文史哲研究：人工智能研究专题

人工智能时代机器道德的可能性追问

吴兴华

摘　要：随着人工智能时代的到来，人工智能尽管对社会发展起着重要的推动作用，但它也存在着巨大的道德风险。为了规避人工智能的道德风险，道德机器的思想开始出现并得到了很多人的支持，但这种观点是出于情感而非理性。机器道德是人类道德在人工智能时代的扩展，而人工智能始终是人工的产物，是不可能拥有人类的社会性的，所以，要求机器能够拥有人类的道德是不可能的。因而，人工智能时代的主要任务不是使机器拥有道德，而是如何破除“人工智能威胁论”，约束人类自身的道德行为，从而促使人机和谐相处。

关键词：人工智能；机器道德；道德智能体；“人工智能威胁论”

作者简介：吴兴华，哲学博士，安徽师范大学马克思主义学院教授、硕士生导师。

基金项目：国家哲学社会科学基金一般项目“他者伦理研究”（14BZX123）；安徽省哲学社会科学规划项目“大数据时代意识形态安全风险及其防范体系建构”（AHSKY2015D）。

中图分类号：B82-057；TP18　**文献标识码**：A　**Doi**：10.3969/j.issn.2095-042X.2019.01.010

随着人工智能技术取得全方位的突破，我们迎来了人工智能时代。今天，人工智能已走进寻常百姓的生活，并给人们带来巨大的福利。然而，人类在享受福利的同时，却又担忧因人工智能的道德缺失而带来的，如“电车困境”和“人工智能算法歧视”等道德风险。“人工智能威胁论”的盛行，就是这种担忧的体现。正如狄更斯所言：这是最好的时代，这是最坏的时代；这是智慧的时代，这是愚蠢的时代。在这个既好又坏的时代，我们如何让机器既能为人类创造更多的福利，而又不至于伤害人类，即人与机器如何实现最优化的相处，这无疑成为人工智能研究者们关注的一个焦点。如美国人工智能和神经科学伦理学专家温德尔·瓦拉赫就提出，设计一种人工道德智能体（AMAs），让机器拥有伦理道德。这种机器道德思想的提出，在学术界也引起了不小反响，而关于人工智能伦理道德的研究也成为学术界研究的一个热点。但问题是，机器道德的思想果真合理吗？机器真的能够拥有道德？这些问题不仅关系到人们对人机关系本质的认识，而且也关系到人对自我的认识，因而，只有正确认识这些问题，才有望消除人类对机器的担忧，实现人机真正的和谐相处。

一、机器道德：人工智能时代伦理道德的扩展

伦理道德问题原本属于社会问题，即主要探讨人与人之间的相处关系。正因如此，伦理

（ethics）一词，无论是在西方还是东方，都拥有浓厚的社会内涵。如希腊文中作为 ehhics 之根的 ethos 就与拉丁文 mores 一样，都表示风俗与习惯。而我国古代，更是把伦理视为人们应该遵循的行为准则，如“三纲五常”等。可见，伦理道德原本是用来协调人与人之间的关系，是人类社会的专属，因而古代的伦理学也只关心人。然而，随着近现代社会的发展，权利主体范围不断拓展，伦理学的范围也在扩展和延伸。例如，环境伦理学的诞生，就同权利主体扩展到大自然这一情况密切相关。由于现代人对自然破坏的加重，自然开始向人类发起报复，这就迫使人类不得不思考和改善“人与大自然的关系”，要求尊重和保护大自然的权利。

如果说，环境伦理学是将人与人的伦理道德关系扩展到“关心动物、植物、岩石、甚至一般意义上的大自然或环境”[1]3，旨在解决人与周遭环境的关系，那么，今天，随着人工智能的普及，人类又出现了一种新型的人与机器的关系（人机关系）。如何协调和处理人机关系？学界开始提出人工智能道德问题，希望以此来解决人机关系问题。所以说，人工智能道德问题的提出并非无稽之谈，而是伦理道德向技术领域的一种扩展。

回顾人工智能的发展史，在 20 世纪 50 年代，人工智能还处于起步阶段，后来遭遇了发展的寒冬，又经历了野蛮生长，最终人工智能迎来了自己的时代。其实，早期人工智能技术的出炉，只不过是一场“头脑风暴”的产物。在 20 世纪 50 年代，一批顶尖学者在达特茅斯学院花园讨论如何制造一台“模拟人类各方面智能”的机器，这便是“人工智能”在人脑中的最初形态。在顶尖级大脑里，“人工智能”是受控于人的，“计算机只能按照编好的程序工作”[2]21，因而人工智能只不过是人类的机器奴仆。然而，经历了短短几十年的发展，人工智能已走进工厂和寻常百姓家，人类也因人工智能而逐渐从繁重的体力劳动中解放出来。2015 年，被誉为智能机器人元年，新一轮技术创新的浪潮正是因人工智能的发展而掀起。正因如此，“互联网预言家”凯文·凯利认为：人工智能是未来 20 年最重要的技术；著名未来学预言家雷·库兹韦尔更预言：2030 年，人类将成为混合式机器人。

今天，随着机器人的机动性、灵敏度以及智能的提高，机器战士、机器教师、机器服务生的出现，人工智能技术正在大规模地改变着我们的生活方式，因而，我们不禁要问：人工智能到底是天使抑或魔鬼？如果说，在人工智能技术发展的初期，人工智能研究的初衷只不过是用机器来代替人的体力劳动，将人从繁重的工作中解放出来，那么，在故事的开始就已埋下了隐患。因为随着人类对于人工智能要求的增长，越来越高级的机器被创造出来，人类又开始担忧机器将会取代人。其实，早在机器在生产领域中出现的时候，卢德主义（Luddism）者就曾因机器替代人导致工人失业，而捣毁机器。在人工智能飞速发展的今天，人类更有理由担忧这种科技的异化。美国思想家瓦托夫斯基就曾指出：“一方面知道科学是理性和人类文化的最高成就，另一方面又害怕科学业已变成一种发展得超出人类的控制的不道德的和无人性的工具，一架吞噬着它面前的一切的没有灵魂的凶残机器。”[3]可见，这场解放似乎又是人类在作茧自缚。

其实，思想家和科学家们对于技术乃至今日人工智能双面性的担忧早已存在。控制论的创始人维纳，早在 1950 年就曾在其著作《人有人的用处》中，认识到“人脑的贬值”。由于“我们没有理由说机器不可以和人相似”[4]，因而，人在与机器的博弈中就有可能会悲惨地出局。德国哲学家弗洛姆也曾指出：“人制造了像人一样行动的机器，培养像机器一样行动的人——有利于非人化的时代。在这个时代里，人被改造成为物，变成生产和消费过程的附属品。”[5]国际象棋大师卡斯帕罗夫败给深蓝，围棋大师李世石败给 AlphaGo，这一切似乎又验证了思想家和科学家们的担忧。人们不禁要问：我们今天输掉了一个象棋大师和一个围棋大师，明天我们还会输掉什么？

今天，人与机器的关系已开始由“机器听从于人的程序和指令”走向了“你必须适应机器人的需求，因为它不会顺从你的需求”[2]37。随着大数据技术的发展以及智能手机的广泛运用，我们的生活似乎已慢慢被机器所掌控。人类要如何躲过这些危险的机械陷阱？如何阻止人工智能的反叛（从“仆人”到“主人”）？面对人工智能的疯狂发展，人类似乎已身处两难的困境。当然，人类要想办法走出困境。正如李开复所言：“人类，要让巨变这一标签作为自己的脚注，而不是被动地成为它的注解。”[2]序言:15那么，人类到底如何才能驾驭自己制造的这头猛兽？正因如此，关于“机器道德”问题的思考便应运而生了。

在一个缺失人类特性，比如意识和情感的人工智能身上，我们又如何能期许它作出道德选择呢？当然，随着人工智能技术的发展，机器已经能够自主学习，从 AlphaGo 身上，人类看到了人工智能的自主性。然而，正如 MIT 的情感计算研究专家罗莎琳德·皮卡德所言：“机器越自由，就越需要道德准则。”[6]18为了在人工智能时代协调好人机关系，就需要机器具有道德。尽管大多数人已就此观点形成共识，但按照近代哲学鼻祖笛卡尔的理解，“机器智能的想法在形而上学上是荒谬的”，“人类把机械的身体和非实物的心灵结合为完美协调的整体，但是实物的机器单独永远不可能有智能的特征”[6]49。所以说，机器是不可能有道德的。所谓道德机器，无非是人类伦理道德向机器的扩展。伦理学家施韦泽就曾认为人类道德要向其他领域扩展，并认为这种扩展不仅是合理的，更是必要的。他本人就“把阐述‘热爱所有创造物的伦理学’的具体要求视为自己的终身职志”[1]71。

二、道德机器：人类的道德幻相

机器道德的提出，缘起于人类对人工智能技术发展的担忧和恐慌，希望机器能够拥有道德，从而使机器在服务于人类的同时，又不会伤害人类。这种思想，表面看来，是以赋予机器道德的方式，来实现人与机器的平等共处，但实际并非如此，它恰恰是一种人类自我中心思想的产物。人类自我中心思想，古已有之，它的存在自有其合理性。因为，任何一个物种都有自保的天性，“自我中心”本来就是万物的本性。早在西方古希腊时代，智者普罗泰戈拉就曾宣称道：“人是万物的尺度，是存在者存在的尺度，也是不存在者不存在的尺度。”[7]而在中国古代《论语》中也有类似人类中心观点的记载。孔子的马棚失火了，孔子从朝中归来，问“‘伤人乎？’不问马”[8]。由此可见，自古以来，人类就以自己为中心来规划世界，因而周遭的世界都是为“我”所用的。实际上，人工智能技术的发展，也是人类“外物为我所用”思想的产物。只是不曾想到，一架模拟自己思维的机器，却已经有了深度学习的能力，这不得不使人类担忧：我们还能否控制和驾驭自己的制造物。美国思想家戴维·埃伦费尔德就曾在批判人道主义时指出：“尽管机器崇拜很流行，但机器并不特别好崇拜。我们的控制装置里没有道德或终极目的，甚至没有个性。”[9]这也就是说，我们虽然能够模拟出人的思维，但我们无法使这架机器模拟人的道德。而机器的深度学习能力又无疑使机器有了自由意志，从而导致人类无法控制机器的行为。因此，道德机器（道德智能体）的提出，无非是出于人类的自保。道德是人类独有的现象，它是属于人的本能的东西，也是人与其他外物区分的一个重要标志。我们常说，没有道德或失去道德，人类就是动物，反之，万物都有道德，那也就没有了独一无二的人类。所以说，机器道德在思想和逻辑上是很难成立的。

既然机器拥有道德很难成立，那么，道德机器为何会受到学界的广泛关注，且诸多学者还积

极地探究如何制作道德智能体？当然，这显然是将人工智能简单地比附于人脑，以及对于意识特别是自我意识错误认识的结果。如维纳的《控制论》一书，就“以数学为纽带，揭示了机器的通信、控制机能与人的神经、感觉机能的共同规律，突破了人机割裂的传统观念”[10]。按照维纳的理解，既然机器与人存在共同规律，那么，我们完全可以制造一架拥有人的行为的机器。正是循着维纳的思路，人们认为图灵机的诞生已证实了机器完全可以模拟人脑的思维。如果说，图灵机只是揭示了机器可以模拟人的思维，那么未来的人工智能就应该也能像库兹韦尔所预言的那样，它“可以超越自然的局限，并依照自身的意志改变世界”[11]1。既然机器有自由意志，那就必然涉及到机器行为的道德性问题。

其实，这里最为关键的问题是机器到底有没有意识？如果机器有意识，那么它与人类的意识是否等同？有意识是否就意味着拥有道德？机器道德倡导者当然是认同机器意识的。但这种认同的问题在于，将人的意识庸俗化。如哲学家约翰·塞尔就曾说过：“意识是一个生物过程，和消化、哺乳、光合作用或有丝分裂一样……大脑是一台机器，确切地说是一台生物机器，但是它自始至终都是一台机器。所以我们首先要弄清楚大脑是如何产生意识的，然后再建造一个人工机器，这个机器要和人一样，具有同样能够产生意识的有效机制。”[12]显然，它是将意识泛化了。然而，意识特别是自我意识是人类所独有的，尽管人工智能拥有深度学习的能力，但并不代表它已具有了自我意识。我们完全可以将一些程序和指令输入机器，但机器决不会拥有它的创造性，因为创造性是源自于人类的“意会知识”（波兰尼语）。在波兰尼看来，意会知识是一种个体知识，它是一种无法言明的东西，是人类创造性的源泉。“如果我们的确是由注意到一些无法言明的事物而认知心理过程，那就意味着我们不可能制造出一台能够做出与我们据以认知这些心理反应完全相同的反应的机器。”[13]所以说，如果承认意识是心理创造的产物，那么超越人类，控制人类的机器就不可能诞生。按照蔡曙山对知识的分类来看，人类认知从初级到高级可以分为五个层级，即“神经认知、心理认知、语言认知、思维认知和文化认知。……前两个层级的认知即神经认知和心理认知是人和动物共有的，称为‘低阶认知’，后三个层级的认知是人类所特有的，称为‘高阶认知’”[14]。这五个层次，人工智能都是望尘莫及的。新西兰生物学家迈克尔·丹顿（Michael Denton）也曾认识到：机器设计原则跟生物学的某些原则存在着巨大的不同，生物形式只能从生物过程中创造，它具有“自我组织性、自我参考性、自我复制性、交互作用性、自我塑造性，以及整体性”[15]，而这是机器所不曾拥有的。

不仅机器无法拥有人类相同的思想和意识，而且就道德本身而言，道德作为人的一种意识或思想观念，它是社会的产物，而机器是不具有社会性的，正因如此，我们很难要求一架机器去作出社会性的道德选择。即便如康德、边沁、密尔等道德大家所推崇的“道德准则应为普世性的”[6]68这一基本观点在特定情境的细节面前也会崩溃，更何况一架机器呢？机器道德论者，要求用道德准则来编程，从而制造道德智能体，然而，“这样的道德准则应该是什么样的？”[2]81机器会不会遵守？我们无从得知。所以说，道德智能体是不可能存在的，所谓的道德机器，也只不过是人类的道德幻相而已。

三、破除“人工智能威胁论”：从机器道德回归人的道德

既然要求机器能够拥有人类的道德，只是人类的一厢情愿，那么，如何来化解人机矛盾，消除人类对机器的担忧呢？这需要我们正视“人工智能威胁论”。伟大的物理学家霍金曾指出：人

工智能的进化速度会比人类更快，而它们的未来还不可预测。SpaceX 公司创始人埃隆·马斯克（Elon Musk）则认为 AI 可能会引发第三次世界大战。卡普兰更在《人工智能时代》一书中，将人工智能的威胁概括为三点：第一，人工智能可能会因为不受道德约束而威胁到人类的生命；第二，人工智能可能取代人类并奴役人类；第三，少数人掌控技术将拉大贫富差距。因而，应“让魔鬼重回瓶子”。这些“人工智能威胁论”者，无疑是导致人们对于机器产生恐慌的始作俑者。然而，正如前文所述，将问题的解决寄望于道德智能体已不可能了，那要消除这种恐慌，实现人机和谐共处，唯有破除“人工智能威胁论”。

其实，“人工智能威胁论”是科技反思的产物，而这种反思在科学诞生之时就已开始。早期的“科技威胁论”是与科学相伴而生的。早在 18 世纪，启蒙思想家卢梭就曾探讨过科技与道德的关系。卢梭指出：“我们的灵魂正是随着我们的科学和我们的艺术之臻于完美而越发腐败。”[16] 卢梭的科技批判，可谓早期的“科技威胁论”。而现代性的批判者，更是对科技异化展开全面批判。法兰克福学派代表人物马尔库塞，在《单向度的人》一书中，揭示发达工业社会是一个单向度的极权主义社会，认为造成极权的根源不是恐怖和暴力，而是技术的进步。在马尔库塞看来，在技术现实的领域内“客观世界正在被改造成一种工具。仍处在工具世界之外的部分——未征服的、蛮荒的自然——如今显然处于科学技术进步力所能及的范围之内”[17]。这是典型的“科技威胁论”的现代工业版本。至于“人工智能威胁论”，也只不过是“科技威胁论”的当代升级版，因而并非什么新生事物。

不仅如此，“人工智能威胁论”在理论和逻辑上也是不成立的。首先，人工智能始终是“人工的”产物，即人创造了人工智能。而人的行为是有目的的，人类之所以制造出人工智能，是希望它为人类服务，这是人类制造人工智能的出发点。所以，50 多年前，艾萨克·阿西莫夫（Isaac Asimov）就提出“机器人三大定律”：“1. 机器人不可以伤害人；或者，通过不作为，让任何人受到伤害。2. 机器人必须遵从人类的指令，除非那个指令与第一定律相冲突。3. 机器人必须保护自己的生存，条件是那样做与第一、第二定律没有冲突。”[6]1 尽管人工智能历经了 50 年的发展，已取得巨大进步，但这三大定律却一直不曾改变。所以，人工智能是超越不了人类的。其次，人工智能永远是人类的他者，无法与人类平起平坐。既然人工智能是人工的产物，那它不仅永远是人类的附庸，而且也不可能获得作为万物之灵的人类的灵性。如果我们出于对自己创造物异化的担忧，非要强行使机器能够与人类一样行为，那么，这恰恰是人类对自我尊严的冒犯，也就是将自己降低到机器的层面。因为，作为人类社会现象的道德，一旦能够存在于机器身上，如果哪一天真的诞生了能够明辨是非的道德智能体，那恰恰是人类非人化时刻的到来。因此，21 世纪最伟大的未来学家库兹韦尔预言奇点的出现，实际上，就是宣判人类历史的终结。由此可见，无论是使机器成为人，还是人成为机器，都是对人类的一种伤害。所以说，“人工智能威胁论”是不可信的，我们没有必要，也没有理由深陷人工智能道德的“高尔丁死结”之中。

“人工智能威胁论”尽管不成立，但它也确实向人类提出了一个重要的问题，即在人工智能时代，我们该如何解决人机矛盾。在“人工智能威胁论”看来，人类已没有退路，我们无法退回到前智能时代，而现在唯一补救的方法就是制造出道德智能体，让机器按照人类的道德标准行动。“人工智能威胁论”的倡导者马斯克说：“既然我之前对人工智能的警告收效甚微，那么好的，我们自己来塑造（人工智能）的发展，让它走向好的一面。”[18] 而这便是人机合一的“赛博格”（cyborg）。然而，正如前文所言，道德智能体思想的提出，不仅是没有理论依据的，逻辑上也是不成立的，而且它没有真正揭示出人工智能时代人机矛盾的实质。

其实，马克思在探究异化问题时，就敏锐地洞察出科技异化的实质问题。实际上，科技异化论并没有真正揭示出这种异化到底是科技的异化还是人的异化。今天的“人工智能威胁论”亦如此。我们知道，在马克思看来，科技带来的问题并非源自科技本身，而是源自挺立在科技背后的人。因而，在谈机器问题时，马克思认为机器以及机器统治人的奥秘，并不是“机器对人的统治”，其实质是人对人的统治。在资本主义社会，就是资本家利用或操控机器来对无产阶级进行统治。所以说，是资本的逻辑掩盖着科技与人之间的真实关系，因而要揭示人机之间的矛盾，同样也只有回到人类自身，即机器道德问题其实折射出的是人类自我的道德问题。

回到问题的起始，机器道德的提出，根源是人类对于机器取代人的担忧，或机器对人类自我尊严遭损害的担忧。这里值得我们去思考的是：我们的当务之急不应是如何使机器拥有道德，而应是人类如何道德地看待和使用机器，即要从机器道德回到人的道德。由于机器是人类制造的，那人类就必须对机器行为负责。我们知道，作为人工产物的机器就犹如人类未成年的孩子，当孩子犯错时，作为监护人的家长就必须对孩子的行为负责。所以说，机器人的道德问题应该是人类的道德问题，作为机器人的监护人，人类理应为机器的行为履行监护的责任，而不是对机器提出道德要求。正因如此，计算机之父图灵“并没有专门思考机器道德问题，而是在思考计算机器到底能否产生原创性的行为的问题”[6]86。

既然机器道德最终还是人的道德，那么，今日人机矛盾的化解，更为恰当和实际的做法，不是制造道德智能体，而是强化人类自我道德的约束，毕竟，站在人工智能背后的，依然是活生生的人类道德主体。所以说，要使人工智能更好地服务于人类，实现“我们制造工具，而工具让我们走得更远”[11]270的愿望，我们就必须要转变思维，从对机器道德的要求回归到对人类自我道德的要求，在善待自己的同时，也善待机器。唯有如此，我们才能创造一个人机和谐共处的美好新时代。

四、结语

综上所述，机器道德的提出，是“人工智能威胁论”者为避免未来机器人对人类的报复而采取的消极应对之策。然而，这种对策是无效的。首先，人工智能到底会不会取代人类？它又是如何取代人类的？这些都是无法确定的问题。因而，人工智能取代人类只不过是一种推测。问题的不明确性，无疑会使问题解决的方案变得无的放矢。其次，机器是在何种意义上拥有道德？拥有道德的机器是不是一个真正的道德主体？我们知道，道德行为的产生是由道德主体造成的。如果机器拥有道德，那么也就意味着，我们已赋予机器道德主体的地位。然而，人类制造人工智能的初衷是要机器服务于人，那么，拥有道德主体地位的机器，无疑具有了与人一样的主体人格。按照康德的伦理学理论，人是目的，千万不要把人当作手段。如此一来，人工智能的出现就成了一个悖论。最后，人工智能的道德是何种意义上的道德？道德是人类社会现象，道德问题是纯粹人的问题，在人工智能时代机器道德的提出，无疑是对传统道德观念的颠覆。正因如此，目前用传统的道德范式来探讨机器道德，用人类的规范伦理来约束机器，这本身就是荒谬的。所以，在人工智能时代，对于机器带来的道德问题，人类当务之急要做的不是制造出道德智能体，而恰恰是恪守自身的道德。

参考文献：

[1] 罗德里克·弗雷泽·纳什. 大自然的权利——环境伦理学史 [M]. 杨通进，译. 青岛：青岛出版社，1999.

[2] 杰瑞·卡普兰. 人工智能时代——人机共生下财富、工作与思维的大未来 [M]. 李盼，译. 杭州：浙江人民出版社，2016.
[3] 瓦托夫斯基. 科学思想的概念基础 [M]. 范岱年，译. 北京：求实出版社，1989：9.
[4] 维纳. 人有人的用处：控制论和社会 [M]. 陈布，译. 北京：商务印书馆，1978：21.
[5] 沈恒炎. 未来学与西方未来主义 [M]. 沈阳：辽宁人民出版社，1989：182-183.
[6] 温德尔·瓦拉赫，科林·艾伦. 道德机器——如何让机器人明辨是非 [M]. 王小红，译. 北京：北京大学出版社，2017.
[7] 北京大学哲学系外国哲学史教研室. 西方哲学原著选读：上卷 [M]. 北京：商务印书馆，2004：54.
[8] 杨伯峻. 论语译注 [M]. 北京：中华书局，1980：105.
[9] 戴维·埃伦费尔德. 人道主义的僭妄 [M]. 李云龙，译. 北京：国际文化出版公司，1988：86.
[10] 路甬祥. 创新的启示——关于百年科技创新的若干思考 [M]. 北京：中国科学技术出版社，2013：128.
[11] 雷·库兹韦尔. 人工智能的未来——揭示人类思维的奥秘 [M]. 盛杨燕，译. 杭州：浙江人民出版社，2016.
[12] 约翰·赛尔. 我们共享的状态——意识 [EB/OL]. (2013-05-08) [2018-12-09]. http://www.24en.com/ted/talk/11487.html.
[13] 迈克尔·波兰尼. 科学、信仰与社会 [M]. 王靖华，译. 南京：南京大学出版社，2004：196.
[14] 蔡曙山. 论人类认知的五个层级 [J]. 学术界，2015 (12)：5-20.
[15] 拥抱“奇点”[EB/OL]. (2017-06-07) [2018-12-09]. http://www.sohu.com/a/146861763_488848.
[16] 让-雅克·卢梭. 论科学与艺术 [M]. 何兆武，译. 上海：上海人民出版社，2005：25-26.
[17] 赫伯特·马尔库塞. 单向度的人——发达工业社会意识形态研究 [M]. 刘继，译. 上海：上海译文出版社，2006：157.
[18] 埃隆·马斯克. 人工智能将引发第三次世界大战 [EB/OL]. (2017-09-05) [2018-12-09]. https://news.china.com/internationalgd/10000166/20170905/31289053.html.

The Possibility of Machine Morality in the Era of Artificial Intelligence

Wu Xinghua

Abstract: With the advent of the era of artificial intelligence (AI), artificial intelligence plays an important role in promoting social development, but it also has great moral risk. In order to avoid the moral risk of artificial intelligence, the idea of moral machine emerges and gets much support. But this view is emotional rather than rational. Machine morality is the expansion of human morality in the era of artificial intelligence. Artificial intelligence is always the product of human without human social attribute. So it is impossible for machines to have human morality. Therefore, the main task in the era of artificial intelligence is not to make moral machines, but how to break away from the “AI threat theory” and restrain human's own moral behavior, so as to promote the human-computer harmonious coexistence.

Keywords: artificial intelligence; machine morality; moral intelligent agent; “AI threat theory”

(**收稿日期**：2018-11-22；**责任编辑**：陈鸿)