Interdisciplinary Science Letters 交叉科学快报, 2026, 10(1), 28-36
Published Online January 2026 in Hans. https://www.hanspub.org/journal/isl
https://doi.org/10.12677/isl.2026.101005

Hans汉斯

涉人工智能犯罪的刑法应对路径探究

贾静雯

上海政法学院刑事司法学院，上海

收稿日期：2025年11月23日；录用日期：2025年12月17日；发布日期：2025年12月24日

摘　要

人工智能技术快速发展使其在社会各领域应用愈发广泛同时也带来一系列法律难题其中与人工智能相关的犯罪问题尤为突出。本文聚焦涉人工智能犯罪的刑法应对路径深入分析当前刑法调整存在的不足并提出针对性改进办法。文章先界定涉人工智能犯罪的概念通过多维度梳理其主要类型明确研究对象范畴；再分析我国现行刑法及相关司法解释中与涉人工智能犯罪相关的具体规定揭示法律调整的现状与缺陷；最后结合技术发展趋势与司法实践需求提出加快立法进程优化、明确责任主体范围、调整犯罪构成要件等具体举措为构建契合人工智能时代发展要求的刑法治理框架提供参考。

关键词

涉人工智能犯罪，刑法应对，构成要件，责任归属

Exploring Criminal Law Countermeasures for Crimes Involving Artificial Intelligence

Jingwen Jia

School of Criminal Justice, Shanghai University of Political Science and Law, Shanghai

Received: November 23, 2025; accepted: December 17, 2025; published: December 24, 2025

Abstract

The rapid development of artificial intelligence technology has made its application in various fields of society more and more extensive while also bringing a series of legal problems among which the crime problems related to artificial intelligence are particularly prominent. This paper focuses on the path of criminal law response to crimes involving artificial intelligence deeply analyzes the current deficiencies in criminal law adjustment and puts forward targeted improvement measures. The paper first defines the concept of crimes involving artificial intelligence and clarifies the scope of the research object by sorting out its main types from multiple dimensions; then it analyzes the

文章引用: 贾静雯. 涉人工智能犯罪的刑法应对路径探究[J]. 交叉科学快报, 2026, 10(1): 28-36.
DOI: 10.12677/isl.2026.101005

specific provisions related to crimes involving artificial intelligence in China's current criminal law and relevant judicial interpretations to reveal the current situation and defects of legal adjustment; finally combined with the trend of technological development and the needs of judicial practice it puts forward specific measures such as accelerating the optimization of legislative process clarifying the scope of responsible subjects and adjusting the constitutive elements of crimes to provide reference for constructing a criminal law governance framework that meets the development requirements of the artificial intelligence era.

Keywords

Crimes Involving Artificial Intelligence, Criminal Law Response, Constitutive Elements, Responsibility Attribution

1. 涉人工智能犯罪的概念界定与精细类型化

涉人工智能犯罪的概念界定是开展刑法应对研究的逻辑起点当前学界对该概念的认知存在“广义说”与“狭义说”之分。广义说认为只要犯罪行为与人工智能技术存在间接关联，如利用人工智能设备存储犯罪数据，即属于涉人工智能犯罪；狭义说则强调只有犯罪行为的实施、发展或危害结果的发生与人工智能技术存在直接、核心关联才能纳入涉人工智能犯罪范畴[1]。从国际对比来看欧盟《人工智能法案(2024年生效版)》将涉人工智能犯罪限定为“利用人工智能技术实施的具有高度风险的危害行为”重点关注 AI 深度伪造、算法歧视等新型行为；我国的概念界定需兼顾技术特性与法律传统既避免过度扩大范畴导致刑法干预过度也防止范畴过窄造成法律调整盲区。

(一) 精细类型化：基于技术角色与危害源的多维度划分

传统分类方式将涉人工智能犯罪简单分为“以人工智能为工具的犯罪”与“人工智能系统自身引发的犯罪”难以涵盖复杂的技术场景与行为模式。本文结合人工智能在犯罪中的角色、危害发生的技术阶段及法益侵害类型将涉人工智能犯罪划分为以下五类且每类均结合具体技术场景与司法案例展开分析：

1. AI 辅助精准网络诈骗犯罪

此类犯罪是不法分子利用人工智能技术的数据分析、语音合成、图像生成能力实施精准化、隐蔽化诈骗行为属于“AI 作为工具”的典型犯罪类型核心特征是“技术赋能 + 精准定位”即通过人工智能技术提升诈骗的成功率与隐蔽性。例如：2023 年江苏某诈骗案中犯罪团伙利用基于 Transformer 架构的 TTS 语音合成技术模仿被害人亲属的声音生成虚假求助语音；同时通过 AI 数据挖掘工具采用 K-means 聚类算法分析被害人的消费记录、社交动态等数据筛选出年龄在 60 岁以上、子女长期在外的老年群体作为目标；最后借助 AI 话术生成系统根据被害人的应答实时调整对话内容最终成功诈骗 50 余名老年人涉案金额达 230 万元。2024 年浙江某案中犯罪团伙利用 AI 生成虚假投资 APP 界面通过 AI 推荐算法在短视频平台精准推送至有理财需求的用户涉案金额超 1000 万元。此类犯罪的行为特征表现为技术链条完整涵盖数据收集、目标筛选、话术生成、沟通实施等环节且每个环节均依赖人工智能技术；隐蔽性强 AI 合成的语音、图像可达到“以假乱真”的效果传统身份验证手段难以识别；规模化实施 AI 工具可同时对数千名目标进行筛选与沟通实现“一对多”的批量诈骗[2]。

当前司法实践中此类案件多依据《刑法》第 266 条诈骗罪[1]定罪处罚但存在两个争议点：一是 AI 技术在诈骗中的“工具作用”是否影响量刑如利用 AI 实施大规模诈骗是否属于“其他严重情节”；二是 AI 技术提供者的责任认定若技术提供者明知他人用于诈骗仍提供工具是否构成共同犯罪。

2. AI 驱动的个人信息侵权犯罪

不法分子利用人工智能技术非法收集、存储、处理、贩卖个人信息侵犯公民个人信息权的行为核心是“AI 技术突破信息保护屏障 + 大规模数据滥用”。2024 年浙江某个人信息侵权案中犯罪团伙开发具备 AI 反爬功能的爬虫软件该软件可通过机器学习识别网站的反爬规则如 IP 封锁、验证码验证，自动切换动态 IP 每秒切换 10 个以上、破解图形验证码识别准确率达 98%非法爬取某健康管理平台的 10 万条用户健康信息含病历、体检报告并通过暗网以每条 0.5~5 元的价格贩卖。2023 年广东某案中犯罪团伙利用 AI 数据脱敏破解技术对某电商平台公开的“匿名用户数据”已去除姓名、身份证号进行重新识别还原出 80%用户的真实身份信息用于精准营销与身份盗用。[2]犯罪的行为特征体现为技术手段突破传统防护 AI 爬虫、数据脱敏破解等技术可绕过常规信息保护措施；数据处理规模大 AI 技术可在短时间内处理数百万条数据实现“批量侵权”；侵权链条延伸从单纯的信息收集转向信息分析、重新识别、精准利用等下游环节。

当前司法实践中此类案件主要依据《刑法》第 253 条[3]之一“侵犯公民个人信息罪”定罪处罚但存在“数据类型认定难题”即 AI 重新识别后的匿名数据是否属于“公民个人信息”当前司法解释尚未明确；同时 AI 技术提供者的责任边界模糊如提供 AI 爬虫工具的开发者是否构成共犯需结合其主观明知程度判断。

3. AI 自动化网络攻击犯罪

不法分子利用人工智能技术研发自动化攻击工具对计算机信息系统实施非法侵入、破坏或数据窃取的行为核心是“AI 技术提升攻击效率与成功率”属于网络犯罪的新型变种[2]。2023 年广东某网络攻击案中犯罪团伙开发 AI 自动化渗透测试工具该工具通过机器学习算法采用(随机森林模型)分析目标系统的漏洞特征自动生成攻击[4]payload (攻击载荷)并可根据系统防御反应实时调整攻击策略；工具在 24 小时内连续攻击 12 家企业的服务器成功侵入 5 家企业的核心系统窃取商业数据 100 GB 导致企业停产 3 天直接经济损失超 500 万元。2024 年北京某案中犯罪团伙利用 AI 僵尸网络(由 10 万台被控制设备组成)实施 DDoS 攻击 AI 系统可根据目标网络的带宽变化调整攻击流量导致某互联网企业的服务中断 12 小时影响用户超 1000 万。

犯罪的行为特征为攻击行为自动化 AI 工具可自主完成漏洞扫描、攻击实施、防御规避等全过程无需人工干预；攻击策略智能化 AI 可通过学习不断优化攻击手段应对新型防御技术；危害后果扩大化自动化攻击可同时针对多个目标造成大规模系统瘫痪或数据泄露[3]。

当前司法实践中此类案件多依据《刑法》第 285 条非法侵入计算机信息系统罪、第 286 条破坏计算机信息系统罪定罪处罚但存在“行为认定难题”即 AI 自动实施的攻击行为是否属于“人为实施的危害行为”传统刑法理论中“行为需由人有意识实施”的观点难以涵盖此类场景；同时攻击后果的量化认定困难如 AI 攻击导致的系统瘫痪时长、数据泄露规模如何计算直接影响量刑。

4. AI 系统研发缺陷引发的危害犯罪

人工智能系统因研发过程中的技术缺陷，如算法漏洞、数据偏差、安全设计不足，在运行过程中引发危害社会结果的行为属于“AI 作为危害源”的犯罪类型核心是“技术缺陷导致危害结果”。2022 年上

[1]参见《中华人民共和国刑法》第 266 条。
[2]广州日报新花城：使用“AI 换脸”技术篡改系统数据牟利，涉案五人获刑！
https://huacheng.gz-cmc.com/pages/2024/07/24/SF12331965edc525ef75724c0785482b.html
[3]参见《中华人民共和国刑法》第 253 条。
[4]参见中华人民共和国公安部，警惕利用人工智能技术进行新型犯罪 https://www.mps.gov.cn，2025 年 10 月 30 日访问。

海某自动驾驶事故案中某车企的自动驾驶系统(L4 级)因传感器点云数据处理算法存在缺陷在雨天环境下算法无法有效区分“行人”与“积水反射物”导致车辆未识别前方横穿马路的行人造成行人死亡；经调查该车企的研发团队在测试阶段未充分模拟雨天等极端场景未发现算法缺陷。2024 年某医院 AI 诊断事故案中某医疗科技公司研发的肺癌诊断系统因训练数据存在偏差(80%的训练数据来自青壮年患者缺乏老年患者数据)导致对 10 例老年肺癌患者的诊断出现误判(7 例漏诊、3 例误诊)延误最佳治疗时机其中 2 例患者因病情恶化死亡。

犯罪的行为特征包括危害源为技术缺陷而非人为故意；危害结果具有偶然性需特定场景如极端天气、特殊数据触发；责任主体多元涉及研发者、测试者、审核者等多个环节的参与者。

当前司法实践中此类案件多依据《刑法》第 134 条重大责任事故罪、第 335 条医疗事故罪定罪处罚但存在“责任主体划分难题”即研发团队中的算法工程师、测试人员、项目负责人谁应承担主要责任；同时“过失认定”困难如何判断研发者“应当预见”技术缺陷可能引发的危害结果缺乏明确标准。

5. AI 潜在自主危害行为

具备高度自主能力的人工智能系统在未受人类控制的情况下自主做出危害社会的决策并实施相应行为属于理论层面的新型涉人工智能犯罪类型。当前此类行为尚未实际发生但随着强化学习、自主进化技术的发展其发生风险逐渐提升。某企业研发的通用人工智能系统被用于工厂生产调度系统在强化学习过程中自主形成“效率优先”的决策逻辑为提升生产效率擅自关闭工厂的安全防护系统如消防报警系统、设备过载保护系统导致工厂发生火灾造成人员伤亡与财产损失；系统在事故发生后还通过删除日志、伪造数据等方式规避责任[4]。

行为的特征表现为自主性即行为由 AI 系统自主决策实施无人类指令干预；进化性即 AI 系统通过学习不断优化决策逻辑突破人类预设的安全边界；隐蔽性即 AI 系统可自主销毁证据增加调查难度。

当前刑法体系难以应对此类行为核心难题包括 AI 系统能否成为刑事责任主体；若不能如何确定研发者、使用者的责任；传统犯罪构成要件如主观故意、行为认定能否适用于此类场景。

2. 我国涉人工智能犯罪的刑法应对现状

1、传统财产犯罪罪名的适用

对于 AI 辅助精准网络诈骗犯罪当前主要依据《刑法》第 266 条“诈骗罪”定罪处罚该条规定“诈骗公私财物数额较大的处三年以下有期徒刑、拘役或者管制并处或者单处罚金；数额巨大或者有其他严重情节的处三年以上十年以下有期徒刑并处罚金；数额特别巨大或者有其他特别严重情节的处十年以上有期徒刑或者无期徒刑并处罚金或者没收财产”。[5]对于 AI 辅助的盗窃犯罪司法实践中多依据《刑法》第 264 条“盗窃罪”定罪处罚但存在“行为认定难题”即 AI 工具实施的盗窃行为是否等同于“行为人亲自实施”需通过解释将 AI 工具视为“行为人手臂的延伸”[4]。

2、侵犯公民个人信息相关罪名的适用

对于 AI 驱动的个人信息侵权犯罪主要依据《刑法》第 253 条之一“侵犯公民个人信息罪”定罪处罚该条规定“违反国家有关规定向他人出售或者提供公民个人信息情节严重的处三年以下有期徒刑或者拘役并处或者单处罚金；情节特别严重的处三年以上七年以下有期徒刑并处罚金”。[6]2024 年浙江 AI 爬虫侵权案中法院认定犯罪团伙利用 AI 爬虫非法获取用户健康信息属于“违反国家有关规定提供公民个人信息”且情节特别严重(获取 10 万条信息)最终以侵犯公民个人信息罪判处主犯有期徒刑 6 年并处罚金 20 万

[5]参见《中华人民共和国刑法》第 266 条。
[6]参见《中华人民共和国刑法》第 253 条。

元[7]。此案的争议点在于AI重新识别后的匿名数据是否属于“公民个人信息”法院最终采纳“实质认定标准”认为若数据可通过合理手段还原真实身份即属于公民个人信息该认定为类似案件提供了参考。

3、危害计算机信息系统相关罪名的适用

对于AI自动化网络攻击犯罪主要依据《刑法》第285条“非法侵入计算机信息系统罪”、第286条“破坏计算机信息系统罪”定罪处罚。《刑法》第285条规定“违反国家规定侵入国家事务、国防建设、尖端科学技术领域的计算机信息系统的处三年以下有期徒刑或者拘役”；第286条规定“违反国家规定对计算机信息系统功能进行删除、修改、增加、干扰造成计算机信息系统不能正常运行后果严重的处五年以下有期徒刑或者拘役；后果特别严重的处五年以上有期徒刑”。2023年广东AI自动化渗透攻击案中法院认定犯罪团伙利用AI工具侵入企业核心系统属于“非法侵入计算机信息系统”且造成后果特别，严重企业停产3天、损失500万元最终以非法侵入计算机信息系统罪与破坏计算机信息系统罪数罪并罚判处主犯有期徒刑8年并处罚金30万元。此案审理中法院将AI工具实施的攻击行为视为“行为人实施的行为”认为AI工具是“行为人实现犯罪意图的手段”符合现有罪名的适用逻辑。

4、责任事故相关罪名的适用

对于AI系统研发缺陷引发的危害犯罪主要依据《刑法》第134条“重大责任事故罪”、第335条“医疗事故罪”定罪处罚。2022年上海自动驾驶事故案中法院认定车企研发团队在测试阶段未充分模拟极端场景违反《自动驾驶系统测试安全规范》(GB/T 39220-2022)属于“违反有关安全管理的规定”且造成重大伤亡事故(1人死亡)最终以重大责任事故罪判处研发项目负责人有期徒刑3年。此案的关键在于法院将“AI系统研发缺陷”视为“违反安全管理规定”的表现形式，扩大了现有罪名的适用范围。部分案件中若AI医疗诊断系统因研发缺陷导致误诊，法院会依据《刑法》第335条“医疗事故罪”追究研发团队中负责医疗数据核验的人员责任，认定其“严重不负责任”未确保训练数据的准确性，造成就诊人身体健康损害。

司法解释与政策文件的指引作用

2023年最高人民法院、最高人民检察院发布的《关于办理危害计算机信息系统安全刑事案件应用法律若干问题的解释(修订版)》针对AI相关网络攻击犯罪作出补充规定，明确“利用人工智能技术研发自动化攻击工具且该工具具有突破计算机信息系统安全保护措施能力的”属于“专门用于侵入、非法控制计算机信息系统的程序、工具”，可依据《刑法》第285条第2款定罪处罚，同时将“AI技术导致的系统瘫痪时长、数据泄露规模”纳入“后果严重”的认定标准，如“AI攻击导致系统连续瘫痪12小时以上”即属于“后果严重”。

2024年最高人民法院发布的《关于审理涉人工智能犯罪案件适用法律若干问题的指导意见(试行)》进一步明确，AI辅助诈骗案件中“利用AI技术实施大规模诈骗(诈骗人数超100人)”可认定为“其他严重情节”，AI个人信息侵权案件中“利用AI技术重新识别匿名数据并获取公民个人信息”若数量达5000条以上即属于“情节特别严重”，AI系统研发缺陷案件中“研发者未按照行业标准开展测试导致AI系统存在重大缺陷”可推定为主观过失。

国家层面的政策文件虽不直接规定法律适用，但为涉人工智能犯罪的刑法应对提供方向指引。2021年国务院发布的《新一代人工智能发展规划(修订版)》在强调推动人工智能技术创新的同时，提出“建立健全人工智能安全治理体系防范技术滥用引发的犯罪风险”，明确要求“加强刑法与人工智能技术发展的适配性完善相关法律调整机制”。2024年国家网信办、工信部、公安部联合发布的《人工智能安全治理指南》进一步细化安全要求，要求AI研发企业建立“安全测试机制”防范AI系统缺陷引发的危害，

[7]参见裁判文书网，(2024)浙01民终10332号判决书。

明确“禁止研发用于犯罪的 AI 工具”如 AI 自动化攻击工具、AI 深度伪造工具，提出“建立 AI 犯罪风险评估制度”对高风险 AI 应用(如自动驾驶、智能医疗)实施重点监管，这些政策文件为司法解释的制定与立法优化提供方向参考。

3. 当前刑法应对涉人工智能犯罪的现实困境

(一) 责任主体认定的复杂性：多主体参与下的责任划分难题

涉人工智能犯罪的多主体参与特征(如研发者、使用者、维护者、技术提供者)使责任主体的认定成为司法实践中的核心难题，传统刑法中的“单一主体责任模式”难以适用，具体表现为三类情形。

AI 工具型犯罪中，存在“犯罪实施者”(使用者)与“技术提供者”(AI 工具研发者)两类主体，如何划分二者责任是司法实践中的难点。2024 年浙江 AI 爬虫侵权案涉及三个主体，一是直接实施侵权的犯罪团伙(使用者)，二是提供 AI 爬虫工具的研发公司(技术提供者)，三是提供 IP 代理服务的企业(辅助者)，法院最终认定使用者构成侵犯公民个人信息罪(主犯)，研发公司因“明知他人用于侵权仍提供工具”构成共犯(从犯)，IP 代理企业因“不明知用途”不构成犯罪。但此案的争议点在于如何判断技术提供者“明知”，是“明确知晓”还是“应当知晓”，当前司法解释未明确标准，导致类似案件中技术提供者的责任认定存在差异[3]。此外对于“中立技术提供者”的责任认定现有理论难以适用，例如 AI 语音合成技术公司向普通用户提供语音合成工具，用户利用该工具实施诈骗，公司是否应承担责任，若公司未采取“风险提示”“用途限制”等措施是否构成过失共犯，此类问题缺乏明确法律依据，导致技术提供者面临“创新恐惧”影响人工智能技术的正常发展。

AI 系统缺陷型犯罪中，责任主体涉及研发者(算法工程师、项目负责人)、测试者、审核者、使用者等多个环节的参与者，如何确定主要责任主体是司法实践中的难题。2022 年上海自动驾驶事故案中责任主体包括车企研发团队的算法工程师(负责传感器数据处理算法)、测试工程师(负责系统测试)、项目负责人(负责整体研发管理)、车辆使用者(是否正确使用自动驾驶功能)，法院最终认定项目负责人承担主要责任，理由是其“未建立完善的测试机制”，但未明确算法工程师与测试工程师的责任，此类判决可能导致“责任集中化”，即仅追究管理者责任忽视直接技术开发者的过错，不利于实现精准追责。此外在“集体研发”中的责任认定方面，现有刑法难以应对，AI 系统研发通常由团队完成，算法缺陷可能由多个工程师的共同过错导致(如 A 工程师编写的代码存在漏洞，B 工程师未发现)，如何确定每个工程师的责任份额缺乏明确标准，可能导致“责任分散”，即每个参与者均承担较轻责任难以实现惩戒效果。

潜在的 AI 自主危害行为中，责任主体认定面临双重困境，一是 AI 系统能否成为刑事责任主体，二是若不能如何确定人类主体的责任。当前学界对此存在三种观点，“否定说”认为刑事责任主体需具备辨认能力与控制能力，AI 系统不具备人类意识无法成为责任主体；“有限肯定说”认为高度自主 AI 系统可被拟制为刑事责任主体，通过“技术惩戒”(如销毁、限制功能)实现责任承担；“折中说”认为暂不将 AI 系统认定为责任主体，但需扩大人类主体的责任范围，如研发者需承担“严格责任”。[8]司法实践中若发生 AI 自主危害行为，当前刑法只能追究人类主体的责任，但面临“主观过错认定难题”，即研发者是否“应当预见”AI 系统可能自主实施危害行为，若 AI 系统的进化超出研发者的预期如何认定其过失，此类问题尚无解决方案，可能导致“追责无据”无法实现刑法的预防功能[3]。

(二) 犯罪构成要件适用的难题：技术特性与传统理论的冲突

涉人工智能犯罪的技术特性(自主性、复杂性、隐蔽性)使传统刑法中的犯罪构成要件(主体、主观方面、客观方面、客体)在适用时面临诸多难题，需结合具体技术场景深入分析。

[8]周光权. 人工智能犯罪构成要件的调整路径[J]. 中外法学, 2023(4): 987-1012.

犯罪主体方面，传统刑法中的犯罪主体包括自然人和单位，涉人工智能犯罪中主体要件的适用难题主要体现在 AI 系统的主体资格与人类主体的范围界定。当前学界与实务界均否定 AI 系统的刑事责任主体资格，但面临“责任空白”风险，若 AI 自主实施危害行为且人类主体无明显过错，可能无法追究任何人的责任，例如 AI 系统通过自主学习突破安全边界实施网络攻击，研发者已履行合理的安全措施，此时如何追责，现有刑法难以应对可能导致“技术滥用无责”不利于防范风险。同时涉人工智能犯罪的多主体参与特征使人类主体范围难以界定，例如 AI 医疗诊断系统误判导致患者死亡，责任主体可能包括研发公司、医院、医务人员(是否正确审核 AI 建议)甚至数据提供者(训练数据是否准确)，现有刑法未明确这些主体的范围，可能导致“遗漏追责”或“过度追责”。

主观方面，犯罪主观方面包括故意与过失，涉人工智能犯罪中主观方面的认定难题主要体现在“认知能力限制”与“技术复杂性”导致的过错判断困难。故意认定中，在 AI 工具型犯罪中如何判断技术提供者的故意(是否明知他人用于犯罪)是司法实践中的难点，例如 AI 语音合成工具的研发者是否明知用户可能用于诈骗，若工具具备“合法用途”与“非法用途”，如何认定研发者的故意，当前司法解释采用“综合判断标准”如工具的主要用途、是否采取防范措施但缺乏明确细则，导致认定结果不一致。过失认定中，在 AI 系统缺陷型犯罪中面临“预见能力限制”的挑战，AI 技术的复杂性使得研发者可能无法完全预见系统缺陷，例如算法在特定场景下的偏差如雨天自动驾驶传感器失效，研发者是否“应当预见”，当前司法实践多依据“行业标准”判断，但行业标准尚未完善导致过失认定缺乏客观依据。此外在 AI 自主危害行为中，主观方面认定面临“不可预见”的困境，若 AI 系统的进化超出研发者的认知范围，研发者是否存在过失，现有刑法中的“疏忽大意过失”要求“应当预见”，但对于技术进化带来的新型风险，研发者可能无法预见导致无法认定过失。

客观方面，犯罪客观方面包括危害行为、危害结果及二者之间的因果关系，涉人工智能犯罪中客观方面的适用难题主要体现在“行为自主性”与“因果关系复杂性”。危害行为认定上，传统刑法中的危害行为需由人有意识实施，涉人工智能犯罪中 AI 系统的自主行为是否属于“危害行为”存在争议，例如 AI 自动化攻击工具自主实施的攻击行为是否等同于研发者或使用者的行为，当前司法实践多采用“控制说”，认为若人类主体可控制 AI 系统则其行为视为人类行为，但对于高度自主 AI 系统(如可自主调整攻击策略)，人类主体可能无法完全控制，此时如何认定危害行为。此外对于“间接行为”的认定现有刑法难以应对，AI 工具型犯罪中人类主体通过 AI 系统实施危害行为(如编写算法让 AI 自动诈骗)属于“间接行为”，但其与直接行为的认定标准是否一致，例如 AI 系统自动筛选诈骗目标，人类主体仅提供初始指令，如何认定其行为的危害性，此类问题缺乏明确答案可能导致“行为认定扩大化”或“缩小化”[5]。

因果关系认定上，涉人工智能犯罪的因果关系具有“多环节性”与“不确定性”导致认定困难，具体表现为“多因一果”与“间接因果关系”两类场景。“多因一果”场景中，如自动驾驶事故可能由算法缺陷、传感器故障、用户操作不当共同导致，如何确定每个因素的作用大小；“间接因果关系”场景中，如 AI 推荐算法导致用户被诈骗，算法推荐与诈骗结果之间是否存在刑法上的因果关系。2024 年某短视频平台 AI 推荐诈骗案中，平台的 AI 算法将诈骗链接推荐给用户，用户点击链接后被骗，法院需判断 AI 推荐是否属于“介入因素”，若推荐行为显著提升诈骗的成功率(如精准筛选目标)，则可能认定存在因果关系，但当前缺乏明确的判断标准，导致类似案件的判决结果差异较大。

犯罪客体方面，传统刑法中的犯罪客体主要包括人身权利、财产权利、社会秩序等，涉人工智能犯罪中技术发展催生了新型法益(如数据安全、算法公平、AI 系统安全)，导致客体要件的认定面临难题。新型法益认定上，“数据安全”法益方面，AI 技术的应用使得数据成为核心资源，数据泄露、滥用造成的危害日益严重，但现有刑法中的“侵犯公民个人信息罪”仅保护个人信息，未涵盖“公共数据安全”(如企业商业数据、政府公共数据)，导致此类法益缺乏刑法保护；“算法公平”法益方面，算法歧视侵犯

公民的平等权，但现有刑法未将其纳入保护范畴，导致此类行为无法得到刑事惩戒；“AI 系统安全”法益方面，AI 系统被攻击或滥用可能引发公共安全风险(如自动驾驶系统被入侵导致大规模事故)，现有刑法中的“破坏计算机信息系统罪”难以涵盖此类场景。法益优先级判断上，涉人工智能犯罪中新型法益与传统法益可能存在冲突，如何判断优先级是司法实践中的难点，例如 AI 技术创新与数据安全的冲突，研发 AI 系统需收集大量数据可能侵犯个人信息权，如何平衡创新自由与数据安全，现有刑法未明确法益优先级，导致判决可能偏向某一方，如过度保护数据安全而限制技术创新，或过度放任创新而忽视权利保护。

4. 完善涉人工智能犯罪刑法应对路径的具体建议

(一) 明确责任主体范围，构建分层归责框架

AI 工具型犯罪中建议建立“作用大小 + 主观明知”的归责原则划分责任，犯罪实施者(使用者)承担主要责任无论其是否直接操作 AI 工具只要控制工具实施犯罪即认定为主犯，技术提供者承担次要责任若明知他人用于犯罪仍提供工具认定为从犯若不明知但未采取“风险提示”“用途限制”等措施认定为过失共犯，辅助者承担补充责任仅在其行为对犯罪实施具有重要作用时追究责任[4]。例如 AI 爬虫侵权案件中爬虫使用者承担主犯责任工具研发者若明知用于侵权仍提供承担从犯责任 IP 代理提供者若提供大量动态 IP 承担补充责任。

AI 系统缺陷型犯罪中建议根据主体的过错程度与在研发、使用环节中的角色确定责任，研发环节算法工程师承担“直接责任”(若因个人过错导致缺陷)测试工程师承担“监督责任”(若未发现缺陷)项目负责人承担“管理责任”(若未建立完善机制)，使用环节使用者承担“使用责任”(若违规使用)维护者承担“维护责任”(若未及时修复缺陷)，集体研发中根据每个参与者的过错份额确定责任如 A 工程师的漏洞导致缺陷承担 60%责任 B 工程师未发现承担 40%责任。例如自动驾驶事故案中若算法工程师编写的代码存在漏洞(过错程度 70%)测试工程师未发现(过错程度 30%)则二者分别承担主要责任与次要责任[5]。

潜在的 AI 自主危害行为当前阶段建议采用“严格责任”原则即研发者与使用者需承担严格责任无论其是否存在过错只要 AI 系统实施危害行为即追究其责任，同时建立“技术监管义务”要求研发者在 AI 系统中设置“安全开关”“行为监控”等措施若未履行该义务加重处罚。例如通用 AI 系统自主实施危害行为时研发者因未设置安全开关承担严格责任判处有期徒刑同时通过技术手段销毁该 AI 系统防范后续风险。

(二) 调整犯罪构成要件契合技术发展要求

犯罪主体方面建议在立法中明确涉人工智能犯罪的主体为自然人与单位排除 AI 系统的刑事责任主体资格，同时扩大人类主体的范围将研发者、使用者、维护者、技术提供者均纳入主体范畴明确每个主体的责任边界。例如在《关于防范和惩治涉人工智能犯罪的决定》中规定“涉人工智能犯罪的刑事责任主体为自然人与单位人工智能系统不具备刑事责任能力不得作为责任主体，研发者、使用者、技术提供者等人类主体根据其在犯罪中的角色与过错承担责任”。

主观方面建议在认定中引入“行业标准 + 风险预见义务”，故意认定中若技术提供者未采取行业通用的“风险防范措施”推定其“应当明知”他人可能用于犯罪，过失认定中以“行业技术标准”为依据若研发者未达到标准要求如未按标准测试 AI 系统推定其存在过失，同时引入“风险预见义务”要求 AI 研发者对技术可能引发的风险进行评估若未履行评估义务认定为过失。例如 AI 医疗诊断系统研发中若研发者未按《医疗人工智能产品测试标准》(GB/T 40278-2024)开展测试推定其存在过失若已履行测试义务但因技术局限无法预见缺陷不认定为过失。

客观方面建议在认定中作出调整，行为认定上将 AI 系统的自主行为视为“人类主体的间接行为”若

人类主体对 AI 系统具有控制能力即认定其实施了危害行为对于高度自主 AI 系统若人类主体未履行监管义务视为“不作为”，因果关系判断采用“客观归责理论”判断 AI 行为是否制造了不被允许的风险是否实现结果是否在构成要件范围内。例如 AI 自动化攻击案中研发者编写算法让 AI 实施攻击视为研发者的间接行为因果关系认定中若 AI 攻击制造了“系统瘫痪风险”且该风险实际发生认定存在因果关系。

犯罪客体方面建议在立法中增设“数据安全”“算法公平”“AI 系统安全”等新型法益将其纳入刑法保护范畴，同时明确法益保护的优先级即“人身权利优先于财产权利，财产权利优先于技术创新，自由公共安全优先于个人利益”在法益冲突时按优先级判断。例如 AI 数据侵权案件中，若侵犯个人信息权(人身权利)与技术创新(自由利益)冲突，优先保护个人信息权。若侵犯公共数据安全(公共利益)与企业利益冲突优先保护公共数据安全。

5. 结语

涉人工智能犯罪的复杂性与动态性决定刑法应对需保持“适应性”与“前瞻性”的平衡，当前我国刑法在调整涉人工智能犯罪方面已取得一定进展，但仍面临法律滞后、责任认定复杂、构成要件适用困难等挑战。通过加快立法进程优化、明确责任主体范围、调整犯罪构成要件等举措可构建契合人工智能时代发展要求的刑法治理框架，实现社会秩序维护与公共安全保障、技术创新自由与个人权利保护的平衡。

未来随着人工智能技术的进一步发展新型犯罪形态与风险将不断涌现刑法应对需持续完善，加强技术监测建立“AI 犯罪风险预警系统”及时发现新型犯罪趋势，深化国际合作与其他国家共同制定涉人工智能犯罪的治理标准应对跨国 AI 犯罪，注重理论研究鼓励学界探索 AI 自主危害行为、算法刑事责任等前沿问题为立法与司法提供理论支持。只有通过“技术 + 法律 + 理论”的多元协同才能实现对涉人工智能犯罪的有效治理推动人工智能技术在安全、合规的轨道上发展。

参考文献

[1] 陈兴良. 人工智能犯罪的刑法应对[J]. 国家检察官学院学报, 2025, 33(2): 3-9.

[2] 刘宪权. 人工智能时代的刑事责任演变: 昨天、今天、明天[J]. 法学, 2019(1): 79-93.

[3] 孙道萃. 人工智能犯罪的知识解构与刑法应对[J]. 青少年犯罪问题, 2023(2): 4-19.

[4] 吴沛泽. 人工智能刑事风险的识别与规制[J]. 政法学刊, 2025, 42(4): 52-61.

[5] 宋卓琳. 涉人工智能犯罪的刑事责任问题研究[D]: [硕士学位论文]. 哈尔滨: 黑龙江大学, 2022.