

【关键词】机器人 人形机器人 伦理学 机器人伦理学

【作者简介】刘永谋，中国人民大学哲学院教授、博士生导师；白英慧，中国人民大学哲学院博士研究生。刘鹏，南京大学哲学学院教授、博士生导师。程林，广东外语外贸大学外国文学文化研究院教授、阐释学研究院兼职研究员。谭笑，首都师范大学政法学院教授。孙圣，西北师范大学哲学与社会学院副教授、硕士生导师。杨庆峰，复旦大学科技伦理与人类未来研究院教授；朱清君，复旦大学社会发展与公共政策学院博士研究生。

【中图分类号】B829 【文献标识码】A

【文章编号】2097-1125（2025）06-0005-55

## 机器人伦理学的拟人论基础<sup>\*</sup>

刘永谋 白英慧

近年来，人工智能、传感器、机器人控制与动力学、云计算与物联网、人工肌肉与柔性材料等技术的不断突破，使机器人更加灵活与自主，并使其应用场景愈加多元化、精细化。同时，机器人的研究、设计、制造和使用面临人类失业、隐私泄露、情感欺骗、责任分配等十分棘手的伦理问题，这些问题对机器人技术的发展及人类社会生活产生了不可忽视的重大影响，必须结合具体情境加以认真研究。在此背景下，机器人伦理学（roboethics）兴起并持续火热。顾名思义，机器人伦理学是研究有关机器人的伦理问题的学问。凯斯·阿布尼（Keith Abney）总结了机器人伦理学研究对象的三层含义：第一，机器人技术专家的职业道德；第二，为自动化机器人编写的道德规范的代码程序，即机器人自己的而非人类的准则；第三，机器人在具备进行伦理推理的自我意识能力时自行选择的伦理准则。<sup>①</sup>换言之，机器人伦理学研究

\* 本文系国家社会科学基金重大项目“现代技术治理理论问题研究”（21&ZD064）的阶段性成果。

① 参见[美]帕特里克·林、凯斯·阿布尼、乔治·A.贝基主编：《机器人伦理学》，薛少华、仵婷译，人民邮电出版社2021年版，第35页。

机器人专家伦理问题、机器人应用伦理问题以及机器人道德地位问题。仔细分析机器人伦理学既有研究可以发现，机器人拟人论是这门学问的基础，甚至可以说，整个机器人伦理学基本上是拟人论思维的产物。因此，有必要对机器人拟人论及其在机器人伦理学中的基础作用进行深刻讨论，特别是要仔细分析“什么是机器人拟人论”“机器人拟人论在机器人伦理学中有何表现”“机器人拟人论何以会成为机器人伦理学讨论的缺省配置”“机器人拟人论的基础作用对机器人伦理学的发展会产生何种后果以及有何启示”等问题。

## 一、作为意识形态的机器人拟人论

拟人论指的是将人类的精神、情感、动机和行为等特性赋予非人对象的思想观念。泛灵论是拟人论的一种典型形态。在宗教领域，泛灵论指向一种对灵魂与精灵的信仰崇拜，将人类的特征投射至非人；在人类学中，泛灵论转向一种破除现代二分的本体论，强调人类与非人的共同特征；在心理学中，泛灵论代表一种缺乏理性或诉诸感性的认知方式，包括对非人的同化或情感依赖。此外，拟人论还被作为一种修辞手法和语言技巧，在日常生活中潜移默化地增强着人类的情感表达，激发着人类的情感共鸣。

在此基础上，机器人拟人论指的是将机器人视为特殊的类人或人类的思想观念。大致来说，机器人拟人论包括人形、人神和人事三个层面的模拟。人形论强调机器人在外在层面上与人类的相似性，尤其是在外形、姿态、声调、表情上像人。比如，Atlas 机器人的行走、跑步、跳跃、翻滚均十分自然，是人形论在实践中的典型技术案例。此外，仿生机器人通过对人类肌肉与骨骼结构的模拟，极大促进了机器人的外形与动作的逼真化、精细化。人神论侧重于机器人在内在层面上对人类意识、语言、道德、情感的模拟。例如，Sophia 社交机器人内嵌自然语言处理模型，能够展现出一定的语言理解与表达能力，好似具有人类意识一般。如今，情感计算技术的不断革新使机器人逐渐具备更加复杂的情感反应，有助于恋爱机器人产业的落地与发展。人事论则主张机器人在交互层面上能够以类似于人类的方式进行社交互动、执行任务和适应环境。在未来，通过开发更为先进的多机器人协作系统，机器人不仅能够与人类协作，还可以与其他机器人共同模拟人类的团队工作模式，而非孤立地执行指令。

与机器人拟人论相比，AI 拟人论主要聚焦于人神论层面，认为 AI 在精神

尤其是智能层面越来越像人。在 AI 拟人论的作用下,当代 AI 文化表现出人类在认知上对 AI 的拟人化、在情感上对 AI 的依赖性以及在信仰上对 AI 的崇拜感的特征。<sup>①</sup>近年来,机器人技术发展迅速,AI 与机器人的融合成为其中的关键。当机器人不仅有智能,而且有物理身体时,这种技术融合便催生了机器人拟人论在人神层面之外的人形和人事层面,为机器人拟人论带来了新的学术议题,如关于“机器人身体与灵魂的关系”“机器人身份同一性”“人机界限”等问题的探讨。

在目前的机器人文化中,存在大量的机器人拟人论意识形态,包括但不限于:“机器人觉醒”,即认为机器人迟早会有意识、知善恶;“机器人权利”,即主张赋予机器人一定的权利与地位;“机器人反叛”,即认为机器人将反抗人类的奴役;“机器人进化”,即认为作为“新生命”与“新物种”的机器人能够不断进化,会像皮诺曹一样,具备进化为人的潜力;“机器超人”,即强调由超级机器和超级 AI 组合而成的超级 AI 机器人将成为尘世的天神,并威胁到人类的生存。

需要明确的是,意识形态不同于公众知识,两者在基础依据、内在特征、主要来源上存在差异。在基础依据上,意识形态依赖于人们的信念、价值观与文化背景,而公众知识依赖于实证研究、科学证据以及普遍事实;在内在特征上,意识形态具有想象性、主观性,而公众知识具备普遍性、客观性;在主要来源上,意识形态受文学、宗教、哲学等文化因素影响较深,而公众知识多源于学术研究、科学发现、公共教育。从这个角度看,一些作为意识形态的机器人拟人论信念“都是由神话、传说、幻想小说作家和大众塑造的,而不是以事实和合理的方式塑造的”。<sup>②</sup>但是,由于其高度的文化象征性和社会意义,某些机器人拟人论已然成为一种集体的意识形态,带有未来主义色彩和对未知技术的恐惧。

## 二、机器人伦理学中的拟人论特征

在既有机机器人伦理学研究中,拟人论特征表现在与机器人相关的术语、方法、问题、设计、角色层面。

<sup>①</sup> 参见刘永谋、白英慧:《科学文化视域下的人工智能恐惧问题》,《理论探索》2024年第4期,第28页。

<sup>②</sup> [美]帕特里克·林、凯斯·阿布尼、乔治·A.贝基主编:《机器人伦理学》,薛少华、仵婷译,人民邮电出版社2021年版,第357页。

在术语拟人化层面，汉语中的“机器人”“智能体”“能动者”“硅基生命”“准人”“拟主体”等概念以人类特征来界定机器。其中，“机器人”“智能体”“能动者”“硅基生命”概念在不同程度上强调了机器具有类人的自主性，而“准人”“拟主体”概念则考虑赋予机器人一定的权利，甚至让其承担责任，具备类人的道德地位和法律地位。此外，在英文中，单词“android”“gynoid”模仿人类性别对机器人进行区分。“android”与“gynoid”的共同后缀“-oid”意为“类似的”，“android”来自希腊语“aner”（男性）与“-oid”的组合，意为“男性机器人”，而“gynoid”来自希腊语“gyne”（女性）与“-oid”的组合，意为“女性机器人”。

在方法拟人化层面，“笛卡尔测试”尝试通过测试机器人是否具备思考能力与意识来确定其是否具备类人心智。“图灵测试”关注机器人是否具备人类特有的思维模式和语言表达能力。“道德图灵测试”评判机器人的整体道德表现是否不低于人类水平。<sup>①</sup>“类脑智能”通过模仿人类大脑的结构与功能，赋予机器人人类认知。“人性测试”源于菲利普·K. 迪克（Phillip K. Dick）的小说《仿生人会梦见电子羊吗？》，在该小说中警方通过向嫌疑人提出旨在引发情绪反应的相关问题并监测由此引发的身体反应来筛选“复制人”机器人。<sup>②</sup>上述方法基于不同的评估维度与目标，为我们提供了多角度衡量机器人拟人化程度的路径。然而，必须认识到，这些方法在实际应用中存在一定的局限性，过度依赖某一维度可能会导致我们对人类的理解过于片面化，进而加深对机器人拟人论的误解。

在问题拟人化层面，技术领域的研究大多聚焦于“机器学习问题”“通用人工智能问题”“具身智能问题”。“机器学习问题”研究致力于使机器人像人一样，通过经验学习、从错误中学习，具备泛化能力，能够进行增量学习与持续学习；“通用人工智能问题”研究尝试使机器人像人一样具备解决各种跨领域的复杂问题的能力；“具身智能问题”研究试图打造类人身体，以使机器人具备模仿人类通过感官获取信息、通过身体行动做出反应的能力。此外，在人文社科领域，讨论与分歧大多集中于“机器人社会地位问题”，此问题可

① 参见 Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong*, New York: Oxford University Press, 2009, pp.70–71。

② 参见 [英] 帕特里克·林、瑞安·詹金斯、基思·阿布尼主编：《机器人伦理学 2.0：从自动驾驶汽车到人工智能》，毛延生、刘宇晗、田野译，上海交通大学出版社 2023 年版，第 354 页。

分解为“机器人是否为应承担责任的主体”与“机器人是否为应具备权利的受体”。学者们在机器人责任与权利问题上看法不一，激进程度不同。无论如何，对相关问题的思考不应脱离实际，“机器人社会地位问题”并非论证问题，而是社会主流价值观演变问题。

在设计拟人化层面，最典型的便是关于“恐惑谷效应”的讨论。“恐惑谷效应”表明，当机器人在外观和行为上与人类越来越相似时，人类对它的情感反应会逐渐变得积极正面；然而，当机器人的外观和行为接近人类的程度达到某一临界点时，继续增加相似度反而会让人产生不适感，形成所谓的“恐惑谷”；如果机器人在这之后进一步接近人类外观和行为，人类对它的情感反应会再次变得正面。这提醒机器人设计者，应避免将机器人过度拟人化。实际上，机器人像不像人并没有那么重要，关键在于实用性而非类人性。因此，在机器人的设计过程中，应采取渐进式的类人化路径，避免对人类的完全模仿，尤其应关注机器人的功能性、友好性、有用性。

在角色拟人化层面，机器人越来越被赋予人类角色，如伴侣、孩童、助理、朋友、恋人等。例如，在军用机器人的实际应用中，不少士兵将机器人视为战友，甚至有士兵愿冒生命危险去营救与其共事的机器人；<sup>①</sup>在陪伴机器人报废后，有人会选择为其举办葬礼。<sup>②</sup>在现有的机器人应用伦理学中，学者们不再仅仅将机器人视为工具，而是越来越多地赋予其多种社会角色，在某种程度上推动机器人融入人类生活的方方面面，并引发了关于机器人社会接受度的广泛讨论。

### 三、机器人拟人论流行的主要原因

机器人拟人论在机器人伦理学中流行的原因至少涉及心理、传播、哲学、文学、学科方面。

机器人拟人论与人类深层的心理结构与认知方式相关。尼古拉斯·埃普利（Nicholas Epley）等提出了拟人化的三因素理论，其中的三因素即“诱发主体知识”、“效能动机”与“社会动机”。具体来说，“诱发主体知识”意味

① 参见[美]帕特里克·林、凯斯·阿布尼、乔治·A.贝基主编：《机器人伦理学》，薛少华、仵婷译，人民邮电出版社2021年版，第201页。

② 参见[美]奥利·洛贝尔：《平等机器：数字技术创造美好未来》，苏苗罕、王梦菲译，上海人民出版社2024年版，第282页。

着，由于关于人类自身的知识比关于机器人的知识更易获取，而且机器人与人类自身具有感知相似性，故而当人类对机器人进行认知时，关于人类自身的知识会被激活、修正并应用于机器人；“效能动机”指的是，将人类自身的知识赋予机器人，能够增加对机器人的理解、预测与控制，从而减少不确定性和焦虑；“社会动机”因素表明，通过将机器人拟人化，人类能够满足自身社会性联系的需要。<sup>①</sup> 值得注意的是，随着机器人自主性的提升，在机器人拟人论的心理层面，投射心理学逐渐让位于关系心理学。这意味着，拟人论不再仅仅是源于机器人对人类情感投射的被动接收，而是更多源于机器人的主动反应及其与人类的双向互动。

机器人宣传术在传播层面极大地推动了机器人拟人论的流行，包括但不限于三种宣传策略。一是拟人化的表达手法。一些科技公司为制造“卖点”，常常从机器人产品的广告语入手，运用拟人的修辞手法赋予机器人类人特征。例如，商汤科技推出的语言大模型“商量”，其宣传口号为“商量商量，都能解决”。二是情感化的叙事方式。打感情牌、走煽情路线是情感化的关键。科沃斯机器人在母亲节的海报文案为“一辈子，你围着我转；这一次，换我围着你转”，这一广告语通过温暖亲切的语言，生动地建立了情感连接、实现了情感传递、引起了情感共鸣，使用户在潜意识中将机器人视为家人，从而增强对其的依赖性与信任感。三是夸张化的神话塑造。一些“标题党”为夺人眼球，大肆宣扬“机器人将接管世界”等言论，此种过度夸张的描述将机器人塑造为具备自主意志与决策能力的存在，赋予机器人近乎“统治神”的角色。

在技术哲学中，机器人拟人论可追溯至恩斯特·卡普（Ernst Kapp）提出的“器官投影说”。卡普将技术视为对人体器官的延伸与强化，即对人体器官的外形、结构与功能的模仿，如锤子是对人手的投影，机器运作是对人类劳动的投影，铁路系统是对人体血液循环系统的投影。<sup>②</sup> 机器人技术的不断进步驱动机器人设计逐渐向更加完整、复杂的“人类模板”演化，这不仅仅是模仿人体的某一单一功能或结构，而且是综合性地模仿人体各个器官的功能和相互作用。进而言之，器官学可以演变为造人学，人类不仅能够使机器人具备物理上的类人性，还能赋予其意识的“觉醒”。

机器人题材科幻作品中的乌托邦与敌托邦图景均源于对机器人的拟人化。

---

① 参见 Nicholas Epley, Adam Waytz and John T. Cacioppo, On Seeing Human: A Three-factor Theory of Anthropomorphism, *Psychological Review*, Vol. 114(4), 2007, pp. 864–886.

② 参见乔瑞金主编：《技术哲学教程》，科学出版社 2006 年版，第 40 页。

在机器人乌托邦中，机器人之所以能够与人类和谐相处，是因为其具备类人的情感、理解、互动能力，从而可以与人类建立持久的、深层的情感连结，实现团结协作；而在机器人敌托邦中，机器人在自我意识和独立决策能力上的拟人化成为其反叛、操纵与威胁人类的根源。不可否认，通过描绘机器人具备类人外貌、行为、情感和自我意识的情节，科幻作品最早塑造了人类对机器人的拟人化理解，并具有深远的影响力。毫无疑问，机器人题材科幻作品中的拟人化倾向推动了机器人伦理学对机器人责任与权利的思考以及对人机关系界限的思考，并为相关规范的制定提供了大量的思想实验和道德框架，如阿西莫夫三大定律。

必须要指出的是，拟人论赋予机器人特殊地位，使机器人伦理学从科技伦理学中分化出来，成为一个独立的应用伦理分支。机器人拟人论预示着机器人愈来愈强的自主性，这在三个方面推动了机器人伦理学的独立。首先，现有的科技伦理学在广义上评估科技对社会的影响，这一传统框架已无法完全涵盖机器人技术发展在责任、权利等问题上面临的困境。其次，鉴于机器人的强大能力，相较其他设计者，机器人设计者面临更加难以预测的风险与不确定性，因此有必要针对机器人设计者构建一套专门的伦理规范，使其伦理责任更加细致化、明确化。最后，通用机器人尚未实现，机器人的落地应用具有较强的专业性，需根据不同的机器人应用场景分别制定有针对性的伦理准则。总之，此三方面均要求机器人伦理学的独立，以便更好地规范机器人产业的发展。

#### 四、机器人拟人论的优缺点与启示

流行于机器人伦理学中的机器人拟人论既有优势，也有缺陷。从理论层面看，拟人化使机器人获得更多的伦理关注，促使人类对现有道德框架进行更深刻的反思与研究。然而，过度拟人化可能导致机器人伦理学观点过于激进、脱离社会现实，进而引发伦理责任的过度扩展、伦理判断的混乱以及道德界限的模糊。从实践层面看，机器人设计者可以利用拟人化的优势，使机器人在人机交互中更易被接纳，但过度拟人化也带来了安全与伦理隐患，给专家学者带来诸多挑战，如责任分配困境、情感欺骗与操控等问题。不可否认，通过分析机器人伦理学的拟人论基础，可以为机器人伦理学的发展提供宝贵的启示。

第一，可借鉴动物伦理学、大地伦理学等发展机器人伦理学。拟人论广泛存在于动物伦理学与大地伦理学中，人们普遍认为动物拥有类人的感知快乐与痛苦的能力，强调自然像人类一样具有“内在价值”。相较机器人伦理

学，动物伦理学与大地伦理学的相关讨论较为成熟，在某种程度上对新兴的机器人伦理学具有借鉴与指导作用。具体来说，应严格区分道德主体与道德受体的概念，道德主体强调责任的承担，而道德受体关注权利的赋予。动物伦理学与大地伦理学承认动物与自然的道德受体地位，认为它们有资格获得道德关怀，应该被尊重与保护，但并未承认它们的道德主体地位。鉴于此，机器人虽然相较动物与自然似乎具有更高的智能与更强的自主性，但是难以具备动物的感知能力，而且对人类而言并非像自然一样完全不可或缺。因此，机器人伦理学在未来或许可承认机器人的道德受体地位，严禁虐待机器人，避免诱发人类的暴力倾向。但是，不应赋予机器人道德主体地位，相关责任必须也只能由人类承担。此外，无论是动物伦理学还是大地伦理学均强调人与非人的和谐共生，机器人伦理学应延续这一理念，通过规范制定、制度构建、技术规约等手段实现人机友好共生。

第二，机器人伦理学的发展目标是以人为本发展“更好的机器人”。现有机器人伦理学聚焦伦理批评，争论机器人地位，我们应清楚地认识到，这只是实现“更好的机器人”的手段，不应将其当作目标。机器人拟人论的优势应得到充分发挥，尤其是应发挥其在增强机器人与人类互动的舒适性与亲切感、避免将人类机器化并拉低到机器的位置方面的作用。拟人化的“伊莱扎效应”表明，人类会将技术物人格化，证明人类适应科技的程度高于科技适应人类的程度。<sup>①</sup>或许，我们应努力尝试发展能够更好适应人类需求的机器人，这意味着机器人伦理学具有很强的操作性，它应指导我们研发、设计、制造有益的机器人，并在此基础上更好地使用机器人，使之为人福祉服务。

第三，机器人伦理学是跨学科的问题学，应研究机器人应用对社会的冲击及其应对，而不是仅仅局限于伦理学与工程学。更准确地说，机器人伦理学应该是机器人 STS (Science, Technology and Society, 即科学、技术与社会) 学。詹马尔科·韦鲁吉奥 (Gianmarco Verrugio) 将机器人伦理学相关领域划分为三个层次：其一，机器人伦理学，这一层次在本质上涉及哲学问题、人文学科和社会科学；其二，机器人伦理 (robot ethics)，这一层次主要涉及科学与技术；其三，机器人的伦理 (robot's ethics)，这一层次主要关乎科幻小说。<sup>②</sup>不

---

① 参见 [美] 帕特里克·林、凯斯·阿布尼、乔治·A. 贝基主编：《机器人伦理学》，薛少华、仵婷译，人民邮电出版社 2021 年版，第 248~249 页。

② 参见 Spyros G. Tzafestas, Roboethics: Fundamental Concepts and Future Prospects, *Information*, Vol.9(6), 2018, p.3。

可否认,机器人伦理学反思会用到哲学、经济学、社会学、管理学、统计学、心理学、文化研究等诸多学科的理论、方法和案例。换言之,机器人伦理学并不属于伦理学的分支,而是属于整个哲学的应用分支。从某种意义上说,机器人拟人论对机器人研究、设计、制造和应用的支配,催生了最主要的机器人伦理学问题,机器人伦理学研究应想方设法解决这些问题,实现“机器人伦理软着陆”,使机器人产业能够健康稳健地发展。

第四,机器人伦理学应具备前瞻性、情境性、动态性。机器人伦理学通常落后于机器人技术的发展。<sup>①</sup>面对机器人拟人化引发的不确定性与极大社会影响,伦理先行尤为关键。在机器人研发阶段,应始终贯穿伦理考量,进行伦理审查与法规制定,主动预测潜在风险,兼顾短期与长期伦理问题。此外,机器人相关伦理问题具有较强的情境依赖性。例如,医疗机器人更多涉及隐私、决策等伦理问题,而工业机器人主要与失业、安全等伦理问题相关。机器人在不同应用场景涉及的伦理规范、价值观和社会影响可能大相径庭,因此机器人伦理学需要针对具体情境进行深入探讨。机器人伦理学的框架需要根据最新的技术成果和社会需求做出相应的调整,积极回应技术创新带来的伦理挑战,建立实时反馈机制,确保灵活性与适应性。

## 如何将机器人安放于人类社会之中?<sup>\*</sup>

### ——一种哲学进路的思考

刘 鹏

机器人小朵是儿童科幻小说《卧底机器人》中的主要角色,她经历了一场时间更久、难度更高的“图灵测试”,需要在人类的学校中“生活”一整个学年

<sup>①</sup> 参见[美]帕特里克·林、凯斯·阿布尼、乔治·A.贝基主编:《机器人伦理学》,薛少华、仵婷译,人民邮电出版社2021年版,第13页。

<sup>\*</sup> 本文系中国科学院学部科技伦理研究项目“数字技术的伦理研究”(XBJLL2023001)、江苏省社会科学基金重大项目“新一代人工智能重大哲学与逻辑问题研究”(24ZD006)的阶段性成果。

· 笔谈 ·

# 机器人伦理学前沿问题

刘永谋等

【主持人语】经过几十年的发展，机器人技术已达到某种“临界点”，表现为人形机器人（humanoid robot）技术加速发展。工业和信息化部在2023年10月印发的《人形机器人创新发展指导意见》开篇即高屋建瓴地判断，人形机器人“有望成为继计算机、智能手机、新能源汽车后的颠覆性产品”。人形机器人的科技研发和应用同样面临科技风险和科技伦理问题，受到全社会的广泛关注。由此，机器人伦理学成为近年来的热门研究领域，尤其是应用伦理学重要的“理论增长点”。本次笔谈由6篇文章组成，聚焦机器人伦理学发展前沿，抛砖引玉，以期推动该领域研究的进一步发展。刘永谋和白英慧讨论拟人论意识形态在机器人伦理学建构中的基础作用，分析机器人拟人论在该领域流行的原因及启示。刘鹏考察了机器人伦理风险的发生机制及治理原则，基于此指出人类社会应通过与机器人的互动、互构，构建一种新型的人机有机结构。程林在跨文化视域下考察了机器人拟人化及恐惑现象，对中国人机协存观念和机器人设计理念提出了建议。谭笑考察了小数据主义技术路线在隐私保护和权力结构均衡等方面的优势，论证了由于所需知识类型不同，这一路线不太适用于社交机器人领域。孙圣引入拟人化分析，阐述了机器人伦理模型的惩罚缺漏不可避免，以此反驳直接将之作为判决标准的可行性，提出拟人度概念，用以划分人机共存社会发展阶段，并论证了划分何以可能。杨庆峰和朱清君考察了人形机器人导致的社会角色的临时替代和永久替代问题，并探讨了由人形机器人角色替代带来的伦理学挑战。

（刘永谋，中国人民大学哲学院教授、博士生导师）

【关键词】机器人 人形机器人 伦理学 机器人伦理学

【作者简介】刘永谋，中国人民大学哲学院教授、博士生导师；白英慧，中国人民大学哲学院博士研究生。刘鹏，南京大学哲学学院教授、博士生导师。程林，广东外语外贸大学外国文学文化研究院教授、阐释学研究院兼职研究员。谭笑，首都师范大学政法学院教授。孙圣，西北师范大学哲学与社会学院副教授、硕士生导师。杨庆峰，复旦大学科技伦理与人类未来研究院教授；朱清君，复旦大学社会发展与公共政策学院博士研究生。

【中图分类号】B829 【文献标识码】A

【文章编号】2097-1125（2025）06-0005-55

## 机器人伦理学的拟人论基础<sup>\*</sup>

刘永谋 白英慧

近年来，人工智能、传感器、机器人控制与动力学、云计算与物联网、人工肌肉与柔性材料等技术的不断突破，使机器人更加灵活与自主，并使其应用场景愈加多元化、精细化。同时，机器人的研究、设计、制造和使用面临人类失业、隐私泄露、情感欺骗、责任分配等十分棘手的伦理问题，这些问题对机器人技术的发展及人类社会生活产生了不可忽视的重大影响，必须结合具体情境加以认真研究。在此背景下，机器人伦理学（roboethics）兴起并持续火热。顾名思义，机器人伦理学是研究有关机器人的伦理问题的学问。凯斯·阿布尼（Keith Abney）总结了机器人伦理学研究对象的三层含义：第一，机器人技术专家的职业道德；第二，为自动化机器人编写的道德规范的代码程序，即机器人自己的而非人类的准则；第三，机器人在具备进行伦理推理的自我意识能力时自行选择的伦理准则。<sup>①</sup>换言之，机器人伦理学研究

\* 本文系国家社会科学基金重大项目“现代技术治理理论问题研究”（21&ZD064）的阶段性成果。

① 参见[美]帕特里克·林、凯斯·阿布尼、乔治·A.贝基主编：《机器人伦理学》，薛少华、仵婷译，人民邮电出版社2021年版，第35页。

# Abstracts

## Frontier Issues in Roboethics

Liu Yongmou et al.

【 Abstract 】 After decades of development, robotic technology has reached a certain “critical point”, manifested in the accelerated development of humanoid robot technology. The Ministry of Industry and Information Technology issued the *Guidelines for the Innovation and Development of Humanoid Robots* in October 2023, which made a high-level judgment at the outset that humanoid robots “are expected to become a disruptive product after computers, smartphones and new energy vehicles”. However, the research, development, and application of humanoid robots also face scientific and technological risks as well as scientific and technological ethics issues, drawing widespread attention from the whole society. As a result, robot ethics has emerged as a hot research field in recent years, particularly as a significant “theoretical growth point” within applied ethics. This special issue features six invited articles that focus on the cutting-edge developments in roboethics, aiming to spark further discussion and advance research in this field. Liu Yongmou and Bai Yinghui discuss the fundamental role of anthropomorphic ideology in the construction of roboethics, and analyze the reasons for the prevalence of anthropomorphism in this field and its implications. Liu Peng examines the mechanisms of the occurrence of robot ethical risks and the principles of governance. Based on this, he points out that human society should build a new type of human-machine organic structure through interaction and mutual construction with robots. Cheng Lin explores the anthropomorphism and uncanny valley phenomenon of robots from a cross-cultural perspective, offering suggestions for the Chinese-style human-machine co-existence and robot design. Tan Xiao examines the advantages of the small data technology roadmap in terms of privacy protection and power structure balance, and demonstrates that this roadmap is not very suitable for the field of social robots due to the different types of knowledge required. Sun Sheng introduces anthropomorphism analysis to demonstrate the inevitability of the retribution gaps of ethical models of robots, thereby refuting

the feasibility of using them directly as a criterion for judgement. He proposes the concept of degree of anthropomorphism as the demarcation of development stages of human-machine coexisting societies and justifies its plausibility. Yang Qingfeng and Zhu Qingjun investigate the temporary and permanent substitution of social roles caused by humanoid robots, exploring the ethical challenges posed by the role substitution of humanoid robots.

(Liu Yongmou, Professor and PhD Supervisor, School of Philosophy, Renmin University of China)

【 Keywords 】 robot; humanoid robot; ethics; roboethics

### **Promoting the Common Values of Mankind: China's Contribution to the Advancement of International Rule of Law**

**Li Lin**

【 Abstract 】 The proposal of new ideas regarding the common values of humanity has not only provided a new value foundation for people of all countries to join hands in building a community with a shared future for mankind, but also offered strong value guidance for China's participation in promoting theoretical, institutional and practical innovations in the international rule of law. Since the founding of the People's Republic of China, particularly since the 18th National Congress of the Communist Party of China, China has upheld the banner of human values and civilizations and made significant contributions to safeguarding world peace and promoting the development of the international rule of law. Hence, as a responsible major power emerging on the global stage and engaging in international affairs, China must adeptly employ the rule of law. In order to employ the rule-of-law thinking and rule-of-law based approach to boost the building of a community with a shared future for mankind, greater emphasis should be placed on the coordinated advancement of domestic and foreign-related rule of law. China is also required to promote the common values of humanity, advance foreign-related rule of law initiatives, and actively engage in the development of the international rule of law, thereby contributing more Chinese wisdom and strength to the progress of international rule of law.

【 Keywords 】 common values of humanity; domestic rule of law; foreign-related rule of law; international rule of law; rule of law civilization