

对谈
无论是在象棋对局或日常生活中，都会出现“人工智能”VS人类？
由“人工智能”思考人活着的意义

酒井邦嘉（东京大学教授） VS 羽生善治（棋士）



酒井邦嘉（东京大学教授）

羽生善治（棋士）

人工智能开拓了新的下法

羽生善治

这几年来，AI也即人工智能的话题变得越来越热门。可以认为，这是因为人工智能终于以肉眼可以看得见的形式而出现，比如人型机器人和汽车自动驾驶等，这是一个巨大的契机。至今为止，它在国际象棋、象棋、围棋等棋盘游戏的世界中不断发展到今天。但是，最近其应用领域变得更为宽广，至今为止只不过是痴人说梦的人工智能，渐渐出现了能活跃于现实世界的可能性。对此，人们在寄予希望的同时，貌似也开始对人工智能凌驾于人类之上的“Singularity（科技奇点）”感到恐惧。

酒井邦嘉

我是语言大脑科学者，对人工智能的思考，其实能够引导我们思考“人是什么”。也就

是说，它和关于大脑功能的思考有着许多相重合的部分，所以我也经常思考并十分关注人工智能相关的问题。而关于奇点，社会上存在着很多误解，所以我们稍后还将继续提到这个话题。

羽生

人工智能和棋士的对战受到举世瞩目，正因为如此，近一年以来，我也增加了很多的机会出席类似与人工智能对决的企划，对此我很是困惑（笑）。在一段时期里，人工智能曾经处在停滞状态，但是，让人们彻底改变看法的是2011年举办的棋士对人工智能的第一届“电王战”，米长邦雄先生输给了人工智能。我觉得就是从那时候开始，人们开始讨论对人工智能的真正研究和运用。

不管是象棋还是日常生活，人们都有自卫本能，只要感觉到有危险，就会做出多余的举动。所以我们棋士在作为职业选手磨练技术的过程中，就会反复不断地修行如何控制住这种恐惧。而与此形成对照的是人工智能并不具备这样的自卫本能和恐惧感。因而它有时会提出一些人不可能想到的创意或灵感。所以我们现在也已经有很多可以从人工智能的棋谱中学到的东西。

比如说在象棋中有一招名叫“相矢仓”的策略，有一次电脑软件找出了其攻略法，而目前尚未发现制衡它的对抗措施。其结果便导致职业棋手使用矢仓的人变得越来越少。那么要说为什么电脑可以编制出这种攻略方法呢，因为电脑的思考方式缺乏一贯性，而人的思维却是你这样来，那我就这样对付，也就是说人的思维具备连续性。而电脑却并非如此，试着跟它下棋，你会觉得它就像是哥伦布与鸡蛋那样，有很多简单却十分出乎意料的下法。

酒井

人是以时间推移为轴线进行思考的。象棋的有趣之处，在于每次局面都会不同，所以每一次都必须重新考虑。

实际上，在与人工智能的对战中，人渐渐地开始无法在象棋及围棋方面取胜了。但是，如果人就此放弃，那这才是真正意义上的败北。二十年前，当时的冠军卡斯帕罗夫输给人工智能的时候，有一些专家对棋盘游戏本身作出了否定性的批判称“国际象棋是低水准的游戏”，这堪比伊索寓言中的“酸葡萄”故事，难以视作正当理由。归根结底，不对游戏内容作出评估，而只是单单讨论赢了或是输给了人工智能，那也未免太过表面化了。失败有时候也能让人从中学到一点什么。

如今，人工智能已具备实力并受到关注，那么如果对战100局又会如何呢？

羽生

会怎么样呢？人与电脑对于时间的概念是不同的。人和人工智能同样，都是越有时间就越好，但象棋必须在有限的时间内作战。也就是说有必要在较短的时间里，保持高质量的判断。我听说象棋人工智能的程序员在其学习阶段，会让它进行一秒钟一局的训练，如此高强度的学习，如果是人，早就累趴下了。所以，对战这件事我认为并不现实。

刚才说到国际象棋的时候提到，当今国际象棋的冠军是一位名叫卡尔森的挪威人，现年二十六岁。现代的年轻一辈棋士在出生时就已经有电脑了，故而在下棋风格中可以看到经过快速计算处理和分析得出的招式。可卡尔森却恰恰与之相反。他采用的反而是考虑到

如何留下更多可能性的下法。乍一看似乎貌不惊人，但其实我注意到，他也是从电脑软件中筛选出富于人类特点的要素，将其运用为自己的优势。今后，象棋也将步入以电脑为前提磨练技巧的时代吧。

在这样的过程中，我有一个疑问，当人与人下象棋时，互相会进行“感想战”。在决出胜败后，两个人一起进行反省和验证，回顾“这一步走得很好，这一步不好”等。但是，软件不可能这么做。也就是说，人们至今为止把重点放在对过程的研究上，那么今后将只是收集数据而已。这样做是否就已足够了呢？

酒井

其实教育也面临同样的问题。现在的年轻人觉得思考太麻烦，因而会使用搜索引擎直接搜索问题的答案。他们很着急地就想知道答案。而不管是象棋也好，学问也好，原本是因为其过程很有意思，所以才去学习，但现在却变得想在短时间内获得结果。这样一来，“学习”本身也就丧失了意义。

在教育第一线，有人反思一味地追求效率和竞争，大学入学考试也出现了新的倾向，就是不再全部都是选择题，而是加入一些记述性的问题，对思考的过程有更多的重视。这是很重要的改变。

人工智能的到来将改变文明？

羽生

回顾以往，这几十年来，人工智能热潮有过好几次的高潮，又恢复平静。我听研究人员说，现在正想跨入真正动用人力与预算的时代。我对人工智能的开发是带着期待的心情来看的，酒井老师您是什么样的想法呢？

酒井

关于人工智能，现在社会上往往以“人的工作会被夺走”等容易带来不安的形式，讲述这个奇点的问题，但是，作为一个从事科学研究的人来说，我想指出这只不过是杞人忧天。人们都说，人工智能超过人将会是在约三十年后，也就是“二零四五年问题”等，这根本没有任何科学依据。说得准确一点，人放弃的瞬间就是奇点。比如说在某项能力上输给了人工智能，那么只要说“不，人还没有输给人工智能”，并且不放弃就可以了。只要人还在继续进行这样的挑战，奇点就绝对不会发生。而且，说到底我们还不知道人究竟是什么，却在对人和人工智能进行比较，这本身就是一个毫无意义的命题。

羽生

您是说，人的极限就是人工智能的极限吧。

另一方面，人这一方也有潜在的危机。因为人工智能是机器，所以很少出现错误。因此人就把所有操作全都交给人工智能，这样就会产生危险，有可能发生无法想象的事故。最容易想象的就是自动驾驶。举例来说，植入了人工智能功能的汽车在进行自动驾驶时，假如有什么动物突然撞上来，或是发生了类似的突发事件。像这种时候，人工智能究竟能

否回避危机？这是因为人工智能的功能其实是通过概率创造而成，经过大量的测试进行学习后，执行被判断为正确答案的程序而已。也就是说，如果出现的事态是未经过测试的，那就无法作出应对。

酒井

忠实地执行指令和不出现错误是两回事。我们不能忘记人工智能是以概率为基础而运行的。

人在依赖人工智能的时候产生的转嫁责任问题十分严峻。假设自动驾驶的汽车引起了人身事故。汽车的主人一定会说：“因为是自动驾驶，所以我没有责任。是汽车不好”这种话吧。但是，如果汽车的程序未见任何故障，那么受害者应该向谁要求承担责任呢？如果是人开车就可以避免的问题，那么选择自动驾驶的人就有可能因为其过失而被问罪。

因为过于重视效率而过于依赖人工智能，那么也会产生人有可能失去正确判断能力的担忧。那就是“开动脑筋，从而使自己可以尽量不用动脑筋”。这是一个奇怪的现象。假如用这种心态去运用人工智能，毫无疑问就会引起文明的衰退。

关于“奇点”的奇怪讨论

羽生

我在某个电视节目中造访的世界人工智能研究第一线，有一家军事相关的企业。现代战争已经运用了人工智能，但因为人工智能不知道会在什么时候“不听话”，所以人也同时监控着人工智能。但是我听说，监控人工智能的人很多都得了忧郁症。说到底，战场本身就是一个非同寻常的地方，在人工智能作出判断并行动的过程中，有些是人无法理解的东西，一直盯着其举动观察，渐渐地人就开始患上了毛病。出于有可能因使用人工智能而造成惨痛的事态，所以就有必要研究人妄用人工智能带来的风险吧。其实这并不像科幻片中所描写的人工智能私自行动并开始攻击人这种故事，而是人在这一方的伦理问题。美国的谷歌公司早在创建公司的初期，就已经设置了伦理部门，可以说他们很有先见之明。

酒井

在进行奇点的讨论时，也会提及这种伦理性的问题，但与此同时，更为明显的论调便是人被人工智能赶走这种担忧。其实这并不是仅限于人工智能的固有问题。当人们获得科学技术的卓越成果时，很悲哀的事便是会出现某些人想要将其作为武器或兵器。今后，我们也有必要制定人工智能的伦理规定等，预见各种各样的事态，这当然很重要，但是，“正因为如此，所以就应该停止所有人工智能的研究”这种极端的主张根本不可能解决问题。

羽生

是的。对包括奇点在内的人工智能各种问题的研究，有一位名叫雷蒙德·库茨魏尔的先驱提出了“收获加速的法则”。其含义是指任何领域的研究在进行到一定程度的时候会出现停滞。但是，在其他领域的研究将会给这种停滞的研究带来新的启发，作为社会整体会越来越加速发展。即使人们停止了对人工智能的研究，应该也会因为周边的研究而孕育出

与人工智能类似、或是同等的研究成果。

酒井

在关于奇点的奇怪论调中，迈克尔·奥斯本提出了“未来的雇佣”这一预想。这是对今后有可能因人工智能的出现而“消失的职业”以及“消失的工作”列出的清单，从根本上来说，这份清单过低评估了工作中所包含的人的能力，在这一点上已经就犯了重大的错误。比如说，这份名单中出现了有必要具备高度熟练度和经验的钟表修理工、照相机的修理工，单单是这一点就让人怀疑其可信度。

服务业所需要的能力，包括工作的安排、对顾客关心等。现在还没有任何证据表明，在不久的将来，人工智能就能实现这些。另外，专业人员的工作中，有着不输机器的准确度和精细度。

羽生

代表着象棋软件的专业人工智能和通用人工智能的同时存在，也是给讨论带来混乱的原因之一，人不经意的动作中，融入了许许多多的判断。比如向一位上幼儿园的小朋友出示正在飞翔的无人飞机。只要他多看了几次，一定能够认识到这和以往看到过的鸟儿有所不同。据说像这种判断其实有着很高的难度，例如，给人工智能看了很多猫咪的照片后，再给它看一张新的猫咪照片，让它判断这是猫咪。人工智能在四年前才刚刚达到这个成果。也就是说，人在不经意间同时进行的学习和推论，而要将它作为算法组合到人工智能中去，将会是难度很高的事。所以说，对至今为止尚无经验的事物，通过适应、磨合、选择，并作出决断是相当难的事。

酒井老师，希望您能为我们作一个讲解。

酒井

人之所以能够同时进行学习和推论，是因为能同时平行使用大脑的不同部位。刚才看到了无人飞机的孩子，能在一瞬间学习到新物体的特征，与此同时，在一瞬间对鸟的形状以及飞翔方法之不同进一步作出推论。这就是人与人工智能的认识之本质差异。

人工智能是模仿脑神经网络得到开发的。近年来发展迅猛的“深度学习”，便是有着类似于大脑视觉皮层那样的大量中间层，各两层的事先学习能发挥效果。即便如此，人工智能的能力与实际大脑相比，还是远远地更为单纯并存在着局限。

顺便说一下，电脑如果不安装特殊的程序，充其量不过是和电子计算器同样的计算机。即便通过电脑在数学上作出了证明，那也不能称为是“判断”。而另一方面，程序有时会因人预想外的错误而出现问题。也就是所谓的“漏洞(bug)”。人的错误根据以往的经验，在一定程度上可以作出预测，但是当人工智能暴走时，有可能无法推测其错误是低次元的还是高次元的东西。于是就产生了一个问题，是否能把对人而言很重要的判断交给这样的人工智能去做呢。说到底这也属于人这一方的问题。

羽生

原来如此。牛津大学人类未来研究所发表了《人类的十二大风险》，其中列举的有气候变化、流行疾病、经济混乱等。这些仅仅只是风险，但是这十二项中也包括了人工智能，虽

然可以考虑到人工智能有许多风险，但它也有可能解决能源问题、粮食问题以及其他所有的风险。人在制定道德规范和规则等上面，的确有很多需要考虑的问题，但是，即使有谁呼吁停止人工智能的开发，其他人也还是会开始的，那么就只有向前进了，我的想法就是这样。

人工智能提问人的心灵和活着的意义

酒井

要想把人的心置换成人工智能，就必须先解决如何看待心这一难题。直到现在，人们在从科学角度看待心这一自我意识及人格上尚未取得成功，现状甚至对“心”还没能作出定义。这是因为心不停地在绕圈、循环，所以即使想要定义普通的心状态，也没法做到。因为没有一定的标准，也就无法作出比较。正如“说谎者悖论”那样，对于“我的精神状态是正常的”这句话，很难在自己身上发掘出为这句话提供保证的明确证据。

羽生

心的不可思议之处，还有一个就是安慰剂的效果。安慰剂似乎对内科性的疾患也有效果，对精神疾患有时也能发挥作用。无法用科学表达的现象会在心中发生。

酒井

人的心极其多样化。不只是由遗传因子决定的部分，而且还有可能是经历了许多经验而变得独具个性。那既然如此，在设计人工智能的时候，就会产生这样一个问题——究竟应该以谁的心作为标准呢？而且，人的心和语言之间还有可能产生偏离，所以猜测他人的心思基本是一件不可能的事。即使人工智能越来越接近人的心，只是这样的话还不能成为商品。如果作为家庭机器人而开发，那么如果他和人一样跟父母吵架或是出走之类的，那就没用了。（笑）

羽生

话说我想问一下，人工智能是否有可能获得语言？我阅读自然语言处理的书本时，原以为会有写到语言的部分，但事实上只有算式。那么就是说，语言也有可能转换为算法。的确我也感到最近电脑的自动翻译功能似乎有所提高。

酒井

从表面来看，你有可能从语言表达中感到人工智能的进步，而对其不足部分作出补充并加以理解的正是人。说到底，以概率、统计以及学习作为基础的人工智能，从原理上来讲是不可能捕捉人类语言的。这是因为人类语言的核心——也即语法判断，完全独立于这些因素。早在六十年前，语言学家诺姆·乔姆斯基就在其著作《语言学理论的逻辑结构》中指出了这一点。乔姆斯基也以构建了人工智能的核心、即自然语言处理的基础而闻名，不知道为什么众多研究人员对这本书置之不顾，却抱有“语言是很容易就能掌握的”这种错觉。

羽生

语言，还需要再加上感情呢。以前我有机会采访一位使用人工智能作曲的研究人员。在作曲之余，他又想到了让人工智能写诗。然而，他感到了一个问题，此举究竟有没有意义呢？也就是说，它应该能够写出一些作品，但是诗应该是反映作者人生与经验的东西，其意义正是在于此。人工智能创作的诗，只不过是文字的排列罢了。

酒井

那如果这样的话，人工智能作的曲也不过是音符的排列而已呢（笑）。正如这样，人工智能让我们重新考虑语言的价值和艺术的价值。真正的人工智能研究，恰恰是了解人的一个手段。

羽生

正如酒井老师所说的“人工智能所思考的东西就是人类思考的东西”。我感觉人工智能就像是人照出自己的一面镜子，通过研究人工智能，就能重新对人活着的意义作出丰富的解释。

[译自《中央公论》2017年4月刊，本文经中央公论新社同意翻译转载。]

酒井邦嘉（东京大学教授）

1964年出生于东京。毕业于东京大学理学部物理学科。修完该大学研究生院理学系研究科博士课程（理学博士）。1997年任东京大学研究生院综合文化研究科助教、副教授。2012年任该校教授。专业为语言大脑科学及脑功能概念。

羽生善治（棋士）

1970年出生于埼玉县。1985年作为有史以来的第三位中学生棋士、职业棋士四段而出道。1989年获得首个龙王战冠军。1994年首次夺取名人，成为将棋史上第一位六冠王。1996年包揽了七大冠。2008年获得十九世名人的永世称号。
