

# 自主型人工智能犯罪的刑事归因与归责

刘芮妍

苏州大学王健法学院, 江苏 苏州

收稿日期: 2026年1月4日; 录用日期: 2026年1月14日; 发布日期: 2026年2月6日

## 摘要

人工智能技术的迭代升级催生出自主型人工智能犯罪的复杂归责难题, 风险社会下人工智能侵害人类法益的行为越来越频繁, 基于其自主决策具备特有的算法黑箱与多端介入性, 导致在侵害结果发生时, 适用传统刑法因果关系理论判定责任归属陷入适用困境。本文通过分析条件说、相当因果关系说和双重筛选条件说在人工智能自主决策犯罪归因中的局限性, 提出三层筛选穿透的归因模型和分层刑事归责的建构路径。查明算法中自主决策与结果侵害间的刑事意义关联和相关主体的互动联系、提取重叠因果中的强黏性标准, 与制度层面的自主人工智能犯罪的法律搭建和过错责任划分结合、辅之以多端化复合治理模式, 实现技术治理与法律规制的深度融合。

## 关键词

自主型人工智能, 因果关系, 自主决策, 刑事归因, 刑事责任

# Criminal Attribution and Responsibility of Autonomous Artificial Intelligence Crimes

Ruiqian Liu

Kenneth Wang Law School, Soochow University, Suzhou Jiangsu

Received: January 4, 2026; accepted: January 14, 2026; published: February 6, 2026

## Abstract

The iterative upgrading of artificial intelligence (AI) technology has given rise to the complex attribution of responsibility for autonomous AI crimes, and AI infringes on human legal interests more and more frequently in a risky society. Based on its autonomous decision-making, it has its own algorithmic black box and multiple interventions, which leads to a dilemma of applying traditional criminal law causation theories in determining responsibility attribution in the case of infringing results. This paper analyzes the limitations of the condition theory, equivalent causation

theory and double screening condition theory in the attribution of artificial intelligence autonomous decision-making crime, and proposes the attribution model of three-layer screening penetration and the construction path of hierarchical criminal attribution. It identifies the criminal significance of the connection between autonomous decision-making in the algorithm and the resultant infringement and the interaction of the relevant subjects, extracts the strong viscous criteria in the overlapping causation, combines with the legal construction of autonomous AI crimes and the fault-based division of responsibility at the institutional level, and is supplemented by the multi-terminalized composite governance model, so as to realize the deep fusion of technological governance and legal regulation.

## Keywords

Autonomous Artificial Intelligence, Causality, Autonomous Decision-Making, Criminal Attribution, Criminal Responsibility

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. 问题的提出

随着智能网联汽车系统、AI 司法辅助决策程序等人工智能技术的深度嵌入，人工智能在智慧司法、自动驾驶、精准医疗等领域的突破性应用，不仅重构了传统产业形态，更衍生出前所未有的技术伦理困境与法律规制难题的复合性挑战，相较于达沃斯会议确立的人工智能初始概念，当代人工智能技术迭代已突破“工具论”范畴，模拟人类某些核心能力，创造出人类才具备的、甚至是超越人类所具有的智能[1]，呈现出自主决策与自我优化的技术主体特征。这种主体性特征，正是当前法律体系面临适应性危机的根源。

近年来，以生成式 AI 伪造生物特征实施诈骗等典型案例，暴露出传统法律理论应对智能技术风险的能力赤字，刑法学界对于人工智能的理论研究也在如火如荼地展开。学界普遍认为，弱人工智能仅可在既定的算法程序内运行，从而不能称之为刑事责任主体；关于强人工智能的刑事主体资格亦有诸多争议，其是否具备辨认和控制能力，及目前可否探讨人工智能犯罪等问题尚未达成共识，各种观点激烈交锋，胜负未定。尽管我国目前尚处于弱人工智能时代，工具利用型的人工智能所带来的便利和对其载入算法程序的信赖，使部分学者认为对人工智能刑事责任问题的相关分析为时尚早，而这并非表明对强人工智能犯罪的探究缺乏实际价值。2024 年欧盟《人工智能法案》第三条中明确了人工智能系统指的是通过机器学习，能够在给定的目标下，自主生成输出的软件系统，并可通过适应性机制改变其行为模式，人工智能因具备自主决策能力，可能衍生出人类难以预判的行为风险，进而引发不良后果的消极侧面[2]。在某信息技术公司与某科技公司不正当竞争纠纷案件中，亦提及模型是计算机从大量训练过程中统计到的输入和输出数据之间的关联关系，使模型可以对其未曾接触过的但分布相似的数据进行预测或决策。<sup>1</sup>尽管学界多数学者认为人工智能并不具备刑事责任能力，但不可否认的是，其完全可能借助深度学习机制发展出自主意识与独立意志[3]。

自主型人工智能实施的危害社会行为，是由其自主生成的程序运行而产生的，这类行为的显著特征在于：基于强人工智能的独立意志，完全脱离人类意识干预或超出人类控制范围，如若将行为后果归责

<sup>1</sup>参见北京知识产权法院（2023）京 73 民终 3803 号二审民事判决书。

于相关主体，则违背了刑法罪责自负的基本原则。从法律属性来看，此时的人工智能已经不再是传统意义上的工具，立足人机协同的视角，其客观层面不仅属于刑法意义上的危害行为，主观层面也具备刑法意义上的故意，可依据新三层筛选模式标准进行归因归责。技术发展的迅猛态势及其蕴含的刑事风险警示，学界不应止步不前，而应未雨绸缪，带有自主决策性的人工智能犯罪的研究与讨论绝非空中楼阁。面对技术演进与规范供给的结构性失衡，刑事司法实践中普遍存在的责任主体虚化、过错认定机制失灵、技术因果证明困境等鸿沟，凸显出构建自主型人工智能犯罪刑事归因和归责模式的迫切需求。

## 2. 自主型人工智能犯罪刑事归因的三层筛选穿透

在高精尖技术发展日新月异的当下，人工智能依托高智能特性与先进的数据运算能力，对法益的侵害正呈现更深层的演进态势。除去执行预先设计的编程外，已经在很大程度上发展出自主学习与分析能力，突破传统工具属性，在处理信息数据同时形成独特的认知逻辑，产生巨大的算法黑箱，加剧对人工智能行为动机的研判难度，更使其犯罪行为的刑事责任归属复杂化。

### 2.1. 既有刑法理论应对因果关系判断的阙如

传统刑法理论持续受到技术形态嬗变而引发的新型犯罪危机的冲刷<sup>[4]</sup>。现有条件说、相当性因果关系说、双重筛选条件因果关系学说理论应用至人工智能犯罪行为领域无法有效洽合，应当结合自主型人工智能超强算法的功能特点重构其犯罪行为的因果关联判断流程。

根据条件说的逻辑推演，数据录入阶段与算法运行中的偏差错误指令生成，均可以构成危害结果发生的必要条件，从而达到“有 A 即 B”的犯罪归责。然而该理论的适用使因果关系的判定必须以人工智能危害行为与结果间的实证关联为基础，这一证明过程受制于人类现有技术水平难以对多发系统故障进行充分抽丝剥茧提取到恰当原因。以深度学习系统为例，其黑箱化运作机制导致即便完整掌握案件事实，仍可能因技术认知局限而无法完成条件关系的有效认定。依赖于既定科学认知框架，难以涵摄超越当前经验法则的新型因果关系，基于当前技术水准作出的判断难免隐含着认知僭越的风险。

相当性因果关系说判断基础是行为人、一般人的认识与预见结果发生的可能上，<sup>[5]</sup>尽管对因果关系的范畴进行限制，仍未改变相当性模糊的界定困境。事实上，一般人在人工智能技术认知层面存在显著差异，这使得法律因果关系判定难以摆脱主观随意定性，不仅致使刑事因果关系的归责要件趋于模糊，更动摇了因果关系认定的客观基础。就行为人对人工智能发出指令的行为与基于自主学习后的系统所做行为而言，其归责主体呈现明显分野。此种“相当性”模式在司法实践中面临多重操作瓶颈，核心矛盾在于过度倚重行为人的主观认知，实则弱化了客观性评价在法律因果关系判定中的核心价值。

适用双重筛选条件说时，第一步先依据客观归责理论中对结果归责判断标准“结果没有超出构成要件的保护范围”进行初筛，第二步根据前两个条件“行为制造了不被允许的危险 + 行为实现了不被允许的危险”进行再选，从而摘挑出对具体结果的发生具有“引起”意义的行为<sup>[6]</sup>。可见该说是完全建立在客观归责理论之上，其对于刑事归责的缜密判断从而搭建的因果关系判断结论诚然可资借鉴，但因太过抽象欠缺具像化操作基准，存在行为随意出入罪的嫌疑<sup>[7]</sup>，这种传统单一的归因理论难以有效回应技术风险社会的归责需求，实质上杂糅了事实与价值判断偏离了纯粹客观因果关系的范畴。

涉自主型人工智能犯罪刑事归因问题中，核心难点在于：当传统法益如人身、财产等遭受侵害时，如何认定基于深度学习形成的自主意识对损害结果的作用机制。而以上学说均不能有效论证，以自动驾驶汽车交通事故作为典型例证。自动驾驶模式下发生交通事故的肇因大多数是基于其设计系统所配置的算法和程序的固有缺陷，即生产者因疏忽未进行设计优化，将本可以降低或避免的产品致损的可预见风险变为现实的不合理危险<sup>[8]</sup>。其因果判断流程如若按照以上三种学说的路径，条件说基于技术回溯数据

缺陷可能是由汽车硬件上传感器故障,或软件上算法误判的多重加持,无法确定技术叠层下的准确验证;相当性说在驾驶员对其自动驾驶理应避险的自主算法的高度信赖,以及智能驾驶系统学习算法的黑箱不可避免等局限,归责认定滑向主观随意;在筛选说中,鉴于自动驾驶汽车事故率无论如何都比人工驾驶的更低,在有无实现不被允许的危险的判定上,通过以利益衡量或不具有行业规则等予以排除构成要件之该当性[9],易阻断刑事归责可能性。

## 2.2. 新三层筛选模式归因犯罪标准的提出

以上理论困境揭示出刑法因果关系认定体系的时代性局限,当技术革新速度超越法律认知更新频率时,就自主型人工智能犯罪领域而言,亟需在吸收客观归责理论和双重筛选条件说的基础之上,构建更具贴合性和体系性的因果关系认定框架,为技术发展的不确定性预留规范评价空间。

在探讨新三层筛选归因模式之前,有必要明确人工智能主体地位的基本立场。根据德国学者观点,刑法教义学的根基与边界严格根植于刑法规范问题,核心在于深度解构规范的概念内涵、逻辑结构及价值,将法律素材通过体系化思维重构,探索寻找概念构成和系统学的新的方法[10]。传统刑法理论中“机器不能被骗”属于既定的教义学规则[11],曾长期主导着司法实践对人工智能的定性逻辑,而今人工智能已经逐渐趋近于或超过人脑的识别技术,“人机关系”再也不可能二元化,而是确实实现了内在的有机融合[12],明确人工智能自主决策的主体性程度可以为其中因果流程的判定起到一定的作用。

### 2.2.1. 自主型人工智能行为产生刑事意义的危害结果

自主型人工智能犯罪的特性在于程序代码算法的开发错误,可能触发不同性质的法律评价,要防止将技术风险简单转化为刑事危害结果的认定,首先必须将非刑事法意义上的评价结果剥离。以自动驾驶为例,在汽车驾驶系统发生故障导致使用者在路上行驶中冲撞行人致人死亡的结果才归属在刑法范畴之内,如若使用者利用自动驾驶系统紧急避让程序主动偏离原定行使路线避免了与行人碰撞的结果,则并非刑法规制范畴中的结果,无须分析使用者的行为与行人事故预防结果之间的因果关系。

### 2.2.2. 存在自主智能算法规范违反结果与主体的互动联系

算法黑箱随着人工智能技术的突飞猛进,算法透明的技术障碍已经被攻克,尽管部分智能模型的算法运行内在透明性欠缺,但这一技术缺陷亦能借助其他事后方法予以补救[13]。人工智能产品需经历从研发设计到应用落地的全周期链条,当智能系统在终端场景引发权益侵害时,其因果关系认定往往面临双重挑战:一是研发主体、运营主体与终端用户间的时空关联呈现非连续性特征,二是算法决策机制、数据交互路径与损害后果之间形成多环节介入的复杂归责路径。因此,其中关键的判断标准即是人的行为对于自主智能决策的控制支配程度和人工智能犯罪时的时间与空间条件是否符合预期,在人工智能的自主决策导致侵害的原因并非是因为研发、生产、使用者的行为引起时,同样不能有效认定自主人工智能决策侵害的归属关系。

### 2.2.3. 重叠因果中智能算法强黏性标准的分离提取

不同于人类犯罪通常的一因一果,人工智能自主决策过程中存在着多因一果或多果一因的情形,前期编程设计到后期的投产运营直至最后流程的使用者都可能成为自主型人工智能犯罪的有关责任人。人工智能行为的源动力根源是人的意志,背后存在多个主体同时承担责任的情况,因此需抽离出对算法植入具有关键强黏性的程序信息,明确多方主体的行为参与损害发生的具体作用导向。尽管有学者认为算法控制研究对象仅包含算法制造和使用行为,并不能作为刑法意义上的行为[14],但在自主型人工智能决策过程中,基于算法代码下使用者与人工智能自主生成的协同效果,打破了这种常规认识局面,通过对犯罪时算法的精准回溯,作出较为准确的方向预测。在此意义上,应根据不同算法等级的影响考量其中



自主决策的支配度。

基于客观归责理论的核心要义，在智能算法介入的场景中，算法输出中断人类行为与结果因果链条的核心判断标准，在于算法自主程度是否形成独立的风险支配力，从而切断人类初设行为的风险，且损害结果完全或主要由算法的自主决策所引发。例如，自动驾驶算法在运行中自主演化出未被人类预设的路径规划逻辑，导致其在特定路况下作出危险变道决策引发事故，此时人类对车辆的初始启动行为与事故结果之间的风险关联已被算法的自主决策切断。另外，若算法输出是人类基于专业认知也无法预见的异常介入因素，且该因素独立导致损害结果，则人类行为与结果的因果链条中断。

此种强黏性算法的判断是相对化的和弹性化的。其要求达到引起最终由于人工智能自主决策的损害结果的程度并不是一个绝对阈值，亦不存在一个普适性的量化标尺。因为对于特定情境中算法对因果流程的支配程度是一个比较性问题，不同行为领域的算法黏性标准不能采用统一的衡量标准。比如，交通事故领域中自主智能算法叠加因果之间的黏性和网络领域中自主智能算法叠加因果之间的黏性就无法采取同一标准，前者因物理时空的闭合性要求算法对损害具备高度的直接支配力，后者则因信息网络传播的即时性、泛化性、交互性的特征，侧重算法对风险扩散的放大效应，二者不可适用同一归责基准。

### 3. 自主型人工智能犯罪分层刑事归责体系的建构

人工智能犯罪形态呈多样化发展，实现自主型人工智能犯罪行为在司法实务中的责任落实关乎刑法教义学中主体适格性、罪责自负原则的理论突破，借助深度学习框架的人工智能体系已经触及传统工具边界，亟需建构划定合理的刑事责任范围体系。

#### 3.1. 搭建自主人工智能犯罪的前瞻性法律体系

生成式人工智能凭借大规模预训练模型，实现智能化水平的跨越式发展，其内容生成机制已超出研发者和使用者的意志控制范围<sup>[15]</sup>。当前人工智能犯罪的刑事处罚体系存在制度性偏差，部分本需纳入刑法规制的人工智能犯罪行为因立法覆盖不足，出现即便在实践中出现具有刑法意义的危害后果，也面临责任追究的规范失灵。立足于刑法教义学基础，坚持积极刑法观，完善技术迭代中周延至算法监管、自主数据运用等问题的处理，实现新时代自主型人工智能犯罪适应性法律体系的专项构建，是解决刑事归责难题的有效路径。

第一，需完善破坏计算机信息系统罪的规制范畴。囿于现行刑法中对涉人工智能犯罪对象局限在计算机信息系统，在面临具有自主决策人工智能系统的规制时已经力有不逮。因此，应当适时拓展犯罪对象的涵摄范围，将“具有自主决策功能的算法模型或人工智能体”纳入系统完整性保护范畴，并重构行为要件，增添“故意干扰算法重要参数”“非法操纵算法技术”等新型行为模式。

第二，采取扩大解释覆盖自主型人工智能相关产品类犯罪行为。参照《刑法》第146条生产、销售不符合安全标准的产品罪，完善对具有自主决策性人工智能犯罪行为的规制体系。从风险链条来看，研发者是风险存在的来源，使用者是风险的实际触发者。在司法实践场景中出现研发者或制造者在设计或制造过程中违反了相关设计标准，或销售者明知自主型人工智能产品并未达到标准仍继续销售，抑或是使用者在明知某自主型人工智能产品的标准不合格的前提下仍使用时，则可能成立该罪名。

第三，构建自主型人工智能责任人监督责任解释条款。作为人工智能系统的实际控制者和利益获得者，研发者与制造者的义务范畴已超越传统责任框架，不仅负有基本算法审查、数据防护等基础监管职责，亦需要通过专业评估算法系统，对运行中可能出现的自主决策偏差等新型风险进行识别与干预。因此，有必要在司法解释中予以补充完善，在行为要素上可以涵盖“未建立算法风险评估系统”“未履行数据溯源义务”等形态，同时规定以构成要件中嵌入履行监管义务即可免责的豁免条件保持刑法谦抑性。

自主型人工智能犯罪的法律完善需建立在现行法适用优先的逻辑前提下，仅在穷尽现行法律规定和法律解释确实无法覆盖相关情形时，才需审慎考虑部分罪名增设的必要。具体而言，可先通过司法解释对“自主算法决策”等核心概念进行匹配扩张，待司法实践积累了充足案例经验后再逐步推动条文修订。另外，可以考虑参照知识产权法院、铁路法院等专门审判机构的经验，设立专门的人工智能案件审理法院，联合科研机构开展法官实训课程；在裁判辅助上，完善技术专家陪审制度，为算法解释、因果关系认定等技术难题提供专业支撑。

### 3.2. 分解自主智能决策中的相对严格过错责任划分

虽然人工智能的类人化显著，算法系统运行的背后却始终关联着人类行为，因而难免存在过失因素。自主型人工智能犯罪的认定应明确一个责任原则理念的前提，即采用过错责任原则，而不是严格责任原则。严格责任原则又称“结果责任”，是只要行为人实施了造成危害结果的行为，不论是否有过错均应当追究刑事责任，出发点是预防重要紧迫的危险，而当下对人工智能自主决策引发的犯罪行为，可通过梳理全链条主体的过错实现追责，无需借助严格责任原则规避举证难题。

研发者的自主算法设计与数据安全保障责任认定。据《生成式人工智能服务管理暂行办法》第7条、第8条相应规定，研发者需采取有效措施识别、过滤、降低数据训练产生的“毒性”[16]，确保算法的可靠性和无偏差，负有算法设计和数据质量的安全保障职责，使系统程序设计符合伦理和法律标准。从技术实践来看，研发者的责任边界需结合人工智能的自主特性加以界定。当智能系统算法运行中出现智能事故与算法编程异化的非预见性变动时，依靠人工智能系统自主识别分析已经生成最优的路径，若该情形超出人类认知阈值，则可基于其不可预见性免责。但这一情形存在严格例外，即如果在自主智能代码出现逻辑漏洞时，研发者在该阶段因选择性忽略而置之不理，且未向后续的制造者和使用者尽到告知义务，构成对注意义务的实质性违反。在此情形下，需由研发者承担最终的归责后果。

制造者的算法质量保障和风险告知责任认定。制造者的责任认定应遵循义务履行—过错归责—技术容错的三重标准，作为自主型人工智能产品研发成果的制造实现者，若其已经在最大程度上确保研发智能产品的制造质量安全落地，向使用者履行完整告知义务，并提供技术支持，便无需承担责任。其通过现有刑法罪名或新设罪名进行刑法追责的情形主要体现在：一是在自主智能产品系统投放市场前，未进行实质性检查；二是因故意或过失未履行风险告知义务。除此之外，需注意不能将对一切在人类认知空白领域或者说虽不符合相关标准，但也被证明其研发产品确实会导致严重损害结果的设计和制造行为绝对禁止[17]，避免过度遏制人工智能发展的技术创新。

使用者规范操作和主体差异化责任认定。自主型智能产品的用户在没有按照系统的设计用途和操作规范使用智能决策系统时，应当承担相应的刑事责任。例如，将自动驾驶汽车用于非设计场景或违反交通规则，或其他在完全可避免事故发生情形下未及时采取任何紧急措施时，承担未尽合理使用的责任。同时，使用者在使用前和使用中也应当尽到风险评估和监控责任，确保安全场景下的使用和必要手段的干预。不容忽视的是，若对人工智能产品的非专业使用者苛以过高的注意义务，则与法律不强人所难的基本原则相悖，过度强制其承担超出能力范围的审慎义务，反而会动摇公众对技术使用的合理信赖。应当合理匹配义务设定中使用者的认知与控制可能性，根据其不同的专业背景、技术使用频次等要素差异化分配责任强度，契合风险社会的认知逻辑。

### 3.3. 辅就自主型人工智能多端化复合治理模式

数字技术时代，自主型人工智能平台方具有刑事风险治理的“看门人义务”[18]，其规范义务应从前端的预防管控、中端的退化处理以及后端的数据熔断强化予以辅助展开，打造有层次且渐进有序地介入智能技术防范的治理体系，确保从多重维度充实人工智能犯罪的规制，这也符合技术发展的基本规律。

自主型人工智能刑事防控义务前端的预防管控。可以借助大数据技术将海量的法律规范条文信息以及相关规制人工智能产品犯罪的相关判例数据进行编码,并将其植入自主型人工智能的算法系统中,对大量数据进行分析 and 深度学习,以人类输入和机器学习的方式,形成预防犯罪的内在观念,构建起针对人工智能自主决策致害的事前预防,在源头上降低犯罪风险,避免人工智能自主决策中的违反犯罪行为,使其更好地服务于人类。

自主型人工智能刑事防控义务中端的退化处理。深度学习能力作为自主型人工智能的核心技术特质,类脑化的处理模式可以使其形成对人类社会规则的适应性理解,更成为主体资格说主张其具备刑事责任主体资格的论证依据。在此背景下,构建能力退化这一技术防控机制,为刑事归责提供了兼具实操性和谦抑性的中间方案,即针对自主智能决策中产生的低危害性犯罪行为,通过算法回溯定位引发危害的学习模块,修改程序以削弱其深度学习权限。在人工智能自主决策中无法判断导致其犯罪深度学习范围时,可采取完全退化其为单一领域的人工智能,使其丧失跨领域的学习能力,仅限定在通过指令设定程序下简单的机器学习。

自主型人工智能刑事防控义务后端的数据熔断。在自主决策型人工智能发生系统性失控或者伦理失范的重大风险刑事事件中,可启动后端治理中的算法销毁措施,即对高社会危害的自主型智能犯罪行为,在采用单纯的算法退化程序修改已经不能完全削弱其核心自主智能决策性时,清除该程序的数据记录,删除其赖以维系代码运行的信息数据和算法编程,以回溯到未产生错误时期,同时保留后续深度学习中的正向数据学习之方式,以此达到清除刑事犯罪风险和避免技术毁坏浪费的效果。

#### 4. 结语

人工智能技术正在以指数级发展态势深刻重构社会运行的基本逻辑,在技术伦理与法律规范的交错地带,自主型人工智能的刑事归因和归责难题,映射出刑法传统理论适用的现实困境。图之于未萌,虑之于未有。当算法决策的自主化程度突破人类控制的边界,法律不应是被动地应对改变,而应随变革的浪潮向前推进,刑法亦须超越工具客体化的固有思维惯性。本文建构的新三层穿透筛选归因模型与分层归责体系,正是通过解构传统刑事因果关系理论的适用困境,以相关主体行为与算法决策的实际交互关系为基础,结合技术发展阶段的算法黏性标准判定因果归属;通过完善法律体系规定和划分行为主体的梯度责任,在技术理性与法律价值之间构建动态平衡的治理框架。相较于被动等待技术发展冲击法律规范体系,更具建设性的路径是在既有法律框架内完善理论的适应性内容要素,唯有如此,才能在科技发展的浪潮中构筑动态平衡的刑事法治屏障,既通过细致化的责任分配为人工智能技术发展预留必要试错空间,又以刑法的谦抑性品格守住社会最后防线。

#### 参考文献

- [1] 刘宪权. 人工智能迭代发展对财产犯罪刑事责任的影响[J]. 国家检察官学院学报, 2025, 33(2): 33-44.
- [2] 于润芝. 人工智能自主决策介入结果发生的因果流程的刑事归责[J]. 苏州大学学报(法学版), 2024, 11(4): 27-39.
- [3] 孙道萃. 人工智能刑法研究的反思与理论迭进[J]. 学术界, 2021(12): 64-76.
- [4] 谢铠. 网络猥亵犯罪的理论建构与刑法适用[J]. 青少年犯罪问题, 2023(5): 95-110.
- [5] 刘志强, 方琨. 论人工智能行为法律因果关系认定[J]. 学术界, 2018(12): 76-92.
- [6] 刘宪权. 涉人工智能犯罪中的归因与归责标准探析[J]. 东方法学, 2020(3): 66-75.
- [7] 刘杰. ChatGPT 类生成式人工智能产品提供者之刑事责任[J]. 法治研究, 2024(2): 61-71.
- [8] 王乐兵. 自动驾驶汽车的缺陷及其产品责任[J]. 清华法学, 2020(2): 93-112.
- [9] 魏超. 自动驾驶汽车交通事故的刑法归责[J]. 环球法律评论, 2024(5): 56-71.

- [10] 耶赛克·魏根特. 德国刑法教科书[M]. 徐久生, 译. 北京: 中国法制出版社, 2001: 53.
- [11] 大谷实. 刑法讲义各论[M]. 黎宏, 译. 北京: 中国人民大学出版社, 2008: 236.
- [12] 李琼琼, 李振. 智能时代“人机关系”辩证——马克思“人与机器”思想的当代回响[J]. 毛泽东邓小平理论研究, 2021(1): 71-79, 108.
- [13] 安晋城. 算法透明层次论[J]. 法学研究, 2023(2): 52-66.
- [14] 皮勇. 论自动驾驶汽车生产者的刑事责任[J]. 比较法研究, 2022(1): 55-70.
- [15] 王迁. 论人工智能生成的内容在著作权法中的定性[J]. 法律科学(西北政法大学学报), 2017(5): 148-155.
- [16] 江海洋. 论人形机器人的刑事主体地位与归责[J]. 东方法学, 2024(3): 26-37.
- [17] 叶良芳, 马路瑶. 风险社会视阈下人工智能犯罪的刑法应对[J]. 浙江学刊, 2018(6): 65-72.
- [18] 储陈城. 间接治理: 刑法应对数字技术风险的基本模式[J]. 法学论坛, 2025, 40(1): 65-77.