

文章编号: 1000-8934(2018)05-0043-06

DOI: 10.19484/j.cnki.1000-8934.2018.05.007

论机器人道德能力的建构

杜严勇

(上海交通大学 科学史与科学文化研究院, 上海 200240)

摘要: 在机器人与人类关系越来越密切的时代背景中, 建构机器人的道德能力可以使机器人更好地与人类相处, 并维护人类的基本利益。机器人道德能力的建构与人类道德能力的培养存在一定的共同之处, 但也有明显的差异。从人与机器人互动的角度看, 需要在考虑人与机器人道德地位不对称性的基础上, 从机器人外观、社会智能、人工情感等方面入手, 建构机器人的道德能力。

关键词: 机器人; 道德能力; 社会智能; 人工情感

中图分类号: N031 **文献标识码:** A

一、引言: 机器人与道德能力

能力一般指能够胜任某项任务的主观条件和才能。道德能力是人认识各种道德现象, 在面临道德问题时能够鉴别是非善恶, 作出正确道德评价和道德选择并付诸行动的能力。⁽¹⁾⁸⁷ 道德能力由道德认知能力、道德判断能力、道德行为能力与道德意志能力等成分构成。⁽²⁾³⁰⁸ 道德能力是现实社会中所有的正常人都具备的一种基本能力, 这种能力主要是通过学习、教育、模仿等途径逐渐发展起来的。毫无疑问, 道德能力在人类的道德生活中发挥着重要作用。

对于机器人而言, 当然可以通过机器学习的方式逐步发展出一定的道德能力。目前的机器学习是建立在海量数据的基础之上的, 阿尔法围棋之所以能够战胜人类棋手, 其重要原因之一就是大量棋局中进行分析学习。但是, 对人类的道德活动进行量化并建立庞大的数据库并非易事, 所以目前发展机器人的道德能力可能主要依赖于人类的建构与机器人的道德实践, 而非机器人基于大数据的自主学习。

建构机器人的道德能力的前提是从理论上厘

清机器人道德能力的构成及影响因素。从以上关于道德能力的界定可以看出, 道德能力不是一种单一的能力, 而是多种能力的集合。机器人的道德能力跟人类会有所区别, 但也应该与人类有某些重合之处。有学者认为, 机器人的道德能力应该包括以下五个方面: (1) 道德词汇; (2) 规范系统; (3) 道德认知和情感; (4) 道德抉择与行动; (5) 道德交流。而且, 目前机器人伦理研究主要关注人们应该如何设计、应用和对待机器人等方面的伦理问题, 机器人道德 (machine morality) 则关注机器人应该拥有哪些道德性能 (moral capacities), 以及这些性能如何实现的问题, 这两种取向目前基本上处于分离状态, 而机器人道德能力的研究可以将两者整合起来。⁽³⁾

建构机器人道德能力的基本目的, 一方面让机器人拥有维护人类利益的能力, 使之行善, 而不会伤害人类; 另一方面是使机器人能够更好地与人类互动, 让人们能够接受机器人。从根本上说, 第一方面是服务于第二方面的。在机器人与人类关系越来越密切的时代背景中, 有不少学者主张应该让机器人成为人工道德行为体。我们认为, 让机器人拥有一定程度的道德能力是它们能够成为人工道德行为体的关键因素之一。当然, 拥有道德能力的机器人并不能解决所有伦理问题, 但道德能力显然是机器人能够解决伦理问题的前提条件。拙文《机

收稿日期: 2017-11-28

基金项目: 国家社会科学基金一般项目“机器人伦理问题研究”(15BZX036)。

作者简介: 杜严勇(1976—), 四川南充人, 理学博士, 上海交通大学科学史与科学文化研究院副教授, 主要研究方向: 科技哲学。

机器人伦理设计进路及其评价》⁽⁴⁾ 主要从设计者的角度初步讨论了机器人道德能力的建构,本文主要从使用者、从人与机器人互动的角度出发,探讨机器人道德能力建构的几个核心因素,由此形成一个较为全面的关于机器人道德能力的结构框架。

二、人与机器人道德地位的不对称性

众所周知,机器人(robot)一词的最初由来,即是取意于“奴隶、奴仆”。现在,许多领域的机器人所做的仍然是人类认为是枯燥、危险的工作。随着机器人智能水平的提高和应用范围的扩大,机器人与人类的关系越来越密切。人们一般倾向于认为,外观跟人类接近、拥有较高智能的机器人,应该拥有一定的道德权利,至少我们应该尊重机器人,而不是把机器人当作纯粹的工具。不过,如果机器人可以拥有道德能力的话,人类对机器人道德能力的预期与普通人的可能会不大一样。在道德能力方面,人类可能对机器人道德能力的期望较高,也就是说人类与机器人的道德地位是不对称的。

这种不对称性至少表现在两个方面。第一,从人类应该如何对待机器人的角度看,人们可能希望机器人对人类的行为更宽容。有学者设计了10个假想情景,让58名大学生分为两组分别对假想情景进行道德评价,在其中一组的情景中受害人不明确,可以假定是人类,而另一组的情景中受害人是机器人。研究发现,对于同样的违反道德的行为,当这些行为是针对人类而不是机器人时,该行为被认为是更不道德的。⁽⁵⁾当然,人们的哪些行为对机器人来说是可以接受的,而对人类自身来说是不可接受的,或者可以接受的程度,这些问题还需要具体分析。

第二,从机器人应该如何对待人类的角度看,人们可能希望机器人道德水平更高,在道德活动中更为积极主动。已有经验研究表明,人们可能对人与机器人采用不同的道德规范。在道德困境中,人们倾向于期望机器人做出符合功利主义的选择。在同样道德的困境中,人们谴责机器人的不作为多于它的作为,但谴责人的作为多于不作为。如果机器人的道德抉择没有符合人们的期望,机器人需要对自己的抉择做出相应的解释以获取人类的信任,由此也需要机器人具备一定的道德交流能力。⁽⁶⁾

毫无疑问,科技工作者应该根据人们对机器人的预期来建构机器人的道德能力,这样才能使机器人更好地被人们所接受,也可以使人们更信任机器人。事实上,从事机器人与人工智能伦理研究的学者,也普遍认为机器人的道德水平、道德标准应该比人类更高。人与机器人在道德地位上的不对称性也表明,人与机器人之间的互动研究可以参考人与人之间的互动规律,但并不能直接照搬,人与机器人之间的互动有其特殊性 & 复杂性。

尽管人们可以对机器人提出较高的要求,但是建构机器人道德能力的目标并不是让机器人成为人人满意的完美道德模范。首先,机器人的道德能力是由人类设计建构的,在建构的过程中不可避免地会有设计者的一些主观偏好融入其中,从而使机器人表现出设计者的某种“偏见”。另一方面,由于人类社会环境的复杂性,设计者不可能预见所有情况,机器人只能在有限的知识与信息的基础上进行道德判断与抉择,由此也决定了机器人道德能力的非完美性。其次,由于不同的人出于各种原因会对同样的道德行为做出不同的道德判断,让机器人成为完美的道德模范是不现实的。第三,即使表面上看来是完美的道德模范,在现实社会中可能并不是最理想的道德主体。任何人都会有一定程度的偏见,人类在与机器人的互动过程中,也会希望机器人与人一样有某种类似的认知人格。事实上,与人类相似的认知偏见在人与机器人长时间的互动过程中发挥着重要作用。也就是说,与没有偏见的机器人相比,人们更愿意与有某些偏见、非完美的机器人互动。⁽⁷⁾

可见,在考虑到人与机器人道德地位的不对称性的前提下,我们应该使机器人的道德判断与行为尽可能符合人类的预期,同时使机器人的道德认知、道德行为跟人类有较多的一致性,包括有某种程度的偏见、偶尔会犯点无伤大雅的小错误,等等,或许这样的机器人才更容易被人们所接受。

三、机器人外观与道德能力

机器人拥有合适的外表,让人产生亲近感,从而使人们更容易接受它们,是建构机器人道德能力的基本前提。对机器人的外观与人类对其的好感度(affinity)之间的关系,影响最大的理论可能是日

本机器人专家森政弘(Masahiro Mori)提出的“恐怖谷(Uncanny Valley)理论”。该理论表明,机器人的外观和行为跟人类越是接近,人们越容易产生积极的正面情感;但是,这种正面的情感到达一个峰值之后,随着相似度的提高,人们会对机器人产生恐怖的感觉,形成所谓的“恐怖谷”;当相似度持续上升到与人类更为接近的程度时,人们对机器人又会重新产生正面的情感。⁽⁸⁾“恐怖谷理论”虽然主要是针对人型机器人的,但对于非人型机器人同样具有一定程度的适用性。

受恐怖谷理论的启发与激励,许多学者对人们关于不同外观机器人的接受与喜爱程度进行了比较研究,因为机器人的应用范围非常广泛,不同应用领域的机器人的能力与外观差异巨大。从外观上看,机器人主要可以分为机器型机器人、动物型机器人、人型机器人等三大类别。对于主要使用对象是少年儿童机器人,机器型、动物型机器人可能更容易被接受。有学者让159名儿童对40张机器人图片进行评价,结果表明,儿童对机器人外观的分类,与成年研究人员的分类基本一致;而且,儿童把人型机器人看作是攻击性的和专横的,而把动物型机器人看作是更友好的和腼腆的。⁽⁹⁾与此类似,有学者对有轻微认知障碍的老人关于机器人外观的态度进行了研究,发现老人更喜欢小型的、某些特征介于人/动物和机器之间的机器人。不过,尽管参与者不太乐意面对类人机器人,但他们确实对某些拥有人类特点的、新颖的小机器人表现出积极的态度。⁽¹⁰⁾青年或中年人可能更倾向于喜欢人型机器人。通过对79名大学生对不同外观机器人的态度研究表明,大多数人喜欢人型机器人,只有少数人偏爱机器型机器人。而且,不同类型的机器人外观也影响人们对机器人性格的认知。⁽¹¹⁾另外,性别、种族、职业、文化程度等多种因素都会影响人们对机器人外观的态度,我们还需要更多深入细致的对比研究。

同时,机器人的外观并不是一个独立的因素,我们需要把机器人的外观与功能、工作环境等多种因素结合起来研究。机器人的外观、举止会深刻影响人们对机器人的认知,以及是否服从机器人指令的意愿。另外,人们通常希望机器人的外观和行为与其工作内容与环境协调一致。通过对127名来自不同年龄阶段、不同文化程度的调查对象的研究表明,人们希望在家里使用的机器人外观像动物,

而在公共场所使用的机器人外观更像人。同时,机器人的功能与外观的关系也非常紧密,比如,虽然人们更喜欢动物型的家用机器人,但如果家用机器人的功能是私人助手的话,人们还是倾向于选择人型机器人。⁽¹²⁾

因此,我们需要针对不同的使用对象、不同的机器人功能定位与应用环境,设计多样化的机器人外观。对于在公共场所使用的机器人,比如在博物馆、旅游景点的导游机器人,它需要面对各种不同的使用对象,这样的机器人的外观也应该尽可能地多样化。而且,要合理设计机器人与人类的相似程度,从而避免出现“恐怖谷”现象。

四、社会智能与道德能力

社会性是人类的基本属性,而且社会性在人类智能的发展中起着至关重要的作用。早在20世纪60年代就有学者研究发现,社会生活是灵长类动物智能产生的前提,并决定其智能的特点。也就是说,灵长类动物的社会生活为其提供了智能进化的环境。⁽¹³⁾我们可以把智能分为低级和高级两类,对过去曾经发生过的事情的模仿只需要低级智能,而只有在事件之间建立新的联系才可能发生的行为,则需要高级智能。灵长类动物拥有的高级智能是适应社会生活的复杂性而不断进化的结果。而且,起初是为了适应解决社会问题而发展起来的思维模式,又进一步丰富了人类和其他灵长类动物的行为。⁽¹⁴⁾这就是“社会智能假说(Social Intelligence Hypothesis, SIH)”。

不同学科对社会智能的界定有一定差别。从系统科学的角度看,人最重要的和最多的智能是在由众多个体构成的社会中进行各种活动时体现出来的,“协作”、“竞争”等等是人类智能行为的主要表现形式,这种社会智能也是人的社会思维涌现出来的群体智慧。⁽¹⁵⁾¹⁶⁵这是从社会群体的角度看待社会智能。与系统科学的角度相反,心理学主要从社会个体的角度研究社会智能。从心理学的角度看,社会智能一般指一种正确认识自我与他人、与人和谐相处的能力,包括言语、情感表达、社会行为调整、社会角色扮演等多种社会参与技能。⁽¹⁶⁾

可见,无论是从智能的起源、进化,还是从社会群体、个体的现实需要角度来看,社会智能的重要

性都是不言而喻的。如果 SIH 是灵长类动物智能进化的一种重要因素,那么它也可能是智能人工物进化的基本原理,而不仅限于生物体。因此,我们应该在个体化的环境中研究机器人之间“一对一”的互动,机器人应该从个体互动中获得经验,彼此认识,从而建立个体间的“个人”关系。⁽¹⁷⁾伦理道德的根本目的是协调与处理人际关系,所有的道德实践毫无疑问也都是在社会环境中进行的。人类的道德能力培养,是在家庭、学校和社会中,通过教育、模仿、实践与思考等多种途径发展起来的。对人类来说,伦理道德的社会性是不言而喻的。从社会智能理论与人类伦理的社会性来看,通过让人类与机器人,以及机器人之间进行互动,使机器人通过模仿人类的道德能力培养的模式来发展机器人的道德能力应该是一种基本途径。

从发展机器人的社会智能的角度建构机器人的道德能力,其必要性还具体表现在以下几个方面。首先,这是由人类道德的本质特征决定的。亚里士多德认为,适度是德性的特点,它以选取中间为目的。“德性是一种选择的品质,存在于相对于我们的适度之中。”⁽¹⁸⁾⁴⁷⁻⁴⁸也就是说,尽管人类拥有道德本能,但道德能力也是通过具体、反复的道德实践活动获得的。因此,通过具体的实践让机器人正确理解“适度”,也是建构其道德能力的必要手段。

第二,这是由人类道德生活的复杂性决定的。比如,尽管人们一般认为存在一些普遍性的道德原则,但是如何在具体的社会环境中进行应用,需要进行反复的道德实践;另一方面,道德原则总会有遇到例外的情况,原则与例外的严重冲突可能导致的所谓“道德困境”也一直是伦理学研究的热门话题,而道德困境的处理只能具体问题具体分析。

第三,通过社会智能的角度发展机器人的道德能力,也符合人工智能发展的基本趋势。目前,人工智能在规则比较明确的领域取得了巨大的成功,阿尔法围棋的强大能力就是有力的证明。但是,机器人在现实的社会生活环境中的灵巧性、适应性,跟人类相比还有很大的差距。要提高机器人的社会适应能力,机器人技术本身的发展进步固然是基础性的,但让机器人拥有更多的常识与社会智能也是至关重要的。让机器人在现实社会中学习掌握更多的关于伦理道德的常识与技能,是建构道德能力的关键因素。

第四,与从认知科学的角度发展机器人的道德能力的一致性。机器人伦理设计的认知科学进路主要是让机器人通过模仿现实社会生活中人类的道德认知,来建构机器人的道德能力。机器人模仿人类的道德认知,其核心内容就是根据社会智能的产生、发展及表现等方面的规律,使机器人在一定程度上模拟人类的社会智能。

五、人工情感与道德能力

人与人之间的情感是人类精神生活的重要组成部分,也是无可替代的。有学者认为,智能的人造物可以取代人的思维能力,尚不能取代人的情感。人的情感不应当由人造物来取代,不应当有“情物”。⁽¹⁹⁾⁴⁶⁹不过,人类很容易对非人的物品和动物产生情感,有时这种情感的强烈程度并不亚于人与人之间的情感。不少研究表明,老人、儿童在与机器人互动的过程中会感到愉快、舒适,机器人在一定程度上确实可以满足他们的情感需要。

从建构机器人的道德能力的角度看,赋予机器人一定的情感能力,是由道德与情感的密切关系决定的。一方面,情感影响人类的道德判断、道德行为。从伦理学的角度看,今天的社会伦理主要是一种以理性规则、道德义务为中心的伦理,但是,我们也要看到同情、怜悯和恻隐之心在人类道德生活中的基础性作用。⁽²⁰⁾¹⁷⁴斯宾诺莎(Baruch de Spinoza)说“善与恶的知识不是别的,只是我们所意识的快乐与痛苦的情感。”⁽²¹⁾¹⁷⁶从认知科学的角度看,理智与情感不是两个独立的过程,情感在人类推理的过程中发挥着重要作用。⁽²²⁾从心理学的角度看,目前的道德情绪研究充分证实了情感在道德行为中的重要作用。⁽²³⁾另一方面,人类的道德判断通常以情感、行为等方式反映出来,比如人们对好坏善恶的不同反应,与人们的喜爱、厌恶等情感存在密切的联系,具体表现在人的心理状态、面部表情与外在举止等方面。

如果机器人能够识别情感,并表现出跟人类类似的丰富情感,可能更容易被人们所接受,这也是人工情感研究的基本目标。人工情感是利用信息科学的手段对人类情感过程进行模拟、识别和理解,使机器能够产生类人情感并与人类自然和谐地进行人机交互的研究领域。人工情感并不是简单

地模拟人的某些情感表达方式和情感识别方式,而是为了使机器人具有像人一样的内在情感,真实地具有像人一样的情感表达能力、情感识别能力、情感思维能力和情感实施能力。⁽²⁴⁾²⁵

目前关于人工情感的研究大多是基于行为主义的思想,也就是让机器人表现出某种行为,使人们认为机器人拥有某种情感。正如情感计算的提出者皮卡德(Rosalind Picard)所说的那样“如何识别一个人的情感状态呢?根据他们的脸、他们的声音、他们的步伐或其他能传达他们感觉的姿态和行为举止。”⁽²⁵⁾¹⁷由于这种表现只是机器人的外在特性,与机器人的内在状态无关,所以有的学者认为这是虚假的情感,甚至是一种欺骗,是不道德的。不过,人类对情感的识别主要是基于他人的外在神态与行为,我们无法读取他人的内心状态。从另一个角度看,人们为了合理地说明某个实体(entity)在给定的社会环境中的行为,通常会在观察的基础上赋予它认知或情感的状态。⁽²⁶⁾

里夫(Byron Reeves)与纳斯(Clifford Nass)甚至认为,人们像对待真实的人那样对待电脑、电视与新媒体,即“媒体等同于真实生活”。通常关于媒体的研究总是关注于人们经验到的信息什么是真的,什么是假的。比如,计算机是否真的拥有智能?电视可以真正拥有一种社会角色吗?然而,诸如此类的问题却忽略了一个重要的事实,即那些像是真的东西通常比确实为真的东西更具影响力。⁽²⁷⁾²⁵³也就是说,与其关注机器人是否拥有表现外在情感的内部状态,还不如关注人类对机器人情感的认知。

当然,随着神经科学、认知科学研究的深入,我们可以在对人类情感的内在机制全面认识的基础上,开发出跟人类类似的、具有与情感相关的内在状态的机器人。有学者认为,把没有意识体验的行为看作情感是一种没有感觉(feeling)的情感,它无法区分行为的哪些方面是情感,哪些不是,而且也无法使我们建立一种情感的意识体验理论。感觉是人类情感的意识体验的一个重要方面,我们应该以建立一种拥有感觉能力的系统为起点,建构可以表现复杂情感的能力,以及表现灵活的、价值驱动的社会行为的能力。⁽²⁸⁾

我们可以把基于行为主义的人工情感称为“初级人工情感”,即使这种情感只是模仿人类情感的外在表象,也是非常有意义的,事实上这正是许多科学家正在进行的研究内容。拥有内在状态、与人

类产生情感机制类似的人工情感,可以称之为“高级人工情感”。初级人工情感通过人类情感的外在行为表现来模拟和识别情感,而高级人工情感研究方向恰好相反,是由机器人内部状态产生情感,由情感来引发各种外在行为。事实上这两种研究方向都是必要的,亦不矛盾,两者的研究成果可以相互借鉴。

为了更好地建构机器人的道德能力,我们需要深入研究情感与道德判断之间的具体机制。此外,我们还需要对人与机器人之间的情感互动进行评价。研究表明,采用人与人之间的情感评价模型来对人与机器人之间的情感进行评价可能是不够的。⁽²⁹⁾

结 语

综上所述,建构机器人的道德能力,需要从人与机器人道德地位的不对称性、社会智能、人工情感、机器人外观等方面出发,建构机器人的道德判断、道德行为、道德评价以及道德交流等能力,其相互关系如图1所示。关于伦理规范的选择及其实现、道德词汇的表征、道德认知、道德评价等问题,拙作《机器人伦理设计进路及其评价》与《机器伦理刍议》⁽³⁰⁾已有粗浅论述,限于篇幅,此处不再赘述。另外,本文的研究亦表明,哲学家在机器人道德能力建构中应当发挥重要作用。当然,要真正使机器人拥有一定的道德能力,我们还需要更多深入细致的经验研究。

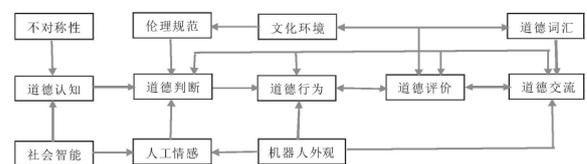


图1 机器人道德能力框架示意图

参考文献

- (1) 蔡志良、蔡应妹. 道德能力论 [M]. 北京: 中国社会科学出版社, 2008.
- (2) 罗国杰. 中国伦理学百科全书(伦理学原理卷) [M]. 长春: 吉林人民出版社, 1993.
- (3) Malle, B. Integrating Robot Ethics and Machine Morality: the Study and Design of Moral Competence in Robots [J]. *Ethics and Information Technology*, 2016, 18(4): 243-256.

- (4) 杜严勇. 机器人伦理设计进路及其评价[J]. 哲学动态, 2017(9): 84—91.
- (5) Lee, S. & Lau, I. Hitting a Robot vs. Hitting a Human: Is it the Same? [C]// *HRI' 11 Proceedings of the 6th international conference on Human - robot interaction*. Lausanne, 2011: 187 - 188.
- (6) Malle, B. et al. Sacrifice One for the Good of Many? People Apply Different Moral Norms to Human and Robot Agents [C]// *HRI' 15 IEEE International Conference on Human - Robot Interaction*. Portland, 2015: 117 - 124.
- (7) Biswas, M. & Murray, J. The Effects of Cognitive Biases and Imperfectness in Long - term Robot - human Interactions [J]. *Cognitive Systems Research*, 2017, 43: 266 - 290.
- (8) Mori, M. The Uncanny Valley [J]. *IEEE Robotics & Automation Magazine*, 2012, 19(2): 98 - 100.
- (9) Woods, S. Exploring the Design Space of Robots: Children's Perspectives [J]. *Interaction with Computers*, 2006, 18(6): 1390 - 1418.
- (10) Wu, Y. et al. Designing Robots for the Elderly: Appearance Issue and Beyond [J]. *Archives of Gerontology and Geriatrics*, 2012, 54(1): 121 - 126.
- (11) Walters, M. et al. Avoiding the Uncanny Valley [J]. *Autonomous Robots*, 2008, 24(2): 159 - 178.
- (12) Lohse, M. et al. Domestic Applications for Social Robots [J]. *Journal of Physical Agent*, 2008, 2(2): 21 - 32.
- (13) Jolly, A. Lemur Social Behavior and Primate Intelligence [J]. *Science*, 1966, 153(3735): 501 - 506.
- (14) Humphrey, N. The Social Function of Intellect [C] // Bateson, P. & Hinde, R. (eds.). *Growing Points in Ethology*. Cambridge: Cambridge University Press, 1976: 303 - 317.
- (15) 载汝为. 社会智能科学[M]. 上海: 上海交通大学出版社, 2007.
- (16) 刘在花, 许燕. 社会智力评估述评[J]. 上海教育科研, 2003(11): 33 - 37.
- (17) Dautenhahn, K. Trying to Imitate—a Step towards Releasing Robots from Social Isolation [C]// Gauthier, P. & Nicoud, J. (eds.). *From Perception to Action Conference*. Los Alamitos: IEEE Computer Society Press, 1994: 290 - 301.
- (18) 亚里士多德. 尼各马可伦理学[M]. 廖申白, 译. 北京: 商务印书馆, 2015.
- (19) 林德宏. 人与机器[M]. 南京: 江苏教育出版社, 1999.
- (20) 何怀宏. 伦理学是什么[M]. 北京: 北京大学出版社, 2015.
- (21) 斯宾诺莎. 伦理学[M]. 贺麟, 译. 北京: 商务印书馆, 2013.
- (22) 费多益. 认知视野中的情感依赖与理性、推理[J]. 中国社会科学, 2012(8): 31 - 47.
- (23) 王云强. 情感主义伦理学的心理学印证[J]. 南京师大学报(社科版), 2016(6): 128 - 135.
- (24) 魏斌, 等. 人工情感原理及其应用[M]. 武汉: 华中科技大学出版社, 2017.
- (25) 皮卡德. 情感计算[M]. 罗森林, 译. 北京: 北京理工大学出版社, 2005.
- (26) Duffy, B. Anthropomorphism and the Social Robot [J]. *Robotics and Autonomous Systems*, 2003, 42(3-4): 177 - 190.
- (27) Reeves, B. & Nass, C. *The Media Equation* [M]. Cambridge: Cambridge University, 1996.
- (28) Adolphs, R. Could a Robot Have Emotion? [C]// Fellous, J. & Arbib, M. (eds.). *Who Needs Emotions? The Brain Meets the Robot*. Oxford: Oxford University Press, 2005: 9 - 25.
- (29) Samani, H. The Evaluation of Affection in Human - Robot Interaction [J]. *Kybernetes*, 2016, 45(8): 1257 - 1272.
- (30) 杜严勇. 机器伦理刍议[J]. 科学技术哲学研究, 2016(1): 96 - 101.

On the Construction of Moral Competence in Robots

DU Yan - yong

(School of History and Culture of Science , Shanghai Jiao Tong University , Shanghai 200240 , China)

Abstract: It is important to build moral competence for robots in order to implement socially acceptable human - robot interaction because human are increasingly confronted with robots in contemporary society. This essay argues that there are similarities and differences between the construction of moral competence in robot and human. There are key elements to construct moral competence for robots, such as robot appearance, social intelligence, artificial emotion, besides paying attention to the asymmetry of moral status between robot and human.

Key words: robots; moral competence; social intelligence; artificial emotion

(本文责任编辑: 崔伟奇)