AI技术浪潮下自由意志面临的伦理冲击与应对策略

王思忆

摘　要：AI技术发展的智能越高级、性能越强、部署越广，其突破人类伦理框架或造成高代价伦理后果的潜在风险也随之增大。本论文聚焦AI技术发展给人类自由意志带来的诸多挑战及其引发的一系列伦理困境，并提出相应的应对策略，以期在AI时代维护人类自由意志，保障社会伦理秩序。

关键词：AI技术；自由意志；伦理困境；应对策略

一、AI技术的发展历史

人工智能这一概念于1956年由“人工智能之父”麦卡锡提出，英文表述是Artificial Intelligence，简称AI。人工智能技术在其发展的过程中大致经历了以下5个阶段：孕育时期（1956前）、形成时期（1956–1970）、暗淡时期（1966–1974）、知识应用时期（1970–1988）、集成发展时期（1986–至今）。[1]其发展过程被划分为三个阶段：第一，弱人工智能（Artificial Narrow Intelligence）是指机器按照设计者给出的既定指令做出相应的反映，按照原先设定好的程序进行活动，不能按照自己的意愿来完成；第二，强人工智能（Artificial General Intelligence）是指人工智能产品拥有情感、意识，能够自己对外界的刺激做出相应的反应，并且能够分析对方给出的刺激，给出一定的应对策略，并且举一反三。第三，超人工智能（Artificial Super Intelligence），超级智能产生于未来，是人工智能的未来发展模式。[2]

二、AI对人类自由意志的冲击

AI技术已成为21世纪最前沿、最热门的高新技术之一，人工智能已经开始进入人类的日常生活，构成了一个庞大的系统，深刻地改变了社会各领域的运行模式，我们的选择自由、个人隐私正被悄然侵蚀，人类作为主体的责任归属也将出现混乱。

首先，AI技术算法对个人自由选择的限制。AI技术强大的数据采集、存储和分析能力，能够对海量的数据进行快速处理，挖掘出其中蕴含的各种信息模式，精准地推送着符合我们过往偏好的信息，例如，社交媒体平台通过分析用户的点赞、评论、分享等行为数据，就能准确把握用户的兴趣偏好，进而在用户的信息流中推送相关内容，它看似满足了我们的需求，实则在不知不觉中限制了我们的视野，我们所接触到的内容越来越单一，思维也逐渐被同质化。AI在为人类提供建议或决策支持时，也可能会限制人类的选择范围，比如，智能驾驶导航系统在车辆行驶的实际应用中，它能够根据实时路况为用户提供最优驾驶路线规划建议，但长期依赖这种建议可能会使用户陷入所谓的“信息茧房”，即用户只接触到系统推荐的有限选择，而忽略了其他可能的路径或决策方式。这种情况长期发展下去，势必会削弱人类自由意志在选择过程中的作用。

其次，AI技术对人类行为的因果解构。人工智能凭借其强大的数据分析能力，对因果关系的挖掘达到了前所未有的深度。它能够通过对海量的心理和行为数据进行分析，找出各种行为之间隐藏的因果关联，这些关联往往是人类自身在日常认知中没有充分意识到的，从而可能造成因果关系重塑的问题。例如，通过分析心理和行为数据，人工智能可能发现某些童年经历和成年后的犯罪行为之间存在很高的关联性。这就引发了一个问题：如果人类的行为可以被归结为一系列复杂的因果关系，而这些因果关系又可以被人工智能所揭示，那么人的自由意志在其中处于什么样的位置呢?

最后，AI技术的数据交互加剧个人隐私泄露。随着大数据技术的不断发展，人类的隐私空间也在不断被压缩，甚至人类已经没有绝对的隐私可言，人类自由意志将面临逐步弱化甚至丧失的困境，因为人们所作的一切

课题项目：西南民族大学中央高校基本科研业务费专项资金项目（人文社会科学类）—功利主义视角下智能驾驶的伦理困境问题研究，项目编号：2025SYJSCX27。

作者简介：王思忆（1993.06——）女，汉族，本科学历，西南民族大学研究生在读，主要从事伦理学方面的研究工作。

选择都将是大数据操作的结果，而并不完全是行为主体在自由意志的支配下所作的选择，这势必会引发责任归属的混乱。例如，Facebook泄露用户隐私事件以及疫情期间全球最大的视频会议软件Zoom泄露隐私事件等等，在这些事件中个人的姓名、联系方式和私人的亲密聊天等被售卖或公开在网上，对用户的个人隐私造成了极其严重的损害。

三、AI技术引发的伦理困境

AI技术对人类自由意志带来的冲击，势必引出相应的伦理困境。

第一，伦理责任划分的困境。在AI参与决策和行动的诸多情境下，由于自由意志的界定变得复杂，责任主体的确定成为一个棘手的问题。从大数据算法推荐来看，当算法基于用户数据进行分析并给出推荐结果，而用户依据这些推荐做出了某种选择并产生了不良后果时，很难明确责任究竟在于算法本身的不合理性（可能存在数据偏差或算法逻辑缺陷），还是在于用户自身没有充分运用自由意志对推荐进行甄别和判断。这种责任归属的模糊性对现有的法律和道德规范形成了挑战，因为传统的法律和道德规范在面对这种新型的责任界定困境时，缺乏明确的指导原则。

第二，人机信任的伦理危机。当人们感知到智能机器人的行为可能受到程序和数据的严格限制而缺乏真正的自由意志时，往往会对机器人的可靠性产生怀疑，从而降低对机器人的信任。例如，在一些服务行业中，当顾客发现服务机器人的回应和行为都是按照预设的程序进行，缺乏灵活性和自主性时，可能会觉得机器人无法真正理解自己的需求，进而对其服务质量产生担忧，不愿意与机器人进行深入的互动。但当机器人表现出过高的自主决策能力（类似拥有自由意志）时，人们又可能因担心其失控而产生不信任感。

第三，道德决策的困境。大数据分析提供的信息往往会影响人们的判断。例如，在医疗领域，当医生在为患者制定治疗方案时，AI技术可能会提供各种不同的治疗案例、成功率等数据，这些数据可能会干扰医生基于自身专业知识和对患者具体情况的了解所做出的判断。医生可能会在遵循自身自由意志（依据专业判断和对患者的关怀）与适应AI影响（参考大数据分析结果）之间产生犹豫，不知道该如何平衡两者以做出最符合伦理的决策。机器人的行为不符合人类传统道德观念时，也会给人类的道德决策带来挑战，比如，在一些涉及机器人参与救援行动的情境中，如果机器人按照程序设定优先救援某些具有特定特征的人员（如根据数据分析认为这些人员生存几率更高），而忽略了其他同样需要救援的人员，这就与人类普遍秉持的公平、正义等价值观念相冲突。

四、应对策略

面对AI技术对人类自由意志的冲击，其核心的问题是，在利用AI为我们带来的便利的同时，如何保证人类主体的自由意志不被算法所引导和束缚。我们可以从以下几个方面努力。

第一，坚持以人为本的技术设计。在技术层面通过提高AI系统的透明度，开发可解释性的AI算法，使得人类能够理解AI系统做出决策的依据和过程。例如，在AI学习模型中，采用可视化技术将神经网络的决策过程以直观的方式展示出来，让用户了解到数据输入、处理到最终转化为输出结果的全过程。这样，当AI系统给出推荐或决策时，用户能够清楚地知道其背后的逻辑，从而在一定程度上维护自己的自主判断能力，避免盲目接受AI的建议。在程序算法设计层面，将伦理原则融入其中，使程序算法设计更符合人类伦理道德观念。例如，在设计资源分配算法时，要遵循公平、正义的原则，避免出现因算法导致的资源分配不公正的情况。对于智能机器人行为模式的设计，要使其行为更加符合人类的道德认知和社会规范，要让其能够表现出尊重、友善等符合人类道德情感的行为，使其在与人类互动过程中能够更好地被接受。

第二，加强对数据隐私的保护。采用加密技术对用户数据进行全链路的加密处理，通过采集端到处理端加密机制，确保数据在采集、传输和存储过程中的安全性，避免源头泄露风险。建立严格的数据访问权限管理制度，限制未经授权的人员对数据的访问和使用，防止数据泄露事件的发生，从而维护用户的隐私权益，间接保障用户的自由意志不受侵犯。同时，需要警惕AI算法推荐可能带来的“信息茧房”效应，防止技术对人们自由意志的隐性操控。从而形成覆盖数据全生命周期的安全防护闭环，从“用户—监管—潜在风险”三个层面筑牢用户数据隐私与信息安全的防线。

第三，提升AI技术开发者和使用者伦理素养。通过加强伦理教育与培训，确保AI技术在符合伦理道德的层面得以开发和使用。对于AI技术开发者而言，通过开展专业的伦理培训和实际案例分析，使他们掌握在设计和开发AI系统及机器人的过程中，应当遵守的伦理道德原

则，了解到不同的算法设计和机器人行为模式可能引发的伦理问题，从而在实际工作中更加注重伦理考量，避免因技术应用而对人类自由意志造成不必要的冲击。对于AI技术使用者来说，通过举办科普讲座、发放宣传资料等方式，培养他们正确的伦理观念和道德责任感，以便在应用AI技术时能够做出合理的决策和选择。

第四，完善AI技术政策法规的制度建设。首先是明确AI系统在不同应用场景下的责任界定标准，针对大数据算法推荐、自动驾驶汽车、机器人执行任务等不同场景，分别制定详细的责任划分规则。例如，在自动驾驶汽车领域，应当明确规定在不同事故情况下，汽车制造商、算法开发者、汽车用户等各主体应承担的责任比例和范围，以便在发生事故时能够迅速、准确地确定责任主体，避免责任归属的模糊不清。其次是出台AI技术应用管理政策，规范AI技术在涉及人类自由意志相关领域的应用范围和方式，特别是对AI技术在医疗、教育、金融等领域的应用进行严格监管，确保符合人类的伦理观念和自由意志保护的要求。最后要加强人工智能技术伦理治理制度的执行和监督，成立独立的管理机构或审查委员会定期对AI系统和机器人的运行情况、伦理合规性等进行检查和评估。对于不符合政策法规要求的AI产品和服务，要责令其整改或予以取缔，从而确保AI技术在健康、有序的轨道上发展，维护社会的伦理秩序和人类的自由意志。

结语

AI技术的发展和广泛应用不仅推动了产业升级、创新，提高了生产效率，一定程度上改变了人们的沟通、出行等生活方式，同时对人类自由意志带来一定的冲击，更衍生出事故责任界定模糊、人机信任危机等伦理困境。采取相应的策略，可以在一定程度上抵御AI技术对人类自由意志的潜在消解，维系社会伦理秩序的存续根基，更可为人工智能技术的规范化发展提供制度框架与价值坐标，使其始终锚定于人类主体地位的维护与社会整体福祉的增进。但在未来的发展中，我们还需要不断关注AI技术的新进展以及其对自由意志的影响，以持续优化应对策略，以确保人类在AI时代既能充分享受技术带来的便利，又能坚守自由意志、承担道德责任，最终实现技术演进与人文价值的辩证统一。

参考文献

[1]Robert Kane, The Significance of Free Will, New York: Oxford Univ Pr on Demand,1998.

[2]蔡白兴，徐光祐.人工智能及其应用（第四版）[M].清华大学出版社，2010：3-9.

[3]赵一秀.人工智能技术的伦理风险及对策研究[D].南京林业大学，2018.

[4]王骁.AI生成内容场景下自然人作者身份认定的对应关系理论[J].西北政法大学学报，2024年第6期，76-88.

[5]周翔.人工智能伦理困境与突围[J].哈尔滨师范大学社会科学学报，2020，11（6）：34-38.

[6]侯苏豫.机器人伦理原则和自由意志影响人机信任的认知与情绪机制[D].闽南师范大学，2024.

[7]陈仕伟.大数据时代的决定论与自由意志及其伦理问题[J].山东科技大学学报（社会科学版），第19卷第6期，2017年12月，19-24.

[8]李天依.大数据时代下的自由意志[J].大众文艺文史哲研究，（2024）15，90-92.

[9]刘燊.量子人工智能中的意识、自由意志与伦理挑战[J].山东科技大学学报（社会科学版），第26卷第6期，2024年12月，1-10.

[10]韩水法.人工智能时代的自由意志[J].社会科学战线，2019（11）：1-11+281.

[11]莫宏伟.强人工智能与弱人工智能的伦理问题思考[J].科学与社会，2018，8（1）：14-24.

[12]傅启宸.探究自由意志的存在性与伦理道德的关系[J].荆楚学术.学术探究.2020年3月.182-183.