

人文解读“四大未知疆域”(太空极地深海网络)系列讲座之网络场·圆桌对话

人文学者江晓原、于海激辩自然科学学者林龙年、危辉

人工智能：边原创冲刺边伦理“刹车”

圆桌对话

于海：刚才危老师的演讲让今天的讨论有了基础。请问你喜欢的“不识庐山真面目，只缘身在此山中”一诗，“庐山”是指人工智能还是人类智能？

何谓不是“一个筐”的人工智能？

危辉：人类智能，它的复杂度被我们严重低估了。所谓“书到用时方恨少”，要模拟它才知道有多难。

“阿尔法狗”并非第一个挑战人类尊严者

于海：今年5月份“阿尔法狗”以3:0战胜柯洁，让许多人开始忧虑人工智能对人类尊严的冲击。人类向来极具尊严，但“阿尔法狗”，这个人类创造的上帝，对人类打击颇大。人类承受自尊心打击远早于人工智能。首先是哥白尼，提出了“日心说”，将人类和地球从宇宙的中心地位移除；接着是马克思，宣扬唯物主义，认为衣食住行这类物质性的东西才是奠定历史的基础，而非人的意识和自由意志；再是弗洛伊德，认为白天人类光鲜亮丽的表现都是受到道德和规范的制约，夜晚的梦才能反映真实的人性欲望，欲望是人类发展的动力。尽管现在AI是热点话题，但其实目前智能能做的事情还是有限。所以，先要明确，当大家都在讨论人工智能时，哪些说法不准确，不属于此范畴？

对人脑技能模拟才是人工智能，机械计算并不是

危辉：有人曾揶揄“人工智能是个筐，什么都能往里装”，虽然刻薄也说明了现状。通常，当解决问题需要推理、决策、理解、学习这类最基本的技能时，我们才认为它跟人工智能很相关。常见的人工智能技术应用有指纹识别、人脸识别、机器翻译等。很多通过机械的计算和机械的记忆实现的东西，一般不把它看成人工智能的本质应用。例如人民大学送快递的智慧小车，核心技术是定位导航和无线通信，属于信息科技的进步，与人工智能关联不大。坦诚而言，现有的大多数人工智能算法只是对人智能的行为层面模仿，并没有进入到脑科学的范围内。

人脑六大认知功能，语言、思维、意识是人独有

于海：林老师曾说过，他最好奇的事情就是人的记忆存在什么部位，怎么进行储藏，又怎么巩固和提取。我知道所有的记忆最后都要落到神经元上，您如何看待类脑和人工智能的关系？

林龙年：“类脑智能”是当今世界性的热议话题，纵观21世纪的人类科技，从微观的基因到宏观的引力波都已获得了飞速发展，但是目前人类对大脑的工作原理仍然知之甚少。现有的脑学科研究主要关注人脑的六大认知功能：感知觉、学习记忆、情绪、语言、思维和意识。前三个功能是人类和动物所共有，后三个一般认为是人类大脑才特有。人类拥有独特的、完整的符号语言体系；在思维上有着严密的逻辑推理，如A大于B，B大于C，A就大于C这类“不证自明”的公理；同时拥有独一无二的意识体验，这也是人类自尊心的重要来源。目前每方面都有一些进展，但仍缺乏完善的理论体系来解释大脑到底是如何处理信息的，大脑智能在本质上又是如何实现的。因此对于类脑的研究也只能处于相当浅薄的层面。

于海：看来无论是人工智能还是脑科学的研究，终极目的是要打开人脑的秘密。

中国在人工智能领域的地位

于海：众所周知，人工智能作为新兴领域，是全世界都在竞相争夺的舞台。中国在此领域的全球大战中是处于“领跑”还是“跟跑”的位置，已有成就和未来发展潜力，都是我们所关心的。有人认为，中国在这一领域有优势与美国竞争，因为我们有大数据。两位专家怎么看？

除去美国，中国比其他国家更有优势，但原创不够

危辉：改革开放以来，国家在科研经费、实验设备的改造、人才培养方面的投入愈来愈多，多年的积累使得中国在国际的人工智能研究平台上表现卓



7月8日，复旦大学计算机科学技术学院教授危辉做客文汇讲堂人文解读“四大未知疆域”（太空极地深海网络）系列讲座最终场，主讲《“深蓝”出现20年后的人工智能》，并与华东师范大学脑功能基因组学研究所教授林龙年、上海交通大学科学史与科学文化研究院院长江晓原、复旦大学社会发展与公共政策学院教授于海对话圆桌论坛。

本次演讲音频请登录文汇微电台（APP喜馬拉雅·听-搜文汇讲堂2017）收听

本版摄影 袁婧、周文强 孙立帆参与本版整理

① 危辉、林龙年、于海、江晓原（从左至右）以不同学科视角对话圆桌论坛《我们如何握手“我们”》
② 现场300余名听众用热情拥抱“人工智能”

越。具体应用上，机器翻译，目前国内有不少企业在人机对话和中英文互翻上已处于世界一流的位置，在相关系统和软件的开发上也风生水起；在声音、文本多媒体技术方面，如大规模的视频监控和图像，都颇有建树。2010年上海世博会中，场馆入场就运用了人脸识别的技术，平均每天要对比近100万人次的人脸，没有娴熟的技术就无法应付自如。

从学术领域的几个观点上来看，首先，中国人在国际人工智能领域发表的论文越来越多。去年美国要召开人工智能的学术会议，时间恰巧与中国春节重叠，组委会决定将会议推迟一周，以便让更多中国学者得以参加。这说明中国学者的参与度直接影响了会议的质量；同时，越来越多的中国学者在世界顶级的期刊上发表论文。就我个人而言，2015年夏季有两篇SCI论文在线发表一周内被下载就超过700次，可见中国学者的研究极具先进性和代表性，因此，国外同行对我们的观点和原创性的研究越来越感兴趣。从这几个角度来看，中国在世界范围的人工智能研究上，虽比美国稍逊一筹，但对比其他国家有极大优势，毕竟我们人才梯队更大。

当然，在原创性的研究上中国还稍欠力度。以“阿尔法狗”为例，从学术上理解，它的研究价值不是很大，但商业价值颇高；目前国内的企业恰恰是在“阿尔法狗”研制成功之后才认识到它的重要性，纷纷跟着成立人工智能研究部。所以我们要要有前瞻性眼光聚焦原创性研究的应用价值，少做跟风出来的研究，不能国外有一个新概念出来，就蜂拥而上，当风头过后一窝蜂撤掉，更要着眼于质量而非数量。

中国要争取下一波人工智能高潮的原创性引领

林龙年：世界强国对脑科学的研究无不重视，美国、欧盟在2013年就已推出了相关脑计划，中国也正在酝酿中。至于人工智能，这一波起源是在加拿大，是当初很受冷落的几位搞神经网络的学者，不断的坚持和持续研究，再加上互联网带来的大数据和电脑计算能力的提升，催生了这波以深度学习为代表的的人工智能的繁荣。中国跟跑非常快，论文数量很多，但缺乏原创性的研究。很显然，这波深度学习绝对不是人工智能发展的顶点，只是发展过渡阶段的一个台阶，关键是要对下一个台阶有前瞻性的引领。

美国国防部有个高级研究计划局

(DARPA)，它经常会设立一些非常前沿的项目，比如早期对互联网和GPS的研究。2008年，它资助了一个很有意思的项目：开发与哺乳动物大脑具有相似形态、功能和构架的认知计算机，分派给IBM、惠普和修斯研究所三家单位来研发。2011年又进行了第二轮资助，一直到2014年，IBM发表论文宣布他们做出了人类历史上第一款仿脑芯片：真北芯片（TrueNorth）。所以，我觉得这种前瞻性的科技布局，是我们国家和企业以及机构更需要优先考虑的。

于海：现在很多国内学者将脑研究和其他学科联系在一起，譬如北京大学的江汀丁教授主持翻译的《神经元经济学》，您觉得神经研究是否不仅关乎人工智能，而且要进入到像经济学这类古老的学科领域，甚至每个学科？

林龙年：人类日常生活都由大脑控制，包括我们的经济活动，以及良心、道德。所以大脑任何的研究成果都与人类生活中的各个方面有关联性。当然，并不是每个学科必须关联，关联度也有大小。

中国工业机器人独步世界，但要重视社会隐患

于海：最后权威的话要让江老师讲，对人工智能和脑科学在中国的发展，您有什么断言或者评价？

江晓原：我补充一点，与他俩所说相得益彰。中国在工业机器人使用方面处于领头羊的位置，但在欢欣鼓舞的同时，我们也要意识到大批工人失业的可能性。在社会没有做好准备前，就一味地鼓励这样做，会存隐患。所以我建议，这种工业用机器人，如果能在产品生产技术上更新换代，则可考虑，若仅仅出于降低经济成本，我觉得地方政府不应鼓励，甚至出台政策，因为结果可能得不偿失。中国现在走得快，要考虑到并存的风险。

类脑是人工智能发展的方向

于海：作为学习者，我想弄清楚图灵测试和类脑两条线是否承前启后，也关心精神和身体是怎样的因果作用？这些只有从大脑工作的机理开始理解。虽然第一步的人工智能还是“很笨”，但不等于它不会进步，因此我们要讨论一下类脑有没有可能，这已经接近思想试验了。

人工智能科学家不会满足于只是模拟智能，还要模拟情绪、意识等人类独有的功能，听说现在已经有了欲望的算法，以后会有情绪算法、恋爱算法、信仰算法吗？神经科学的进步，会促使这个模拟走到哪一步呢？

历史教训：流行40年的逻辑主义流派的短路

危辉：做类脑研究的计算机专家，已经依据视觉研究的一些成果来改进算法，即严格仿照大脑的机制来设计计算模型，受益匪浅，可能以后的计算结构和此前标准的图灵计算范式完全不一样。

在1950年代，计算机界非常推崇按照逻辑主义做人工智能，到1990年代还很盛行，如非单调推理等，后来发现很多复杂的问题解决不了。当时人工智能界认为，因为我们放弃了对于鸟的模拟，所以造出了飞机，由此推论出如放弃对大脑的模拟也能取得人工智能的成功。但逻辑主义流派的多年实践表明此路走不通，还是回归到模仿大脑之路。目前来讲，这条线路还是非常高效——要研制和大脑机制很像的计算机。

深度学习和大脑机制关系不大，未必是方向

这样来审视一些应用，显然存在方向性误区。目前特别流行的技术EEG和功能磁共振技术，更准确地说是现代信号处理和数值计算手段来分析取自于脑外部的数据，而不是类脑计算；深度学习，与大脑机制的相似性很牵强。第一大脑神经元分两种类型且分布不同，而深度学习的网络里，计算单元完全不分类型；第二大脑层数就六层，而深度学习动辄就上百层；第三大脑是很多脑区协同工作，有很多自上而下的反馈，但深度学习模型里只是往前推，没有关键的反馈机制。如果一定要把深度学习放到类脑的路线图里，我觉得过了几年会失败，因为它解释不了大脑的工作机理。

大脑尚未研究清楚，无法判断类脑能走多远

林龙年：我觉得深度学习模拟大脑把网络分层，是一个进步。脑科学家虽然获得了视觉相关脑区的网络层次结构图，但说实话对它的功能，还不能很完整地解释，只能给计算机界提供一些概

念。从脑科学角度来看，目前的类脑研究还在起步阶段，“类脑”这个叫法也主要流行在人工智能界。让我评判类脑可以走多远有点勉为其难。但是，人脑奥秘没有研究出来，是否就一定不能研究类脑呢？也许并非完全相悖，在某些方面或许可以相互促进。

神经元研究是否会扩展到文化元和社会元

于海：视觉识别研究很接近大脑机制，并取得局部的前沿成绩。那是否可以从人脸识别，走到人脸美丑识别或人品识别？所以，我在想，科学家想通过神经元的研究，把大脑的智力过程看透，是否还该把人性看透，这样研究就会从智性走向德性走向情性，神经元研究要扩展到文化元和社会元更广的领域里，辐射到人的善恶、美丑、真伪等价值领域。

科学的双刃剑是否可以克服？

于海：尽管危辉一开始认为人工智能事实上比较“笨”，还有待完善，但现实已经带来了极大的便利。我是这个年龄段最早上社会学英语课程的老师，现在已是复旦大学内教英文社会学课程最多的一位了，当年曾在英文上苦下功夫，但现在发现这些并不那么重要，人工智能已经能既快又准地做双语翻译，哪怕有口音。这是否意味着以外语为业就此收场？这也让我们深思，科学的双刃剑是否可以避免？

用马克思主义学说来分析——科学是人这个主体创造的，但会背叛主体、伤害人类，这一节标题特别——我们和“我们”，就提示了这种关系的微妙。5G即将出台，手机就是最好的AI的终端，它改变生活也在改变人的心理。比如，有人就只愿意和人工智能机器人结婚。所以江老师曾言，人类在玩两把最危险的火，一把是生物技术，一把就是人工智能。

如果中国人不做，其他国家科学家也会做，怎么办？

江晓原：人工智能专家们面对公众对其导致广泛失业危害的质疑时，经典的回答是：一个新技术导致旧岗位失去，但会产生一些新的岗位。但他们又说将来90%以上的工作岗位都可以由人工智能替代，那新岗位是不是包括在这90%中？我认为广泛失业必然导致人类社会不可逆转的不稳定，最终甚至是人工智能毁灭人类。

于海：对人类社会来说最高的原则是存在，中国文化讲“天地之大德曰生”。如果人工智能最后的结果是让人类灭亡，就有悖于这个最高原则。

林龙年：不一定。如果他们灭了，或他们把我们灭了，各50%的概率。

江晓原：如果这样，现在干嘛还要把它们研究出来呢？这会让双刃剑两个刃越来越不对称，而且祸害一端越来越大。过去，我们认为科学技术是无限美好、尽善尽美，但现在需要引入一个维度——科学技术会带来危险，因此有些技术要从源头上阻止它发展。

林龙年：但如果中国不发展人工智能，美国、英国科学家依然会加速研究，当一个国家科学技术停止发展时，后果只能是处于中国百年前的“落后挨打”或被灭掉的局面。

对于脑科学来说，一个终极的问题是研究智能，我们利用动物的脑研究人类有无意识，按江老师的说法，显然是恶，因为我们杀了动物，确实是！但我们获得的是脑科学的知识，反过来可以拯救脑疾病的患者。

江晓原：确实，我们不做别人就会做。爱因斯坦是造原子弹的“曼哈顿工程”的推动者，他写信给罗斯福总统说，纳粹德国在研究原子弹，会威胁美国这样的自由世界，由此，大家看到了目前的发展现状。但晚年他签署了宣言，要求禁止使用核武器。因为爱因斯坦认识到了，这件事情本身是恶的，当时不得已做了。

我在新书《今天让科学做什么》中把科学技术分成了四种类型，有一类是充满争议没有必要做的，还有一类是充满争议现在不得不做的，人工智能属于后者。比如说美国现在最热衷的就是军事用途的人工智能，和当年研发原子弹相类似，所以如果我们发展军事用途的人工智能，也应该抱着当年搞“两弹一星”一样的心态。

为何要把老虎养到吃人时刻，而非最初就预防？

江晓原：刚才说我是技术悲观论者，的确。请问危教授，您有一张PPT谈到无人驾驶的汽车出事故无法定罪，既然看到了这些危害，您觉得能为此做些什么呢？

危辉：说实话，现在对于这些危害，我们的认识还很浅，因此，我们会继续把技术往前推。

江晓原：你不觉得有问题吗？

危辉：的确有问题。以无人驾驶为例，2016年在拉斯维加斯全球消费电子展上的自动驾驶技术，汽车过十字路口要看红绿灯，理论上装一个摄像头即可，但现实场景中很多的红颜色和绿颜色与红绿灯很相似从而干扰感知，怎么解决？到路口的时候，GPS会告知经纬度，我把所有的十字路口都做一个IP地址，通过GPS指引来访问交通信号灯网站，得知当前的红绿灯状态，以此决定是否开车过去。这件事情理论上听起来非常完美，但只要有人黑掉了网站，就把红绿灯的信号改变了，车就会发生事故。但对于技术设计者来说，他觉得这一新手段没尝试过，就会去试试。

江晓原：谢谢，我听到了一个很生动的回答。我曾用“养虎为患”来形容人工智能的危害，但总有人抱着侥幸心理：“老虎还小，离吃人还很远”。现在很多人有同样的逻辑。在考虑好这些隐患之后再决定养不养虎，这样的逻辑不是更合理吗？

危辉：在风险尚小或还能控制时，我们更多的考虑的是技术的挑战性或可行性。

科学是中立的！双刃剑存在于使用者的立场？

于海：圆桌就在于讨论，科学家认为用中性方法做出来的科学成果，它即便也有祸害，只要所得多于所失就值得。江老师所说的“曼哈顿工程”就是一例。所以今天的克隆技术，就有伦理限定，当技术可能破坏掉我们日常遵守的伦理底线时，科学家是否要考虑可能的后果，还是依然觉得科学的真理高于一切？

林龙年：科学的双刃剑问题永远存在。毫无疑问，祸是使用它的人造成的，不是科学家造成的。我们现在对于科学危害的任何恐惧，对民族的生存没有任何帮助，整个人类社会的发展不会因为 we 害怕它而止步不前。

如克隆技术般立法，人工智能不触碰“意识”底线

危辉：于老师所说的克隆技术已经立法，做了很好的典范。人工智能是否要和克隆技术一样有立法？我们做类脑计算，从模拟人的计算到推理到人的记忆，一旦模拟到人的意识时，就可能要刹车。

有两个很好的例子，第一是电影《机械公敌》，男女主角拼命要保护一个机器人，但公司要把它毁了，因为它是唯一进化出自我意识并今天会做梦的机器人。第二是电影《AI（人工智能）》，主人公小男孩是人工智能，父母因为事故失去了儿子特别想念他，用号称最先进的的人工智能技术复制了儿子，他除了永远长不大，和真实的孩子一模一样。但是最后他的父母死了，这个孩子依然年轻且永远不死，他沉湎在对父母无限的思念中，最后不得不把自己沉到海里求灭亡。

所以，这两个科幻例子在告诉我们，研究人工智能应该有底线——不可触碰“人的意识”。

江晓原：补充一点，真正心怀不轨的人工智能，有可能约定好一个暗中的原则，就是不要通过图灵测试，明明能通过也故意不通过，这样就可以避免引起人类的警惕。

林龙年：我觉得有意识的机器人会是未来类脑智能研究的一个非常重要的发展方向，各国肯定会投大量的资源去做这件事。

人工智能发展维度：多学科并进、化危为福

于海：我们今天在台上，看上去好像互相有点冒犯，这就是学术讨论。科学家要解决问题，非常务实，我和江老师是哲学人文训练，喜欢脑洞大开的思维。

今天的人工智能，要达到人的灵巧度和人的智慧度依然路漫漫但它会进步，我们必须关心走下去的两个限度。其一是伦理限度，如果机器人真的有意识，具有的智能超过人类，与机器人该怎么相处？其二是技术限度，是否会走到这一步，需要多久？把这些问题提出来，今天并无现成的答案，但作为银河系的高等物种必须思考。

同时，未来人工智能并非科学的单兵突进，它必然需要多学科并进，并以人类现有的文化、政治、协商等制度和法规去约束和规范。如今天两位专家所说，生物历经40亿年进化而成就的人脑，才展现了今天低能耗高智慧的奇妙，应了恩格斯所说“人类思维是人类进化史、地球进化史最美的一朵花”。所以，对于人工智能，人类应施展科学探索和伦理考量的双臂去拥抱。（明日将刊发危辉主讲及整场互动）