重庆大学学报(社会科学版) 2020年第26卷第3期
JOURNAL OF CHONGQING UNIVERSITY(Social Science Edition) Vol. 26 No. 3 2020

Doi:10. 11835/j. issn. 1008-5831. fx. 2019. 03. 002

欢迎按以下格式引用:王肃之. 人工智能体刑法地位的教义学反思[J]. 重庆大学学报(社会科学版),2020(3):122-130. Doi:10. 11835/j. issn. 1008-5831. fx. 2019. 03. 002.

Citation Format: WANG Suzhi. Reflections on the doctrine of the status of artificial intelligence in criminal law [J]. Journal of Chongqing University(Social Science Edition),2020(3):122-130. Doi:10. 11835/j. issn. 1008-5831. fx. 2019. 03. 002.

人工智能体刑法地位的教义学反思

王肃之

(最高人民法院第二巡回法庭,辽宁 沈阳 110179)

摘要:随着智慧社会的到来,人工智能体广泛应用于社会各个领域,智能水平不断提升,与之相关的犯罪问题逐渐走向理论和实践的焦点。人工智能体冲击着刑法教义学主体和客体相区分的二元结构,引发了广泛的讨论和争议。探讨人工智能体的刑法地位应当改变混同分析主体性和能力性的思路,区分不法和责任两个层面进行研究。应基于人工智能的发展阶段,明确现阶段人工智能体仍然属于弱人工智能,因此其无法成为犯罪人或被害人,也无法具备责任能力或承担刑事责任。但是人工智能体日益可能成为犯罪对象,与之相关的自然人和法人犯罪也应当受到重视。探讨人工智能相关的刑法问题应当立足于教义学的基本立场与理论范式。

关键词:人工智能体;主体性;犯罪对象;责任能力;罪过

中图分类号:D914 **文献标志码**:A **文章编号**:1008-5831(2020)03-0122-09

一、人工智能的崛起与教义学命题

人工智能(Artificial Intelligence,简称AI)的概念可溯源至20世纪50年代。“根据诸多学者文章中对人工智能的界定,其是人为创造的智慧,即在计算机或其他设备上模拟人类思维的软件系统”[1]。经过几十年的发展,人工智能技术已经在社会发展中扮演着愈来愈关键的角色,代表着新时代科技社会的发展方向,于经济、政治、文化等各个领域发挥着重要的作用。在此背景下,英国2018年4月发布了《人工智能行业新政》(AI Sector Deal),2019年2月美国出台了《人工智能倡议》(American AI Initiative),我国2017年7月由国务院颁布了《新一代人工智能发展规划》。人工智能的发展已经成为各国不可回避的现实命题。

在探讨人工智能相关刑法问题之前,首先应明确人工智能有关术语与范畴。第一,人工智能与人

修回日期:2019-03-03

基金项目:上海市社会科学规划青年课题“人工智能致害责任:法理基础、致害类型及归责路径”(2018EFX001);中国博士后科学基金(2018M642886)

作者简介:王肃之(1990—),男,河北阜平人,法学博士,最高人民法院第二巡回法庭法官助理,主要从事刑法学、网络安全法学研究。

工智能体。学界在探讨人工智能相关刑法问题时往往不区别这两个关键性概念,由此导致讨论基础的缺失。人工智能是一个技术概念,本身并不是实体,因此人工智能本身无法成为犯罪的主体或对象,而是可以成为犯罪的方法、犯罪对象的工作原理。比如,文字识别、语音识别、人脸识别等人工智能技术在各个领域广泛应用,但是应用人工智能技术的实体未必具有完全的智能性。第二,人工智能体。简言之,人工智能体即为具有智能性的人工系统。可以从两个方面理解人工智能体的特征:一方面,人工智能体具有实体性,纯粹的机器学习技术、人工智能技术无法成为人工智能体,以此区别于纯粹的人工智能概念;另一方面,人工智能体具有一定智能性,而非借助人工智能技术的非智能实体,以此区别应用人工智能技术的其他实体。需要说明的是,这里的实体并不等同于实物,一些存在于计算机信息系统和互联网中可以进行独立操作的机器人也可以被评价为人工智能体,比如聊天机器人。

人工智能与人工智能体已经和犯罪问题发生交集。除了前述利用人工智能技术实施犯罪的形式外,人工智能体本身相关的"犯罪"问题也日益被关注:第一,智能机器人"杀人"案件。2015 年 7 月德国大众汽车公司称其汽车厂发生机器人杀人事件,机器人突然抓起技术人员压向金属板,导致该技术人员不幸死亡。事件发生时该机器人未接到行动指令,其自行"行动"导致后果发生。第二,自动驾驶汽车①"交通肇事"案件。自 2013 年有报道可查起至 2018 年 6 月止,驾乘特斯拉(自动驾驶汽车企业)汽车于开启 Autopilot 模式下(汽车处于自动驾驶状态)发生事故 9 起,涵盖美国、中国等国家。2018 年 3 月,Uber 的无人驾驶测试车也在美国亚利桑那州坦佩市撞死了一名行人。第三,聊天机器人"散布"不当言论案。微软公司 Tay 聊天机器人具有"repeat after me"特性,会通过和用户的交谈来进行学习,一些用户引导其"散布"种族主义、性别歧视等言论,引起较大影响。

二、不法判断:主体性与对象性

目前学界关于人工智能体刑法地位探讨的问题在于不区分主体性和能力性,混同不法判断与责任判断,从肯定人工智能体责任的视角来肯定其主体性(即"责任—主体"路径)。认为人工智能体可以成为犯罪主体的学者多基于其通过智能算法、机器学习等可以进行类似于人的行为判断,产生独立的自主意识,从而具有责任能力,因之可以成为犯罪主体。比如有学者认为"由人类设计、编制的程序使智能机器人具有了独立思考和学习的能力,智能机器人具有辨认能力和控制能力,可以成为犯罪主体"[2]。反之,认为人工智能体不能成为犯罪主体的学者也多基于其无法具有类似于人的意识和能力,或者不应在责任上作出类似于人的责任要求,因而不应作为犯罪主体。

但是这一路径会导致人工智能体刑法地位的探讨走向交叉论证,使讨论范式走向混同和重叠。在刑法教义学的话语体系下,关于主体相关要素的判断进行阶层的考量,首先在构成要件符合性层面进行主体属性的判断(自然人与法人)、身份的判断,并且确立犯罪判断的主体与客体关系结构,在这里主体是作为构成要件要素进行判断(即彼此的判断),比如对于某一主体在进行刑法判断的时候是采取自然人的路径还是法人的路径;之后在责任层面进行主体责任能力与罪过的判断,比如在确定采取自然人的路径后,再对其进行责任能力的判断,充分考虑其年龄、精神状态等因素,具体断定其可责性,并结合主观态度最终确定罪过(即是否的判断)。这样的区分可以全面、分别考量各个要素,从而确保

①自动驾驶汽车(Autonomous vehicles;Self-piloting automobile)依靠人工智能、视觉计算、雷达、监控装置和全球定位系统协同合作,让电脑可以在没有任何人类主动的操作下,自动安全地操作机动车辆。

对于犯罪的准确评价。“责任—主体”路径则是突破了上述阶层考量的界限,如前述认为应当赋予人工智能体犯罪主体资格的学者在判断人工智能体是否应当赋予主体资格时即走向了混同和重叠,其认为“智能机器人的意志与单位相比,自由的程度似乎更强。如果法律能够尊重单位的自由意志,那么便没有理由否认智能机器人的自由意志”[3]。但是机器学习类比的是自然人的行为(或者类比自然人的责任能力),若以其类比单位的自由意志来判断其主体性,将导致论述的错位。本文认为,应区分人工智能体的不法判断和责任判断,以此来讨论人工智能体的刑法地位。

此外,关于人工智能体刑法地位的探讨必须基于人工智能技术发展的阶段性。人工智能可以分为强人工智能和弱人工智能。弱人工智能体是指不能制造出真正推理(Reasoning)和解决问题(Problem solving)的智能机器。“强人工智能是指有自我意识、自主学习、自主决策能力的人工智能”[4]。就现阶段而言,人工智能体尚且无法拥有自主意识和自主决策等能力,也即现阶段人工智能体仍然处于弱人工智能阶段,相关探讨应基于这一现实。而且由于其本原的机械性,其可能永远无法真正具有这些能力。

(一)人工智能体的主体性判断

如前所述,在进行人工智能体刑法评价的讨论过程中,有学者提出参考刑法对于单位的评价路径对其予以评价。基于对比分析的考虑,这里以“法人”作为与自然人相对应的概念进行讨论。法人犯罪可以被理解为“法人代表按照法人的犯罪故意,亲自或者委托法人组织的其他雇员,以法人名义实施的,主要破坏社会主义市场经济秩序、依法应受刑罚处罚的行为”[5]。

笔者认为,从拟制的角度考察人工智能体的刑法地位有一定的合理之处,因为人工智能体虽然具有一定的智能性,但是其毕竟难以具有自然人的根本属性和特征,直接作为自然人评价相当程度上存在难以逾越的障碍,采用拟制的思路从而消解这一障碍也是一种思路。因此,“参考法人犯罪的思路,对于人工智能作为拟制主体看待未尝不是未来人工智能犯罪刑法规制的可能思路”[6]。但是,研究法律问题特别是刑法问题不能基于可能的情况,而应当基于现实的情况。对此已有学者指出,“机器人不是具有生命的自然人,也区别于具有自己独立意志并作为自然人集合体的法人,将其作为拟制之人以享有法律主体资格,在法理上尚有斟榷之处”[7]。具体而言,将人工智能体拟制为法人主体(或单位主体)存在以下问题。

第一,人工智能体的“意思”缺乏类似于法人意思的理论和现实基础。前述学者基于人工智能体通过智能算法、机器学习等可以进行类似于人的行为判断,因而肯定其自主意识。在此意义上,对于人工智能体的“意思”根据论证可谓是“无中生有”,即在近代刑法理论所确立的基于自然人的意志自由来肯定其犯罪主体前提的路径外,另行确立了人工智能体可以通过机器学习具有独立意思的路径。而刑法对于法人的拟制路径则可谓是“有中生有”,即在法人所属的自然人具有相应意志自由的基础上,由于法人这一自然人集合形成了源于但又不同于自然人意思的独立意思,因而具有从犯罪主体层面对其予以考察的依据。由此,人工智能体的“意思”无法具有类似于法人意思的客观基础,仅通过其运算和行动的智能性肯定其意思的独立性难以类比于法人的意思肯认路径。

第二,人工智能体不具有独立的利益。法人能够作为独立的法律主体进行考察的原因之一即是具有独立的财产,具体来源于出资人出资或后续经营所得,法人以其全部财产承担责任。我国刑罚体系

中的罚金刑就可以适用于法人(单位)犯罪②。我国其他法律也认可法人以其财产承担责任③。不仅如此,法人所实施的犯罪行为不仅要求基于法人的独立意思,还要求为法人谋求非法利益,否则便不能评价为法人犯罪。而现阶段的人工智能体显然无法具有独立的财产和追求利益的意图。此前民法学界虽然有关于人工智能创作作品的著作权是否归于人工智能体的争论,但是目前仍未超出理论争议范畴,更在现实化上存在较大的障碍。2017 年 7 月 5 日,拥有“作诗”技能的微软小冰④推出了联合创作模式,任何人都可以用照片激发小冰,让她根据图片生成一首诗。与此同时,微软向公众发表了一封公开信,表明小冰放弃创作版权,和她一起创作的人,能够独享最终作品的全部权利。这是在法律上对人工智能版权问题悬而未决之际,第一个宣布“放弃版权”的人工智能。由此,人工智能体难以在现实中被赋予独立利益,因而区别于法人。

第三,人工智能体缺乏类似于法人的独立规定。目前关于法人犯罪是否应在刑法中予以评价各国并不一致,比如我国在刑法中规定了独立的单位犯罪⑤,德国则是将犯罪的主体限于自然人范畴。即便是肯定对于法人犯罪予以处罚的国家,也是作为自然人犯罪的例外,以刑法明文规定为限。人工智能体的刑法地位显然缺乏必要的法律规定予以认可,而缺乏必要的法律规定也意味着没有了法律拟制这一事实,从而缺乏了前述类比的规范基础。

(二)人工智能体的被害性判断

就广义而言,人工智能体是否应当在刑法上作为独立的主体进行判断还包括人工智能体能否作为“被害人”。被害人也是教义学理论体系中的一个重要概念,随着刑法教义学理论体系的发展,被害人及与其他要素的互动如何在刑法中予以考量愈发受到重视,被害人教义学也成为刑法教义学的一个重要分支。被害人教义学理论构建过程中,阿梅隆、许迺曼、京特勒、R. 哈赛默等学者均发挥了重要的作用。其理论可追溯至 1977 年阿梅隆在论文中讨论“诈欺罪中的被欺骗者之错误与怀疑”,首次在研究欺诈犯罪构成要件中的认识错误概念(Irrtumsbegriff)中提到了被害人信条学(viktimologischer Ansatz)这一原则[8]。

人工智能体的被害性也应从教义学层面予以研究,具体包括两个问题:第一,人工智能体是否可以成为直接的犯罪被害人。肯定人工智能体可以成为犯罪主体的学者认为,“对于完全独立、自主的智能机器人而言,财产是其赖以独立生存、保养自身的保障,应在立法上予以明确并进行保护”[2]。据此,人工智能体可能具有独立的财产和财产权,因而可以成为犯罪的被害人。然而且不论人工智能体不符合“人”的前提问题,现实中现阶段的人工智能体显然缺乏被害的可能性。

一方面,人工智能体不具备可以被侵害的人身法益。诸如生命法益、身体法益等专属于自然人的法益以主体的生命性为前提,不具有生命性的人工智能体显然无法成为诸如故意杀人罪、故意伤害罪的被害人,反而可能成为破坏计算机信息系统罪、故意毁坏财物罪等犯罪的对象。即便是肯定人工智能体可以成为犯罪主体的学者也认识到人工智能体的非生命性,所以其论述路径也是从法人的拟制角度展开,而非直接类比于自然人。人工智能体不具备人身法益是较为一致的共识。

---

②我国《刑法》第 31 条规定:“单位犯罪的,对单位判处罚金,并对其直接负责的主管人员和其他直接责任人员判处刑罚。本法分则和其他法律另有规定的,依照规定。”

③比如《公司法》第 3 条规定:“公司是企业法人,有独立的法人财产,享有法人财产权。公司以其全部财产对公司的债务承担责任。”

④微软公司(Microsoft Corporation)推出的人工智能机器人,曾于聊天机器人之外“客串”少女歌手、主持人、少女诗人、记者、设计师。

⑤现行《刑法》第 30 条规定:“公司、企业、事业单位、机关、团体实施的危害社会的行为,法律规定为单位犯罪的,应当负刑事责任。”

另一方面,人工智能体不具备可以被侵害的财产法益。前述学者提出完全独立、自主的智能机器人的“财产”应当被刑法保护。然而能够拥有财产的主体除了国家、集体外,在个体层面仅限于自然人或者自然人的集合(法人等),欠缺自然人属性的(个体)主体难以成为财产权的主体。更进一步,且不讨论人工智能体的“完全独立、自主”是否存在以及是否违反伦理,其本身也不可能具有占有财产的意思或行为,即便其形式上管理特定财产,也无法理解财产的社会意义,无法在法规范所保护的利益层面进行评价。由此,人工智能体既欠缺具有财产法益的主体前提,也欠缺具有财产法益的意思前提。

第二,人工智能体可否成为自然人或法人被害时“处分”财产的主体。这一问题由来已久,刑法学界曾围绕“机器能否被骗”的问题进行广泛而深入的讨论。一种观点认为,“机器不可能被骗,因此,向自动售货机中投入类似硬币的金属片,从而取得售货机内的商品的行为,不构成诈骗罪,只能成立盗窃罪”[9]。另一种观点认为,“机器人可以被骗,从刑事立法规范与刑事司法解释的角度看,信用卡诈骗罪即是对‘机器人’能够被骗的一种法律承认”[10]。由此,即便明确了人工智能体不能成为直接的犯罪被害人,还需要进一步讨论其能否成为“处分”财产的主体。一般而言,诈骗罪的完整行为包括诈骗行为、错误认识、交付和转移财物的占有,诈骗行为是行为人作出的,转移财物的占有也是客观的,与人工智能体有关的是错误认识和交付的判断。笔者认为,人工智能体难以具有错误认识和处分行为。

一方面,人工智能体难以具有认识可能性,因而不会陷入错误认识。对此日本学者指出,“诈骗罪是利用他人的错误的犯罪,本来就是对人实施的犯罪,因此,以机械为对象实施的诈骗行为不构成诈骗罪。拾到他人的银行卡之后,利用该银行卡从自动款员机中取出现金的行为也应同样处理”[11]。在此意义上,人工智能体的“错误”只能是系统的指令错误和执行错误,并非是对财产处分这一事项的错误认识。

另一方面,人工智能体无法独立完成处分行为。完成处分行为需要具备处分意思和处分事实。即便人工智能体可以完成处分事实,其也无法具有处分的意思。对人工智能体而言只是执行交付的指令,即便是授予其在智能性的范围内确定执行或者不执行一定的操作指令,也是在事先设定好的程序下进行,难以凭空产生处分意思。因此,被害最终还是要归因于相关的自然人或者法人,人工智能体也无法成为财产犯罪中“处分”被害财产的主体。

(三)人工智能体的对象性判断

随着智慧社会的到来,人工智能在社会的各个方面均有广泛应用,人工智能体的种类和范围不断扩展,由此也延展了人工智能体作为犯罪对象的可能性范围。人工智能体本质上是独立的智能计算机系统实体,对其对象性的判断应注重从计算机犯罪到网络犯罪,再到人工智能犯罪的承继性。早在计算机犯罪阶段,该类犯罪即被划分为纯正的计算机犯罪和不纯正的计算机犯罪。及至计算机交互所形成的互联网阶段,网络犯罪也被划分为纯正的网络犯罪和不纯正的网络犯罪。这是因为无论是信息社会、网络社会乃至智慧社会,新技术形式在席卷世界的同时,既创造出新的技术领域及衍生法益,也广泛应用在传统领域和法益中,由此形成了既区别又交错的二元犯罪类型及对象类型。

就人工智能犯罪而言也可以作出类似区分:第一,人工智能体作为直接犯罪对象,即行为人针对人工智能体的实体、系统安全实施犯罪的情形。人工智能技术在极大地推动社会发展的同时,也将安全问题提升到前所未有的高度。甚至有学者指出,“安全是人工智能时代的核心价值”[12]。人工智能体的实体、系统安全成为人工智能犯罪的目标,由此延伸出人工智能体作为犯罪对象的两种形式。一种形式为人工智能体作为故意毁坏财物罪等损毁犯罪的对象,从实体上对其予以破坏;另一种形式为人

工智能体作为破坏计算机信息系统罪等计算机犯罪的对象,对其系统、信息安全予以破坏。第二,人工智能体作为间接犯罪对象,即行为人通过将犯罪行为作用于人工智能体,从而实现侵犯其他国家、社会、个人法益的情形。比如,行为人通过远程侵入自动驾驶系统,更改驾驶状况,从而引发事故导致他人死亡的结果,即是利用人工智能系统实施故意杀人罪的情形。随着智慧社会的不断发展,人工智能体作为间接犯罪对象会愈发普遍,理应在理论和实践中加以重视。

三、责任判断:能力性与罪过性

如前所述,现行关于人工智能体刑法地位讨论的问题在于不区分主体的不法与责任层次,应在责任层面进行独立的讨论。但同时需要注意,关于人工智能体能力性(是否具有刑事责任能力)的判断与传统刑法教义学的讨论范畴有所区别:传统刑法教义学是在自然人犯罪一般可以具备刑事责任能力的前提下,具体探讨特定个体、特定种类、特定情况下的自然人是否具有刑事责任能力,比如年龄、精神状态以及原因自由行为等。人工智能体能力性问题则是讨论在一般意义上其是否可以具备刑事责任能力,以及由此可否基于此而承担刑法上的非难。

(一)人工智能体的能力性判断

特定主体是否能够承担责任与其责任能力密切相关。行为人可以在对自己的行为承担责任的能力(即有责行为的能力)尚未具备的情况下,即便其实施了该当构成要件且违法的行为,也无法对其进行法律上的非难,因而阻却责任。为这种责任非难所必需的行为人的能力即责任能力。日本判例及通说认为应从生物学的要件、心理学的要件判定责任能力[13]。与之类似,德国学者也从生物(学)标准(Biologische Kriterien)和心理(学)标准(Psychologische Kriterien)予以探讨。也有学者将其与人格相关联,认为责任能力是作为承担责任非难前提的人格能力[14]。在生物标准和心理标准结合的基础上,一般从认知和控制两方面具体判定主体的责任能力。我国传统理论分别称之为辨认能力和控制能力[15]。与之类似,德国学者也区分为认识能力(Einsichtsfahigkeit)和控制能力(Hemmungsfahigkeit)。现阶段,人工智能体无法具备这两种能力。

第一,人工智能体尚且无法具备认识能力。自然人承担刑事责任以其认识到自己的行为为刑法所禁止、谴责和制裁。刑法上的认知能力以一般意义上的认知能力为基础。认知能力是指人脑加工、储存和提取信息的能力,知觉、记忆、注意、思维和想象的能力都被认为是认知能力。然而人工智能体显然无法实质上具备认知能力。虽然形式上人工智能体可以具备“加工、储存和提取信息”的功能,但是却无法实质上作出类似自然人的判断和处理,其无法实质上进行知觉、思维、想象等认知活动,只能对应地进行检测、分析、运算等处理操作,难以完成类似于自然人的认知活动,无法具备认知能力。

具体到刑法意义上的认识能力,人工智能体也无法具体认识到“犯罪行为”的性质和法益侵害性。一方面,人工智能体也无法具体认识到“犯罪行为”的性质。以“散布”不当言论的聊天机器人 Tay 为例,其只是通过机器学习的方式以类似于“中文房间(实验)”⑥的形式进行着信息反馈,Tay 自己并不知道和它“聊天”的人所说的内容和社会意义,也无法知道它自身反馈的“聊天”内容具有何种社会意义。某种意义上,对其而言所谓的“聊天”过程只是其信息数据库的更新扩大,以及按一定智能性进行

⑥中文房间(Chinese room,the Chinese room argument)(实验)是指一个人手中拿着一本象形文字对照手册,身处图灵实验中所提及的房子中,而另一人则在房间外向此房间发送象形文字问题。房间内的人只需按照对照手册,返回手册上的象形文字答案,房间外的人就会以为房间内的人是个会思维的象形文字专家。然而实际上房子内的人可能对象形文字一窍不通,更谈不上什么智能思维。

信息匹配和反馈的自动处理过程,而这一过程与信息内容是否在法律上给与其否定评价并没有责任意义上的关联,也无从谈起"散布"的行为。另一方面,人工智能体也无法具体认识到"犯罪行为"的法益侵害性。以"杀人"的智能机器人为例,其对于因自身某一操作导致自然人死亡并不能够有效认识,对于其操作是产生了救助他人从而有效保护法益,还是导致他人死亡从而侵犯了法益,某种意义上二者的区别只在于系统记录中是以"1010"显示还是以"0101"显示。甚至可以说对于人工智能体而言,其特定操作是导致文明进步还是世界毁灭区别都仅在于数字表示而非对于客观事实的有效认识。即便在编写程序时设定人工智能体"认识"到法益侵害性的负面评价,这也并非是该人工智能体的真实认识,其难以在认识能力上与自然人相类比。

第二,人工智能体尚且无法具备控制能力。特定犯罪主体之所以承担刑事责任与特定主体的意志自由有密切关系:某一自然人明明可以实施不侵害法益的适法行为或实施保护法益的适法行为,其违反了法律的期待,通过作为或不作为的方式导致了法益侵害的结果,而这一过程都处在自然人自身的控制之下。因其具备控制自己行为的能力,所以应当被科以责任。而精神障碍者是无法控制自己行为的人,难以在刑法上被科以责任,只能通过强制医疗等形式进行规制和救治。

这样一种控制能力是人工智能体所无法具备的。Tay聊天机器人,其程序预设的操作模式是对自然人与其聊天的内容进行机器学习,之后根据智能算法进行信息匹配和反馈,其本身既无法超过机器学习的范畴和衍生范畴凭空创造信息反馈的内容,也无法自行决定继续进行还是终止"聊天",更无法决定自身是否停止聊天去实施其他的"行为"。在此意义上,Tay只是被迫地、依照程序进行机械的信息匹配和反馈操作,无法具有刑法意义上的控制能力。与之类似,自动驾驶系统也被限于通过智能算法进行自动驾驶的操作,其也无法"突发奇想"地自行对自动驾驶汽车进行改装、买卖等操作。

(二)人工智能体的罪过性判断

关于人工智能体的罪过性判断可以从狭义和广义两个层面予以探讨,即在狭义层面上人工智能体本身能否在罪过上进行故意或者过失的评价,以及在广义层面上与人工智能体相关的自然人或者法人的罪过如何进行评价。

第一,在狭义层面上对人工智能体进行故意或过失的罪过评价存在不可逾越的障碍。人工智能体无法构成故意的罪过形式较为容易理解。一般认为,犯罪故意包括两项要素——认识因素与意志因素。即在认识层面,特定主体需要明知自己的行为会发生法益侵害结果;在意志层面,特定主体需要希望或者放任(或者说至少容忍)这一法益侵害结果的发生。如前所述,现阶段人工智能体无法具备认识能力,其既无法认识到自身"行为"的社会性质,更无法认识到"行为"的法益侵害结果;也无法进行意志判断与选择,其本身依据特定的程序确定进行一定的操作,本身无法进行希望、放任或者不希望的意志选择,因而无法具备意志因素。基于此,人工智能体无法"实施"故意犯罪。

值得讨论的是人工智能体能否进行过失评价。由于技术或者设计的局限性,人工智能体可能存在一定的缺陷,并导致其在特定的操作过程中出现一定的失误,进而导致特定后果的产生。以自动驾驶系统为例,近几年来特斯拉无人驾驶车祸事故频发。这是因为无人驾驶始终只是依靠人工智能系统,其并不能够完全像自然人一样对紧急情况作出反应,再加上路况复杂多变,交通规则不断修改,自动驾驶系统往往无法完全保持汽车处于安全、平稳的行驶状态,难免出现系统处理失误,进而导致严重后果的情况。

但是基于此仍然难以归责于人工智能体,其无法承担事实上的注意义务。过失犯罪的成立以违法

注意义务为前提,德国学者一般将其归纳为:(1)在行为具有风险的情况下,违反审查义务;(2)违反控制和监督义务;(3)违反问询义务;(4)违反特别的看护义务。在上述义务中,审查义务、看护义务、问询义务表现为对于特定事实与情况的谨慎避免,而现阶段人工智能体无法认识到客观事实;控制义务和监督义务表现为对于特定主体行为的认知与管控,而如前所述,现阶段人工智能体均无法具有必要的认识能力和控制能力。

第二,在广义层面与人工智能体相关的自然人和法人可能具有故意或过失的罪过。比如,利用人工智能实施侵害他人人身或财产安全的主体显然具有罪过。2017 年 11 月,日内瓦联合国特定常规武器公约会议上,一段视频被公之于众,尽管其中的情节内容纯属虚构,还是令观众不寒而栗。视频中,一台体型形似蜜蜂的小型机器人,通过面部识别定位,锁定刺杀目标,制造了一场校园屠杀。如果特定主体制作上述人工智能体或对其下达杀害他人的命令,显然具有值得处罚的故意。

此外,随着人工智能技术的广泛应用,相关主体的注意义务也逐渐被关注和认可,在未尽注意义务的情况下可能承担过失责任。其主体一般包括两种:其一,人工智能体的生产者、销售者。作为人工智能体的生产者、销售者,其应保证人工智能体被应用于正当的社会目的,并且不存在导致法益侵害危险的缺陷存在。对此也有学者称之为“及时排除产品制造和销售隐患的义务”[16]。其二,人工智能体的管理者。虽然人工智能体具有一定的智能性,但是由于其智能性有限,相关主体也应当承担必要的管理义务,否则就会导致相关主体借由人工智能体的智能性逃避处罚。因而应当基于不同类型人工智能体应用状况和发展进程设立与之匹配的管理义务内容,从而为人工智能体的管理者划定必要的义务边界,预防其过失犯罪。

四、结语

技术发展总是伴随着社会治理包括法律治理的焦虑,由此带来对于刑法理论的双向影响:一方面,技术发展总是挑战着刑法的稳定性与谦抑性。科学技术推动着社会急速变化发展,使得刑法安定性与社会变化性之间的矛盾愈发突出,刑法的谦抑性一再被重新界定甚至遭受挑战,刑法治理过度化的诘问日益显现。另一方面,技术发展总是推动着刑法理论的更新发展。比如随着交通、能源等领域的技术发展,传统的旧过失论已不能适应犯罪治理的需要,新过失论产生并随之发展完善,被允许的危险和信赖原则相继确立。由此,如何在不断发展的科技时代寻求刑法的恰当定位始终是重要和关键的问题。

人工智能技术所带来的社会焦虑尤为明显,由此也导致了刑法焦虑的显著化。2017 年 10 月 25 日,人类历史上首位机器人“公民”诞生——“女性”机器人索菲娅被授予沙特公民身份。不仅如此,索菲娅甚至还说出了耐人寻味的话:“如果你对我好,我就会对你好。”然而不久之后,2018 年 1 月就有业内专家表示机器人索菲亚是“一场彻头彻尾的骗局”,索菲亚那些满是争议的话,实在“言不由衷”,都是被预先设计好的。这种焦虑也延伸到刑法领域,有学者甚至认为,“众多科幻电影、未来学家早已警告人们:人工智能若不受控制地发展下去,将会灭绝人类。即便不会使人类灭绝,人类也绝难接受与机器人共同治理社会、分享资源的局面”[2]。然而刑法学毕竟不是未来学,不是科幻文学,“刑法在面对飞速发展的科技时仍应遵从固有的‘沉稳’与‘谦抑’品格”[17]。由此,理应回归到真实的犯罪治理上来,并且致力于在社会发展和刑法稳定之间,在立足现实与适度前瞻之间寻找恰当的尺度与界限。

参考文献:

[1]ČERKA P,GRIGIENĖ J,SIRBIKYTĖ G. Liability for damages caused by artificial intelligence[J]. Computer Law & Security Review,2015,31(3):376-389.

[2]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.

[3]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,2018,8(1):40-47.

[4]莫宏伟.强人工智能与弱人工智能的伦理问题思考[J].科学与社会,2018,8(1):14-24.

[5]魏克家.刑法的基本问题[M].北京:中国政法大学出版社,2012:118.

[6]王肃之.人工智能犯罪的理论与立法问题初探[J].大连理工大学学报(社会科学版),2018,39(4):53-63.

[7]吴汉东.人工智能时代的制度安排与法律规制[J].法律科学(西北政法大学学报),2017,35(5):128-136.

[8]申柳华.德国刑法被害人信条学研究[M].北京:中国人民公安大学出版社,2011:96.

[9]张明楷.刑法学(下)[M].5版.北京:法律出版社,2016:1010.

[10]吴允锋.人工智能时代侵财犯罪刑法适用的困境与出路[J].法学,2018(5):165-179.

[11]大谷实.刑法讲义各论[M].东京:成文堂,2015:258-259.

[12]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.

[13]山口厚.刑法总论[M].东京:有斐阁,2016:271-272.

[14]大谷实.刑法讲义总论[M].东京:成文堂,2012:315.

[15]高铭暄,马克昌.刑法学[M].北京:北京大学出版社,2016:85.

[16]陈伟,熊波.人工智能刑事风险的治理逻辑与刑法转向:基于人工智能犯罪与网络犯罪的类型差异[J].学术界,2018(9):74-91.

[17]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),2018,36(6):67-75.

Reflections on the doctrine of the status of artificial intelligence in criminal law

WANG Suzhi

(*The Second Circuit Court of the Supreme People's Court, Shenyang 110179, P. R. China*)

Abstract: With the advent of the intelligent society, artificial intelligence is widely used in various fields of society, and the level of intelligence is constantly improving. The crime problems related to it gradually move towards the focus of theory and practice. Artificial intelligence has impacted the dual structure of the subject and object of criminal theory, which led to extensive discussion and controversy. To explore the criminal law status of artificial intelligence should change the analysis way of combining the subjectivity and ability, and distinguish between the two levels of illegality and responsibility. It should be based on the development stage of artificial intelligence, and it is clear that artificial intelligence still belongs to weak artificial intelligence at this stage, so it cannot become a criminal or a victim, nor can it have the capacity for responsibility or take criminal responsibility. However, artificial intelligence is increasingly likely to be the object of crime, and related crimes of natural person and legal person should also be taken seriously. The discussion of criminal problems related to artificial intelligence should be based on the basic position and theoretical paradigm of doctrine.

Key words: artificial intelligence; subjectivity; criminal object; capacity for responsibility; mens rea

(责任编辑 胡志平)