

负责任创新视域下人工智能赋能脑机接口的伦理治理

马鑫宇, 刘晓宇

锦州医科大学马克思主义学院, 辽宁 锦州

收稿日期: 2026年9月22日; 录用日期: 2026年10月13日; 发布日期: 2026年10月26日

摘要

人工智能技术的嵌入显著提升了脑机接口在神经信号采集、解码、反馈与个体适配方面的能力, 推动脑机接口由实验室研究加速走向临床康复、智能交互和人体功能修复等应用场景。然而, 人工智能赋能并非单纯的技术增效过程, 也使脑机接口风险呈现出复合化、隐蔽化和扩散化特征。算法黑箱可能导致神经信号误判, 意识解析可能冲击主体能动性, 神经数据处理可能引发隐私泄露与数据滥用, 责任归属和权属边界也因“人脑-算法-设备”共同作用而趋于复杂。负责任创新理论强调在科技创新全过程中嵌入预测、反思、协商与回应机制, 为人工智能赋能脑机接口的伦理治理提供了重要分析框架。基于此, 应从前瞻性风险评估、算法透明与可审计、神经数据全生命周期保护、多主体协同审查和责任规则完善等方面构建治理体系, 以推动人工智能赋能脑机接口技术安全、规范、可持续发展。

关键词

人工智能, 脑机接口, 负责任创新, 伦理风险, 神经数据, 伦理治理

Ethical Governance of AI-Enabled Brain-Computer Interfaces from the Perspective of Responsible Innovation

Xinyu Ma, Xiaoyu Liu

School of Marxism, Jinzhou Medical University, Jinzhou Liaoning

Received: September 22, 2026; accepted: October 13, 2026; published: October 26, 2026

Abstract

The embedding of artificial intelligence technologies has significantly enhanced the capabilities of

文章引用: 马鑫宇, 刘晓宇. 负责任创新视域下人工智能赋能脑机接口的伦理治理[J]. 哲学进展, 2026, 15(10): 331-336. DOI: 10.12677/acpp.2026.1510496

brain-computer interfaces (BCIs) in neural signal acquisition, decoding, feedback and individual adaptation, accelerating the transition of BCIs from laboratory-based research to application scenarios including clinical rehabilitation, intelligent interaction and human-body function restoration. Nevertheless, AI-enablement is not merely a process of technical efficiency improvement; it also renders risks associated with BCI compound, covert and diffuse. The algorithmic black box may give rise to misinterpretation of neural signals; consciousness decoding may challenge subject agency; neural data processing may trigger privacy breaches and data misuse. Furthermore, the attribution of responsibility and demarcation of ownership boundaries have grown complicated due to the joint interaction among the human brain, algorithms and devices. Emphasizing the integration of foresight-oriented prediction, reflection, deliberation and responsiveness throughout the whole process of scientific and technological innovation, the theory of Responsible Innovation provides a vital analytical framework for the ethical governance of AI-enabled brain-computer interfaces. Accordingly, a governance system should be constructed from the perspectives of proactive risk assessment, algorithmic transparency and auditability, full-life-cycle protection of neural data, multi-agent collaborative review and improvement of liability rules, so as to promote the safe, standardized and sustainable development of AI-enabled BCI technologies.

Keywords

Artificial Intelligence, Brain-Computer Interface, Responsible Innovation, Ethical Risk, Neural Data, Ethical Governance

Copyright © 2026 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

脑机接口是在人脑与外部设备之间建立直接信息交流通道的前沿技术。传统意义上的脑机接口主要通过采集、解码和转化脑神经信号,使使用者能够不依赖外周神经和肌肉系统完成对外部设备的控制[1]。随着神经科学、人工智能、材料科学和医学工程的发展,脑机接口已经在运动康复、交流辅助、神经调控、意识障碍评估等领域展现出重要价值[2]。尤其是在人工智能技术的支撑下,脑机接口系统对复杂神经信号的识别、分类、预测和反馈能力显著提升,技术应用边界不断拓展。

现有研究已经从医学伦理、生命伦理、算法伦理、数据安全和主体哲学等角度讨论脑机接口的风险问题。有学者指出,脑机接口可能引发生理安全、心理安全、隐私保护、自由意志、公平分配和身份认同等问题[3][4];也有研究从人的自主性、自我延展和身体替代等角度揭示脑机接口对主体存在方式的影响[5]-[7]。近年来,随着人工智能算法深度嵌入脑机接口系统,风险的生成机制发生了新的变化。人工智能不仅提升脑机接口的运行效率,也使技术风险从单一设备风险转向算法、神经数据、主体能动性和社会公正交织的复合风险。

负责任创新理论强调,科技创新不应只关注技术效率和经济价值,还应在创新全过程中主动识别、反思和回应潜在伦理后果[8]。这一理论与脑机接口治理具有较强契合性。脑机接口直接关联人的大脑、意识、身体和人格价值,人工智能的加入进一步放大了其伦理敏感性。因此,有必要从负责任创新视角重新审视人工智能赋能脑机接口的风险类型与治理路径。

2. 人工智能赋能脑机接口的技术逻辑与应用价值

2.1. 提升神经信号解码能力

脑机接口运行的核心环节包括信号采集、信号解码、指令输出和反馈控制。由于人脑神经信号具有

微弱性、复杂性、动态性和个体差异性, 传统算法在处理高维脑电数据时容易受到噪声干扰, 进而影响识别精度。人工智能特别是机器学习和深度学习算法, 能够从大量脑电图、皮层电图和神经元活动数据中提取特征, 建立信号与意图之间的对应关系, 从而提升脑机接口的解码能力[9]。这使脑机接口能够更准确地识别使用者的运动意图、交流需求和认知状态, 为临床康复和智能交互提供技术基础。

2.2. 推动脑机交互实时化

脑机接口要真正应用于临床康复和外部设备控制, 必须实现较低延迟的实时反馈。人工智能算法能够压缩神经信号处理流程, 提高数据分类和指令生成效率, 使脑机接口系统更快地响应使用者意图。在脑卒中康复、脊髓损伤功能代偿、神经调控等场景中, 实时反馈不仅影响使用体验, 也直接关系到治疗安全和康复效果[2]。因此, 人工智能赋能脑机接口的重要价值不仅在于提高识别准确率, 还在于使脑机交互从“可实现”走向“可持续使用”。

2.3. 增强系统个体适配能力

不同个体的大脑结构、神经活动模式、疾病状态和训练经验存在明显差异, 同一用户在不同情绪、疲劳程度和环境刺激下也可能产生不同信号特征。人工智能算法的自适应学习能力有助于脑机接口系统根据个体差异持续调整模型参数, 提升长期使用中的稳定性和适配性。由此, 脑机接口不再只是固定程序控制的技术装置, 而逐渐成为能够与使用者共同训练、共同调整的动态系统。这种技术特征为个性化医疗和精准康复提供了可能。

2.4. 拓展临床与社会应用场景

脑机接口最初主要服务于严重运动障碍、闭锁综合征和神经系统疾病患者, 具有明显的治疗和辅助属性[4]。但随着人工智能能力增强, 脑机接口的应用逐渐从治疗修复扩展到人体增强、智能交互、教育娱乐、工业控制甚至军事领域[3]。这种应用扩展提升了技术社会价值, 也使伦理风险更加复杂。治疗型应用主要强调患者利益和医学安全, 而增强型、商业型和军事型应用则可能涉及公平、滥用、权力支配和国家整体安全等更广泛问题。

3. 人工智能赋能脑机接口的伦理风险

3.1. 算法黑箱与信号误判风险

人工智能赋能脑机接口首先带来算法黑箱风险。脑机接口中的人工智能模型往往依赖复杂的机器学习算法, 其内部运行机制并不完全透明。算法如何从神经数据中提取特征、如何判断使用者意图、为何输出某一控制指令, 普通用户甚至研发人员都可能难以完全解释。王高峰等指出, 算法的不透明性、不确定性和易误导性, 会在脑机接口场景中引发安全、隐私、责任和歧视等问题[10]。如果算法错误识别用户意图, 可能导致外部设备误操作; 在医疗场景中, 误判还可能影响康复训练、神经刺激或临床决策, 造成身体和心理伤害。

3.2. 意识解析与主体能动性风险

脑机接口不同于一般人机交互技术, 它直接面向人的大脑神经活动。人工智能算法对神经信号的深度解析, 使技术系统可能触及个体的情绪状态、认知倾向、行为意图甚至潜在心理活动。虽然当前技术尚不能完全“读取思想”, 但其发展方向已经对思想隐私和精神自主提出挑战。魏郡一认为, 脑机接口技术可能改变人的自主性结构, 使人的行为决策受到外部技术系统影响[5]。金晨、张贵红则从自我延展

角度指出, 脑机接口可能使外部设备参与个体自我表达和行动过程, 从而改变自我边界[6]。因此, 人工智能赋能脑机接口的伦理风险不仅是数据风险, 更是对主体能动性、人格价值和自我同一性的深层挑战。

3.3. 神经数据泄露与滥用风险

神经数据是脑机接口运行的核心资源。与一般个人信息相比, 神经数据具有高度敏感性、个体唯一性和深度人格关联性, 可能反映个体的健康状况、认知状态、情绪变化和行为倾向[11]。人工智能算法能够通过大规模数据训练和模式识别, 从神经数据中推断更多潜在信息, 这使神经数据的二次利用风险显著增加。王静指出, 脑机接口数据安全风险贯穿采集、传输、解码和利用全过程, 表现为采集合规风险、公开泄露风险、归集滥用风险和写入操纵风险[12]。一旦神经数据被泄露、买卖或用于未经授权的商业分析, 个体隐私和人格权益将面临严重威胁。

3.4. 责任归属与权属边界风险

人工智能赋能脑机接口的行为结果通常不是单一主体完成的, 而是由使用者意图、算法解码、设备执行、医生操作和企业设计共同构成。当脑机接口出现误判、失控或损害后果时, 责任应由用户、医疗机构、算法开发者、设备生产者还是数据处理者承担, 往往难以清晰界定。与此同时, 脑机接口采集的原始神经数据、人工智能处理后的衍生数据以及由脑机接口生成的内容, 也存在权属不明问题。徐娟指出, 神经数据难以完全纳入一般个人信息处理逻辑, 应当作为具有高度人格关联性的特殊信息类型加以理解[13]。如果不能明确数据权属、使用边界和收益分配, 技术商业化过程可能引发新的法律和伦理冲突。

3.5. 技术增强与社会公平风险

脑机接口在医疗领域的初衷是修复功能、改善患者生活质量。然而, 随着人工智能能力提升, 脑机接口可能进一步用于认知增强、运动增强和感官增强。治疗与增强边界的模糊, 可能导致技术资源分配不公。马兰指出, 脑机接口不仅具有治疗疾病和促进产业发展的价值, 也可能削弱个体自主性、加剧社会不公并带来国家整体安全风险[3]。如果脑机接口增强技术只被少数高收入群体、特定机构或军事系统掌握, 可能造成新的能力差距和社会分层。这种风险已经超出个体伦理范围, 进入社会正义和公共治理层面。

4. 负责任创新视域下人工智能赋能脑机接口的治理路径

4.1. 预测维度: 建立前瞻性风险评估机制

人工智能脑电信号识别精度不足, 不仅源于算法模型泛化能力有限, 也受传统植入电极材料性能缺陷的制约, 二者共同导致信号采集噪声大、失真率高、解析偏差明显等问题。为精准解决该类技术短板, 需采取算法迭代与材料优化双向并行的优化方案。在算法层面, 引入自适应深度学习与迁移学习算法, 使模型可根据不同用户的个体脑电差异、实时生理状态动态调整参数, 有效过滤环境与生理噪声, 降低识别误差; 同时扩充多场景、多人群的标准化脑电数据集, 提升模型的通用性与容错率, 并增设结果实时校验机制, 进一步减少解码错误。在硬件材料层面, 摒弃传统硬质、易损耗的电极材料, 研发高柔性、高生物相容性、高信号传导率的新型电极材质, 减少植入后脑组织排斥、电极磨损、信号衰减与回撤问题, 从源头提升脑电信号采集的完整性与清晰度。通过软硬件协同优化, 全面提升人工智能对脑电信号的解析精准度, 保障脑机接口系统稳定、高效运行。

4.2. 反思维度: 推动算法透明与伦理嵌入设计

负责任创新强调研究者和技术开发者应持续反思创新目标、技术路径和社会影响。人工智能赋能脑

机接口的伦理治理, 应推动算法透明度、可解释性和可审计性建设。推动算法透明度是确保脑机接口技术负责任发展的重要环节, 这有助于帮助政府构建结构完善的治理体系, 推动科学研究的发展从而加强公众对脑机接口技术的信任。推动算法透明可以从以下三个方面进行:

第一, 政府监管部门应该制定一套系统完善的适用于脑机接口技术相关的社会、法律、伦理等问题的监管政策。监管部门在制定政策时还应该考虑到脑机接口技术应用不同场景时呈现出的不同发展特点, 以灵活性、多样性、包容性的原则推动脑机接口技术的发展。

第二, 脑机接口技术的研究者对于高风险医疗脑机接口, 应建立算法模型备案、训练数据说明、误差率披露和第三方审计机制。王高峰等提出, 应通过算法审计和用户紧急制动权降低脑机接口算法风险[10]。在设计脑机接口技术的过程中, 还应引入“伦理嵌入”理念, 将隐私保护、最小必要采集、用户授权、异常预警和人工干预机制内置于系统之中, 使伦理要求不只是外部规范, 而成为技术结构的一部分。

第三, 对于脑机接口技术的使用用户以及公众, 政府应建立公开平台倾听使用者及其家属和公众的观点, 推动所有的利益相关方在该平台公开透明的交流, 鼓励公众积极参与治理过程。欧盟委员会启动了名为“BNCI 地平线 2020”的脑机接口社区协调和支持行动, 鼓励公众参与[14]。

4.3. 协商维度: 构建多主体参与的伦理审查机制

脑机接口涉及患者、医生、科研人员、企业、监管机构和公众等多方主体, 单一主体难以完成有效治理。负责任创新要求通过开放协商机制吸纳不同利益相关者的意见。对于人工智能赋能脑机接口研究, 应完善伦理委员会审查制度, 特别关注受试者知情同意、风险收益评估、弱势群体保护和退出机制。由于许多脑机接口临床对象可能是严重残障患者或意识障碍患者, 其同意能力和真实意愿表达需要受到特别保护。伦理审查不能仅停留在形式化签字, 而应确保受试者及其家属充分理解技术风险、数据用途和可能后果。

4.4. 回应维度: 完善神经数据全生命周期保护

负责任创新的回应维度要求治理制度能够根据技术发展和社会反馈不断调整。神经数据治理应覆盖采集、存储、传输、解码、共享、商业使用和删除全过程。在采集阶段, 应坚持明确目的、单独同意和最小必要原则; 在存储和传输阶段, 应采用加密、匿名化、去标识化和访问权限控制; 在共享和商业化阶段, 应建立用途限定、备案审查和收益规则; 在侵权发生后, 应完善责任追溯和救济机制。对于能够反映认知状态、情绪倾向和行为意图的神经数据, 应比一般健康数据采取更高等级的保护。

4.5. 协同治理: 形成法律、伦理、技术与行业标准联动

人工智能赋能脑机接口的风险具有跨学科、跨场景和跨主体特征, 仅依靠伦理倡议难以实现有效治理。应构建法律规制、伦理审查、技术标准和行业自律相结合的协同治理体系。政府部门应推动脑机接口数据保护、临床试验、设备安全和算法审计等规则建设; 医疗机构应加强临床应用规范和患者保护; 企业应承担产品安全、数据合规和算法透明义务; 科研机构应在研究设计阶段嵌入伦理评估; 公众和患者组织也应参与技术应用边界讨论。通过多主体协同, 才能在促进创新与防范风险之间形成动态平衡。

5. 结语

人工智能赋能脑机接口技术具有重要医学价值和社会价值, 它能够提升神经信号解码能力, 推动脑机交互实时化和个体化, 为神经系统疾病诊疗、功能康复和人机协同提供新的可能。然而, 人工智能的嵌入也改变了脑机接口风险的生成方式, 使其不再只是单纯的设备安全或医学伦理问题, 而是算法黑箱、神经数据、主体能动性、责任归属和社会公平共同构成的复合治理问题。负责任创新理论为这一问题提

供了较为系统的分析框架。未来, 人工智能赋能脑机接口的发展应坚持以人为本, 将人的尊严、自主、隐私、安全和公正置于技术创新的中心位置, 通过预测、反思、协商和回应机制, 实现技术进步与伦理治理的同步推进。

参考文献

- [1] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G. and Vaughan, T.M. (2002) Brain-Computer Interfaces for Communication and Control. *Clinical Neurophysiology*, **113**, 767-791.
[https://doi.org/10.1016/s1388-2457\(02\)00057-3](https://doi.org/10.1016/s1388-2457(02)00057-3)
- [2] 支迪暄, 金磊, 王逸鹤, 等. 脑机接口临床实践与展望[J]. 中国现代神经疾病杂志, 2026, 26(1): 13-20.
- [3] 马兰. 脑机接口: 应用价值、伦理风险与治理框架[J]. 中国医学伦理学, 2026, 39(2): 159-166.
- [4] 李佩瑄, 薛贵. 脑机接口的伦理问题及对策[J]. 科技导报, 2018, 36(12): 38-45.
- [5] 魏郡一. 脑机接口技术: 人的自主性问题及其伦理思考[J]. 医学与哲学, 2021, 42(4): 27-31.
- [6] 金晨, 张贵红. 脑机接口对自我的延展[J]. 医学与哲学, 2024, 45(15): 1-5, 21.
- [7] 刘光明. 脑机接口替代身体“限度”的哲学审视[J]. 医学与哲学, 2024, 45(9): 11-15.
- [8] Stilgoe, J., Owen, R. and Macnaghten, P. (2013) Developing a Framework for Responsible Innovation. *Research Policy*, **42**, 1568-1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- [9] Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. and Arnaldi, B. (2007) A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces. *Journal of Neural Engineering*, **4**, R1-R13.
<https://doi.org/10.1088/1741-2560/4/2/r01>
- [10] 王高峰, 张志领. 算法伦理视域下的脑机接口伦理问题研究[J]. 自然辩证法研究, 2022, 38(7): 68-73.
- [11] 王静. 脑机接口应用的数据安全风险及其治理[J]. 中国科技论坛, 2026(4): 109-117.
- [12] 陈龙. 脑机接口的伦理问题思考[J]. 医学与社会, 2025, 38(2): 75-82, 92.
- [13] 徐娟. 脑机接口技术应用中神经数据的法律属性与共享边界[J]. 河北法学, 2026, 44(7): 127-147.
- [14] Blank, R.H. (2023) United States Policy on BCIs: Funding Research, Regulating Therapies, and Commercializing Consumer Technology. In: Dubljević, V. and Coin, A., Eds., *Policy, Identity, and Neurotechnology*, Springer, 189-206.
https://doi.org/10.1007/978-3-031-26801-4_11