

人工智能在国际法框架下的责任归属问题探究

张景

海南大学，海南 海口 570228

摘要： 本研究旨在探讨人工智能在国际法框架下的责任归属问题，分析其法律属性与现存法律体系的适用性。通过界定人工智能的概念及其分类，深化了对 AI 特点及其法律属性的理解。本研究比较分析了传统责任归属原则，如侵权责任、合同责任与新兴理论，强调人工智能的不可归责性和模糊性对传统法律框架构成的挑战，提出如主动责任等新的责任归属机制的必要性与可行性。通过对比我国与美国、欧盟在人工智能责任立法方面的实践，以期对未来研究提供有益启示，指引国际法与人工智能的深入融合与发展。

关键词： 人工智能；国际法；责任归属；技术代理；自主决策

Research on the Attribution of Responsibility of Artificial Intelligence Under the Framework of International Law

Zhang Jing

Hainan University, Haikou, Hainan 570228

Abstract： This study aims to discuss the attribution of responsibility of artificial intelligence under the framework of international law, and analyze its legal attributes and applicability of the existing legal system. By defining the concept of artificial intelligence and its classification, the understanding of AI characteristics and its legal attributes is deepened. This study compares and analyzes traditional liability attribution principles, such as tort liability, contract liability and emerging theories, emphasizes the challenge posed by the non-liability and ambiguity of artificial intelligence to the traditional legal framework, and proposes the necessity and feasibility of new liability attribution mechanisms, such as active liability. By comparing the practice of China, the United States and the European Union in the legislation of artificial intelligence responsibility, we hope to provide useful inspiration for future research and guide the in-depth integration and development of international law and artificial intelligence.

Keywords： artificial intelligence; international law; attribution of responsibility; technical agent; autonomous decision

在当今社会，人工智能技术的快速发展已成为推动各行各业革新的核心动力。随着深度学习（Deep Learning）、机器学习及自然语言处理（Natural Language Processing）等技术的迅猛演进，AI 的应用范围不断扩展，涵盖医疗、金融、交通等多个领域。^[1] 然而，伴随这一技术革命而来的是关于其法律责任归属的严峻挑战，特别是在国际法的框架下，该问题愈发凸显出其复杂性的重要性。

传统法律体系在认定责任主体时，通常依赖于个体行为的可识别性与可追溯性。但在 AI 所参与的决策过程中，自动化与自主性使得技术代理（Technological Agency）愈发模糊，难以清晰划分责任归属。

一、人工智能及其法律属性概述

（一）人工智能的定义与分类

人工智能作为一种复杂的技术现象，其定义与分类不仅涉及技术层面，更对法律适用形成深远影响。^[2] 在当前的学术讨论中，人工智能的定义多种多样，主要可以归纳为两类：一类是将人工智能视为具备模拟人类智能特征的系统，另一类则从其应用功能出发，认为人工智能是用于支持或替代人类决策、理解和操作的工具。^[3] 这种定义上的多元性直接影响到法律界对其责任归属的界定，尤其是在自主决策与辅助决策系统之间的法律适用。

自主决策 AI 系统指的是那些能够在缺乏人类干预的情况下，依据输入数据进行判断和决策的技术，例如自动驾驶车辆。^[4] 这种系统通常通过复杂的机器学习算法处理海量数据，形成预测模型并制定行动方案。在这种情况下，若发生事故，法律责任的归属问题显得尤为复杂，因为其决策过程是高度自动化且不可预测的。例如，在一次知名的自动驾驶汽车事故中，车辆根据实时路况做出避让操作，最终却未能避免碰撞，相关责任的界定引发了广泛讨论，涉及到车辆制造商、软件开发者以及交通管理部门的责任。

（二）人工智能的法律地位与属性

针对人工智能的法律地位与属性，我们必须理解其在法律体

系中的角色。^[6] 法律属性分析 (Legal Attributes Analysis) 为我们提供了一个系统化的视角, 通过不同法律框架下对 AI 的定位, 揭示其法律责任的归属机制。

需要指出的是, 当前的法律框架尚未完全适应 AI 的特性与发展, 尤其是在其独立性、自主性方面。传统法律概念主要基于自然人或法人, 而 AI 的出现则打破了这一界限。例如在《中华人民共和国民法典》中, 法律并未明确赋予 AI 这一新的法律主体以权利或义务, 导致在行为发生归责时, 法律的空白空间愈发明显。

(三) 人工智能在国际法中的适用范围

人工智能 (AI) 作为一种具有高层次算法驱动的技术, 其在国际法领域的适用范围日渐受到关注。^[6] AI 的集成不仅重塑了技术行业的运作模式, 亦对国际法的实施产生了深远影响。尤其是在数据处理、智能决策和自动化执行等方面, AI 及其应用成为法律适用性分析的关键领域。

AI 在国际法中的适用范围可以涵盖多个领域, 如国际贸易法、国际人权法和国际环境法等。在国际贸易法领域, AI 技术被广泛运用于合同管理、风险评估和合规监测, 通过智能化的数据分析与情境预测, 提高了贸易互动的效率和透明度。^[7] 例如 AI 驱动的智能合约系统, 使得交易各方在未必完全信任的情况下, 仍能保持交易的安全性和合同的执行率。

在国际人权的适用方面, AI 技术同样展示出其潜在的贡献。通过大数据分析与机器学习算法, 国际法庭可以更精确地对大规模人权侵犯案例进行筛查与评估。例如, AI 算法可以帮助相关机构快速识别侵权行为模式, 从而更高效地进行干预与改善。这表明, AI 不仅能够有效提升法治实施的效率, 而且为公平正义的实现奠定了更为坚实的基础。

二、国际法框架下的责任归属理论

(一) 传统责任归属原则

在国际法的传统框架中, 责任归属的基本原则通常基于“国家责任法”的形式。根据《国际法委员会关于国家责任草案》, 国家的国际责任可归因于其违反国际义务的行为, 这种义务包括对他国公民或其他国家利益的保护义务。然而, 随着“人工智能”等新兴科技的发展, 这种传统责任归属理论面临诸多挑战。

传统的责任归属原则难以有效适用于人工智能主体的行为。与以往主要依靠人类决策的法律框架相比, 人工智能的自动决策和推理能力使得其行为的复杂性显著增加。例如在自动驾驶汽车发生事故的责任承担问题, 这种情形挑战了人类主导的责任归属模式, 突显出在界定责任时的模糊性与不确定性。

当前的法律体系往往依赖于明确的人类行为与法律后果之间的因果联系, 然而, AI 系统在自主性和决策智能上的日益增强, 使得这一传统的法律分析框架面临前所未有的困境。^[8] 例如, 当一款自驾车因系统失误导致交通事故时, 究竟该由制造商、软件开发者, 还是用户承担法律责任, 成为法律实务中的难点。此类问题之所以复杂, 源于多主体参与及其交互作用的高度不确定性。

(二) 责任归属的新兴理论与机制

在探讨国际法框架下的责任归属问题时, 尤其是在针对人工智能 (AI) 技术的快速发展与应用背景下, 责任归属的新兴理论与机制愈显重要。当前, 随着人工智能逐渐融入诸多领域, 包

括自动驾驶、医疗辅助、金融决策等, 其自主性与决策能力的提升, 使得传统的责任归属理论显得日趋不足。在此背景下, 提出创新的责任归属理论显得尤为必要。

现有的责任归属理论如“主观责任”与“客观责任”在面对人工智能的自我学习能力时, 难以明确锁定责任方。例如, 自动驾驶汽车发生事故时, 判断责任归属通常围绕于驾驶者、制造商及软件开发者之间展开。然而, 人工智能在涉及决策时可能脱离人类的直接控制, 从而导致责任归属的复杂性加大。对此, 需引入“分层责任理论”, 该理论提出根据行为的不同层级, 划分责任归属, 进而明确各方在特定情境下的责任。

三、各国在人工智能责任归属方面的立法与实践

(一) 美国的相关立法与案例

在探讨人工智能在国际法框架下责任归属问题时, 必须注意到美国联邦与州立法在该领域存在的显著差异。

联邦法律在人工智能责任归属问题上实行的是较为统一的标准, 而州立法则因各州经济、文化和法律传统的差异, 展现出显著的多样性。例如, 联邦层面的《自动化责任法案》初步确立了在特定情况下人工智能系统所引发的损害责任由开发者或使用者承担的基本框架。^[9] 例如, 加州的《机器人法律责任法》明确规定当 AI 系统造成但因算法偏见引发的不公平或有害后果时, 其责任应由相关利益方共同承担, 这一立法明确了算法透明性的重要性。

自动驾驶汽车作为 AI 应用的一个突出案例, 突显了现行法律体系对责任归属问题的适应性和局限性。例如 2018 年发生的 Uber 自动驾驶车辆致人死亡事件, 引发了广泛的法律与伦理讨论, 不仅暴露了 AI 系统在决策过程中的复杂性, 还引发了对于企业与开发者在技术失误或事故中责任的法律解读。

(二) 欧洲的相关立法与案例

随着人工智能技术的不断快速发展, 欧盟在此背景下所提出的 AI 法案对于各国法律框架中的责任归属问题具有重要的指导意义。该法案的目标在于确保人工智能系统与创新与安全之间的平衡, 同时明确界定责任的归属。通过利用政策评估框架, 可以深入分析该法案在责任归属方面的具体条款和应用效果。

AI 法案对于不同风险等级的人工智能系统进行分类, 以便于制定相应的法律责任条款。该法案将人工智能系统划分为高风险和低风险两类, 其中高风险系统需满足严格的合规要求, 包括技术文档的编制、风险评估和监控机制的建立。^[10] 这种分类不仅关注了 AI 系统的潜在危险性, 更加突出了在责任归属方面的明晰性。例如, 如果一个高风险的自动驾驶汽车因技术故障导致交通事故, 法案将界定制造商在此案例中的潜在责任, 从而提升企业在技术开发阶段所承受的法律义务。

(三) 我国的相关立法与案例

我国在国家层面关于人工智能治理的政策方针具备了相对全面的体系, 主要体现在《新一代人工智能发展规划》《国家数字经济发展战略纲要》等重要文件。通过政策框架的分析, 可以发现国家层面的立法者正在探讨以何种形式将现行法律制度适用于人工智能的责任归属问题。这一过程不仅涉及到对现有法律条文的解读能力, 更涵盖了对新兴技术所带来的法律空白的填补工作。以《新一代人工智能发展规划》为例, 该文件提到要加强人工智

能的伦理治理，确保技术与法律法规的协调。

而在地方立法的具体案例中，可以观察到不同地区因其经济发展水平、技术应用场景及法律传统的差异，展现出各异的立法策略。例如，某东部沿海城市针对无人驾驶技术的推广，出台了《智能交通管理办法》，明确规定了自动驾驶汽车在事故发生时的责任划分。这项法规通过细致的条款设定和责任主体界定，为地方道路交通管理提供了法律依据。同时，该法规结合了事故数据的分析，评估了潜在的社会风险，确保技术应用的安全性。

四、国际合作与人工智能责任归属的前景

（一）国际组织的作用与影响

在当今全球化进程加速的背景下，人工智能的迅猛发展为国际法带来了前所未有的挑战与机遇，联合国作为全球治理的重要机构，肩负着制定和协调人工智能相关国际规则与标准的责任。具体而言，联合国不仅在推动各国间的对话与合作方面发挥了引导作用，还通过其附属机构和专项委员会积极构建多边治理框架，努力应对人工智能所引发的伦理、法律以及社会问题。

联合国在人工智能治理中的举措体现在其促进国际法律框架修订的倡导上。例如，联合国国际法委员会已开始审视现有的国际法律体系如何适应快速发展的 AI 技术，其研究重点包括 AI 决策过程透明度及其责任归属问题。根据 2019 年联合国大会通过的有关决议，联合国的各项机制正在为跨国界的法律适用问题提供指导与建议，确保在全球范围内对 AI 进行有效监管，推动建立全面且一致的法律规则。

在全球化的背景下，人工智能的迅猛发展对国际经济体系及法律框架的构成产生了深远的影响。在此背景下，世界贸易组织作为国际贸易领域的重要机构，其在推动人工智能相关贸易法则的建立方面扮演了至关重要的角色。这一过程不仅需要各成员国的积极参与与合作，还要求有针对性的规则制定，以应对技术变革带来的新挑战。例如，《关贸总协定》和《服务贸易总协定》等现有框架为新兴技术的贸易问题提供了法律依据，确保各国在技术进出口方面的公平性与合理性。

（二）多边条约的可能性

在人工智能日益渗透于各国社会各个领域的背景下，传统的国内法律体系面临着巨大的挑战与矛盾，尤其是在责任归属方

面。因此，建立一个有效的跨国法律框架显得尤为迫切。这一框架不仅有助于填补现有法律真空，还能够为跨国技术应用提供清晰的法律指导，保障各方利益。

现行的国内法律体系往往在面对人工智能这一新兴技术时显露出适用范围的局限性。各国法律对责任归属的定义及适用往往不一致，这可能导致在发生损害时，当事人无从寻求有效救济。

（三）各国间的协作机制

在全球化日益加深的背景下，随着人工智能技术的迅猛发展，国际社会面临着如何有效监管和分配人工智能责任的严峻挑战。各国间的双边与多边合作模式在这一领域显得尤为重要。双边合作模式能够为国家间建立紧密的合作关系提供有利平台。例如，某些国家通过签署专门的“人工智能责任协议”，实现数据共享与标准统一，从而在从事共同的研究与开发时能够清晰地界定责任界限。

相较于双边合作，多边合作模式则展示了更广泛的国际协商机制，尤其是在面对跨境人工智能应用所引发的复杂法律问题，显得尤为关键。以“人工智能全球治理倡议”为例，该倡议凝聚了一系列多国政府及相关利益方，通过建立共同的框架与规范，助力各国在人工智能责任归属问题上达成共识。多边机制不仅促进了技术层面的信息交流，更加强了法律层面的适用性与执行力度，使得各国在行使监管职能时更加有效。

五、结语

传统的法律原则在面对人工智能的自主决策机制、算法复杂性与数据处理能力时，表现出显著的局限性。通过对自主决策与辅助决策 AI 系统的界定，我们发现，前者的责任划分更加复杂，涉及到制造商、开发者与用户等多个主体，而后者则通常难以推脱由人类决策者承担主要责任。在不同国家和地区，针对人工智能的责任归属也存在一定的差异性，反映各国在推进科技与法律协调发展方面的努力尝试。

因此，在未来的法律制度建设中，各国立法机构、监管部门与行业协会应加强合作，形成跨国责任网络，通过相互借鉴与协调，确保在人工智能愈发普及的背景下，责任机制能够正确适应技术发展的新要求。

参考文献

- [1] 韩雨潇. 生成式人工智能作品版权归属问题研究 [J]. 扬州大学学报 (人文社会科学版), 2023, 27(04): 52-64.
- [2] 李享, 罗天宇. 人工智能军事应用及其国际法问题 [J]. 信息安全与通信保密, 2021, (01): 99-108.
- [3] 丁继珂. 人工智能机器人新闻作品著作权归属问题的探讨 [J]. 中国新通信, 2021, 23(24): 87-89.
- [4] 邓义伟, 邓新军. 网络暴力中平台责任的刑法问题探究 [J]. 江西警察学院学报, 2023, (06): 100-109.
- [5] 夏凉, 周鼎坤. 经济犯罪基础问题之反思——以法律责任归属为核心 [J]. 经济刑法, 2022, (00): 3-18.
- [6] 王华平. 自动驾驶汽车的责任归属问题研究 [J]. 人民论坛·学术前沿, 2021, (04): 40-48.
- [7] 房慧颖. 人工智能犯罪刑事责任归属与认定的教义学展开 [J]. 山东社会科学, 2022, (04): 142-148.
- [8] 孟歌. 人工智能的道德责任归责问题探究 [J]. 湖北经济学院学报 (人文社会科学版), 2022, 19(07): 94-98.
- [9] 沈伟, 赵尔雅. 数字经济背景下的人工智能国际法规制 [J]. 上海财经大学学报 (哲学社会科学版), 2022, 24(5): 123-137.
- [10] 张胜飞. 数据治理视阈下人工智能的国际法规制探讨 [J]. 互联网天地, 2023(12): 29-33.