合适的答案。**

下面就涉及到具体怎么做了，我不详细说了，我就说结果，**结果是有了常识以后，性能有了显著的改善，对话的质量提高了。**这篇文章已经发表，有兴趣可以去阅读。

另外是**准符号模型，深度学习、神经网络主要用来模拟感性行为**，感性行为是一般很难采用符号模型，因为感性（感觉）没法精确描述。比如「马」，怎么告诉计算机什么叫做马？你说马有四条腿，什么叫做腿？你说细长的叫做腿，什么叫细？什么叫做长？没法告诉机器，因此不能用符号模型。目前用的办法就是我们现在说的神经网络或者准符号模型，也就是用人类同样的办法，学习、训练。我不告诉机器什么叫做马，只是给不同的马的图片给它看，进行训练。训练完以后，然后再用没见过的马的图片给它看，说对了，就是识别正确了，说不对就是识别不正确，如果 90% 是对的，就说明它的识别率是 90%。后来从浅层的神经网络又发展到多层的神经网络，从浅层发展到多层有两个本质性的变化，一个本质性的变化就是输入，深层网络一般不用人工选择的特征，用原始数据就行。所以深度学习的应用门槛降低了，你不要有专业知识，把原始数据输进去就行了。第二个是它的性能提高很多，所以现在深度学习用得很多，原因就在这个地方。

通过数据驱动建立的系统能不能算是有智能呢？必须打一个很大的问号，就是说你做出的人脸识别系统甚至识别率会比人还高，但是我们还不能说它有智能，为什么呢？这种**通过数据驱动做出来的系统，它的性能跟人类差别非常大，鲁棒性很差，很容易受干扰，会发生重大的错误，需要大量的训练样本。**我们刚才已经说过，给定一个图像库我们可以做到机器的识别率比人还要高，也就是说它可以识别各种各样的物体，但是这样的系统，我如果用这个噪声输给它，我可以让它识别成为知更鸟，我用另外的噪声输给它，可以让它识别成为猎豹。换句话讲，**这样的系统只是一个机械的分类器，根本不是感知系统。**也就是说它尽管把各种各样动物分得很清楚，但是它不认识这个动物，它尽管可以把猎豹跟知更鸟分开，但是它本质上不认识知更鸟和猎豹，它只到达了感觉的水平，并没有达到感知的水平，它只是「感」，没有上升到「知」。我们的结论是，只依靠深度学习很难到达真正的智能。这是很严峻的结论，因为如果有这样的问题，在决策系统里头是不能用这样的系统，因为它会犯大错。我在很多场合讲过，**人类的最大的优点是「小错不断、大错不犯」，机器最大的缺点是「小错不犯，一犯就犯大错」。**这在决策系统里头是不允许的，这就显示人跟机器的截然不同，人非常聪明，所以他做什么事都很灵活，这就使得他很容易犯各种各样的小错。但是他很理性，很难发生大错。计算机很笨，但是很认真，小错误绝对不会犯，但是它一犯就是天大的错误。刚才把那个把噪声看成知更鸟，这不是大错吗？你把敌人的大炮看成一匹马，不是大错吗？但是人

类不会发生这种错误，人类只会把骡看成驴，但是计算机的识别系统会把驴看成一块石头。原因在哪儿？原因还是 AI 的理解能力问题。

我们看这个自动驾驶，过去讲得很多，而且讲得很乐观，我们看看问题在什么地方。我们现在是这样做，我们通过数据驱动的学习方法，学习不同场景下的图象分割，并判别是车辆还是行人、道路等，然后建立三维模型，在三维模型上规划行驶路径。现在用硬件已经可以做到实时，请问大家，这样能不能解决问题？如果路况比较简单，行人、车辆很少，勉强可以用。复杂的路况就用不了。什么原因？非常简单，好多人总结出这个经验，行人或者司机都会有意无意破坏交通规则，包括外国人也一样，中国人更严重一点。这就使得数据驱动方法失效，比如说我们可以用数据驱动方法来了解各种各样行人的行为，我们可以通过大量进行训练，都训练完以后，如果出现新的情况呢？计算机能理解这是人从底下钻过来，很危险吗？所以你不可能把所有情况都训练到。自动驾驶不可能对付突发事件，如果这个突发事件它没见过，它就解决不了。怎么来解决这个问题呢？实际上**就是要解决从「Without」到「With」理解的问题**。人工智能现在有两种基本方法，一种是用符号模型来模拟理性行为，符号模型可以表达信息的内容，所以它是在一个语义的符号空间里头，但是非常不幸，这个离散的符号表示，数学工具很难用，很多数学工具用不上去，所以它发展很慢。在模拟感性行为的时候，我们用的是特征空间的向量，向量就是数，可以把所有的数学工具都用上，优化的工具、概率统计的工具全部用上。所以数据驱动方法这几年发展非常快，再难的问题，下围棋非常难吧，计算机也可以「算」出来。但是它有一个非常大的缺陷，它是在特征空间里，缺乏语义。我们用数据去训练一个模型，所谓「黑箱学习法」，加上你的数据质量不高，很难学出有用的东西。什么叫概率统计？重复多了就是真理。如果数据质量差，充满了「谎言」。谎言重复多了，就变成真理了。

我们现在想出的解决办法是这样的，就是把这两个空间投射到一个空间去，这个空间叫做语义的向量空间。也就是说我们把符号变成向量，同时把特征空间的向量变成语义空间的向量。怎么做？一是通过 Embedding（嵌入）把符号变成向量，尽量保持语义不变，可惜现在的方法都会引起语义的丢失，我们只能在投射的过程中让语义丢失得少。第二方面做的工作比较少，就是 Raising（提升），把特征空间提升到语义空间去，这主要靠学科交叉，靠跟神经科学的结合。只有这些问题解决以后，我们才能够建立一个统一的理论，因为过去的感知和认知是不同的处理方法，大家说不到一块，如果我们能够投射到同一空间去，我们就可以建立一个统一的理论框架，这是我们的目标。在语义空间处理就可以解决理解问题，但是这项工作是非常艰巨的。

介绍一项我们现在做的工作。人工神经网络为什么不能得到语义信息呢？人脑的神经网络为什么可以呢？差别就在这里，我们现在用的人工神经网络太简单了，**我们正想办法把脑神经网络的许多结构与功能加进去**，我们这里只用了「稀疏发电」这一性质，就可以看出一些效果，人脸、大象或者鸟的轮廓，神经网络可以把它提取出来。

还有一个办法就是把数据驱动跟知识驱动结合起来。刚才讲了，人的智能没法通过单纯的大数据学习把它学出来，那怎么办？很简单，加上知识，让它有推理的能力，做决策的能力，这样就能解决突发事件。我们现在做的工作就是把这些结合起来，这是我们的基本思路，知识也好，数据也好，都投射到同一空间，然后都用同样的数学方法进行处理，这方面我们已经做了不少工作。

最后做一个总结，我们从这个坐标看人工智能，横轴代表领域的宽窄，从单领域到多领域、到开放领域。纵轴代表信息的确定性与完全性，从完全到不完全、从确定到不确定。在左下角代表最容易的，就是刚才讲的符合 5 个条件的，现在人工智能在这部分解决得非常好，我们用白色来表示它，AlphaGo 在这里，深蓝在这里，工业机器人在这里。现在我们正在向灰色地区去走，打牌，信息不完全，现在打德州扑克，一人对一人，计算机能战胜人类，多人对弈，计算机还不行，这是灰色地带，我们还可以做，为什么可以做？尽管打牌是不确定的，但是它在概率意义下是确定的，你拿的这副牌的概率，可以算出来，同花的概率是多少，排成顺的概率是多少，既然概率能算出来，最终人类肯定会被计算机打败。Watson 在右边，它的领域比较宽，但是它是确定性的，所以是在灰色的区域。往右上方去就比较难了，自动驾驶、服务机器人、大数据分析，它是一个大框，有的简单，有的困难，就自动驾驶来讲，专用道、行车很少，路况简单等，在白色或者灰色区，如果路况复杂就到了黄色区域，黄色区现在计算机还解决不好。最远的在哪儿呢？右上角，图灵测试。大家对图灵测试有很多误解，其实图灵测试是开领域问答，很难！索菲亚做得怎么样？很糟糕。自然语言理解也在这里，复杂环境下的决策在偏左一点的地方，这也是很难的。所以我们人工智能现在是从左下角往右上角走，我们现在处在出发点附近。有的人想把它用一些名词来区分人工智能的不同发展阶段，有专家问我，你的看法怎么样？我建议不要用新词，用新词往往说不清，很麻烦，有的人说现在是弱人工智能，以后是强人工智能，也有人说现在叫增强智能（Augmented Intelligence）也是 AI……概念太多说不清，还是简单一点，「我们正在通往真正 AI 的路上」，现在走得并不远，在出发点附近，人工智能永远在路上，大家要有思想准备，这就是人工智能的魅力。大家为什么这么重视人工智能？因为我们永远在路上，这就吸引我们去解决这些问题，这些问题一旦解决了，人类的社会进步、人类的生活就会发生本质上的改变。

最后我用中文写最后一段作为总结，可惜我翻译不了。周穆王西巡狩，路遇匠人名偃师。翌日偃师谒见王，偕来一个假人。「趋步俯仰，信人也」。「领其颊，则歌合律；捧其手，则舞应节。千变万化，惟意所适。王以为实人也，与盛姬内御并观之，技将终，倡者瞬其目而招王之左右侍妾。王大怒，要杀这个偃师。偃师大惧，立剖其倡者以示王，皆傅会革、木、胶、漆、白、黑、丹、青之所为。穆王始悦，诏贰车载之以归。

这是 3000 年前我们古人对机器人的想象，看看现在的人工智能做得怎么样呢？索菲亚是我们现在达到的水平，可是她不会唱歌、不会跳舞，只会说英文，周王也听不懂，肯定没有印象。现在我们假设索菲亚「瞬其目而招王之左右侍妾」，向周王的姨太太们送去秋波，王会如何呢？我认为没反应，因为索菲亚是女的，他用不着吃醋。但是我们假设索菲亚「瞬其目而招王」，向大王送去秋波，王会大悦，立即神魂颠倒，坠入爱河？我认为不会，因为索菲亚根本不像人，它最近才刚刚安上手脚，走路都不利索，怎么行呢？所以我的结论是，「索菲亚通不过穆王的测试，当然它更通不过图灵测试」。

我们的结论是什么？**人工智能刚刚起步，离真正的 AI 还很遥远**，大家共同努力吧，我们任重道远。