论强人工智能的刑事责任

陈洋

青岛科技大学 法学院法律系，山东青岛,266061;

摘要：随着人工智能技术的迅猛发展，其法律地位问题日益成为社会各界关注的焦点。在当前的弱人工智能阶段，人工智能主要被视为人类的辅助工具，尚未具备独立承担法律责任的能力。然而，当我们展望强人工智能时代，如果人工智能能够自主决策并产生重大影响时，是否应该赋予其刑事责任主体地位便成为了亟待解决的问题。这不仅是对现行法律体系的重大挑战，也是科技伦理与社会责任的交汇点。因此，考虑科技发展的进程，展望人工智能与刑事责任主体的未来，赋予人工智能刑事责任主体资格的可能性逐渐显现。

关键词：人工智能；法律人格；刑事责任能力

DOI：10.69979/3029-2700.25.08.065

1 问题的提出

随着人工智能技术的蓬勃发展，无论智能机器人还是生成式人工智能都早已渗透到社会的各个角落。华西口腔医院完成国内首例国产手术机器人辅助下的口腔颌面外科手术；武汉智能数字驾驶舱能够结合城市各项指标，快速生成分析报告并提供全天候、智能化的决策服务;Autodesk Generative Design 这款产品在产品设计和开发中利用生成式人工智能算法探索解决方案的所有可能变体快速生成满足重量、强度、材料使用和成本等标准的设计替代方案。这些实例无一不彰显了人工智能科技领域的迅猛进步与日新月异的发展态势。然而，这一进程并非全然乐观，伴随着人工智能的广泛应用，也浮现出其潜在的负面影响。2016 年特斯拉自动驾驶汽车不幸酿成了全球首例由自动驾驶系统直接导致的致命车祸；紧接着在 2017 年，俄罗斯军方更是推出了一款能够借助人工智能技术自主识别目标并判断是否发起攻击的机器人枪械系统，这一系列事件引发了深刻的法律反思。当人工智能涉足刑事案件时，应由谁来为这些人工智能的行为承担刑事责任？现行的刑法体系又是否足以适用于这些非传统主体？如何对人工智能进行定罪量刑？现行刑法面对这些问题新问题似乎无从下手，为了避免法律滞后性所带来的弊端，应对人工智能所带来的法律问题，应在综合考虑人工智能的危害之后，根据强人工智能的特性，设置一套专门对强人工智能的进行处罚刑罚处罚体系。

2 强人工智能取得法律主体地位的法理基础

鉴于强人工智能作为非生物实体的特性，我们探讨其是否有可能成为承担刑事责任的主体。这一探讨的核心在于从自由意志的角度出发，分析强人工智能是否能被赋予法律上的主体地位资格。同时，如果确实存在这种可能性，我们还需进一步探究应采用何种形式的法律人格来界定其在法律体系中的地位。

2.1 强人工智能与自由意志

自由意志，作为法律主体的核心要素，其存在与否成为了人工智能是否能被赋予主体地位的争论焦点。康德认为，在现象界（即我们所能经验到的世界）中，一切事物都遵循自然因果律，但人类却拥有一种超越这种必然性的能力，即自由意志。他提出了“物自体”的概念，认为物自体是现象界的自由因，而人的自由意志正是由此而来的一种先验的自由。这种先验自由是超出自然因果性之外的，它不能被认知，但可以被设定。黑格尔在继承和发展康德思想的基础上，进一步探讨了自由意志的问题。他将自由视为绝对精神的体现，认为个体的自由是在绝对精神自我发展和实现的过程中获得的。在黑格尔看来，自由不是孤立存在的，而是与整个宇宙的精神发展紧密相连的。康德和黑格尔都试图回答自由意志的来源，康德通过引入“物自体”和先验自由的概念，为自由意志的存在提供了理论基础；而黑格尔则通过绝对精神和个体自由的关系，进一步阐述了自由意志的发展和实现过程。但康德与黑格尔并未对自由意志的具体来源进行了深入的探讨。人脑是否是自由意志的唯一具体产生来源？这个问题至今仍困扰着哲学界，引发了无休止的争议。无可置疑，自由意志与人脑之间存在着深刻的内在联系，它紧密关联着个体的思考过程、决

策制定以及行动能力。但问题在于，随着科技的飞速发展，强人工智能在智能领域不断突破，甚至在某些方面超越了人类。它们在深度学习和进化的过程中，是否可能会孕育出类似于人类的自由意志？ 因此，我们不能简单地否定强人工智能产生自由意志并成为刑事责任主体的可能性。

2.2 强人工智能的法律人格

在探讨强人工智能是否应具备法律人格时，学者们提出了多种观点。首先，欧盟提出了“电子人”的概念，认为应赋予高级自主智能机器人独特的法律地位，使其成为介于自然人和法人之间的新型法律主体，享有特定权利并承担相应义务。其次，袁曾主张人工智能应具有有限的法律人格，因为其行为能力有限，需要特殊的法律规范和侵权责任体系[1]。杨清望和张磊则提出“次等法律人格说”，认为应从人类权益出发，通过法律拟制技术赋予人工智能次等法律人格，并建立登记备案机制和法律责任体系[2]。张绍欣则提出了“位格加等说”，认为可以通过提升智能机器人的法律位格，使其类比于自然人，但又不完全相同于自然人和法人[3]。尽管这些观点在赋予人工智能法律人格的程度和方式上存在差异，但它们都基于人工智能技术的快速发展和法律的预见性需求，旨在为人工智能确立法律地位。然而，也有学者持反对意见，认为在当前弱人工智能阶段，人工智能的责任归属最终指向人类，赋予其法律人格似乎多余。然而，随着人工智能技术的进一步发展，尤其是当人工智能达到强人工智能水平时，其自主意识和决策能力将显著增强，为了有效规制其行为后果，赋予其相应的法律人格将成为必然选择[4]。

3 强人工智能刑事主体证成

在强人工智能可能具有法律人格的基础上，还需要进一步从行为、、刑法对其处罚是否能达到刑罚的目的以及构建刑罚处罚体系等方面探讨人工智能的刑事主体资格。

3.1 人工智能的刑事主体地位

刑法意义上的行为是指能够被法律评价并可能受到处罚的行为。传统上，这些行为被认为是人的行为，但强人工智能的行为是否也应被视为刑法意义上的行为，这是一个新的法律问题。行为理论的发展历史包括因果行为论、社会行为论、目的行为论、人格行为论以及消极的行为概念，这些理论都对刑法上的行为有所界定。因果行为论将行为视为由意识支配的生理或物理过程，社会行为论强调行为的社会意义，目的行为论关注行为的目的性，人格行为论则认为行为受人格特质和心理机制影响，消极的行为概念则指能够避免而没有避免的行为。学者们通常将刑法上的行为主体限定为人，主要原因是排除非自由意志支配的行为和限于生命体实施的行为[5]。然而，法律主体不一定是生命体。强人工智能虽然不是有机体，但具有许多人类特征，如意识、认知和情感，通过不断进化，它们可能会获得媲美甚至超越人类的自主意识和决策能力。因此，强人工智能的行为可能符合刑法意义上的行为的特征，包括有体性、有意性和社会危害性，从而应纳入刑法的规范范畴。

3.2 强人工智能的处罚可能性

在探讨强人工智能是否可能承担刑事责任时，我们可参照单位犯罪的处罚可能性。犯罪主体，界定了哪些实体能因其危害社会的行为而负起刑事责任。这些主体涵盖自然人及单位（诸如公司、企业等组织）。对于自然人而言，成为犯罪主体的门槛在于达到法定年龄并具备刑事责任能力，即能够认知并控制自身行为的后果。此外，特定类型的犯罪还附加了对行为人身份的特殊要求。反观单位，当其以集体名义涉足犯罪行为时，同样被视为犯罪主体。我国 1979 年颁布的《刑法》中，并未前瞻性地纳入单位作为犯罪主体的概念。然而，随着国家经济的蓬勃发展和市场环境的复杂化，单位犯罪的案例逐渐浮出水面并呈现增长态势，这无疑对社会的和谐稳定构成了新的挑战。单位成为犯罪主体可以说是时代进步和法律完善的产物。这一法律演进的趋势与强人工智能技术的迅猛发展在某些方面不谋而合。展望未来，当强人工智能技术达到前所未有的高度时，其自主性和决策能力日益增强时，强人工智能也面临着与单位犯罪同样的境遇，即如何界定强人工智能在犯罪行为中的责任地位？是否应将其视为独立的犯罪主体加以规制？关于单位犯罪的刑事责任分配，各国法律实践各异，主要体现为单罚制、双罚制和代罚制三种模式，其核心差异在于是否将法人与自然人视为独立且不同的责任主体进行惩处。具体而言，单罚制仅对法人施以刑罚；双罚制则同时追究法人和相关自然人的责任；而代罚制则仅限于自然人受罚。基于上述分析，笔者建议借鉴刑法中单位的刑事责任理念，将强人工智能纳入刑事责任主

体的范畴，与自然人和单位并列。在具体操作上，可参照我国现行的单位犯罪双罚制原则，即在追究强人工智能犯罪的同时，也应对其背后的直接责任人——如研发者、销售者或使用者——实施相应处罚。这样的设计旨在确保当强人工智能的自主决策与实际行为符合刑法规定的犯罪构成要件时，能够有一套合理的机制来对其进行法律评价和责任追究。然而，在实施过程中需特别注意区分传统针对自然人的刑罚方式与适用于强人工智能的新模式。对于涉及复杂犯罪形态的情况，若强人工智能的行为既体现了其自身的意志又掺杂了自然人的操控因素，我们应采取综合考量、分别定责的策略，既要追究强人工智能的责任（尽管这在当前技术和法律框架下尚具挑战性），也要对相关的自然人依法严惩，以实现法律的公正与效率。[6]

4 强人工智能的刑法规制

在确认强人工智能具备刑事责任能力，并能承担相应责任的前提下，我们深入探索如何针对强人工智能适用刑法规范，以法律手段进行有效管理。鉴于强人工智能与自然人及单位在本质上的显著差异，构建其刑罚体系时必须采取独特的路径。在我国当前的刑法体系中，刑罚的种类主要划分为财产刑、自由刑、资格刑以及生命刑这四大类别。然而，在强人工智能的语境下，这些传统刑罚方式面临诸多不适用之处：首先，强人工智能既不具备对财产权的实际需求，也不依赖于财产来维持其运作功能，因此，诸如罚金或没收财产的处罚手段对于它们而言是毫无意义的。其次，鉴于强人工智能并不参与政治生活领域，所以剥夺其政治权利这一惩罚措施同样无法适用。再者，强人工智能的“行为”源于人类的设计与编程，缺乏独立的人身权和人身自由权概念，故传统的限制自由型刑罚（如管制、拘留、有期徒刑等）均不适用。最后，鉴于强人工智能从根本上缺乏生命特征，死刑这一极端且旨在剥夺生命的刑罚手段自然无法对其执行。面对上述种种挑战与限制，我们迫切需要构建一套专门针对人工智能的刑罚体系。在遵循罪责刑相适应原则的同时，也要充分考虑到成本控制的因素。基于这些考虑，我提出以下针对人工智能的刑法措施：一是算法调整与优化，通过修改违法人工智能系统的核心算法，预防其再次实施违法行为；二是系统程序干预，包括删除或修改关键软件部分，以此影响其行为能力，类似于对强人工智能进行“精神制裁”；三是功能与能力限制，利用技术手段限定强人工智能的活动范围或禁止特定功能，减少其再犯风险；四是技术隔离与禁用，对于严重违法的强人工智能系统，采取网络隔离乃至永久禁用的严厉措施；五是物理销毁与程序破坏，当其他方法不足以消除其潜在威胁时，考虑对其进行彻底的物理与逻辑摧毁。

5 结语

刑法理论是随时代发展而不断发展的产物，并不断解决时代所产生的新问题。科技发展的速度远远超过人类最乐观的预测与想象。展望未来，在即将到来的强人工智能时代，这些强人工智能在超越人类预设的设计与编程框架后，极有可能萌发出自主意识，进而依据其萌发的自由意志，涉足犯罪行为的领域。因此，我们有充分的理由将此类强人工智能视为刑事责任主体，并采用特定的方式对其进行制裁。唯有如此，我们方能在科技发展的浪潮中保持清醒的头脑，既充分利用人工智能带来的无限可能，又有效防范其潜在的负面效应，更好地服务于社会的整体福祉与进步。

参考文献

[1]袁曾.人工智能有限法律人格审视[J].东方法学,2017,(05):50-57.
[2]杨清望、张磊.论人工智能的次等法律人格.载《中国法理学会 2017 年年会论文集》.
[3]张绍欣.法律位格、法律主体与人工智能的法律地位[J].现代法学,2019,41(04):53-64.
[4]刘瑞瑞.人工智能时代背景下的刑事责任主体化资格问题探析[J].江汉论坛,2021,(11):105-110.
[5]张明楷.刑法学[M].第六版.183-185.
[6]王燕玲.人工智能时代的刑法问题与应对思路[J].政治与法律,2019,(01):22-34.

作者简介：陈洋（2001 年-），男，山东聊城人，硕士，青岛科技大学法学院法律系，主要研究方向为刑法学.