论机器人的道德承受体地位及其规范意涵

杨通进

摘　要：我们应当区分机器人享有的两种不同道德地位，即作为道德行为体的道德地位与作为道德承受体的道德地位。基督教、佛教与儒家都为我们认可机器人的道德承受体地位提供了终极关怀层面的理由。在规范共识层面，我们可以为机器人的道德承受体地位提供五个重要理据，即间接义务理据、自我建构理据、行为主义理据、人机共同体理据，以及潜在道德行为体理据。区分机器人的两种不同道德地位具有重要的实践与规范意涵：我们应当尊重机器人的人性、不能把机器人当成普通商品来对待、应当把人视为机器人的道德监护人、应当持续扩展道德关怀与伦理共同体的范围。

关键词：道德行为体；道德承受体；终极关怀理据；规范共识理据

中图分类号：B82　文献标志码：A　文章编号：2095-0047(2019)06-0014-20

机器人的道德地位问题是机器人伦理学的核心问题之一，也是近年来机器人伦理学研究的持续热点问题。因为机器人道德地位问题不仅涉及人对机器人负有何种道德义务的问题，而且还涉及一系列其他的社会、政治与伦理问题，如人类是否应当研制具有道德自主性的机器人，机器人对人类与其他机器人是否负有以及负有何种道德义务，人类是否愿意进入一个人(类)—机(器人)共处的后人类时代。2016年，欧洲议会法律事务委员会给欧洲议会提交的“机器人民事法律规则建议书”中明确建议欧洲议会考虑“赋予机器人特定法律地位的可能性，以便至少可以确认最复杂的自动机器人作为享有特定权利、负有特定义务的电子人身份”；2017年，沙特政府正式授予机器人索菲亚“沙特公民”身份。因此，确认机器人的

作者简介：杨通进，广西大学马克思主义学院哲学系教授，中国特色社会主义道德文化省部共建协同创新中心首席专家。

道德地位是摆在我们面前的重要伦理与法律问题。[①] 区分机器人道德地位的两种含义（作为道德行为体的机器人与作为道德承受体的机器人）是本文展开相关论述的基本策略。目前，许多学者都主张，作为具有充分道德自主性的机器人是完全意义上的道德行为体，并享有道德行为体的基本权利。本文对这一较强的命题持一种开放的态度，而主张一种较弱的命题：机器人享有作为道德承受体的道德地位，人类对机器人的行为存在道德意义上的正确与错误之分；这部分是源于具有道德能动性（moral agency）的机器人的行为表现，部分是源于人类对自身本质特征的确认。

一、何种机器人？哪种道德地位？

设想一下这样的场景：2118 年，我们已经进入人（类）—机（器人）共处的后人类时代。我们每天都要与各种机器人（护理机器人、服务机器人、助理机器人、伴侣机器人、家用机器人等）打交道。[②] 它们像其他人那样与我们正常交往：用清晰而明确的语言与我们交流与沟通 / 遵循基本的社会交往规则与规范 / 遵循基本的道德原则（至少不会蓄意做出不道德的行为）。从外在行为表现的角度看，它们与我们人类基本没有什么区别；它们的理性思维能力完全可以与我们人类媲美，在某些方面甚至超过人类；根据其外貌、语言与动作，我们几乎无法判断，站在我们面前的是机器人还是真人。人类与机器人唯一明显的区别是：人类是被生出来的，具有一副生理机体，而机器人是被制造出来的，以机械的躯体为主（某些机器人可能会拥有部分的生理机体）。

很明显，这种机器人只能是强人工智能意义上的机器人。这样的机器人不是某种只会“干活”不会说话的“哑巴”；相反，它们能够与人交流。某些机器人可能以某种或几种专门功能为主，但是，它们也不是现在的 AlphaGo 这样的专用机器人。它们更是像教师那样的教育机器人，像护士那样的护理机器人，像保姆那样的家用机器人，像节目主持人那样的文娱机器人，像战士那样的军用机器人，像异性伴侣那样的性爱机器人（甚至法律意义上的机器爱人），等等。它们基本上以通用机器人的面目出现在人们的公共生活与私人生活领域中，并与它们身边的人（包括其他机器人）形成良好的互动与协同关系。

这样的机器人可被称为社会机器人或社交机器人（social robots）。“社会机器人

① Susan Anderson，“Asimov’s ‘Three Laws of Robotics’ and Machine Metaethics”，*AI & Society*，Vol.22，No.4，2008，pp.477—493.

② 一些学者认为，我们会比人们所预期的更早地生活在这样一个世界，在其中“每个人都拥有一位机器人”。Declan Butler，“A World Where Everyone Has a Robot：Why 2040 Could Blow Your Mind”，*Nature*，Vol.530，No.7591，2016，p.398a.

就是具身化的自动行为体，它能够在社会的意义上与人交流与互动。社会机器人通过社会纽带实现与人的交往，它们展示了适应性的学习能力，能模仿各种情感状态。我们与它们的交往遵循社会行为模式，它们被设计出来是为了鼓励人机之间的情感联系。”① 本文所讨论的机器人即这类以“强人工智能”为基础、具有人类外表与智能、深度介入人类的公共生活与私人生活的智能机器人或人工智能行为体（artificial intelligent agents）。② 某些这样的机器人已经以“雏形”或“初级”的形式出现在人们的生活中，如医院的护理机器人，陪伴老人或儿童的“看护机器人”或“机器宠物”，以及充满争议的性爱机器人（它们已经在欧洲某些红灯区“提供服务”）与军用机器人（联合国人权高级委员会已经提议禁止研发军用机器人）。随着机器人技术的提高与完善，社会机器人将在未来逐步以更完整的形式参与人们的社会生活与私人生活。

这样的机器人将会具有或者能够享有何种道德地位呢？

道德地位（moral standing，moral status）指的是一个道德主体在道德共同体中占有的位置、享有的资格、或承担的角色。“拥有道德地位就是在道德上应给予关怀，或拥有道德资格。拥有道德地位的实体就是这样的实体，即道德行为体（moral agent）对它负有、或能够负有道德义务。如果一个实体拥有道德地位，我们就不能为所欲为地对待它；在进行慎思的时候，我们有道德义务赋予它们的需要、利益或福祉一定的分量。”③ 道德地位意味着某种道德资格（moral entitlement），即做出某种行为与获得某种待遇的资格。“一个实体拥有道德地位，当且仅当它能够被错误地对待。因此，当且仅当一个实体拥有道德地位时，一个道德行为体的某些行为对该实体所产生的影响才能直接（即不取决于这些行为对其他实体的影响）或间接地成为对这些行为进行道德评估的决定因素。”④ 因此，道德地位与一个道德主体在人们

① Kate Darling，“Extending Legal Protection to Social Robots：The Effects of Anthropomorphism，Empathy，and Violent Behavior Towards Robotic Objects”，https：//paper s.ssrn.com/abstr act=20447 97.

② 著名的人工智能专家库兹韦尔认为，人工智能发展的奇点将在 2045 年出现。蔡斯认为，强人工智能“最有可能在本世纪下半叶出现，而它一旦出现，就很有可能在几个月甚至几周之内演化为超级智能，不太可能需要几年之久”（蔡斯：《人工智能革命：超级智能时代的人类命运》，张尧然译，北京：机械工业出版社 2018 年版，第 123、173 页）。人工智能乐观论者认为，整合了各种功能的“超级智能机器人”（通用机器人）在所有的方面都将超越人类，而“悲观主义者的立场正在退缩：在每一个单独的领域，［机器人］都有可能超越人类，不一定会在整体上全面超越人类”（尼克：《人工智能简史》，北京：人民邮电出版社 2018 年版，第 225 页）。“日渐加速的人工智能创新已经让莱斯大学的摩西·瓦迪等计算科学家预言，最迟在 2045 年，在一些非常重要的工作中，人类将退出历史舞台。更有甚者，有人提出，计算机将快速进化，最终将用一代人或最多两代人的时间超越人类智能”（马尔科夫：《人工智能简史》，郭雪译，杭州：浙江人民出版社 2017 年版，第 10 页）。

③ Mary Anne Warren，*Moral Status*：*Obligations to Persons and Other Living Things*，Oxford：Clarendon Press，1997，p.4.

④ Justin Sytsma and Edouard Machery，“The Two Sources of Moral Standing”，*Review of Philosophy & Psychology*，Vol.3，No.3，2012，pp.303—324.

的道德慎思中占有何种地位有关。

在道德共同体中，存在着两种不同类型的道德地位：作为道德行为体（moral agent）的道德地位与作为道德承受体（moral patient）的道德地位。[①] 作为道德行为体的道德主体指的是具有道德能动性（moral agency）的主体，它们能够进行道德推理，它们的行为遵循道德原则的引导；它们具有道德自主性（moral autonomy），需要对自己的行为承担道德责任。“道德行为体对那些拥有道德地位的实体负有义务或责任。”[②]

道德承受体指的是具有道德承受性（moral patiency）的主体，它们有资格获得道德关怀；道德行为体对它们负有道德义务与责任。“不仅像拥有我们这样的意识（consciousness）的存在物是道德承受体，非人类动物也是道德承受体，在进行道德慎思时，我们有责任把它们的利益考虑进来。”[③] 更重要的是，我们之所以有道德义务这样做，不是因为对它们的保护对我们或其他人有利，而是因为它们的需要在我们的道德慎思中拥有道德上的重要性。[④] 道德承受体最重要的特征是，它们没有能力进行道德慎思，不能做出符合道德要求的选择，但是它们有资格获得道德关怀。如果说道德行为体是道德的生产者与提供者（moral producer），那么，道德承受体就是道德的接受者（moral receiver）与消费者（moral consumer）。

很明显，道德承受体的范围要大于道德行为体。所有的道德行为体都是道德承

① 在当代英语学术界，尤其是应用伦理学界，moral agent 与 moral patient 是两个具有紧密联系的概念。moral agent 和 moral patient 都是 moral subject，因此，不宜把 moral agent 译为“道德主体”（moral subject），或把 moral patient 译为“道德客体”（moral object）。如果我们把道德共同体比喻为某个餐厅或商场，那么，moral agent 类似于其中的服务员，moral patient 则类似于其中的顾客。moral agent 提供的道德服务即“道德关怀”；他们有义务对所有的 moral patient 提供这种服务，不得对后者采取歧视态度。moral patient 不能提供道德服务，但是，他们有资格获得 moral agent 提供的道德服务。基于这一理解，笔者早年曾把 moral agent 与 moral patient 分别译为“道德代理人”与“道德顾客”。此外，在环境伦理学的语境中，人类是唯一具有道德思维能力的存在者，他们代表地球上的所有存在物提出某种能够获得理性辩护的道德体系，并向那些有资格获得道德关怀的“道德承受体”（或道德顾客）提供道德服务。因此，在环境伦理学的语境中，把 moral agent 与 moral patient 分别译为“道德代理人”与“道德顾客”不会引起太大的歧义。（有学人曾对笔者的理解提出异议。参见姚晓娜：《moral agent 是道德代理人吗》，载《道德与文明》2010 年第 1 期。）但是，在机器人伦理学的语境中，机器人也可能是 moral agent。因此，本文遵循多数学人的惯例，把 moral agent 译为“道德行为体”，并相应地把 moral patient 译为“道德承受体”。王小红等把 agent 译为“智能体”、把 artificial moral agents 译为“人工道德智能体”，亦值得参考（参见瓦拉赫、艾伦《道德机器：如何让机器人明辨是非》，王小红等译，北京：北京大学出版社 2017 年版）。需要提醒注意的是，moral agent 与 moral agency（道德能动性，道德思维与道德行为能力）密不可分，而 moral patient 与 moral patiency（道德受动性，道德承受能力）之间也有紧密的联系。机器或机器人的 moral patiency 问题在国内学界尚未得到充分探讨。

② Justin Sytsma and Edouard Machery, “The Two Sources of Moral Standing”.

③ John Basl, “Machines as Moral Patients We Shouldn’t Care about (yet): the Interests and Welfare of Current Machine”, *Philosophy and Technology*, Vol.27, No.1, 2014, pp.79—96.

④ Mary Anne Warren, *Moral Status: Obligations to Persons and Other Living Things*, Oxford: Clarendon Press, 1997, pp.9—10.

受体，但是，并非所有的道德承受体都是道德行为体。同时，所有的人都是道德承受体，但是，只有那些具有正常思维能力的成年人才是完全的道德行为体，才需要对其行为承担全部道德责任。有些人（如青少年）是不完全的道德行为体，只对其行为承担部分道德责任。有些人（如婴儿、智障人士）则不是道德行为体，不对其行为承担道德责任；他们是无需提供道德服务，但却有资格享有道德服务的道德承受体。

人之外的存在物是否享有作为道德承受体的道德地位？这是随着当代环境伦理学的深入发展才逐渐变得明晰起来的问题。根据当代环境伦理学的研究，几乎所有重要的宗教伦理学对这一问题的回答都是肯定的。基督教、犹太教、伊斯兰教、佛教等宗教至少都从神学的角度为某些动物、植物或天使（天神）的道德承受体地位提供了某些证明。现代世俗伦理学已经达成的一个基本共识是：具有苦乐感受能力的动物也是道德承受体，人类对它们的行为需要接受道德的约束。高等动物的这种道德承受体地位还在相关的动物保护法，尤其是动物福利法中得到了确认。

因此，某些非人类存在物有资格享有作为道德承受体的道德地位。那么，人之外的存在物（不论是自然存在物还是人工存在物）是否也有资格享有作为道德行为体的道德地位呢？这取决于我们如何理解享有作为道德行为体地位的必要条件。毫无疑问，要成为道德行为体，至少要具备理解道德原则的能力与根据道德原则调节自己行为的能力。但是，是否还需要其他条件，如自由意志、自我意识、意图等，学术界则存在争议。①

在讨论机器人的道德地位时，我们需要明确区分机器人的两种不同道德地位：作为道德行为体的机器人与作为道德承受体的机器人。人类与这两类具有不同道德地位的机器人之间的伦理关系是有区别的。本文接下来的部分将分别探讨机器人享有道德承受体地位的理据及其规范意蕴。

二、作为 moral patient 的机器人：终极关怀层面的理据

如前所述，当代学术界关于道德行为体与道德承受体的概念区分（特别是关于道德承受体之范围的扩展），是随着当代环境伦理学关于人对人之外的非人类存在物的道德义务的探讨而逐渐进入人们的学术视野的。近现代西方的主流伦理学（尤其是康德的伦理学）的一个未明言的前提是，只有人类才享有道德地位，才有资格获得道德关怀。主流伦理学在人类那里寻找到许多独特的属性（properties）——诸如人能够感受苦乐、使用语言、拥有理性、从事道德推理，等等——来为人类的这

① John Basl, “Machines as Moral Patients We Shouldn’t Care about (yet): The Interests and Welfare of Current Machine”.

种独特的道德地位提供辩护。然而，要在人类身上寻找到某种所有的人类个体都拥有、人类之外的任何自然存在物都不拥有的属性是不可能的。有些属性（感受苦乐）虽然所有的人类个体都拥有，但是，人类之外的其他高等动物也拥有；有些属性（道德推理能力）虽然只有人类拥有，人类之外的其他自然存在物不拥有，但是，并非每一个人类个体（如婴儿、严重智障人士、精神病患者、植物人等）都拥有。① 因此，即使是在人类内部，也并非每一个人都具备作为道德行为体的道德地位（如婴儿与严重智障人士）。但是，我们不能因为一个人不是道德行为体、不能履行道德义务就否认他享有作为道德承受体的道德地位。这些缺乏道德理性（或只具备部分道德理性）的人之所以享有道德承受体之道德地位、有资格获得道德行为体的道德关怀，乃是由于他们能够感受苦乐，或拥有自己的利益，或是道德共同体的成员。因此，当代环境伦理学扩展了道德承受体的范围，使道德行为体的道德关怀对象从人扩展到了人之外的动物（动物解放 / 权利论）、植物（生物中心主义）乃至作为整体的生态系统（生态中心主义）。②

当代环境伦理学对人之外的自然存在物之道德承受体地位的确认，为当代机器（人）伦理学继续扩展道德关怀之范围、确认机器（人）之道德承受体地位提供了必要的“智识桥梁”。依据当代环境伦理学与机器（人）伦理学的研究成果，我们至少可以从两个层面（终极关怀层面与规范共识层面）为机器人的道德承受体地位提供辩护。

世界上存在着许多关于世界的本源与本质、人性、生活的目的与意义的系统化的学理体系，即系统化的哲学与宗教，罗尔斯称之为完备性学说（comprehensive dorctrines）。任何一种理性化了的完备性学说至少包含两个层面的内容，即终极关怀层面的学说与公共规范层面的学说。终极关怀层面的学说涉及人们对生活的意义与世界的终极实在的理解，它们往往以某种历史悠久的文化传统和宗教信仰的形态表现出来。它们为人们建构并理解人与世界的关系，以及人生的目的与意义提供基本的框架与参照系统。限于篇幅，本文仅简单探讨基督教、佛教与儒家思想从终极关怀层面对于机器人之道德承受体地位所可能提供的辩护思路。

（一）上帝与机器人

“人们通常认为，基督教对技术持一种反对的态度。但是，实际情况并未总是如此。”③ 在中世纪，天主教是永动机（automata）研制计划的主要资助者，并设计和

① 罗尔斯顿：《环境伦理学：大自然的价值以及人对大自然的义务》，杨通进译，北京：中国社会科学出版社 2000 年版，第 85—105 页。

② 参见温茨：《现代环境伦理》，宋玉波、朱丹琼译，上海：上海人民出版社 2007 年版；Louis Pojman，*Global Environmental Ethics*，Mountain View，CA.：Mayfield Publishing Company，2000.

③ Gabriele Trovato，Loys De Saint Chamas，et al.，“Religion and Robots：Toward the Synthesis of Two Extreme”，in *International Journal of Social Robots*（May 2019）. https：//doi.org/10.1007/s12369-019-00553-8.

制造了机械化的天使与能够喷火的魔鬼。[①]16 世纪，教廷还制造了一个能够敲钟的机械化的修道士（monk），献给西班牙的菲利普二世。这个机械修道士可被视为最早的“牧师机器人”，他的外表形象是圣迭戈（San Diego），木质结构，能够做出祈祷等动作。[②] 中世纪基督教的这类实践为基督教参与当代的“机器人文化”建设提供了重要的思想资源。在基督教的话语体系里，我们至少可以发现三种认可机器人道德承受体之道德地位的思路。

1. 生态神学进路（ecotheology approach）

在价值观层面，传统的主流基督教虽然是人类中心主义的，但是，基于其潜在的“上帝中心主义”背景，当现代环境伦理学对人类中心主义发起挑战时，现代基督教很快就从生态神学的角度做出了回应。在生态神学家看来，上帝关怀“自然中的所有事物，有生命和无生命的，人类和人类之外的，动物和植物，枯枝，空气，水，岩石：一切事物”[③]。上帝的目的是拯救整个宇宙；不仅人的灵魂，所有的事物都是上帝拯救的潜在对象；甚至原子与离子也属于获救者的行列。[④]因此，“人们如果毁灭或从根本上改变了岩石、树木、空气、水、土地、动物的存在状态，那他们就是超越了其神圣权利的界限——就如同他们谋杀其同胞时那样”[⑤]。

与传统的神学相比，生态神学有两个相当引人注目的特点。第一，人们对人类之外的存在物的行为受到某些伦理原则的约束；换言之，上帝创造的人类之外的存在物都拥有内在价值，享有道德承受体的道德地位。第二，生态神学把枯枝、岩石、空气、水、土地等无生物也纳入了道德承受体的范围。既然枯枝等无生物都是人类道德关怀的对象，那么，把在外貌、信息交流、行为方式，甚至情感表现等方面都与人非常接近的机器人纳入道德关怀的范围并接受其道德承受体的道德地位，在逻辑上就不会存在太大的障碍。事实上，根据当代过程神学与万有在神论的观点，世界在上帝之中，同时，上帝也在世界之中；上帝既超越世界又内在于世界；“上帝不呈现之处无物存在”[⑥]。如果我们把机器人理解为万物之一，那么，上帝的神性当然也应当在机器人之中；上帝不在机器人之中，机器人也不会存在。此外，如果我们把人的理性理解为上帝的理性的延伸，那么，我们也可以把机器人理解为是上帝通过人的理性间接地创造出来的。这样一来，也可以说，在终极意义上，机器人是上帝创造的。

---

① Jessica Riskin, *The Restless Clock: A History of the Centuries-long Argument over What Makes Living Things Tick*, Chicago: University of Chicago Press, 2016.

② Jessica Riskin, “Machines in the Garden”, in Mario Biagioli and Jessica Riskin (eds.), *Nature engaged: Science in Practice from the Renaissance to the Present*, Washington: Palgrave Macmillan US, 2012, pp.229—248.

③ 转引自纳什：《大自然的权利：环境伦理学史》，杨通进译，青岛：青岛出版社 1999 年版，第 121 页。

④ 同上书，第 119、130—131 页。

⑤ 转引自纳什：《大自然的权利》，第 131 页。

⑥ 黄铭：《过程思想及其后现代效应——科布神学思想研究》，北京：宗教文化出版社 2010 年版，第 82 页。

2. 逻各斯神学进路（logos theology approach）

传统的基督教认为，人是由肉体与灵魂两部分构成的。肉体与灵魂是两种完全不同的实体。肉体有生有灭，灵魂则是永恒的。灵魂被植入肉体中，但人死后灵魂就脱离肉体继续存在下去。灵魂是使所有的生命（不管是人、动物还是植物）具有活力、成为生命的依据。根据正统的基督教神学，我们似乎很难把基督教的肉体与灵魂理论扩展应用于机器人，因为，机器人的肉体是机器，而非有机的生命形式，无法作为接纳灵魂的物质基础。

尽管如此，在基督教中还是存在着另一种神学传统，即约翰的逻各斯神学（Johannine logos theology），这种神学能够对机器人革命作出积极的回应。这种神学把逻各斯（logos）或“道”（the Word）视为世界的本源；《新约·圣经》的《约翰福音》开篇即宣称：“太初有道（the Word），道与神同在，道就是神。”有的学者把这里的“道”理解为logos，并依据希腊哲学家赫拉克利特（前535—前475）的学说，把logos理解为理性。logos不仅创造了一个智能的宇宙，还是宇宙运行的原理。[①] 根据《新约》的经文，“logos被认同为耶稣基督”[②]。这样一来，我们就可以把机器人的源代码视为logos的一部分，视为理性的延伸。“基督教教会的成员应当与他们的护理机器人、看护机器人或伴侣机器人肩并肩地共同敬畏logos——一种神性的算法（divine algorithm），它弥漫在他们的世界中，即人的世界与数字世界中。”[③]还有的学者认为，基督教视域中的“肉体”（flesh）一词可以适当地扩展，使之“包括有机体之外的其他种类的物质”，从而把人体与机器人的躯体也视为可接纳logos的物质。[④] 总之，根据逻各斯神学进路，我们可以把神理解为“道”，把“道”理解为逻各斯，把逻各斯理解为算法，并把肉体的范围加以扩展，从而实现人与机器人的认同与统一。

3. 启示论人工智能进路（apocalyptic AI approach）

根据基督教的启示论，人类生活于其中的充满了苦难历史的世界将随着弥赛亚的降临而终结，那时，所有信仰上帝的人都将复活，在至福状态中生活一千年，然后升入永恒的天国，永享完美的幸福。当然，为了升入天国，每个人都需要一个新的身体；这个身体是由上帝提供的完美的身体；这个身体是永恒的、不朽的。先知以赛亚告诉世人，上帝将“造新天新地”（《旧约·以赛亚书》76：17）。使徒保罗宣称：“血肉之躯不能承受神的国，必朽坏的不能承受不朽坏的……死人要复活，成为

① Vladimir de Beer, “The Cosmic Role of the Logos, as Conceived from Heraclitus until Eriugena”, *Philosophy & Theology*, Vol.27, No.1, 2015, pp.3—24.

② James McBride, “Robotic Bodies and the Kairos of Humanoid Theologies”, *SOPHIA*（2017）, https: // doi.org/10.1007/s11841-017-0628-3.

③ Ibid.

④ Gâbor Ambrus, “In the Beginning Was the Word: Theological Reflections on Language and Technical Media in the Context of the Gospel of John”, *Acta Universitatis Carolinae Theologica*, Vol.6, No.2, 2017, pp.135—151.

不朽坏的，我们也要改变。”（《新约·哥林多前书》15：50、52）

沿着启示论的思路，一些著作家与神学家认为，以生物有机体形式存在着的人类有着极大的局限性，如较短的生命周期，力量与忍耐力也非常有限；被局限在身体中的心灵艰难地学习，缓慢地思考，知识的传递有缺陷，记忆不仅有限，还往往不准确。随着机器人技术、纳米技术与基因工程技术的聚合，人类将研制出完美的机器人（类似于上帝提供给人类的新的完美身体）；那时，人们的意识将被下载并存储在机器人中。人就是他 / 她自己的意识；一旦这种意识被在不同的机器之间拷贝与复制，人不仅将克服作为生物物种的人类的所有局限，而且还变成了真正不朽的存在。那时，人就是机器人，机器人就是人；人与机器人之间的界限完全消失。① 总之，通过认同于具有全新形态与能力的智能机器人，启示论人工智能论者不仅实现了人与机器人的统一，还把机器人视为超越并替代生物学人类的新人类 / 后人类 / 超人类物种。

（二）佛与机器人

在佛教的话语体系中，某种类似于万有在神论的“佛性遍在说”可以为机器人之道德承受体地位提供支持。根据佛性遍在说，万物皆有佛性。《金刚经》明言：“万法是真如，真如是万法”；“真佛体在一切法”；“我及众生皆有此性故名佛性，其性遍造、遍变、遍摄。”我国天台宗的湛然大师（711—782）也明确指出，“无情有性”，即不仅有情识的动物拥有佛性，没有情识的植物、无机物、山川、草木、大地、瓦石等也具有佛性。中国的禅宗强调，“郁郁黄花无非般若，青青翠竹皆是法身”。日本的禅宗亦认为：“山脉和河流，石头与树木，花与鸟全都具有开悟以及一起踏上成佛之路的潜质……自然本身就是佛。”② 同时，佛教承诺的是一种整体主义的本体论与宇宙论，认为每一个微小的事物中都完整地包含着大千世界的本质、本性，极其微小的芥子、毛孔与无限的宇宙是相互含摄的。③

根据佛教的上述观点，作为万法之一的机器人无疑也具有佛性。曾提出“恐怖谷理论”的日本机器人工程师森政弘（Masahiro Mori）也认为，机器人不仅内在地就拥有佛性（Buddha-nature），而且还拥有成为佛的潜能。④ 此外，有学者把佛教“八识”中的最后一“识”即“阿赖耶识”理解为一种独立的精神本体，某种可以持续的意识状态，它先于个人的出生而存在，在人出生的时候进入人体，在人们死亡的时候又离开

① Robert Geraci，“Apocalyptic AI：Religion and the Promise of Artificial Intelligence”，*Journal of the American Academy of Religion*，Vol. 76，No. 1，2008，pp.138—166.

② 史蒂夫·奥丁：《日本人有关环境伦理的自然观与奥尔多·利奥波德的环保美学》，载安乐哲等主编：《佛教与生态》，何则明等译，南京：江苏教育出版社 2008 年版，第 92—108 页，引文见第 105 页。

③ 魏德东：《佛教的生态观》，载《中国社会科学》1999 年第 5 期。

④ Masahiro Mori，*The Buddha in the Robot*：*A Robot Engineer's Thoughts on Science and Religion*，Tokyo：Kosei Publishing Company，1989.

死者的身体。[①] 如此一来，佛教徒就可以宣称，人工智能技术也“可以把那种意识储存在无生命的客体中，而不是仅限于储存在有机体上；这种观点与基督教的下述观点并无多大区别：上帝（作为圣灵）寄居在所有的事物甚至餐具中”[②]。既然机器人拥有佛性，它们当然就享有佛性拥有者的道德地位（至少是作为慈悲与怜悯的对象）。

（三）儒家视野中的机器人

儒家思想强调人在天地万物中的特殊地位，强调人是道德价值的践行者与维护者。但是，儒家并不是狭隘的人类中心主义者。儒家强调“人禽之辨”，主要是为了凸显人作为道德行为体的地位，并不是为了把动物当成单纯的工具来对待；相反，儒家强调人要对动物有怜悯之心。换言之，儒家在凸显人作为道德行为体之道德地位的同时，并不否认动物作为道德承受体之道德地位。

事实上，儒家不仅要求人要怜悯动物，而且要求人成为“天民”，即把人当作宇宙中唯一的道德物种来定位，并要求人自觉地承担起对宇宙万物的关怀与看护义务。孟子要求人“上下与天地同流”。《中庸》要求人“成己成物”，“赞天地之化育”。王阳明在总结儒家“万物一体说”的基础上明确指出，“大人者，以天地万物为一体者也”，人的“休惕恻隐之心”不仅要扩展到其他人身上，还要扩展到“鸟兽”“草木”与“瓦石”身上。真正有道德的人对鸟兽要有“不忍之心”，对草木要有“怜悯之心”，对瓦石要有“顾惜之心”。“君臣也，夫妇也，朋友也，以至于山川鬼神鸟兽草木也，莫不实有以亲之，以达吾一体之仁，然后吾之明德始无不明，而真能以天地万物为一体矣。”[③] 既然山川、草木、瓦石等无机物都可以成为人类仁爱关怀的对象，享有道德承受体的地位，那么，对儒家来说，把这种仁爱关怀扩展到在很多方面都与人相似或相同的机器人身上，并使后者享有道德承受体的道德地位，就不应该存在任何逻辑上的困难。

三、作为 moral patient 的机器人：公共规范层面的理据

不同的完备性学说在终极关怀层面存在着相当大的差距。一种完备性学说的信奉者往往难以接受或理解另一种完备性学说的终极关怀理念。但是，在公共规范层面，不同的完备性学说之间却存在许多共识，因为公共规范涉及的是公共生活的基本规范、理想的社会制度、相互交往的基本原则等。而这些规范是每一个社会的人们都需要的。由于人们的公共生活存在许多共同性，因而，不同文化与信念传统

① Laurence Tamatea, “Online Buddhist and Christian Responses to Artificial Intelligence”, *Zygon*, Vol.45, No.4, 2010, pp.979—1002.

② Noreen Herzfeld, *In Our Image: Artificial Intelligence and the Human Spirit*, Minneapolis: Fortress, 2002, p.90.

③ 《王阳明全集》，上海：上海古籍出版社 1992 年版，第 968、969 页。

的人们也会遵循许多大致相同的规范（如尊敬老人、不伤害他人、反对杀人与偷盗等）也就不足为怪了。随着不同文化传统的人们之间在政治、经济与文化方面的交往在全球层面日益深入，人们在全球层面达成的公共规范也在日益增加。因而，前述关于机器人之道德承受体地位的终极关怀层面的理据，虽然往往只对特定完备性学说的信奉者有效，但是，在公共规范层面，我们还可以发现许多支持机器人之道德承受体地位的理据。这些理据能够获得不同完备性学说之信奉者的认可与接受，或者他们至少难以合理地加以拒斥。

（一）间接理据

所有文明的现代社会都反对虐待动物。一些人把反对虐待动物视为人对动物负有的一种直接义务。例如，根据功利主义学说，凡带来痛苦的行为就是恶的，带来快乐的行为就是善的；动物能够感受苦乐，因而给动物带来痛苦的行为就是恶的或错误的。① 当然，更多的人是从间接义务论的角度来论证不虐待动物的义务，即我们不伤害动物不是因为伤害动物本身是错误的，而是因为对动物的伤害会导致对他人的伤害，而伤害他人是错误的。阿奎那（1225—1274）较早地表达了这种观点："基督教的任何教义看起来都禁止我们去残忍地对待那些不能开口说话的动物，比如，禁止杀死幼鸟；这或者是因为人们会把这种思维转移到残忍地对待其他的人——既然会对动物残忍，也就会对他人残忍；或者是因为对动物的伤害会导致对他人的现世伤害——或是实施行为，或是其他暴行。"② 阿奎那的这两个理由其实可以归结为一个理由，即对动物的残忍会导致对他人的伤害。康德也明确提出了类似的观点，"对动物残忍的人在处理他的人际关系时也会对他人残忍。我们可以通过一个人对待动物的方式来判断他的心肠是好是坏。"③

根据上述思路，我们完全可以推出相似的结论：我们拥有不虐待机器人的义务，因为，对机器人的虐待行为会导致我们虐待其他人。人们对网络暴力游戏的一个普遍担忧是，这种游戏会让游戏者养成"暴力思维"的习惯，并对现实中的暴力行为变得麻木不仁。网络游戏中的暴力还只是一种"虚拟的暴力"。可是，如果人们在现实生活中以虐待机器人为乐，那么，这种虐待就不是一种"虚拟的虐待"，而是"实实在在的虐待"，是对虐待行为的实施、练习与强化。如果网络中的"虚拟暴力"都值得我们担忧，那么，人机交往中的这种"真实的虐待"就更值得我们担忧。因此，为了使人们避免养成虐待他人的思维方式与行为习惯，我们必须禁止对机器

① 辛格：《所有动物都是平等的》，江垭译，载《哲学译丛》1994 年第 5 期；另载杨通进、高予远主编：《现代文明的生态转向》，重庆：重庆出版社 2007 年版，第 154—166 页。

② 阿奎那：《理性生物与无理性生物的差别》，载辛格、雷根编：《动物权利与人类义务》，曾建平、代峰译，北京：北京大学出版社 2010 年版，第 7—10 页，引文见第 10 页。

③ 康德：《对于动物的责任》，载辛格、雷根编：《动物权利与人类义务》，第 25 页。

人的虐待行为。

人们或许会辩解说，如果我是单身，我购买了一位伴侣机器人，他/她永远都只在我的私人空间（我的私人住宅）中与我交往，只是我的私人财产。我与他/她的关系只涉及我自己，不涉及任何第三者。根据密尔的自由原则，我的行为中只有涉及他人的部分，才需要遵循相关的社会规范。我与我的机器人的关系，完全不涉及他人。因此，我想如何与我的机器人相处，社会完全无权干预。一般来说，密尔的自由原则是无可非议的。但是，对自由原则的运用并非毫无边界。首先，这一原则不适用于儿童。就像我们对儿童的个人自由施加某些限制那样，我们也可以对人们虐待机器人的个人自由施加某些限制。其次，即使在一个自由至上的社会中，人们也会认为某些行为是不可接受的，即使这些行为发生在私人空间，如某些性行为。最后，哪怕是绝对的私人财产，我们也有义务不予浪费；地球上的资源是有限的；挥霍财富无异于挥霍地球上的有限资源；因此，哪怕是虐待作为财产的机器人也属于暴殄天物的不当行为。更何况，智能机器人不是椅子、桌子那样的私人财产，它们有着人的外表与行为方式。同时，它们的运行依赖强大的公共网络与云数据系统；每一台智能机器人都与公共网络系统保持着紧密的海量数据联系，它们在某种意义上已然构成公共网络系统的一部分，属于社会公共资源的一部分。因此，智能机器人已经不属于传统意义上的私人财产；用户对其拥有的智能机器人的自由行为不受密尔的个人绝对自由原则的保护。“所以，即使对最为宽容的社会来说，对虐待机器人的行为施加某些限制也是言之成理的。”①

（二）自我建构理据

康德虽然主张，“我们对于动物的责任仅仅是我们对于人类的间接责任”，但是，在论证这一观点时，除了诉诸对他人的伤害理据，他还指出，履行对动物的义务，是我们的人性的展现。如果一条狗长期服务于它的主人，那么，“当这条狗老了不能再为它的主人服务时，它的主人应该照顾它，直到它死去”。相反，“如果一个人因为他的狗没有能力再为它服务而把它给杀了……[那么，]他的行为就是不人道的，是对他自身的人性的损害，而他有义务向他人展现出这种人性”②。在这里，不承担对动物的义务的行为所涉及的，就不是对他人的间接伤害，而是对人性本身的损害。所以，康德实际上是把关心动物当作人性的内在要素加以理解和建构的。换言之，一个正常的理性的人必须要把对动物的关怀当作一种直接的义务来认可与承诺。这关乎的不仅仅是动物的“应得”，更关乎人性的完整。关心动物是人性的内在情感，一个人如果“不想扼杀他的人性情感（human feelings），他就必须要以仁

① Blay Whitby，“Sometimes It's Hard to be a Robot：A Call for Action on the Ethics of Abusing Artificial Agents”，*Interacting with Computers*，Vol.20，No.3，2008，pp.326—333.

② 康德：《对于动物的责任》，载辛格、雷根编：《动物权利与人类义务》，第25页，译文略有改动。

慈（kindness）的方式对待动物”[①]。同理，如果人们虐待为他们提供服务、能够与他们交流、并能遵循人类行为规范的机器人，那么，这也是对人们的人性的一种损害。以这样一种形象展现在他人与机器人面前的人，建构与认可的是一种有缺陷的、不完美的人格。

英国学者布里森认为，机器人应当是奴隶，即人类的仆人；它们应当被制造、被营销，并在法律上被当作奴隶来对待。[②]根据前述康德关于人类自我建构的观点，布里森此处论点的错误之处在于，通过把机器人建构为人类的奴隶，布里森建构了一个错误的人类自我，即作为奴隶主的人类自我；布里森把奴隶主的属性建构成了人性的内在要素。奴隶主拥有的是一种等级性的、压迫性的、傲慢性的人格特征。如果人们在与机器人交往时展现的是这样一些人格特征，那么，我们由此建构的就是一个具有严重缺陷的人性概念。要想把人类建构成一个理性的、有美德的道德物种，我们就不应当把在外表与行为等方面都与人相似的机器人当作奴隶来对待。我们的心灵不可能被“分区”为两个互不影响的“磁盘”：一个“磁盘”按照平等公民的逻辑运行，另一个“磁盘”按照主人—奴隶的逻辑运行。因此，我们在虐待机器人时，所伤害的就不仅仅是机器人，我们还扭曲了自己的人格，减损了人性本身的光辉。

（三）行为主义理据

康德在讨论人对动物的义务时曾提到，一条狗对它的主人提供的服务，类似于人的服务；这条狗的服务应当得到回报，即“当这条狗老了不能再为它的主人服务时，它的主人应该照顾它，直到它死去”。康德这里提到的实际上是人与动物之间的相互性或互惠性（reciprocity）。在实际的生活中，我们都会对那些长期为我们提供服务的家畜或伴侣动物抱有某种感情；我们会把它们当作值得我们从道德上加以关怀的对象来对待；人们还会为那些“义犬”“勇敢的战马”等树碑立传。机器人给我们提供的服务无疑比这些动物还要周到细致，而且，它们还能遵循人类交往的基本礼仪与道德规范。根据康德的相互性理念，机器人享有的道德地位不应比动物更低。

人们可能会说，机器人即使遵循了人类的道德原则，它们也不是道德行为体，因为它们并不理解那些道德原则的内容，而且，它们并不是在依据道德原则行事，而是按照事先编程的软件系统在运行。这里涉及几个目前正在展开的理论争论。首先，现代伦理学在评价一个行为或政策（以及法律或制度安排）的道德价值时，大多采取后果主义立场，即更多地关注行为的后果，而不去追问行为主体的内在动

① 康德：《对于动物的责任》，载辛格、雷根编：《动物权利与人类义务》，第 25 页，（译文略有改动）。

② Joanna Bryson and Philip Kime, “Just a Machine: Why Machines Are Perceived as Moral Agents”, *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, Barcelona: Spain, 2011, pp.1641—1646; Joanna Bryson, “Robots Should be Slaves”, in Yorick Wilks (ed.), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, Amsterdam, The Netherlands: John Benjamins Publishing Company, 2010, pp.63—74.

机或主观意愿。例如，在依据“不伤害原则”来评价一个行为或制度安排的道德价值时，人们主要关注的是是否有伤害的后果发生，是否有人遭受了实际的伤害。其次，自然人的主观动机比较容易识别，但是，法人（如公司，以及各类政府与非政府组织）的动机却难以确定；法人没有自然人那样的中枢神经系统；法人本身没有苦乐感受的能力。在对法人的行为或决策进行道德评价时，我们主要根据该行为或决策是否符合或遵循了大多数人所认可的原则或价值，是促进还是伤害了相关主体的合法利益。再次，根据行为主义理论，我们无法直接把握人们的主观动机与内在想法，只能通过其外在行为来推论其主观动机；我们无法在人们的外在行为与其主观动机之间做出截然的区分。因此，在操作的意义上，我们可以把人或机器人的外在行为等同于其内在动机。在评价机器人行为的道德价值时，我们只需关注其行为的后果，无需追问机器人的主观动机。最后，一个合理的决定论与道德生活是相容的。苏格拉底曾说，无人有意作恶；人们之所以作恶，乃是由于他们不知道什么是善。知识即道德，所以，美德是可学的。换句话说，人们之所以做出了符合道德的行为，乃是由于他们能够认知和把握道德的具体要求，他们获得了做出正确决策所需的正确的信息；更重要的是，他们养成了正确的思维方式，而这种正确的思维方式又是由他们过往生活所发生的各种内外因素模塑而形成的。因此，苏格拉底实际上是把道德实践问题变成了知识与推理问题，把正确的道德选择视为正确的知识、完备的信息与正确的推理模式相结合的结果。只要人们的道德推理模式是正确的，那么，从正确的大前提（正确的道德原则）出发，加上特定选择环境所需的完备的信息，我们就能推出特定环境中的正确道德决策。所以，从决定论的角度看，人的道德决策与机器人的道德决策之间的区别并没有人们想象的那样大。当代的机器伦理学实际上正是从这种决定论的角度来理解、设计与制造“道德机器人”的。总之，从行为主义的角度看，只要机器人做出了与人大致相当的行为，我们就应当把机器人视为与人大致相当的主体来对待。有的学者甚至认为：“如果某个行为体在行走、交谈与行为方式方面都足够与我相似，那么，我即使不能合理地认为它拥有心灵（mind），我也有义务把它当作它好像是道德行为体那样来对待。”①

（四）人机共同体理据

当代机器人伦理学的一个争议热点是，我们是否应当研制性爱机器人（sex robots）。美国学者勒维在2008年曾著书预言，2050年左右，人们不仅能与机器人做爱，他们还想与机器人结婚，与机器人保持浪漫的伴侣关系。② 反对性爱机器人的英国学者理查森认为，研制性爱机器人有物化女性与儿童的嫌疑；性爱机器人都

① Kenneth Himma，“Artificial Agency，Consciousness，and the Criteria for Moral Agency：What Properties Must an Artificial Agent Have to be a Moral Agent?”，*Ethics and Information Technology*，Vol.11，No.1，2009，pp.19—29.

② David Levy，*Love and Sex with Robots*，New York：Harper Perennial，2008.

是“性爱活动”的被动接受者，因而对性爱机器人的使用和消费会助长人类的性虐待行为，长此以往会使人们对其性伴侣失去同情心；使用和消费性爱机器人会导致性交易的泛滥，进而对人与人之间的关系产生负面影响。① 此外，目前所研发的性爱机器人缺乏“知情同意”的程序设置。在这种情况下，人们对性爱机器人的使用与消费不仅带有“强奸”的色彩，而且还会扭曲许多人的心灵，使他们误以为在与潜在的人类性伴侣发生性行为时，后者的知情同意可有可无。②

那么，我们应当如何对待包括性爱机器人在内的社会机器人（如护理机器人、助理机器人，以及具有多种功能的全能机器人）？它们仅仅是一个物品吗？这些社会机器人实际上都扮演着某些社会角色。我们与它们实际上构成了一种新的共同体——人机共同体。我们每一个人虽然都同时生活在多种不同的共同体中，在不同共同体中扮演不同的角色，发挥不同的作用。但是，我们都属于同一个伦理共同体；作为道德行为体，我们在与共同体中的其他成员交往时一定要接受某些伦理原则的约束。例如，我们在与作为伦理共同体之成员的其他性伴侣（不管是自然人还是机器人）发生性行为时必须要征得对方的同意。我们不能把性爱机器人仅仅当成物品来对待。伦理共同体不能接受任何包含“强奸文化”的要素；我们的伦理共同体也不能接纳这样的新成员，人们在与这些新成员发生性行为时可以不征求它们的知情同意。我们期望建立和维护的是这样一种共同体，这个共同体奉行的文化是，在与性伴侣发生性行为时，必须要征得对方的知情同意。③

很显然，作为理性的成熟的道德行为体，我们必须要认可和接受性爱机器人的道德承受体地位。否则，我们的共同体就必须接受某种“强奸文化”。在这种文化中，作为性别歧视主义的态度、制度与行为模式的结果，非知情同意的性活动变得常态化，或得到明显的或潜在的认可。如果我们无法否认性爱机器人的道德承受体地位，那么，我们同样无法否认护理机器人、看护机器人、助理机器人、演艺机器人等社会机器人的道德承受体地位。我们与这些社会机器人所建立的共同体同样属于伦理共同体；这样的伦理共同体不能奉行否认其他成员之道德承受体地位的伦理歧视主义的文化。

（五）潜在道德行为体理据

艾伦等人认为：“研发人工道德行为体（AMAs）的终极目标应当是制造道德上

① Kathleen Richardson, “The Assymmetrical Relationship: Parallels between Prostitution and the Development of Sex Robots”, *Acm Sigcas Computers and Society*, Vol.45, No.3, 2015, pp.290—293.

② Sinziana Gutiu, “The Robotization of Consent”, In Ryan Calo R et al eds., *Robot law*, Northampton, MA.: Edward Elgar Publishing, 2016, pp.186—212.

③ Lily Frank and Sven Nyholm, “Robot Sex and Consent: Is Consent to Sex between a Robot and a Human Conceivable, Possible, and Desirable?”, *Artificial Intelligence and Law*, Vol.25, No.1, 2017, pp.305—323.

值得称赞的行为体。”[①]安德森亦指出，机器伦理学的核心目标是制造自主的道德机器。[②]事实上，人工智能与机器人伦理学的主流声音是，“AMAs 是必须的也是必然的”[③]。因此，研究与开发人工道德行为体不仅是可欲的，而且势不可挡。

研发 AMAs 的理由主要有：（1）研发 AMAs 是人类的必然追求。这种追求的动力主要来自四个方面：科学家与工程师的求知欲与创新冲动；消费者对性能更为完备的智能机器人的偏爱；企业占领新兴市场份额的压力；现有国际制度安排下民族国家追求科技制高点的战略选择。[④]（2）防止对人伤害。机器人的智能、力量、耐力等都高于作为个体的自然人。为了防止机器人在特定情境中做出伤害其身边的人类的决策或行为，机器人就必须要能够遵循人类的道德原则，不做出违背人类道德的决策或行为。[⑤]（3）为了获得公众的信任。智能机器人拥有巨大的破坏能力；只有让机器人变得具有道德能力，才能消除公众的不安，提升公众对机器人的接受程度。（4）应对复杂多变的环境。随着交往系统与社会生活变得越来越复杂化、多样化，机器人必须要能够在特定的情景下自主地做出符合人类道德的选择，因为，工程师不可能预计到每一种道德场景并事先把解决路径编入程序中。因此，机器人需要变得具有道德能力，以致即使身处无法预计的全新环境中也仍然能够用符合道德规则的方式来约束其行为。[⑥]（5）防止不道德的利用。机器人拥有巨大的潜能。为了防止被利用于不道道的目的（例如充当杀人魔王），机器人必须要变成道德行为体，能够识别并抵制其用户或伙伴发出的违背人类道德的指令，从而防止别有用心的人对它的误用或滥用。（6）弥补人的弱点与缺陷。机器人的知识储备远远高于具体的个人，它们收集和处理信息的能力既快速又准确，同时，机器人在做道德决策时不会受到情感或偏见的干扰和影响，因而它们的道德决策会比那些有道德缺陷的人更加完美、更加理性、更加公道与前后一致；甚至在战场上，机器人也能够克服人类士兵的许多弱点（例如不会强奸或洗劫钱财）。（7）加深并推进对人类道德的理解。为了能够把道德思维与道德推理编入机器人的初始程序中，我们必须要更为准确地理解道德语言的内涵，更为全面地把握人类的道德思维与推理模式。因而，

① Colin Allen, Gary Varner and Jason Zinser, “Prolegomena to Any Future Artificial Moral Agent”, *Journal of Experimental and Theoretical Artificial Intelligence*, Vol.12, No.3, 2000, pp.251—261.

② Susan Leigh Anderson, “Machine Metaethics”, in Michael Anderson and Susan Leigh Anderson eds., *Machine Ethics*, Cambridge, UK: Cambridge University Press, 2011, pp.21—27.

③ 瓦拉赫、艾伦：《道道机器：如何让机器人明辨是非》，第 33 页。

④ Colin Allen and Wendel Wallach, “Moral machines: Contradition in Terms of Abdication of Human Responsibility?”, in Par Lin et al.(eds.), *Robot ethics: The Ethical and Social Implications of Robotics*, Cambridge, MA.: MIT Press, 2011, pp.55—68.

⑤ Mathias Scheutz, “The Need for Moral Competency in Autonomous Agent Architecture”, in Vincent Müller(ed.), *Fundamental Issues of Artificial Intelligence*, Synthese Library, 2016, pp.517—527.

⑥ Colin Allen et al., “Why Machine Ethics”, *IEEE Intelligent Sysytem*, Vol.21, No.4, 2006, pp.12—17.

研发具有道德推理能力的机器人将最终导致对人类道德的更好理解。①

鉴于研发具有道德能力的智能机器人是机器人技术的既定目标，因而很多学者认为，机器人不仅是道德承受体，还能够成为道德行为体。例如，美国索诺马州立大学（Sonoma State University）的哲学教授萨林斯认为，自主性、意向性、履行责任是成为道德行为体的三个充分条件；将来的机器人能够拥有这三种能力。“可以肯定，只要我们追求这种技术，那么，未来高度复杂的、具有交往能力的机器人将是拥有相应权利与责任的道德行为体；然而，在某种（但不是所有）抽象层面，即使是当今普通的机器人也能够被视为道德行为体，并应获得道德关怀。”②

当然，即使机器人可以成为道德行为体，我们也要弄清，机器人能够成为何种意义上的道德行为体。穆尔曾区分了四种伦理行为体：（1）其行为具有伦理后果的伦理行为体（ethical impact agents），即代替人类从事重复、枯燥或危险工作的行为体。（2）隐性的伦理行为体（implicit ethical agents），即安全性与可靠性程度较高、对它的使用不会给人类带来负面影响的行为体。（3）显性的伦理行为体（explicit ethical agents），即伦理原则与规范已经编入其软件系统、且能依据“道义逻辑”进行推理的行为体；获得沙特公民资格的索菲亚已初步具备显性伦理行为体的雏形，而电影《机械姬》中的“爱娃”则是合格的显性伦理行为体，已经属于享有道德行为体地位的机器人。（4）充分的伦理行为体（full ethical agents），即能够做出清晰的道德判断、具有较大的道德自主性的行为体；一般认为，只有具有意识、意向性与自由意志的存在者才能成为充分的伦理行为体。③不过，严格说来，第一、二类行为体都不是道德行为体，因为它们的行为不是出于特定伦理原则或道德规范的引导。第三种行为体符合道德行为体的基本定义，因而能够发展成为人工道德行为体。第四种伦理行为体的标准比较高，只有那些正常的理性的人类个体（而非所有的人类个体）才会成为这类道德行为体。“鉴于我们目前对道德推理、人工智能、认知机制等的理解，我们至多能够制造出显性的伦理行为体，这种伦理行为体能够做出某些道德判断（这些判断不是作为软件预先编入其系统中的），并有能力对它们为何会做出那些道德判断做出解释。”④因此，我们似乎不应期待能够制造出作为充分的道德行为体的机器人。

① Aimee van Wynsberghe, “Critiquing the Reason for Making Articifial Moral Agents”, *Science and Engineering Ethics*, Vol.3, No.3, 2019, pp.719—735; Eric Dietrich, “Homo Sapiens 2.0: Why We Should Build the Better Robots of Our Nature”, *Journal of Experimental & Theoretical Artificial Intelligence*, Vol.13, No.4, 2001, pp.323—328.

② John Sullins, “When Is a Robot a Moral Agent?”, *International Review of Information Ethics*, Vol.6, No.1, 2006, pp.24—30.

③ James Moor, “Four Kinds of Ethical Robots”, *Philosophy Now*, Vol.72, 2009, pp.12—14.

④ Vincent Wiegel, “Bulding Blocks for Artificial Moral Agent”, 2006, https: / /www.researchgate.net /publication/228615030.

未来的机器人虽然很难发展成为充分的道德行为体，但是至少可以发展成为显性道德行为体。根据前文关于“所有的道德行为体都是道德承受体”的命题，作为显性道德行为体的机器人也享有道德承受体的资格。

四、实践与规范意涵

即使人们在机器人是否享有道德行为体之道德地位的问题上尚存在分歧，根据前面的论述，我们至少可以认定，以强人工智能技术、基因工程技术与纳米技术为基础的机器人能够享有作为道德承受体的道德地位。对机器人之两种不同道德地位的这种区分和厘定在机器人的设计、人机关系的厘定、人机共同体的建构等方面至少具有下述四个方面的实践与规范意涵。

第一，尊重机器人身上的人性。一提到“机器人”（robot）一词，人们联想到的就是某种与人的“形象”（human image）有关的对象或实体，而不仅仅是某种纯粹的工具或机器。任何一种实体，只要具有了人的形象，它就获得了某种特殊的意义，如历史人物的雕像、艺术家创造的人物雕像、佛像、圣母或天使雕像，等等。在面对这些实体时，人们的内心会涌起某种或崇敬（崇拜）或亲切的情感。对这些实体的蓄意毁坏或亵渎会在人们内心激起某种愤怒的情感或谴责的意向，因为这类行为亵渎了这些实体所蕴含的“人的形象”及其所蕴含的人的尊严与价值。

高度发达的智能机器人不仅有着“人的形象”，而且还具备或展现了许多属于人的属性（property）：符合人类礼仪的言谈举止、较快的推理与思维能力、对人类的法律与道德原则的遵守，等等。机器人身上的这些属性展现了人的尊严与价值，因而，理应享有道德承受体的道德地位，有资格获得道德行为体的道德关怀。根据康德关于“人是目的”的著名命题，我们也可以说：道德行为体在任何时候都不仅要把自己身上的人性，而且要把机器人身上的人性永远当作目的本身，而不能仅仅当作手段来看待。我们不仅要尊重自己身上的人性，而且要尊重机器人身上的人性。机器人之道德承受体地位所施加给道德行为体的义务之一，就是以尊重的态度对待机器人。

第二，机器人不是商品。由于机器人展现的是人的形象，因而它不是普通的商品，这意味着，首先，我们不能把机器人当作普通的商品来设计。设计者要区分两类不同的自动机械系统，一类是以人的形象出现的自主系统（如性爱机器人），一类是以完成特定工作任务为目标的智能系统（如无人机）。对于后一类自动机械系统，我们最好不要把它们设计成人的模样，也不要称之为机器人。例如，对于高度自动化的武器系统，我们最好不要把它们设计成军人的模样。我们宁可把它们直接设计成自动驾驶军用飞机、自动化坦克，等等。对于前一类自主系统，我们要尽量把它

们设计得与人相似，使得它们能够按照人类的规则来与人类交往。例如，在设计性爱机器人时，我们就不能仅仅把它们当成“性奴”或“情趣用品”来设计。我们需要把人际性爱交往的某些基本规则（如知情同意原则、相互接受原则）纳入性爱机器人的程序中。同样，在设计机器人士兵（soldier robot）或机器人警察（police robot）时，我们需要避免把它们仅仅当作杀人机器人（killing robot）来设计。机器人士兵与机器人警察的道德判断与行动都直接涉及人的生命，因而它们的设计与生产不仅需要透明、公开，还需要接受某个公正的全球机构的监管；必须要把对人类核心道德的维护与遵守作为强制性条款编入这类机器人的程序中；机器人士兵与机器人警察的设计需要遵循某些共同的全球标准。其次，消费者也不能把机器人当做普通的商品来消费。由于机器人不是普通商品，因而对机器人的消费与使用也不同于普通商品。（1）对机器人的消费主体要加以适当的审查；机器人用户应接受相关的培训与学习。只有那些通过了相应审查与培训要求的消费者才有资格购买和使用机器人，就像美国的成年公民只有通过了相关审查才有资格购买和使用枪支一样。（2）机器人的使用者要把机器人当作道德承受体来对待。例如，性爱机器人的使用者不能把性爱机器人当作“性奴”来使用，机器人士兵或机器人警察的使用者与也不能仅仅把它们当作“炮灰”或“替死鬼”来使用。（3）应成立“机器人保护协会”（类似于动物保护协会或自然保护协会），对消费者使用机器人的情况定期进行访问，引导用户以“文明而体面”（civil and decent）的方式与机器人交往。

第三，人是机器人的道德监护人。关于机器人能否，以及如何成为负责任的道德行为体，是机器人伦理学当前争论比较激烈的一个问题。本文认可机器人能够成为道德行为体的立场。但是，我们只能指望机器人成为显性道德行为体，而不能指望它们能够成为像成熟人类个体那样的充分的或完全的道德行为体。人类是机器人的道德监护人。首先，作为显性的（而非充分的）道德行为体，机器人能够履行常规的道德责任。但是，当面临复杂的道德境遇或需要做出艰难的道德判断与道德选择时，正常而理性的用户或专家需要帮助机器人做出相关的判断和决定。其次，由于不是充分的道德行为体，机器人所承担的责任类似于 16—18 岁的青少年所承担的责任。我们只能用青少年的责任标准来要求机器人。最后，机器人承担的是一种共享的道德责任（shared responsibility）。机器人的设计者、生产者与使用者都需要为机器人的行为承担部分道德责任。同时，我们也有义务通过让机器人参与人们之间的人际互动来学习人类交往规则。这意味着，我们要把具有道德承受体地位的机器人当作我们的伦理共同体的成员来对待。

第四，持续扩展道德关怀与伦理共同体的范围。把机器人纳入道德关怀的范围，这是人类道德所面临的又一次革命性的变革。前一次变革发生于人们把人之外的动物（以及所有的生命与大自然本身）纳入道德关怀范围之时。人们反对把机

器人纳入道德关怀范围的理由之一是，机器人感觉不到痛苦。[①] 然而，植物人也感觉不到痛苦，但他 / 她们仍然是道德承受体；没有神经中枢系统的法人（各类组织与机构）也是我们道德关怀的对象。况且，一个实体是否有资格获得我们的道德关怀，同时还取决于我们想与它们建立何种交往关系，取决于我们想建构一种什么样的共同体文化。如前所述，我们反对把性爱机器人建构为“性奴”，是因为我们反对一个共同体奉行“强奸文化”；我们反对把机器人士兵当作炮灰或杀人恶魔来使用，是因为这种使用方式没有恰当地尊重（机器人的）“人的形象”所承载的人的尊严与价值。更重要的是，我们对他（它）者的态度涉及我们对人性的建构。

随着机器人越来越多地介入我们的生活，人类不可避免地要进入人机共处的时代，我们不可避免地要与机器人“比邻而居”。为了使人类能够更加顺利地适应这样一个新时代，也为了使人类变得更完美，我们需要建立人机伦理共同体，把机器人纳入道德关怀的范围。为此，我们需要在全球范围内培养并弘扬一种理性而健康的机器人文化（robots culture）。我们需要普及关于机器人的完整知识，使人们养成对待机器人的正确态度与行为方式，树立正确的机器人道德观。

我们不应认为，伦理共同体范围的扩展会给人类带来威胁。人类仍然是独特的，仍然是拥有特殊的道德地位——充分的道德行为体——的唯一主体。只有人类能够最为充分地理解道德的本质与内涵。这是人类的独特之处，也是人性的光辉之点。

（责任编辑：肖志珂）

① 甘绍平：《机器人怎么可能拥有权利》，载《伦理学研究》2017 年第 3 期；雷瑞鹏、冯君妍：《机器人是道德行动者吗》，载《道德与文明》2019 年第 4 期。

Abstracts and Key Words

- Human Subjects and Quasi-Human Subjects: How Do Humans Go Along with AI? Reflection on the Accidents of the Boeing 737 MAX Air Crash *GAO Zhaoming*

Abstract: The AI technology has been changing human ordinary life and human nature through changing the conditions of human survival. We should pay much attention to the problem of how to deal with or go along with AI. Although AI has its own intelligence as well as capacity of self-studying and autonomously making choices, it is not a human subject. At most, it is a quasi-human subject due to the fact that it has no the meaningful world and freedom spirit of itself. The relationship between Humans and AI is not the inter-subjective one, but the one between humans and their artificial creatures or their own living world. AI is short of imagination and critical thinking as well as a meaningful world. There is considerable limitation on its cognition, judgment, and choice-making. Humans' trust on AI cannot be without conditions. Humans cannot give up their right of choices for their future completely to AI without conditions. Humans' attempts to put AI within their own control might create a new paradox for survival: only few human professionals are able to control AI and it is possible for them to control the rest of humankind by controlling AI; and thus when humans try to get rid of one predicament of being controlled, they might fall into another new predicament of being controlled.

Key words: human subject; quasi-human subject; relationship between humankind and AI; trust; paradox for survival

- On Robot's Moral Status as Moral Patient and Its Normative Implications *YANG Tongjin*

Abstract: We should distinguish the two different moral statuses of robot, that is robot as moral agents and robot as moral patient. Christianity, Buddhism and Confucianism provide some arguments from the perspective of ultimate concerns for us to recognize and embrace robot as moral patients. From the perspectives of normative consensus, we can offer at least five arguments for the robot's moral status as moral patients, that is indirect duty argument, the self-construction argument, behaviorism argument, human-robot community argument, and implicit moral agent argument. The practical and normative implications of distinguishing the two different moral statuses of robot include that we should respect the humanity in robot, should not treat robot as common goods, should be moral custodian of robot, and should expand continuously the scope of moral concerns and moral community.

Key words: moral agent; moral patient; ultimate concern arguments; normative consensus arguments

- Do Robots Have Empathy? From the Perspective of Confucian Ethics of *Qing* *FU Changzhen*

Abstract: Does the coexistence of human being and artificial intelligence mean a breakthrough of new civilization? Contemporary sentimentalism ethicists believe that no matter how humanized a robot is, it is impossible for AI to build a meaningful world with empathy that is closely linked to life. From the perspective of Confucian ethics, the robot's emotion is only the performance state of facing different situations rather than actually having empathy abilities, because the one-dimensional robot is separated from the ethical relations of personal practice and the experience of social survival. Therefore, whether or not the singularity will come, human beings should pay more attention to the human nature and dignity, explore the human ability and emotional uniqueness, the Guardian and the separation between the robot and the inter-phase. In the face of the new ethical relationship in the post-human era, Confucian emotional ethics may still have a positive impact on human civilization.

Key words: artificial intelligence; robot; empathy; receptivity; Confucian ethics of *Qing*