人形机器人的角色替代及其伦理挑战*

杨庆峰　朱清君

从本质上说，人形机器人是以具身形式存在的 AI 智能体 (agent)，它在家庭和社会中的角色和作用值得思考。人形机器人不仅能填补某个成员暂时缺失导致的情感空缺，还能在一定程度上替代永久性缺失的家庭成员。然而，这种双重替代作用也带来了一些伦理挑战，包括对原生家庭结构的破坏、是否能有效替代人类角色等。为了全面理解人形机器人的角色替代及其带来的社会和伦理问题，我们需要对角色替代进行深入研究。

一、作为 AI 智能体的人形机器人

智能体这个概念最初被用来定义人工智能，但是从 2024 年开始，这个概念仿佛成为一个新概念，被当作代表人工智能发展下一个动向的表达。智能体完全可以被看作分析人形机器人的基础概念。

（一）作为智能体的人形机器人的含义

智能体曾经被作为人工智能的本质定义。斯图尔特·罗素（Stuart Russell）等、戴维·普尔（David Poole）等将智能体作为理解人工智能的一个重要概念。在罗素等看来，理性智能体（rational agent）的概念是研究人工智能的方法的核心，“任何通过传感器（sensor）感知环境（environment）并通过执行器（actuator）作用于该环境的事物都可以被视为智能体（agent）”。① 在普尔等看来，智能体可以进行推理。“智能体由感知、推理和行为组成。”② 当前，智能体被看作未来人工智能发展的趋势。赞恩·杜兰特（Zane Durante）

*　本文系“科技企业战略发展联盟研究”（20241H06030）的阶段性成果。

① ［美］斯图尔特·罗素、彼得·诺维格：《人工智能：现代方法》，张博雅等译，人民邮电出版社 2023 年版，第 32 页。

② ［加拿大］普尔、麦克沃思：《人工智能：计算 Agent 基础》，董红斌等译，机械工业出版社 2015 年版，第 6 页。

等“把智能体定义为一类交互系统，能够感知视觉刺激、语言输入以及其他基于环境的数据，能够产生有意义的具身行动”。该研究提出了由与规划任务和技巧观察有关的环境与知觉、学习、记忆、行动和认知五个模块构成的新的智能体范式，并将智能体划分为通用智能体、具身智能体（行动智能体和交互智能体）、模仿和环境智能体、生成智能体、知识和逻辑推理智能体以及大语言模型和视觉大模型智能体六类。①

人形机器人被看作智能体意味着至少四个方面的含义。

其一，人形机器人是多模态智能体。这是一种原初的、基本的具身描述。根据杜兰特等的看法，多模态智能体带来一种范式更新，强调多模态。以往的智能体范式只是基于单向的输入 – 输出以及人类监督的智能体的。本文第一作者也曾经指出，通用人工智能具有多任务、多语境与多模态的“三多”特征。②

其二，人形机器人是具身智能体。在一定意义上，大语言模型可以被看作理性智能体。对大语言模型来说，人的需求是单向度的，人以提示词（prompt）的形式表达自身的困惑和问题，然后大模型做出有效回应。而具身性则是人形机器人的一个特性，如具有人形的外貌、流畅的声音以及柔软的肌肤触感。

其三，人形机器人是环境智能体。人形机器人要能够迅速适应各种自然环境、技术环境和社会环境，要能够表现主体间性和交互能力。杜兰特等把环境智能体与模仿放在同一类别，这对人形机器人来说并不适合。虽然当前的人工智能研发背后的逻辑是模仿类似，从理性角度实现机器人的推理能力、感知能力、决策能力和行动能力，但是对人形机器人的更高的要求是社会性的类似，即能够像人一样构建社会关系，继而表现出利他的特征。

将人形机器人看作具身智能体和环境智能体，依据的是杜兰特等对智能体的分类。笔者认为，具身性是人形机器人的特性，没有身体，一些具身任

① 参见 Zane Durante, Qiuyuan Huang and Naoki Wake et al., Agent AI: Surveying the Horizons of Multimodal Interaction, https://arxiv.org/abs/2401.03568, pp. 2, 15, 20–24。在分析六类智能体时，该研究使用了一种现象学风格的划分：primary subject topics /secondary subject topics。在该研究看来，多模态智能体是原初智能体，也就是通用智能体。次生智能体包括具身、知识和逻辑推理等五种。在对记忆的划分中，胡塞尔使用过原初记忆和次生记忆的概念。

② 参见杨庆峰：《通用人工智能是多模态吗》，《哲学动态》2024 年第 9 期，第 43~48 页。

务，如陪护机器人与人类的必要的身体接触，是无法实现的。然而，杜兰特等的分析忽略了社会性，多为功能性的分析，而社会性并不是功能性，更多地是对本质属性的规定。因此，人形机器人也应是社会智能体。

其四，人形机器人是社会智能体。在计算机领域，大多数学者把智能体看作理性智能体，杜兰特等也注意到智能体的语境（context），而语境是社会性得以建立的基础，但杜兰特等忽略了这个规定性。人形机器人必然要求社会性，而且这种社会性必然是基于语境的，而不是根据某种逻辑规则或伦理规则构建出来的。

（二）人形机器人社会性的实现类型

要融入人类社会，人形机器人必然要具有社会性和语境性；从智能体角度来说，社会智能体和语境智能体是未来人形机器人构建的基本要求。

社会性要求人形机器人能够构建自己的社会关系。费希特与马克思对人类构建社会关系做出了阐述。前者认为需要构建社会关系，后者认为人的本质是一切社会关系的总和。随着数智时代的到来，机器人交往、虚拟交往变得日益重要。此外，人类社会中一种普遍的情结，即强者对弱者的照顾也是重要的规定性。因此，根据社会性的双重规定，人形机器人构建的特点就体现为以社交机器人和陪护机器人为主。①

语境性要求人形机器人能够学习、理解并置身于某种情境，即做到同情与共情。同情与共情不仅可以被看作人类道德的起源，而且可以被作为人类交往的深层根据。社会交往的外化导致了社交互动机器人的出现，它们通过模拟面部表情和三维头部特征表达情绪状态，对人类用户产生了深远影响。②这些情绪状态可能并不“真实”，而是源于将机器人拟人化，即将人类特有的属性、动机、意图或情感赋予非人类实体。拟人化的三因素理论着重考虑了三个核心的心理决定因素：以人为中心的知识的可及性与适用性、效能动机和社交动机。③

---

① 目前，机器人的类型有多种，如 social robots、assistive robots、companion robots、care robots，国内对此的翻译较为混乱。

② 参见 C. Breazeal, Emotion and Sociable Humanoid Robots, *International Journal of Human-Computer Studies*, Vol.59(1–2), 2003, pp.119–155。

③ 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。

二、角色替代：临时替代与永久替代

根据皮尤研究中心的数据，美国 18 岁以下儿童中，近四分之一（23%）生活在单亲家庭中。① 相较欧美国家，中国的单亲家庭占比较低，但随着社会观念的变化和离婚率的上升，预计未来可能会有所增加。同时，随着人口老龄化加速，中国老年人群体的照护需求日益突出。截至 2023 年底，全国 60 岁及以上老年人口已达 2.97 亿，占总人口的 21.1%。② 预计到 2035 年左右，这一数字将突破 4 亿，中国将进入重度老龄化阶段。③ 可以看出，社会发展对社交机器人特别是陪护机器人的需求更为紧迫。

从本质上看，社会交往、陪护他人是智能体社会性的体现。从功能上看，社交机器人旨在与人和其他机器人构建新的社交关系；陪护机器人旨在成功融入家庭环境，并充分发挥潜在的照护与陪伴作用。这会产生角色替代的问题，如以下可能的情况：陪护机器人取代了原先的、居于千里之外无法照顾双亲的子女，这是临时替代；社交机器人则能够让社交恐惧症人士建立一种非人的社交关系网络，从而实现平衡，满足其内在的心理诉求，这接近一种永久替代。

对社会和家庭来说，人形机器人的角色替代会表现为两种类型：临时替代和永久替代。机器人在临时替代中替代因工作繁忙或其他原因无法参与家庭事务的成员，在永久替代中则替代缺失家庭成员从而导致形成新的家庭结构。而划分这两种替代的重要基础是替代是否为必然事件。

首先，永久替代涉及机器人永久替代某个成员以致形成新的家庭结构。这是一种必然事件。原生家庭成员因为疾病或者意外事故离世，这种情况会导致成员缺失，给其他成员带来极大伤害。修复伤害和处理创伤记忆就成为亲属必须面对的课题。因此，机器人被赋予替代因死亡或其他不可逆原因而

① 参见《美国单亲家庭儿童比例接近 1/4，全球最高》，https://www.guancha.cn/internation/2019_12_15_528462.shtml，2025 年 3 月 14 日。

② 参见《2023 年度国家老龄事业发展公报》，https://www.gov.cn/lianbo/bumen/202410/content_6979487.htm，2025 年 3 月 14 日。

③ 参见《国家卫健委：2035 年左右 60 岁及以上老年人口将破 4 亿 占比将超 30%》，https://news.cctv.com/2022/09/20/ARTInjejQDvmMaZi5jzTPHYT220920.shtml，2025 年 3 月 14 日。

长期缺失的家庭角色的必然性任务，这种情况就是永久替代。然而，这会带来情感替代难度、伦理问题以及家庭动态变化的挑战。此外，机器人无法完全替代人类关系的深度与复杂性，尤其是在处理情感依恋方面。

其次，临时替代涉及机器人短期替代因工作或其他原因无法参与家庭事务的成员，这是一种偶然事件。电影《非诚勿扰 3》展示如下情节。年过半百的主角秦奋的妻子因追求环保事业而离开家庭，十年过去了，他的妻子还没有回来。秦奋的朋友为了弥补秦奋的情感需求和照顾他的生活，购买了类似他的妻子的人形机器人。直到有一天，秦奋真正的妻子回来了，于是人机博弈就此开始。这里的人形机器人起到了临时替代作用，因为真正的妻子最终会回到这个家。临时替代存在依赖风险和角色割裂的潜在问题。为此，应限定机器人的替代时间，明确其临时职责；设计双向互动机制，让缺席的家庭成员能远程参与家庭事务；对家庭其他成员进行教育，明确机器人的角色和边界，如将其功能定位于鼓励儿童的独立思考和社交互动，而非替代父母的指导作用。

此外，我们还可以通过道德责任的转让与否，对永久替代和临时替代做出区分。费希特在《伦理学体系》中将道德责任区分为普遍职责和特殊职责，前者指主体不能转交的职责，而后者指主体可以转交的职责。[①] 对主体来说，法律规定的责任义务是无法转交的。但是，随着技术的发展，这类责任也有可能发生变化。因此，所有被替代的责任都属于特殊职责，即可以转交给机器人的主体道德职责。在永久替代关系中，主体的道德责任永久地由人转让给机器人，原先的主体退场，机器人成为道德责任的承担者；而在临时替代中，道德责任不是永久转让给机器人，而是临时转让，换句话说，没有完全转让给机器人。

当然，在功能替代中，机器人承担身体或认知能力受限的家庭成员无法完成的功能性任务，这可能会引发个体尊严维护、依赖以及家庭角色变化的问题。因此，需要强化机器人的辅助角色，避免永久替代；进行功能个性化设计，满足被替代者的具体需求；设计心理支持系统，让被替代者感受到尊重和支持。对临时替代来说，原有成员会因机器人替代对家庭的重塑而产生不满，由此带来伦理与家庭关系问题以及恶性替代风险的挑战。

① 参见［德］费希特：《伦理学体系》，梁志学、李理译，商务印书馆 2010 年版，第 270 页。

三、双重替代的四重伦理挑战

正如前面提到的，角色替代存在的前提之一是拟人化，即为非人类事物赋予人类特征的倾向。拟人化作为一种“进化适应”存在，帮助早期人类辨别友敌，是一种关键的生存机制。事物与人类相似度越高，拟人化现象就越可能发生。基于拟人化，人形机器人角色替代将带来如下挑战。

第一，人形机器人的安全性是技术层面的挑战。需要严格考量机器人的尺寸设计、报警功能及错误管理机制，这些直接关系到不同年龄人群的安全和健康。目前，许多项目更多关注机器人的功能设定及商业化产品的技术要求，而忽视了护理的本质——一种涉及照顾者与被照顾者关系，兼具社会、道德和政策维度的活动。机器人应被视为护理过程的辅助元素，而非照顾者的替代品。当前政策多聚焦于机器人的经济影响及物理安全，忽视了社会文化因素及长期人机互动的实际影响。负责任的机器人技术发展需要重新审视任务分配，平衡机器人与人类的角色。

第二，机器人社会性的合法化是角色替代面临的根本难题。用于临时替代与永久替代的人形机器人（如为有故去亲人或长期缺席亲人的家庭成员提供陪伴的机器人）面临的核心伦理挑战是如何承认机器人的社会性。当人形机器人的同理心和同情心足够强时，它可能会成为用户情感依赖的对象，特别是在家庭成员长期缺席的情况下。这种情感依赖可能导致依赖者难以形成健康的人际关系，对机器人“感情”产生错觉，把机器人当作“他者”看待。有研究表明，大语言模型在理解错误信念、解读间接请求、识别讽刺和失礼行为等方面可以模拟人类。[①] 例如，在替代故去的父母角色时，机器人能够理解死者的离去以及亲属的失落感，这就成为人形机器人被当作他者的关键。

第三，忽视人形机器人的语境性是实现角色替代面临的文化难题。在设计和部署这些机器人时，还需要充分考虑个人语境、文化情境和社会背景。大模型在对社会情境的理解上可以超过人类，能够适应和回应他人的行为。[②] 社交

① 参见 James W. A. Strachan, Dalila Albergo and Giulia Borghini et al., Testing Theory of Mind in Large Language Models and Humans, *Nature Human Behaviour*, Vol. 8(7), 2024, pp. 1285−1295。

② 参见 Justin M. Mittelstädt, Julia Maier and Panja Goerke et al., Large Language Models Can Outperform Humans in Social Situational Judgments, *Scientific Report*, Vol.14, 2024, p.27449。按照杜兰特等的分类，大模型属于第六类智能体，可用于规划任务。

机器人被视作非控制环境下的护理补充，旨在陪伴、支持独立生活、减轻孤独感等。然而，社交机器人的概念常被简化为技术实现，从而忽视了其社会性和文化情感维度的复杂性。工程师主导的研发往往缺乏对社会需求的深入理解，导致机器人虽能模拟社交行为，却难以达到真正的人类社交互动水平。不同文化对拟人化和人与机器互动的接受程度各异，由此可能产生伦理冲突。

第四，对人形机器人的过度依赖可能导致自主性丧失是替代过程面临的伦理难题。人类可能为机器人赋予人类特征，进而对其行为产生不合理的期望。当机器人无法满足这些期望时，可能引致失望或心理伤害。同时，机器人在照护过程中的同理心表现尤其是情感识别和反馈的能力，是评估其伦理合规性的重要标准。[①] 当前，人形机器人通常依靠算法分析面部表情、语音语调和身体语言，试图以此推测人的情感状态。然而，机器人在情感识别方面的能力仍然存在局限，尤其是在处理复杂情感时。例如，虽然机器人可以识别出哭泣等明显的情感表现，但很难理解其中的细微差别，如悲伤与失望的不同，或者缺乏合适的安慰反应。更为重要的是，机器人无法真正体验情感，也不具备人类的情感共鸣能力，这使机器人在情感支持上的局限性更加明显。[②] 这种“虚假同理心”现象引发了对伦理误导的担忧，即技术的高度模拟可能会误导用户相信机器人能够提供如人类一般的情感支持，使用户产生过强的情感依赖。

总而言之，人类的时间和注意力有限，而在当今社会，这种有限性正被“非生产性的虚假社交娱乐”特别是与机器人的关系发展占据。这种趋势可能会导致我们忽视真实人类之间的交往，从而引发社会关系失衡的“机器人时刻”。因此，需要思考人形机器人带来的新的人机关系的真正本质。机器人的角色替代可能会导致社会孤立，这是因为技术只能提供一种表面的陪伴感。[③]

（责任编辑：李润东）

① 参见 Patrick Lin, Keith Abney and George A. Bekey, eds., *Robot Ethics: The Ethical and Social Implications of Robotics*, Cambridge, Massachusetts: MIT Press, 2012。

② 参见张晶晶、吴鹏、曹琪等：《基于认知科学的社交媒体用户情感建模研究综述》，《信息资源管理学报》2021 年第 1 期，第 59~69 页。

③ 参见 Joanna J. Bryson, Robots Should Be Slaves, in Yorick Wilks, ed., *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, Amsterdam and Philadelphia: John Benjamins Publishing, 2010, pp. 63−74。

·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*

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。① 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 Superviser, 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