

Doi:10.11835/j.issn.1008-5831.fx.2021.04.001

欢迎按以下格式引用:孙道萃.人工智能刑法主体地位的积极论——兼与消极论的答谈[J].重庆大学学报(社会科学版),  
2022(4):216-229. Doi:10.11835/j.issn.1008-5831.fx.2021.04.001.



Citation Format: SUN Daoctui. Positivism for AI subject in criminal law: Also answer to the negativism[J]. Journal of Chongqing University  
(Social Science Edition), 2022(4):216-229. Doi:10.11835/j.issn.1008-5831.fx.2021.04.001.

# 人工智能刑法主体地位的积极论

## ——兼与消极论的答谈

孙道萃

(中国政法大学 国家法律援助研究院,北京 100088)

**摘要:**智能技术及其应用的迅猛发展,对以人类的主体性地位为根本存立前提的当代法律体系形成强烈冲击,智能主体的刑法地位成为关键问题。目前,消极论和积极论处于相持阶段,加速了现行刑法理论体系的知识变革进程。“人造物”等不同程度的消极论,以现行刑事法律制度难以直接接纳及其所裹挟的潜在的刑事风险等为由,虽有其可取之处,但其逻辑机理正是过度释放人类中心主义的预设效应。完全按照人类中心的立场进行审视会得出无法调和的结论,也显示了智能技术应用的工具属性被深度放大之倾向,但上述消极事由不尽然合乎规制人工智能犯罪的现实需要,也与法律主体制度的演进规律不符,甚至会压制刑法主体的自主进化。“电子人”等不同形式的积极论,率先走出绝对的人类中心主义及其观念束缚,以发展的眼光看待人工智能与刑法的互动及其动向,更加契合了人工智能犯罪的发展动态。而意志自由、刑事责任能力、道德伦理规则、智能程度与智能主体类型、刑事责任的客观存在、算法的特殊地位与意义等既有的知识框架与新的要素累积,意味着人工智能可以在发展的刑法体系中实现衔接与契合。经由不真正完全背离刑法原理的认识扭转,亦供给了学理层面的支撑。而且,遵循功利主义的理路,以及在一系列有关人工智能主体方面的先行立法之做法的引领下,应分阶段、类型化、动态化厘定智能主体的刑法地位,使其在法律拟制的路径上可以继续延续和拓展。在人工智能的刑法主体地位得以澄清下,有必要接受智能主体可以享受有一定的新兴权利之现实情状,但应具体地确证权利类型等具体内容。在现阶段的技术代际之策动下,智能主体的权利内容与范围目前是限制性的,无法采取与“人”对等的保护策略,保护方式也应有别。这种“降维”保护有其现实合理性,通过积极制度设计与规则配置,能够最大限度策应人工智能主体的刑法地位之更迭态势。

**关键词:**人工智能主体;刑法地位;消极论;人类中心主义;积极论;功利主义;权利逻辑

**中图分类号:**D914;TP18

**文献标志码:**A

**文章编号:**1008-5831(2022)04-0216-14

**基金项目:**国家社会科学基金重大项目“健全社会公平正义法治保障制度研究”(20ZDA032);国家社会科学基金青年项目“人工智能时代的刑法前瞻与应对研究”(18CFX041)

**作者简介:**孙道萃,中国政法大学,Email:sundaocui2012@126.com。

## 一、未竟的知识颠覆

人工智能技术及其应用的增速,使目前仍被认为是“人造物”的“智能主体”<sup>①</sup>(智能机器人、人工智能体等,下同),因(日益/高度接近)“类人”的智能程度与行动能力等,开始被贴上“主体”的新标签,使人的主体性地位开始遭受双面震荡:人继续作为“造物主”,人的主体性意识、人的主体性发展均受益;但也严重冲击人的主体性地位、能力、道德伦理,制约人的主体性发展。这种辩证观察显示了智能主体的到来及其地位的升级,既是人的主体性之延续,也加速销蚀人的主体性地位。在这场拉锯战中,人的主体价值观显现出陷入悖论的迹象。“人”与“智能主体”的法律地位之争是隐性效应,并且会弱化“人创制法律”的认识观、世界观与价值观以及法律制度。

目前,理论上关于智能主体的法律身份与地位的看法不一。实践中存在一种强烈的声音:应用中的智能机器人以及更高级的智能主体,应被认为是像“人”的法律主体,并以此构建智能主体格局下的主体范式、关系范式、权利话语体系。但这种前瞻性的“建构”,也被认为是人类中心主义<sup>[1]</sup>。不仅暗示所谓“智能主体”必然缺乏道德主体资格,也使智能主体通过法律设立的权利“形同虚设”或无法真正行使。毕竟智能主体是“人造物”,具有鲜明的“工具属性”,是为保护人类的利益而存在的,非真正的法律主体权利,与人的主体性本质与要求完全相斥。智能主体能否成为“(类)人”?关键在于确认“(类)人性”(人的社会现实性、历史文化本性等)。这是智能主体法律地位之“难题”的核心和实质。甚至智能技术的“工具性意义”与“机器人是否会威胁或代替人类”的主体性危机相比,是更实际与真实的当下担忧。但消极论脱离真实的诉求。最新的探索已走在前面<sup>②</sup>。

理论观望还在持续中,但智能技术引发的“人类主体性危机”已悄然蜕变和加剧,甚至将成为人类社会最大的“人造危机”。当代法律体系、特别是刑法体系,正面临已经迫近的根本性制度危机。目前,关于智能主体的刑法地位问题,主要分为消极论和积极论,两方纷争不止,但未能提出面向未来的建设性方案;关于智能主体的权利问题,则疏于讨论,导致研究深度与广度不足。应当从理论与现实层面,对智能主体的刑法地位课题进行回顾、反思并再构。这不仅关系人的主体性地位与刑法的历史定位,也关系当代刑法的生死存亡与功能发挥。特别是对于人工智能刑法的前沿问题研究,应容忍必要的学术“想象力”,以此稀释浪漫主义与虚无主义法律的偏一风险,进而推动人工智能刑法理论研究趋于理性与可行。甚至在一定程度上,主动畅想“人工智能刑法学”也是有依据和意义的前沿探索<sup>③</sup>。基于此,对于人工智能的刑法主体地位及其相关问题的讨论与归结,不宜完全只停留于实定法的既定框架,而需植入必要的前瞻性,以面向未来的姿态,开启积极的建构性研究。

## 二、消极论:解析与回应

### (一) 消极论的主要事由与解构

#### 1. 消极论的主要事由

在法理上,智能主体的法律人格问题颇具争议。在民法领域,也存在不同看法<sup>[2]</sup>。客体说包括产品责任说、类推适用动物说等。主体说包括代理说、电子人格说等。折中说认为,智能主体的民法地位

<sup>①</sup>智能主体是现实功能主义立场下的称谓,智能机器人及产品等是人类中心主义立场的称谓。但此不做区分。在现阶段,对智能主体诸问题的前瞻讨论,不免仍然需要借助现行刑法话语体系,如刑事责任能力等概念。

<sup>②</sup>国际社会对于智能主体的法律地位目前呈现为松绑的积极姿态与做法,并体现在立法与实践方面。

<sup>③</sup>参见:孙道萃《人工智能刑法研究的反思与理论进阶》(《学术界》,2021年第12期第64-76页)。

属于“人工类‘人格’”,属于物的范畴,是权利客体,而非民事主体。在智能技术与智能主体迅猛发展,智能程度不断递增等情况下,客体说具有显著的“人类中心主义”倾向,其滞后性会逐渐显现。主体说的立场较为激进,但未能考虑人工智能技术应用尚处于弱人工智能时代的现实。折中说所指的“人工类‘人格’”,虽提及“人”的主体因素,但归根结底是“人造性”,结论实质上是“物”,而非独立的法律人格体。按此逻辑,AI不存在法律主体的逻辑问题。这些看法对刑法领域的理解有一定的辐射作用。

目前,刑法领域的消极论主要表现为:(1)全面消极论。人工智能缺乏对外在行为及其社会属性的认识、规范评价意义上的意志自由、理解并遵从法规范的能力、对不法的认识可能性与抵制意志,不能按照法人进行类比<sup>[3]</sup>。(2)刑法对象论。人工智能具有绝对的客体性。人工智能犯罪的可归责主体,只能是作为算法安全社会关系主体的“自然人—法人”二元主体<sup>[4]</sup>。按照塞尔“中文房间模型”认知理论,人工智能不具备刑法意义的辨认能力与控制能力,应当视为犯罪工具或犯罪对象<sup>[5]</sup>。(3)刑事主体资格条件的缺乏论。根本不可能在认知、意识、情感方面设计与人类相媲美的具有人类心智的人工智能主体,更不能实施犯罪与承担刑罚,不是适格的受罚主体与刑事责任主体<sup>④</sup>。(4)非法定主体论。人工智能是高级工具,无法被赋予“权利义务统一性”,不具有刑法意义上的辨认、控制能力,不能实现报应、预防等目的。犯罪主体只能是自然人或单位<sup>[6]</sup>。

## 2. 消极论的解构

消极立场的理由,基本上可以概括为:在现行刑法中,“人的主体性地位”及其专属性、不可替代性、不可动摇性、天然的法定性等牢不可摧。(1)意志自由的缺失与刑事责任能力不匹配。从生物学看,人的刑事主体地位的形成,是建立在意志自由与刑事责任能力的基础之上。前者是哲学层面的人的意志自由决定,后者是规范层面的行动自由,共同决定是否可以承担刑事责任。但讨论智能主体的意志自由与刑事责任能力,面临尴尬的技术局限、语境受限以及规范缺失等问题。智能主体的意识、意思或者对行为的违法性认识等问题,与人的主观方面存在本质的区别。和法人相比,虽有相似之处,但差异很大。智能主体暂时不具有刑法评价意义上的行为认识与控制要素的自主性,不具有人的目的理性所支配的可答责基础,欠缺作为刑事责任主体的本质要素。这些因素共同决定无法以“人类中心主义”的逻辑来讨论智能主体的法律地位问题。(2)行为理论的不对等。行为是最基础的要素,是刑法评价的直接对象。智能主体的行为,是在其相应的意识、意思以及能力的作用下,自主决定是否实施,并应对该行为负责。只是相比于人类的行为逻辑以及所实施的犯罪行为,智能主体的实行行为表现为“算法”及其运行的新情状,二者有明显差异。(3)与刑事制裁不吻合。智能主体不能直接适用目前刑法规定的刑罚措施。而确认智能主体的刑法地位,首要目的是追究刑事责任。与之相适应的刑事制裁措施尚处阙如,容易使该目的落空。智能主体缺乏“类人”的可罚性特质,使其不能成为刑法主体。(4)与基本理论不对恰。智能主体目前不是法定的刑法主体类型。即使确认智能主体的身份,但由于理论体系等未能及时更新或无法同步确立,必然会出现深度的理论对冲。(5)定罪和量刑的紊乱。按照现行刑法规定,在定罪量刑活动中,很多规定无法适用于智能主体,即使定罪的对象、依据等内容失效,也使量刑情节适用等活动无法进行。例如,对智能主体的归责,目前只能主要考虑损害结果,却模糊人工智能作为主体的主观恶性与行为时的故意、过失之分,容易导致刑罚圈的恣意扩大与具体刑罚适用的失当。(6)法律适用的不对接。智能主体是否存在主观恶性与人身危险性等内容尚无法确定,刑罚预

④参见:张镭《人工智能体可罚性辩疑》(《南京社会科学》,2018年第11期82-87页);叶良芳《人工智能是适格的刑事责任主体吗?》(《环球法律评论》,2019年第4期67-82页)。

防措施难以有效配置;无法或者无需严格区分故意与过失及其定罪量刑要求,难以体现故意与过失的区分意义;自首、立功等量刑情节亦无法适用,犯罪特殊形态如犯罪预备、中止、未遂等情形,套用在智能主体上非常生硬。(7)刑事责任的转嫁性。肯定智能主体的刑法地位,则可以独立承担刑事责任。智能主体与研发者等人类主体相互独立或剥离,在实践中可能出现不同行为的交错与重叠。不排除可能出现智能主体滥用技术优势并转嫁风险给人类个体,逃避刑事责任的追究。

## (二)人类中心主义的理性消解

### 1. “人造物”观与工具思维的叠加

消极论遵循“人类中心主义”思维。所谓人类中心主义,是以人为宇宙中心的哲学立场。核心内容是一切以人为中心,一切以人为尺度,一切从人的利益出发,最终为人的利益服务<sup>[7]</sup>。在自然人、自然人集合体等组成法律主体格局下,智能主体缺乏“人的生命属性”,无法与“法人”直接类比,智能主体尚不足以取得独立的法律地位。从“人造物”的主客体关系看,智能技术的发展程度无法改变作为人类劳动创造的“技术客体”,主张是“物”有其客观理由。由此,基于智能技术应用的“工具属性”,“物”的角色才是智能主体进入法律规范的身份,是与法律交互的前提。这反映人类的“创制”地位对界定智能主体的“映射”效应,“工具属性”使智能主体无法获得独立的法律地位。但是,法律主体及其范围呈扩张趋势。故,积极说才是应然选择。可以参照“法人”之做法,有限度地运用“法律拟制”之立法技术,赋予智能主体的法律能力与地位并无法律技术障碍。

### 2. 人类中心主义过度化的“加持”

从绝对到相对的人类中心主义,也是认识观与实践观的理性迭进<sup>[8]</sup>。完全按照人类逻辑消极看待智能主体的法律地位,不免存在认识论上的误区。人工智能并未对法律基础理论、法学基本教义提出挑战,受到挑战的只是如何将传统知识适用于新的场景<sup>[9]</sup>。在人类中心主义导向下,不仅否定当前讨论智能主体刑法地位问题的真实性与必要性,也对智能主体这一元问题提出根本性的质疑。在方法论上,当前对智能主体的“法律地位”之判断,套用人类中心主义的固有逻辑,用“符合人类的法律主体判断标准”而非智能时代的特有逻辑来判断智能主体的法律地位问题。这在认识论上隐藏立场失当的隐忧:根据人应当具备的认知、道德、意思等具体理性能力与特质为内容所形成的判定标准体系,对智能主体进行实质与形式的双重“匹配”,是人类中心主义主导下的“认识”。可以“顺理成章”推导出智能主体缺乏生命价值、生命脆弱、感知痛苦等人的具体特质,与人类应当具备的一些判定标准体系“绝缘”,故不能作为法律主体。对于智能主体法律地位的判断,应当以算法、深度学习等要素为基础,根据智能社会对法律人格判定的专属标准体系,设定独立于“人”的具体特质加以判断。只有建立既不完全脱离人类的现有法律体系,但又符合智能时代的特定发展需要的“判断规则”,才能在过程与结论上,更恰当地对智能主体法律地位作出符合实际的判断。

### 3. 对否定论的审正

针对消极论的质疑与担忧,应当澄清以下问题:(1)作为智能主体存在基础的道德伦理正在形成。人的生理性与社会性,共同决定“人之为人”。人的社会性表现为社会规则担当起维系社会发展的重任,同时以道德伦理作为最重要的基础。智能主体是否具有道德伦理,是决定其法律属性与刑法地位的前提与基础,并与自然人的关系、法律属性、法律制度等相互映照。目前对于智能主体的道德伦理问题更多的是担忧。特别是在“人类中心主义”的思维下,套用人类建立的规则体系,几乎没有肯定智能主体拥有道德伦理的空间。然而,从国际社会的动态和一些发达国家的做法看,对智能技术伦理的讨论更趋于乐观,也形成一些共识和积极成果。逐步建立与智能时代相适应的道德伦理规则体系并非不

可能,也为确定智能主体的刑法地位预留制度空间。(2)法律主体的历史变迁规律。智能主体已然不同于普通机器,存在赋予其法律人格的正当化依据。将没有生命体的智能主体,作为法律主体,具有技术可能性和现实需求,不会实质性地背离法学基础理论。超越程序设计与编制范围,基于独立的独立意志进行活动的智能主体完全可能出现,使其具备法律人格的基础并成为法律主体。这是对“人”的概念的增补而非致命的冲击。法律中“人”的概念并非僵化不变。立法上赋予智能法律人格与权利义务并非完全不可能。世界范围内已有先行的立法活动。(3)法律拟制的刑法功能与意义。从法律发展的规律看,法律是以人类为规制对象。经过长期实践与发展后,天然地不适用于人工智能。目前,人工智能不具备也无法模拟人的全部行为与活动。将智能主体拟制为法律主体,直接降低人的主体性地位,也可能使“人造物”的概念出现异化。但从人作为法律主体的形成逻辑,法人作为犯罪主体的历史规律与法律逻辑看,基于人工智能技术的发展与算法规则的发达,以及高度智能化的特征等因素,智能技术的理性延伸具备相应前提,并会赋予行动能力的基础。完全寄希望于既有法律主体规定的有效性,显然是偏于保守。法律拟制可以依法赋予智能主体的法律地位,与刑法中的法人无异。而且,法律拟制的方式具有人的可控性特征,使人对智能主体的刑法地位之预设与调整处于安全的状态。(4)功利主义的司法意义延展。智能主体的刑法地位问题,是由新兴犯罪态势客观催生的,而不是纯粹主观臆测的猜想。为了确保刑法目的与刑法机能的正常运行,有效规制智能技术及其应用过程中形成的刑事风险,应从根本上谋求更合理的解决方案。从功利主义的现实角度看,智能主体刑法地位的阙如是当前困扰的首要症结。对智能主体刑法地位予以清晰界定,从源头上回答智能主体是否可以独立实施犯罪行为,从功能上明确智能主体是否可以承担刑事责任,这才是理性的风险控制观。

### 三、积极论:匡正与立基

积极接纳智能主体的刑法地位是新的供需关系所使,关切到人类生存与发展的重大问题<sup>⑤</sup>。在功利主义的引导下,通过立法逐步确认智能主体刑法地位具有可行性基础与现实必要性。

#### (一)积极论的主要情状与展开

##### 1. 积极论的几种样态

理论上形成以下看法:(1)完全积极论。与现有刑事责任主体的差别、无法证成“强智能机器人不具有自由意志”、处罚上的不能、确立强智能机器人刑事责任主体地位引发的推卸刑事责任等不成立。强智能机器人具有辨认能力和控制能力,是刑法主体<sup>⑥</sup>。(2)相对积极论。智能代理可以成为道德代理。人工智能具备法律人格是成为犯罪主体的基本条件,可以通过智能代理获得犯罪主体资格的条件。而且,人工智能具有可罚性<sup>⑦</sup>。(3)一般积极论。弱人工智能引起刑法质变的可能性仍然抽象。如若无法引起刑法的突变,“人工智能的刑法规制”极可能是伪命题。只有假想的超级人工智能,才可能导致刑法发生突变。但突变后的刑法,显然已不再是当下的刑法<sup>[10]</sup>。(4)功能发展论。人工智能的“人格化”具有方法论依据、智能性的本质要素、法理基础等。人工智能作为“人”会冲击传统刑法教义

<sup>⑤</sup>这里暂不深究“人之为人”“人是世界主体”等哲学问题及其旷世纠葛,侧重从法律上探讨现实可能性。

<sup>⑥</sup>参见:刘宪权《人工智能时代的刑事风险与刑法应对》(《法商研究》,2018年第1期3-11页);刘宪权《对强智能机器人刑事责任主体地位否定说的回应》(《法学评论》,2019年第5期113-121页);刘宪权《智能机器人工具属性之法哲学思考》(《中国刑事法杂志》,2020年第5期2034页)。

<sup>⑦</sup>参见:彭文华《人工智能的刑法规制》(《现代法学》,2019年第5期135-151页);彭文华《自由意志、道德代理与智能代理——兼论人工智能犯罪主体资格之生成》(《法学》,2019年第10期18-33页);[英]拉塞·夸尔克、王德政《人工智能机器人的刑事可罚性》(《中州学刊》,2020年第10期48页)。

学,但有助于解决智能主体实施的“犯罪”及其疑难社会问题<sup>[11]</sup>。

在对人类中心主义进行一定的认识观修正后,积极论的意义在于:(1)有助于明确智能时代的犯罪诸问题。智能时代新型犯罪问题已经开始出现,主要以智能“工具型”“对象型”犯罪为主。智能时代的犯罪形态仍在发展,犯罪类型等基本问题尚不确定,影响对刑事责任的理解与认定。传统犯罪形态与智能时代的新型犯罪不尽相同。智能主体与现行刑法主体的实质差异是其缩影。明确智能主体的刑法地位,有助于认识和界定智能时代的犯罪这一前提问题,也对解决好刑事责任、刑罚等问题有积极意义。(2)可以解决智能主体的刑事责任之客观问题。目前,利用智能技术应用平台实施新型犯罪已经浮出水面<sup>[12]</sup>,也包括针对智能主体的犯罪与智能主体自主实施犯罪的情况。一律直接借助现行法律规定或理论体系的规制未必有效和可行。例如,在智能驾驶情形中,出现是否存在真正的驾驶主体、驾驶主体究竟是人还是智能系统以及背后的制造者等问题。发生了重大交通安全事故后,涉嫌产品犯罪、公共安全犯罪、重大监督过失犯罪,甚至故意犯罪等犯罪。传统犯罪规定不足以解决智能驾驶过程中的新型犯罪问题。而明确智能主体的刑法地位是解决问题的关键。(3)有助于智能主体与“人”的责任分担。智能主体的刑事责任,与“人”(研发者、设计者、经销者、所有者、使用者等)的责任分担相关。完全由“人”来承担替代责任显然不可取。不同智能程度的智能主体,其与人“共同”犯罪的主次关系变得复杂。合理确定智能主体的身份与地位是前提。

## 2. 积极论的规范分解

在积极论的前提下,对智能时代的意志自由、刑事责任能力等规范要素,应当采取更“包容”的接纳态度,与智能程度、算法发展等积极因素相互叠加,共同夯实积极论的成立基础与有益条件。为此,应阐明人工智能与现行刑法之间的融合度。(1)意志自由的有无问题。人的意志自由问题,主要表现为自主地决定自己的意志并实施行为。而其刑法学意义是指犯罪主体是在所有因素的综合下,仍可以自主地决定是否实施犯罪行为。意志自由是刑法主体的必备要素。基于“人类中心主义”及“工具属性”的语境,智能主体是人实施犯罪的新工具或方法,没有可以和人平等的“意志自由”,也就没有作为法律主体的独立性、自主性基础。但是,这种理解是狭隘的:一是忽视随着智能技术的发展以及智能程度的升级,可能出现完全智能的主体。从理论看,智能主体在认识能力和意志能力上,可能接近甚至超越人类水平,完全可能成为犯罪的主体,直接造成社会危害结果。二是简单固化智能主体的刑事风险,仅限于“人的滥用风险”,而忽视其他刑事风险类型。这是“工具属性”思维左右智能犯罪类型认识的具体反映。在初期阶段,由于算法水平等因素的限制,应用功能较为单一,仍属于“工具”的范畴,与传统的“产品”无异;在深度应用阶段,可能持续挑战人类社会的认知底线,自主性迅速增量;在完全独立应用阶段,智能主体的地位趋于巩固,所带来的颠覆性后果完全超出现有法律的约束能力。因此,讨论智能主体的刑法地位,应秉持发展性、阶段性的持续性修正理念。三是智能技术的“算法”规则及其运行是以“智能化”为本质特征,而智能技术的应用是为了“解放人的劳动力”。“工具属性”与“智能属性”之间的博弈关系,随着智能技术的发展,倾斜于后者的趋势显而易见。完全否定智能主体的“自主性”及其“意志自由”,从技术本身、应用范围、应用目的以及实际情况等方面看,都难以成立。(2)刑事责任能力及其判断。目前,人作为刑法主体所必需的具备刑事责任能力要素,是否可以“等值”输入到智能主体内,作为评价刑法主体资格的核心要件备受争议。刑事责任能力包括犯罪能力和承担刑事责任能力,具体由辨认能力与控制能力所组成。智能主体由程序设计和编制而成,不是具有生物学意义上的生命体。然而,智能技术的发展,会使智能主体无限接近“人”;与动物相比,智能主体具有辨认能力与控制能力;与普通机器人相比,智能主体可能超越人类设计和编制的程序产生自主的意识和意志。

从法律属性上,可以将智能主体定位为经程序设计和编制而成的,可以通过深度学习产生自主意识和意志的不具有生命体的“人工人”。如今,单纯将智能主体看成是“执行强制指令而无法独立作出判断的计算工具或编程”已不合时宜。为了使军用机器人比人类更具有人性,应当在机器人系统中设计“人工良心”,并公开征求机器人应当遵循的道德规范<sup>[13]</sup>。不仅可以逐步建立起独立的智能主体道德伦理体系,也可以基于智能程度和算法系统等能力条件,形成独立的思考与行动系统。(3)智能化程度与智能主体的类型划分。决定智能主体地位及其能力的首要因素是智能程度,可以根据智能程度来确定刑事责任能力程度。基于“行为与责任能力的同时性”原则,在方法论上可以将人工智能的智能化程度作为衡量“刑事责任年龄”的指标依据,判断人工智能体是否具备刑事责任能力。例如,智能水平分为很低、较高、很高,可以分别比拟自然人的完全无刑事责任、相对有刑事责任、完全有刑事责任的情形。间歇性系统异常或中病毒的人工智能体,可以按照类比精神病人或醉酒的人<sup>[14]</sup>。这种看法有其合理性。根据智能程度确定智能主体的类型以及刑事责任范围是目前较为可行的做法。据此,可以对不同智能阶段的智能主体,有区别地施以具体的刑罚。但前提是攻克技术难题,制定具有共识性、正当性的判断标准。目前,主要依靠研发者、设计者等“外部主体”,根据政策与现实情况,设计具有操作性的标准。但最终应当通过立法规定或发布司法解释,建立法定化、公开化、统一化的规范判断标准体系。(4)算法作为核心能力的自主优化升级。算法是智能技术获得优势地位的基础,算法规则、伦理以及运行过程等,是智能主体的“(数字)大脑”。目前,智能技术应用的认识论基础是“可计算化的认知”。它最终可以通过智能技术的应用方式,对人脑功能予以“完全仿真”模拟。从现行法理看,通过法律拟制智能“电子人”的主体地位,必须同时对智能算法的设计原理、与人(设计者和使用者)之间的互动模式,以及决策的自主程度等问题进行详细的研究、分类,并基于应用场景等的需要,完善算法的规则设计,才能具备可操作性。

## (二) 功利主义的功能演绎

### 1. “电子人”诸说与“去人类中心化”动向

人工智能具有自主性、主动性,是已非完全受人支配的行为客体,应当设定为法律上的“电子人”<sup>[15]</sup>。人工智能是介于人与物之间的客观存在,财产性而非工具性是其本质。应当以财产性人格为基础,将人工智能拟制为电子法人,解决责任分配问题<sup>[16]</sup>。在智能技术迅猛应用下,这些折中的积极论有相当的“分量”。

“电子人”等积极看法,肯定了智能主体的法律人格及其主体性能力,以满足实际需要。按照功利主义对现实合理性的侧重,应当合理地“去人类中心”的绝对标准,考虑将法律标准“等值”摄入,并遵循智能时代的规律与本质,以便更理性地看待智能主体的法律地位。也即:(1)从法理看,法律主体与法律关系、权利主体、义务承担者以及法律行为的对象,都有非常直接的内在关系。按照人类中心主义的立场,人类及个体的生理要素、意识能力和独立意志的心理要素、独立的社会角色的社会要素等,都是法律主体必须具备的条件。按照现行法律规定,智能主体暂时仍无法具备传统法律主体地位所需的基本条件;反而,“工具属性”更容易得出“物”的界定立场。然而,“人造物”等看法,脱离“主客分离”的主体识别规则,实质混淆法律关系的主体与客体之概念。人工智能产品目前可以认为是人类创造的产物,是法律关系的客体或标的。但是,智能主体未来可以是独立的法律主体。(2)仅从生物学角度否认智能主体的法律人格之逻辑似乎成立,但不能过度放大。立法者通过“法律拟制”的立法技术,赋予“法人”相应的刑法主体地位。相比之下,虽然临界点并未到来,但是随着智能主体的智能程度、自主性意识等不断提升,“传统法律主体格局”及其规范意义也不免会逐渐出现失效或失灵等现象。以

阶段性、发展性的思维审视“人造物”语境,将智能机器人作为独立的主体,可以形成自然人、法人(单位)、智能机器人的并列格局,有助于拓宽法律主体的内涵。这是对人类中心主义主导人类主体地位及其标准等问题进行“去中心化”的必然反映,也是社会主体地位及其法律体系根据社会时代变化而调试的正确做法。

## 2. 功利主义与法律拟制的契合

对于机器人的不道德行为,只有在具备“人工道德”的前提下,才可以被非难并承担相应的责任。在技术过渡期,“人类中心主义”与智能主体的道德体系有所“对冲”。参照人类法律体系的基本逻辑,对智能主体法律地位进行评判,主要涉及如何看待智能主体的道德伦理、智能主体与自然人的关系等新问题。当智能主体能够且应当拥有道德时,便存在可谴责的社会伦理基础,才能对其不道德的行为进行规范上的非难。赋予智能主体道德伦理能力的前提是其可以在“拟人”的道德判断情境中,作出与人类实质相似的决策过程,并会作出合法的行动。应当采用智能时代应有的思维与逻辑以及道德伦理等一般性社会关系准则与规范,通过法律拟制的方式,对智能主体的法律地位作出相适应的“等值”评判。高度模拟的情景不是完全的“人类”场景,而是由法律拟制而成并具有法律意义的智能伦理体系。

尝试肯定或承认智能主体专属的道德伦理问题有显著的意义,将赋予具有自主意识的智能机器人,拥有判断某种行为正当与否的基本标准及准则。而且,确认智能主体的法律地位,有助于解决智能主体与研发者、设计者之间的责任分配等问题。此乃功利主义的思考逻辑。在法律责任方面,欧盟赋予某些智能机器人以法律人格或“电子人格”,可以享有特定的权利和义务,可以追究法律责任。但是,如若单纯为了解释智能主体的行为效力和可以追究法律责任,便赋予其法律主体资格,这种遵循功利主义的做法,也未必具有目的的正当性。即便立法创设“电子人”,而算法规则、人工语言等相关技术无法跟进,“电子人”仍无法理解按照以人类方式表述的法律规则。对于这些困难和疑问,需要技术层面提供支持。

## (三) 立法动向与启示

### 1. 实践先行的参照意义

实际上,对于智能主体的刑法地位,域外与我国当前不乏一些先行的积极做法。(1)2016年5月与10月,欧洲议会法律事务委员会先后发布《就机器人民事法律规则向欧盟委员会提出立法建议的报告草案》<sup>⑧</sup>《欧盟机器人民事法律规则》认为,智能机器人已经开始具有自主性和认知特征,具有从经历中学习并独立自主作出判断的能力,可能实施侵害行为,须要承担法律责任。应当制定新的规则,人工智能主体应适用“电子人格”身份。欧盟委员会法律事务委员会向欧盟委员会提交动议并要求,将最先进的自动化机器人定位为“电子人”,赋予其“特定的权利和义务”;建议对智能自动化机器人进行身份登记,以便为其纳税、缴费、领取养老金的资金账号。(2)2016年,Google无人驾驶汽车在美国加州山景城测试时,与一辆公交大巴相撞而发生事故。美国公路交通安全管理局(NHTSA)表示,关于交通事故责任认定,谷歌驾驶系统可以被视为“司机”。(3)《合伦理设计:利用人工智能和自主系统最大化人类福祉的愿景(草案)》[国际标准化组织(IEEE)]指出,如果AI依靠人类的交互,实现新内容或发明创造,使用AI的人应当是作者或发明者。在知识产权保护上,与未借助AI进行的创作和发明的

<sup>⑧</sup>该报告第50(f)项建议:“从长远来看要创设机器人的特殊法律地位,以确保至少最复杂的自动化机器人可以被确认为享有电子人(electronic persons)的法律地位,有责任弥补自己所造成的任何损害,并且可能在机器人做出自主决策或以其他方式与第三人独立交往的案件中适用电子人格(electronic personality)”。

具有同等性<sup>[17]</sup>。(4)2017年,百度和博世共同开发一款自动驾驶汽车。交警认为该车的试驾涉嫌“违规驾驶”<sup>[18]</sup>。(5)2017年,沙特阿拉伯对机器人“索菲亚”授予“公民”身份,并向“她”发放护照。

## 2. 法律功利主义的立法导向

边沁认为,全部道德体系,整个立法体系,都建立在一个唯一的基础上:关于痛苦和快乐的知识。快乐和痛苦是关于这一主题的清晰观念的唯一基础<sup>[19]</sup>。这是法律功利主义的核心旨趣。功利主义按照势必增大或减少利益有关者之幸福的倾向,亦即促进或妨碍此种幸福的倾向,来赞成或非难任何一项行动。无论什么行动,其不仅是私人的每项行动,而且是政府的每项措施<sup>[20]</sup>。法律是有目的性的活动,立法是为了实现立法者的既定目标。边沁明确将法律作为实现功利主义的工具。“对于主权者,他在采纳法律过程中所考虑的目的,或外在动机,基于功利原则,仅仅是社会的最大福利”<sup>[21]</sup>。刑法立法必须遵循功利主义,在犯罪、刑事责任以及刑罚之间,寻求最优组合、最佳配置与最高效益。“政府的业务在于通过赏罚来促进社会幸福。由罚构成的那部分政府业务尤其是刑法的主题”<sup>[22]</sup>。促进人类的最大幸福,需从立法入手,尤其是需从刑法入手。

功利主义是关于法律制定的重要思想。为了达到通过立法增进人类幸福的目的,立法者应当运用道德算术,计算苦与乐,使法律草案的起草过程融入功利主义的元素。而针对立法草案规定的苦与乐的立法计算<sup>[23]</sup>,功利主义之于法律,最直接的启示意义与实践逻辑,便在于为立法者是否以及如何立法提供最重要的“权衡之术”。不仅包括价值取舍层面,也涉及立法技术的最优化标准。可以归纳其最核心的立法指导逻辑为功利主义应作为一项原则性、概括性、框架性的观念,引导立法者更主动、积极地通过立法满足变革需求。从方法论看,解决智能主体刑法地位问题,相对温和的措施是立足现有法律体系,激活扩张解释的司法张力;而相对激进的措施,是通过立法的方式直接加以确认或规定。应当以智能技术为基本前提,以智能应用为分析基础,对不同智能程度的智能主体予以类型化考虑,充分兼顾身份的一般性与特殊性、功能的现实性与前瞻性、价值的法定性与拟定性。分阶段、有层次、类型化地确认或规定不同智能程度的智能主体,在不同时期、不同应用情形、不同行为环境下的刑法主体属性及其能力。

## 3. 立法建言

在现阶段,可以考虑暂时将智能主体视为限制性的法律主体。既承认具有人的智慧特质之工具属性,又强调是可以做出独立意思表示的特殊法律主体。进言之,高度智能化的主体,虽具有工具属性的特质,但在一定程度上已经具有独立的自主意识,应享有限制的法律人格,有限地承担法律责任。对智能主体拟制的法律能力虽作了一定的限制,但与自然人或公司等法律拟制的人格不同。分阶段思维既不需要全面突破现行法律规定,又同时可以解决一些新出现的问题。处于不同智能程度的智能主体,其与“人”(研发者、设计者、所有者、使用者、销售者、改造者等)的关系也是有所差异的。例如,在不同的智能驾驶模式中,应对研发者、设计者、所有者、使用者、销售者、监管者等主体进行差异化的归责,以妥善解决智能产品责任犯罪、智能驾驶公共安全犯罪等突出问题<sup>⑨</sup>。概言之,刑法意义上的人工智能主体,可以分为:一是基于“工具属性”关系,智能主体完全是“人”的新技术工具。二是“过渡性主体”,指智能主体虽有“工具属性”,也是“自主主体”。这两者的关系是动态的,要根据实际情况来判断。三是“拟制主体”与“人”的法律平等地位。“智能主体”基于法

<sup>⑨</sup>参见:孙道萃《人工智能驾驶的刑法解构与回应路径》(《西安交通大学学报》(社会科学版),2022年第2期第158-166页)。

律拟制,被确立为独立的法律主体,在法律意义上与“人”是平等的。四是“新独立主体”,指智能主体在终极层面获得统治地位,取代“人”并成为未来的唯一“法定主体”。

在功利主义的策动下,现阶段对智能主体刑法地位予以规范性的界定与确认,都处于应然的“拟制”与“循证”之理性主义阶段,在规范层面尚未形成有效法律文本与适法依据。但从立法上予以明确,才能将智能主体作为刑法规制的对象或作为刑事法律关系的主体,才能从根本上对智能时代的刑事风险予以正面回应。

## 四、权利论:证成与前瞻

### (一) 智能主体权利的法理课题

智能主体逐步稳固其法律主体地位,则智能主体的权利保障问题必然会提上议程。有效保障智能主体的新生权利,是确认智能主体刑法地位的集中体现。

#### 1. 智能主体权利的由来与现状

关于智能主体的“权利”,当前处于“无法律规定”的蛮荒时代。但智能主体权利问题已出现在以下领域:(1)人格权<sup>[25]</sup>。运用智能技术收录人的声音、表情、肢体动作等利用人格权的客体的行为,隐藏侵犯人格权的危险,甚至导致人格权保护可能出现空挡。(2)数据财产。智能技术的应用是以信息、数据以及算法等为基础的。但数据、信息的法律性质尚不确定,数据作为权利对象或权利内容仍是空白。(3)著作权。智能创作技术及其应用日益成熟,亟待明确智能主体的著作权利保护。智能主体的真正完全独立,将使传统著作权领域的主体格局发生巨变<sup>[26]</sup>。在刑法领域,对强人工智能的生成物,不乏主张与自然人应当被同等保护<sup>[27]</sup>。(4)生命权利。国外已经出现机器人伴侣现象,人类虐待、侵害智能机器人的生命权、人格权的,应否承担民事责任以及精神损害赔偿等问题开始浮出水面。总之,智能主体刑法地位的“替代”过程必然触发智能主体的权利问题,并与人的权利体系时代相行渐远。当前,对人工智能及其相关权利问题进行保护的制度安排与法律支持体系明显不够<sup>[28]</sup>,讨论的法理基础、规范依据等储备尤为不足。而“人类中心主义”在该问题上延伸的“坚硬性”效应仍在蔓延。

#### 2. 作为新兴权利的法理释明

晚近以来,以人为绝对主导核心地位的权利格局开始出现变化。这就是新兴权利及其保护问题。新兴权利的出现乃至泛化,正危及传统权利的基本观念。其证成标准是同时具备合理性、合法性与现实性<sup>[29]</sup>。对于新兴权利,其法理的接纳难点在于认识论、价值论等方面的重大颠覆。而法律技术层面的规范难题在于“如何证成”。对于正崭露头角的智能主体之权利问题,也面临同样的制度困境。其中,最大的阻碍就是人类中心主义立场与人的主体性地位效应。从现代法理的逻辑看,法律主体地位与权利配置是正向的关联性或者一致性。智能主体的刑法地位之确认,是决定其权利问题的最基本前提。新兴权利的出现及其保护问题,也反过来推动对权利主体资格与能力的法律确认进程。对智能主体权利问题的过度绝对化看法不妥。而且,从当代法律主体的权利发展与历史演变进程看,权利的内容与形式是发展的,新兴权利与传统权利的类型是交替互进的。承认并赋予智能主体相应的权利,既是权利范畴发展的内在规律,也是智能技术推动社会发展的必然结果。智能主体在刑法领域享有权利并作为新型的权利类型,将改写现行法律体系中的权利生态格局。

不乏否定观点认为,智能主体不是人,不会制造或产生道德问题,无法自行独立作出道德判断,不能具有权利<sup>[30]</sup>。显然,否定智能主体的法律地位,也就没有权利存在的必要与空间。“当机器说出它们的感受和感知经验,而我们相信它们所说的是真的,它们就真正成了有意识的人。而大多数道德和法律制度,也是建立在保护意识体的生存和防止意识体受到不必要的伤害的基础上的”<sup>[31]</sup>。从现阶段的智能技术水平及其应用程度看,智能主体可以享有限制性权利。智能主体的权利具有立法的拟制性、利他主义属性、功能差异属性,融合了工具性的价值与利他的特性,负有服务人类的预设功能。应当根据智能主体的类型与具体功能,采取差别化的权利拟制与保护机制。“限制性权利”是由智能主体地位的相对性与限制性决定的,也巩固法律地位与权利赋予之间的内在联系性。在此观念下,可以赋予智能主体一定的自由权,使其具有一定的自主决定权;但是,与人类的自由权仍有差别,应当加以限制,使智能主体的权利及行使更符合实际情况。

## (二) 刑法保护的原则与路径

当代刑法对智能主体的保护,是对其主体地位与法定权利的“实践认同”。不能完全参照刑法对人的保护格局,应当根据权利的类型设定科学的保护措施。

### 1. 平等保护的异议与匡正

人类中心主义与功利主义作为两种不同的认识论,在智能主体权利的保护原则上仍将继续“角力”。也即智能主体与自然人暂且无法对等,也几乎无法获得完全、不受限制、与自然人平等的主体地位。智能主体可以拥有受限的权利,在权利形态及保护范围等方面存有差距。“平等原则”对自然人与智能主体之间的关系而言并不奏效。“平等保护”看似合理,实际因遵循人类的标准来匹配与设定智能主体的权利保护原则,故而是“人类中心主义”思维的具体产物。基于人类的特殊身份而确立人的主体性地位及其保护逻辑,无法“天然”地适用于智能主体。即使强行“划上等号”,由于法律能力等不尽相同,刑法保护的要求不同,是否“对等”人类的标准予以保护,该类比意义很是微弱。甚至可以说,“平等原则”实则是一个伪命题。具有独立、自主意识的智能主体,并非自然人的财产或单纯的产品。不应按照人类的标准,对智能主体权利进行“等值”保护。对其权利形态进行保护时,应当遵循智能时代的价值观与标准体系,尽量按照智能时代的需求进行专属保护。当前,若自然人利益与智能主体利益发生冲突的,原则上应当优先保障人类利益。

### 2. 权利保障的渐进策略

以技术代际的更迭为前提,智能主体的权利内容应当是发展性的。智能主体的刑法保护范围与方式取决于权利的内容与范围。在现阶段,可以着重围绕三个方面展开:(1)跳出“人”的思维窠臼审视智能主体是否具有“类人”的基本权利。在现阶段,对发展中的智能主体之权利,及其与人的权利形态之间的关系,既要跳出“人类中心主义”的思维窠臼,也要导入功利主义的现实考虑。智能主体不是一般意义上的生命体,没有生命权、人身自由权,但可以拥有财产权利<sup>[32]</sup>。实际上,智能主体是否可以享有生命权的技术难点在于能否按照人的生命体及其特征,界定智能主体的生命形态及其法律性质、外部载体方式等。人的生命权与传统的保护思维不再继续有效,那么,需要重新审定剥夺或侵害智能主体生命的方式、刑事制裁等问题。智能主体的升级换代,是通过更新程序等方式,改变(或延长)智能主体的生命;达到使用年限后的重新格式化是结束生命的方式。技术升级使智能主体的存续方式更持久。无论从技术还是智能系统的独立性看,智能主体具有特殊的“生命权”,以及基于生命权所形成的独立的电子人格、自由权、追求幸福的权利、避免被长时间进行奴役

性的劳动权利等。(2)立足智能时代审视智能主体权利及其保护的的特殊性。尽管在认识智能主体权利问题上,目前没有先例可供遵循,在知识论上不得不参照“人类”法律立场。但不能沦为简单的“类比”或“复制”。例如,智能主体权利的基本类型数据资源的共享权利、个体数据的专有权利或排他性权利,以及基于社会功能约束的自由权和获得法律救济的权利<sup>[33]</sup>。立足智能技术、技术应用领域等因素,设定数据类的权利形态是可取的。该观点不完全依循“人类中心主义”及“人”的权利格局,是典型的智能时代“专属思维”。但是,未能根据智能技术应用水平与智能程度、智能主体类型等,精细化地阐明权利的内容,并设置刑法保护措施。智能主体的权利内容不限于数据形态方面。即使主要肯定智能主体的数据权利,仍需明确数据权利的基础、类型、权限,以及在权利体系中的地位等相关问题。(3)现阶段刑法保护的主要策略。一是权利形态的兼容性。可以适当比拟“人的权利”逻辑,为探讨智能主体的权利范围提供思考空间。例如,生命权等具有高度“人的专属性”,是否可以“移植”为智能主体的权利形式。可以结合智能技术、算法运行、应用载体等要素,进行功能的“等值”判断。又如,对于智能主体的著作权,目前的应用已经很成熟,可以考虑优先确认。合理借鉴“人的权利”逻辑,可以对智能主体的权利范围有更清晰的“前瞻性思考”。二是权利范围的专属性。智能主体的权利内容与形式具有显著的专属性,包括基于算法运行而生成的一系列专属性数据权利。当前,基于数据形成的权利体系是智能主体权利的主要来源与表现形式,但具体类型不断变化。智能主体专属的数据权利,不仅与算法等智能因素息息相关,也与研发者、设计者、所有者、使用者等“人”的(数据)权利相互竞合,不能忽视与人作为权利主体之间的实质界限。三是权利的限制性。智能主体并不必然享有专属于人的权利。智能主体的权利行使受限,相互冲突时,一般优先保护人的权利。四是权利结构的优化。既包括实体性权利,也包括程序性权利,如智能主体作为法律意义上的原告与被害人等。既要赋予权利内容,也要赋予救济性权利,避免权利处于司法保护的空白状态。五是保护措施的完善。通过立法增加与智能主体权利保护相适应的刑事制裁措施,或同步修正现有的刑事制裁措施,更好地保护智能主体的权利或权益<sup>[34]</sup>。

#### 参考文献:

- [1]刘振宇.人工智能权利话语批判[J].自然辩证法研究,2018(8):31-37.
- [2]杨立新.人工类人格:智能机器人的民法地位:兼论智能机器人致人损害的民事责任[J].求是学刊,2018(4):84-96.
- [3]时方.人工智能刑事主体地位之否定[J].法律科学(西北政法大学学报),2018(6):67-75.
- [4]魏东.人工智能犯罪的可归责主体探究[J].理论探索,2019(5):5-13.
- [5]刘宪权.对人工智能法学研究“伪批判”的回应[J].法学,2020(1):3-14.
- [6]冀洋.人工智能时代的刑事责任体系不必重构[J].比较法研究,2019(4):123-137.
- [7]余谋昌.走出人类中心主义[J].自然辩证法研究,1994(7):8-14,47.
- [8]邱耕田.从绝对人类中心主义走向相对人类中心主义[J].自然辩证法研究,1997(1):15-18,58.
- [9]刘艳红.人工智能法学研究的反智化批判[J].东方法学,2019(5):119-126.
- [10]董玉庭.人工智能与刑法发展关系论:基于真实与想象所做的分析[J].现代法学,2019(5):120-134.
- [11]周详.智能机器人“权利主体论”之提倡[J].法学,2019(10):3-17.
- [12]皮勇.全国首例撞库打码案的法律适用分析[J].中国检察官,2019(6):7-9.
- [13]杜严勇.现代军用机器人的伦理困境[J].伦理学研究,2014(5):98-102.
- [14]马治国,田小楚.论人工智能体刑法适用之可能性[J].华中科技大学学报(社会科学版),2018(2):108-115.

- [15]郭少飞.“电子人”法律主体论[J].东方法学,2018(3):38-49.
- [16]张志坚.论人工智能的电子法人地位[J].现代法学,2019(5):75-88.
- [17]孙加顺.以伦理为基准的设计 通往 AI 时代的至高美德:访 IEEE 标准协会首席执行官康斯坦丁诺·卡拉卡琉斯博士 [J].中国标准化,2018(1):18-21.
- [18]李鹏,李彦宏“违法”无人驾驶背后法规该如何跟上新技术脚步[N].北京科技报,2017-07-10(4).
- [19]边沁.立法理论[M]李贵方,译北京:中国人民公安大学出版社,2004:35.
- [20]边沁.道德与立法原理导论[M].时殷弘,译.北京:商务印书馆,2000:58.
- [21]边沁.论一般法律[M].毛国权,译.北京:上海三联书店,2008:41.
- [22]亨利·西季威克.伦理学史纲[M].熊敏,译.北京:江苏人民出版社,2007:203.
- [23]张宏生,谷春德.西方法律思想史[M].北京:北京大学出版社,2000:284.
- [24]谢晖.论新型权利的基础理念[J].法学论坛,2019(3):5-19.
- [25]王利明.人工智能时代提出的法学新课题[J].中国法律评论,2018(2):1-4.
- [26]孙占利.智能机器人法律人格问题论析[J].东方法学,2018(3):10-17.
- [27]刘宪权.人工智能生成物刑法保护的基础和限度[J].华东政法大学学报,2019(6):60-67.
- [28]季卫东.人工智能开发的理念、法律以及政策[J].东方法学,2019(5):4-13.
- [29]雷磊.新兴(新型)权利的证成标准[J].法学论坛,2019(3):20-29.
- [30]甘绍平.机器人怎么可能拥有权利[J].伦理学研究,2017(3):126-130.
- [31]雷·库兹韦尔.人工智能的未来:揭示人类思维的奥妙[M].盛杨燕,译.杭州:浙江人民出版社,2016:203-205.
- [32]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.
- [33]张玉洁.论人工智能时代的机器人权利及其风险规制[J].东方法学,2017(6):56-66.
- [34]孙道萃.智能时代的刑法立法:人类中心主义与现实功利主义的取舍[J].学术交流,2020(4):69-80.

## Positivism for AI subject in criminal law: Also answer to the negativism

SUN Daocui

(College of National Legal Aid, China University of Political Science and Law, Beijing 100088, P. R. China)

**Abstract:** The rapid development of intelligent technology and its application has a strong impact on the contemporary legal system based on the fundamental premise of human subjectivity, and the criminal law status of artificial intelligent subjects has become a key issue. At present, the negative theory and the positive theory are in a stalemate, accelerating the process of knowledge change in the current criminal law theoretical system. Different degrees of negativism, such as “artificial objects”, have their merits on the grounds that the current criminal legal system is difficult to directly accept them and they carry potential criminal risks. Its logical mechanism is the presupposition effect of excessive release of anthropocentrism. Examining from an anthropocentric standpoint will lead to irreconcilable conclusions, and it also shows the tendency of the tool attributes of intelligent technology applications to be deeply magnified. However, the above-mentioned negative reasons are not necessarily in line with the actual needs of regulating artificial intelligence crimes, they are also inconsistent with the evolution law of legal subject system, and it will even suppress the autonomous evolution of the subject of criminal law. Different forms of positive theory such as “cyborg” are the first to break out of

absolute anthropocentrism and its conceptual constraints, and view the interaction and trends of artificial intelligence and criminal law from a developmental perspective, which is more in line with the development trend of artificial intelligence crimes. The accumulation of existing knowledge frameworks and new elements, such as freedom of will, criminal responsibility, moral and ethical rules, intelligence level and types of intelligent subjects, the objective existence of criminal responsibility, and the special status and significance of algorithms, means that artificial intelligence can develop in the future to achieve cohesion and fit in the criminal law system. This reversal of understanding, which does not really completely deviate from the principles of criminal law, also provides support at the theoretical level. Moreover, following the rationale of utilitarianism, and under the guidance of a series of advanced legislation on artificial intelligence subjects, the criminal status of intelligent subjects should be determined in stages, types, and dynamics, so that they can continue and expand on the path of legal fiction. Under the clarification of the subject status of artificial intelligence in criminal law, it is necessary to accept the reality that intelligent subjects can enjoy certain emerging rights, but the specific content such as the type of rights should be specifically confirmed. Under the instigation of the current technology generation, the content and scope of the rights of intelligent subjects are currently limited, and it is impossible to adopt a protection strategy equivalent to that of “people”, and the protection methods should also be different. This kind of “dimension reduction” protection has its practical rationality. Through active system design and rule configuration, it can coordinate with the changing situation of the criminal law status of artificial intelligence subjects to the greatest extent.

**Key words:** artificial intelligent subject; criminal law status; negativism; anthropocentrism; positivism; utilitarianism; logic of rights

(责任编辑 胡志平)