

判”，^①需要在认清新技术可能带来的各类问题的基础上，开启建设性的工作。如同拉图尔所说的，对人类社会中新技术带来的各种“怪物”和各种新型的“弗兰肯斯坦”，要热爱它们，要“像对待我们的孩子一样悉心照料”它们，不能因为技术会犯错就放弃，也不能溺爱技术而对之疏于管教。^②只有这样，才能在潘多拉魔盒的底部找到希望。

克服恐惑谷效应：跨文化视域下的 中国人机协存社会^{*}

程 林

人工智能伦理研究者斯文·尼霍姆（Sven Nyholm）曾假设，如果对无具体面容的人形机器人进行改造，赋予其中国人的外貌、名字和行为方式，那么人们对待改造前后的机器人的态度就会非常不同。^③不难理解的是，名称、外观、背景故事和设计理念会影响我们对机器人的感知，进而影响人机关系和伦理。与之类似，是以拟人化乃至仿真化为宗旨来设计机器人，还是出于某种异化模拟带来的不安而对拟人化产生抗拒，这既是机器人设计的现实问题，也是人文学界关心的审美、心理与伦理话题。机器人恐惑现象或曰“恐惑谷效应”（uncanny valley effect）的存在，可能会从审美、心理和情感方面导向对类人机器人乃至人机共存观念的排斥。恐惑谷效应亦与社会文化、人机关系和伦理联系紧密，这在既有研究中尚未得到足够重视。本文从机器人

① [荷] 马克·舒伦伯格、里克·彼得斯编：《算法社会：技术、权力和知识》，王延川、栗鹏飞译，商务印书馆 2023 年版，第 13 页。

② 参见 Bruno Latour, Love Your Monsters: Why We Must Care for Our Technologies as We Do Our Children, in Michael Shellenberger and Ted Nordhaus, eds., Love Your Monsters: Postenvironmentalism and the Anthropocene, Oakland: Breakthrough Institute, 2011。

* 本文系广东外语外贸大学外国文学文化研究院 2025 年度招标课题“文明互鉴视域下的东西方 AI/ 机器人文化研究”（25ZBKT03）的阶段性成果。

③ 参见 [瑞典] 斯文·尼霍姆：《序言·致中国读者》，《人与机器人：伦理、行动与拟人论》，刘铮译，上海交通大学出版社 2024 年版，第 2 页。

拟人化、恐惑现象与恐惑谷效应出发,在日本和欧美跨文化视域下讨论关于仿真机器人及恐惑现象的文化差异,最后落脚到中国文化中的相关现象,进而讨论中国式机器人文化、人机协存观念与机器人设计理念。

一、拟人化倾向：为何机器人被设计得像人？

不同程度上的仿人和拟人是多数机器人特别是仿人机器人的基本特征。从外观类人程度看,仿人机器人可以分为两类:人形机器人(humanoid robot)和类人机器人(android)。前者强调对人类外形基本特征和功能的模仿,是当今科技研发的热点;而后者强调外观的高度仿人和逼真,它长久以来都是科幻想象,但目前也有技术实践。从机器人人文角度讲,机器人可被理解为对人之整体或部分的机器模拟。在中文日常语用中,“机器人”概念的内涵被过度地拟人化,我们要么接受它极其广泛的所指,要么感到困惑。一方面,它嫁接了“机器”与“人”两个迥异的概念,其名词性偏正结构在字面上宣示着“机器人”是“机器式的人”而非“人式的机器”。另一方面,即便我们将“机器人”等同于“人式机器”,仍要面临一个问题:人们不仅将“通用 robots”(generalist robots,实为通用仿人机),而且将四足机器狗、自动扫地盘、胃镜胶囊、工厂机械臂等“专用 robots”(specialist robots,实为专用自动机)乃至聊天程序也叫做“机器人”或“××机器人”,这种能指与所指的错位造成了认知困惑,这是因为“机器人”这一能指在我们脑海中触发的直接想象多是仿人机器人,即人之整体的投影而非其他。

实际上,拟人化是人在想象和设计机器人时的常见“冲动”,并有两个维度:一是人们明知某种机器没有生命,但仍将其人格化,待其如生命体,如“伊莱扎效应”(Eliza effect);^①二是人们在想象和设计机器人的过程中为机器人赋予类人外观、特征以及“机器人设定”或社会角色。人类自身的机器人化有时令人避之唯恐不及,让机器仿人却很常见——即便在部分场景中,实现仿人不仅技术难度高、操作难度大,而且成本高、耗能大,也没有必要性。扫地机等专用机器人就是如此,手持扫帚的人形机器人在现阶段不会比自动扫地盘更为实用。与之相比,非人形设计在功能优化、成本控制、特定场景

^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。

适应性以及减少情感依赖和伦理阻力等方面具有优势。然而，人形机器人在当下产业界和媒体上的热度不减，科幻机器人形象的类人程度也经常很高，那么为何人们会赋予机器人人形或类人特质？

从人类思维角度讲，让机器人仿人，不管是“仿身”、“仿智”还是“仿生”，都是人类想象和制造机器时难以抑制的人类中心主义式本能和执念，都“源自人类以自我为万物尺度的思维和认知惯式”，^①最终有助于人类的自我认知定位与思辨。机器仿人的思维惯式在科幻作品中尤为明显，它经常转化为机器人希望变成人的愿望。但这种愿望在山本弘的小说《诗音翩然到来之日》中也会被讽刺：在面对“你不想变成人类吗？”这个问题时，诗音回应道，“我不想变成人类。我不是铁臂阿童木。我看了那部漫画，但是我无法理解为什么铁臂阿童木想变成人类。那是人类想出来的故事”。^②

从人机交互角度讲，拟人化设计有助于增强人们在面对机器人时的熟悉感和信任感，使机器人更好地扮演特定社会角色，更符合人们以人之预期、习惯和逻辑进行交互的需求，这在情感、陪伴与教育领域尤为明显。有学者指出，“人们更愿意用人际交往的方式与机器打交道”，而且“相比具有机器外观的机器人，人们对人形机器人有着更高的共情程度……社会成员会把人际交互的社交脚本应用到人机交互上”。^③在电影《机械姬》中，凯莱布曾询问为何智能机器不是灰盒子而是有性别的机器人。与之相应，当智能机器以漂亮女性面目出现时，人机交互的内容和方式都会变化，因为其中掺杂了情感等因素。

从日常应用角度讲，人形或类人机器人成为当下的科技与社会热点，这见证着人类对机器人赋能的拓展与转变，即从专能到多能的拓展、从纯粹工具到潜在伙伴的转变。人形机器人在部分日常场景中更具人类环境适配性，这是因为仿人才能更好、更通用地适应人类的物理世界特别是城市空间。在此情况下，机器人可直接利用现有的人类工具与环境而无须另配专用系统。

此外，在科幻叙事中，人形或类人机器人经常是世俗造物主诉求的显影。这是因为，在创造类人机器人时，科学家等世俗造物主通常并非旨在制造机器，而是意在创造生命。这类科幻叙事常能进一步加深人们在想象机器人时

① 程林：《仿身、仿智与仿生：一段未完待续的仿人机器文化史》，《东岳论丛》2025年第4期，第71页。

② [日]山本弘：《艾比斯之梦》，张智渊译，新星出版社2023年版，第257页。

③ [丹麦]马尔科·内斯科乌编：《社交机器人：界限、潜力和挑战》，柳帅、张英飒译，北京大学出版社2021年版，第302、305页。

的拟人化倾向。从哲学角度出发，拟人而异质的机器人最易引发经典的审美、伦理与哲学讨论。例如，日本机器人工程师石黑浩希望制造尽可能像人的机器人，旨在探索人与机器复制之间的对比与交互以及人机之间的模糊边界。而从科技和商业资本的宣传角度讲，仿造人的机器人设计能够营造一种超过真实科技实力的表象，常被用于夸大宣传。

二、拟人化风险：机器人与恐惑谷效应

应对机器人的拟人化予以辩证审视：一方面，拟人既是人类思维惯性，有时又可以用于改进人机交互；另一方面，拟人不仅会改变人们对待机器人的方式，而且更易引发情感和伦理问题，还可能揭开机器人现象潜在的一个阴暗面，即当机器人在外表等方面过度拟人时，就有跌入“恐惑谷”的风险。机器人想象中的拟人化冲动，是恐惑谷效应出现的前提。仿真机器人可能会令人感到审美心理意义上的不安，这在国外文化史中有较长传统。除去缺乏心理深度的历史简略记载，机器人恐惑现象经历了文学审美、心理分析和工程设计三个阶段。

其一，文学审美阶段。在19世纪初德国浪漫主义时期，E. T. A. 霍夫曼（E. T. A. Hoffmann）在小说《仿人自动机》和《沙人》等早期机器人叙事中就塑造了扮演人类角色、混入人类日常生活的仿人机器形象，令人疑惑于其为人而异人的边界状态，并在字里行间营造了令人恐惑不安的美学氛围。

其二，心理分析阶段。在20世纪心理分析的兴起时期，德语文化区学者恩斯特·延齐（Ernst Jentsch）和西格蒙德·弗洛伊德（Sigmund Freud）分别在《恐惑心理学》和《论恐惑》中以霍夫曼的《沙人》中的仿人自动机形象为案例，讨论了恐惑心理学与美学现象。例如，延齐认为，仿人机器随着拟人程度的提高只会变得更令人恐惑不安：“那些像真人一样大小、能完成敲鼓或跳舞等复杂任务的仿人自动机，能轻易地让人感觉不安。这种机械装置越是精密，对人形的模仿越是到位，就越能引起这种特殊效果。”^①

其三，工程设计阶段。日本机器人工程师森政弘在随笔《恐惑谷》中指出，人对机器人的好感会随着机器人人类程度的提高而提升（至第一峰），但

^① Ernst Jentsch, Zur Psychologie des Unheimlichen, *Psychiatrisch-Neurologische Wochenschrift*, No. 23, 1906, p. 203.

这种正相关不会一直持续，这是因为当机器人在非常像人但又与人存在差异时，其带来的“亲和感”可能会骤然下降，跌入恐惑谷；只有当机器人与人类的差异缩小至人机莫辨时，人的好感才会再次上升（至第二峰）。森政弘建议，机器人设计应以适度仿人为目标。^①以拟人化为标尺，森政弘的恐惑谷设想中的第一峰、谷底和第二峰分别对应适度、过度与完全拟人。

恐惑谷是“关于人们对机器人外表可能产生的情感变化的假设”，^②是一种“宽谱”现象，即人们在相对较大的范围内或较多的情况下都会跌入恐惑谷。恐惑谷效应可能导致仿人机器人的他者定位和对人机协存理念的排斥。但在仿真机器人与恐惑现象方面，日本和欧美社会及学界有着不同的姿态。

三、跨文化视域：东西方文化中的机器人恐惑现象

克服恐惑谷效应有规避和跨越恐惑谷两条路径。关于恐惑谷是否能被跨越，学界也有两种看法。一是欧美文化理念中的不可跨越论，如“恐惑墙”（uncanny wall）假设，即随着时间的推移，人类对技术的识别能力不断增强，跨越恐惑谷的难度也会不断提高，这会让恐惑谷的右侧形成一堵没有上限的墙。^③如果从欧美文化观念出发，将仿人自动机、类人机器人、CGI（计算机生成图像）形象、数字人乃至赛博格或后人类视为“技术他者”（technical other）在不同时期的表征，并将恐惑谷效应归结于人在异化复制和科技他者面前的自我认知冲击以及不信任和无共情，那么恐惑谷效应似乎难以消解，并且会以更新面目复现。美国未来学研究者吉米斯·卡西奥（Jamais Cascio）还曾提出“第二恐惑谷”（the second uncanny valley）设想，即当今的健康人在演变为未来的“超人”或“后人”时也可能跌入恐惑谷，这意味着未来人与机器人的恐惑谷现象会以当今的健康人为轴互为镜像。^④此外，机器人在智能、情感和行为模拟等诸多方面都可能会引发审美心理层面的恐惑现象，仅聚焦外观是不全面的。

① 参见[日]森政弘：《恐惑谷》，江晖译，《外国文学动态研究》2020年第5期，第85~92页。

② 江晖：《恐惑谷：错觉还是科学？》，《信睿周报》2024年第13期，第1页。

③ 参见Angela Tinwell et al., The Uncanny Wall, *International Journal of Arts and Technology*, Vol.4(3), 2011, pp. 326-341。

④ 参见Jamais Cascio, The Second Uncanny Valley, http://www.openthefuture.com/2007/10/the_second_uncanny_valley.html, 2025年3月14日。

二是东方学者眼中的可跨越论。例如，有中国学者指出，“森氏恐惑谷无须专门跨越，或者正在被跨越……这种情况也再一次证明技术接受与技术拒绝始终是历史性、情境性、地方性和建构性的现象”。^①从工程和实用角度考虑，森政弘建议规避恐惑谷，但作为人机关系思考者和佛教徒的他曾至少三次提及跨越恐惑谷问题：一是森政弘在随笔《恐惑谷》中提及的理论可能，即当机器人与健康人无异时的情形；二是他在2005年提出，佛像应取代健康人站在第二峰，这是因为佛像面孔“高尚而雅致，不受尘世的困扰，散发庄严的光芒”，是“比人类更加迷人、更和蔼可亲的事物”；^②三是他在2011年提到，以日本青年女性为模型的仿人机器人HRP-4C已可算作跨越恐惑谷的案例，尽管其在仔细观察时仍令人感觉奇怪。^③其中，第二点尤其值得关注。森政弘还希望借助佛教理念和慈悲胸怀来形塑人机关系，认为机器人像万物一样有佛性，只有当机器人与人类平等时，人机协作才能产生好的结果，^④因此其机器人理念又被称为“佛教机器人学”（Buddhist robotics）。^⑤对他而言，跨越恐惑谷既是机器人工程学议题，也需要宗教等文化因素的介入。

石黑浩与德国哲学学者马库斯·加布里埃尔（Markus Gabriel）的对话同样展现了机器人恐惑现象的跨文化维度。尽管承认恐惑谷效应存在，但石黑浩仍执着于研发“双子机器人”（geminoid），即如同双胞胎或镜像般类人的机器人。然而其“机器人胞弟”在欧洲巡展期间令不少德国观众难以接纳，加布里埃尔亦有同感。他认为，人与人性的定义对德国人来说确定的，高仿真机器人意味着对人之固有定义和人性的挑战，人之概念的去稳定化和人性

① 刘永谋：《森氏恐惑谷无须专门跨越，或者正在被跨越》，《信睿周报》2024年第13期，第9页。

② Karl F. MacDorman, Masahiro Mori und das unheimliche Tal: Eine Retrospektive, in Konstantin D. Haensch, Lara Nelke and Matthias Planitzer, eds., *Uncanny Interfaces*, Hamburg: Textem Verlag, 2019, p. 233.

③ 参见 Norri Kageki, An Uncanny Mind: Masahiro Mori on the Uncanny Valley and Beyond – An Interview with the Japanese Professor Who Came Up with the Uncanny Valley of Robotics, <https://spectrum.ieee.org/an-uncanny-mind-masahiro-mori-on-the-uncanny-valley>, 2025年3月14日。

④ 参见 Masahiro Mori, *The Buddha in the Robot: A Robot Engineer's Thoughts on Science and Religion*, trans. Charles S. Terry, Tokyo: Kosei Publishing, 1999, pp. 173–182.

⑤ 参见 Wayne Borody. The Japanese roboticist Masahiro Mori's Buddhist inspired concept of "The Uncanny Valley", *Journal of Ethics and Emerging Technologies*, No. 1, 2013, p. 36.

的破坏有可能会带来威胁，因此“永远不要让人们对人性的认识成疑”。^① 在此，加布里埃尔不仅认为高仿真机器人令人不安，还将其上升到了哲学层面，认为这会动摇人之为人的信念。但石黑浩以人之科学定义尚不存在为出发点，认为人机边界模糊是好事，新物种也会产生。^② 与加布里埃尔视仿真机器人为威胁人类自身认知的他者不同，石黑浩追求的是人在自身镜像面前的自知、自适以及人与机器人的协同进化。

从以上差异可见，东西方文化传统和学界对仿真机器人及其恐惑现象的理解有所不同，森政弘与石黑浩对恐惑谷效应的超越或搁置主要依靠佛教信仰与人机协同进化理念，而加布里埃尔则坚持对人和人性的传统理解，并不认同石黑浩的理念。双方既有不同学科领域的观念差异，也有社会文化、人性和人机伦理观念的差异。在对机器人的态度如负面感知方面，德国、美国（代表欧美传统）和日本社会都有其代表性观念或现象。

除了加布里埃尔对高仿真机器人及其恐惑现象的上述解读，在早期机器人的恐惑现象方面，德语文化区在文学叙事（如霍夫曼的早期机器人叙事）、心理学与美学（如延齐的《恐惑心理学》与弗洛伊德的《论恐惑》）以及早期电影艺术〔如弗里茨·朗（Fritz Lang）执导的《大都会》〕中都扮演着引领角色。此外，在当今的助老机器人设计领域，以及在仿人机器人与心理学的交叉研究中，机器人恐惑现象在德国学界仍是重要议题。

与德国类似，美国同样有焦虑型机器人文化，恐惑机器人是这种文化的表征之一。此外，美国文化特别是科幻文艺还凸显了对机器人奴仆及其失控的想象，这成为“机器人威胁论”和“机器人恐惧症”^③ 出现的原因之一。反乌托邦文艺还被认为“对西方民众技术治理的成见形成产生了最重要的影响”。^④

在日本，森政弘提出了恐惑谷效应，它被日本机器人工程学界接受，但在森政弘和石黑浩等人机关系思考者看来，恐惑谷效应有被克服或缓解的可

① 参见两者对谈纪录片《欲望时代的哲学》，<https://www.bilibili.com/video/av31555231/>，2025年3月14日。

② 参见 Hui Jiang, Lin Cheng and Hiroshi Ishiguro, The Blurring of the Boundaries between Humans and Robots is a Good Thing and a New Species would be Born: An Interview with Hiroshi Ishiguro, *Technology and Language*, No.1, 2022, pp. 40–46。

③ 参见程林：《终结者的幽灵：美国科幻电影中机器人恐惧症的表征、渊源与争议》，《当代电影》2025年第4期，第55~60页。

④ 刘永谋：《试析西方民众对技术治理的成见》，《中国人民大学学报》2019年第5期，第145页。

能，两者希望借助佛教机器人学或人机协同进化论等理念来消弭人机关系中潜在的张力。在信仰观念、日式伦理观、现代化进程和科幻文艺的促进下，畅想人机协存的愿景式机器人文化在日本社会文化中占据主流。

由此可见，欧美与日本社会文化都承认机器人恐惑现象的潜在可能，但对它的理解不同。尽管科技全球化进程会一定程度地减少不同国家在科技实践方面的差异，但机器人恐惑现象在不同文化中仍有不同的表现与应对，文化因素在考察机器人恐惑现象和人机关系方面扮演着重要角色。与之相应，不同社会文化理解人机关系的方式亦有差异，德国学者托马斯·拉姆格（Thomas Ramge）的观点就值得进一步探讨：“机器人在欧洲、美国和日本分别是‘敌人’、‘奴仆’和‘朋友’，在中国则是‘同事’。”^①

四、中国机器人文化与人机协存社会

人们对仿人机器人和恐惑现象的认知既有时间维度，即随着人机交互时间和经验的增加，机器人对人之心理的影响程度会发生变化，也有文化和跨文化维度，例如欧美和日本文化对仿真机器人及恐惑现象的认知具有异质性，其原因在于两者的伦理观念、物我关系、现代化进程、技术文化等均有差异。对上述异质性的研究应服务于中国社会文化的自我观察。在人形机器人越来越普及的趋势下，机器人恐惑现象是否会在中国社会文化和应用实践中成为显性问题呢？

本文的回答是：不会。这是因为在社会实践方面，中国的现代化进程和社会文化总体上决定了社会整体偏向科技乐观态度，并积极推动机器人等前沿科技的应用。机器人是当代中国社会在展示科技实力与“中国智造”时最常用的名片，这在北京冬奥会的“智慧冬奥”场馆及相关新闻报道中展现得淋漓尽致。在中国社会中，机器人常见的命名方式包括“小化名”（“小×”，如小爱）和叠音名（“××”，如佳佳），这类昵称传达了好感、亲昵或期许，而非相反。

在人造人认知传统方面，恐惑谷效应常常源自一种因为辨别真假人类而引发的自我认知失调或归类困难，对真与假、本体与复制差异的纠结在欧美文化传统中尤为明显，但这在中国文化中并不显著。中国古代“类机器人故

^① Thomas Ramge, *Mensch und Maschine: Wie Künstliche Intelligenz und Roboter unser Leben verändern*, Ditzingen: Reclam, 2018, p. 20.

事”（人偶故事）并没有深度与细腻地描写个体心理，但也未将人造人想象为“恐惑他者”。从《列子·汤问》中偃师制造的歌舞人偶，到清代宫廷中的写字人偶，再到当代各类杂耍机器人或春晚表演机器人，在中国传统和当代社会中机器人并非恐惑他者，而是常常表现为娱乐技术。

在大众文艺方面，恐惑谷效应被认为是“爵迹”系列电影票房失败的原因之一，^①但整体而言，这种效应在中国当代文艺中的存在感并不强。从魏雅华的小说《温柔之乡的梦》到情景剧《编辑部的故事》第22集《人工智能人》，从春晚小品《机器人趣话》再到电影《非诚勿扰3》，这些作品均塑造了人机莫辨的女性机器人形象，但恐惑谷现象均未在其中出现。从《人工智能人》和《机器人趣话》及其创作背景可见，当时社会倾向于认为，人有各种缺点，但机器人是完美的——这“与中国社会的现代化进程以及科技进步所被寄予的厚望有关”，人们“相信机器人及其所代表的新科技能为社会和个人发展带来福利”。^②中外科幻作品对机器人形象的塑造也有差异。在瑞典科幻剧《真实的人类》及其英美改编版《人类》中，均有令人感到恐惑不安的仿真机器人的特写镜头，但其中国改编版《你好，安怡》弱化了“芯机人”的这种效果。瑞典和英美版片名都言说着人类与复制人之间纠缠不清的模拟、虚实与替代关系，而中国版则用“你好”向名为安怡的机器人打招呼。在购置家政机器人的情节中，只有中国版中有以下特写镜头：在“芯机人”卖场的入口，面对琳琅满目的机器人，每个人脸上都写满了惊叹、兴趣与期待——仿佛他们此刻带着这种姿态迈入了人机协存的未来。

在日常语用方面，中文“机器人”概念本身似乎让中国人更习惯于机器人的类人属性。在20世纪80年代，研究人员刘海波曾从技术从业者视角提出，robot与译名“机器人”之间存在含义错位，并从词源、含义和读音三方面考虑，建议使用“劳伯”的译法，^③但“劳伯”未能撼动“机器人”，理工与人文学界对机器人认知的割裂也仍在继续。实际上，“机器人”概念出现在1911年的中国报刊上时，本就指代高度仿人的德国自动机器，而非指向当时

① 参见陈涛：《论“人机交互”式数字电影表演——以“恐惑谷”效应为中心》，《华中师范大学学报》（人文社会科学版）2024年第4期，第71页。

② 程林：《跨文化视域下的中国式“机器妻”想象：从〈机器人趣话〉到〈非诚勿扰3〉》，《上海文化》2024年第6期，第76页。

③ 参见刘海波：《Robot译名诤议》，《国外自动化》1985年第4期，第43~44页。

还未面世的 robot 概念。^① 如上所述,“机器人”的能指与所指存在错位,并过度凸显拟人化。我们不妨做以下思想实验:考虑如下三类人,一类是日常使用“机器人”概念的中国人,另一类是将“劳伯”挂在嘴边的中国人,第三类是日常使用 robot 概念的外国人,三者特别是前者相较后两者感知同样的仿人自动机器的方式以及对它的初始伦理认知是否会有差异?笔者认为,三者会潜移默化地出现感知差异,这是因为不同能指虽可具有同样的指称义(denotation),但其内涵义(connotation)会有差别——只是“机器人”这一能指到底如何影响国人对仿人机器的感知乃至伦理设定,是否会导致中国人比欧美人更倾向于赋予仿人机器类主体或类人定位,还值得在语言学、哲学、心理学和社会学等的跨学科视域下进一步探析。

以上现象表明,虽然机器人恐惑现象是跨越社会文化的存在,国内学界和媒体也已论及恐惑谷效应,但中国文化整体而言对机器人的恐惑潜能并不敏感。然而,人机共存社会将催生大量的人机交互场景,不同赋能也需要不同仿人程度的机器人,机器人恐惑现象似乎难以完全避免。因此,本文从人文文化和跨文化机器人学角度出发,建议在文化互鉴中进一步促进人机协存关系,即在宏观上构建符合中国社会文化和时代精神的中国式机器人文化及人机协作共存观念,并在微观上通过构建自己的机器人设计理念以更好地规避恐惑现象等负面效应。

第一,在宏观层面,见证中国式机器人文化的形成,在文化互鉴中形塑符合中国传统文化、社会需求和时代精神的人机关系理念。从中国传统文化与当下社会实践和技术文化看,中国的机器人和人机关系认知整体上比“欧美焦虑型文化”更积极乐观,比“日本愿景式文化”更务实,中国社会正在生成的机器人文化,即“折中、务实、积极并倡导人机和存(协存)的‘中国机器人文化’”,^②也即欧美与日本机器人文化之外的又一种方案。

在大致认同拉姆格“同事说”的基础上,本文认为,机器人在中国社会文化中的定位倾向于处在“同事”与“工具人”之间,根据功能的差异有时也可能被视为“朋友”或“劳仆”。中国传统文化倾向于促成对机器人以及对人机协存未来的接纳态度,而非将机器人视为恐惑他者。此外,已有学者在

① 参见程林:《跨文化视域下的中国式“机器妻”想象:从〈机器人趣话〉到〈非诚勿扰3〉》,《上海文化》2024年第6期,第69~70页。

② 程林:《德、日机器人文化探析及中国“第三种机器人文化”构建》,《上海师范大学学报》(哲学社会科学版)2022年第4期,第108页。

尝试用中国传统文化资源特别是儒家和道家思想理解或构建人机关系，双翅目等科幻作家也借助中国儒家理念来设计仿生机器的行为方式。人机协存式机器人文化的动态生成需要避免欧美焦虑型机器人文化，这并非意味着对各类机器人不假思索的拥抱态度或对智能机器之优点的过度宣扬，而是意味着应始终坚持务实、向善、以人为本和为人服务的理念，并合理应对机器代人、认知异化、情感沉溺、隐私泄露等伦理冲击。在此过程中，机器人不需要成为另一种主体，人与机器人应成为以人为本、为人服务的关系共同体和任务协作体。

第二，在微观层面，中国学界无须盲从于恐惑谷效应，而应催生符合中国审美的机器人设计理念。在国内学界习惯不假思索地将恐惑谷效应作为“恐怖谷理论”应用之后，近年来国内学界开始重新审视恐惑谷现象，认为恐惑谷可被跨越或正在被跨越，或建议不应纠结于机器人恐惑现象，而应将注意力放在制造友好型机器人和构建和谐人机关系上。在当今的机器人设计领域中，高度类人化设计并不少见，如“脸面机器人”（face robot）、“表情机器人”（expression robot）、“模拟情绪表达机器人”（simulative emotional expression robot）以及“人头机器人”（humanoid head robot）等，此类设计因为服务于日常人机交互，需要动态情感表达，所以更易令人惊奇与共情，但有时也更易跌入恐惑谷。中国业界亦有对此类机器人的研发。为避免或缓解机器人恐惑现象，本文提出如下建议。

首先，机器人设计应有原创理念，追求符合中国文化审美需求的机器人设计创意，无论这种设计遵循去拟人化原则，还是试探高度拟人化。以仿真机器人形象为例，面部风格写实的索菲亚机器人令人感到诡异不安，而在外观上有时以中国古风特色示人的佳佳机器人则不会带来类似感觉。可见，机器人的高度拟人化并非机器人设计绝对不可触碰的禁区，关键在于如何设计机器人并赋予其何种理念和故事。上述两种机器人仿真效果对比带来的第一条启发就在于，中国机器人设计首先应有自己的创意和理念。

其次，机器人设计理念需要叙事，即讲好机器人故事，做好类似“人设”的“机器人设定”和“人机关系设定”，为人机关系赋予和谐的理念，给予看似“冰冷”的机器人人文温度。对塑造机器人形象而言，机器人设计、人机交互理念以及机器人叙事都很关键。具体到中国机器人科幻叙事，它可回应人类普遍关心的科技世代议题，也可与中国传统文化及当下时代精神有机融合，但无须继续重复国外叙事俗套，在恐惑机器人形象及人机对立叙事方面尤其如此。

· 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 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