

文章编号: 1000-8934(2021)010-0055-06

基于情境体验的社交机器人伦理: 从“欺骗”到“向善”

王亮

(西安交通大学 马克思主义学院 西安 710049)

DOI:10.19484/j.cnki.1000-8934.2021.10.010

摘要: 社交机器人“欺骗”伦理问题是机器人伦理学研究的热点,但传统的研究大多是基于实体属性观的视角,通过分析社交机器人的“拟人化”属性而得出其会产生“欺骗”伦理风险的判断。科克尔伯格则立足于人机交互的情境体验,悬置了社交机器人的“拟人化”属性,进而将社交机器人的“欺骗”伦理转向了“人类善的伦理”。

关键词: 社交机器人; 情境体验; 欺骗; 向善

中图分类号: N031 文献标识码: A

“拟人化”是社交机器人的重要属性,但它包含着一种“真实”与“虚假”的冲突,由此而导致了一些学者对社交机器人会产生“欺骗”伦理风险问题的担忧。如果我们不从本体论的角度来深究社交机器人的“真实”与否,而是关注人机交互过程中人类的体验,我们就会对“欺骗”有一种新的理解。在此,本文通过引入科克尔伯格的情境体验视角来重新解读社交机器人“欺骗”伦理问题。

一、社交机器人的“欺骗”伦理问题

1. “欺骗”的源起: 社交机器人的“拟人化”属性

随着社交机器人的普及,其所引发的伦理风险也受到越来越多的关注,其中一个重要的关注点就是社交机器人的“欺骗”伦理问题。社交机器人,顾名思义,是指能和人类进行交流、互动的自主机器人,“拟人化”是社交机器人的重要属性。来自麻省理工学院(MIT)媒体实验室的布雷泽尔(Cynthia L. Breazeal)认为,“社交机器人能够以人的方式与我们交流和互动,理解我们,甚至与我们建立关系。它是一个具有类人社交智能的机器人。我们与它互动,就好像它是一个人,甚至是一个朋友。”⁽¹⁾同样来自MIT的达菲(Brian R. Duffy)也认为,“机器人能够与人进行有意义的社会交往,这本身就要

求机器人具有一定程度的拟人化或类人属性,无论是在形式上还是在行为上,或者两者兼而有之。”⁽²⁾¹⁷⁸事实上,从目前市面上所流行的社交机器人来看,它们不仅具有人形的外表,而且还可以通过“言语”“面部表情”“肢体语言”等来模拟人的情感。⁽³⁾可以预测,随着人工智能技术的发展,社交机器人的“拟人化”属性会越来越凸显。不仅如此,达菲还认为,“当今许多机器人专家的终极目标是制造一个完全拟人化的合成器。”⁽²⁾¹⁷⁷然而,当人类正为机器人的“完全拟人化”而付出努力,并为之自豪时,“拟人化”所带来的伦理风险也不应被低估。来自都柏林大学智能实验室的扎韦斯卡(Karolina Zawieska)一针见血地指出了问题所在,她认为,“拟人化的核心是幻觉,因此,社交机器人的一个关键伦理问题就是欺骗。”⁽⁴⁾来自英国谢菲尔德大学的阿曼达·夏基(Amanda Sharkey)和诺埃尔·夏基(Noel Sharkey)也认为,“设计机器人来鼓励拟人化的属性可能被视为一种不道德的欺骗行为。”⁽⁵⁾³⁵

为什么社交机器人的“拟人化”属性能产生“欺骗”伦理问题呢?这需要对“欺骗”的理解说起。斯坦福哲学百科全书将欺骗定义为,“使人相信虚假的东西”⁽⁶⁾。可见,对于被欺骗者而言,所面向的事物或者其属性具有“虚假”性,这是欺骗得以产生的根源。斯派洛(Robert Sparrow)通过比较机器人与生物体之间的区别,指出了机器人的“虚假”性,

收稿日期: 2021-3-21

基金项目: 国家社会科学基金青年项目“跨文化视角下人工智能的伦理嵌入机制研究”(19CZX018)。

作者简介: 王亮(1985—),湖北黄冈人,哲学博士,西安交通大学马克思主义学院副教授、硕士生导师,主要研究方向: 马克思主义理论、科技伦理等。

他认为,“机器人没有任何感觉或体验。”⁽⁷⁾³¹³来自维也纳大学的科克尔伯格(Mark Coeckelbergh)也从三个方面总结了情感机器人的“不真实”性,“1. 情感机器人企图用他们的‘情感’来欺骗。2. 机器人的情感是不真实的。3. 情感机器人假装是一种实体,但它们不是。”⁽⁸⁾³⁸⁹由此可见,无论机器人的“拟人化”程度多高,和人类自身相比,其仍然是虚假的和不真实的,而这无疑和欺骗定义中的“虚假的东西”高度契合。但只有“虚假”还不足以将“欺骗”上升到道德伦理的高度。卡森(Thomas L. Carson)认为,“欺骗需要某种意图来导致他人产生错误的信念。”⁽⁹⁾⁴⁷“欺骗在道德上的显著特征是它涉及到故意地使他人产生错误的信念。”⁽⁹⁾⁴⁹很显然,社交机器人必须要“故意地”隐瞒自己“情感”“语言”“构件”等方面的虚假性,才能产生“欺骗”伦理问题。然而,由集成电路、元器件等组合而成的社交机器人本身并没有任何“意图”,但是,它们“拟人化”的属性却自带“意图”,即诱导人相信它们看起来更像人的意图。扎韦斯卡等人认为机器人本质上只是一种机器,“它利用人类拟人化的倾向来制造生命的幻觉,而不是变得有生命。”⁽¹⁰⁾⁷⁹可见,对“人类(将物体)拟人化的倾向”的利用就是一种“故意”,它所指向的就是“制造生命的幻觉”的欺骗“意图”。因此,结合欺骗的定义以及社交机器人的“欺骗”意图来看,确实存在社交机器人的“欺骗”伦理问题。那么,这些伦理问题主要是如何体现的呢?

2. 社交机器人“欺骗”引起的伦理风险

需要注意的是,当我们在谈论社交机器人的“欺骗”行为时,并不是说它们在某一个事件中欺骗我们,而是强调它们的拟人化特征所引起的人类的“自欺欺人”式的行为。固然“自欺欺人”和人自身十分相关,但结合上述的讨论,也可以将其归因至社交机器人,“因为它预设或鼓励了这种欺骗”⁽⁷⁾³⁰⁵。社交机器人的产生是为了服务人类,尤其是从情感上让人得到满足,为了达到最佳效果,社交机器人必须具有拟人化的“人工情感”。也正因此,在人机交互中社交机器人所引起的人类的“自欺欺人”主要体现为人类对社交机器人的“情感泛滥”和“情感沉溺”,而斯派洛指出,“在我们与机器人的关系中产生的情感是道德上可悲的多愁善感的典型例子。”⁽⁷⁾³¹⁵为什么人类对社交机器人的“多愁善感”是“道德上可悲的”或者说不道德的

呢?斯派洛进一步指出了其深刻的原因,即,“沉溺于这种多愁善感违背了我们必须自己准确理解世界的(薄弱)责任。”⁽⁷⁾对于人而言,情感丰富本身并没坏处,然而再丰富的情感也需要适度,也是有其作用范围的,当我们将人类宝贵的情感作用于非生命体时,就成了“情感泛滥”,而一旦沉溺于其中就容易产生幻想,这不仅有碍于我们准确、客观地理解真实的世界,而且还会浪费我们宝贵的时间、情感等,而这些都是“我们有责任去避免的。”⁽⁷⁾³¹⁵显然,对于大多数人来说,“准确理解世界的(薄弱)责任”能够保证我们活得真实,并且使我们的人生充满意义和价值。正如斯派洛所强调的,“我认为我们直觉的力量反映了我们的信念,即虚幻的经历在人的一生中没有任何价值。这里明显不道德的是欺骗人们或鼓励他们自欺欺人的意图。”⁽⁷⁾³¹⁶

另外,社交机器人的“欺骗”伦理问题还体现在,当我们“信以为真”地和机器人进行人机互动,交流情感时,我们会变得机械化、简单化,进而出现某种程度的功能“退化”。正如特克尔(Sherry Turkle)在《群体性孤独》一书中所描述的那样,与机器相处久了,我们不仅会把我们的情感“缩减到机器可以制造的范围内”,同时还会降低“对所有关系的期待”。⁽¹¹⁾扎韦斯卡也提出了类似的看法,她认为,“当我们继续将机器人拟人化时,我们也冒着否定人的属性的风险。”⁽⁴⁾从个体的角度来看,人之所以为人就在于其丰富而深刻的社会性和关系性,当个体完全脱离社会或者与社会渐行渐远,则有违人的本质,是对自己的“背叛”;从人类的发展史来看,人应当有一种使命感,即通过自身的努力来为人类文明、人的进化尽一己之力,从深度和广度上来完善人的内涵,而非相反,否则也是一种“背叛”。此外,阿曼达·夏基和诺埃尔·夏基还具体分析了社交机器人“拟人化”的欺骗对老人和儿童的心理、行为等方面产生的负面影响。⁽⁵⁾

针对社交机器人“欺骗”的伦理风险,学者们提出了不少解决的方法,传统的解决方法大致分为两条路径:一是继续增强社交机器人的“拟人化”程度,让它们更“像人”,功能更为完善;二是在一定程度上限制社交机器人的发展。例如,朔伊茨(Matthias Scheutz)呼吁相关机构应当“尽可能制定出未来立法者可以用来对可部署的社交机器人类型施加明确限制的原则”⁽¹²⁾²¹⁷。总体来看,这两条路径各有优势,但也有不足。就第一条路径来看,它对

技术的要求非常的高,以致遭受到了众多质疑,例如,戈德贝希尔(Rich Firth - Godbehere)从人的认知的“模糊性”、复杂性、人类进化的独特性等方面对机器人高度“拟人化”的可行性提出了质疑。⁽¹³⁾就第二条路径来看,随着科技的飞速进步和人对社交机器人依赖程度的增强,“限制”社交机器人某方面的功能,甚至“限制”某种类型的社交机器人的发展不太现实。因此,上述两条路径都不能彻底地解决社交机器人因“拟人化”属性而产生的“欺骗”伦理问题。达菲认为,“人与机器人的交互作用是拟人化在机器人系统中应用的主要动机。”⁽²⁾¹⁷⁸进而可以看出,社交机器人本身就是一个矛盾体,一方面,它需要通过“拟人化”来促进人机交互,服务于人;另一方面,“拟人化”又造成了一系列的“欺骗”伦理问题。面对这一难题,难道真的就束手无策了吗?科克尔伯格的情境体验视角给了我们有益的启发。

二、“欺骗”的消解:真实的人机交互情境体验

首先,科克尔伯格分析了社交机器人“欺骗”伦理问题的立论根据,“第一,在情感交流中,机器人在我们看起来可能是有情感的实体,例如,像人类一样,但实际上它只是一台机器;第二,这种‘伪装’在道德上是不可接受的。”⁽⁸⁾³⁹¹紧接着,他对这种依据进行了质疑,认为这一论点是从“柏拉图式的形而上学”传统出发,而这一传统预设了这样的前提,即“实体是什么和实体在我们看来是什么之间有一个明显的区别”⁽⁸⁾³⁹¹。也就是说,支持社交机器人具有欺骗性的结论是从“本体论的真实”与否得出来的,尽管机器人“看起来”真实但从本体论意义理解的机器人实体并不真实,因此构成“欺骗”。我们能否改变策略,不去深究“本体论的真实”问题呢?科克尔伯格提出,如果我们从现象学的视角来看待上述问题,“本体论的真实”与“看起来”真实的冲突性就不会那么紧张,他认为,“根据另一种哲学认识论传统(现象学),在真实与表象之间做出如此明显的区分是不可能的:我们对真实的看法总是经过中介或构建的,我们所认为的真实是呈现给我们的真实。”⁽⁸⁾³⁹¹可惜的是,科克尔伯格并没有深入分析这种“呈现”的真实从何而来,这里就涉及到真实人-

机体验得以发生的第一个前提,我们可以从海德格尔的理论中找到答案。海德格尔认为,在“一种知识(认识)的关系尚未建立起来”之前,“还只有一种前理论的关系”⁽¹⁴⁾²。具体而言,“对象性的东西、被认识的东西,本身是疏远的,是从本真的体验中被提取出来的。”⁽¹⁴⁾¹²在海德格尔看来,认识是属于“理论的东西”,体验属于“前理论的东西”,而“‘理论的东西’是以‘前理论的东西’为基础的”。⁽¹⁴⁾³⁻⁶而我们对于机器人是否是真实的判断就是一种关于机器人的“知识(认识)的关系”,它取决于我们的人-机体验,或者说“起因于”(motiviert in)“真正的体验世界”⁽¹⁴⁾⁶。既然我们在人-机交互的过程中将机器人看作是“像人类一样”,那么取决于这种人-机体验的对机器人真实与否的判断也必定和体验自身相一致,即同样认为机器人“像人类一样”,所以,是体验的真实决定了“我认为机器人像人类一样”的“呈现”的真实。

然而,人在人-机交互的过程中为什么能够获得真实的体验呢?这里就涉及到真实人-机体验得以发生的第二个前提。海德格尔为了说明个体体验的不同,阐述了一个“讲台体验”案例:对于学生和“我”而言,“几乎一下子就看见了”讲台,能够获得“讲台体验”,而对于黑森林农民和经常居住在小木屋里的塞内加尔黑人朋友而言“看到的东西就难以细说了”,甚至会把讲台看作是“一个‘他不知道拿它怎么办’的东西”。⁽¹⁴⁾⁸⁻¹⁰为什么对于同一种东西,“我”的看和塞内加尔人的看会如此不同呢?或者说会有不同的体验呢?张一兵教授通过“固有之物”对此进行过分析,他认为,“这里‘固有之物’不是说一种原来存在的现成东西,而是随‘我’和(塞内加尔)黑人生活中固有的一切可情境化的行为、经验和概念构架。正是这个建构性的故有生成当下的建构性回波和复境。”⁽¹⁵⁾⁶可以理解为,个体的“固有之物”是造成不同“体验”的根本原因。而“固有之物”并非一成不变之物,它是“建构性的”,并且与我们生活中的“可情境化的行为、经验和概念构架”紧密相关。就人类与社交机器人进行交往互动而言,人类的“拟人化”能力正是这一特定情境体验中的“固有之物”。“拟人化”能力既是人类与社交机器人能进行互动的前提,又是人类能够通过人机交互而获得相关体验的基础。设想一下,如果没有“拟人化”能力,人类只会把机器人看作是“机器”,正是凭借着“拟人化”能力,人类能够将无生命

的机器“生命化”,让它们能够成为人类生命和生活的一部分,让自身对于生命和生活意义的追求在人机交互的过程中“回响”。

真实人机交互体验得以发生的第一个前提要求我们要投入到人机交互的现实情境中,要相信体验给我们的“呈现”,而非纠缠于客体化的机器人本身。因此,我们应当关注“关于机器人的体验现象学”。⁽¹⁶⁾⁶⁶科克尔伯格认为,我们与机器人交互的体验不仅“产生于实验室和研究室”,而且“它首先出现在医院、无人驾驶飞机控制室、人们的家里——在机器人能发挥作用的所有地方。”⁽¹⁶⁾⁶⁶⁻⁶⁷“我们的经验是,当我们使用这些设备时,只有一种体验和真实,它们是我们生活的一部分,并与我们的生活纠缠在一起。”⁽¹⁷⁾⁷⁶具体的生活情境才是社交机器人与人构建关系,实现人机交互的主战场,而非研究机器人的实验室或者书斋。社交机器人从实验室走进生活的重要意义就在于,情境体验是向着生活和生命的,而不是与生活 and 生命体验无关的客观对象化认识过程,即情境体验拒绝抽象地探讨“本体论的真实”。换言之,人们不关心抽象本体论意义上的机器人与人之间的关系(“复制”或“模拟”)不会在意“本体论的真实”与否,相反,人们会关心与机器人交互过程中产生的体验的真实性。具体而言(在现实中)“当我们与智能设备、类人的自主机器人、语音界面(如谷歌 Assistant)等交互时,我们通常不会将其体验为‘虚幻’,反而体验为‘真实’。”⁽¹⁷⁾⁷⁶当我们从情境体验的角度来探讨社交机器人“欺骗”伦理问题时,基于本体的机器人的“虚假”属性被“疏远”了,取而代之的是体验的真实。这样一来,在人机交互的情境体验中,起因于“本体论的真实”的社交机器人“欺骗”也被消解了。

真实人机交互体验得以发生的第二个前提要求我们充分发挥人类的“拟人化”能力。前面已经分析过社交机器人本身的矛盾性,一方面,人类需要将它拟人化,赋予它迎合“拟人化”的属性,来增强人-机互动;另一方面,这种“拟人化”属性又是虚假的。而关于社交机器人的“欺骗”伦理问题只看到了矛盾的第二个方面,相反,如果我们不是紧盯着“欺骗”,而是进一步去考察矛盾的第一个方面,即人类为什么要将机器人拟人化,我们就会从一种更为宏观的视角来理解社交机器人伦理。依据之前的论述,人类之所以能够投入到人机交互体验中并且获得真实的感受,是因为我们的“固有之

物”——“拟人化”能力,它对于我们来说意义何在?心理学家卡波雷尔(L. R. Caporael)认为,“拟人化”是在人类长期进化过程中形成的,它“可以在更大的范围内将群体‘重新编织’到社会中”⁽¹⁸⁾²²⁶。显然,就目前的人类“拟人化”能力而言,它已经失去了它原始的作用意义。但为什么人类还保留着这种“认知偏差”的残余呢?或者说,它目前还有存在的意义吗?扎韦斯卡等人认为,“拟人化”之所以被广泛应用于木偶和玩偶中,是因为人们“在使用拟人化道具的同时,积极地为创造生命的幻觉做出贡献”⁽¹⁰⁾⁸⁰。也就是说,“拟人化”至少还能够为人类创造有意义的“生命的幻觉”。具体到特定的人机交互情境体验之中,扎韦斯卡认为,“社交机器人的角色不是将拟人化的投影强加给机器人用户,而是作为人-人类社会互动和人类对人的意义的探索的催化剂。”⁽⁴⁾可见,人类的“拟人化”能力在人类与社交机器人进行互动的过程中为人类创造积极的“生命的幻觉”和为促使“人类对人的意义的探索”发挥着重要作用。当社交机器人作为一种人类进行自我探索的工具和人类追求其自身意义的“催化剂”时,我们就会去追问诸如这样的一些问题,“人机交互(关系)能促进人类的繁荣和幸福吗?这些交互能构成友谊、爱情或关系吗?他们(人和机器人)能共同塑造一个繁荣的共同体吗?”⁽¹⁹⁾²²⁰很显然,这些问题所关心的是人类自身的道德追求和实践,而非“欺骗”伦理之类的机器人自身的道德问题,是“人-机器人-人的交互”⁽⁴⁾模式在伦理道德领域的具体体现。科克尔伯格把这一模式下的社交机器人伦理称为“人类善的伦理(ethics of human good)”,他认为“人类善(human good)”“关注的是我们应该如何生活,我们应该养成什么样的道德习惯和道德品质”⁽¹⁹⁾²²⁰。

三、社交机器人伦理的新解“人类善”

在社交机器人与人类共存的高科技时代,我们到底“应该如何生活”呢?这一问题的回答离不开对“人类善”的深刻理解。也只有深刻理解了“人类善”的本质,我们才能真正把握情境体验视域下社交机器人伦理的要义。科克尔伯格认为,在理解“人类善”时,会出现两种不同的方法,“一种方法是从人类的善、人类的繁荣、幸福、友谊和爱的某种概

念出发。”⁽¹⁹⁾²²⁰在这种方法下,我们可以分析“人类的善、人类的繁荣、幸福、友谊和爱”等概念的内涵,并以此为基础,“提出一份我们用以判断个人机器人的伦理方面的标准清单。”⁽¹⁹⁾²²⁰对照着“人类善”的伦理标准清单,我们很容易去评价社交机器人的行为是否符合伦理标准。然而,科克尔伯格并不认同这一方法,他认为,“善不是独立于实践之外的;它只有在实践中才能存在和繁荣。”⁽¹⁹⁾²²⁰于是,他提出了第二种考察社交机器人伦理的方法,“我们需要从人机交互的具体体验和想象开始,然后讨论善在性能清单方面是如何被理解的,而不是从作为先验道德规范的性能清单开始。”⁽¹⁹⁾²²⁰很显然,第二种方法更符合具体的人机交互情境,更注重人机交互过程中的“体验”和“想象”这与科克尔伯格的情境体验理论一脉相承。具体而言,如何实现这种“人类善”呢?科克尔伯格认为我们应当发挥人类的“道德想象力”,去“研究、想象和塑造”能够有利于“人类善”的“人类与机器人共同生活的具体情境”并且要将“增进人类繁荣和福祉的人-机共同生活”作为“社交机器人的设计、使用和管理”的目标。⁽¹⁹⁾²²⁰总而言之,要使社交机器人伦理回归到人类生活本身,要从人机交互情境体验的整体出发来考察“人-机共同生活”的道德意义,让人类的繁荣、福祉成为衡量人机交互关系好与坏的道德落脚点,成为指导人与社交机器人该如何共同生活的标尺。

如果说社交机器人的“欺骗”伦理是警醒我们尽量去回避与机器人的“亲密接触”,还机器人以非生物性的真实面目,让自己远离幻想,那么社交机器人的“人类善的伦理”则是鼓励我们尽可能多地、深入地进行人机交互活动,不仅如此,我们还要去挖掘内心最真实的交互体验,同时展开想象的翅膀去构想未来更好的人机交互情境,最终构建“增进人类繁荣和福祉的人-机共同生活”。从“欺骗”伦理到“人类善的伦理”不仅仅是研究范式的转向,更重要的是对待社交机器人态度的转变,从消极回避到积极拥抱。科技进步及其与我们生活的高度融合是未来社会发展的必然趋势,科克尔伯格将“人-机共同生活”看作是人机交互最广阔的情境,人类应当在此情境之中充分地发挥主观能动性,去学会与机器人一起生活,重要的是去观察、聆听、想象、践行其中的真善美,最终使得人类的价值观念和道德观念在未来人机共生的“技术-道德的世

界”中得到全新的阐释。⁽⁸⁾³⁹³因此,科克尔伯格不仅通过情境体验回应了社交机器人的“欺骗”伦理问题,而且还将社交机器人的伦理问题推向了新的高度,即关乎人类未来命运、福祉、繁荣的“人类善的伦理”高度。

四、结语

社交机器人到底是“欺骗”还是“向善”取决于我们对待社交机器人的态度和分析社交机器人伦理的角度。如果说传统的“欺骗”伦理问题是向“后”看,注重本体论追溯,那么科克尔伯格则是向“此”和向“前”看,注重情境体验,注重发展着的人机未来关系。需要注意的是,我们不能因为自身看问题方式的转变而忽略了社交机器人现实中所存在的问题。比如,尽管我们肯定了人类运用“拟人化”的能力,并希望积极发挥这种特有的能力与社交机器人进行良好的互动,并共同构建美好的生活,但不能否认,考虑到当前人工智能技术的发展水平和一些特定人群的身心特点,人机交互更多的可能是“欺骗”而非“向善”。阿曼达·夏基和诺埃尔·夏基认为,婴幼儿和特殊老人“都有强烈的社交欲望”,但因为受认知能力的限制,他们“无法理解机器人行为背后的机制”⁽⁵⁾³⁵。这样的结果必将会使婴幼儿和特殊老人更容易“被动”地被社交机器人塑造,而非“主动”地利用其“拟人化”能力来构建美好的人机交互体验,进而会对婴幼儿的身心发展,老人的尊严等产生负面影响。同时也要注意,我们不能因为人与社交机器人交互过程中出现的一些问题而放弃了人机共同构建美好生活的初衷和理想。研究表明,社交机器人确实能够消除老人的孤独感,并能够智能地呵护老人的健康,它也能为孩子带来更多的欢笑,它能帮助我们处理一些日常琐事,成为人类生活的助手,为人类的美好生活做出积极贡献。一个不可否认的事实是,社交机器人已经走入我们日常生活,并且未来人机有深度融合的趋势,面对这一人类未曾有过的社会发展前景,我们固然要用人类的伦理道德观念来调整和完善,但是这些伦理规范不应是抽象和绝对的,而应该有丰富的情境支撑和深刻的价值内涵。社交机器人伦理应当是一种既立足于人机交互的情境体验,又指向人类繁荣与福祉的“技术-道德”规范。

参考文献

- (1) Breazeal C L. *Designing Sociable Robots* [M]. Massachusetts: The MIT Press, 2002: xi.
- (2) Duffy B R. Anthropomorphism and the Social Robot [J]. *Robotics and Autonomous Systems*, 2003, 42(3-4): 177-190.
- (3) Fong T, Nourbakhsh I, Dautenhahn K. A Survey of Socially Interactive Robots: Concepts, Design, and Applications [J]. *Robotics and Autonomous Systems*, 2003, 42(3-4): 143-166.
- (4) Zawieska K. Deception and Manipulation in Social Robotics [EB/OL]. [2015-03](2019-12-10). https://www.researchgate.net/publication/272474319_Deception_and_Manipulation_in_Social_Robotics.
- (5) Sharkey A, Sharkey N. Children, the Elderly, and Interactive Robots [J]. *IEEE Robotics & Automation Magazine*, 2011, 18(1): 32-38.
- (6) Stanford Encyclopedia of Philosophy. The Definition of Lying and Deception [EB/OL]. [2015-12-25](2019-12-05). <https://plato.stanford.edu/entries/lying-definition/#TraDefDec>.
- (7) Sparrow R. The March of the Robot Dogs [J]. *Ethics and Information Technology*, 2002, 4(4): 305-318.
- (8) Coeckelbergh M. Are Emotional Robots Deceptive? [J]. *IEEE Transactions on Affective Computing*, 2012, 3(4): 388-393.
- (9) Carson T L. *Lying and Deception: Theory and Practice* [M]. New York: Oxford University Press, 2010.
- (10) Zawieska K, Duffy B R, Strońska A. Understanding Anthropomorphisation in Social Robotics [J]. *Pomiary Automatyka Robotyka*, 2012, 16(11): 78-82.
- (11) [美] 雪莉·特克尔. 群体性孤独 [M]. 周涛, 刘菁荆, 译. 杭州: 浙江人民出版社, 2014: 136.
- (12) Scheutz M. The Inherent Dangers of Unidirectional Emotional Bonds Between Humans and Social Robots [C]// Patrick Lin, Keith Abney, and George A Bekey. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: The MIT Press, 2012: 205-221.
- (13) Rich Firth - Godbehere. Emotion Science Keeps Getting More Complicated. Can AI Keep Up? [EB/OL]. [2018-11-29](2019-11-05). <https://howwegettonext.com/emotion-science-keeps-getting-more-complicated-can-ai-keep-up-442c19133085>.
- (14) [德] 海德格尔. 形式显示的现象学: 海德格尔早期弗莱堡文选 [M]. 孙周兴, 译. 上海: 同济大学出版社, 2004.
- (15) 张一兵. 作为发生事件(Ereignis)的生命体验——关于青年海德格尔早期弗莱堡讲座的构境论解读 [J]. *现代哲学*, 2011(5): 1-9.
- (16) Coeckelbergh M. The Moral Standing of Machines: Towards a Relational and Non-Cartesian Moral Hermeneutics [J]. *Philosophy & Technology*, 2014, 27(1): 61-77.
- (17) Coeckelbergh M. How to Describe and Evaluate “Deception” Phenomena: Recasting the Metaphysics, Ethics, and Politics of ICTs in Terms of Magic and Performance and Taking a Relational and Narrative Turn [J]. *Ethics and Information Technology*, 2018, 20(2): 71-85.
- (18) Caporael L R. Anthropomorphism and Mechanomorphism: Two Faces of the Human Machine [J]. *Computers in Human Behavior*, 1986, 2(3): 215-234.
- (19) Coeckelbergh M. Personal Robots, Appearance, and Human Good: A Methodological Reflection on Roboethics [J]. *International Journal of Social Robotics*, 2009, 1(3): 217-221.

Social Robot Ethics Based on Situational Experience: From “Deception” to “Good”

WANG Liang

(College of Marxism, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: The ethical problem of social robot's “deception” is a hot topic in the research of robot ethics. But most of the traditional researches are based on the view of entity attribute, through analyzing the “anthropomorphic” attribute of social robots, to get the judgment that social robots will produce the ethical risk of “deception”. Based on the “situational experience” of the interaction between human and social robots, Coeckelbergh suspended the “anthropomorphic” attribute of social robots, and then turned the “deception” ethics of social robots to “ethics of human good”.

Key words: social robots; situational experience; deception; good

(本文责任编辑: 崔伟奇 郑泉)