中国心理学前沿
2024 年 9 月第 6 卷第 9 期

恐怖谷效应与机器人伦理结合探究

胡凝湘

广西师范大学，桂林

摘　要 | 恐怖谷效应是指当机器人或仿真角色的外表和行为与人类高度相似但又不完全相同时，引起的人类反感和不安的心理现象。许多实验研究都在一定程度上验证了恐怖谷效应。可以从人类进化、认知发展和认知过程的特点等多个角度来解释恐怖谷效应，但是各项研究对于其解释各不相同得出的结论也有所差异。但是随着科技发展至今，人机合作已是大势所趋，我们更加需要了解并探索恐怖谷效应对人机合作的影响。

关键词 | 恐怖谷效应；机器人；伦理

1　引言

恐怖谷效应（Uncanny Valley）是森政弘在1970年提出的概念，描述人们对接近人类但并不完全相似的机器人或虚拟角色产生的不安和反感情绪。该情绪的表现是由于机器人与人类之间的同频气息相似增加而后降低，进而形成“谷”中波纹。该效应在心理学、认知科学、机器人学和人工智能领域引起了广泛关注，并对机器人设计产生了影响。机器人伦理关注机器人设计、制造和使用中的道德原则，涉及权利、责任、隐私、安全等方面，对技术发展和人类社会价值观具有重要意义，恐怖谷效影响人们对机器人的接受度和信任感，对机器人的社会融入和伦理设计具有重要影响，需要在设计中寻求技术进步与伦理责任的平衡。

2　恐怖谷效应的界定

2.1　恐怖谷效应的定义

“恐怖谷”一词源于UVH对“新和”概念的模糊定义，森昌弘（Mori）创造了这个日语新词，用简

作者简介：胡凝湘，广西师范大学，研究方向：学校与教育心理学。
文章引用：胡凝湘. 恐怖谷效应与机器人伦理结合探究 [J]. 中国心理学前沿，2024，6（9）：1594-1602.
https://doi.org/10.35534/pc.0609178

单的术语来描述对各种类人物体的情感体验的积极和消极特征；UVH沿着人类相似的维度定义这些物体（DHL）。虽然与情感体验有关，但这一概念的不良规范导致了许多争论和对其含义的各种渲染，随着对UVH和shinwakan的调查，如愉快、舒适水平、怪异、熟悉（即熟悉感与陌生感）、可爱性、和同理心（例如，Mac Dorman and Ishiguro，2006[1]；Bartneck et al.，2007[2]；Seyama and Nagayama，2007，2009[3]；Green et al.，2008[4]；Mac Dorman et al.，2009[5]，2013；廷威尔、格里姆肖，2009；迪尔等，2012；Burleigh et al.，2013）[6]。森昌弘（Mori，2012）最近将情感体验的积极和消极特征重新定义为亲和力，增加或反映UVH概念的模糊性。[7]

其可能与恐怖相关的研究中或多或少地依赖于特别开发的自评量表来评估情感体验有关（例如，Hanson，2006；[8] Mac Dorman，2006；[5] Green et al.，2008；[4] Mac Dorman et al.，2009a，b；Seyama and Nagayama，2007；[3] Tinwell et al.，2011[9]）。特设自评量表在离奇研究中受到青睐，因为它们价格低廉且易于管理（参见，Ho et al.，2008）。但是，这些测量作为shinwakan或shinwakan概念被认为捕获的情感体验指标的心理测量效度和可靠性尚未得到证明，这使得对迄今为止的研究结果的解释和综合变得困难。

与DHL的概念化和操作化有关（Cheetham et al.，2011）[10]。森昌弘对UVH的说明使用了一个单一的人类样本来代表人类类别（Mori，1970），[7]正如一些经验和理论工作所反映的那样（Tinwell and Grimshaw，2009）。[9]但这种概念化有效的假设在DHL的人类类别中没有物理或心理相似空间的变化。对沿着DHL的感知判别和类别处理的检查（使用连续态来表示人类和非人类类别范例之间物理相似性空间的线性维度）表明，这种假设是不正确的（Cheetham et al.，2011，2014）。[10]使用线性形态连续体来表示DHL并不新鲜（例如，Seyama and Nagayama，2007），[11]但是对连续体的不良控制可能导致了研究结果的不一致。连续形态一直受到各种实验混淆的影响，这些实验混淆可能会对沿DHL的物体产生系统偏见的主观体验（有关批判性讨论，请参见Cheetham and Jancke，2013）。[12]这些混淆包括使用不同并置的连续形态来表示DHL，从而沿着连续形态产生感知不连续（Hanson et al.，2006；[8] Mac Dorman and Ishiguro，2006）[1]对变形噪声（例如，连续体的连续变体之间面部特征对齐的差异）的影响，正如最近对类别模糊对主观体验的影响的研究所表明的那样（Ramey et al.，2005）。[13]至关重要的是，变形可能会改变DHL人类–非人类类别结构的认知表征，并可能本身影响主观体验（Cheetham and Jancke，2013）。[12]

2.2 恐怖谷效应的测量方法

肖苏衡在对恐怖谷效应的研究中采用了三种测量方法：拟人度评分、喜爱度评分和视角采择任务。拟人度评分利用ABOT数据库中的机器人图片，结合被试的主观评价进行量化；喜爱度评分则是被试基于对机器人的喜爱程度给出的0到100的评分；视角采择任务则评估被试在视觉或空间任务中是否倾向于从机器人的视角出发。统计分析显示，尽管喜爱度评分反映出恐怖谷效应，视角采择任务却没有显示出这一效应，表明情感反应和行为模式可能存在差异。[14]董安利在硕士论文《恐怖谷效应对机器人道德判断的影响》中，通过三个实验探讨了恐怖谷效应对道德判断的影响。实验一使用ABOT数据库图片，量化了恐怖谷效应；实验二通过道德困境评估对机器人代理的道德判断；实验三基于双加工理论，探讨

https://doi.org/10.35534/pc.0609178

了恐怖谷效应的认知机制。研究结果表明，在特定条件下，如机器人作为受害者时，恐怖谷效应会显著影响人们的道德判断。[15]

3 恐怖谷效应的成因

3.1 人类的认知机制

在《恐怖谷效应对机器人道德判断的影响》一文中，董安利深入探讨了恐怖谷效应的成因及其对道德判断的影响。她指出，恐怖谷效应是由机器人与人类在外观和行为上的相似性引起的人类反感和不安，这种效应与人类的认知机制有着密切的联系。具体来说，当机器人的拟人化特征与人类接近但又不完全一致时，人类的视觉感知系统会将这些特征与大脑中存储的人类形象进行比较，从而产生认知冲突和不适感。[15]此外，人们的注意力会被机器人的某些非人类特征所吸引，触发与人类特征相关的记忆，进一步加深了对差异的感知。在道德决策方面，当机器人的行为与人类的道德预期不符时，这种不确定性和不可预测性可能会导致人们在决策时产生厌恶感。董安利的研究还发现，恐怖谷效应对道德判断的影响取决于受害者的类型，即当受害者是机器人时，效应更为显著。

3.2 情感反应

丁晓军等人在其研究《人们对人工智能做道德决策的厌恶感之源及解决之道》中探讨了人们对人工智能进行道德决策时的厌恶感。研究指出，这种厌恶感可能源于人们认为AI缺乏完整的心智，尤其是心智感知的能动性和体验性两个维度。[16]能动性包括自我控制、道德判断和记忆等能力，而体验性涉及对情感如痛苦、恐惧和快乐的体验。这两个维度与亚里士多德关于道德行为主体和受动者的理论相联系，分别关联责任和权利。

研究还发现，恐怖谷效应并非普遍适用，例如儿童对机器人的喜好并不完全遵循成人的恐怖谷模式。此外，人们对机器人的好感度与人类相似度之间的关系可能并非简单的三次函数，而是更符合线性或二次关系，表明好感度随相似度增加的变化可能更为复杂。这些发现挑战了恐怖谷效应的普适性，并提示在设计拟人机器人时应考虑更广泛的因素，以避免单一依赖相似度作为接受度的唯一指标。

3.3 社会文化因素

恐怖谷效应可能在不同文化中有不同的表现，因为不同文化对于人类特征的理解和接受度可能有所不同。还可能与文化背景、价值观、宗教信仰，以及对科技的接受程度有关。

在《服务机器人与顾客补偿性消费——社会支持的干预作用》中提到，服务机器人在中国社会背景下的应用可能会受到特定的文化影响，例如，中国消费者对于服务机器人的接受程度可能与西方消费者有所不同（张凯铄，2021）。[17]

肖苏衡在2021年的研究中分析了中西方文化差异对人类与机器人交互的影响。结果表明，中国文化中的“天人合一”思想使得人们更倾向于与机器人合作，而西方的“物我对立”观点可能导致人们在交互时保持距离。实验进一步发现，中国被试更可能从机器人视角考虑问题，且随着机器人拟人度的提

升，中国被试的喜爱度增加，但并未出现恐怖谷效应，表明拟人度提高并未减少从机器人视角进行交互的倾向。这些发现对于定制化人机交互设计具有重要意义。[14]

4 机器人伦理的框架

4.1 机器人伦理的基本原则

恐怖谷效应描述了人们对于外观和行为与人类高度相似但并不完全相同的机器人产生的不安和反感情绪。这种心理现象可能导致人们回避与高度仿真机器人的互动，影响机器人的社会接受度。恐怖谷效应的存在使得在设计机器人时，需要特别考虑如何减轻用户的不适感，同时确保机器人的设计和行为符合伦理标准。

岳璿在其著作《人工智能伦理：规制、理论与实践难题》中提出了人工智能伦理的基本原则，强调了透明性、可解释性、稳健性、可靠性、安全性，以及可追责性的重要性。[18]他指出，这些原则对于确保人工智能技术的负责任发展，保护社会价值和个体权利至关重要。透明性和可解释性要求AI系统决策过程必须清晰，用户和利益相关者能够获得明确解释。稳健性、可靠性和安全性确保AI系统在各种情境下均能稳定运行，保障公共安全。可追责性原则确保在AI系统引发道德争议或损害时，能够明确责任归属并承担相应后果，从而促进人机和谐共存。支振锋和刘佳琨在《伦理先行：生成式人工智能的治理策略》中提到，生成式人工智能的治理需要伦理先行，强调人工智能新技术新应用创新中留下充分空间的重要性。[19]此文中伦理治理的方式可以推动技术社区、产业群体、社会组织、社会公众、治理机构共同进行探索磨合，并凝聚共识。在实践经验充足、立法时机成熟时，再将共识性规范纳入法律框架之中。

4.2 伦理决策模型

机器人伦理的框架关注于设计、制造和使用机器人时应遵循的道德原则和规范，其中伦理决策模型是核心组成部分，它确保机器人的行为不仅透明和可解释，而且明确责任归属，保护用户隐私，并保障操作的安全性。

杜严勇在《机器人伦理研究论纲》中提到，机器人伦理研究内容十分丰富，涉及伦理决策模型的构建。他强调，机器人伦理研究不仅仅是抽象的理论研究，更需要做大量的调查研究，掌握公众的态度与期望。[20]杜严勇认为，为了使机器人伦理研究更有指导意义，可以将技术评估与针对用户的愿景评估结合起来，使机器人技术更好地为社会服务。这表明，在伦理决策模型的设计和应用中，需要考虑公众的道德观念和期望，确保机器人的设计和行为与社会伦理标准相一致。[21]

李晔和黄心语在《人机信任校准的双途径：信任抑制与信任提升》中提出，在人机交互中，通过信任抑制和信任提升来校准个体对机器人的信任水平，以促进有效的人机合作。[22]这与森政弘提出的恐怖谷理论相辅相成，该理论指出，机器人的过度拟人化可能引起人们的不适感。因此，在伦理决策模型中，设计者应考虑通过透明度和解释性AI来建立适当的人机信任，同时平衡机器人的拟人化程度，以规避恐怖谷效应，确保用户的信任和接受度。

https://doi.org/10.35534/pc.0609178

4.3 伦理规范的实施

伦理规范的实施是确保机器人设计和行为遵循既定伦理原则和标准的关键过程。这一过程要求从开发者到制造商，再到用户在内的所有相关方在机器人的整个生命周期内采取积极措施以符合伦理要求。

支振锋和刘佳琨在其论文《伦理先行：生成式人工智能的治理策略》中提出在生成式人工智能的治理中，应优先考虑实施伦理规范，而非急于立法限制，以促进技术的健康发展并保护社会价值和个人权利。他们认为，伦理治理应先于法律规制，确保技术发展的同时，适时将伦理规范整合入法律体系。[19] 龙龙在《敏捷治理理念下人工智能软法之治的问题与对策》中提到人工智能领域采用敏捷治理理念，通过体系化治理、多元协商、自治能力提升、全过程规范实施和软硬法融合等措施，以确保AI技术的伦理发展和社会责任。[23] 江必新和刘倬全在其著作《论数字伦理体系的建构》中提出，在构建数字伦理体系时，需要平衡规范与发展、问题导向与系统思维、操作性与分级分阶段、宣传与机制约束、国家主权与全球数字命运共同体等五大关系，并确保各方利益主体的实质性参与，建立全面的制度体系。[24]

5 恐怖谷效应与机器人伦理的相关研究

5.1 恐怖谷效应对机器人设计的影响

许多研究强调了拟人化设计在机器人外观和行为设计中的重要性（Mori，1970；[7] Bartneck et al.，2007[2]）。拟人化是机器人设计中提升用户信任和亲近感的关键策略，但需适度以免触发恐怖谷效应。不同文化对机器人拟人化反应各异，某些文化可能更易接受高拟人度机器人。设计时，考虑目标用户的文化背景至关重要，以找到拟人化的恰当平衡点，确保机器人既有亲和力又不引起不适 。

为了规避恐怖谷效应，机器人设计师可以采用特定的情感设计策略，如通过调整机器人的表情、动作和语言来管理用户的期望和感知（Shin et al.，2019）。[25] 机器人设计不仅要考虑外观，还要考虑其功能。有时，强调机器人的工具属性而非人类相似性，可以减少恐怖谷效应的影响。通过优化人机交互体验，例如提高机器人的交互自然性和适应性，可以减轻恐怖谷效应，增强用户的正面情感（Bartneck et al.，2008）。[2] 随着机器人越来越深入人类生活，其设计还需要考虑道德和伦理问题，如隐私保护、责任归属等，这些都可能间接影响用户对机器人的情感反应。

5.2 机器人伦理对恐怖谷效应研究的指导

本文旨在深入探讨恐怖谷效应的基本概念、心理机制及其在现代机器人设计中的应用，同时考察不同文化和社会背景下的恐怖谷效应表现，并探索通过设计和技术创新减轻或避免恐怖谷效应的策略，以提高人们对机器人的接受度和信任感。此外，研究还将分析机器人伦理在应对恐怖谷效应时的角色，以及如何在技术发展与伦理责任之间找到平衡点，从而为机器人设计提供理论指导，为机器人伦理的实践提供参考，并为人机交互和机器人技术的未来发展趋势提供科学依据。

随着机器人和虚拟角色的人类相似性增加，它们是否应获得更高的道德地位成了伦理学讨论的焦点。麦克多曼和石黑（MacDorman and Ishiguro，2006）提出，人类对机器人的同情心可能会随着其人类

相似性的增加而增加，但当相似性达到一定程度时，可能会引发恐怖谷效应，导致同情心下降。在设计机器人时，设计师需要考虑恐怖谷效应对用户情感体验的影响。[1] 例如，伯利等人（Burleigh et al.，2013）的研究表明，设计者可能需要避免创造处于恐怖谷中的机器人，以减少用户的不适感。[6]

5.3 跨学科视角下的恐怖谷效应与机器人伦理

5.3.1 心理学视角

恐怖谷效应是森政弘提出的心理学现象，指机器人或虚拟角色与人类外观和动作相似度增加时，人们的情感反应先正面上升，达到一定峰值后急剧下降至谷底（产生不安和反感），随后随着相似度继续提高，情感反应再次上升。这一效应揭示了人类对类人物体的复杂情感变化。波利亚科夫等人（Poliakoff et al.，2013）发现恐怖谷效应在假手上也会出现，相对机械手、真手和与真手相似度高的假手相比，与真手相似度低的假手会让被试产生更怪异的感觉。还有日本学者在婴儿身上也发现了恐怖谷效应，研究结果发现，当婴儿面对母亲的面部图片、陌生人的陌生图片、陌生人与母亲二者合成的图片时，婴儿更愿意面对母亲的脸，其次是陌生人的脸，最后才是二者合成的脸。[18] 有学者发现，当被试在虚拟社交网络上面对介于真实头像与卡通头像之间的虚拟头像时，被试与虚拟头像之间的互动会给被试带来更为怪异的感觉。

5.3.2 社会学视角

在恐怖谷效应与机器人伦理的研究中，社会学视角关注于人类如何感知和评价机器人的道德行为。董安利在其硕士论文中探讨了恐怖谷效应对机器人道德判断的影响，发现人们对于外观类似人类的机器人的道德判断可能会受到恐怖谷效应的影响。这表明，机器人的设计和外观可能会影响人们对其道德行为的评价。

恐怖谷效应的研究已经跨越西方文化界限，开始关注不同文化背景下的表现形式。肖苏衡的研究特别强调中西方文化差异对机器人拟人度、喜爱度，以及视角采择的影响，揭示文化背景对恐怖谷效应可能产生的调节作用，导致不同文化中人们对机器人的接受程度和道德评价出现差异。[21]

5.3.3 哲学视角

机器人伦理是哲学领域关注的一个新兴议题，它涉及机器人的道德责任、权利，以及其在社会中的角色。在文件中，张玉洁的研究《论人工智能时代的机器人权利及其风险规制》深入讨论了机器人的权利地位，提出了机器人权利的法律拟制性、利他性和功能性等权利属性。讨论为机器人伦理提供了哲学基础，探讨了机器人在道德和法律体系中的地位。[27] 恐怖谷效应与机器人伦理的研究试图理解人类对机器人外观和行为的道德判断。董安利的实验发现，恐怖谷效应可能影响人们对机器人行为的道德评价，表明道德感知受外观等因素影响。[15]

6 案例研究

6.1 恐怖谷效应在机器人设计中的应用案例

契瑟姆等人（Cheetham et al.，2014）探讨了感知歧视难度（perceptual discrimination difficulty，PD）

https://doi.org/10.35534/pc.0609178

与恐怖谷之间的关系。研究发现，与恐怖谷假设相反，当PD难度增加时，并不必然伴随着负面情绪的增强。这一发现对于机器人设计具有重要意义，因为它表明设计师可以通过调整机器人的人类相似性来避免恐怖谷效应。[28]

契瑟姆等人（Cheetham et al.，2015）使用电生理和自我报告方法来测试恐怖谷假设。结果表明，沿着人类相似性维度的类别模糊性并不特别与增强的负面情绪体验相关联。这一发现进一步质疑了恐怖谷假设，并为机器人设计提供了新的思路。[29]

费里等人（Ferrey et al.，2015）提出了一个新的假设，即恐怖谷效应可能与刺激相关的抑制过程有关。研究通过实验测试了这一假设，发现即使在非人类刺激中也出现了恐怖谷效应，表明恐怖谷效应可能与认知抑制和情感贬值有关，而不仅仅是与人类相似性有关。[30]

6.2 机器人伦理在实际案例中的挑战与解决方案

伊丽莎程序是由麻省理工学院（MIT）的计算机科学家约瑟夫·魏泽鲍姆在1964至1966年期间开发的，它通过预设脚本理解基础自然语言，并模拟人类对话，使用户产生信任感。程序名称源自萧伯纳的剧作《卖花女》，暗喻了通过语言的力量实现转变。魏泽鲍姆的创造象征着人工智能的拟人化设计，旨在激发用户的积极反应。这种设计策略，即“伊丽莎效应”，提升了AI的互动体验，减少了人们对机器进行道德判断的抵触情绪。现代AI助手，如苹果的Siri、亚马逊的Alexa和谷歌助手，运用自然语言处理（NLP）和机器学习技术，以类似人类的交流方式理解和回应用户，进一步强化了用户的接受度和互动意愿。这些技术的发展不仅展示了AI的交流能力，也使虚拟助手成为人们日常生活的一部分，证明了伊丽莎效应在当代人工智能设计中的持续影响。

在机器人伦理的实践中，降低恐怖谷效应的策略涉及多方面的考虑。设计时应将人工智能限制在辅助角色，避免让其成为最终决策者，以减少用户的不信任感。同时，通过提升AI的专长性和体验性，但避免过度拟人化，可以增强用户的信任和满意度。确保透明度，提供知情同意，并优化机器人的外观设计，有助于用户理解和接受机器人的功能性和限制。此外，通过个性化设置、逐步适应和增强互动性，可以提升用户对机器人的接受度和舒适度。情感设计和教育也起到了关键作用，通过正面情感引导和提升用户理解，进一步减少恐怖谷效应。综合这些策略，设计者和开发者可以创建出既符合伦理标准又易于用户接受的人工智能和机器人产品。

7 结论

本文探讨了恐怖谷效应与机器人伦理的交集，强调了在人工智能发展中考虑设计和伦理决策的重要性。恐怖谷效应指的是人们对高度仿真机器人的不适感，这一现象在机器人设计时必须关注，以免负面影响用户体验。研究通过调整机器人的拟人化特征来减轻恐怖谷效应，并利用机器人伦理框架确保设计符合道德标准。跨学科研究揭示了心理学、社会学和哲学在全面理解和应对恐怖谷效应中的作用。案例研究展示了如何在实际应用中通过设计和管理策略规避恐怖谷效应，并在机器人伦理指导下解决实际问题。文章总结指出，设计优化和伦理治理在人工智能领域至关重要，未来研究需继续探索，以促进技术在符合伦理原则的基础上健康发展。

参考文献

[1] Mac Dorman K F，Ishiguro H. The uncanny advantage of using androids in cognitive science research [J]. Interact Stud，2006（7）：297–337.
[2] Bartneck C，Kanda T，Ishiguro H，et al. Is the uncanny valley an uncanny cliff? [M] // in Proceedings of the 16th IEEE，RO–MAN（Jeju），2007：368–373.
[3] Seyama J I，Nagayama R S. The uncanny valley：effect of realism on the impression of artificial human faces [J]. Presence Teleoperat Virtual Environ，2007（16）：337–351.
[4] Green R D，Mac Dorman K F，Ho C–C，et al. Sensitivity to the proportions of faces that vary in human likeness [J]. Comput. Hum. Behav，2008（24）：2456–2474.
[5] MacDorman K，Green R，Ho C C，et al. Too real for comfort? Uncanny responses to computer generated faces [J]. Comput. Human Behav，2009（25）：695–710.
[6] Burleigh T J，Schoenherr J R，Lacroix G L. Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces [J]. Comput. Hum. Behav，2013（29）：759–771.
[7] Mori M. Bukimi no tani [The uncanny valley] [J]. Energy，1970，7（4）：33–35.
[8] Hanson D. Exploring the aesthetic range for humanoid robots [M] // in Proceedings of the ICCS/CogSci–2006 Long Symposium：Toward Social Mechanisms of Android Science（Vancouver），2006：39–42.
[9] Tinwell A，Grimshaw M，Nabi D A，et al. Facial expression ofemotion and perception of the uncanny valley in virtual characters [J]. Comput. Hum. Behav，2011（27）：741–749.
[10] Cheetham M，Suter P，Jancke L. The human likeness dimension of the “uncanny valley hypothesis”：behavioral and functional MRI findings [J]. Front. Hum. Neurosci，2011（5）：126.
[11] Seyama J，Nagayama R S. The uncanny valley：effect of realism on the impression of artificial human faces [J]. Presence（Cambridge，MA），2007（16）：337–351.
[12] Cheetham M，Jancke L. Perceptual and Category Processing of the Uncanny Valley Hypothesis’ Dimension of Human Likeness：Some Methodological Issues [J]. Journal of Visualized Experiments：JoVE，2013.
[13] Ramey C H. The uncanny valley of similarities concerning abortion，baldness，heaps of sand，and humanlike robots [M] //in Paper Presented at the Proceedings of Views of the Uncanny Valley Workshop：IEEE–RAS International Conference on Humanoid Robots，（Tsukuba）2005.
[14] 肖苏衡. 恐怖谷与非恐怖谷：机器人拟人度对喜爱度及视角采择的影响 [D]. 南京大学，2021.
[15] 董安利. 恐怖谷效应对机器人道德判断的影响 [D]. 西北大学，2022.
[16] 丁晓军，喻丰，许丽颖. 人们对人工智能做道德决策的厌恶感之源及解决之道 [J]. 自然辩证法通讯，2020，42（12）：80–86.
[17] 张凯铄. 服务机器人与顾客补偿性消费 [D]. 华中科技大学，2021.
[18] 岳璠. 人工智能伦理：规制、理论与实践难题——学术史梳理及其问题域考察 [J]. 东南大学学报（哲学社会科学版），2024，26（3）：32–41，150.
[19] 支振锋，刘佳琨. 伦理先行：生成式人工智能的治理策略 [J]. 云南社会科学，2024（4）：60–71.
[20] 杜严勇. 机器人伦理研究论纲 [J]. 科学技术哲学研究，2018，35（4）：106–111.
[21] 杜严勇. 恐怖谷效应探析 [J]. 云南社会科学，2020（3）：37–44，187.
[22] 黄心语，李晔. 人机信任校准的双途径：信任抑制与信任提升 [J]. 心理科学进展，2024，32

https://doi.org/10.35534/pc.0609178

（3）：527–542.
[23] 龙龙. 敏捷治理理念下人工智能软法之治的问题与对策 [J]. 江汉论坛，2024（5）：128–134.
[24] 江必新，刘倬全. 论数字伦理体系的建构 [J]. 中南大学学报（社会科学版），2024，30（1）：38–49.
[25] Shin L M，Liberzon I. The neurocircuitry of fear，stress，and anxiety disorders [J]. Neuropsychopharmacology，2009（35）：169–191.
[26] Poliakoff E，Beach N，Best R，et al. Can looking at a hand make your skin crawl? Peering into the uncanny valley for hands [J]. Perception，2013（42）：998–1000.
[27] 张玉洁. 论人工智能时代的机器人权利及其风险规制 [J]. 东方法学，2017（6）：56–66.
[28] Cheetham M，Pavlovic I，Jordan N，et al. Category processing and the human likeness dimension of the Uncanny Valley Hypothesis：eye–tracking data [J]. Front Psychol，2013（4）：108.
[29] Cheetham M，Wu L，Pauli P，et al. Arousal，valence，and the uncannyvalley：psychophy siological and self–report findings [J]. Front. Psychol，2015（6）：981.
[30] Ferrey A E，Burleigh T J，Fenske M J. Stimulus–category competition，inhibition，and affective devaluation：a novel account of the uncanny valley [J]. Front Psychol，2015（6）：249.

Exploring the Valley of Terror Effect in Conjunction with Robotics Ethics

Hu Ningxiang

Guangxi Normal University, Guilin

Abstract: The Valley of Terror effect is the psychological phenomenon of human revulsion and unease caused when robots or simulated characters look and behave highly similar to humans but not exactly the same. Many experimental studies have verified the Valley of Terror effect to some extent. The Valley of Terror effect can be explained from a variety of perspectives, such as human evolution, cognitive development, and the characteristics of the cognitive process, etc., but the conclusions of the various studies on the interpretation of the Valley of Terror effect are also different. However, with the development of science and technology, human-machine cooperation has become a major trend, and we need to understand and explore the impact of the Valley of Terror effect on human-machine cooperation.

Key words: Uncanny valley; Robots; Ethics