

弱人工智能时代的刑事责任主体和责任承担简析

凌 洁

华东政法大学, 研究生院, 上海

收稿日期: 2025年6月15日; 录用日期: 2025年7月8日; 发布日期: 2025年7月17日

摘 要

本文探讨了弱人工智能时代刑事责任主体和责任承担的问题。随着人工智能技术的快速发展而带来的刑事风险对传统刑事责任理论提出了新的挑战。本文首先界定了弱人工智能机器人和强人工智能机器人的划分, 其次分析了在当前所处弱人工智能机器人时代下, 弱人工智能作为刑事责任主体的可能性, 及其设计者、生产者和使用者之间的责任承担。最后, 提出了完善弱人工智能时代刑事责任体系的建议, 包括立法完善、司法应对和技术保障等方面。本研究对于应对人工智能技术带来的刑事法律挑战具有重要的理论和实践意义。

关键词

人工智能, 犯罪风险, 刑事责任

A Brief Analysis of Criminal Legal Subject and Criminal Liability in the Era of Artificial Narrow Intelligence

Jie Ling

Graduate School, East China University of Political Science and Law, Shanghai

Received: Jun. 15th, 2025; accepted: Jul. 8th, 2025; published: Jul. 17th, 2025

Abstract

This article explores the issues of criminal legal subject and liability assumption in the era of artificial narrow intelligence. The criminal risks brought about by the rapid development of artificial intelligence technology pose new challenges to the traditional criminal responsibility theory. This

paper first defines the division between Artificial Narrow Intelligence and Artificial Superintelligence. Secondly, it analyzes the possibility of Artificial Narrow Intelligence Robot being the subject of criminal responsibility in the current era, as well as the criminal liability among designers, producers and users. Finally, suggestions for improving the criminal responsibility system were put forward, including aspects such as legislative improvement, judicial response and technical guarantee. This research has significant theoretical and practical significance for addressing the criminal legal challenges brought about by artificial intelligence technology.

Keywords

Artificial Narrow Intelligence (ANI), Criminal Risk, Criminal Penalty

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

当今时代，随着人工智能三要素：算法、算力和数据的大幅度提升和规模化的发展，把我们带入了一个史无前例的人工智能时代。可以说人工智能逐渐渗透到我们生活的方方面面，例如人脸识别、自动驾驶、智能穿戴、智能诊断等逐渐被应用于我们的生产生活领域，提升了生产效率、大数据优化算法有助于精准决策、更强的算力支持科研技术领域的创新[1]。与此同时，我们应当认识到人工智能也存在风险，我们也看到诸如自动驾驶事故带来的公共安全隐患，网络平台算法带来的侵犯公民个人信息的隐患，利用人工智能技术操纵证券市场而给社会经济和秩序带来的隐患，以及生成式人工智能在创造文艺作品过程中产生可能侵犯他人所有的著作权的问题等等。

2. 弱人工智能刑事犯罪的刑事责任主体

当下，学界多数认同以人工智能机器人是否具有独立意识和意志为依据，把人工智能划分成弱人工智能机器人与强人工智能机器人。并且，当前的人工智能普遍为弱人工智能机器人，是针对特定功能进行编程的并且不具备超越预定指令的能力的机器人。面对弱人工智能机器人产生的刑事责任，法律应当如何对其进行界定和规制？也就是“谁”来承担法律责任、承担怎样的法律责任，以及弱人工智能机器人本身是否可以成为刑法规范下承担责任的主体呢？

首先，由于弱人工智能机器人仅仅是执行设计、生产者所赋予其的代码和指令，本质上其只是一个拥有先进算法的机器人，其行为体现的是设计、生产者的意志和意识，不具备独立的自主意志和意识，不具有独立的辨认和控制能力，因而不能视为独立的责任主体。我们知道，法律责任是由特定法律事实所引起的对损害予以补偿、强制履行或接受惩罚的特殊义务。弱人工智能机器人的行为产生的损害责任应当由设计者或/及生产者或/及使用者承担。具体可以分为以下情况。

3. 设计者责任、生产者责任

设计者或生产者的犯罪故意。具有犯罪意图的研发者完全可以在设计和编制程序的时候植入自己的犯罪计划[2]。设计者在故意设计开发出弱人工智能机器人并且利用之进行犯罪，这种情况下设计者具有犯罪的故意，应当结合其具体的犯罪行为认定刑事罪名。如果人工智能的生产者明知设计者具有违法犯罪的目的，仍然按照设计者提供的设计进行生产组装人工智能机器人，此时可以认为生产者与设计者

共同犯罪，属于帮助犯，承担相应的责任。具体来说设计者、生产者的故意基于其明知其行为必然会产生危害结果，并且希望和积极追求该危害结果的产生。

设计者或生产者的犯罪过失。设计者或生产者违反法律规定其应当负有的注意义务导致危害结果的发生。具体可以分为：(1) 其应当预见，但因疏忽大意未能预见的情况以及；(2) 其应当预见且已经预见，但因过于自信轻信其能避免。例如，在达芬奇人工智能医疗机器人的中国代理商在 2018 年《对内窥镜手术机械控制系统主动召回》的事件报告中称对中国销售的 900 多个内窥镜手术控制系统进行召回。可见，为避免发生后续产品刑事责任，设计者或生产者有义务对可预见将产生危害结果的流通产品进行召回。具有缺陷产品召回义务的生产商如果没有及时召回缺陷产品，其将可能承担相应的产品行政责任、民事责任和刑事责任。

4. 使用者责任

使用者的犯罪故意。设计者或生产者出于合法的目的设计或者生产了弱人工智能机器人，但使用者出于故意的非法目的利用该人工智能机器人实施了犯罪行为，造成了危害结果的产生，使用者对此承担相应的刑事责任，而设计者或生产者对此不承担责任。例如，2024 年 2 月 3 日，上海市静安区人民法院对一起利用 AI 换脸进行诈骗的案件作出一审判决，该案件中犯罪分子利用人工智能换脸技术“假冒”某知名演员获取被害人信任后以各种理由索要钱款[3]。AI 换脸技术是一种基于人工智能和深度学习的图像处理技术，属于生成式人工智能，它的产生和发展在更低成本的情况下带动了影视创作中的艺术创新以及更多的娱乐和虚拟现实的应用，然而一旦被别有用心使用者恶意利用成了犯罪工具，那么其使用者应当承担刑事责任。

使用者的犯罪过失。使用者在使用弱人工智能机器人的同时，应当对其负有监管和管理责任。使用者违反法律规定其应当负有的注意义务导致危害结果的发生。具体可以分为：(1) 其应当预见，但因疏忽大意未能预见的情况；(2) 其应当预见且已经预见，但因过于自信轻信其能避免。近些年，伴随着智能驾驶的快速发展，智能汽车的道路交通事故也屡有发生，一般认为当今我国智能驾驶的发展处于“人机混驾”阶段，即由人类驾驶员与自动驾驶系统共同完成智能汽车的驾驶工作。该阶段，如果使用者在智能汽车自动行驶中违反安全使用注意义务，则应当承担责任。此类安全使用注意义务主要包括：(1) 对智能汽车自动驾驶功能进行正确操作与使用的义务；(2) 对智能汽车的接管义务，即在自动驾驶系统请求或在驾驶员认为有必要接管汽车时立即接管汽车的义务；(3) 是对自动驾驶汽车的警觉义务，即在智能汽车自动驾驶期间，保持必要的警觉，以便能随时履行接管义务，避免危害结果的发生[4]。因此，当驾驶人员在驾驶过程中“全程看手机”“全程睡觉”等违反上述注意义务而发生重大交通事故时，驾驶人员无疑作为智能汽车的使用者应当对其过失承担责任。

5. 设计者、生产者和使用者共同责任

根据我国《刑法》第二十五条：共同犯罪是指二人以上共同故意犯罪。二人以上共同过失犯罪，不以共同犯罪论处；应当负刑事责任的，按照他们所犯的罪分别处罚。当设计者、生产者和使用者持有共同故意，利用弱人工智能机器人进行犯罪，则应当认定共同犯罪。弱人工智能机器人设计者明知其设计的弱人工智能机器人将被用于犯罪而为犯罪分子提供技术支持可能构成帮助犯。例如，伴随着人工智能的不断发展，汽车行业正在逐步迈入智能汽车时代。我们享受其带来的安全、高效、舒适、节能等各项智能化优势，但如果自动驾驶汽车的设计者、生产者与使用者共谋利用自动驾驶汽车进行运输非法物品的违法犯罪活动呢？或者共谋利用自动驾驶汽车对特定目标建筑或者人员等进行攻击呢？面对这样共同的恶意地滥用人工智能技术行为，需要我们的法律制度对其进行规范，其设计者和使用者将构成共同犯

罪, 根据其犯罪的客观和主观要件可能构成运输危险物质罪、危害公共安全罪或者其他财产、人身犯罪等。又例如, 在证券交易领域, 人工智能技术凭借其高速的预测和分析能力以及强大的算法功能近年来逐渐广泛被使用于证券分析和交易领域。设计者开发出能够自动操纵证券市场的人工智能交易算法, 并且将如何利用其进行证券市场操纵的手法传授给使用者, 那么设计者和使用者的行为已经触犯了我国《刑法》第 182 条规定的操纵证券、期货市场罪, 应当承担相应的刑事责任。

6. 意外事件

当我们在考虑设计者、生产者和使用者之间的责任承担问题时, 有一点值得我们考虑: 依照当时的技术发展水平来衡量设计者、生产者是否对危害结果的发生具有预见可能性。根据《刑法》第 16 条之规定, 行为在客观上虽然造成了损害结果, 但不是出于故意或者过失, 而是由于不能抗拒或者不能预见的原因所引起的, 不是犯罪。如果受限于当时的科学技术水平, 设计者或者生产者难以识别或者修正已存在的技术或产品瑕疵, 则设计者或者生产者不应当承担刑事责任。还应当考虑到如果使用者在正常、合理、审慎使用弱人工智能机器人的过程中发生难以预料的机器程序运行错乱, 则使用者不应当承担刑事责任。应当属于意外事件[5]。

7. 结语

纵观人类的发展进程, 人工智能从二十世纪五十年代诞生起至今只有区区几十年, 尽管如今人工智能时时刻刻环绕在我们身边, 但是从时间的计量上看还只是宛如一个出生的小婴儿般, 具有发展潜力。面对其引发一些社会问题及产生一系列的风险, 以及未来可能的强人工智能时代的挑战, 法律既为科技提供包容性框架以释放创新潜力, 又通过底线规则防范技术风险, 加强人工智能相关法律、伦理、社会问题的研究具有重大的价值和意义。

参考文献

- [1] [以]尤瓦尔·赫拉利. 智人之上: 从石器时代到 AI 时代的信息网络简史[M]. 林俊宏, 译. 北京: 中信出版集团, 2024.
- [2] 刘宪权. 人工智能时代的刑事风险与刑法应对[J]. 法商研究, 2018, 35(1): 3-11.
- [3] 透视“AI 换脸”侵犯个人信息公益诉讼案: 换脸, 不是想换就换[EB/OL]. https://www.spp.gov.cn/spp/zdgz/202308/t20230823_625773.shtml, 2023-08-23.
- [4] 皮勇. 人工智能刑事法治的基本问题[J]. 比较法研究, 2018(5): 149-166.
- [5] 房慧颖. 涉人工智能犯罪刑法規制问题研究[D]: [硕士学位论文]. 上海: 华东政法大学, 2020.