

人工智能类人特征引发的伦理风险与应对路径

许胜晴

江苏海洋大学文法学院, 江苏 连云港

收稿日期: 2025年2月17日; 录用日期: 2025年3月17日; 发布日期: 2025年3月25日

摘要

类人特征是人工智能发展程度的重要判断标准。随着算法和芯片等硬件基础的不断发展, 人工智能在运行逻辑、外在形态和行为方式等方面逐渐接近甚至超越人类的状态。人工智能的类人特征在为生产生活带来便利、效率的同时也带来了诸多伦理风险。人工智能类人特征引发的伦理风险主要包括智能生成引发的诚信风险、社交替代造成的社会疏离、智能决策带来的主体冲突等方面。为预防和减轻相应伦理风险及其损害, 应着重强化人工智能标识制度与生成物使用规范、完善人工智能时代的社交保障机制、完善决策冲突的处置机制等。

关键词

人工智能, 类人特征, 伦理风险, 社会关系

Ethical Risks Caused by Humanoid Characteristics of Artificial Intelligence and the Coping Approaches

Shengqing Xu

School of Humanities and Law, Jiangsu Ocean University, Lianyungang Jiangsu

Received: Feb. 17th, 2025; accepted: Mar. 17th, 2025; published: Mar. 25th, 2025

Abstract

Humanoid characteristics are important criteria for judging the development of artificial intelligence. With the continuous development of software and hardware foundations such as algorithms and chips, artificial intelligence is gradually approaching or even surpassing the human state in terms of operation logic, external form and behavior. The human-like characteristics of artificial intelligence bring convenience and efficiency to production and life, but also bring many ethical risks.

The ethical risks caused by artificial intelligence humanoid characteristics mainly include the integrity risk caused by intelligent generation, the social alienation caused by social substitution, and the subject conflict caused by intelligent decision-making. In order to prevent and mitigate the corresponding ethical risks and damages, emphasis should be placed on strengthening the artificial intelligence labeling system and the use norms of products, improving the social security mechanism in the era of artificial intelligence, and improving the resolution mechanism of decision conflicts.

Keywords

Artificial Intelligence, Humanoid Characteristics, Ethical Risks, Social Relations

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着数字信息技术的进步，人工智能已经过多次迭代升级快速向接近于人类内在思维和外在活动的阶段迈进。人工智能的类人化甚至超人化是人工智能技术进步的标志。计算机领域的图灵测试也是将人工智能与人类判断的接近程度作为评价计算机智能化水平的标准。近年来，人工智能在图像识别、自然语言理解、强化学习等领域因算法和芯片等软硬件基础的进步而实现质的飞跃。人工智能类人特征的不断强化在促进人工智能服务于社会生产生活和科学研究需求的同时也带来了诸多伦理风险，影响了人机互动的深度和广度，也进一步影响了人类自身的认知、情感和发展以及社会关系形态。

2. 人工智能及其类人特征

人工智能是技术发展的重要领域。就理论来看，人工智能具有与人类类似的功能特征，能够替代或者辅助人类从事实践活动，用于解决诸多复杂问题和各种人类需求。但是，人工智能的类人特征也给人类社会带来严重挑战。孙伟平认为“智能机器人的快速发展模糊了人机界限，对人的本质、人的主体地位等形成强烈的冲击，令‘人是什么’和人机关系凸显为挑战哲学常识的时代难题”[1]。人工智能在自然科学、社会科学等研究领域以及在交通、医疗、金融等应用场景均展现出强大的技术生产力。脑认知科学与人工智能的融合促进，进一步为人工智能的发展提供理论基础和设计理念[2]。

在多学科和各类技术综合演进的背景下，人工智能的类人特征进一步强化。人工智能的类人特征主要体现在人工智能在外在形态、内在运行机制、刺激反射活动、行为操作表现、问题解决逻辑等方面逐渐向人类思维和行为模式接近，甚至在计算功能和信息检索、加工和提取方面体现出更加强大的能力。在形态方面，人形机器人与人的外形和仪态可以做到极其接近并具有相应的人机交流功能[3]。在解决复杂运算和信息提取方面，人工智能依靠大数据和相应算法支撑更是可以极大缩减人力和时间消耗。随着各类人工智能进入生产生活的方方面面并逐渐普及，人类与人工智能共存情况下如何处理人机关系以及人际关系如何影响人际关系成为亟待反思和探究的问题。

3. 伦理风险的表现形式

人工智能技术的进步使得判断人工智能先进程度的指标不断突破，但也给人类社会带来新的问题。田海平认为人工智能“在‘以人为镜’或‘以人类智能为锚定目标’时，产生不断切近乃至超越人类智能的演化方向或行动目标，从而对‘人何以为人’的决定性属性带来挑战”[4]。人工智能的类人特征给人

类对人机关系的传统认知带来了颠覆性影响，甚至冲击甚至动摇人类的主体地位，并进而带来一些列严重的社会影响[5]。当前在法学领域探究的人工智能的权利主体、责任主体等问题就反映了这种新的认知已经进入到强制性社会规范的视角。除了人的主体性地位之外，人工智能活动所产生的数据信息也会带来的新的社会问题。例如有学者以生成式人工智能(AIGC)为例探究了人机交流产生的社会问题，提出话语失范与污染带来的攻击性语言交流、机器沉迷引发的社会越轨行为、隐私侵犯行为与信息茧房效应的加剧等问题[6]。人工智能介入社会关系将带来进一步的伦理风险。当前，人工智能的伦理风险主要体现在智能生成引发的诚信风险、社交替代造成的社会疏离以及智能决策带来的主体冲突等方面。

3.1. 智能生成引发的诚信风险

诚信是人际交往的重要标准也是社会主义核心价值观。在网络时代，在线交流和获取信息已经成为重要的社会交往和信息获取方式。人工智能在生成图像、视频和声音方面已经达到相当高度。实践中出现了大量因人工智能介入而带来的欺诈问题。一方面，人工智能为诈骗者提供了更加多样和强大的工具，包括网络钓鱼攻击、深度伪造、嵌入恶意代码等，极大增加了辨别真伪和调查求证的难度。受害者基于伪造的视频形象、声音和签名等信息落入陷阱并遭受人身和财产损失。基于深度伪造技术等人工智能方法，不法分子传播虚假信息，侵犯相关主体权益，给社会稳定带来威胁。另一方面，人工智能技术也为各类社会活动和市场交易中弄虚作假提供了可利用的工具。例如：以人工智能作品混淆原创作品或者运用人工智能侵害他人知识产权引发著作权诚信风险；经营者在商业宣传中利用深度伪造技术进行虚假宣传或者诋毁其他主体商誉损害等[7]。有学者指出，虚拟生物识别信息的再造违背人的意愿，改变了个体决策隐私、意向隐私和信息隐私的权利，从而在行为和思想表达方面带来混淆[8]。在视频网站中，依靠深度伪造技术实施的人脸或身体拼贴、再造景观等视频则往往引发信息盗取、网络勒索和诈骗等危害公共安全的行为[9]。

3.2. 社交替代造成的社会疏离

社会关系和社会交往关系人的本质特征和生存生活状态。人工智能类人化的高度发展使得人际社交的部分内容逐渐被人机交互替代，引发了人工智能对社会关系的社交替代。有学者提出，人类与人工智能的关系随着人工智能与人类相似程度的变化而呈现U型变化，即好感度从高到低再到升高，并在二者高度类似时产生移情现象[10]。这种移情现象是社交替代的重要促成因素。社会交往带来的精神生活体验被人工智能设计的交互体验替代。这一现象既有其积极一面，但也蕴含深刻的风险。对于本身缺乏社会交往机会或因其他客观原因无法与他人进行正常交往的情况下，这种社交替代能够起到补充作用，有助于相关主体获得替代性的社交体验。但是，过度沉溺于虚拟的空间容易进一步带来孤独感和自我封闭倾向，进而增加社会疏离程度，淡化人际交往的功能，甚至会增加与伴侣或其他亲属的冲突等问题[11]。在人工智能时代，社会交往借助虚拟技术进行数据交互即可完成。因人工智能而兴起的虚拟偶像，反映了现代工业社会中大众对商品的偶像崇拜，并带来幻化的社会交往，即虚拟偶像与其粉丝群体之间的社会交往形态因技术的虚构而幻化[12]。

3.3. 智能决策带来的主体冲突

智能决策是人工智能和大数据技术发展的重要领域。有学者指出：智能决策综合了人工智能、决策支持系统、知识管理系统、专家系统以及管理信息系统等多元系统中的知识、数据、业务流程等内容，通过模型库、方法库、专家库等进行分析、推理，从而帮助解决复杂决策问题的系统[13]。随着社会的发展变化，各类决策活动在频度、广度及复杂性和不确定性日益增加，传统的数据分析方法以及基于人工经验的决策逐渐难以满足实践需求，而各类人工智能技术在医疗、金融、交通甚至司法等领域的运用更

是直接带来智能决策与人工经验决策的冲突[14]。在智能决策过程中，算法本身的缺陷以及算法黑箱和算法偏见等问题会影响决策的合理性[15]。以人工智能参与公共决策为例，人工智能虽有助于提高决策效率，但决策中权责体系失衡。人工智能在分担人类决策任务的同时，逐渐取代了人的主体性地位，而技术却无法像人一样承担决策的相关责任[16]。在司法领域，人工智能虽在一定程度上能够减少同案不同判现象或提升司法效率，但此类裁量多是基于文本和数据，其对个案的特殊情况的判断会因样本局限性和因果关系的复杂性而难以充分理解和裁断。正如有学者指出，司法人工智能系统的裁量是以大数据要素之间的相关关系而不是因果关系作为决策基础逻辑的，其仅仅是以计量的方式将事件和法律后果相联系[17]。在与人为裁判出现冲突时，应该如何协调二者关系以及如何确定决策责任成为人工智能介入决策或裁判时需慎重解决的问题。

4. 主要应对路径

4.1. 强化人工智能标识制度与生成物使用规范

人工智能标识制度是实现透明化治理的重要手段。对人工智能生成物进行标识是保障受众知情权的必要选择。当前国内外均把人工智能标识制度作为人工智能治理的有机组成部分。例如欧盟的《人工智能法案》、美国加州的《数字内容溯源标识法案》等[18]。在平台参与的情况下，该制度还要求强化相关平台在人工智能内容标识方面的责任。我国《互联网信息服务深度合成管理规定》要求深度合成服务提供者需对生成内容进行标识。除服务提供者外，《人工智能生成合成内容标识办法(征求意见稿)》第10条规定：“用户向提供网络信息内容传播平台服务的服务提供者上传生成合成内容时，应当主动声明并使用平台提供的标识功能进行标识”。但是，在相应的法律责任机制方面仍需进一步强化，明确未主动声明或标识需承担的行政责任。另外，对于人工智能生成物，为防止其滥用还需制定更为具体的使用规范，明确其可适当适用和不应当使用的情形并设置相应法律责任。例如，在以原创性为核心要求的创作和学术领域应适当限制对人工智能生成物的使用，并要求相关主体作出诚信声明。

4.2. 完善人工智能时代的社交保障机制

交往是人类特有的活动和存在方式，是人与人之间发生社会关系的一种中介。交往不仅确证人的本质，而且决定人的发展，而且人的全面发展取决于交往的普遍发展[19]。“传统社会交往在构建社会结构、发展生产力和促进人的发展方面发挥着独特的作用”[20]。人工智能对社交的替代一定程度上将对相关主体的社会关系产生影响。为趋利避害，应完善人工智能时代的社交保障机制。相应部门应进一步加强精神文明建设，大力开展群众广泛参与的文化娱乐、旅游体育等互动。在城乡规划中为社会交往提供更多的空间和设施支持，并基于人工智能带来的效率提高为社会交往提供充分的节假日保障。同时，还需开展相关公益性心理健康服务，减轻人工智能时代的网络沉迷、自我封闭和盲目追求虚拟体验的现象。

4.3. 完善决策冲突的处置机制

在智能决策与决策人员的认知存在冲突时应当建立决策冲突处理机制。应首先认识到智能决策也存在局限性。基于人工智能的计算或信息提取并不能取代专业人员基于理论和实践知识作出的判断。人工智能应主要定位于辅助决策的角色。一旦产生决策冲突应由对决策负有责任和能够实际履行责任的决策者拥有最终决策的话语权并承担相应责任。需要注意的是，决策者应对人工智能决策的不合理之处进行辩驳以充分考虑基于大数据和算法所作出的计算或决策结果。在公共领域，对于存在重大冲突的情况，可考虑通过听证会或者冲突处置委员会等形式作出决定。在智能驾驶等特定应用领域，应在设计阶段充分考虑决策冲突，并就其算法选择进行充分论证且公开其冲突选择，保障用户和相关公众的知情权。

5. 结论

人工智能的类人特征具有促进社会发展和引发伦理风险的双重属性。在人工智能治理中应采取有效治理框架和措施预防和控制相关伦理风险。因与人类思维和行动具有相似性,人工智能给社会诚信、社会交往和各类决策行为带来了重大的挑战。人工智能社会治理应秉持以人为本的理念,在促进技术发展的同时关照人自身的发展及其相关的社会角色、社会责任等社会性因素。在人工智能技术高速发展的当下,尤应关注其所带来的伦理风险,从而实现人类和社会发展的趋利避害,也为人工智能本身的发展提供更为合理的设计理念和方向。

基金项目

本研究为2024年江苏省基础研究专项资金(软科学研究)项目“江苏人工智能安全及伦理风险治理框架研究”(BR2024027)的阶段性成果。

参考文献

- [1] 孙伟平. 人工智能与人的“新异化”[J]. 中国社会科学, 2020(12): 119-137+202-203.
- [2] 薛贵, 刘德建. 脑认知科学和人工智能驱动的未来教育变革[J]. 人民论坛·学术前沿, 2024(17): 24-40.
- [3] 姜华. 类人想象与杂合属性: 论人工智能的演进及社会影响[J]. 华东师范大学学报(哲学社会科学版), 2024, 56(2): 20-30.
- [4] 田海平. 人与类人之间的道德前景——从意向性看人工智能的伦理之“锚”[J]. 华东师范大学学报(哲学社会科学版), 2025, 57(1): 8-21.
- [5] 孙伟平. 智能机器人的主体地位、权利、责任和义务问题[J]. 华东师范大学学报(哲学社会科学版), 2024, 56(1): 1-8.
- [6] 董浩. 生成式人工智能时代人机对话的传播伦理风险及其应对[J]. 阅江学刊, 2024, 16(1): 104-114+174.
- [7] 龙俊, 王天禹. 人工智能深度伪造技术的法律风险防控[J]. 行政管理改革, 2024, 3(3): 69-79.
- [8] 张灿. 人工智能深度伪造技术的伦理风险与虚无困境[J]. 山东科技大学学报(社会科学版), 2023, 25(2): 11-19.
- [9] 王璇, 宋春龙. 基于深度伪造技术的“二创”视频伦理风险及规制治理研究[J]. 西南民族大学学报(人文社会科学版), 2024, 45(5): 165-173.
- [10] 沈佳骏. 基于恐怖谷理论的机器人拟人化认知评估研究[D]: [硕士学位论文]. 天津: 河北工业大学, 2023.
- [11] 吴海江, 武亚运. 人工智能与人的发展——基于马克思人学理论的考察[J]. 学术界, 2019(3): 75-81.
- [12] 付茜茜. 偶像符号的编码: 人工智能虚拟偶像消费文化研究[J]. 学习与实践, 2021(2): 132-140.
- [13] 王继业. 智能决策[J]. 电力信息化, 2011, 9(7): 5.
- [14] 于洪, 何德牛, 王国胤, 等. 大数据智能决策[J]. 自动化学报, 2020, 46(5): 878-896.
- [15] 孙继文. 人工智能算法决策与民法公平原则的冲突与协调[J]. 河南司法警官职业学院学报, 2024, 22(2): 72-76.
- [16] 王锋, 刘玮. 人工智能参与决策过程的挑战与图景[J]. 求实, 2023(3): 18-29.
- [17] 王文玉. 司法算法决策和裁量正义的冲突与融合[J]. 江汉学术, 2024, 43(6): 16-27.
- [18] 马永强. 人工智能生成合成内容的标识义务及制度优化[N]. 民主与法制时报, 2024-10-16(003).
- [19] 安世遨. 构建交往机制促进人的发展[N]. 光明日报, 2018-01-30(013).
- [20] 李成保, 李鹏飞. 新时期社会交往评价问题研究[J]. 天津师范大学学报(社会科学版), 2016(5): 72-76.