

智能机器外观形象的伦理问题

——从拟人化到环境假设的道德图像

The Moral Problems of Intelligent Machines' Appearance:

Transforming the Anthropomorphic Moral Self-Image to the Moral Image of the Environmental Hypothesis

张正清 /ZHANG Zhengqing

(北京航空航天大学马克思主义学院, 北京, 100191)
(School of Marxism, Beihang University, Beijing, 100191)

摘要: 智能机器的外观设计存在拟人化倾向, 拟人化设计背后是一种对自我的内在假设, 认为智能机器想要合乎道德共同体的道德标准, 就必须具有与共同体内部成员相似的道德图像。本文认为这种对道德图像的认识导致了一种自我谬误。这种错误理解会在个体层面导致道德身份同一的问题, 在群体层面产生中心化的问题。而自我谬误反映到智能机器的道德自我上, 表现为自我与他者混淆、绝对对象化和道德图像的不对称; 反映到外观设计上就是内在假设的自我主导机器的道德图像。而环境假设的自我可以克服自我谬误带来的问题, 其设计上主张客体图像的独立性、它者行为的差异性和互动关系的去道德化, 从而构建一种承认它者性道德的设计路径。

关键词: 智能机器 道德图像 设计伦理

Abstract: The appearance design of intelligent machines has the trend to be more anthropomorphic than before, in order to produce more affinity for machines to be accepted by their users. The inherent hypothesis of self behind this anthropomorphism claims that only the moral image similar to that of the members of moral communities can secure the appropriate performance of a moral-self of machines. This paper considers this moral-self as a fallacy of self. It creates the problem of moral identity at the individual level, and the problem of centralization at the group level. For the hypothetical moral-self of intelligent machines, the fallacy of self causes the mixture of self and others, the absolute objectification, and the asymmetry of the moral image. For the design of appearance, the fallacy led to a moral image of machines dominated by the inherent hypothesis of self. The self of environmental hypotheses is a way to overcome the fallacy of self. In design, it proposes the independence of others' image, the dissimilarity of others' action and the de-moralization of interrelationships. Based on that, therefore, we can establish a design approach that admits the morality of otherness.

Key Words: Intelligent machine; Moral image; Design ethics

中图分类号: N0 文献标识码: A DOI: 10.15994/j.1000-0763.2019.10.004

在关怀机器人或治疗机器人的设计中, 其外观或互动方式常以儿童或者动物的形象出现, 其外形偏向于球体或者毛绒、柔软的设计。表面上看来, 这是为了使得用户自然地产生亲近感。但实际上, 其蕴含了人对智能机器这种技术物的道德图像的思考。

一、智能机器的道德图像

社会机器人 (social robot) 是目前与普通人在生活中产生最多交际的智能机器, 而具有治疗或者陪伴功能的关怀机器人是其中最典型的一类。

基金项目: 国家社科基金一般项目“‘后经验转向’时代的技术伦理规约机制及模式研究”(项目编号: 13BZX024)。

收稿日期: 2018年11月4日

作者简介: 张正清 (1989-) 男, 河北辛集人, 北京航空航天大学马克思主义学院讲师, 研究方向为科学技术哲学与技术伦理。

Email: 290563434@qq.com

关怀机器人的伦理问题近年来得到普遍关注，首先是因为来机器人学的飞速发展，智能机器人本身带来的伦理问题得到了人们的重视；其次是由于老龄化的加速，使得关怀机器人成为一种可以预见的社会必需品。^{[1], [2]}其中主要的伦理问题有：1. 隐私问题。机器人在与人进行互动时需要采集数据加以分析，不论是对人的直接摄像，还是行为模式的数据化都会带来隐私问题。2. 减少人与人之间的交往并影响社交质量。关怀机器人会强化与使用者之间的关系，减少使用者与其他社会成员之间的交往时间，并且关怀机器人本身的陪伴质量也受到人们的质疑。3. 改变了照料本身的含义。从效率出发，照料的提供者与照料的接受者之间需要通过机器中介来进行，由此如何有效的利用关怀机器人，是使用者与设计者双方共同的问题。^[3]

基于以上的伦理问题，有些学者认为现在的社会机器人存在伦理缺陷。例如在医疗机器人的使用中，如果病人真的认为关怀机器人是在照料他们，那么他们就会处于风险之中，甚至会威胁到病人的尊严和自主性，造成感情上的伤害。^[4]人之所以相信机器人是在照料他们，是因为机器人具有某种机器人其实并不具有的属性。正是由于这种错误的信念，人才会有意无意地感受到机器人的关怀与爱，并认为自己从中受益。^[5]而有些学者支持关怀机器人的使用，认为我们在深入思考人与机器的关系时，也必须清除对人与人之间关系的成见。^[6]就像机器人将要大规模应用的今天一样，人与人的现代性关系也经历过一个大规模的推广、应用时期，并且这种非传统关系一直延续到今天。

在这些伦理争论的背后，有一个看起来表面化、但却事关人机互动本质的问题，那就是外观设计问题。在面对电子通讯的图像或者机器人时，我们更为注重其外观所呈现的东西。^[7]所谓外观，不仅指其静态的形象，还指其行为的模式，总的来说就是机器人的外在表现，包括外形、材质、行为举止，甚至表达的逻辑。我们之所以觉得机器的行为可以在伦理维度讨论，一个重要的原因就是我们把机器进行了拟人化 (anthropomorphism) 的处理。我们只有觉得是人，而不是机器在照顾我们，才会从“关怀”的经典意义上来考虑这个问题。因此，拟人化的外观设计就是把机器人的外观形象设计得无限接近于人。雪莉·特克

尔 (Sherry Turkle)、马蒂亚斯·舒尔茨 (Matthias Scheutz) 等人指出把机器人拟人化并在此基础上建立人机关系，是对人与人之间关系的破坏。([4], p.173) 但是现实情况是，机器人已经在众多领域成为了社会行动的一分子，人不可避免地与人建立了亲密关系。彼得·辛格 (Peter Singer) 甚至指出在一些战地故事中，士兵会冒着生命危险去拯救机器人。([4], p.175) 在这些事例中，拟人化通过起名字等表达接纳的行为而实现，人们会以一些欢迎仪式和准入方式来赋予一个纯粹技术物以同伴的身份。

因此，如果一个关怀机器人在拟人化上非常成功，我们似乎就可以接受其为社会的一员，并建立亲密关系；而如果拟人化不成功，我们会认为这个机器人不能承担与社会关系相关的任务。在伦理上存在这样一个逻辑：外观上越拟人，我们就容易将其纳入到道德的考虑范围内；而外观与人差距越大，我们越不会考虑其道德方面的问题。这个逻辑被应用到社会机器人设计之中，导致所有社会机器人的外观设计，都倾向于具有某种拟人的、使人亲近的特点。实际上这是我们在根据自己的道德图像要求，来构建自己的人工物。

这种设计理念的来源是人对自我道德图像的认识。不论一个人的行为是否经常符合道德要求，这个人自己都希望自己的形象是正面的、道德的。这就是个人的道德图像 (moral self-image, MSI)，道德图像提供了一种道德向心力，使人为了竭力维持自我的道德图像而努力。^[8]同样，人在与他者交往的过程中，也在理解、感知与期望着他者的道德图像。想要把机器人纳入到社会交往活动中，就需要机器人自己的道德图像。这不只是为了塑造其本身的道德特性，更是为了满足我们自身认知、感受它者道德图像的需求。

那么智能机器人的道德图像是由什么构成的？我们如何构建作为一个独特道德能动者的自我图像，又是怎样帮助其他所谓的非道德能动者构建既属于它们、但又只为我们而存在的道德图像？只有考虑了这些问题之后，才能从根本上回答拟人化设计所带来的伦理问题。

二、智能机器人外观设计中的自我谬误

想要讨论智能机器人的道德图像，就要先梳理

人的道德图像。自我位居于道德图像的中心,有什么样的道德自我,就有什么样的道德身份。道德身份直接构成道德图像,人的道德身份与其所处的道德共同体(moral community)息息相关。而道德共同体的排他性使得人的道德图像在具有朝向自我的向心力的同时,也具有排斥其他非共同体内部成员的排他倾向,这实际上这是一种自我谬误的体现。

1. 道德自我

道德准则常常具有很大范围的普遍性,是一种需要共同遵守的规定。例如康德就认为道德准则应该在本质上是普遍有效的,否则就不能算是真正的道德准则。但是关于个人化准则的讨论也没有停止过,从古希腊开始的道德讨论很多都关注个人道德选择,例如斯多葛学派。西塞罗就说,“只有卡托这个人的自杀是道德的”。^[9]这种对伦理准则的认识是基于独有的自我观念。所以说,个人化准则的根据在于第一人称的观点,即从自我的经验与身体知觉出发。([9], p.37) 这时自我可以分解为两个自我思考的方式,一个是从个性(individuality)出发,讨论自我意识与主体性;另一个是从心理功能的统合性出发,讨论人际与社会的参与能力、对象性理解。([9], p.37) 第一种自我是我们常说的作为道德能动者的自我,第二种自我是一种共同性自我,强调社会成员之间在伦理取向上的共通之处。但这种共同性是一种假设,因为只有第一种自我才是具体讨论自我内部,而第二种自我只是基于自我的存在而直接进行道德活动的功能性假设。这种共同性假设的核心就是,我们相信自己具有某种本质,这种本质与其他主体的本质之间有某种共同性。我们之所有觉得社会中的他人或者他者具有某种道德上的可讨论性,或者直白地说,值得我们考虑,就是因为我们认为其具有某种共通性的个人品质,某种被共同体所接受的“个人塑造的品质(the person-making qualities)”。^[10]

道德自我与道德目的通常是分开的。道德向心力(moral centrality)是考察自我与动机关系时的一个重要概念。并不是因为特定的道德目的驱动,人就可以做出符合自我需求的道德行为。^[11] 分层解释理论(Construal Level Theory, CLT)认为对行为原因不同层面的解释,可以模拟出非理性的自我目的设定。^[12] 不论如何精细化地分析道

德目的,道德目的并不直接构成道德图像,只有道德向心力才是直接原因。道德向心力不为实现直接的、特定的道德行为而存在,而是为了形成共同性假设下的群体内在一致性而存在。从个人层面讲,这种向心力形成了道德身份;而从群体层面讲,这种向心力形成了道德共同体。

道德身份是自我的独特性所在,在之前的哲学讨论中,身体上的延续性与一致性是身份同一的核心。而2000年之后,人们开始把心理上的延续性作为身份同一的重要方面。^[13] 人希望在与他者的互动中不断强化自我观念,这时对他者的经验就是对一种“自我对象(self object)”的经验。^[14] 要想在机器身上获得如此强大的自我认同,就需要机器有极高的拟真度,即人越是能在机器身上获得自我经验,机器就越接近于人,机器就越接近道德能动者。这是人工道德能动者讨论中一个重要的错误观念,称之为自我谬误。自我谬误即体现了道德身份同一性的问题,又中心主义的问题。

2. 中心主义的道德共同体

伦理学家也试图让伦理包含他者,但是人类中心主义的,或者在道德认知意义上的自我中心主义的方法论框架,在处理他人、动物、环境、机器等他者时,采取了一种还原论的态度,强行把他者通过类比或曲解的方式,纳入到传统的自我中心主义框架中。^[10] 这表面上扩大了伦理思考的范围,实际上仍是一种只把自我属性作为可考虑内容的伦理倾向。斯拉沃热·齐泽克(Slavoj Žižek)把这种趋势描述为:“他者性的本质,就是另一种共同性,一种展现我们要控制他者的决心的结构”。([15], p.161) 这种结构的现实表现就是道德共同体。

道德共同体是中心化的具体呈现,这种中心化是超越所谓的人类中心主义的。共同体的成员没有区别人与非人,更没有许诺把所有人类都纳入到共同体。共同体的标准更具有地方性,可以区别地看待不同道德共同体之间的文化、社会差异,而不像是人类中心主义那样预设了一个抽象的人类共同体。人类中心主义在行使歧视性标准的同时,本身就是一种对人类内部的歧视。历史上西方社会就一度认为白人男性是道德共同体的成员,而女性和非白人不是。^[16] 这不是一种理性的标准,而是感性的成见。这种划分与我们前面提到的排他性过程一

致，但是区别在于，排他性过程遵循一种内在的理性逻辑，其可以历史地看成是人类道德知识演化的过程；而道德共同体成员标准的变化，没有严格遵循这个演化过程，它可以是按照非理性的标准，例如种族主义；也可以是按照理性的标准，例如是否具有自我意识与理性等。但是，两者的共同之处都在于，其有排他性倾向，并保持总是有标准之外的它者存在，而这些它者因为共同体或中心主体的存在而没有道德地位。

排他性道德共同体的存在本身就是挑战每个成员内在的道德性。每一次道德共同体的扩大，都是对原来成员自身的道德本质的质疑，迫使共同体内部的每个人思考这样两个问题：1. 共同体的道德标准。这个道德标准是在下降还是上升？是我们接纳了新成员，还是被接纳为新团体？我们是修改了自己的道德标准，还是重新认识了自己的道德本质？不论理论上应该如何回应，但是事实上，道德共同体扩大的过程，也是道德等级形成的过程；加入道德共同体越晚，在道德上的地位越低。2. 共同成员的道德本质。为什么原来只有这样的主体有资格被道德地考虑，而现在那样的主体却获得了同样的资格？新成员与旧成员之间在道德本质上有什么异同？新成员在自我层面是否还具有同样朝向的向心力？现如今，道德本质问题直接导致了我们的未进入的道德主体，以及刚进入的道德主体仍持有怀疑态度。我们认为这些主体的道德能力只是表征层面或者功能上成立，实际上其仍然没有从其自我层面拥有成为共同体一员的资格。这两个问题构成了中心论意义上的自我谬误。

3. 自我谬误带来的人工物道德困境

综上所述，自我谬误的现实问题，就是我们依靠旧共同体的道德标准，来考虑内部成员的道德自我，以及外部它者的道德自我。因此，自我谬误会非常直接地受到三个挑战：

(1) 人的自我同一问题

古典哲学认为，创造人工物的活动就是自我见之于自我的活动。对这个观点的误读导致人们认为，人工物就是一个自我客体。这忽视了黑格尔等古典哲学家们普遍具有的“过程”思想。创造活动是一种见之于自我的活动，但人工物不是见之于自我的客体。首先，创造活动是人能动性的反映过程，在对象性中断时，人全心全意地投

入到自己延展自我的活动中，是一种见之于自我的过程。然后，当人工物形成时，物质性与对象性重新回归，人开始审视这个异己之物与异己之自我，它者正式形成。人工物在这个意义上与自然物并没有本质的区别，物质性与对象性带来的异己感是其核心。因此不能说人工物是一个自我客体，而只能说人工物的诞生过程是出于人见之于自我的需要。

(2) 控制问题

人真的能从人工物或自然物之中加强自我观念吗？这种尝试一直存在，但是效果欠佳，因为人类历史上的强化自我概念的行为都是从社会关系来实现的。是主客、统治与被统治等传统关系的不断强化，使人能从控制他者的活动中获得自我的强化。控制，而非创造，才是经典自我观念的外化过程。旧的伦理原则也是基于此来建立的，所有的传统伦理原则都强调人与人之间的关系，并且是倾向于维持关系而不是破坏关系。这种设置在今天视为先验的、理所应当的，其根本来源就是自我的控制需求。但是现代性开始以后，人类创造的过程依旧被理解为控制的过程，导致了用控制的伦理来理解创造的伦理，必然面临问题。而人工物即便在今天取得了巨大突破，也在因为控制的伦理而不断被束缚与误解。

(3) 不确定性问题

确定性是哲学的一大追求，直到20世纪，杜威的实用主义还在追求技术社会中的确定性。确定性是与追求自我的观念相联系的，而不确定性是与追求他者的观念相联系的。确定性最后的丧失，与人们转而追求非结构的、后现代的自我观念，预示了他者伦理将要取代自我伦理。而人工物伦理作为一种过渡，既是自我外化的表现，又是他者性的载体。需要先建立人工物伦理，再建立他者伦理，从而扭转中心论、自我论的伦理。因此，人工物也不应该在自我伦理的意义上是一种道德能动者。因为能动性作为自我道德能力的一种表现，对于道德他者性没有贡献。即便自我之外的它者可以在传统意义上成为某种程度的能动者，也只是在自我的序列中争取到了一定的地位，远获得他者应有的伦理地位。

三、自我的环境假设对自我谬误的反思

道德自我与个人、记忆和欲望都有直接的关系,甚至与未反思、加工的认知体验和知觉也有关联。但最主要的是,道德自我与一个在社会环境中所处的位置有紧密联系。^[13]很多学者认为,机器目前没有意识与自我,因此也没有道德状态,也就没有自我道德图像。这种观点是站着一种道德能动性的自我视角之上,认为自我观念与意识功能是道德个体的基础。这种观点存在两个问题,首先个体概念不以个体能力为基础,我们承认个体存在是以异己性为根据,而不是以其能力大小以及与我们的相似程度为依据。对于植物人等类型的病人,我们依然承认其个体性与生存资格;对于具体的植被,甚至抽象的生态圈层,现在环境伦理都承认其道德地位。显然以与人相同的能力作为道德可考虑性的根据,已经过时。其次,道德的概念不以自我为基础,而是以互动和关系为基础。我们考察具体情境中每个能动者对共同体规则的遵守程度,而不是考察每个能动者的内部道德状态。

如果经典的道德个体观念存在问题,那么对于机器等人工物的道德能动性就会有新的解释。目前智能机器还没有自我观念和类人的意识能力,但是机器在被使用的过程中处于一种互动关系中。对于作为使用者的我们,需要从它者的视角,而不是自我的视角,来区分它者的两种道德自我。一种是“内在假设的自我”,即基于他心内部的道德状态来生成的道德身份,来构建和维持的他者的道德图像。另一种是“环境假设的自我”,即自我的道德形象体现在具体的互动环节中,我们会依靠肢体动作、表情、话语等因素来理解与构建他人的道德图像。对于“环境假设的自我”的疑问在于,人可以抛开内部状态来单纯的讨论意义环境吗?其实,互动中的道德图像可以不基于内在状态。梅洛庞蒂就认为在主体互动关系中,内在状态并不扮演主要角色。梅洛庞蒂没有否定互动关系就有内在性,但是他在对话这种具体、精细的互动环境的限制条件下指出,所谓的互动意义,只来自于互动关系本身,而并不来自双方的主体状态。^[17]

拟人化作为人为智能机器构建道德图像的常见方式,恰恰是假设了机器具有某种内在的自我。这看起来是矛盾的,因为机器没有内在自我与意识能力,为什么还要遵从内在自我的假设来设计

其外观呢?这正是因为我们只承认存在内在假设的自我,所以机器、尤其是智能机器的设计,正是朝着让机器具有内在自我的方向前进。如果一个道德图像的背后是内在假设的自我,那么全方位的、不分次序地为其赋予自我道德能力,也是我们现在努力的方向。因为我们想让机器进入我们的排他性共同体,而唯一的路径就是让它者获得我们自身道德图像与道德自我所展现的内容。

所以,人在维持自己作为道德共同体成员的图像时,也在极力维持互动型机器人的拟人形象。拟人既包括趋近于人的形象或审美,也指对人的感受性的亲近。由于智能机器没有内在假设的自我,所以正确的途径应该是从环境假设的自我来考虑。但是目前的智能机器的设计者并没有按照环境假设的思路来构建机器的道德图像,仍旧遵循拟人化的“内在假设的自我”,这就比自我谬误更进了一步,会带来一系列道德图像的问题:

首先,混淆了自我与他者,导致一种自我中心论的设计缺陷。模糊、封闭的自我标准会影响道德沟通,使得他者的道德身份对互动关系的塑造无法彰显,导致实际互动中的道德沟通障碍。在承认异己的物质性存在的基础上,先假设地把机器作为道德共同体的一员,来设想其形象,然后又否认其内外一致性,最终会产生一般化的自我道德图像与设计他者道德图像之间的矛盾。

其次,绝对对象化的他者道德图像,导致自我的道德图像也不能获得正确的地位。本来对象化是绝对道德中立的,因为对象化就是主体外化自身,为自己创造客体的过程。^[18]但是从女性主义反对物化开始,对象化在道德上就开始偏向贬义。对象化揭示了现代性意义上自我与他者的关系,即他者就是披着主体性面纱的客体,或者说有着客体性本质的类自我、类主体(quasi-subject)。这是一种强调自我、虚化他者的理论,给人机互动带来本体论的问题。

最后,构建与维持道德图像之间的不对称。如果共同体不承认智能机器的道德地位,就会导致自我道德图像的构建与维持之间的不对称。没有绝对独立存在的、理论上的自我道德图像,自我图像是在自我的主体性与赋予他者的主体性之间维持着图像身份的统一。如果抛开这个自我与他者之间的关系,道德身份同一就面临困难。智能机器的道德图像就只能在认知上具有可理解性,

但是在身份与地位上无法取得认同。

四、从漠视到舒适：它者的道德图像

恐怖谷 (The uncanny valley) 指出了在机器人在拟人过程中出现的使人感到不适的阶段。([19], p.25) 机器人的外观会随着不断接近于人而使人感到亲近感, 在当时的日本, 外观的含义更为单调, 只是其形象与行为是否与人相似。即便在这个情况下, 在非常接近人的外表时, 我们仍然会把注意力从相似之处转向细微的不同之处, 从而产生巨大的诡异感与恐怖感。可以想象在今天, 假如机器人除了在形象上相似之外, 还在行为举止、甚至表达上与人相近的话, 也会相应地出现道德图像的恐怖谷, 我们会对机器产生道德上的不信任感与交往上的排斥感。这些直观上的感受都是因为我们希望把机器纳入到道德共同体之中, 但是自我道德图像上的不同一又无法克服拟人所带来的差异感。这种嫌隙实际上也发生在每次共同体扩大时所纳入的新成员与旧成员之间, 只是面向人工物的这一次跨越尤为使得共同体成员感到不适。

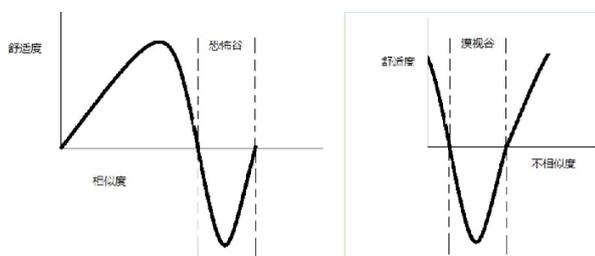


图1 恐怖谷与漠视谷

要克服这种恐怖谷带来的不适感, 按照内在假设的自我, 就要进一步提高机器的拟人化程度, 并在内在自我的层面进行突破。且不说内在假设本身的问题, 就算在突破机器自我的技术层面, 对目前的技术水平来说也是不现实的。因此, 拟人化的、强调自我道德图像的路径是走不通的。机器作为社会互动关系中一员, 首先是一个独立的它者, 其次才是可能被旧的道德共同体吸纳的潜在成员。我们不能本末倒置, 认为只有成为共同体成员, 才是道德上独立的个体。所以智能机器人的外观设计首先应该是按照一个它者的道德图像来设计, 而不是按照一个与人同一的道德图像来设计。不相似才是设计的核心, 而不是相似。

现在社会中应用的社会机器人, 大多数以拟

人为设计初衷, 如果去拟人化, 必然会带来舒适度的下降 (见图1中漠视谷)。这种舒适度下降带来的是差异感, 由于人只关心趋向于中心的道德主体, 因此去拟人化带来的远离中心的道德图像会使人产生淡漠感。就像沙漠中的沙子一样, 变成人毫不关心的客体存在。但是, 如果坚持这种去拟人化的路径, 同时保持社会机器人的互动角色, 最终人们会在漠视它者的中心感中走出来, 正视这种作为它者的、保持与共同体中道德图像的绝对差异的、却又如实地在道德活动中发挥作用的机器。此时, 因为道德丰度的提高, 舒适度又会上升。基于这种逻辑, 关怀机器人的设计需要遵循如下三个原则:

第一, 客体图像的独立性。作为他者出现的机器的道德图像, 不应该成为我们共同性道德图像中的新成员; 或者说, 不要以我们自己道德共同体的准入标准来要求机器, 至少使其处于一个道德性不被排他性共同体评判的位置。这样使得机器的伦理问题是一个机器伦理问题, 而不是人伦理问题。

第二, 它者行为的差异性。要使得机器的行为具有机器的特点而不是无限趋近于人的行为特点, 从而让我们可以注意到一个异己的存在。这样不只是为了提示潜在的道德伤害, 更是为了避免人的道德的机器化, 或者机器行为的人化。只有这样才能保证机器的道德地位与互动中的身份同一。

第三, 互动关系的去道德化。人机关系具有社会性特征, 这是智能机器与人互动的表征特点。^[20]但是这种社会性不意味着人与机器之间存在道德关系。按照固有的道德共同体认定方式, 内部成员与共同体之外的他者之间是没有道德关系的, 但是拟人设计违反了这一点, 因此首先要使得共同体的标准可以维持自身的一致性。而接下来, 我们要反对道德共同体的概念, 强调一种一般性的、包含道德它者与道德自我的道德共在标准, 从而使得互动关系获得新的道德性。

总的来说, 社会机器人的设计应该遵循人工物本来的外观, 而不是拟人的外观。这会产生另一个问题: 关怀、陪伴机器人本来就是为补足人类劳动力不足而产生的, 如果外观上去人化, 是否与其初衷相悖? 面对老龄化社会, 我们确实面临从事照料活动的劳动力不足的问题。机器人代

替人从事服务行业,也确实从一定程度上缓解了这个问题。而事实是,每个接受关怀机器人的老人,也并没有真的把机器人当做人;而我们更不能认为有了机器人,对老龄人口的人文关怀就可以变成机器关怀。我们既要满足对照料活动的需求,也要如实地区别人与机器这两种不同的照料服务提供者。假设机器是共同体的一员,或者强行为机器塑造一个类人的自我,都是自欺欺人的行为。更重要的是,人有自己的道德地位,需要共同体内部成员的关怀;机器也有自己的道德地位,也需要作为它者而被尊重。

总之,智能机器的道德图像既不能因为一种中心化的共同体准入标准而创造,也不能为了服务共同体成员的便利,而假设其具有类人的道德自我。智能机器的道德个体性不因其是否属于某个共同体而存在,机器的道德性也不依赖人赋予其的内在假设的道德自我而存在。只有把机器放入道德互动的社会语境中,从与其他社会成员的互动性上来看,才能正视机器的道德地位。

[参考文献]

- [1] Borenstein, J., Pearson, Y. 'Robot Caregivers: Harbingers of Expanded Freedom For All?'[J]. *Ethics & Information Technology*, 2010, 12(3): 277-288.
- [2] Nagai, Y., Tanioka, T., Fuji, S., et al. 'Needs and Challenges of Care Robots in Nursing Care Setting: A Literature Review'[A], *International Conference on Natural Language Processing and Knowledge Engineering*[C], IEEE, 2010, 1-4.
- [3] Royakkers, L., Est, R. V. *Just Ordinary Robots: Automation from Love to War*[M]. Boca Raton: CRC Press, 2015.
- [4] Lin, P., Abney, K., Jenkins, R. *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*[M]. London: Oxford University Press, 2017, 97.
- [5] Sharkey, A., Sharkey, N. 'Granny and the Robots: Ethical Issues in Robot Care for the Elderly'[J]. *Ethics & Information Technology*, 2012, 14(1): 27-40.
- [6] Coeckelbergh, M., Brey, P., Briggie, A. 'Care Robots, Virtual Virtue, and the Best Possible Life'[A], Spence, E., Brey, P., Briggie, A., Spence, E. (Eds) *The Good Life in a Technological Age*[C], New York: Routledge, 2012, 281-293.
- [7] Pols, J. 'Wonderful Webcams: About Active Gazes and Invisible Technologies'[J]. *Science Technology & Human Values*, 2011, 36(4): 451-473.
- [8] Jordan, J., Leliveld, M. C., Tenbrunsel, A. E. 'The Moral Self-image Scale: Measuring and Understanding the Malleability of the Moral Self'[J]. *Frontiers in Psychology*, 2015, 6(1878): 1878.
- [9] Remes, P., Sihvola, J. *Ancient Philosophy of the Self*[M]. Dordrecht: Springer Netherlands, 2008, 37.
- [10] Gunkel, D. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*[M]. Cambridge, Massachusetts: The MIT Press, 2012, 64.
- [11] Krettenauer, T. 'The Dual Moral Self: Moral Centrality and Internal Moral Motivation'[J]. *Journal of Genetic Psychology*, 2011, 172(4): 309-328.
- [12] Bello, P., Bringsjord, S. 'On How to Build a Moral Machine'[J]. *Topoi*, 2013, 32(2): 251-266.
- [13] Strohminger, N., Nichols, S. 'The Essential Moral Self'[J]. *Cognition*, 2014, 131(1): 159-171.
- [14] Anderson, M., Anderson, S. L. *Machine Ethics*[M]. Cambridge: Cambridge University Press, 2011.
- [15] Žižek, S. *The Plague of Fantasies*[M]. New York: Verso, 1997, 20.
- [16] Neely, E. L. 'Machines and the Moral Community'[J]. *Philosophy & Technology*, 2014, 27(1): 97-111.
- [17] Coeckelbergh, M. 'The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics'[J]. *Philosophy & Technology*, 2014, 27(1): 61-77.
- [18] Lucas, P. *Ethics and Self-knowledge Respect for Self-Interpreting Agents*[M]. Dordrecht: Springer Netherlands, 2011, 1.
- [19] Dumouchel, P., Damiano, L., DeBevoise, M. *Living with Robots*[M]. Cambridge, Massachusetts: Harvard University Press, 2017, 25.
- [20] 陈巍、赵翥. 社会机器人何以可能? ——朝向一种具身卷入的人工智能设计[J]. *自然辩证法通讯*, 2018, 40(1): 17-26.

[责任编辑 李斌 赵超]