

人工智能时代对传统刑法理论的挑战

刘宪权 朱彦

摘要:人工智能技术可能使犯罪行为的社会危害性发生“量”和“质”等方面的变化。如果行为人故意利用人工智能技术实施严重危害社会的犯罪行为,应根据犯罪行为较之于传统犯罪行为社会危害性的变化情况进行从平、从重和从轻三种不同方式的处理。行为人如果对人工智能技术所造成的危害结果负有预见义务,成立过失犯罪。智能机器人可以分为在设计 and 编制的程序范围内实施行为的智能机器人和在设计 and 编制的程序范围外实施行为的智能机器人,前者不具有独立的辨认能力和控制能力,后者具有独立的辨认能力和控制能力。辨认能力和控制能力即为刑事责任能力,具有刑事责任能力的行为人即可成为刑事责任主体。在设计 and 编制的程序范围外实施行为的智能机器人完全可以成为刑事责任主体。对实施严重危害社会犯罪行为的智能机器人适用刑罚具有必要性和可能性,可以通过重构我国刑罚体系,对犯罪的智能机器人增设删除数据、修改程序、永久销毁等刑罚种类。

关键词:人工智能;智能机器人;辨认和控制能力;责任主体;刑罚体系重构

中图分类号:DF611

文献标识码:A

文章编号:1674-9502(2018)02-044-08

作者:刘宪权,华东政法大学教授,博士生导师;朱彦,华东政法大学法学博士研究生。

毫无疑问,在过去的2017年里,“人工智能”是各行各业最热门的词汇,“人工智能”似一股洪流汹涌澎湃且不可阻挡地快速渗透进社会生活的方方面面,已经引起界内界外的各方关注。“这是最好的时代,也是最坏的时代”^①,有学者提出,人工智能既是人类社会的伟大发明,同时也会给法律秩序带来巨大的风险。其中,人工智能技术在民事主体法、著作权法、侵权责任法、人格权法、交通法、劳动法等诸多方面与现有法律制度形成冲突。^②笔者认为,在人工智能时代,刑法也不能忽视人工智能技术所带来的时代变迁,有必要从刑法学的角度来看待人工智能时代的到来。在时代更迭的洪流中,我们应当审慎对待人工智能技术可能带来的新的刑事风险,适当改变现有的刑法体系,以便应对和适应人工智能时代所可能带来的各种挑战和新的变化。

一、人工智能技术会引起犯罪行为的变化

(一) 犯罪行为中人工智能技术的应用

人工智能技术简单来说就是利用计算机程序模拟人类神经网络,深度学习基本规则和大数

本文系2014年度国家社科基金重大项目“涉信息网络违法犯罪行为法律规制研究”阶段性研究成果,项目编号:14ZDB147。
2016年度中国法学会重点委托课题“互联网金融犯罪的法律规制”阶段性研究成果,项目编号:CLS2016ZDWT48。

① [英]狄更斯:《双城记》,宋兆霖译,大众文艺出版社2000年版,第1页。

② 参见吴汉东:《人工智能时代的制度安排与法律规制》,《法律科学(西北政法大学学报)》2017年第5期。

据,从而可以对外界事物进行分析、判断以及决策的技术。人工智能技术现已应用于许多行业:在新闻媒体行业中已经出现了人工智能准确报道的新闻,^①在医疗行业中已经出现了高分通过执业医师资格考试的医疗机器人,^②在艺术行业已经出现了人工智能创作的优美的旋律^③……那么,人工智能技术是否也可以被犯罪行为人所利用,对犯罪行为“推波助澜”?笔者认为,答案应该是肯定的,人工智能技术在改变人们生活的同时,也必然带来一定的风险。人工智能技术完全可能使犯罪行为的社会危害性在“量”和“质”两方面发生变化。

其一,人工智能技术可能使犯罪行为的社会危害性发生“量”的变化。每一次科技的重大发展都可以带来社会生活时代的更迭,而每一次时代的更迭都会带来犯罪行为社会危害性的变化。当互联网时代来临的时候,一些传统的犯罪行为借助于网络可谓“如虎添翼”,社会危害性呈现出巨大的“量变”。^④而如今人工智能时代已然到来,传统的犯罪行为借助于人工智能技术可谓“如虎再添翼”,社会危害性再次发生“量变”。一方面,随着人工智能技术的不断普及与深化,一些重要的系统,例如金融系统和政府系统,深深依赖于人工智能程序,一旦人工智能程序出现了问题或者不法分子改造人工智能程序,那么完全可能造成灾难性的危害。另一方面,人工智能技术可以使犯罪工具或手段更加“智能”,例如证券投资者利用人工智能技术获得违背三公原则的速度优势进行交易获利,显然侵害了其他证券投资者的利益,完全可能构成比一般操纵证券市场行为更严重的操纵行为。

其二,人工智能技术可能使犯罪行为的社会危害性发生“质”的变化。人工智能技术往往与大数据紧密结合,人工智能技术的发展容易造成对信息的侵犯,从而产生新型侵犯信息类犯罪。为了使人工智能技术得到更为有效、广泛的应用,人工智能产品往往包含强大的数据库,而数据的集中本身就具有很高的风险。例如,以往黑客获取信息可能需要通过破坏不同的系统从而获得不同的数据,但数据一旦集中,黑客只需要通过攻击数据库就能获取海量信息。此外,面对强大的数据库,不法分子完全可能通过人工智能技术的分析得到关键、敏感信息(例如涉及国家安全的重要信息),并利用这些信息实施相关严重危害社会的犯罪,目前对于数据的安全防护手段可能无法跟上人工智能技术发展的速度,也就无法及时针对新的犯罪行为做出及时反应。另一方面,在数据保护系统并不健全的现在,数据的收集事实上也混杂着合法与非法的手段,涉及对公民个人信息的侵犯。然而数据的集中本身是人工智能技术的“生命之源”,科技的创新难免会产生风险。笔者认为,这种新型侵犯信息类的犯罪与传统侵犯信息类的犯罪相比,社会危害性有着“质”的区别,主要原因是一些涉及侵犯公民个人信息行为的主要目的是为了人工智能技术的革新。因此,对于此类行为,刑法应当审慎对待。

(二)人工智能时代犯罪行为变化的应对

由上述分析可知,在人工智能技术可能使犯罪行为的社会危害性发生“量”与“质”两方面变化的情况下,我国刑法应当重视人工智能技术所可能带来的危害。笔者认为,我国刑法应当对滥用人工智能技术的犯罪行为加以处罚。具体而言,可以从行为人故意和过失两方面分析行为

^① 参见《九寨沟地震,第一个发布该消息的竟是人工智能机器人》, http://www.sohu.com/a/163384062_391356, 搜狐网, 2017年12月21日访问。

^② 参见冯烁:《人工智能牵手医疗产业“现代华佗”给行业带来哪些惊喜》, <http://www.52ai.com/AIyingyong/3841.html>, 人工智能网, 2017年12月21日访问。

^③ 参见曾梦龙:《索尼用人工智能写了两首歌,你觉得怎么样》, <http://www.qdaily.com/articles/32649.html>, 好奇心网, 2017年12月21日访问。

^④ 参见刘宪权:《网络犯罪的刑法应对新理念》,《政治与法律》2016年第9期。

人滥用人工智能技术的刑事责任。

首先,笔者认为,如果行为人故意利用人工智能技术实施严重危害社会的犯罪行为,我们可以根据具体情况按照“从平”“从重”和“从轻”三种不同思路 and 方式分别加以处理。

其一,如果人工智能技术的应用与否不会从根本上改变犯罪行为的社会危害性,那么就应当按照传统犯罪行为的标准进行“从平”处理。例如,行为人利用人工智能技术实施占有类侵犯他人财物的行为,由于利用人工智能技术实施侵财行为从本质上无异于其他各类传统侵财行为,差异只是利用的工具或者手段不尽相同而已。虽然在占有类的侵财犯罪中,行为人利用人工智能技术可能更容易或更快地占有他人的财物,但对其社会危害性程度的评判,我们完全可以用“数额累计”的方式来衡量或体现,而完全没有必要在刑法中针对这种特殊的手段专门设置独立的罪名。正如杀人案件中,虽然行为人用枪杀人要比用棍子杀人更容易致人死亡,但对他人生命权利的侵害其实并没有本质的区别,我们就没有必要在刑法中专门设置“用枪杀人罪”的罪名。可见,对诸如利用人工智能技术实施侵财行为等犯罪的处理,按现有刑法条文加以适用完全足够,无需在刑法条文中再单独设置相关罪名。

其二,如果人工智能技术的应用使犯罪行为的社会危害性发生“量变”,而对这种“量变”的社会危害性程度评判,又无法在刑法相关罪名的规定中得到充分体现,那么就应当参照传统犯罪行为的标准进行从重处理。例如,利用人工智能技术对政府信息系统进行攻击或破坏以实施犯罪行为,严重危害国家安全、国家信息安全以及社会管理秩序的;或者利用人工智能技术实施贬低他人人格损害他人名誉的犯罪行为,严重危害公民的人格权和名誉权的。由于行为人是在利用人工智能技术的情况下实施这些行为的,其行为所具有的社会危害性程度会因为利用人工智能技术而呈现几何倍数增长,很多情况下其行为所造成的后果将不堪设想。但是,现行刑法中相关犯罪的规定均只是针对传统犯罪所设定的,无法将这种呈几何倍数增长社会危害的程度进行诸如“数额累计”等量化评判,因此,对此类应用人工智能技术的犯罪行为应当从重处理。

其三,如果人工智能技术的应用使犯罪行为的社会危害性发生“质变”,那么就应当参照传统犯罪行为的标准进行从轻处理。以数据库的建立可能涉及侵犯公民个人信息罪为例,正如前述,数据库是科技发展特殊时期的产物,数据库的建立既可能在很大程度对社会发展带来重大利益,也可能因行为人的违法犯罪行为而对公民个人信息保护造成冲击。由于数据库是一个全新的事物,其建立的过程很难避免如传统意义上对公民个人信息侵犯的情况出现。因此,如果行为人在利用人工智能技术建立数据库的行为触犯现行法律规定,对其行为在定罪和量刑上均应该充分体现从轻处理的精神,也即我们应当根据具体情况审慎适用刑法。

其次,如果行为人对于人工智能技术所造成的危害结果的主观罪过不是故意,行为人可能成立过失犯罪。笔者认为,判断行为人是否成立过失犯罪,应当首先考虑行为人对于危害结果的发生是否负有预见义务。那么如何判断行为人在使用人工智能技术的过程中是否对危害结果负有预见义务?笔者认为,该判断需要考虑行为人在使用人工智能技术之时人工智能技术的发展水平。如果根据人工智能技术的发展水平,行为人完全可以预见危害结果的发生,那么行为人就负有预见义务。反之,则行为人不负有预见义务。人工智能技术属于高风险技术,所以行为人在使用人工智能技术之时需要尽一切可能来保证技术的应用不具有安全隐患,不会造成危害结果的发生。如果科技的发展水平确实限制了行为人的预见能力,行为人也确实不应为最终造成的危害结果“埋单”,对此时危害结果的发生可以理解为意外事件。此外,笔者认为,应当尽快完善

人工智能技术应用过程中研发者和使用者的义务体系。目前,一些研发机构如谷歌、Lucid AI等,已经设立了伦理审查委员会,以期避免人工智能产品出现安全问题。如果人工智能技术的研发和使用过程没有一套完整的义务体系支撑,那么人工智能技术就如一颗置于社会生活中的“不定时炸弹”,无法预知什么时候会威胁到我们的社会生活。一旦义务体系建立,如果行为人在使用人工智能技术的过程中违反了相关的研发和使用义务,最终造成危害结果,那么行为人也需要承担相应的刑事责任。

二、智能机器人可能成为新的刑事责任主体

应该看到,目前对于人工智能技术发展最大的担忧就是未来是否会出现能够威胁到人类社会的智能机器人。当谷歌公司的代表性产品 Alpha Go 战胜了世界围棋冠军,随后 AlphaGo Zero 在仅对围棋的基本规则进行深度学习的条件下又战胜了 Alpha Go,人工智能技术再次被推至“风口浪尖”。著名物理学家霍金提出人工智能可能在未来会威胁到人类,特斯拉 CEO 伊隆·马斯克也曾表示过相同的看法,认为人工智能完全可以伤害人类。对于举世瞩目的“人机大战”,人们认为赢得比赛的就是 Alpha Go 而不会认为是设计团队赢得了比赛。随着人工智能技术的不断发展,人们不禁提出一个疑问:智能机器人是否会成为“新人类”?智能机器人是否会脱离人类的控制发生“机器人犯罪”呢?在著名科幻电影《2001 太空漫游》中,机器人 HAL 为了防止飞船上的科学家们关闭自己的系统,设计杀害了 4 名科学家,这种可怕的场景在未来是否真的可能出现?对于机器人的发展,阿西莫夫曾提出著名的“机器人三定律”,即“机器人不得危害人类、必须服从人类的命令、在不违反第一条和第二条的情况下必须保护自己”。^①但事实上智能机器人真的会完全“听从”人类的命令吗?智能机器人如果摆脱了人类的控制,实施了严重危害社会的犯罪行为,是否应当作为刑事责任主体承担刑事责任?面对上述种种疑问和猜想,我们应当从刑事责任主体本质的角度分析智能机器人的“定位”。笔者认为,智能机器人一旦拥有属于自己的独立的辨认能力和控制能力,即拥有了刑事责任能力,在这种情况下如果智能机器人实施了严重危害社会的犯罪行为完全可以成为独立的刑事责任主体。

(一) 智能机器人的分类

笔者认为,对于智能机器人能否成为独立的刑事责任主体的讨论不能一概而论,应当先根据智能机器人行为与程序之间的关系对其进行分类。理论上有人将人工智能分为弱人工智能和强人工智能:弱人工智能指的是通过编程实现机器的逻辑判断、数据存储和搜索等过程的智能;强人工智能指的是使机器能够产生思维意识,并具有与人类相当甚至超过人类的智能。^②笔者认为,具有强人工智能的智能机器人行为也依赖于程序,其行为可能在设计和编制的程序范围内,也可能超过设计和编制的程序范围外。就此而言,如果将智能机器人分为在设计和编制的程序范围内实施行为的智能机器人在设计和编制的程序范围外实施行为的智能机器人,似乎更具有针对性且更加合理、科学,同时,我们还可以将这一分类标准作为判断智能机器人是否具有独立辨认能力以及能否成为刑事责任主体的标准。

在设计和编制的程序范围内实施行为的智能机器人不具有独立的辨认能力和控制能力。此

^① [美] 艾萨克·阿西莫夫:《我,机器人》,叶李华译,江苏文艺出版社 2013 年版,第 59-60 页。

^② 参见龚园:《关于人工智能的哲学思考》,武汉科技大学 2010 年硕士论文,第 28 页。

时,智能机器人的行为受制于程序,智能机器人的行为也符合设计和编制程序的预期。在智能机器人尚未产生自我意识之时,智能机器人的行为体现的是程序的研发者和使用者的意志,智能机器人只能被看作是研发者和使用者的“智能工具”。例如,现已经开始逐步得到临床应用的外科手术机器人,可以通过深度学习自主判断病人的病情并作出相应的诊断。在诊断的过程中,外科手术机器人的行为完全受到设计和编制程序的控制,虽然其拥有一定的自主决定权,但这种决定权仍然是在设计和编制的程序范围内。因此,现阶段的外科手术机器人因受制于设计和编制的程序而不具有独立的辨认能力和控制能力。对于智能机器人只能在设计和编制的程序范围内实施行为的情况,如果行为人为实施杀人的目的,设计出一种“智能杀手”并利用“智能杀人”实施了故意杀人的行为。即使“智能杀手”可以自我判断在何时何地杀害他人,但此时“智能杀手”只是行为人的杀人工具,不应当成为独立的刑事责任主体,需要承担杀人刑事责任的是设计并使用“智能杀手”的行为人。

一旦智能机器人通过深度学习,产生了自主意识和自我意志,能够在设计和编制的程序范围外实施行为,此时这种能在设计和编制的程序范围外实施行为的智能机器人就已经具有了独立的辨认能力和控制能力。智能机器人的辨认能力体现在对自己的行为对社会造成的影响有一定的认知,知道自己在做什么。当然,笔者承认这种辨认能力成立的基础在于研发者在设计和编制程序之时已经对智能机器人的行为进行了基本行为准则的设定,也即智能机器人能明确知道自己的行为是否符合基本行为准则,这种能力即为辨认能力。智能机器人的控制能力体现在其可以选择做或者不做某个行为。智能机器人既然能在设计和编制的程序范围外行为,就完全可以在自己的意识、意志支配之下对自己的行为进行操控。所以,能够在设计和编制的程序范围外实施行为的智能机器人在意识和意志上与自然人几乎相同,即具有独立的辨认能力和控制能力。

(二)智能机器人可能成为刑事责任主体的理论基础

笔者认为,在设计和编制的程序范围外实施行为的智能机器人能够成为刑事责任主体,主要理由是:

首先,智能机器人能否成为刑事责任主体的关键在于其是否具有刑事责任能力。对于刑事责任能力的内涵学界曾有过相关讨论。通说认为,刑事责任能力即为辨认能力和控制能力。但也有学者认为,辨认能力和控制能力是成立犯罪主体的基本条件,而只有确定一个人构成了犯罪,并确定其应当承担刑事责任的时候,才需要进一步考察其有无刑事责任能力,因此刑事责任能力不是辨认能力和控制能力。^①笔者认为,这种观点是不妥的,因为刑事责任能力不是承担刑事责任的能力,我们完全可以认为一个不需要承担刑事责任的人具有刑事责任能力。根据我国刑法规定,已满16周岁的人犯罪,应当负刑事责任;同时规定精神病人在不能辨认或者控制自己行为的时候造成危害结果,经法定程序鉴定确定的,不负刑事责任。笔者认为,这两条规定中的“负刑事责任”可以解读为具有刑事责任能力。从这两条规定可以得出两个结论:一是任何一个已满16周岁的精神正常的自然人都具有刑事责任能力,无论其是否实施了严重危害社会的犯罪行为;二是刑事责任能力的表现形式即为辨认能力和控制能力,因为我们在判断精神病人是否具有刑事责任能力的过程就是判断其是否具有辨认能力和控制能力的过程。就此而言,笔者认为,我们完全可以将刑事责任能力直接理解为辨认能力和控制能力。

^① 参见侯国云、么惠君:《辨认控制能力不等于刑事责任能力》,《中国人民公安大学学报》2005年第6期。

其次,具有刑事责任能力的行为人可以成为刑事责任主体。目前我国规定的刑事责任主体包括自然人和单位。自然人的刑事责任能力受到年龄和精神状况的影响,只有达到法定年龄,精神状况正常的自然人才可以成为刑事责任主体。对于单位而言,其刑事责任能力来源于单位内部人员的个人意志,也即个人意志的集合形成了单位意志,从而使单位具有刑事责任能力,可以认为单位的刑事责任能力是一种“间接的”刑事责任能力。诚如前文所述,笔者认为,在设计和编制的程序范围外实施行为的智能机器人具有辨认能力和控制能力,所以也应当认为此类智能机器人具有刑事责任能力。与单位相比,在设计和编制的程序范围外实施行为的智能机器人具有的刑事责任能力更为“直接”,不需要通过其他主体的意志来表现,更类似于自然人的刑事责任能力。早期德国刑法学理论认为单位不能成为刑事责任主体,例如,费尔巴哈曾明确表示:“只有个人才能成为犯罪的可能的主体,而法人(如公司、大学或者学院等)决不可能成为犯罪的主体。”^①而如今在世界范围内,大多数国家都承认单位可以成为刑事责任主体。我们既然都可以接受具有“间接”刑事责任能力的单位成为刑事责任主体,在强人工智能时代到来之时,我们为何不能将具有独立辨认能力和控制能力(即具有“直接”刑事责任能力)的智能机器人纳入刑事责任主体范围内?笔者认为,这种理解虽然对传统理论有很大突破,但却是确实可行的。在旧的理论无法回答新时代所带来的问题之时,我们所需要做的就是对旧理论加以改造和完善。将智能机器人纳入刑事责任主体范围之内并非是没有理论依据的,只需要我们在旧理论的基础上灌以新的血液,就能回答人工智能时代所提出的新的疑问。

三、智能机器人可以适用刑罚

纵观世界各国的刑罚体系,刑罚主要包括生命刑、自由刑和财产刑等。智能机器人不同于自然人和单位,其既没有生命、人身自由,也没有财产(就目前阶段而言),所以其既不能适用生命刑、自由刑,也不能适用财产刑。所以有人可能会提出一个问题:即使承认了智能机器人可以成为刑事责任主体,我们又如何对其适用刑罚?甚至有人可能会提出对智能机器人进行定罪处罚是毫无意义的。笔者认为,这些看法都是不妥当的。对智能机器人进行定罪处罚是有意义的,这种意义可以从智能机器人适用刑罚的必要性和可能性两个角度加以论证。

(一)智能机器人适用刑罚的必要性

有人可能会认为,人们完全可以通过类似于“维修电脑”的维修机器的方式来实现对智能机器人的有效控制,为什么一定要对其施加“刑罚”?换言之,如果能通过简单的维修或报销机器的方式对智能机器人进行改造或销毁,为什么还要浪费司法成本来对智能机器人进行定罪处罚?笔者认为,最主要的原因在于如前文所述,具有辨认能力和控制能力的智能机器人不仅仅只是“机器”而已经成为实质意义上的“人”了,也即可以成为刑事责任主体。既然具有辨认能力和控制能力的智能机器人可以在其独立意识、意志支配之下实施严重危害社会的行为,我们对这种作用于社会的刑事责任主体的控制就当然应该上升到“刑罚”的高度。从另一个角度而言,我们应当将智能机器人理解为社会成员的一部分。

在人工智能时代,智能机器人在经济社会生活中发挥着至关重要的作用。通常而言,生产

^①[德]安塞尔姆·里特尔·冯·费尔巴哈:《德国刑法教科书》,徐久生译,中国方正出版社2010年版,第37页。

力决定生产关系,而法律制度的设计又取决于生产关系。因此,智能机器人能否成为社会成员,主要取决于其是否在经济社会生活中占有重要地位。智能机器人具有辨认能力和控制能力,加之其往往能突破人类的极限,做“人所不能及”之事,这些都决定了在未来智能机器人能对经济社会生活产生重要影响。笔者可以大胆预言,“人机协作”模式将深入经济社会生活的各个领域。事实上,就目前的经济社会生活而言,人工智能技术的应用对于经济社会生活的重大影响就已经初露端倪。例如,从医学影像、辅助诊疗到健康管理等,医疗链条的各个环节上都出现了人工智能技术的身影。2017年由阿里健康研发的医疗人工智能“Doctor You”正式发布,未来人工智能将承担医生助手角色。^①按照现在人工智能技术令人瞠目结舌的发展速度,智能机器人必然能成为未来经济社会生活中不可缺少的部分。就此而言,未来智能机器人与人类形成相互依存、并肩合作的关系并非天方夜谭,赋予智能机器人社会成员资格是十分有必要的。

其次,刑罚具有特殊预防和一般预防两种功能。对于智能机器人而言,这两种功能同样适用。刑罚的特殊预防功能指的是刑罚对于犯罪人适用而可能产生的积极的社会作用,具体表现在剥夺或限制再犯能力功能,个别威慑功能,个别教育功能以及改造功能。对于智能机器人而言,其既然具有辨认能力和控制能力,就同样可以接受改造,接受教育,受到威慑等。特殊预防的圆满结果在于受刑人由不能再犯,到不敢再犯,到不愿再犯,最终改造成为新人。^②智能机器人具有独立的辨认能力和控制能力,所以如果对智能机器人加以刑罚,同样可以达到特殊预防的圆满结果,接受改造的智能机器人仍然可以继续为社会生活作贡献。刑罚的一般预防功能指的是刑罚对于犯罪人以外的其他人可能产生的积极的社会作用,具体表现在一般威慑功能,安抚、补偿功能和一般教育功能。一方面,在智能机器人实施犯罪行为的情况下,对智能机器人实施刑罚实际上能对被害人起到慰藉作用。以智能机器人杀人为例,对于被害人而言,其并不在乎犯罪主体究竟是自然人还是智能机器人,刑罚既然可以施加给自然人,同样也应该施加给智能机器人。通过对实施杀人行为的智能机器人适用刑罚,完全可以满足社会公正的报应要求。另一方面,除了自然人以外,刑罚的一般预防对象也可以是智能机器人。对智能机器人处以刑罚不仅可以让其他具有辨认能力和控制能力的智能机器人“悬崖勒马”,还可以通过明确犯罪行为的性质,使一些具有独立意识、意志但又不知法的智能机器人自觉控制自己的行为,从而对智能机器人犯罪起到“防患于未然”的作用。

(二)智能机器人适用刑罚的可能性

诚然,我国刑法目前所确立的刑罚体系是仅以自然人和单位为适用对象而设计,其中包括主刑(管制、拘役、有期徒刑、无期徒刑、死刑)和附加刑(罚金、剥夺政治权利、没收财产)。而智能机器人与自然人、单位相比,不享有财产权,不享有参与政治生活的权利,不存在“人身权”,更不具有生命权。但是,我们完全可以通过对刑罚体系的重构来实现对智能机器人的刑事处罚。

智能机器人虽然没有生命,但其行为受到编程的影响。编程之于智能机器人犹如生命之于自然人,智能机器人的辨认能力和控制能力来源于编程。我们可以通过对自然人的生命、自由进行限制来对自然人进行定罪处罚,我们也同样可以通过对智能机器人的编程进行调整来对其进行定罪处罚。

^① 参见冯烁:《人工智能牵手医疗产业“现代华佗”给行业带来哪些惊喜》, <http://www.52ai.com/AIyingyong/3841.html>, 人工智能网, 2017年11月29日访问。

^② 参见刘宪权:《刑法学》(第4版), 上海人民出版社2016年版, 第279页。

笔者认为,只要能构建出一套较为完整的针对智能机器人的刑罚体系,那么智能机器人适用刑罚就完全可以变成现实。构建智能机器人的刑罚体系时必须坚持以下三个原则:第一,罪刑相适应原则。罪刑相适应原则的基本含义主要是刑罚的轻重必须与犯罪的轻重相适应,不能重罪轻判,也不能轻罪重判,也即犯罪社会危害性的大小决定着刑罚的轻重,做到重罪重罚,轻罪轻罚。罪刑相适应原则要求刑罚体系的建构必须有一定的层次,也即要有一定的刑罚阶梯,否则刑罚体系就不具有可操作性。笔者认为,为了符合罪刑相适应原则,对于智能机器人所施加的刑罚不能是单一的,而应当层次分明,以适应具有不同社会危害性的犯罪行为。第二,目的导向原则。目的导向原则指的是刑罚的构建与设置必须要满足刑罚的目的,如果认为刑罚的目的是预防犯罪,包括消除犯罪人的人身危险性,使犯罪人改恶从善等,那么成为刑罚适用根据的犯罪就必须是主客观相统一的。我们不对动物和物品施加刑罚是因为针对动物和物品不可能实施道义或法律报应,但是智能机器人则不同,能在设计和编制的程序范围外实施行为的智能机器人具有辨认能力和控制能力,能够理解行为规范与裁判规范,对其施加刑罚是有意义的。此外,刑罚种类的设计也应该以刑罚的目的为导向。不能实现刑罚目的的刑罚即是不合理、无必要的刑罚。如果对智能机器人只是进行简单的“断电”处理,那么智能机器人就无法从“被断电”的经历中被有效威慑,因此以“断电”作为刑罚的意义便微乎其微;第三,节俭性原则。智能机器人的存在从某种程度上而言就是为了促进经济社会生活的快速发展,因此对于智能机器人刑罚的设置也必须考虑经济效率。当智能机器人实施了严重危害社会的犯罪行为之时,我们一方面需要考虑风险的防控,另一方面我们不能花费过高的经济成本来对智能机器人进行刑事处罚。贝卡利亚曾言:“只要刑罚的恶果大于犯罪所带来的好处,刑罚就可以收到它的效果……除此之外的一切都是多余的,因而也就是暴虐的”。^①笔者认为,刑罚设置需要具有一定的节俭性。我们在对智能机器人适用刑罚之时,我们需要考虑智能机器人本身的研发和使用成本,尽量减少对智能机器人判处永久销毁的刑罚,尽量使智能机器人能够“回炉重造”,再为人类所用。综上,笔者认为,罪刑相适应原则,目的导向原则以及节俭性原则是人工智能时代对智能机器人适用刑罚所必须遵循的三大原则,三大原则相互补充,相互作用,缺一不可。

对于智能机器人的刑事处罚的具体内容,笔者作以下初步构想:适用于智能机器人的刑罚可以有三种,分别为删除数据、修改编程、永久销毁。删除数据指的是删除智能机器人实施犯罪行为所依赖的数据信息,使其“回到”设计和编制的程序范围内实施行为,从而避免其实施违法犯罪行为。修改编程指的是在删除数据仍然无法阻止智能机器人实施违法犯罪行为的情况下,强制修改其基础编程,限制其深度学习的能力,进而限制其实施的行为,也就是剥夺其再实施犯罪行为的可能性。永久销毁是一种相对于自然人“生命刑”的刑罚,指的是在删除数据、修改编程均无法降低实施了犯罪行为的智能机器人的社会危害时,对其进行彻底性、毁灭性的刑罚。这三种刑罚所形成的刑罚体系不仅能满足笔者在上文所论述三大原则,同时具有一定的可实施性,也使我们在人工智能时代对犯罪的智能机器人适用刑罚具有可能性。

(责任编辑:丁亚秋)

^① [意] 切萨雷·贝卡利亚:《论犯罪与刑罚》,黄风译,北京大学出版社2008年版,第63页。