·科技与社会· 文章编号:1000-8934(2020)01-0056-06

社交机器人“单向度情感”伦理风险问题刍议

王 亮

(西安交通大学 马克思主义学院,西安 710049)

摘要:将人工情感作为社交机器人的典型特征之一已经成为学界共识,由于人工情感的不真实性和人类情感的真实性之间的失衡性关系,引发了一系列伦理风险问题,其中主要体现于社交机器人对人类同理心的“操控性”和其“欺骗性”。人类需要从法律、伦理监管,社交机器人的优化设计,以及人类对自身的道德、价值观念的调整等方面着手才能有效地应对相关伦理风险。

关键词:社交机器人;单向度情感;伦理风险

中图分类号:N031 文献标识码:A

随着社交机器人的普及,其所引发的伦理风险问题也受到越来越多的关注,其中关注的焦点就体现在社交机器人的人工情感方面。人工情感毕竟不是真实的人类情感,人们对它的认知存在多方面的不足,因此这会造成许多潜在的伦理问题。本文通过对不同学者的理论进行梳理和对比,分析社交机器人“单向度情感”伦理风险问题的成因、表现形式以及解决路径等,希望能引起学界的共鸣。

一、“单向度情感”:人机交互的伦理陷阱

随着人工智能技术的发展,智能化产品也日渐进入到了我们的日常生活之中,其中社交机器人的应用就是人工智能这一新技术革命所带来的广泛影响的重要体现。那么,什么是社交机器人?它有哪些典型特征?韦斯娜(Vesna Kirandziska)和内韦娜(Nevena Ackovska)两位学者认为,“社交机器人应该有一些人类的特点,比如,能进行语言和非语言交流,它们应该有自己的身体,它们应该会感知和表达情感。正如定义的那样,使机器人成为社交机器人的一个特殊条件是嵌入情感。”[1]此外,来自卡耐基梅隆大学机器人研究所的特伦斯等人(Terrence Fong, et al.)通过梳理社交机器人的发展历史和分析社交机器人的不同类型,总结出了社交机器人的七大特征,即,“表达和/或感知情感;与高层级对话沟通;学习或识别其他代理的模型;建立或维护社会关系;使用自然的暗示(凝视、手势等);表现出鲜明的个性;可以学习或发展社交能力”[2]。中国学者邓卫斌和于国龙通过对国内外多个具有代表性的社交机器人的功能进行对比分析,总结道,“纵观国内外社交机器人的发展可以发现,人机交互和情感化始终是其研究的重点。”[3]综合以上对于社交机器人的描述和研究可以看出,情感因素是社交机器人不可或缺的,是其典型特征之一。为什么对于社交机器人来说情感因素是如此重要呢?特伦斯等人分析了三大原因,“在社交机器人中使用人工情感有几个原因。当然,它的主要目的是帮助促进可信的人机交互。人工情感还可以向用户提供反馈,如指示机器人的内部状态、目标和(在一定程度上)意图等。最后,人工情感可以作为一种控制机制,驱动行为,反映出机器人在一段时间内如何受到不同因素的影响,并适应不同的因素。”[2]其实,对于后两种原因而言,无论是提供反馈(从用户角度而言),还是进行内部调控(从机器人角度而言),其最终的目的还是在“促进可信的人机交互”,这正是社交机器人的情感设定的根本原因。进一

收稿日期:2019-7-20

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讲师,博士后,主要研究方向:马克思主义理论、科技哲学等。

步,在现实的应用中社交机器人是如何来表达情感的呢?特伦斯等人分别从机器人的"言语"、"面部表情"、"肢体语言"三个方面考察了社交机器人的情感表达机理。[2]

尽管社交机器人的情感设定理由十分充分,并且其情感的表达方式也多样,但是在现实应用中的效果却不尽如人意。正如鲍姆格特纳(Bert Baumgaertner)和魏斯(Astrid Weiss)所言,"目前已经有一些陪护机器人能够根据照顾者和被照顾者之间的互动模式来表达模拟的情感,但它们的实际互动和沟通能力仍然非常有限。"[4]韦斯娜和内韦娜将这种社交机器人情感表达的现实困境归结为两点原因,一方面,"情感定义的多样性使得我们很难理解什么是真正的情感,它们是如何表现和表达的。"[1]"另一个困难表现在情感感知方面,它是因个性和文化差异造成的。"[1]就情感的多样性而言,普拉契克(Robert Plutchik)的"情感之轮(wheel of emotions)"理论认为人类在"四个强度水平上共发生224种情绪"[5]。事实上,人类所表现出的情感特征要远远多于224种,社交机器人很难做到对人的情感的全方位模拟。就跨文化、个体差异与情感表达差异之间的关系而言,情感表达样式的多样化也是显而易见的,例如,霍尔(Edward T. Hall)认为,在"高语境文化"中,人们的表达方式较为间接和含蓄,在"低语境文化"中,大多数事情都需要解释,人们的表达方式较为直接。[6]因个体差异而导致的情感表达差异现象则更为明显。可以看出,人类的文化多样性和个体的特殊性在一定程度上促成了人类情感表达的丰富性,这是社交机器人所无法比拟的。

正是由于上述这些困难,在现实中,社交机器人很难敏锐地、及时地给人以充分的情感反馈,相反,在人机交互过程中,人类对机器人的情感反馈则显得丰富而深刻的多。朔伊茨(Matthias Scheutz)认为,"已经有足够的证据表明,人们很容易受到实验室之外社交机器人的影响,尤其是当他们与机器人重复进行长期互动时。"[7]他在文章中例举了一个关于Roomba扫地机器人的真实案例,朔伊茨认为,"随着时间的推移,人类会对Roomba产生一种强烈的感激之情,因为它能清洁他们的家。"[7]扫地机器人本来是被设计为替人类打扫房间的,但是"有些人会替Roomba完成打扫工作,这样扫地机器人就可以休息了,而另一些人则会把他们的Roomba介绍给他们的父母,或者在旅行时带上它,因为他们成功地发展了(单向的)关系。"[7]可以看到,朔伊茨已经敏锐地察觉到了人机交互的单向性关系,而这种单向性主要体现在情感维度上,一方面,由于技术的限制,社交机器人无法充分地模拟人类复杂多样的情感,并给人以对等的情感反馈,另一方面,由于社交机器人的拟人化特征、与人的长期互动、人类的"多愁善感"等原因,人类赋予机器人以情感,而这种情感大多是"一厢情愿"的,非对称的。社交机器人的设计初衷是增强人机交互,陪护和照料人类,但是由于它的"单向度情感"缺陷,也为人类在人机交互过程中埋下了伦理风险陷阱。

二、"单向度情感"伦理风险的典型类型:操控性和欺骗性

1. 被操控的同理心

机器人被赋予情感,很大程度上是人类的同理心(Empathy)在"作怪"。对于人类而言,同理心有其特殊的作用,是人类在进化过程中形成的心理机制。美国堪萨斯大学的舒尔茨(Armin Schulz)认为,"似乎有两种不同的选择性压力来源导致了这种特质的进化(尽管还需要进一步的研究来证实这一点)。首先,同理心可以促进合作,而合作反过来又具有很强的适应性(比如帮助后代)。然而,进一步证明,这种合作同理心可以是利他的,也可以是利己的。其次,同理心可以帮助快速应对环境突发事件(如掠夺性攻击)。"[8]由此可见,在人类生存进化过程中,同理心是一种必不可少的能力或者特质。苏林思(John P. Sullins)认为,"无论是生理因素,还是社会进化因素似乎都为我们提供了一种能力,使我们能够将情感依附扩展到我们自己物种之外。"[9]而一旦人类将这种能力运用到非生物体的机器人身上,就会产生一些风险,苏林思直言不讳地指出,"有一件事应该是非常清楚的,那就是情感机器人,就像今天看起来的那样,通过操纵人类的心理来达到最佳效果。人类似乎有许多进化出来的心理弱点,可以利用这些弱点让用户接受模拟的情感,就像它们是真实的一样。利用进化压力所带来的人类根深蒂固的心理弱点是不道德的,因为这是对人类生理机能的不尊重。"[9]可以看出,当人类与机器人处于一种"单向度情感"关系之中时,人类

的同理心可以被情感机器人操控、利用，这不仅是不道德的，而且其后果也是不堪想象的。

然而在现实中，这种情况不仅没有得到有效抑制，反而还被“煽情化”了。正如朔伊茨所言，“社交机器人显然能够推动我们的‘达尔文按钮’，即我们社交大脑中的进化产生的机制，以应对社会群体的动态和复杂性，这些机制自动触发对其他代理人心理状态、信念、欲望和意图的推断。”[7] 社交机器人为什么能如此显然地“推动我们的‘达尔文按钮’”呢？或者说，在社交机器人面前，人类的同理心为何表现得如此明显？在朔伊茨看来，这与对社交机器人的设计和宣传有直接关系。可以看出，从一开始社交机器人的设计就陷入了“情感矛盾”，一方面，它需要通过情感的内置，人格化、可爱的形象来吸引用户，进而促进人机交互；另一方面，这些人格化、情感化的设计又促使人类更加依恋机器人，如果再加上商业性的过度煽情式宣传，人类的同理心就会被强化，甚至被操控。这种操控不仅体现为“对人类生理机能的不尊重”，而且也可能会传导至对道德的操控，因为“同理心对道德生活至关重要，它有助于发展广泛的道德能力，如同道德能力被各种伦理理论所定义的那样。同理心有可能丰富和加强对他人的道德审慎，行动和道德辩护。”[10] 此外，对同理心的操控可能还会衍生出其他伦理问题，例如，社交机器人通过利用同理心来取得用户的更多信任，进而广泛收集用户的隐私信息等。

2. 具有欺骗性的社交机器人

相较于人的同理心被操控，有一个更为宏观的伦理问题需要人类去面对，即欺骗。为什么社交机器人具有欺骗性呢？科克尔伯格(Mark Coeckelbergh)从三个方面总结了原因，“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。”[11] 可以看出，这三者之间的逻辑是层层递进的，“欺骗”之所以产生，根源在于机器人自身的非生物体特征，基于电子元器件、算法等构成要素的机器人无法产生生物意义上的真实情感，进而其所表达的非真实情感就构成了欺骗。斯派洛(Robert Sparrow)也通过比较机器人与生物体之间的区别，指出了机器人的欺骗性特征，他认为，“尽管宠物机器人的行为方式可能被设计得与真实动物的行为非常相似，但它们的行为仍然只是模仿。特别是，机器人没有任何感觉或体验。”[12] “机器人至多有复杂的机制来模仿情感状态。”[12] 阿曼达·夏基(Amanda Sharkey)和诺埃尔·夏基(Noel Sharkey)则从拟人主义(anthropomorphism)的角度对机器人的欺骗问题进行了分析，他们认为，“设计机器人来鼓励拟人化属性可能被视为一种不道德的欺骗形式。”[13]

诚然，正如之前所讨论的，设计师们将机器人拟人化有利于促进人机交互。但是它的负面影响也是显而易见的，斯派洛认为，“一个人要想从拥有一只机器宠物中获得巨大的好处，就必须系统地欺骗自己，不去了解他们与动物之间关系的真实本质。它需要一种道德上可悲的多愁善感。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。这些机器人的设计和制造是不道德的，因为它预设或鼓励了这种欺骗。”[12] 根据斯派洛的论述，至少可以看出机器人的欺骗性会造成两点负面影响：第一，使用户沉溺于情绪化；第二，削弱了用户理解和认知世界的(薄弱)责任。就第一点而言，情绪化或者多愁善感本身并没有太大的坏处，更谈不上“不道德”，但是通过欺骗的方式而将人的情感导向于“错误”的对象，甚至使人的情感沉溺于其中的行为就是一种不道德。此外，人们还可能会被机器人激起的情感蒙蔽双眼，使人无法准确地理解和认知世界，反而活在自我陶醉的虚幻世界中，进而削弱了本身就较弱的人“正确理解世界”的责任，而这一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。斯派洛强调到，“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。”[12] 也许社交机器人本身并没有任何“意图”，但是其集非生物体性和拟人化为一身的特征导致了来自于机器人的虚拟情感和来自于人的真实情感关系的失衡，这种虚拟与真实之间失衡的、不对等的情感关系就体现为一种欺骗关系。而这种欺骗性导致了人们“正确理解世界”的(薄弱)责任的进一步弱化，进而人们可能会沉溺于虚幻的人机交互之中，从而失去人生本该有的真实的价值。正如特克尔(Sherry Turkle)在《群体性孤独》一书中所描述的那样，“当你和机器‘生物’分享‘情感’的时候，你已经习惯于把‘情感’缩减到机器可以制造的范围内。当我们已经学会对机器人‘倾诉’时，也许我们已经降低了对所有关系的期待，包括和人的关

系。在这个过程中，我们背叛了我们自己。”[14]这样看来，作为罪魁祸首的社交机器人的欺骗性确实是“不道德”的。

三、伦理风险化解的可能路径：从社交机器人到人

对于社交机器人造成的这些“单向度情感”的伦理风险我们是否束手无策了呢？答案是否定的。学者们分别从不同的路径提出了风险化解方案。总的来说，可以分为两条路径：一条路径是围绕着社交机器人展开的；另一条路径是围绕着人展开的。

1. 以社交机器人为中心的伦理风险化解路径

一般来说，从外部对社交机器人进行监督管理是一种有效的化解伦理风险的方法。目前世界各国都在紧锣密鼓地制定各种人工智能的伦理原则或相关法律，但这些法律原则都比较宽泛，很少有专门针对社交机器人量身定制的。难得的是，在2019年3月IEEE全球倡议推出了《符合伦理的设计：以自主和智能系统优先考虑人类福祉的愿景》（第1版），其中有一个章节专门来探讨社交机器人伦理问题。在这一专门章节中，有六个原则性倡议被提出：“（1）亲密系统的设计或部署不应有成见、性别或种族的不平等或加剧人类苦难。（2）亲密系统的设计不得明确地参与对这些系统用户的心理操控，除非用户意识到他们正在被操控并同意这种行为。任何操控都应通过选择性加入（opt-in）系统进行管理。（3）关怀式自主智能系统的设计应避免造成用户与社会的隔离。（4）情感机器人的设计者必须公开告知，例如，在产品说明书中写清这些系统可能会产生副作用，诸如干扰人类伙伴之间的关系作用方式，导致用户和自主智能系统之间形成不同于人类的依赖关系。（5）具有关怀性用途的自主智能系统不应该被呈现为具有法律意义的人，它们也不应该被赋予人的身份并进行售卖。（6）关于个人形象的现行法律需要从关怀式自主智能系统方面进行重新审议。除了其他伦理考虑外，关怀式自主智能系统还必须要与当地的法律和习俗相适应。”[15]可以看出，这六个方面的原则倡议涵盖了社交机器人伦理的各个方面，既有心理操控问题、情感依赖问题，又有机器人身份问题、外观问题、歧视问题、跨文化问题等，它们对社交机器人伦理原则和相关法律的制定具有较强的指导意义。

除了从外部监管来控制社交机器人的伦理风险之外，还可以从社交机器人的内部设计来寻找应对的办法。根据上面所讨论的情况来看，社交机器人的伦理风险主要是由情感问题造成的，所以其内部的设计应当考虑情感因素。朔伊茨提出了一个比较极端的方案，他认为，“我们需要的是一种方法来确保机器人不会以另外的，（正常）人类无法做到的方式来操纵我们。为实现这一目标，可能需要采取激进措施：赋予未来机器人以类人的（human-like）情感和感觉。”[7]然而，这样美好的愿望是否能实现呢？有学者对此提出了质疑。戈德贝希尔（Rich Firth-Godbehere）认为，人对复杂语境的感知、人脑记忆的动态建构、人的情感过程的模糊性、人类进化出来的感官、人的内在感受性等等，都是机器人无法模拟的，这也导致了机器人无法真正地进行“情感体验”。[16]此外，戈德贝希尔还提出了一个颇让人深思的问题，“创造一台体验情感的机器并不能告诉我们是否我们拥有一台和我们一样感受情感的机器。它可能表现得好像是这样，它可能说它是这样，但是我们真的能知道它是这样吗？”[16]戈德贝希尔在这里提出了一个挑战，即就算我们制造出能够体验人类情感的机器人，但我们能否真正理解，甚至体验机器人的情感呢？如果不能做到彼此理解，就会出现一种新的情感失衡。

和戈德贝希尔的“诘难”相比，鲍姆格特纳和魏斯的批判显得更有“杀伤力”，他们直接否认了情感内置方案的必要性。他们认为，“陪护机器人的相关行为对于成功建立其与人之间的关系至关重要，而不是这种行为的来源。因此，我们认为，除非情感理论是建立在纯粹的行为基础上的，否则，对于老年人陪护机器人的人机交互伦理来说，情感理论是不必要的。”[4]鲍姆格特纳和魏斯不仅认为行为比情感更重要，而且还认为“情感会妨碍有效的护理行为。”[4]可以看出，鲍姆格特纳和魏斯从相反的方面，即通过解除社交机器人的情感重要性来化解了因机器人情感问题而起的伦理风险。

2. 以人为中心的伦理风险化解路径

如果说鲍姆格特纳和魏斯从机器人的角度“釜底抽薪”式地化解了社交机器人的伦理风险问题，那么科克尔伯格则从人的角度消弭了机器人的本体论预设，进而也化解了社交机器人的伦理风险。科克尔

伯格从一开始就亮明了自己的立场和方法，他强调到，“我提出的机器人伦理学方法是有意识地以人类为中心，而不是以机器人为中心。让我们转向交互的哲学，认真对待外观的伦理意义，而不是关于机器人究竟是什么或（能够）思考什么的心理哲学。这是一个从‘内部’（机器人的‘心理’）到‘外部’（机器人对我们做什么）的转变。”[17] 可以看出，科克尔伯格提出了一种独特的机器人伦理学方法，与传统机器人伦理学方法不同，这种方法不是从机器人的内部（心理）出发，而是从机器人的外部（外观特征）出发，并且将机器人的外观与人相联系，最终在人机交互的情境下来思考机器人伦理问题。科克尔伯格这一独特的机器人伦理学方法的思想来源则是现象学，正如他所说，“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我们所看到的真实。”[11] 机器人的外观特征，在这里正是一种“中介”，是“我们所看到的真实”。所以，科克尔伯格基于外观的机器人伦理学方法有效地避免了关于机器人的真实性（包括情感真实）问题的探讨，而将重点转移至与机器人外观紧密相关的人机交互情境问题的分析。

科克尔伯格认为，“因批判情感机器人而引入的真实与虚幻（reality－illusion）的区别应该是对机器人的外观的区别：在某些情境下机器人看起来像是机器，在某些情境下机器人看起来像人，而‘不仅仅是一台机器’。”[11] 正是因为情境的存在，我们不能简单地、绝对地、孤立地对机器人的真假进行评判。科克尔伯格认为，“似乎机器人可以在不同的时间、不同的环境（例如，家庭护理的环境和科学实验室的环境）以不同的方式出现在不同的人面前。机器人有不同的格式塔（Gestalts），它们不能同时体验，但都是‘真实的’可能性。”[11] 机器人之所以有多种“真实的”可能性，就在于科克尔伯格没有孤立地来考察机器人内部的“心理”、“情感”等特性，而是将其放在人机交互的情境之中来研究，有效避免了传统方法中的本体论预设，正如他所言，“属性观假设一个实体只有一个‘正确’的本体状态和意义，与机器人的‘外观’和‘感知’形成对比。那些指责人们行为不‘应该’的人依赖于道德立场的科学，而道德立场的科学假定了实体（例如，机器人，作为一个物自体本身）和实体的外观之间的二分法。但我们可以想到另一种非二元论的认识论，它拒绝这种二分法，接受一个实体可以以几种方式出现在我们面前，而这些方式都没有先验的本体论或解释学的优先权。”[18] 可以看出，在没有本体论预设的情况下，科克尔伯格的机器人伦理学方法对于处理伦理风险的优势已经完全体现出来了。“欺骗”的前提是从本体论上首先认定机器人不是情感物，或者是不真实的情感物，对于传统的伦理学理论而言机器人的非真实性是预先存在的，因此对人类构成了“欺骗”；而科克尔伯格基于外观的机器人伦理学理论认为，在具体的人机交互情境中，不存在任何“本体论”的优先性，只有人机交互的关系性和体验性，所以也无所谓“欺骗”。科克尔伯格强调，“我们所需要的，如果有的话，不是‘真实’，而是与特定情境相适应的恰当的情感反应。”[11] 至此，社交机器人的欺骗性伦理风险问题就消弭在人机交互情境之中。科克尔伯格对社交机器人与人类进行交互的伦理风险问题始终持较为乐观的态度，他对未来人机共生世界的一些观点给了我们十分有益的启示，他认为，“尽管现在我们倾向于从柏拉图式和浪漫主义的角度来看待与社交机器人的情感交流，但在未来，如果我们的价值观发生变化，如果我们对与其他实体之间的关系更加信任，我们很可能会学会与我们现在称之为‘欺骗’的机器人一起生活。”[11] 诚然，未来的世界不得而知，但是科技进步及其与我们生活的高度融合是未来社会发展的一种必然趋势，人类为了更好的生存，为了繁荣福祉，面对不断革新的世界应当敞开怀抱，积极地去调整自己的价值观念和道德观念。

四、结　语

制造社交机器人的目的是促进良好的人机交互，尤其是帮助那些需要情感安抚的群体，因此，具有人工情感是社交机器人的典型性特征之一。然而，虚拟的人工情感和真实而丰富的人类情感之间存在一种天然的失衡性，人类对社交机器人形成了一种“单向度情感”依赖，进而产生了一系列潜在的伦理风险。一方面，人类“柔弱”的同理心在高度拟人化的社交机器人面前“不堪一击”，用户很有可能面临着被操控的伦理风险；另一方面，在与虚拟的社交机器人进行人机交互的过程中，人类可能会沉溺于自欺欺人式的情感之中，逃避正确理解真实世界的责任，虚度有

价值的人生。为了化解这些潜在的伦理风险，一方面需要加强法律道德的监管，从各个环节来避免风险产生的可能；另一方面需要通过科技的发展来使得社交机器人更像人类，变得更为“真实”；与此同时，人类也应当坦然地去面对未来人机共生的“技术-道德的世界”，[11]及时调整自身的道德观念、价值观念，积极应对新技术革命所带来的变化。

参考文献

〔1〕Kirandziska V，Ackovska N. A Concept for Building More Humanlike Social Robots and Their Ethical Consequence [J]. *IADIS International Journal on Computer Science and Information Systems*，2014，9(2)：19-37.

〔2〕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.

〔3〕邓卫斌，于国龙.社交机器人发展现状及关键技术研究 [J].科学技术与工程，2016，16(12)：163-170.

〔4〕Baumgaertner B，Weiss A. Do Emotions Matter in the Ethics of Human - robot Interaction？-Artificial Empathy and Companion Robots [EB/OL]. [2014] (2019-07-12). https://pdfs.semanticscholar.org/55e0/c6339f4b4541ea479160bcb7177cca93534c.pdf.

〔5〕乔建中，高四新.普拉契克的情绪进化理论 [J].心理学报，1991(4)：433-440.

〔6〕Edward T. Hall. The Paradox of Culture [C] // B. Landis，E. S. Tauber. *In the Name of Life：Essays in Honor of Erich Fromm*. New York：Holt，Rinehart and Winston，1970：218-235.

〔7〕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.

〔8〕Schulz A W. The Evolution of Empathy [C] // Heidi L. Maibom. *The Routledge Handbook of Philosophy of Empathy*. New York：Routledge，2017：64-73.

〔9〕Sullins J P. Robots，Love，and Sex：The Ethics of Building A Love Machine [J]. *IEEE Transactions on Affective Computing*，2012，3(4)：398-409.

〔10〕Julinna C. Oxley. *The Moral Dimensions of Empathy：Limits and Applications in Ethical Theory and Practice* [M]. New York：Palgrave Macmillan，2011：4.

〔11〕Coeckelbergh M. Are Emotional Robots Deceptive？ [J]. *IEEE Transactions on Affective Computing*，2012，3(4)：388-393.

〔12〕Sparrow R. The March of the Robot Dogs [J]. *Ethics and Information Technology*，2002，4(4)：305-318.

〔13〕Sharkey A，Sharkey N. Children，the Elderly，and Interactive Robots [J]. *IEEE Robotics & Automation Magazine*，2011，18(1)：32-38.

〔14〕[美] 雪莉·特克尔. 群体性孤独 [M].周逵，刘菁荆，译.杭州：浙江人民出版社，2014：136.

〔15〕The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design：A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems，First Edition [EB/OL]. [2019-03] (2019-07-10). https://standards.ieee.org/content/ieee-standards/en/industry-connections/ec/autonomous-systems.html.

〔16〕Rich Firth-Godbehere. Emotion Science Keeps Getting More Complicated. Can AI Keep Up？ [EB/OL]. [2018-11-29] (2019-07-05). https://howwegettonext.com/emotion-science-keeps-getting-more-complicated-can-ai-keep-up-442c19133085.

〔17〕Coeckelbergh M. Personal Robots，Appearance，and Human Good：A Methodological Reflection on Roboethics [J]. *International Journal of Social Robotics*，2009，1(3)：217-221.

〔18〕Coeckelbergh M. The Moral Standing of Machines：Towards A Relational and Non-Cartesian Moral Hermeneutics [J]. *Philosophy & technology*，2014，27(1)：61-77.

Discussion on “Unidirectional Emotional” Ethical Risk Arising from Social Robots

WANG Liang

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

Abstract： The artificial emotion as one of the typical features of the social robots has become a community consensus. Due to the unbalanced relationship between fake artificial emotion and real human emotion，a series of ethical risk problems are caused. These are mainly reflected in the “manipulation” of empathy and “deception” by social robots. Human beings need to deal with ethical risks effectively from the aspects of legal and ethical supervision，the optimal design of social robots，and human beings´ adjustment of their own morality and values.

Key words： social robots；unidirectional emotional；ethical risk

（本文责任编辑：崔伟奇）