

人工智能的伦理缺失：情感与道德的空白

徐楚楚

湖北大学哲学学院，湖北 武汉

收稿日期：2023年12月2日；录用日期：2024年1月2日；发布日期：2024年1月11日

摘要

自人工智能问世以来，国内外学者就其伦理问题进行了广泛而深入的讨论。然而，对于人工智能自身的伦理问题，探讨相对较少。本文主张，尽管当前人工智能技术取得了显著进步，但其本质仍然是工具，而非具有道德自主性的实体。因此，人工智能引发的伦理问题主要源于人类的决策和行为，而非人工智能自身。本文通过探讨人工智能在情感和道德层面的缺失，阐明其与人类在伦理决策上的根本区别，旨在为理解和应对人工智能时代的伦理挑战提供新的思考和见解。

关键词

人工智能伦理，人工智能，伦理学

The Ethical Void in Artificial Intelligence: The Absence of Emotion and Morality

Chuchu Xu

School of Philosophy, Hubei University, Wuhan Hubei

Received: Dec. 2nd, 2023; accepted: Jan. 2nd, 2024; published: Jan. 11th, 2024

Abstract

Since the advent of artificial intelligence, there has been extensive and in-depth discussion among scholars globally regarding its ethical implications. However, discussions specifically about the ethical issues intrinsic to artificial intelligence itself have been relatively scarce. This paper contends that, despite the significant advancements in current artificial intelligence technology, its fundamental nature remains that of a tool, not an entity with moral autonomy. Consequently, the ethical issues associated with artificial intelligence mainly originate from human decisions and behaviors, not from the artificial intelligence itself. By examining the lack of emotional and moral dimensions in artificial intelligence, this paper highlights the fundamental differences in ethical

decision-making between AI and humans, aiming to provide new insights and perspectives for understanding and addressing the ethical challenges in the era of artificial intelligence.

Keywords

Artificial Intelligence Ethics, Artificial Intelligence, Ethics

Copyright © 2024 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

人工智能(Artificial Intelligence, 简称 AI), 指的是由计算机系统或机械设备执行的类似于人类智慧的行为。在这个充斥着数字化的时代, 人工智能不仅已经深入渗透到我们的日常生活中, 而且在从基础的数据处理到复杂的决策制定等多个领域展现了其令人瞩目的能力。然而, 伴随着其应用范围的不断扩大, 人工智能在伦理、情感及道德方面的局限性逐渐成为显著的问题。本文将深入探讨人工智能是否具备情感 and 道德判断能力, 以及这一点是否在伦理层面引发了新的问题。

尽管人工智能在某些方面似乎模仿了人类的行为模式和决策过程, 但我们必须认识到, 它们实际上缺乏真正的情感体验和道德认知。这一事实在本文的后续讨论中至关重要。本文的目的是从伦理学的视角出发, 通过论证人工智能在情感和道德层面的不足, 来阐明人工智能与人类在伦理决策上的本质区别。同时, 本文也将分析这种缺失如何影响人工智能的应用和其未来发展的趋势, 旨在为理解和应对人工智能时代的伦理挑战提供新的思考和见解。

2. 人工智能与人类伦理的区别

2.1. 人工智能的基本原理和能力

人工智能的核心是基于数据驱动算法, 这些算法能够通过模式识别和学习来执行各种任务, 包括但不限于语言理解、图像识别、预测分析和决策支持。然而, 尽管人工智能在处理逻辑和数据密集型任务方面表现出色, 它仍然缺乏对其处理内容的真正理解。人工智能缺乏人类的情感和主观经验, 这限制了其在某些领域的应用, 特别是那些涉及复杂人际互动和情感判断的领域。

2.2. 人类伦理、情感和道德的特点

人类的道德和伦理决策包含了多个复杂且相互关联的层面, 这些层面超越了纯粹的逻辑和理性分析, 深入到情感、同情心和道德直觉的领域。这些特点包括: 情感共鸣, 人类在做出道德判断时, 通常会考虑到他人的感受和福祉。这种共鸣不仅是对他人情感状态的认知理解, 还包括对他人处境的情感投入和同理心。例如, 在考虑是否援助他人时, 人们可能会基于对对方困境的情感共鸣来做出决策; 道德直觉, 这些直觉往往是快速且自动的, 源于长期的社会化过程和个人经验。道德直觉使人们能够迅速识别某种行为是否合适或不当, 甚至在深入分析具体情节的情况下也是如此; 以及情境灵活性, 人类的道德和伦理判断具有高度的情境依赖性。同一道德原则在不同的情境下可能会导致不同的判断。例如, 诚实大多数情况下被视为美德, 但在某些特定情境下, 为了保护他人的感受或安全, 说谎可能被认为是合理的。而这些是目前的人工智能所无法实现的。

2.3. 人工智能与人类伦理决策的差异

人工智能在做出决策时，无法像人类那样考虑道德和伦理层面的因素，这导致了在处理复杂伦理问题时的局限性。首先，人工智能通常缺乏真正的情感理解和共鸣能力，让一串代码与人类共情这件事听起来就显得匪夷所思。人工智能可以被编程用来识别和响应人类情感表达的模式，但这种识别是基于算法和数据，并非真正的情感体验。其次，人工智能虽然可以通过编程设置道德规则和指南，但人工智能无法像人类那样拥有基于直觉的道德感知。人工智能的道德判断完全基于其被编程的规则，而不是内在的、自发的道德感知。最后，人工智能在面对不同情境时的灵活性有限。虽然人工智能可以在给定的参数和规则内作出调整，但它们通常无法理解复杂情境的深层道德和伦理含义，也无法像人类那样根据情境的细微变化灵活调整其判断。简而言之，人工智能在执行任务和处理大量数据方面展现出了卓越能力，但在伦理决策上，它们与人类相比仍然有显著差距。这一限制是人工智能技术短期内无法解决的短板。

3. 情感与道德在人工智能中的缺失

3.1. 人工智能情感与道德缺失的现状

尽管人工智能技术取得了显著进展，但其在情感和道德方面的能力仍然有限。

十七世纪的哲学家约翰·洛克认为情感起源于感官输入和经验的深思熟虑。他指出，通过反思自身以及事物如何影响我们的内心，我们能够洞悉情感的产生。洛克进一步阐述，无论是感官受外界刺激还是内心的思考，都能引发我们的快乐或痛苦，“我们的情感观念是由感觉和反省得来的”[1]。休·布莱克则认为情感始于感官印象，即通过五感感知外界。想象力在情感生成中扮演中介角色，使个体能够重新处理和内化感官印象，创造更丰富和复杂的情感体验。按照他的理论，当人们观赏一幅美丽的风景画时，会通过想象力将这幅画与个人记忆、情感和愿望联系起来，产生更深刻和情感丰富的体验。从生物学角度看，情感的产生是一个涉及多个大脑区域的复杂过程，包括杏仁核、海马、前额叶等。早在1939年，Kluver和Bucy的研究发现，切除猴子双侧颞叶会导致一系列特征性行为改变，包括情感反应的缺失，这一现象被称为“Kluver-Bucy综合征”[2]。这一发现清楚地表明，颞叶结构在情感调控中起着至关重要的作用。换言之，这个实验充分证明情感是大脑生理过程的产物。但人工智能只是由数据代码组成的实体，缺乏生物体的生理基础和人类思想的感官与想象能力，因此无法产生情感，限制了其模拟和理解人类情感的能力。

在道德地位的问题上，瑞典哲学家尼克·博斯特罗姆在《人工智能的伦理学》一文中指出，尽管普遍认为人工智能系统缺乏道德地位，但仍存在对这一地位确切属性的不确定性。他提出，知觉和智慧是判断道德地位的两个关键因素。根据“基材无歧视原则”，若人工智能系统具有感知能力和类似成年人的智慧，理论上应当享有与人类相同的道德地位[3]。然而，由于人工智能的非实体性质和感知能力的缺乏，目前的技术状态使其无法获得道德地位。荷兰特温特理工大学的布瑞认为，道德主体应具备基于对错进行推理、判断和行动的能力，以及对自己行为后果的责任感。他将人工智能视为一种“准道德主体”，可能具备进行道德决策的能力，但不具备道德思考的能力。他将当前的人工智能视为“准道德主体”，可能具备一定的道德决策能力，但缺乏深层的道德思考能力[4]。塞拉斯等人则认为，自主性、意向性和责任感是成为道德主体的必要条件。由此来看，目前的人工智能由于缺乏自主思维、决策能力和承担责任的能力，因而不符合道德主体的标准。

综上，人工智能既缺乏情感产生的基础条件又不是道德主体，这表明人工智能在做伦理决策时无法依据情感或道德，只能依赖于程序员设定的指导原则或算法。

3.2. 缺失对人工智能决策的影响

人工智能决策过程中情感和道德的缺失，对于其应用和效果产生了深远的影响。首先，由于缺乏情感，人工智能无法完全理解和回应人类情感需求。例如，在医疗或心理咨询领域，人工智能可能提供技术上准确的建议，但缺乏对患者情感状态的理解和同情，这可能影响治疗效果和患者满意度。

其次，由于缺乏道德判断力，人工智能可能在决策过程中无意中促成不公正或歧视性的结果。例如，在人力资源管理中，如果人工智能在招聘过程中使用的数据存在性别或种族偏见，它可能做出歧视性的招聘决策，即使其程序设计本身并无此意图。这不仅违背了道德和公正的原则，也可能导致法律责任和公众信任的损失。

此外，人工智能在处理复杂的道德问题时可能表现出缺乏灵活性和创新性。“人工智能不可能超出所编程序规定的范围，依靠自主意识进行独立判断并实施行为，不可能在实际生产生活中面对复杂的情况，作出道德决策。” [5]例如，在自动驾驶汽车中，如何在潜在的交通事故中做出道德选择(例如，保护乘客还是行人)是一个复杂的道德问题，目前的人工智能技术还无法完全解决。

3.3. 案例分析：具体应用中的伦理挑战

考虑到人工智能在情感和道德方面的局限性，我们可以通过一个具体案例来分析其伦理挑战：慈善捐款的决策。当人们通过媒体或社交网络看到饥饿儿童的慈善宣传时，他们可能会受到同情心和同理心的触动，从而做出捐款的决策。这种情感反应源于对儿童困境的理解和共鸣。然而，若这一决策过程被交由人工智能执行，可能会引发多种伦理问题。

3.3.1. 排除特定受益人群

若人工智能的历史数据中存在对某些群体的偏见，例如性别、种族或社会地位的歧视，人工智能可能无意中排除这些群体，使他们无法获得捐款支持，违背了慈善捐款的初衷。这不仅违背了慈善捐款的初衷，也暴露了人工智能在伦理决策中的缺陷。

3.3.2. 缺乏灵活性

人工智能在处理援助项目时，通常按照设定的标准参数运作，如预算上限、预期效益和执行期限等。然而，在面对如灾难援助等复杂情境时，由于人工智能难以准确预测和适应不断变化的状况，其决策可能无法及时调整，导致救援行动延误或不当。

这些案例表明，人工智能在伦理决策中的主要局限性在于其缺乏情感和价值判断能力。与人类不同，人工智能无法真正理解和感受情感，也缺乏对伦理问题的内在判断。它的决策依赖于预设的规则和数据输入，而不是基于自主的伦理判断。因此，未来的研究和实践需要在情感智能和人工智能之间寻求平衡，以满足伦理需求并尊重人类的人道价值观。

4. 伦理问题的根源：人类决策

与人工智能相关的伦理问题主要体现在人工智能系统的设计、开发和应用都直接或间接地受到人类伦理观、价值观和决策的影响。我们明确人工智能自身是没有伦理问题的，伦理问题源于人类决策和行为。

4.1. 人类在设计中的应用中的责任

在人工智能的背景下，人类在设计中的应用中的责任至关重要。这种责任不仅包括在开发过程中考虑伦理问题，如确保人工智能系统的公平性、透明度和尊重用户隐私，还包括在人工智能系统部署和运行

期间的持续监督和管理。首先，设计师和开发者在构建人工智能时，需要确保它们不会无意中加剧现有的不平等或偏见。这意味着必须在数据集的选择、算法的设计和最终应用的实施过程中考虑到多样性和包容性。其次，尊重隐私是设计和应用人工智能时不可忽视的方面。这意味着必须在数据收集、处理和存储过程中严格遵守隐私保护标准，确保个人数据不被滥用或泄露。最后，设计者和开发者还需要考虑他们的作品可能对社会造成的长远影响，比如就业市场的变化、社会结构的调整以及伦理和法律规范挑战。他们应该与政策制定者、社会学家和伦理学家合作，以确保人工智能的发展不仅技术上先进，而且在道德和社会层面上也应负责任。

4.2. 政策和法规在伦理指导中的作用

政策和法规在提供人工智能伦理指导方面发挥着关键作用。政府和监管机构负责制定确保人工智能负责任和伦理使用的框架和指南。这些政策和法规通常包括但不限于隐私保护和数据安全、防止歧视、透明度和可解释性以及问责制和人类最终决策权。这些措施旨在创建一个支持创新同时确保社会福祉和道德标准的环境。这不仅指导了人工智能的开发者 and 使用者，还为在伦理违规情况下的问责和执行提供了机制。这种平衡对于获得公众信任、促进技术进步和保护社会价值至关重要。

5. 结论

人工智能，本质上是一套遵从预设逻辑与编程指令的代码体系，它无法自发形成意识、情感，也缺乏独立的道德判断能力。因此，人工智能本身并不涉及伦理问题；相反，伦理挑战源于人类对这些系统的设计和应用。为了确保人工智能的负责任使用，我们必须深入反思人类设计中的伦理困境，并采取相应措施。确保透明度、保持信息来源的多样性以及进行持续的伦理评估，是解决这些挑战的核心方法。这些措施将有助于保障人工智能技术在道德和社会责任方面的正当性，确保其符合伦理标准。同时，我们也不能忽略人工智能带来的潜在风险，应当积极探索和实施各种策略，以保证伦理意识与技术进步能够协调并进，共同促进社会的健康发展。这不仅是技术创新的要求，也是对人类自身责任和智慧的考验。

参考文献

- [1] [英]洛克. 人类理解论[M]. 关文运, 译. 北京: 商务印书馆, 1983.
- [2] Kluver, H. and Bucy, P.C. (1937) 'Psychic Blindness' and Other Symptoms Following Bilateral Temporal Lobectomy. *American Journal of Physiology*, **119**, 254-284.
- [3] Bostrom, N. and Yudkowsky, E. (2014) The Ethics of Artificial Intelligence. Frankish, K. and Ramsey, W., Eds., *Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, New York, 316-334. <https://doi.org/10.1017/CBO9781139046855.020>
- [4] Brey, P. (2014) From Moral Agents to Moral Factors: The Structural Ethics Approach. In: Kroes, P. and Verbeek, P.P. Eds., *The Moral Status of Technical Artifacts*, Springer, Dordrecht, 125-142. https://doi.org/10.1007/978-94-007-7914-3_8
- [5] 张今杰. 人工智能体的道德主体地位问题探讨[J]. 求索, 2022(1): 58-65.