

doi: 10.15936/j.cnki.1008-3758.2020.01.012

AI 刑事责任主体否定论的法理与哲理证成

——兼论“人”是什么

冯文杰¹, 李永升²

(1. 东南大学 法学院, 江苏 南京 211189; 2. 西南政法大学 法学院, 重庆 401120)

摘 要: AI 刑事责任主体肯定论隐藏着将问题缥缈化、对策简单化及法理忽视化的问题之虞。当前的 AI 刑事责任主体否定论存在忽视肯定论的具体语境问题而稍显批判力不足之虞,应对其进行法理与哲理证成。弱 AI 产品与强 AI 产品都不应是刑事责任主体,而是人类可控制的依靠电能存续的 AI 机械产品,应被作为人类实现自由而全面发展的工具。人是万物的尺度,肯定论不当坚持了一种机械式“人”论。刑事法适用的核心部分和刑事归责的基本原理不会因 AI 技术的出现发生根本变化。对 AI 技术发展的风险,不仅需依靠算法的顶层设计,且需将法治管理嵌入 AI 发展的每个环节,以防止造成消极后果。法律人须承认自身知识的局限性、保持适度克制、不能误将情感或联想等同于“现实”。

关 键 词: AI 刑事责任主体肯定论; AI 刑事责任主体否定论; 法理证成; 哲理证成

中图分类号: DF 62 **文献标志码:** A **文章编号:** 1008-3758(2020)01-0090-09

The Jurisprudence and Philosophical Proof of the Negative Theory of AI Criminal Responsibility Subject

—— Also on What “Human” Is

FENG Wen-jie¹, LI Yong-sheng²

(1. School of Law, Southeast University, Nanjing 211189, China; 2. School of Law, Southwest University of Political Science and Law, Chongqing 401120, China)

Abstract: The affirmation of AI criminal responsibility subject hides the problems of floating the problem, simplifying the strategy, and neglecting the legal principle. The current negative theory of AI criminal responsibility subject has the problem of neglecting the specific context of affirmation and the lack of critical power. It should be legalized and philosophically proved. Neither the weak AI products nor the strong AI products should be the subject of criminal responsibility. They are only intelligent mechanical products that humans can control and that rely on electric energy for survival. They should be used as tools for human beings to achieve free and comprehensive development. Man is the measure of all things. The affirmation theory insists on a mechanical “human”. The core part of the application of criminal law and the basic principles of criminal imputation will not change fundamentally due to the emergence of AI technology. The risk of the development of AI technology depends not only on the top-level

收稿日期: 2019-04-10
基金项目: 国家社会科学基金重大资助项目(16ZDA060); 教育部人文社会科学研究规划基金资助项目(15YJA820015); 重庆市教委人文社会科学一般资助项目(18SKGH007)。
作者简介: 冯文杰(1991-),男,河南项城人,东南大学博士研究生,主要从事中外刑法学研究; 李永升(1964-),男,安徽怀宁人,西南政法大学教授,博士生导师,主要从事中外刑法学研究。

design of the algorithm, but also on the integration of rule of law management into every aspect of AI development to prevent negative consequences. Legal persons must acknowledge the limitations of their knowledge, maintain moderate restraint, and must not mistake emotion or association for “reality”.

Key words: the affirmation of AI criminal responsibility subject; the negative theory of AI criminal responsibility subject; jurisprudence proof; philosophical proof

一、问题的由来:AI 刑事责任主体论的兴起

近年来,人工智能(以下简称为 AI)刑事责任主体问题逐渐兴起,无论赞同还是反对,这都已经成为刑法学者不得不认真对待的问题。毋庸置疑,若强 AI 产品能够出现,“面对研发和使用人工智能程序中研发者和使用者实施的或者智能机器人超越人类智慧独立实施的严重危害社会的犯罪行为,刑法绝对不应该无动于衷甚至束手无策”^[1]。刘宪权教授从不同的视角,依据不同的论据,提出了一些值得深思的结论:刑法学必须抛开以往的成见或偏见,正面应对 AI 刑事责任主体问题,将强 AI 产品作为适格的刑事责任主体。因为强 AI 产品在自主意志支配下实施不在人类设计和编制的程序范围内的“犯罪行为”时,实现的是自身意志,而非研发者等人的意志,这与一般的具有刑事责任能力的人实施犯罪行为的情形并无不同,故而应赋予强 AI 产品刑事责任主体地位。强 AI 产品也可能“与自然人责任主体、其他强人工智能产品构成共同犯罪,针对强人工智能产品的犯罪,有必要在刑法中增设删除数据、修改程序、永久销毁等刑罚种类”^[2]。与之相反,时方博士指出,从意志自由、刑罚目的及法人特殊的运作机理等方面而言,AI 刑事责任主体肯定论难以成立^[3]。其结论无疑是合理的,但其忽视了 AI 刑事责任主体肯定论的强 AI 产品刑事责任主体地位这一特殊语境,由此而来的批判稍显不足,有必要对其进行法理与哲理证成。更为重要的问题在于,AI 机器人是否能够摆脱人类控制自主实施法益侵害行为?如果不能解决这个技术性问题,则所谓的刑事对策很可能属于“无病呻吟”;反之,

则当前的刑事对策就属于高瞻远瞩的未雨绸缪。AI 产品通过人工赋予的“学习”生成并不断更新换代,机器学习是典型的数据驱动的思维模式:从数据出发,通过各种计算方法来理解数据,并建立适当的算法模型以整合数据,从而得出结论。当前具有自由意志的强 AI 产品尚未出现^①,能够自主实施法益侵害行为的强 AI 产品更未出现。

早在 2001 年,张保生教授就指出,人一机系统联合方案立足于人与机器的功能互补,目的是打破对于人的自由而全面的发展的束缚,并服务于国家治理功能的优化^[4]。毫无疑问,人类发展 AI 技术的目的应当是实现人类自由而全面的发展。在新事物产生之前,当前学界本着负责之心对其可能出现的刑事风险展开规制措施研究,有一定的必要性。但若研讨的刑事问题及对策没有法理与哲理支撑,则往往不为公众所认同。具体而言,AI 产品是否能够作为刑事责任的归责主体,这一问题方兴未艾;若无法合理解决之,则由之衍生的许多问题皆不能被妥善解决。故而本文特就这一问题展开反思性的法理与哲理分析,进而合理解决 AI 刑事责任主体肯定论提出的具体问题,确立一种合理的“人”论,从而促使 AI 技术应用在一个理性的路径上健康、有序发展。

二、AI 刑事责任主体否定论的法理与哲理证成

针对 AI 产品是否能够成为刑事责任主体,当前学界形成了针锋相对的肯定论与否定论。肯定论的见解虽不乏真知灼见,但隐藏着将问题缥缈化、对策简单化及法理忽视化的问题之虞。当前的否定论的见解虽有让人赞赏之处,但有忽视

① 当前学界将 AI 产品划分为不同类型,有三分法与两分法之争。三分法认为,AI 产品有弱 AI 产品、强 AI 产品及超 AI 产品;二分法认为,AI 产品有弱 AI 产品与强 AI 产品。不论是两分法还是三分法,区分的基本标准都是 AI 产品是否具有自主意识、是否能够自主实现意志。鉴于当前刑法学界的主流观点是两分法,且只要合理认知划分标准,则两分法与三分法对于解决刑法问题并无重要差异,故而本文坚持两分法。

肯定论的具体语境问题而稍显批判力不足之虞,应对其进行法理与哲理证成。

1. 针锋相对:AI 刑事责任主体论的正反聚讼

总体而言, AI 刑事责任主体肯定论认为, 应将强 AI 产品作为适格的刑事责任主体, 将弱 AI 产品作为辅助人实施法益侵害行为的工具。因为第一, 强 AI 产品能够依靠对于外在事物的认识及评价控制自身行动, 当其为实现自身意志而自主实施犯罪行为时, 应当受到相应的刑罚惩罚, 由此能够将强 AI 产品作为刑事责任主体。第二, 强 AI 产品能够“独立”实施法益侵害行为, 这是以刑法规制其行为的基本前提, 因为“无法益侵害, 则无犯罪行为”。第三, 相较于动物与单位而言, 强 AI 产品与自然人的实质相似程度相当高, 应当对其自身的所作所为承担相应责任^[5]。第四, 既然强 AI 产品能够自主实施犯罪行为, 则必须坚持罪责自负原则, 对其进行刑事规制, 否则, 即违背了罪责自负原则。第五, 强 AI 产品与人一样, 二者都具有理性, 即“用智识理解和应对现实的(有限能力)”^[6]。第六, 强 AI 产品的出现不是天方夜谭, 而是有着充分科学依据的。换言之, 科技能够实现强 AI 产品的生产^[7]。肯定论主张的诸如强 AI 产品具有理性等支撑理由奠基于强 AI 产品具有自由意志之上。由此可见, 肯定论的立证之基是强 AI 产品具有自由意志。若要以直接反驳方法驳倒其结论, 则需从否定强 AI 产品具有自由意志展开; 若要以间接反驳方法驳倒其结论, 则需从反驳其具体刑事规制措施展开。

大体而言, 当前的 AI 刑事责任主体否定论认为, 不应赋予 AI 产品以刑事责任主体地位, 应将其作为能够为人利用而实施法益侵害行为的工具。因为其一, 弱 AI 产品不具有自由意志, 仍属于能够为人利用而实施法益侵害行为的辅助工具。其二, 所谓的强 AI 产品具有自身的自主意识和意志之论属于伪科学, 不具有现实可能性, 由此无法将其作为刑事责任主体^[3]。其三, 所谓的强 AI 产品尚未出现。虽然“未来已来, 但不是说来就来”^[8]。当前应当脚踏实地地研究具有现实性的问题, 比如由自动驾驶衍生出的过失犯认定问题, 还比如由 AI 产品“生产”的诗歌、小说等事物是否能够被认定为著作权法上的作品问题, 再比如如何提高 AI 辅助量刑系统精准度的问题, 不宜研究过于缥缈的问题。由此可见, 当前的否定论主要是以直接反驳方法反驳肯定论的, 立论

之基为 AI 产品不具有自由意志, 科技不可能生产出强 AI 产品。

无论是肯定论还是当前的否定论, 都认为弱 AI 产品不具有自由意志, 无法作为刑事责任主体, 只可能作为人或单位(在单位犯罪中, 具体实施法益侵害行为的主体仍然是人)实施法益侵害行为的辅助工具。针对 AI 刑事责任主体肯定论, 当前的否定论的反驳虽已相当有力, 但有忽视肯定论的具体语境问题而稍显批判力不足之虞, 且尚有一些值得深究之处: 其一, 如何针对肯定论基于 AI 技术发展的设想进行反驳; 其二, 如何针对肯定论提出的“增设删除数据、修改程序、永久销毁等刑罚种类”的规制措施进行反驳; 其三, 能否在哲理上反思肯定论的背后体现出何种“人论”, 其是否合理; 若不合理, 则人们应当坚持何种“人论”, 从而合理认知 AI 刑事责任主体问题。

2. 直接反驳之具体展开

应当认为, 强 AI 产品不具有自由意志, 不宜被作为刑事责任主体对待。从直接反驳角度而言, 主要从如下四个方面展开。

第一, 肯定论以 AI 机器人可以在技术和专业知识的帮助下获得“道德”为例, 证明 AI 产品可以成为具有自主意识的强 AI 产品。具体而言, 人的道德暨善恶观念由来源于后天的自身实践感悟和他人的培养, 而不是一种生来即有的意识。经过 AI、脑神经科学等技术的智能化“训练”后, AI 机器人能够具有道德上的善恶观念。此外, 虽然当前的 AI 机器人尚未具有道德观念, 但将来的 AI 机器人能够具有道德观念。何况人们需要 AI 机器人能够具有道德观念, 从而更好地服务于人们的生活。比如家庭服务型机器人在照看幼儿及服务家庭成员的工作过程中, 需要认知伤害与救助、打扫家务与清除垃圾等活动的各自内涵; 否则, 难以服务好人们的生活^[9]。论者的观点有所不当。其一, 应当认为, 若要使 AI 机器人能够更好地造福于人类生活, 则必须使其尽可能以“自然法”要求的方式实施行动。但即使人类通过 AI 等技术的应用, 使其具有作出符合法律要求的行动的功能。这也仅仅意味着, 其具有由大数据图像、语音、语义识别等技术赋予的“他律”行动功能, 却无法实施“自律”行为。其二, 部分肯定论者承认, AI 机器人的“道德是基于数据和算法的线性‘思维’形成的”^[10]。换言之, 人们“给人工智能系统设定一套代码形式的价值观”^[11], 以使

得 AI 机器人的行动选择结果不致偏离基本善恶观。既然是这样的一种事物,就无必要称其具有道德观念,其仅仅具有人类通过数据分析与算法改进等技术赋予的行动选择功能,难以认为这样一种功能是道德观念的象征。其三,论者存在混淆“需要”与“现实”之虞。虽然人们不仅需要 AI 机器人能够区分伤害与救助、打扫家务与清除垃圾等活动的不同内涵,且需要其具有善良的道德观念,从而能使其更好地服务于人们的生活,但仅仅根据“需要”不能产生“现实”,这是两个不同的范畴。论者存在以人们需要 AI 机器人具有道德观念推导出其能够具有道德观念的问题之虞。应当认为,即使 AI 机器人在外在上能够区分善恶观念,也不代表其在内在上具有了道德观念。

第二,虽然强 AI 产品在表象上可以作出法益侵害行为,但对此必须区分不同情形进行审慎分析,否则,难免有以偏概全之虞。首先,刑法中的行为不是指任何表象世界中的活动,而是经一定价值判断后所厘定的特殊行为,即“行为主体实施的客观上侵犯法益的身体活动”^[12],比如动物的袭打行为不是刑法意义上的行为,还比如恶劣天气产生的雷击活动也不是刑法意义上的行为。刑法中的行为主体仅仅包含自然人,单位与自然人构成了刑法上的归责主体。刑法能够对单位进行刑事归责的客观依据是,单位是由诸多自然人组成的联合体,单位犯罪实际上是由自然人具体实施的。故而必须是适格的行为主体实施的法益侵害行为才会被作为刑法上的行为。其次,当强 AI 产品实施法益侵害活动是由于受到自然人或单位的控制时,这种活动才能够因附着于自然人或单位而被作为刑法上的行为。换言之,这种行为事实上是相关自然人或单位的行为,刑事归责主体也不是强 AI 产品,而是相关的自然人或单位。当强 AI 产品并未受到任何适格刑事责任主体的控制而自主实施法益侵害活动时,必须在查明原因后进行合理评价,原因无外乎两种:其一,技术人员等相关主体的过失。其二,完全是被害人的行为导致了强 AI 产品的误解。当原因为“其一”时,强 AI 产品的行为是刑法上的行为,必须依据过失犯的归责原理认定相关技术人员等是否构成相应的过失犯罪;当原因为“其二”时,强 AI 产品的行为不是刑法上的行为,为了避免类似事故的再次发生,必须由相关技术人员改进技术^[13]。最后,肯定论以强 AI 产品能够自主实施

法益侵害行为为由,认为这是必须赋予强 AI 产品刑事责任主体地位的一个条件之说,显然是未深入地区分不同类型的结果,有所不当。

第三,肯定论认为,科技发展必将促成具有自由意志的强 AI 产品的出现,甚至认为在未来 10 年至 20 年间便能够成功^[7]。不得不说,这一结论为时过早。依照当前 AI 技术界的通说而言,这一结论并不客观。当前受制于技术等条件的限制, AI 技术发展遭遇了一些尚无法克服的瓶颈。美国国家科技委员会认为,强 AI 产品在未来数十年都难以出现。数百名世界顶尖 AI 专家认为,人类级别的 AI 产品暨强 AI 产品的出现至少需要等到 2040 年^[14]。“AI 之父”阿兰·图灵在 1936 年提出了“图灵机”设想,由此在数理逻辑上为计算机技术的创新发展作出了新的尝试,并在 1950 年提出了“图灵测试”:若一台机器能够与人类展开对话且不能被辨别出其机器身份,则这台机器就是智能的。其实,即使“图灵测试”成功,也仅仅证明了 AI 机器人的研发将仿生学运用得相当精湛,难以由此认为这种 AI 机器人具有自由意志^[15]。AI 本身是一个模拟人类能力和智慧行为的跨领域学科,涉及到诸如计算机科学、大数据分析、算法、语言学、仿生学、神经生理学等多个领域。若人类在上述领域无法取得突破性进展,则 AI 技术必将难以实现突破性发展。换言之,“人工智能的一些领域似乎特别具有挑战性,如语言、创造力和情感。如果人工智能不能模拟它们,要实现强人工智能就好似做白日梦”^[16]。当前任何一台按照现有方式加以编程的计算机都无法完成“全面模拟人类的心智”这项任务,因为其没有人脑所特有的“意识”(awareness)。英国物理学家彭罗斯由此认为,无论是“强 AI 概念”还是“弱 AI 概念”,都已远远高估了 AI 技术的发展潜力,应当以“更弱 AI 概念”描述当前的 AI 技术^[17]。此外,当前对于 AI 技术的应用皆指向工具范畴,并未显现涉及刑事责任主体问题的可能。毋庸置疑,以现在预测未来总是困难重重,因为以现在否定未来总会给人以若干诟病。但从目前的技术及肯定论的支撑理由而言,所谓的强 AI 产品的出现之论不具有现实基础。

某种科学的出现需要一定的客观基础,如果这种科学的实现根本不具有现实可能性,则人类大可不必相信其会成为现实。譬如人类希冀“长生不老”,但“细胞的分裂是有极限的,不可能使人

的生命永存”^[18],这种希望至多只是一种美好愿望而已。法律人不会研究这一问题:如果人类实现了“长生不老”的愿望,则应当适用何种刑罚措施规制其法益侵害行为。“如果没有特定的价值秩序基础,把什么行为都认为是正确的,那么法律是不可能存在的。”^[19]因为“如果一种法律没有规则或者其规则未得到有序的遵循,那么这种法律就不能成为我们所了解的那种法律制度。它是一种随意决断的非正式的‘制度’”^[20]。法律不存在的后果必然是社会的混乱不堪,并使其成为个人与个人之间的战争的社会。应当相信具有自由意志的强 AI 产品不可能出现。即使出现了比弱 AI 产品更“强”一些的强 AI 产品,后者仍不具有自由意志,仅仅是前者的升级版,即在“深度学习”“决策判断”“智慧模拟”等方面能力更高一些。对于由强 AI 产品衍生的刑事风险而言,宜在坚持故意与过失二分法的情形下进行具体判断。一方面,不具有故意的强 AI 产品研发者和使用者可能触犯相关过失犯罪,应在明晰技术水平的前提下,审慎考量信赖原则、预见可能性等刑法原理,若相关责任主体的行为符合相应过失犯罪的具体规定,即以相关过失犯罪定罪量刑即可;另一方面,若自然人或单位积极利用或负有作为义务的相关责任主体容忍强 AI 产品实施法益侵害行为,则依照规范与事实的对应原理,将目光在法规范与案件事实之间循环往返,以相应故意犯罪对责任主体进行定罪量刑即可。由此可见,刑事责任的承担主体仍然是相关责任人,而非强 AI 产品。此外,域外相关立法也赞同上述结论,比如欧洲议会《在《机器人民法规范》中认为,“至少在现阶段,责任必须由人而不是机器人承担”^[21]。

第四,肯定论认为,可以设置删除数据、修改程序、永久销毁等刑罚措施对强 AI 产品自主实施的值得刑法处罚的法益侵害行为进行规制。这种观点虽不乏启示作用,但仍有所不当。因为即使强 AI 产品能够在外在上展示一种痛苦感受,也不意味着其具有如人一般的心理感受。何况上述痛苦感受仅仅是由相关技术人员通过多种技术所赋予的,强 AI 产品只是由一堆机器部件和编制程序组成的智能体。刑罚的本质是一种伴有痛苦性的权益剥夺的报应,没有痛苦性权益剥夺的措施无法被称之为刑罚措施^[22]。普罗大众既不会相信上述“刑罚”的施加,能够造成强 AI 产品有类似于人的痛苦感受,也不会认可上述所谓的

“刑罚”能够被称之为真正的刑罚措施。如此一来,施加上述“刑罚”不仅无法取得公众认同的效果,且很可能造成公众质疑司法不公正的结果^[23]。当前的 AI 产品无法理解自身的程序预设目标之外的意义或价值,比如在 AI 技术应用较为成熟的自动驾驶领域,自动驾驶系统仅仅能够根据对具体环境的感知,按照预设程序发出的指示作出行动,无法作出自动驾驶系统中的预设程序之外的行动。同时,人们在驾驶装载了自动驾驶系统的车辆时,对于系统本身的感知和反馈模式拥有绝对控制权。由此可见,AI 产品并无自主行动功能,仅仅具有工具属性。此外,对于 AI 产品实施的法益侵害行为,在其他部门法尚未展现系统的规制措施时,刑事法学者便倡导由刑法进行规制,有违刑法谦抑性原则(辅助性的法益保护原则)。因为刑法是最为严厉的惩罚法,不能动辄以刑事手段惩罚法益侵害行为,“只有更为缓和的国家手段(像民事制裁、公法上的禁令、运用违反秩序法或其他社会政策性措施),无力维护和平和自由的时候,才允许刑法的介入”^[24]。

3. 间接反驳之具体展开

如果仅仅以直接反驳方法反驳肯定论,肯定论可能继续反驳称,否定论的反驳与其阐述的观点不在同一语境内,由此势必降低否定论的说服力。基于此,以下从间接反驳角度对肯定论进行反驳。

其一,肯定论提出赋予强 AI 产品刑事责任主体地位观点的基本前提是,强 AI 产品“不可控制”^[25],这有所不当。如果强 AI 产品给人类带来的是无法逆转的灭绝性结果,则有必要像“禁止克隆人”一般禁止相关技术人员再对其进行研究与生产。若要使得肯定论的观点合理化,则其认定的基本前提应当是,上述仅仅带给人类灭绝性结果的强 AI 产品不会出现,出现的仅仅是偶尔自主实施严重法益侵害行为的强 AI 产品,且这种法益侵害结果必须能够为人类所控制。因为首先,人类从事物质活动和精神活动的首要目的是生存,没有生存,其他一切便失去了必要基础。如果强 AI 产品的出现将给人类带来灭绝性结果,则人类必将反对生产这种产品。其次,虽然人类在生产强 AI 产品时,可能无法预料这种强 AI 产品是否将带给人类灭绝性结果,但若这种给人类带来灭绝性结果的强 AI 产品出现,则人类必将走向灭绝,继续谈论刑事规制措施便没有了必要

性。因为即使制定了应对的刑事规制措施,也无法具体实施。最后,对于强 AI 产品具有自主意识这一观点,虽然无论相信还是反对,都只是一种猜测,并无绝对的对错之分,仅仅有合理与否之分,但强 AI 产品必须能够为人所控制,否则,谈论刑事规制措施便无意义。此外,设置“删除数据、修改程序”的刑罚措施证明了,在肯定论看来,强 AI 产品仍然是可以被人类控制的,这便与其逻辑上的“不可控制”出现了相矛盾之处,而“在追求真理的主张和理论中,矛盾是一种致命的缺陷”^[26]。由此可见,肯定论对于基本前提的界定有所不当,难以使人相信其刑事规制措施具有充分的合理性。

其二,如果强 AI 产品完全不受人类的控制,自主攻击人类所不允许伤害的物体,则人类可以、且应当直接将其消灭,没有必要设置删除数据、修改程序、永久销毁等刑罚种类。这类似于战场上的对我双方,处于你死我活之状态,没有中间道路可走。换言之,这样的强 AI 产品已经不是“人民”,而是“人民”的“敌人”,对待战场上的敌人的方法应是彻底征服,设置刑事程序既无必要,也会造成行动受困的不力结果。此外,肯定论提出的刑罚措施中的“永久销毁”本身就是一种直接消灭强 AI 产品的应对对策,而不宜称之为一种刑罚措施。

其三,即使人类通过诸多技术的改进及应用,使得强 AI 产品能够在外在上发出痛苦状的撕裂声,也难以由此认为应当为其设置“删除数据、修改程序、永久销毁”的刑罚措施。有理由相信公众不能接受这样的刑罚措施可以称之为真正的刑罚措施。同时,如此设置刑罚措施存在过于简单化之嫌。比如强 AI 产品自主实施法益侵害行为后,对于被害人的赔偿如何进行,这一问题并未解决。由这种强 AI 产品衍生出一系列冲击现有司法制度的问题,并未被肯定论所解决,比如被害人是否相信自己有必要“谅解”强 AI 产品等问题。

三、“人”意宣示:AI 刑事责任主体问题的背后

AI 刑事责任主体肯定论与当前的 AI 刑事责任主体否定论的聚讼点在于强 AI 产品是否具有自由意志,双方争论的背后忽视了对于“人”的考察。如果不懂得“人”意为何,则势必无法从哲理

上廓清 AI 刑事责任主体论的认知障碍。基于此,以下展开对于“人”意的哲理分析。

针对“人”是什么这一问题,虽然古今中外的哲人思来想去,但当前尚无统一答案。对于某个事物,作出概念上的定义本身就是困难的,因为概念思维总是存在着某些漏洞^[27],何况要对世间最为“深不可测”的“人”作出定义,就更是难上加难。苏格拉底认为,“人是一个对理性问题能给与理性回答的存在物”^[28]。但理性是一个含义隽永之词,其往往带来标准的模糊不清,且“理性不足以让人理解人类文化生活形式的全部丰富性和多样性,或者说,不足以表征这种丰富性和多样性的统一性”^[29];故而不宜将人界定为理性的存在物。基于此,卡西尔将人界定为“可以利用符号去创造文化的动物”^[30]。这种定义仍然显得较为模糊不清,因为“符号”“文化”本身就是一些歧义横生的语词。人区别于其他动物的显著特征在于,前者不仅具有后者所具有的感受及效应系统,且具有后者所不具有的符号系统。由此亚里士多德认为,人是一种能够依靠语言表达价值观念而优越于其他动物的政治动物^[31]。虽然人类语言和动物语言都具有自我表达与发出信号这两种较低级的语言功能,但人类语言更为高级的地方在于,其具有描述符合于事实的观念的描述功能与对描述本身加以批判、反思的论证功能^[32]。在这个意义上而言,人不同于动物。此外,人优越于其他动物的依据不仅仅在于,人知晓凭借语言符号进行价值意义上的交流与沟通;且在于,人知晓运用主观能动性获得身体支配能力和精神生活水平的提高。故而亚里士多德对于人的界定有所不当。从比较角度而言,卡西尔和亚里士多德对于人的界定都是站在功能论角度而言的,而非站在本体论角度而言的^[33]。但仅仅在功能论意义上对人进行界定,并不合理,因为人不仅具有多种功能,且人首先是目的。由此可见,功能论层面的“人”意界定不仅失之于片面性,且忽视了对于人是目的这一视角的分析。

鉴于我国深受马克思主义的影响,对于“人”是什么这一问题,有必要考察马克思主义的观点。马克思主义认为,人是一种具有主观能动性的动物,人的本质是社会关系的总和。AI 产品本质上是人的工具,即人的器官的延伸,故而无法在法律关系中真正成为一个主体。生产力是社会发展的根本动力,其由劳动者、劳动工具、劳动对象三要

素构成。在这三个要素中,劳动者处于支配地位,而 AI 产品属于劳动工具,为人所造且为人所用。总之,“人工智能无法复制、模拟和超越人类主体性”^[34]。虽然这种观点极具启发意义,但仍有不足之处。因为上述观点并未指出人的规范化含义^[35]。现代社会是一种奉行规则之治的社会。“规则的统治意味着,权利是受到严格限定的,而公民的义务则是有限的。因此,对规则的精心设置产生了对官员行动一致性和公平性的各种期待。”^[36]社会中的人是一种规范化的物种,受规范支配、约束和保护^[37]。早在古希腊时期,哲学家普罗泰戈拉便提出了“人是万物的尺度”这一论断。对于“万物”的评价而言,人们需要综合考量各个具体事物、抽象属性、感觉属性以及诸如正确与谬误等评价概念^[38]。这一论断的主要目的在于,使得公民理性地懂得,如何在社会的剧烈转变时期,使得自身主动而有效地参与社会实践,并促使价值多元化的社会现实中存在基本共识^[39]。这种思维对于 AI 时代的法律规制极具借鉴价值,面对 AI 时代的诸多风险,有必要在法律规制上形成基本共识,以有效促使 AI 技术的健康、有序发展。尽管人对于许多事物的产生原理及运行状态都不清楚,但人仍然需要、能够且应当是自我构建的目的。由此应当认为,人是万物的尺度,可以决定某一事物的价值高低。比如环境利益属于一种刑法法益的原因在于,它是有利于人类自身的诸如生命、身体、自由等利益实现的“现实存在”;如果没有这种关联性,则环境利益不应当成为一种刑法法益^[40]。此外,康德亦认为,“现代人不应受到本能的指挥,……他应当从自身中创造一切事物来”^[41]。其“实践理性”原则阐述了具有主观能动性且应受“自律”支配的人,而技术理性提升了人的身体的支配能力、拓宽了人的身体的支配范围^[42]。由此可见,应当将人作为衡量万物的尺度且具有主观能动性与受规范约束的物种。

“所有的文化或文明都是人的‘能在’与‘应在’不断从潜能到现实,从应然到实然不断反复交替地辩证运动的结果,都是人的历史的组成部分。”^[29]从具有理性、属于政治动物、会使用语言符号、具有主观能动性、受规范化约束等角度对“人”进行的界定,都在不同程度上受到了“人是万物的尺度”这一论断的影响^[43]。首先,人们认为,只有人类自身才是衡量万物价值以及属性的应然

尺度;其次,人类作出价值以及属性评判所依靠的内在工具是理性,外在工具则是语言符号;再次,人类依靠社会化的维持及其良性发展维护已固定的价值观念,通过综合性奖惩措施维护规范的公众认同。总之,唯有人才能够决定某种物种是否具有价值及其价值高低。AI 刑事责任主体肯定论之所以认为,应当赋予强 AI 产品刑事责任主体地位,是没有持有一种合理的“人”论。肯定论持有的是一种机械式“人”论,不当混淆了“表象”与“内在”之区别。不应将强 AI 产品作为一个具有自主意识的人对待,正如人们不会将动物作为人对待一样,因为其无法在某种程度上本能地意识到并通过行动实现自身意愿。本文不能保证文中所述观点,可以让人们得到 AI 刑事责任主体问题被合理解决的所有预期得以实现的结果,却能够接近想要破解的答案,可以保证整个应对机制在一种合理的、递进的期待中得以维持、改进。当然,偏向于法理与哲理论证的策略确受自身能力所限,因此自知仅靠法理与哲理论证维系基本的判断力明显不及那些既取法于法理与哲理论证又工于 AI 技术挖掘的方法选择,好在有人早早地为法理与哲理论证作了以下辩护,“紧张的意识很可能在理性全然缺失的情况下产生,……仅仅由不着边际的幻象和欲望所充斥的心灵看起来能万无一失地追求某些东西,但是它们并不具备人类灵魂的高贵特性;因为上述追求不会因目标的任何前景得到启发”^[44]。

四、结 论

弱 AI 产品与强 AI 产品都不应是刑事责任主体,而是人类可控制的依靠电能存续的 AI 机械产品,应被作为人类实现自由而全面发展的工具。对于由 AI 产品衍生的刑事风险而言,当行为人故意利用了 AI 技术或产品实施犯罪行为时,则具体利用人,而非 AI 产品,应受到刑法惩罚。对于科研人员等主体过失造成 AI 产品实施了犯罪行为时,则依据预防可能性等过失犯原理进行规制即可。法律有其固有的调整范围,其既不是虚拟之物,也不是联想或幻想的产物,而是社会现实及其要求的理性反映。刑事法适用的核心部分及刑事归责的基本原理不会因 AI 技术的出现而发生根本变化。法律人必须承认自身知识的局限性,面对不稳定、不清晰的事物,在法规规范层

面上,宜以原则性条款应对,不宜作出细致性规定。AI 刑事责任主体肯定论是某种情感或幻想的产物,不当坚持了一种机械式“人”论,应当将人作为衡量万物的尺度且具有主观能动性受规范约束的物种。

弱 AI 产品正逐渐在医疗卫生、生物工程及司法审判等领域为人类带来“福音”。比如在当前进行的“智慧法院”建设中,基于诉讼的便捷化、审判的公正化、执行的高效化等目的考量,利用大数据分析实现了海量文书检索等目标,充分展现了将 AI 技术与法律相结合所带来的“红利”。但人类仍需时刻铭记自己是万物的主人,并警惕弱 AI 产品的现有及潜藏缺陷。比如当下的 AI 技术难以消除算法歧视,可能通过形成“自我实现的歧视性反馈循环”进一步固化歧视,并以算法决策的方式损害规则适用的公正性,甚至扩大刑事司法的不公正^[45]。鉴于“刑事司法判断,有可能以人工智能提供的裁判规则作为依据……理应设置严格的禁区,防范可能的风险”^[46]。无论 AI 产品掌握的数据如何精准、“思维”运算多么快捷,司法机关和社会公众都不会将案件的审判权赋予 AI 产品,其只能被作为辅助法官公正审判的角色。积极促进 AI 技术的发展,并对其可能带来的风险进行防范本属常识,必须“坚持以法律为边界,避免偏离有益于人类社会发展的方向”^[47]。当前宜由相关部门在对行业发展状况、技术水平限制、社会利益维护、个人权利保障等因素进行综合考量的基础上,遵循法治程序,确立相关的行业技术标准、安全义务标准和个人数据保护标准。换言之,不仅需“依靠算法的顶层设计”^[48],且需将法治规范嵌入 AI 发展的每个环节。

参考文献:

- [1] 刘宪权. 人工智能时代的刑事风险与刑法应对[J]. 法商研究, 2018(1):3.
- [2] 刘宪权. 人工智能时代的“内忧”“外患”与刑事责任[J]. 东方法学, 2018(1):134.
- [3] 时方. 人工智能刑事主体地位之否定[J]. 法律科学, 2018(6):67-75.
- [4] 张保生. 人工智能法律系统的法理学思考[J]. 法学评论, 2001(5):21.
- [5] 王耀彬. 类人型人工智能实体的刑事责任主体资格审视[J]. 西安交通大学学报(社会科学版), 2019,39(1):131.
- [6] 博登海默 E. 法理学——法律哲学与法律方法[M]. 邓正来,译. 北京:中国政法大学出版社, 2004:473.
- [7] 刘宪权,胡荷佳. 论人工智能时代智能机器人的刑事责任能力[J]. 法学, 2018(1):40.
- [8] 储陈城. 人工智能可否成为刑事责任主体[N]. 检察日报, 2018-04-19(3).
- [9] 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):42.
- [10] 彭文华. 自动驾驶车辆犯罪的注意义务[J]. 政治与法律, 2018(5):96.
- [11] 李帅. 人工智能威胁论:逻辑考察与哲学辨析[J]. 东北大学学报(社会科学版), 2019,21(1):18.
- [12] 张明楷. 刑法学(上)[M]. 5 版. 北京:法律出版社, 2016:142.
- [13] 储陈城. 人工智能时代刑法归责的走向——以过失的归责间隙为中心的讨论[J]. 东方法学, 2018(3):27-37.
- [14] Lehr D, Ohm P. Playing with the Data: What Legal Scholars Should Learn About Machine Learning[J]. U. C. Davis Law Review, 2017,51(2):653-718.
- [15] 高奇琦,张鹏. 论人工智能对未来法律的多方位挑战[J]. 华中科技大学学报(社会科学版), 2018,32(1):86-96.
- [16] 玛格丽特·博登. AI:人工智能的本质与未来[M]. 孙诗惠,译. 北京:中国人民大学出版社, 2017:69.
- [17] Penrose R. Shadows of the Mind: A Search for the Missing Science of Consciousness[M]. Oxford: Oxford University Press, 1994:12.
- [18] 杜孝义,杜孝玺. 人不可能“长生不老”[J]. 科学与无神论, 2002(5):43.
- [19] 石佑启,李锦辉. 生存与合作:进化论视角下法律的元价值[J]. 世界哲学, 2018(5):99.
- [20] 阿蒂亚 P S,萨默斯 R S. 英美法中的形式与实质——法律推理、法律理论和法律制度的比较研究[M]. 金敏,陈林林,王笑红,译. 北京:中国政法大学出版社, 2005:61.
- [21] 张建文. 格里申法案的贡献与局限——俄罗斯首部机器人法草案述评[J]. 华东政法大学学报, 2018,21(2):38.
- [22] 邱兴隆. 刑罚理性辩论——刑罚的正当性批判[M]. 北京:中国检察出版社, 2018:1.
- [23] 魏超. 巨额财产来源不明罪法益与主体新论——信赖说之提倡与国家工作人员之证立[J]. 东北大学学报(社会科学版), 2018,20(4):400.
- [24] 克劳斯·罗克辛. 刑事政策与刑法体系[M]. 2 版. 蔡桂生,译. 北京:中国人民大学出版社, 2011:72.
- [25] 刘宪权. 人工智能时代我国刑罚体系重构的法理基础[J]. 法律科学, 2018(4):47-55.
- [26] 艾伦·诺里. 刑罚、责任与正义[M]. 杨丹,译. 北京:中国人民大学出版社, 2009:29.
- [27] 吴从周. 概念法学、利益法学与价值法学[M]. 北京:中国法制出版社, 2011:22.
- [28] 色诺芬. 回忆苏格拉底[M]. 吴永泉,译. 北京:商务印书馆, 1984:28.
- [29] 张志平. “人是什么”:一种语义学和现象学的分析[J]. 江海学刊, 2015(4):16-17.
- [30] 卡西尔. 人论[M]. 甘阳,译. 上海:上海译文出版社, 1985:34.
- [31] 亚里士多德. 政治学[M]. 吴寿彭,译. 北京:商务印书馆,

1983:7.

[32] 卡尔·波普尔. 客观知识:一个进化论的研究[M]. 舒炜光,卓如飞,周柏乔,等译. 上海:上海译文出版社, 2005:139.

[33] 石福祈. 卡西尔回答了“人是什么?”这一问题了吗? [J]. 现代哲学, 2015(2):74-82.

[34] 张劲松. 人是机器的尺度——论人工智能与人类主体性[J]. 自然辩证法研究, 2017,33(1):49-54.

[35] 韩东晖. 人是规范性的动物——一种规范性哲学的说明[J]. 中国人民大学学报, 2018(5):2-8.

[36] 彼得·斯坦,约翰·香德. 西方社会的法律价值[M]. 王献平,译. 北京:中国法制出版社, 2004:3.

[37] 王国有. 西方理性主义及其现代命运[J]. 江海学刊, 2006(4):55-60.

[38] 策勒尔. 古希腊哲学史纲[M]. 翁绍军,译. 济南:山东人民出版社, 1992:87.

[39] 赵本义. “人是万物的尺度”的新解读[J]. 人文杂志, 2014(6):6-12.

[40] 钟宏彬. 法益理论的宪法基础[M]. 台北:台湾元照出版公司, 2012:227.

[41] 韦恩·莫里森. 法理学——从古希腊到后现代[M]. 李桂林,李清伟,侯健,等译. 武汉:武汉大学出版社, 2003:139.

[42] 康德. 实践理性批判[M]. 邓晓芒,译. 北京:人民出版社, 2016:21.

[43] 张法. 人是什么:中、西、印思想的不同向路[J]. 学术月刊, 2010(10):6.

[44] 乔治·桑塔亚纳. 常识中的理性[M]. 张沛,译. 北京:北京大学出版社, 2008:4.

[45] 王禄生. 大数据与人工智能司法应用的话语冲突及其理论解读[J]. 法学论坛, 2018(5):139.

[46] 黄京平. 刑事司法人工智能的负面清单[J]. 探索与争鸣, 2017(10):88.

[47] 皮勇. 人工智能刑事法治的基本问题[J]. 比较法研究, 2018(5):149.

[48] 李彦宏. 智能革命:迎接人工智能时代的社会、经济与文化变革[M]. 北京:中信出版集团, 2017:312.

(责任编辑:王 薇)

(上接第 89 页)

[5] Beever A. The Structure of Aggravated and Exemplary Damages[J]. Oxford Journal of Legal Studies, 2003,23 (1):87-110.

[6] Cunnington R. Should Punitive Damages be Part of the Judicial Arsenal in Contract Cases? [J]. Legal Studies, 2006,26(3):369-393.

[7] 王泽鉴. 王泽鉴法学全集(第 14 卷)[M]. 北京:中国政法大学出版社, 2003:97.

[8] 王利明. 美国惩罚性赔偿制度研究[J]. 比较法研究, 2003(5):1.

[9] 杨立新. 侵权责任法[M]. 上海:复旦大学出版社, 2010:87.

[10] Calabresi G. The Cost of Accidents: A Legal and Economic Analysis [J]. American Political Science Association, 1973,67(4).

[11] 何勤华. 法的移植与法的本土化[M]. 北京:法律出版社, 2001:107.

[12] 马克思. 德谟克利特的自然哲学和伊壁鸠鲁的自然哲学的差别[M]//马克思,恩格斯. 马克思恩格斯全集(第 1 卷). 北京:人民出版社, 2002:76.

[13] 韩登池. 司法三段论——形式理性与价值理性的统一[J]. 法学评论, 2010(3):140.

[14] 小詹姆斯·A. 亨德森,查理德·N. 皮尔森,道格拉斯 A. 凯撒,等. 美国侵权法——实体与程序[M]. 王竹,丁海俊,董春华,译. 北京:北京大学出版社, 2014:591.

[15] 文森特·R. 约翰逊. 美国侵权法[M]. 北京:中国人民大学出版社, 2004:68.

[16] 柏拉图. 法律篇[M]. 上海:上海人民出版社, 2001:382.

[17] 威廉·M. 兰德斯,理查德·A. 波斯纳. 侵权法的经济结构[M]. 王强,杨媛,译. 北京:北京大学出版社, 2005:204.

[18] Holmes O W. Review of C. C. Langdell, Summary of the Law of Contract[J]. American Law Review, 1880, 14 (1):233-235.

[19] 白江. 我国应扩大惩罚性赔偿在侵权责任法中的适用范围[J]. 清华法学, 2015(3):111.

(责任编辑:王 薇)