Dispute Settlement 争议解决, 2025, 11(12), 11-17
Published Online December 2025 in Hans. https://www.hanspub.org/journal/ds
https://doi.org/10.12677/ds.2025.1112364

Hans汉斯

人工智能过失风险的刑法规制研究

——以新过失论为理论视角

庞惠子

北方工业大学文法学院，北京

收稿日期：2025年11月7日；录用日期：2025年11月26日；发布日期：2025年12月9日

摘　要

在人类社会迈入人工智能时代的背景下，人工智能技术在为生活提供便利的同时，其引发的侵权与过失风险问题亦随之凸显，对刑法的适应性调整提出了新的挑战。面对日益增多的人工智能法律争议，亟需对人工智能事故的责任主体界定与归责规则等核心问题予以规范，因此有必要引入新过失论理论，以保障科技的持续发展，避免因刑罚风险的不当干预而阻碍技术创新。基于此，应在客观注意义务与主观注意义务的内容构建上，进一步限缩过失犯罪的成立范围，并完善我国关于人工智能涉过失风险的法律规范体系。

关键词

人工智能，过失犯罪，新过失论，过失风险

Research on the Criminal Law Regulation of Artificial Intelligence's Negligence Risk

—From the Theoretical Perspective of the New Negligence Theory

Huizi Pang

School of Humanities and Law, North China University of Technology, Beijing

Received: November 7, 2025; accepted: November 26, 2025; published: December 9, 2025

Abstract

With human society entering the era of artificial intelligence, while AI technology brings convenience to life, the infringement and negligence risks it triggers have become increasingly prominent,

文章引用: 庞惠子. 人工智能过失风险的刑法规制研究[J]. 争议解决, 2025, 11(12): 11-17.
DOI: 10.12677/ds.2025.1112364

posing new challenges to the adaptive adjustment of criminal law. Faced with the growing number of AI-related legal disputes, it is urgent to regulate core issues such as the identification of responsible subjects and liability attribution rules for AI accidents. Therefore, it is necessary to introduce the new negligence theory to ensure the sustainable development of science and technology and avoid hindering technological innovation due to improper intervention of criminal law risks. Based on this, we should further restrict the scope of negligent crimes in constructing the contents of objective and subjective duties of care, and improve China's legal normative system concerning AI-related negligence risks.

Keywords

Artificial Intelligence, Negligent Crime, New Negligence Theory, Negligence Risk

1. 人工智能过失风险分析

随着时代的发展，人类技术的进步，人类社会已经经历了以蒸汽技术为基础的第一次技术革命、以电力技术为本体的第二次技术革命以及以计算机及其信息技术为标志的第三次技术革命，当前进入了大数据、云计算等并驾齐驱的第四次技术革命，即人工智能时代[1]。人工智能时代的来临，在一定程度上给人类生活社会提供了高效化、便利化的同时，在研发、制造、使用人工智能的过程中可能会带来一系列刑事风险，如危害国家安全和公共安全、侵犯公民人身权利和财产权利、破坏经济秩序和社会秩序等。而现在各国现行的法律尚不能够对人工智能可能带来的刑事风险形成完备的应对规则系统，故本文将结合过失犯罪的相关学说，针对人工智能能否成为犯罪主体，是否有责任能力，如何构成过失犯罪，对其犯罪行为责任如何认定等一系列的问题进行研究讨论。

人工智能体的功能已逐渐完善，逐渐从弱人工智能向强人工智能进步，即人工智能逐渐从协助人类的工具，演变为已经能够替代人类工作的存在，伴侣机器人代替妻子的角色，自动驾驶汽车替代驾驶员的角色[2]。人工智能在当前社会的广泛应用在给人类提供便利的同时，也带来许多风险，例如在教育领域，ChatGPT 等生成式人工智能逐渐成为教师教学和学生学习的主要工具，既能够帮助教师编写教学大纲，也能帮助学生完成作业，据相关报道，美国超八成的学生在运用 ChatGPT 完成作业，但是过分地依赖人工智能带给自己的便利，从而避免使用自己的思考能力，未尝不是一种欺骗自己的行为；在交通领域，现代汽车的自动驾驶功能，给驾驶员一种新型的轻松开车方式，在实际驾驶过程中也存在许多难以预见的问题而难以避免交通事故的发生，2016 年 1 月 20 日，京港澳高速河北邯郸段发生全球第一起自动驾驶车祸致死事故，随后几年在全球各地陆续发生数起类似自动驾驶车祸事故，一定程度上危害人类的社会生活安全；在知识产权领域，AI 画师通过 AI 完成的作品屡次被发现抄袭真人画师的元素，将数位画师的作品元素进行整合而完成一部新的“作品”，频繁涉及侵权等问题，除此之外在医疗器械，食品安全等领域都存在个人信息泄露、知识产权侵权及危害公众安全等触犯刑法风险的问题[3]。

而关于人工智能制造出的风险，主要包括内生性风险和外来性风险。前者是指这种风险源于人工智能本身，而后者是指人工智能程序被人为破坏或黑客侵入，导致出现危害人类的后果，这种风险源于人工智能外部[4]。

2. 人工智能的主体困境及解决

关于人工智能的过失风险认定，在以下几个方面存在困难与障碍：首先是对人工智能过失风险中责任主体的认定存在一定困难，究竟谁是能够承担责任的主体？人工智能体本身能够成为刑法上的责任主体吗？以及人工智能与过失犯罪中各学说中提到的客观注意义务与结果预见义务的认定标准上存在许多障碍，危险的分配影响着注意义务的分配，从而影响责任的归属[5]，下面将对以上问题进行分析讨论，并更深入的分析过失理论各发展阶段是否能够适用人工智能过失风险的认定。

2.1. 主体困境

面对逐渐增多的涉及人工智能的刑法相关问题，各国均对人工智能是否拥有主体地位而进行讨论与研究。如前文所说，人工智能所制造出的风险分为其本身的过失刑事风险和人类利用人工智能创造的过失刑事风险，对于后者来说，人工智能体本身就是研发者或者使用者的犯罪工具，与通常的犯罪没有什么区别，可以直接适用刑法的有关规定予以定罪处罚；而对于前者来说，人工智能体本身的过失而导致的犯罪，其主体难以确定，有关人工智能的讨论，主要集中在民商法、法理法史等领域，专门从行政法、刑法角度探讨的论文屈指可数。

根据美国 Narrative Science (《叙事科学》)预测，未来 15 年将有 90%的新闻稿件由机器人完成，大量的美术、音乐等艺术作品也将出自人工智能创作，在民商法、知识产权领域的部分支持确立主体者认为，人工智能体应当被确立主体地位，以应对日渐庞大的潜在风险问题，甚至在有些国家，某些情况下已经具备了民事主体身份，如沙特在 2017 年赋予了智能机器人索菲亚以公民身份[6]，而同样有反对者认为，人工智能体的作品只是数据逻辑运算的结果，没有感情描述和表达，与蕴含情感、理想和思想的人类作品有本质区别，难以被认定为主体。

在刑法领域中，随着人工智能体逐渐取代人类，当人工智能产品致害事故发生后，不仅要考虑犯罪主体是否是使用者，还要考虑研发设计等主体是否实施了相关危害行为。因此有关人工智能过失风险谁能够作为刑事责任主体、人工智能体本身是否能够成为刑法上的责任主体、与人工智能过失风险所涉及的刑法结构体系、刑事责任能力等问题，应当进行进一步深入的研究才能够回答。

2.2. 主体困境的解决

对于人工智能体是否应该具有主体资格的问题，有学者提出“机器人主体地位的趋势已经形成，我们应当正视这一社会伦理现象，考虑赋予其适当的法律资格与地位，制定并完善相关的法律制度。”[7]

但实际上，首先人工智能尚且没有认识行为规范含义的能力：就人工智能来说，它对外界信息的获取与认识主要是通过“电子眼”、“电子耳”来实现事实层面的认识[8]，并且现阶段的人工智能还难以实现自由选择合法行为，如果有设计者在自动驾驶的程序中写入“撞击行人”等类似的指令，作为自动驾驶的人工智能技术是没有办法对抗指令设计者的命令的，说明人工智能体不具有真正的辨认和控制能力；其次人工智能体与单位、法人或其他组织不具有类比性，将单位、法人或其他组织拟制为法律主体符合法律主体利己性的前提假设，人工智能体服务于人类社会的存在目的体现了其工具性的根本价值，而工具性价值有决定了其具有天然的利他属性[1]；最后，对人工智能体进行刑罚处罚达不到刑法的目的，根据费尔巴哈的“心理强制说”，如果在法律上规定有犯罪必有刑罚，刑事主体就会为了避免刑罚所产生的“痛苦”，而选择因抑制犯罪行为所导致的“小的不愉快”[9]，而人工智能体无法基于其内心感知来比较“痛苦”与“小的不愉快”从而做出选择，他既不会“害怕”刑罚也不会因刑罚而“痛苦”。综上所述，人工智能体缺乏承担罪责的能力，也无法理解惩罚的意义，惩罚人工智能体本身起不到一般防御和特殊防御的刑罚预防作用[10]。

根据以上内容，笔者认为人工智能体尚且不能够作为刑法中具有刑事责任的主体。由于否定了人工智能体的刑事主体地位，对于人工智能的过失风险将从人工智能体的研发者、制造者以及使用者的过失责任进行规制，下面将对过失犯罪的各理论学说寻找解决人工智能归责问题的途径。

3. 人工智能过失刑事风险认定困境与解决

3.1. 有关客观注意义务的困境

以违反注意义务为核心解释过失被认为过失脱离罪过形式判断的表征，就此意义而言，违反客观注意义务就是过失实行行为的问题。注意义务的概念分为“注意”和“义务”两部分，“注意”是一种心理活动事实，是与行为这一外在概念所对应的内在方面，“义务”则是一种“主体性”和“规范性”相结合的产物，但是更加侧重一种客观的规范。而当“注意”与“义务”相结合时，便构成了过失犯的构成要件，即客观上没有履行相应的义务和主观上没有尽到相应的注意[11]。客观注意义务是指在不法阶段对一般人提出的抽象要求，而对于人工智能体来说，有学者提出，由于人工智能深度学习的应用，智能机器人的行为将是非线性的，实际上是不可预测的[12]，既然其不能够作为刑法中具有刑事责任的主体，那么在其产品设计与使用环节的规范性标准的缺失，就对客观层面刑法注意义务的判定带来难题[2]。

过失犯的实行行为就是违反客观注意义务的行为，这种客观注意义务来源于法律规范或社会条理的要求，在具体内容上包含着客观的预见义务与客观的结果回避义务。2021 年 9 月 1 日由交通运输部等部门联合印发的《智能网联汽车道路测试与示范应用管理规范(试行)》对自动驾驶汽车道路测试与示范应用管理机构、车辆、驾驶人等进行了规定，却并不能满足自动驾驶汽车交通肇事后刑事规则的需求[2]。在现阶段，我国关于自动驾驶汽车在公共道路行驶的规定缺失，《道路交通安全法》(2021 年修正)未明确自动驾驶汽车的法律地位，与自动驾驶汽车相关的驾驶要求仍游离在法律之外。违反客观注意义务须以具有客观注意义务之履行可能性为前提，即行为人应当具有法益侵害危险的预见可能性或法益侵害结果之回避可能性[13]。由于人工智能体的主体地位已经被否定，亦无法依据在相关行政法律规范中的注意义务，对人工智能产品的设计、生产、制造以及使用主体所具有的刑法上客观注意义务进行判断，从而对人工智能产品致害所涉的过失刑事责任认定产生严重阻碍。

3.2. 引入新过失论解决客观注意义务

通过上文的分析已将人工智能体的主体地位予以否定，故此人工智能导致的过失风险应当由相应的自然人，如研发者、设计者或使用者承担责任，根据当前人工智能的发展现状，以及关于注意义务的阻碍，对解决其所涉过失风险的问题或许可以通过引用过失理论的内容解决，下面将对过失理论的各种学说进行分析，并判断何种学说更为适合解决有关人工智能过失风险的问题。

刑法第 15 条第 1 款规定：“应当预见自己的行为可能发生危害社会的结果，因为疏忽大意而没有预见，或者已经预见而轻信能够避免，以致发生这种结果的，是过失犯罪。”据此，过失犯罪是指过失实施的犯罪。犯罪过失，则是指应当预见自己的行为可能发生危害社会的结果，因为疏忽大意而没有预见或者已经预见而轻信能够避免，以致发生这种结果的心理态度。

在理论中，关于过失论大致经历了旧过失论、新过失论以及新新过失论的发展。旧过失论是以“预见可能性作为过失犯的责任基础，也即，只要发生了结果，在行为的构成要件符合性、违法性上与故意犯并无区别，两者的不同仅仅为：故意是对结果的认识、预见，而过失犯是对结果的认识可能性、预见可能性。”[14]主张结果预见义务说，单纯考虑行为人的结果预见可能性。到了 20 世纪初，大量交通事故的出现，新过失论粉墨登场，其核心观点为：即使对结果具有预见可能性，但如果行为人履行了结果回避义务，就不成立过失犯[15]，主张结果回避义务说，不能追究能够预见危害结果，并且履行了结果回

避义务的行为人的过失刑事责任。而后随着公害事件的不断出现，无论是依赖于旧过失论的“结果预见可能性”还是新过失论的“结果回避义务”，都无法有效地遏制此类犯罪的蔓延，由此新新过失论应运而生，其认为，只要行为人对危害结果的发生存在不安感、危惧感就够了，而不要求行为人具有具体的预见可能性。

如表 1，相较于只考虑行为人的结果预见可能性，以旧过失论的标准来认定过失犯罪的做法，扩大了过失犯的处罚范围且与社会公平正义、经济效率理念相悖，旧过失论由于过度忽视过失“行为”的层面，几乎等同于仅追究结果责任[17]；新新过失论则是要求只要对危险发生具有危惧感，就产生了消除这种危惧感的结果回避义务，没有履行这种义务而致使结果发生，就违反了注意义务而应付过失责任[18]，而新过失论在解决人工智能过失刑事风险方面展现了其他学说所不具备的理论优势。

Table 1. Comparison of three different negligence theories [16]
表 1. 三种不同的过失理论的比较[16]

	旧过失论	新过失论	新新过失论
注意义务	以预见义务为核心	以回避义务为核心	以回避义务为核心
预见可能性	具体	具体	不安感
处罚范围	——	限定	扩大

首先，新过失论能够保证科技的发展，以不至于使科技发展止步于刑法对风险的不正当干预，若在科技创新领域贯彻旧过失论的理论，就会出现这样的局面：只要科学家对新的科技创造成果有结果预见可能性，就会成立过失犯罪的话，大部分科研人员可能就会拘泥于现有的刑事政策，而不敢进一步的进行技术创新，任何科技发展都伴随着风险，当科研人员都不敢去进一步研究的时候，我国科学技术就会面临停滞不前的局面；若贯彻新新过失论的理论，即使是不能具体的预想到会有什么样的危险发生，只要有了危惧感，就应对过失风险承担责任，尽到注意义务，日本学者藤木英雄认为：在科学技术的无限发展和扩大的同时，未知的危险对我们人类的威胁越来越大，要想利用科学技术，就必须负有社会性的责任[19]，但是这对于无论是研发者还是使用者的要求都太过苛刻，并不利于科研者进行大胆的尝试，而新过失论在限制过失犯的处罚范围上比旧过失论更为完善，其为了限制处罚的范围，也需要对其回避义务以及过失行为进行实质的判断，在主观上需要行为人的具体预见为基础[15]，在限制过失犯的处罚范围上并不故步自封，为了避免社会陷入停滞而特别强调允许的危险理论，正因为如此，对社会有益的、不可缺少的行为，即便其自身是侵害法益的危险行为，但在一定范围内也承认这种危险行为的合法性[20]，如此一来，对科研人员的限制变小，能够有利于科研领域的发展。

其次，新过失论与其他理论相比，其对处罚范围进行进一步的限定。当人工智能发生事故之后，若按照旧过失论的观点，一旦发生事故，那么该技术的研发者、制造者、使用者都必然违反了注意义务，会不当的扩大处罚范围[21]，反观新过失论对过失犯的判定在主客观两个方面都有所要求，当研发者和制造者在研发和制造的过程中严格遵守业务规范，则不会违反结果回避义务，就不构成过失犯罪，通过新过失论的双重限定，能够限定过失犯的成立范围，既不会阻碍科研人员对新科技的持续研发，也能够保证他们不会因为别的原因而发生事故而承担莫须有的罪名。

综上，相较旧过失论和新新过失论，新过失论更适合解决人工智能涉及过失风险方面的问题，在对人工智能的过失犯罪上可以更合理的提出修改法律规范，对人工智能体背后的自然人进行归责。

3.3. 解决措施

在对人工智能体本身的主体进行否定后，意味着不能够对人工智能本身进行处罚，那么迎面而来的问题便是对其背后的研发者、制造者和使用者的自然人在何种程度上成立过失犯罪进行探讨。下面将以医疗领域与自动驾驶领域对自然人的过失认定为例，结合主观过失义务和客观过失义务，阐述部分解决措施。

首先在医疗领域，一方面对 AI 医疗器械进入临床诊断的技术标准进行更严格的要求，AI 诊断相较于自然人拥有三大特点：第一，其诊断水平已接近最高水平的医务人员；第二，其拥有深度学习和自我迭代的能力；第三，其具有自动化的特点[22]。我国《侵权责任法》规定，医务人员的注意义务是以当时的医疗水平为限，所以为了最大限度的保障患者的权益，对 AI 诊断主体要求应当高于“当时的医疗水平”[23]。另一方面，我国的医务人员在就业前都需要经过考核并取得医师执业资格证，而在 AI 医疗器械广泛运用的现在，对医务人员的要求不应只限于取得医师执业资格证，还应对其能否规范使用 AI 医疗器械进行考核，重视相关的医疗法规范所确立的禁止性规范，在使用的过程中医疗人员能否对相关的风险具有预见能力，注意采取结果回避措施，才能够避免违反客观注意义务。

另外在自动驾驶领域，只要自动驾驶过程中的车辆，在遵循交通法规的情形下行驶，即使发生了危害结果，其自动驾驶的车辆也没有违反客观注意义务，则不需要由行为人承担刑事责任；若没有遵循法规行驶，则要分开讨论研发者和驾驶人是否采取了结果回避措施，对自动驾驶的车辆出现误判后有无采取回避行为，进行责任的认定[24]。另外在主观注意义务的认定上，要根据其个人的具体情况判断其是否具有预见结果的能力，如果具有预见的能力，则成立主观的注意义务[15]，当驾驶人却无违法注意义务的时候，再去追究该技术的研发者和制造者在研发过程与制造过程中是否有违反结果预见可能性或结果回避义务的注意义务，完成归责问题的归属。

关于人工智能过失犯罪，专门的法律法规、技术规范以及规章制度等几乎是确定注意义务的唯一来源，是追究刑事责任的决定性因素[25]。没有刑法的强有力保护，既不利于社会秩序的维护，也不利于人工智能相关领域的科研发展，刑法对人工智能产品领域研发设计主体、生产销售主体以及使用操作主体的过失行为进行规制，目的是规范其自身行为，保障社会大众利益[2]。

4. 结语

人工智能的时代给人类带来许多便利这一点不可置否，但是也不可避免地带来一系列法律风险，给我们现在的法律适用提起许多挑战。通过本文的研究，对人工智能的风险进行评析，针对人工智能当前难以通过现行法律规范认定的困境，否定了人工智能体本身的主体资格，应当对其背后的自然人进行追责；通过对比各阶段的过失理论，最终认可新过失论的主张，并以此为基础分析人工智能过失风险所牵扯的各种问题；最后在对人工智能体过失风险的构成要件上，依据新过失论的主张提出解决办法。综上所述，在对人工智能所涉及的法律风险问题仍需更多的理论支撑和解决方案，在人工智能时代中由于人工智能的使用，对人类社会来说将面临更多新的难题，于刑法领域来说还需要更长久的改革之路要走。

参考文献

[1] 许钟灵, 吴情树. 人工智能体过失刑事风险的因应[J]. 法治社会, 2020(6): 93-103.

[2] 姚瑶. 人工智能时代过失犯理论的挑战与应对——以自动驾驶汽车交通肇事为例[J]. 浙江社会科学, 2022(12): 59-67+157.

[3] 储陈城, 魏培林. 生成式人工智能犯罪中研发者刑事责任的认定——以 ChatGPT 为例[J]. 重庆理工大学学报(社会科学), 2023, 37(9): 103-113.

[4] 赵运锋. 人工智能不应作为刑事责任主体[J]. 检察风云, 2019(6): 28.

[5] 张明诚, 吴勃. 自动驾驶交通肇事中使用者注意义务及过失犯罪认定[J]. 合肥工业大学学报(社会科学版), 2021, 35(2): 53-59.
[6] 王乐兵. 自动驾驶汽车的缺陷及其产品责任[J]. 清华法学, 2020, 14(2): 93-112.
[7] 刘宪权. 人工智能时代机器人行为道德伦理与刑法规制[J]. 比较法研究, 2018(4): 40-54.
[8] 刘宪权. 人工智能时代的“内忧” “外患”与刑事责任[J]. 东方法学, 2018(1): 134-142.
[9] 张明楷. 刑法的基本立场[M]. 修订版. 北京: 商务印书馆, 2019: 17.
[10] 廖兴存. 无人驾驶汽车交通肇事过失刑事责任论纲[J]. 湖南社会科学, 2021(3): 122-130.
[11] 纪康. 注意义务违反与过失不法的规范限缩[D]: [博士学位论文]. 上海: 华东政法大学, 2021.
[12] [美]瑞恩·卡洛, 迈克尔·弗兰金. 人工智能与法律的对话[M]. 陈吉栋, 译. 上海: 上海人民出版社, 2018: 8.
[13] 杜宇, 毕海燕. 论过失犯实行行为的识别[J]. 清华法学, 2023, 17(1): 116-136.
[14] [日]西田典之. 刑法总论[M]. 第 2 版. 沈阳: 弘文堂, 2010: 258.
[15] 姚万勤. 新过失论与人工智能过失刑事风险的规制[J]. 法治研究, 2019(4): 98-107.
[16] [日]前田雅英. 刑法总论讲义[M]. 第 6 版. 曾文科, 译. 北京: 北京大学出版社, 2018: 184.
[17] 王霖. 自动驾驶场景下过失犯归责困境巡检与路径选择——以规范归责模式为视角[J]. 河北法学, 2020, 38(3): 93-108.
[18] 陈兴良. 过失犯论的法理展开[J]. 华东政法大学学报, 2012(4): 30-47.
[19] [日]藤木英雄. 公害犯罪[M]. 丛选功, 等, 译. 北京: 中国政法大学出版社, 1992: 62.
[20] 黎宏. 日本刑法精义[M]. 第 2 版. 北京: 法律出版社, 2008: 210.
[21] 张依楠. 自动驾驶交通肇事中研发者过失犯罪的责任限缩研究[J]. 重庆广播电视大学学报, 2021, 33(2): 48-54.
[22] 李兴臣. 人工智能医疗服务的法律责任[J]. 医学与法学, 2018, 10(4): 8-12.
[23] 郑志峰. 诊疗人工智能的医疗损害责任[J]. 中国法学, 2023(1): 203-221.
[24] 龙敏. 自动驾驶交通肇事刑事责任的认定与分配[J]. 华东政法大学学报, 2018, 21(6): 77-82.
[25] 彭文华. 自动驾驶车辆犯罪的注意义务[J]. 政治与法律, 2018(5): 86-99.