

人工智能能否超越人类智能

王晓阳

(上海交通大学 哲学系,上海 200240)

摘要:半个多世纪以来,人工智能研究似乎取得了迅猛发展,但批评之声也一直不绝于耳。本文亦是一项关于人工智能的批判性分析,针对的是一个关系到人工智能研究基础的问题,即“人工智能能否超越人类智能”。我将尝试构造一个基于集体人格同一性(collective personal identity)的新论证,从而表明:首先,该问题原则上无法仅在经验科学的框架中获得有效解决;其次,该问题其实是一个没有认知意义的问题;最后,如果该问题没有认知意义,那么人工智能研究的一个重要目标,即“生产出一种新的能以人类智能相似的方式做出反应的智能机器”,恐怕也无法实现。

关键词:人工智能;人类智能;集体人格同一性;形而上学;人工智能哲学

中图分类号:N031 **文献标识码:**A

一、引言

人工智能(artificial intelligence, 以下简记为 AI)“是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。AI 是计算机科学的一个分支,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器,该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等”。^[1]

以上这段引文来自因特网“百度百科”中的“AI”这一词条,也是一个几乎可以得到普遍认可的关于 AI 的标准理解。从上述引文,我们不难发觉 AI 研究的一个重要目标,那就是“生产出一种新的能以人类智能相似的方式做出反应的智能机器”。

可以说,自麦卡锡(John McCarthy)在 1955 年达特茅斯会议上首次提出“AI”这个概念之日起,绝大多数尝试制造智能机器或计算机来模拟人类智能(human intelligence, 以下简记为 HI)的研究都是朝着这一目标努力的。事实上,尽管在涉及和模拟人类的高级认知能力或智能方面(比如,情绪的表

达、艺术的创造力等),AI 研究似乎一直鲜有进展,然而在涉及和模拟人类某些特定的智能方面(比如,逻辑推理和计算),AI 研究无疑已取得了令人瞩目的伟大成就。这方面广为人知的一个例子就是,IBM 公司近十几年来一直在大力推动的“人机大战”。例 1,1997 年国际象棋世界冠军卡斯帕罗夫在前五局战平的情况下,最后一局(第六局)输给 IBM 的国际象棋计算机“深蓝”。例 2,2011 年 IBM 专为多年来一直风靡美国的智力竞赛电视栏目“危险边缘”(Jeopardy!)所设计的计算机“沃森”与该节目历史上两位最成功的选手肯-詹宁斯和布拉德-鲁特进行了一场终极对决,最终“沃森”完胜。沃森的总成绩为 77147 分,远高于两位人类选手 21600 分和 24000 分的成绩。

在许多人看来,如果说,例 1 开始让人们认识到 AI(电脑)在某些方面(如,推理和计算)可以与人类较量的现实,那么例 2 就足以表明,AI 不仅在哪些方面已经可以与人类较量,甚至开始超越人类了。然而在我看来,关于 AI 在某些方面(如,推理和计算)所取得的成就以及未来进展所抱有的这种乐观甚至推崇的态度,不免会引起如下这个值得我们

收稿日期:2015-04-14

基金项目:2011 年度国家社科基金青年项目“意识研究”(11CZX023),2014 年度上海市教委科研创新重点项目“解释鸿沟难题研究”(14ZS069),2014 年度国家社科基金重大项目“欧洲生命哲学的新发展研究”(14ZDB018)。

作者简介:王晓阳(1978—),江苏泰州人,上海交通大学哲学系副教授,主要研究方向:分析哲学、形而上学、认知科学。

以上两个例子参考了互动百科“人机大战”的相关报道。^[2]

进一步考虑的问题,那就是:人类有可能制造出比人类自身智力水平更高的事物吗?

二、问题分析

上述问题是本文将要处理的目标问题。不难理解,这个问题无疑也是一个涉及 AI 基础的关键问题。在弄清楚这个问题之前,我们似乎有必要事先澄清如下两个相关问题:第一,这里的“制造”是什么意思?对于这个问题,需要说明的是,本文在通常的意义上使用“制造”这个词,但是“制造”这个词的确切含义在以下的讨论中并不重要。因为,这篇文章的目的并不在于讨论有关“AI”的一些技术性问题。我所关注的仅仅是一种哲学上的“可能性”。这种需要加以考察的“哲学上的可能性”就是,是否有可能制造出一个比人类自身具有更高智力水平的事物?换句话说,本文并不打算从技术的层面对“如何制造智能机器或智能物”进行具体的可行性论证,而是试图从哲学层面来考察“AI 是否(原则上)有可能超越 HI”这个涉及到 AI 基础的问题。第二,如何来界定这里的“智力”?在本文中,我仍沿用当前学界通行的思路,即凭借某种已获得特定权威机构或者科学共同体认可的或可靠的智力表来测试受试者的智力。利用这种测试方法所得到的结果是一些经验数据或数值,我们再将这些数值与智力测试表中的标准数值进行比较,从而就可以判断受试者智力水平的高低。这种做法其实就是凭借特定的智力表来测量受试者的智力水平。因此,如果我们愿意接受这种判定智力水平的方式的话(这也是目前学界通行的做法),那么在上节末尾提到的问题就可进一步转换为如下更为明确的表述,即我们是否有可能制造出一个(或一类)事物 K,再利用特定的智力表 S 进行测试,则得出这样的结论:这个(类)事物 K 的智力水平比人类中任何一个成员的智力水平都要高?对此问题,可区分如下三种不同的情形论述之:

第一种情形。无论什么原因,如果我们始终无法提供任何一份合适的智力表 S,用以测量事物 K 的智力,那么,我们根本就无法判断 K 具有怎样的

智力水平,更不用谈所谓智力高低的问题了。

第二种情形。如果我们制定出了这样一份合适的智力表 S,并且从世界各地随机找来各种各样的人(他们有着不同的信念、职业或性格等等),与事物 K 进行比照测验,经过一系列长期反复的严格测试,直至满意为止,结果显示:K 确实具有比找来参与测试的任何一个人高的智力水平。由于我们相信为此而设定的智力表 S 的可靠性以及所选取的比照组(即随机找来的各种各样的人)的客观性,因此可得出这样的结论:K 具有比人类高的智力水平。但是,即便这样,就可以宣称已经制造出一个(类)比人类智力更高的事物 K 吗?我认为还不行。首先,我们可以说智力表 S 并不是合适的,不能如实地反映事物 K 的实际智力情况。其次,即使认可智力表 S,也无法忽视一种未来的可能性,即似乎目前没有任何正当的理由可以阻止我们相信如下这种可能性:未来有可能出现一种我们目前还无法获知和拥有的技术,帮助人类实现智力水平的巨大飞跃,并且经由 S 测试,未来的 HI 仍高于 K。可见在这种情形下,只能表明,经由 S 测试,K 的智力水平高于至测试时为止的人类智力水平。因此,如果这种可能性存在的话,那么严格地讲,由于没有考虑时间因素,“可以制造出比人类智力水平更高的事物 K”这一说法,仍然是站不住脚的。

第三种情形就是在第二种情形的基础上,把时间因素考虑在内。这种情形就是,不管通过何种途径,假设我们能够制定出一份有效测试人类成员的智力水平的智力表 S,不仅现在有效而且将来任何时候都有效。下面我们来具体考察这种情形,可进一步再分为下述两种情况来讨论:(1)假定我们已经制造出了事物 K,不管是一台复杂的机器还是某种奇怪的生物体,我们要测试 K 的智力水平,该怎么办?按照规定,唯一的途径就是对照智力表 S。但是,S 是否适用于 K 呢?按照上述规定,我们可以用 S 来测量人类成员的智力水平,并且我们也知道 K 是我们的制造物,但是由于 K 不同于人类,因此我们无法判断 S 是否也适用于 K。也就是说,在这种情况下,我们根本无法断定是否可以用 S 来测试 K 的智力水平,更谈不上和人类的智力水平加以再比

在哲学因特网百科全书“人工智能”这一词条中,^[3]Hauser 建议将有关 AI 的问题区分为两大类:第一类问题是“机器究竟能否思考”?这类问题由阿兰·图灵(Alan Turing)首先提出,而著名的图灵测试(Turing Test)则是关于此类问题的一种解答方案;另一类问题是,“机器智能能否接近或超越人类的智能水平”?前者是 AI 的初阶问题,而后者则是高阶问题。本文重点关注的是后者。

请注意,本文一开始就预设了 S 是我们可以用来测试和衡量人类成员智力水平的唯一方式。

较了。换句话说,在这种情况下,我们对于“我们可以制造出比自己智力水平更高的事物”这一说法,仍然无法判断。(2)如果S也适用于K,那么我们可以认为K的智力与人类的智力之间具有一种“可比较关系”(comparable relation, 以下简记为CR)。注意:这里的CR是关键。如何理解CR呢?我的建议是,可以利用以下双向条件句来进行刻画:

(CR) 事物t1和事物t2之间具有一种“可比较关系”,当且仅当t1和t2之间具有某个或某些相类似(analogous)的特征(character)C。

例如:这朵花比那朵红,当且仅当这朵花和那朵都具有相类似的红色特征,这样我们才可以比较二者之间的红色程度。

同理,当且仅当K和人类在智力方面具有某些类似特征,则K和人类之间才具有CR。然后再借助S,我们就可以测量和比较二者之间的智力水平了。可以说,至今为止AI几乎全部的研究工作都是基于承认或接受CR这个前提之上的。据我所知,尽管对于AI的批评一直不绝于耳,然而到目前为止几乎还没有人怀疑这个前提的必要性。换句话说,至今为止,无论是AI的支持者还是反对者,似乎都普遍接受如下基本信念:不承认K与人类之间具有CR,则我们所理解的AI研究是不可能开展下去的。这一点从本文开头所引的关于AI的标准理解中也不难看出。然而在我看来,这个必要前提或基本信念却是可疑的,因为不得不牵涉到形而上学层面的人格同一性论题。可以说,至今为止,我们关于AI的研究工作或争论(无论AI支持方还是反对方)似乎都没有注意到人格同一性论题与AI研究之间的关系。因此本文的重点就是:有别于目前对AI的常见质疑,将尝试借助人格同一性论题来构造一个从不同角度质疑AI研究基础和研究目标的全新论证,即“基于集体人格同一性的论证”(argument based on collective personal identity)。通过该论证,我将要为三个密切相关论点提供辩护:首先,目标问题(AI究竟能否超越HI)原则上无法仅在经验科学的框架中获得有效解决;其次,目标问题其实是一个没有认知意义的问题;最后,如果目标问题没有认知意义,那么AI研究的一个重要目标,即“生产出一种新的能以HI相似的方式做出反应的智能机器”,很可能也无法实现。

三、关于AI的常见争论

在给出新论证之前,先简要回顾一下关于AI的常见争论或许是有益的。按照Hauser的梳理,^[3]当前学界对AI的种种质疑,可归为三大类。第一大类是围绕计算主义的争论。有如下三种相互争论的观点:计算主义、二元论和心脑同一论。简言之,计算主义认为,可以将人脑看成数字计算机,因此HI或心灵本质上就是可计算的程序。而按照丘奇-图灵论题,作为一种通用图灵机,数字计算机可以完成一切可能的计算,因此AI是可能的。二元论认为,人脑是能产生主观经验(subjective experience)的独特装置而非数字计算机。心脑同一论则认为,人脑是神经活动的生理器官而非数字计算机,因此(计算主义意义上的)AI是不可能实现的。

第二大类是围绕AI理论框架可行性的争论。常见有如下两种反对意见:来自哥德尔不完备性定理的反驳,来自遵循规则行为或规则约束行为的反驳。第一种反对意见认为,按照哥德尔不完备性定理,在一个一致的算术系统中,至少有一条无法在该系统内部获得证明的定理。但是人类凭直觉可以知道这条无法被证明的定理为真。^[4]因此任何计算机原则上无法实现HI,即AI原则上是不可能的。第二种反对意见则认为,由于人在应对各种复杂或突发的自然现象和社会现象时,常不按某种既定的计划来执行。即使有预先谋划,也常随时调整,甚至也常常凭直觉和经验行事。这就表明,HI具有很强的机动性、直觉性和适应性,并非是某种既定的算法程序。正如德雷福斯(Hubert Dreyfus)所言,人类的很多高级行为(high-level behavior)(如,直觉反应)是不可以被编码的(encoded),是不可能被还原成遵循规则的行为(rule-following behavior)。^[5]可见,任何仅由算法或规则程序约束的机器是不可能做到完全模拟或类似于HI的。因此AI不可能实现。

第三大类是围绕人类高级意识活动的争论。HI除了能体现出推理、计算这些意识活动之外,还能体现出情感、感受质、意向性、自由意志等一些更高级的意识活动。后面这些高级的意识活动涉及主观的维度,而任何关于AI的科学研究显然都只能是客观的活动。例如,查莫斯(David Chalmers)认为存

这种情况其实相当于上述第一种情形,即我们提供不出既适用于人类又适用于K的智力表S的情形。

在两类意识问题：^[6]一类是意识的易问题，另一类是意识的难问题。前者指的是，关于那些能按照其因果角色(causal role)来做功能化解释的特定意识活动(如：学习、推理、回忆、信念等)中的问题。(原则上)认知科学的方法可以处理这类意识问题。后者指的是，不承担任何因果角色，因而不能给出功能化解释也不能用认知科学方法处理的那类意识问题。如：感受质(qualia)难题。如果他的区分是合理的，那么 AI 也是不可能实现的。因为按照查莫斯的区分，AI 研究充其量只能处理意识的易问题。对于像感受质这样的难题，AI 无能为力。又如，塞尔(John Searle)正是基于人类意识活动具有意向性的特点提出了中文屋论证。他认为，“被例示的计算程序仅凭自身永远不能成为意向性(产生)的充分条件”。^{[7]417} 因此，试图仅凭算法程序来模拟 HI 的强 AI 是不可能实现的。

限于篇幅和主旨，不再对上述争论及其各种回应做进一步考察。但是不难看出，以上简要回顾应足以得出如下结论：目前学界围绕 AI 的种种争论尽管角度各不相同，却有一个共同的前提，那就是，至今为止，几乎所有的争论(无论支持或反对)都认为，AI 与 HI 之间存在“可比较关系”(CR)(在本文第二节，我利用一个双条件句对 CR 进行了刻画)。分歧仅在于：支持的一方认为，HI(原则上)是人工可复制的(artificially replicable)，^[3]因而 AI 完全可以接近或超越 HI；反对的一方则认为，由于 HI 具有某些独特性，因而(原则上)AI 无法接近或超越 HI。

四、基于集体人格同一性的论证

基于上一节的考察，现在不难理解，HI 不同于 AI 是这两者之间存在 CR 的一个必要条件。换言之，仅当 HI 有别于 AI，HI 和 AI 之间才有可能具有 CR。因此，我们首先需要将 HI 与 AI 区别开来，才有可能进一步将二者进行比较。但是问题恰恰在于：究竟出于怎样的理由，我们认为 AI 有别于 HI 呢？是由于 AI 是人为的，因而不同于 HI？还是由于 HI 具有某种独特性，而 AI 并不具有，因而二者不同？或者其他理由？

在我看来，不管出于怎样的理由，除非我们可以将人类与具有智能的人造物一劳永逸地区别开，

否则我们就不能有效回答本文的目标问题(人类有可能制造出比人类自身智力水平更高的事物吗？)。注意，这里的“一劳永逸”是关键。因为任何宣称已经或将要制造出高于人类智力水平的造物(K)的个人或机构，都避不开如下一个质疑：出于什么理由，我们可以一劳永逸地认为 K 有别于人类？不难理解，除非这类理由是某种必然为真的知识，否则我们就不可能做到一劳永逸地将 K 与人类区别开来。

那么，我们是否能拥有这样的必然知识呢？在回答这个问题之前，请先设想如下情形：一般来说，如今任何一个智力水平正常的人都会毫无困难地分辨出人与狗。因为人与狗之间的不同不仅是显而易见的，而且也可以找到科学的精确证据(例如，染色体的数目不同或基因图谱的区别)。但是如果将来有一天，人类经历了一场可怕的灾难。例如：某种未知病毒 x 的突然爆发，该病毒危害性大、传染性强而且传播迅速。病毒 x 致使人类身体大面积溃烂并导致迅速死亡，我们根本来不及应对。整个人类种族面临灭绝的危险。然而，我们幸运地发现狗对此病毒具有先天的免疫力。而且，届时的科技已经非常发达，人类已能够将人脑成功移植嫁接到别的哺乳类动物身上。假定，时间紧迫而且为躲避灭顶之灾，我们不得已采用这种技术。于是，经历这场浩劫之后，满街都是狗身人脑的生物。如果这个时候再问：人与狗之间有什么不同？这个问题似乎就很难回答。但是有一点可以肯定：面对此混合体，我们以往拥有的关于人与狗之间差别的所有常识或证据似乎都将失效。

这种可怕的情况也许永远不会降临到人类头上，但无论如何都是可以设想的。而且，上述“人狗混合体”思想实验带给我们的一个直接启示就是，我们一直持有的关于人与狗之间差别的牢靠无疑的知识，其实是偶然真的。同理，我们也可以认为，“K 有别于人类”的知识，也并非是必然真的。因此，如果我们无法一劳永逸地区别开 K 与人类，那么我们就不能宣称 K 与人类之间具有 CR。而如果 K 与人类之间不具有 CR，那么我们将无法判断 AI 是否可以超过 HI。也就是说，人类有可能制造出比人类自身智力水平更高的事物吗？这个问题根本不可能得到明确的回答，因而是一个没有认知意义的问题。

有关 AI 论证的更多详细介绍，可参考玛格丽特·博登^[8]和 Hauser^[3]的有关论述。前者是一个囊括了众多支持和反对 AI 论证的经典论文集，后者则对各种常见论证及最新进展一一进行了梳理、分类和论析。

题。同理,如果K与人类之间不具有CR,那么AI研究的一个重要目标——“生产出一种新的能以HI相似的方式做出反应的智能机器”——也是不可能实现的。为便于理解,可将这里的论证整理如下:

(1) 如果本文目标问题(人工智能能否超越人类智能)是一个有认知意义的问题,那么AI和HI之间具有CR。

(2) 如果AI的重要目标(生产出一种新的能以人类智能相似的方式做出反应的智能机器)可能实现,那么AI和HI之间具有CR。

(3) 如果AI和HI之间具有CR,那么存在“K有别于人类”的必然知识。

(4) 不存在“K有别于人类”的必然知识。

【集体人格同一性难题】

(5) 由(3)(4)可得, AI和HI之间不具有CR。

【Modus Tollens】

因此,

(6) 由(2)(5)可得, AI的重要目标不可能实现。【Modus Tollens】

(7) 由(1)(5)可得, 本文目标问题不是一个有认知意义的问题。【Modus Tollens】

简单分析一下上述论证。首先, 该论证并不打算去质疑AI的种种具体方案是否可行, 而是专门针对有关AI的各种争论中普遍承诺的一个共同前提——AI和HI之间具有CR。其次, 由于“AI和HI之间具有CR”的一个必要条件是, 存在“K有别于人类”的必然知识。但是借助对“人狗混合体”思想实验的分析, 我们有理由相信, 由于集体人格同一性难题的存在, 因而并没有“K有别于人类”的必然知识。最后, 由于本文的目标问题(人工智能能否超越人类智能)是否具有认知意义, 以及AI的重要目标(生产出一种新的能以人类智能相似的方式做出反应的智能机器)是否能实现, 均取决于AI和HI之间是否具有CR。由此可进一步得出结论, 不但本文目标问题并不是一个有认知意义的问题, 而且AI的重要目标也不可能实现。此外, 不难看出, 该论证中的第(4)步很关键: 借助于集体人格同一性难题, 才得以否决存在“K有别于人类”的必然知识。可见, 这也是其有别于其他关于AI论证的一个新颖之处。因此, 我将该论证称之为“基于集体人格同一性的论证”。

下面再回应一下该论证很有可能碰到的两个

质疑:

【质疑一】也许有人会说, 我们完全可从物理-生理层面获得某种关于人类的本质的必然知识。如, 按照当代分子遗传学的说法, 人是具有某种特定基因图谱g的物种。于是我们就可以宣称, 凡是具有基因图谱g的个体都可以称为人, 反之则不是。这样, 我们就获得了关于K有别于人类的一劳永逸的标准, 即K有别于人类的必然为真的知识。尽管这种必然知识不是先天的(a priori), 而是与我们的经验科学的发现有关, 因此可以看做类似于“水是H₂O”这样的后天(a posteriori)必然知识。

【对质疑一的回应】质疑一的关键是, 先区分两种必然知识: 先天必然知识和后天必然知识。然后指出, 我们没有理由不相信, K有别于人类的知识不可以是一种类似于“水是H₂O”这样一类的后天必然知识。但是在我看来, 这个类比是无效的。理由就是, 像水这样的简单无机物, 我们似乎的确能通过揭示其微观物理结构而获得关于其本质的必然知识。然而我们有充足理由相信, 能够展现出智能行为或社会性行为的存在物, 不能仅通过揭示微观结构的方法而获得关于其本质的后天必然知识。基于集体人格同一性论证的关键就在于此。具体而言, “集体人格同一性”, 关注的不是由多个个体按照生物学意义上的某些生物性或物理性共同特征所组成的特定种群(population)或物种(species)——例如, 智人——的人格同一性, 而是由多个个体按照某些伦常性或社会性共同特征所组成的特定集体(collective)的人格同一性。换句话说, 集体人格同一性探讨的是, 究竟出于怎样的理由, 我们应将某个特定集体称为人类呢? 也就是说, 我们不应该出于某种物理-生理层面的原因, 而将K与人类区别开来。因为, 在我看来, 如果K的确具有了和我们相似的智能以及随之而来的复杂智能行为或社会性行为, 那么我们就有理由认为, K和我们应同属于一个集体, 而非分属两个不同的集体。而且, 由于任何集体所具有的社会性特征总会随着特定的社会环境和历史文化背景的变迁而不断变化, 因此, 不管K与人类是否属于同一个集体, 这个或这些集体也都不会具有某种恒定不变的本质特征或在不同时空中均保持同一的社会性特征。

【质疑二】质疑一中的“必然知识”是一种能在所有情况下或在所有可能世界中将K与人类区别

限于篇幅, 暂且处理在我看来最可能出现的两个质疑。

开的知识,但这样的要求或许太强了,可做如下弱化处理:我们也许不能获得 K 有别于人类的必然知识,但是,只要在某些可能的情况下或某些可能世界 w 中,有理由将 K 与人类区别开,那么我们就可以说,至少在这些可能世界 w 中,我们有关于“K 有别于人类”的知识 e。尽管这些知识 e 不是必然的(因为 e 不在所有可能世界都为真),但却是某种起实际效用的偶然知识。因此,在这些可能世界 w 中,“K 的智能是否可以超越 HI”这样的问题仍有可能获得有认知意义的回答。

【对质疑二的回应】做这种模态上的弱化,是否就能(从认识论层面)成功“挽救”上述问题呢?表面上看,这个“弱化”处理似乎可行,甚至还较符合现实情况(因为现实情况下,我们似乎就是这样来区分 K 和人类的)。但是进一步分析将表明,其实这种做法也是行不通的。理由就是,基于上述对质疑一的回应,如果承认并不存在“K 有别于人类”的必然知识,那么我们又凭什么认为可以有某种“K 有别于人类”的偶然知识 e 呢?在我看来,这种偶然知识 e 充其量只能算作某种人为的协议(protocol)或因某些历史文化因素而约定俗成的习俗(convention)。因为我们仍可以进一步追问:做出这类协议或形成这类习俗,究竟源于怎样的深层理由或进一步原因呢?对于这个问题,似乎仅有如下两个可能的回答:要么根本没有深层理由或进一步原因,就是随机约定或形成的(例如,我们对河外星系中某颗恒星的随机命名),要么所谓的深层理由或进一步原因依旧需要依赖或源于某种恒定不变的本质特征。然而,上述针对集体人格同一性的分析业已表明(参见对质疑一的回应),这样的本质性特征并不存在。因此,完全有理由怀疑,这类协议或习俗可以得到任何可靠的辩护(justification)。换句话说,如果 e 仅是一类无法得到可靠辩护的协定或习俗,那么我们就有理由相信,e 没有资格被称为知识。因此可以认为,即使仅在某些可能世界 w 中,“K 的智能是否可以超越 HI”这样的问题,也不能获得具有认知意义的回答。

五、结 语

总结一下,如果基于集体人格同一性的论证是一个有效的论证,那么与本文目标问题密切相关的三个论题已经得到了辩护:(1)“AI 是否能超越 HI”(本文的目标问题),并不是一个纯粹的经验科学问题。因为它免不了要涉及集体人格同一性这个形而上学问题,所以原则上无法仅在科学的框架中获得有效解决。(2)本文的目标问题其实是一个没有认知意义的问题。因为,该目标问题具有认知意义的一个必要条件是,我们能拥有“K 有别于人类”的必然知识。然而基于集体人格同一性的论证表明,该必要条件不可能成立。(3)“生产出一种新的能以 HI 相似的方式做出反应的智能机器”(AI 研究的重要目标)是无法实现的。因为如果智能机器的确具有了和人类相似的智能,并伴随有和人类相似的智能行为或社会性行为,那么似乎就没有理由再认为它们和人类分属两个不同集体,而是应该同属一个集体。因此,严格地讲,AI 研究的这个目标是无法实现的。换句话说,这个所谓的研究目标充其量只能做到增加同一个集体中的个体数量罢了。

不难设想,在下个一百年里,我们的物理-生理外形也许不会经历太大的变化。但是,下一个一千年呢,或下一个一万年呢?按照目前计算机科学和生命科学迅猛发展的趋势来看,等到下个一千年或下一个一万年,我们后代的物理-生理外形又会是怎样?对此,多数人似乎会觉得难以设想。因此,凭什么来确定他们是我们人类的后代,而不是一个全新的物种?反之,又有什么可靠的理由说,他们不是一个全新的物种,而是与我们并没有本质区别,只是他们在其能力的诸多方面(包括智能)都已经远远地超越了我们而已?

参考文献

- [1] 人工智能,引自“百度百科”,URL: <http://baike.baidu.com/subview/2949/5816869.htm>, 2013-11-19.
- [2] 人机大战,引自“互动百科”,URL: <http://www.baike.com/z/c/>

这将涉及关于“什么是知识”的争论。限于篇幅,不便展开。这里需要注意如下两点:(1)由于不可能得到有效的辩护,因此我们有理由认为 e 并不能算作知识;(2)若有人针对上述“有效辩护”提出异议,认为 e 即便不能获得认知层面的有效辩护,也有可能获得某种实践层面或伦常层面的有效辩护,因而 e 可以算作某种实践知识或默会(tacit)知识。但是在我看来,这么做仍旧无济于事。因为,即使 e 可算作实践知识或默会知识,我们真正关心的“K 的智能是否可以超越 HI”这样的问题,也不会因此而获得任何有认知意义的回答。

- AI/index.html?prd=shouye_toutiao , 2013-11-19.
- [3] Hauser, L.. Artificial Intelligence, In International Encyclopedia of Philosophy. <http://www.iep.utm.edu/art-inte/> , 2014-01-12.
- [4] Godel, K.. On Formally Undecidable Propositions of Principia Mathematica and Related Systems [M]. New York: Dove Publishers, 1992.
- [5] Dreyfus, H.. Intelligence without representation - Merleau Ponty's critique of mental representation [J]. Phenomenology and the Cognitive Sciences, 2002, 1(4). Special Issue: Hubert Dreyfus and the Problem of Representation, Jacobson, A.J., Ed., Kluwer Academic Publishers. <http://www.class.uh.edu/cogsci/dreyfus.html>
- [6] Chalmers, D.J.. Facing Up to the Problem of Consciousness[J]. the Journal of Consciousness Studies, 1995, 2(3):200-219.
- [7] Searle, J.. Minds, Brains and Programs [J]. Behavioral and Brain Sciences, 1980, 3(3): 417-457.
- [8] 玛格丽特·博登. 人工智能哲学[M]. 刘西瑞, 王汉琦, 译. 上海译文出版社, 2001.
- [9] Olson, Eric.T.. Personal Identity, in Stanford Encyclopedia of Philosophy. <http://plato.stanford.edu/entries/identity-personal/>, 2014-01-12.
- [10] Russell, S. J., Norvig, P.. 人工智能: 一种现代方法[M]. 第3版. 北京: 清华大学出版社, 2011.

Can Artificial Intelligence Surpass Human Intelligence?

WANG Xiao-yang

(Department of Philosophy, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: More than half a century, although artificial intelligence research seems to develop rapidly, the criticisms never disappear. In this essay, I will discuss a question concerning the foundation of artificial intelligence(AI). The question is “can AI surpass human intelligence at all”? Based on collective personal identity, I will successively argue for the three arguments related to the question. Firstly, the question is actually a philosophical question, so it cannot be addressed in the framework of empirical science. Secondly, we cannot give any cognitively meaningful answer to the question. Finally, if any answer is cognitively meaningless, then we have reason to believe that it is impossible to produce any intelligent machine which can approach the human level.

Key words: artificial intelligence; human intelligence; collective personal identity; metaphysics; philosophy of artificial intelligence

(本文责任编辑: 费多益)