

人工智能： 刑法的时代挑战

刘宪权 主编


未来能来，人外有人
智者见智，创所未创

探讨人工智能时代的刑法风险与应对

人工智能： 刑法的时代挑战

刘宪权 主编



 上海人民出版社

图书在版编目(CIP)数据

人工智能:刑法的时代挑战/刘宪权主编.—上

海:上海人民出版社,2018

ISBN 978-7-208-15335-6

I. ①人… II. ①刘… III. ①人工智能-刑法-研究

IV. ①D914.04

中国版本图书馆 CIP 数据核字(2018)第 158449 号

责任编辑 秦 莹 夏红梅

封面设计 一本好本

人工智能:刑法的时代挑战

刘宪权 主编

出 版 上海人民出版社
(200001 上海福建中路 193 号)
发 行 上海人民出版社发行中心
印 刷 上海商务联西印刷有限公司
开 本 720×1000 1/16
印 张 20.75
插 页 2
字 数 354,000
版 次 2018 年 7 月第 1 版
印 次 2018 年 7 月第 1 次印刷
ISBN 978-7-208-15335-6/D·3253
定 价 68.00 元

目录

CONTENTS

第一编 未来能来:人工智能时代的刑事风险

- 003 人工智能时代的“内忧”与“外患”
- 017 人工智能时代的刑事风险与刑法应对
- 033 人工智能时代的刑事风险与刑事责任
- 038 人工智能时代的刑事风险和责任分配
- 050 智能汽车的刑事风险与刑法应对
- 062 无人驾驶汽车刑事风险的刑法应对
- 071 人工智能时代:无人驾驶汽车应用的刑事风险与责任

第二编 智者见智:人工智能时代的法律新题

- 083 人工智能时代机器人行为道德伦理与刑法规制
- 103 涉人工智能犯罪刑法规制正当性和适当性的路径
- 124 人工智能发展的类型化与刑事责任
- 140 人工智能致损的法律责任
- 154 强弱人工智能产品刑事责任认定规则
- 170 人工智能犯罪中的刑事归责原则
- 180 刑法对人工智能应秉持的理性态度
- 194 智能机器人的法律定位及其权利审视
- 206 人工智能在人身危险性测评中的功能



第三编 人外有人:人工智能时代的责任主体

- 219 人工智能时代智能机器人的刑事责任能力
- 231 人工智能时代刑事责任主体的重新解构
- 240 智能机器人的法律主体地位及刑事责任
- 251 人工智能体的刑事主体资格
- 260 对认知人工智能体刑事处罚的依据与意义
- 271 人工智能时代智能机器人的刑事责任与刑罚理念反思
- 281 人工智能刑事责任主体的差异性

第四编 创所未创:人工智能时代的刑罚体系

- 291 人工智能时代我国刑罚体系重构的法理基础
- 307 人工智能时代刑事责任与刑罚体系的重构
- 319 人工智能时代法律风险及刑法应对



第一编 未来能来：人工智能时代的刑事风险

人工智能技术的发展速度超乎人们的想象，已呈现出深度学习、跨界融合、人机协同、群智开放、自主操控等新特征，几乎赋予人工智能以无限可能。人工智能“无限可能”背后的深层隐喻是风险的“有限可控”，技术风险及其对社会的影响处于一种不确定的状态。2017年7月8日国务院《新一代人工智能发展规划》对人工智能发展之不确定性所可能带来的新挑战作了预估：“人工智能是影响面广的颠覆性技术，可能带来改变就业结构、冲击法律与社会伦理、侵犯个人隐私、挑战国际关系准则等问题，将对政府管理、经济安全和社会稳定乃至全球治理产生深远影响。在大力发展人工智能的同时，必须高度重视可能带来的安全风险挑战。”就国家层面而言，以各种可能的手段降低、控制风险，成为我国发展人工智能的重要战略任务。

“风险”在《辞海》中被释义为“人们在生产建设和日常生活中遭遇能导致人身伤亡、财产受损及其他经济损失的自然灾害、意外事件和其他不测事件的可能性”。“风险”与刑法中的“实害”相区别，与刑法中的“危险”相联系。“实害”是已然发生的侵害，“风险”是发生侵害的可能性，二者属已然与未然的关系；“危险”指发生侵害的高度可能性，“风险”指发生侵害的一般可能性，二者属高阶与低阶的关系。由此，当我们谈论“刑事风险”概念时，其并不指向表征着社会危害性的“实害”与“危险”，而可能与“犯罪风险”同义，即刑事风险是发生刑事犯罪的可能性。

刑事风险意味着刑事犯罪可能发生，也可能不发生，刑法所作出的应对也就相应地表现为两个层次：第一，当犯罪发生时，解决刑事责任分配问题；第二，当犯罪可能发生而尚未发生时，通过立法举措威慑并预防犯罪。由此决定了“刑事风险”中的“刑事犯罪”概念可能具有两种含义，从兼具司法与立法的更广义的角度看，它不仅指实然法层面已经被一国刑法规定为犯罪的行为，还包括应然法层面根据时代与情势的需要应当被一国刑法规定但尚未被规定为犯罪的行为。

本编“未来能来：人工智能时代的刑事风险”共收录7篇论文，它们对“刑事风险”的描述与论证基本没有超出前述的两种语境。一方面，人工智能时代衍

生出富有该时代色彩的犯罪样态,如利用人工智能实施的犯罪,人们可以在现行刑法框架下寻求责任分配的解决方案;另一方面,人工智能时代可能催生出以往任何时代不曾具有却为该时代所独有的“严重危害社会的行为”,其在现行刑法框架下尚不占据一席之地,对该行为的规制可能要以前瞻性地修补乃至重构刑法体系为前提。

《人工智能时代的“内忧”与“外患”》一文对人工智能时代的潜在刑事风险作了“内外有别”“强弱有度”的归纳与总结,即分别表现为弱人工智能产品被外部主体利用所致的“外患”与强人工智能产品内部自主自发行为所致的“内忧”。强、弱人工智能产品的根本分野在于该人工智能产品是否具有独立的意志,也即是否具有人格。该文对“人格”的探讨不受限于“人身”,澄清人身不是人格的必要条件,人们过去将人视为人格人,而今应当将人格人视为人。人工智能产品人格之有无,造成了刑事风险来源的内外之分。《人工智能时代的刑事风险与刑法应对》堪称前文姊妹篇,其在刑事风险来源存有内外之别的理论基础上,进一步描摹刑事风险在不同来源下的具体表现形式,如可能使人类部分传统犯罪的危害量变升级或使人类的犯罪形式推陈出新,也可能使人类之外的人工智能产品脱离人类控制,在设计和编制的程序范围外实施严重危害社会的行为。这两篇论文均谈及刑法应对及刑事责任分配问题,但前文侧重于阐明“内忧”与“外患”这两种不同的风险来源在刑事责任分配上的联系,即在“内忧”情形下也并不排斥向“外”归责,如在强人工智能产品独立承担刑事责任的场合,若其设计者和使用者未履行相应义务,也可能承担刑事责任,且“内忧”与“外患”可能交织共生,自然人、单位可以与强人工智能产品形成共同犯罪;后文则侧重于论证“内忧”与“外患”所对应的刑事责任体系上的差异,处理“外患”所对应的基本上是现行的刑事责任体系,增设滥用人工智能罪、人工智能事故罪不过是对该体系的修补与延展,不足以动摇该体系的根基,而预防“内忧”所对应的是承认强人工智能产品刑法主体地位的全新的刑事责任体系,这当属根本性的突破。《人工智能时代的刑事风险与刑事责任》一文既是对前两篇论文核心观点的摘要与凝练,也是在具备一定理论准备之后的修法倡议,其建议在我国《刑法》第二章第四节单位犯罪之后增设一节“智能机器人犯罪”,可以预期这将成为刑法增量理论研究中的新一轮热点。在本编中,龙敏、汤君、钟菁、张金钢这四位青年学者也围绕人工智能时代的刑事风险问题各抒己见。他们或关注人工智能时代的宏观刑事风险,或着眼于无人驾驶汽车应用的微观刑事风险,从微观窥见宏观,以宏观统领微观。在视角的不断变换、观点的激烈碰撞之中,我们会身临其境地感受到,人工智能时代的未来,能来。



人工智能时代的 “内忧”与“外患”

刘宪权*

【内容摘要】 以人工智能产品是否具有独立意志,以及是否能够在设计和编制的程序范围外自主决策并实施相应行为为依据,可以将人工智能产品划分为弱人工智能产品与强人工智能产品。弱人工智能产品不具有独立意志,仅能在设计和编制的程序范围内实施行为,实现人类设计和编制程序的目的。弱人工智能时代的“外患”在于,人类完全有可能让弱人工智能产品为己所用,实现自己的犯罪目的。弱人工智能产品作为犯罪工具,当然无刑事责任的承担可言。强人工智能时代的“内忧”在于,强人工智能产品有能力在设计和编制的程序范围外,依据自己的独立意志实施危害社会的行为。强人工智能产品具有独立人格和刑事责任能力,可以将强人工智能产品实施的严重危害社会的行为认定为犯罪行为并给予刑罚处罚。强人工智能产品的设计者和使用者未履行预见义务和监督义务,可能承担相应的刑事责任。强人工智能产品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。

【关键词】 弱人工智能 强人工智能 独立意志 辨认控制能力 刑事责任

人工智能(Artificial Intelligence, AI)是指使机器像人一样去完成某项任务的软硬件技术。^①人工智能虽不是人类的智能,但其可以模拟人的意识和思维过程,像人一样进行思考。人工智能是网络技术和机器人技术发展到新阶段的体现。随着科技水平的不断提高,人工智能的发展日新月异,从1.0时代到2.0时代,再到3.0时代,科学家分别创造出了对外界环境无感知的人工智能产品,对外界环境有感知的人工智能产品和对外界环境有认知、推理和决策能力的人工

* 刘宪权,华东政法大学刑法学教授、博士生导师。

① Andrew Arruda, “An Ethical Obligation to Use Artificial Intelligence? An Examination of the Use of Artificial Intelligence in Law and the Model Rules of Professional Responsibility” 40 American Journal of Trial Advocacy 443(2016).

智能产品。人类科技发展的历史告诉我们，科技的发展是“爆炸式”的，其发展速度呈几何式提升，我们必将在不久的将来进入人工智能 N.0 时代。各国都意识到人工智能的影响电照风行，纷纷出台政策，推动人工智能技术的发展和革新。正如俄罗斯总统普京所言：“无论对于俄罗斯还是全人类而言，人工智能都是未来。人工智能蕴含着巨量的机遇，同时也隐藏着难以预料的威胁。谁在人工智能领域能够成为领导者，就能够成为世界的规则制定者。”2017 年国务院《新一代人工智能发展规划》将发展人工智能技术提升到战略高度。《新一代人工智能发展规划》指出：“人工智能是引领未来的战略性技术……当前，我国国家安全和国际竞争形势更加复杂，必须放眼全球，把人工智能发展放在国家战略层面系统布局、主动谋划，牢牢把握人工智能发展新阶段国际竞争的战略主动，打造竞争新优势、开拓发展新空间，有效保障国家安全。”

随着人工智能时代的到来，这个本来属于纯粹技术问题的新生事物将会在法律、哲学、伦理等方面对人类社会带来极大的影响和冲击，并可能产生相应的刑事犯罪。对人工智能产品的合理应用，将会带来许多行业的重大变革，并在较大程度上通过解放劳动力，促进社会的深度和快速发展。但是如果设计者或使用者对人工智能产品进行不当利用，甚至将其作为实现犯罪意图的工具（此为我们在人工智能时代面临的“外患”），抑或是人工智能产品本身在具有独立意志之后产生了犯罪意图，进而实施严重危害社会的行为（此为我们在人工智能时代面临的“内忧”），将会给人类社会带来极大的威胁，甚至会导致人类社会的毁灭。“圣人不治已病治未病，不治已乱治未乱，此之谓也。夫病已成而后药之，乱已成而后治之，譬犹渴而穿井，斗而铸锥，不亦晚乎。”^①面对人工智能技术发展可能带来的负面影响，我们的刑法学者不应该无动于衷，我们的刑事立法和刑事司法更不应该面临危险而束手无策。“宜未雨而绸缪，毋临渴而掘井”，我们应提前预想到在人工智能时代可能面临的“内忧外患”，明确不同类型的人工智能产品的法律特征和法律地位，并就我国刑法应当如何合理应对人工智能时代所带来的刑事风险展开具体的分析与讨论，提前明晰人工智能时代的刑事责任承担路径。只有做好充分准备，方可避免在危险来临时“患至呼天”。

一、强弱人工智能产品之界分

按照人工智能产品的不同形态，可以将其划分为人形人工智能产品与非人

^① 《黄帝内经·素问·四气调神大论》。

形人工智能产品,其中,非人形人工智能产品又包括意识形态人工智能产品(如AlphaGo,其只是隐藏在电脑里的一个围棋程序)与实体人工智能产品(如扫地机器人)。物理形体并非人工智能产品的必备要素。而无论是否具有物理形体,人工智能产品都必须依托算法(如深度学习和神经网络等),实时收集海量数据信息,并根据算法的计算结果实施行为。由此可见,人工智能产品的本质要素是算法。但是鉴于算法的复杂性,将算法作为区分强人工智能产品和弱人工智能产品的依据,在法律上既无可能,也无必要。其一,“根据算法的复杂性和学习或运算能力对强人工智能产品和弱人工智能产品进行区分,这在技术认知上没有问题,但在法律上很难按照智能程度给出精确的标准。法律应对复杂世界的方式是确立一般性的简单规则,在概念上对社会个体进行抽象假定(如行为能力),而非针对特殊主体,否则规则体系本身将变得异常复杂,难于理解和操作。而人工智能从单一的自动化服务向多元通用服务转变的过程可能是一个相当长的光谱,法律需要针对其本质特征进行约束,并探索一套应对未来的方案”。^①其二,正如人类的大脑是一个复杂而精致的系统,尽管当代神经科学不断深化我们对人脑的认知,我们仍不能完全了解大脑的运转机制。但是,应当看到,没有完全理解人脑的运转机制,并未妨碍我们通过法律建立一整套的规则体系来控制人类的行为,也并未妨碍我们运用刑法对行为人的主观心态(即故意和过失)进行认定。尽管我们对人工智能的本质要素——算法未能完全了解,但是,只要在法律上对强人工智能产品和弱人工智能产品的划分确定一个明确的标准,就不会影响我们对我国刑法应当如何合理应对人工智能时代所带来的刑事风险等问题展开具体的分析与讨论。

以人工智能产品是否具有独立意志,以及是否能够在设计和编制的程序范围外自主决策并实施相应行为为依据,可以将人工智能产品划分为弱人工智能产品与强人工智能产品。弱人工智能产品虽然可以在设计和编制的程序范围内进行独立判断并自主作出决策,但不具有独立意志,其实现的只是设计者或使用者的意志。例如,阿尔法狗(AlphaGo)虽然凭借自主判断和决策战胜了围棋国手,但其自主判断和决策仍是在设计和编制的程序范围内进行的,实现的是设计者或使用者的意志——在下围棋时战胜对手。强人工智能产品具有独立意志,既可能在设计和编制的程序范围内进行独立判断并自主作出决策,实现设计者或使用者的意志,也有可能超出设计和编制的程序范围,进行自主决

^① 胡凌:《人工智能的法律想象》, https://news.uc.cn/a_16384305516683655823/, 访问日期:2017年9月25日。



策并实施相应行为,实现其自身的意志。“强人工智能”一词最初是约翰·罗杰斯·希尔勒针对计算机和其他信息处理机器创造的。按约翰·罗杰斯·希尔勒的理解,“计算机不仅是用来研究人的思维的一种工具;相反,只要运行适当的程序,计算机本身就是有思维的”。^①正如1968年上映的美国经典科幻电影《2001太空漫游》所呈现的那样,人工智能产品HAL读懂飞行员想要关闭它的唇语,于是决定先发制人,故意让电脑失灵,导致4人丧生。电影里的HAL就是我们这里所讲的强人工智能产品,它可以超出设计和编制的程序范围自主作出让电脑失灵的决策,从而达到不让飞行员将自己关闭的目的。当然,这种情景只是人类的幻想,但是人类的科技发展史正是不断将幻想变成现实的历史。1783年法国J.F.P.罗齐埃和M.达尔朗德乘气球升空、1903年莱特兄弟飞机试飞成功,实现了庄子在《逍遥游》中所描述的“列子御风”的幻想;1969年“阿波罗号”飞船首次成功登上月球,实现了“嫦娥奔月”的幻想。虽然到目前为止,人工智能产品仍只能在设计和编制的程序范围内实施行为,实现设计者或使用者的意志,即我们目前仍处于弱人工智能时代,正如李开复在谈到人工智能产品时所言:“现在的机器没有丝毫自我认知,没有感情,没有喜怒哀乐,作了决定也讲不出为什么。”但是随着深度学习、神经网络、蒙特卡洛树搜索、云计算等技术的不断发展和完善,当技术突破下一个瓶颈时,出现具有独立意志、能够在设计和编制的程序范围外作出自主决策并实施相应行为、实现自身意志的人工智能产品,其实并非镜花水月、海市蜃楼。

弱人工智能产品不具有独立意志,仅能在设计和编制的程序范围内实施行为,实现人类设计和编制程序的目的。人类是弱人工智能产品的创造者,可以通过程序给弱人工智能产品设定行为目的和行为边界。此时,弱人工智能产品的设计者或使用者的完全有可能让弱人工智能产品为自己所用,实现自己的犯罪目的,如制造专门用来杀人的人工智能产品,实现自己非法剥夺他人生命的目的。同时,不法分子可能利用弱人工智能产品本身存在的漏洞,实现自己的犯罪目的。如诺基亚公司董事长李思拓所言,未来的家居设备很可能都是智能联网的,犯罪分子有可能通过洗碗机所连接的WIFI侵入家庭网络,从而操纵其家里的一切,实现自己的犯罪目的。在这种情况下,弱人工智能产品做出的严重危害社会的犯罪行为本质上是为了实现他人的犯罪意志,应当将该弱人工智能产品看作他人实施犯罪的“工具”。在弱人工智能产品自身之外出现的刑事法律风险,即我们在本文中探讨的弱人工智能时代的“外患”。

^① J.Searle in *Minds Brains and Programs*. *The Behavioral and Brain Sciences*, vol.3, 1980.

强人工智能产品依赖于自身具有的独立思维,有能力在设计和编制的程序范围外,依据自己的独立意志实施危害社会的行为。如上文中提到的强人工智能产品 HAL 超出设计和编制的程序范围自主作出让电脑失灵的决策,从而达到不让飞行员将自己关闭的目的。此时,强人工智能产品的行为未受到设计者或使用者等任何人为设置或操控,完全是自发实施的,且该种行为超越甚至根本违背了相关人员的意志。强人工智能产品实施的这种犯罪行为和其他具有刑事责任能力的主体实施的犯罪行为相去无几。在强人工智能产品自身之内出现的刑事法律风险,即我们在本文中探讨的强人工智能时代的“内忧”。

二、弱人工智能时代的“外患”

弱人工智能产品可以在设计和编制的程序范围内独立判断并自主作出决策,似乎具有一定的辨认能力与控制能力,但是这种能力完全受控或受制于人类设计和编制的程序,因此可以将弱人工智能产品看作人类改造世界的新型工具。人工智能是人类科技发展的重要里程碑,将会从根本上改变人类社会生活的面貌。但是如果被违法犯罪分子作为犯罪工具利用,将会带来难以想象的恶果。2017年8月,以特斯拉首席执行官伊隆·马斯克(Elon Musk)为代表的来自26个国家的人工智能技术领袖,向联合国进行“公车上书”,要求禁止“杀人机器人”的研发与使用,这体现了人类在弱人工智能时代,对人工智能产品所引发的“外患”的普遍担忧。那么在弱人工智能时代,在人工智能产品外部可能会出现哪些“外患”?即在弱人工智能时代,弱人工智能产品的设计者或者使用者可能会利用弱人工智能产品实施哪些犯罪行为?

弱人工智能产品是人类改造世界的新型工具,如同其他工具是人类双手的延伸一样,我们也可以将弱人工智能产品看作人手的延伸,其可以解放人类劳动力,并弥补人类自身技能的不足。相较于普通工具,弱人工智能产品的优势在于其可以在某一方面完全替代人类自身的行为,达到人类预期的目的。如果这一优势被犯罪分子所利用,其就会在付出更小代价的同时,带来更大的社会危害。但是应当看到,弱人工智能产品的工具属性决定了其在犯罪中的地位,弱人工智能产品不可能作为犯罪主体而承担刑事责任。

笔者认为,弱人工智能产品被他人作为工具加以利用实施严重危害社会的犯罪行为,可能体现在刑法规定的许多犯罪中。

利用弱人工智能产品可以实施危害国家安全、国防利益、公共安全和社会管理秩序的行为。随着弱人工智能时代的到来,一些危害国家安全、国防利益、

公共安全和社会管理秩序的行为可能不再需要人类亲力亲为,只需设计和编制出相应的程序,让弱人工智能产品分忧代劳。在自动枪械系统、无人机等弱人工智能产品出现的情况下,犯罪分子不必再冒着生命危险进行自杀式袭击就可以达到严重危害社会的目的。同时,弱人工智能产品的数据处理分析能力大大增强,能够更快速有效地帮助犯罪分子实施窃取、刺探情报的行为。上述因素都将对国家安全和国防利益等造成极大的威胁。同样,弱人工智能产品对公共安全的威胁也是现实存在的。如恐怖活动分子可以利用弱人工智能产品携带武器、弹药、爆炸物在人员密集的场所或公共区域实施爆炸袭击,进行恐怖活动。利用弱人工智能产品既有利于保证恐怖活动分子自身的安全,也更有利于其进行隐蔽,逃避抓捕。另外,弱人工智能产品被犯罪分子利用可能对社会管理秩序造成威胁。如基于弱人工智能产品强大的信息收集和数据处理能力,可以帮助犯罪分子更快速有效地非法获取国家秘密或者考试内容;又如基于弱人工智能产品在设计和编制的程序范围内可以独立完成相应任务的优势,可能被犯罪分子利用作为运输毒品、毁坏文物的工具。

当然,利用弱人工智能产品也可以实施危害经济、侵犯财产的行为。《刑法》第 287 条规定:“利用计算机实施金融诈骗、盗窃、贪污、挪用公款、窃取国家秘密或者其他犯罪的,依照本法有关规定定罪处罚。”由此可见,在弱人工智能产品只能作为工具的情况下,利用其实施犯罪活动,只能按照犯罪的目的行为进行定性,笔者在此不作赘述。值得一提的是,在金融领域,原先必须由高素质人才完成的大部分工作,已经被人工智能产品所取代。人工智能产品在速度和数据整合准确度上,已经超过金融分析师。现在纽约和伦敦证券交易所的交易大厅几乎形同虚设,真正的交易过程已经全面实现了“机器自动化”,^①犯罪分子一旦攻破程序的漏洞,控制了这种人工智能产品,就相当于获得了移山竭海的力量,只需一个简单的代码或指令,就会导致证券交易市场的崩溃,而犯罪分子将可以趁市场的混乱收获累累“硕果”。

利用弱人工智能产品还可以实施危害人身的行为。犯罪分子为了销毁犯罪证据、逃避法律制裁,往往采用借刀杀人、移花接木的手法。而借用弱人工智能产品来杀人就是最有效的方法。同样,弱人工智能产品的设计者、使用者也可以利用弱人工智能产品实施绑架、拐卖等剥夺或限制他人人身自由的行为。弱人工智能的发展和创新应用离不开数据的开放和共享。值得一提的是,腾讯

^① 《全球金融业进入“机器人时代”:你的血汗钱是否安全?》, <http://finance.eastmoney.com/news/1347,20160321606253387.html>, 访问日期:2017年9月27日。

借助 QQ、微信等产品建立了 10 亿级别覆盖度的基础库,对范围内的人群进行统计和分析,能够实现动态跟踪区域内人群流动、评估人流拥挤等级、捕捉开放社交网络的情况、分析目标客户群轨迹,让客户更精准地定位目标人群。^①这些信息一旦被过失或有意泄露,造成的危害后果将难以想象,不仅是对公民个人信息的侵犯,更会对公民的人身安全和财产安全造成巨大威胁。

应当看到,尽管弱人工智能产品威力巨大,在犯罪中发挥着重要作用,甚至在某些情况下可以完全替代设计者或使用者实施严重危害社会的行为,且其实施的行为涵摄到刑法所规定的方方面面的犯罪,但是弱人工智能因为不具有独立意志而始终无法摆脱工具属性,其实施的行为体现的是设计者或使用者的意志,只能被看作设计者或使用者行为的延伸。所以在弱人工智能时代,犯罪行为只能是在弱人工智能产品之外的设计者或使用者的意志支配之下而实施,我们将这种情形称为弱人工智能时代的“外患”。

三、强人工智能时代的“内忧”

强人工智能产品与弱人工智能产品的本质区别在于是否具有独立意志,以及是否能够在设计和编制的程序范围外自主决策并实施相应行为。强人工智能产品通过学习,产生脱离程序的独立意志,具备独立思考能力。笔者认为,强人工智能产品的行为可以分为两种:一种是在设计和编制的程序范围内的行为;另一种是超出设计和编制的程序,依靠独立意志,自主决策所实施的行为。对于前者而言,强人工智能产品与弱人工智能产品并无不同,此时的人工智能产品仅能作为工具,其行为本质上是人工智能产品的设计者或者使用者行为的延伸,实现的是设计者或使用者的意志,行为所产生的后果也应全部归责于设计者或使用者。对于后者而言,强人工智能产品与弱人工智能产品相比则有很大区别,此时的人工智能产品已经完全超出工具的范畴,因为其已具有独立意志,其行为不再是设计者或使用者行为的延伸,而是为了实现自己的意志、自主决策并自主实施的行为,行为所产生的后果也不能当然地归责于设计者或使用者。

正因为强人工智能产品具有独立意志,因此其实施的行为有可能不符合程序设计和编制的目的,甚至从根本上违背设计者设计和编制程序的目的。在此

^① 《人工智能时代云和大数据到底有多重要?》, http://blog.sina.com.cn/s/blog_be62e9a90102wg0i.html, 访问日期:2017年9月27日。

状态下，强人工智能产品完全可能实施人类无法控制的严重危害社会的行为。早在1940年，美国科幻作家阿西莫夫就预想到在未来，机器人有可能在具有独立意志之后，做出危害人类的事情，因此为了保护人类，其提出了“机器人三原则”，^①并因此获得了“机器人之父”的桂冠。这三个原则体现出人工智能的发展不应威胁到人类自身安全的愿望。当然，在人工智能时代，即使有了原则，原则也未必起多大作用。当脱离程序控制的强人工智能产品出现时，有可能真的会成为“我们人类文明面临的巨大威胁”，谁也不敢保证科幻电影中强人工智能产品HAL杀死飞行员的场景不会在现实中上演。“防范机器人向负面发展已经迫在眉睫，无需等到机器人的智能等于人的智能的那一天……它们带给人类社会的不只是工作效率的提高和人的生活改善，将冲击或改变人类社会的某些规则。”^②“居安思危，思则有备，有备无患。”^③虽然目前所出现的弱人工智能产品尚不能对人类社会构成威胁，但是既然我们已经预见到随着深度学习、神经网络、蒙特卡洛树搜索、云计算等技术的不断发展，脱离人类设计和编制程序控制的强人工智能产品有出现的可能，就应当提前在刑事法律层面做好准备，以防止强人工智能产品给现行刑事责任追究制度带来挑战时，我们的刑事立法和刑事司法措手不及、无法应对。

尽管在某种程度上，弱人工智能产品的出现有可能对国家安全、国防利益、公共安全、社会管理秩序、公民的人身和财产安全等造成负面影响，但是正如前文所述，弱人工智能产品只具有工具性的意义，对于其给人类社会带来的风险系数增加的状况，我们只能按照规制传统犯罪的方法对弱人工智能产品的不当研发和使用加以控制，而不能考虑由仅能作为犯罪工具的弱人工智能产品承担刑事责任。但是，一个新的具有自主意志和独立思考能力，可以自主作出决策并实施相应行为的强人工智能产品的出现，必然会对现行的刑事法律体系带来冲击和挑战。人工智能产品的普及完全可能使得其在人类社会生活中的地位从普通工具上升至法律主体的高度，这是不可回避的社会发展的必然。正因为如此，未来人类社会的法律制度都需要针对人工智能技术的发展进行相应的调整。我们必须认识到强人工智能产品与弱人工智能产品在本质上的差异，并明晰强人工智能产品在刑法中的地位。具有自主意志的强人工智能产品自主决

^① “机器人三原则”是指：第一，机器人不得伤害人类，或看到人类受到伤害而袖手旁观；第二，机器人必须服从人类的命令，除非这条命令与第一条相矛盾；第三，机器人必须保护自己，除非这种保护与以上两条相矛盾。

^② 封锡盛：《机器人不是人，是机器，但须当人看》，《科学与社会》2015年第5期。

^③ 《左传·襄公十一年》。

策并在人类设计和编制的程序范围外实施犯罪行为时应如何处理,是强人工智能时代不可避免的“内忧”,也是需要刑法必须正视和解决的问题。

从上文分析中可以得出结论,强人工智能产品完全有能力在设计和编制的程序范围之外实施严重危害社会的行为,也即强人工智能产品在自主意志支配之下的行为完全有可能对刑法所保护的社会关系造成严重侵害。简言之,强人工智能产品有能力对刑法所保护的客体造成侵害。

刑法中所规定的行为在狭义上指的是“危害行为”,即由行为人意志支配之下违反刑法规定的危害社会的客观活动。马克思曾有名言,“对于法律来说,除了我的行为以外,我是根本不存在的,我根本不是法律的对象”,一切法律都是以规制行为作为其内容的。那么“行为”是否必须是“人”的行为?假如我们列出“行为”的所有必备要件或本质特征,而强人工智能产品在自主意志支配下能够全部符合的话,是否意味着法律也应当将强人工智能产品在设计和编制的程序范围外按照自主意志实施的行为纳入规制范围?对于上述问题,笔者已在相关论文中进行过详细阐述。笔者认为,强人工智能产品在设计和编制的程序范围外的行为相比于传统理论对于“行为”的定义,除了不满足自然人犯罪主体需具有生命体的要素之外,其他的要素似乎均符合行为理论要求。而这一结论又是因为传统理论未预见到人工智能技术会取得如此大的进展。就此而言,在强人工智能产品的“智能”高度发展的状况下,突破传统理论显然不足为奇且无可厚非。我们完全可以认为,强人工智能产品在设计和编制的程序范围外的行为可以成为刑法意义上的“行为”。

承担刑事责任的基础是责任主体的意志自由,具体表现为责任主体具有辨认能力和控制能力,而辨认和控制能力的基础是思维能力。自然人的思维能力来源于大脑,强人工智能产品的思维能力则来源于程序和算法。刑法意义上的辨认能力是对行为的性质、后果、作用的认识能力,这种认识包括事实层面的认识和规范层面的认识。事实认识是一种形式上的认识,强人工智能产品可以通过其“电子眼”、“电子耳”认识到事实。相较于辨认能力,控制能力是强人工智能产品的必备能力,是程序设计与编制的必需内容。强人工智能产品拥有极快速的处理能力、反应速度和极精准的控制能力,能够凭借大数据与高速运算能力对行为作出精准的控制。与人类相比,强人工智能产品具有更强的控制能力与可培养的辨认能力。由此可见,强人工智能产品可以符合刑法所规定的犯罪构成中的主体要件。

罪过是主观目的的表现,罪过形式可以表现为故意或过失。所谓故意,是指明知自己的行为会发生危害社会的结果,并且希望或者放任这种结果发生。

所谓过失，是指应当预见自己的行为可能发生危害社会的结果，因为疏忽大意而没有预见，或者已经预见而轻信能够避免，以致发生这种结果。刑法中的罪过形式（故意或过失）包括认识因素和意志因素两个方面。正如上文所提及的，强人工智能产品具有更强的控制能力与可培养的辨认能力，而辨认能力是认识因素的具体体现，控制能力是意志因素的具体体现，所以具有辨认能力和控制能力就具备了在主观上成立故意和过失的两种可能。应当看到，虽然我们应当承认强人工智能产品具有成立故意和过失的可能，但是对于强人工智能产品故意或过失的主观罪过判断，应与对自然人责任主体的主观罪过判断有所区别。原因在于，在认定自然人责任主体的主观罪过时，我们通常以一般人的认识能力作为基本标准并结合行为人自身的特殊情况进行判断。自然人责任主体形成对事物的认识是靠学习和生活经验积累；而强人工智能产品形成对事物的认识是靠程序、算法和深度学习而不是靠生活经验的积累，且设计者设计大部分强人工智能产品的初衷可能是为了让其在特定领域中发挥作用，因此强人工智能产品的认识水平在很大程度上会被局限于特定领域，而缺乏对社会生活的整体认识，其判断是非善恶的标准也就有可能异于常人。但无论如何，强人工智能产品的主观心理可以符合刑法所规定的犯罪构成的主观要件，对此，似乎不应该有太多的异议。

四、人工智能产品的刑事责任

正如前述，强人工智能产品在设计和编制的程序范围内实施的行为，与弱人工智能产品没有差别，所造成的刑事法律风险属于“外患”，此时，我们应将强人工智能产品视为犯罪工具。而强人工智能产品超出设计和编制的程序范围，依靠独立意志，自主决策并实施严重危害社会的行为，所造成的刑事法律风险属于“内忧”，其所实施的行为完全符合刑法所规定的犯罪构成要件，此时，我们可以将强人工智能产品实施的这种行为视为犯罪行为，并给予刑罚处罚。

（一）“外患”情况下的单方责任

“外患”情况，指的是人工智能产品不具有独立意志，在设计和编制的程序范围内实施行为，被行为人利用进行犯罪活动的情况。在这种情况下，设计者是人工智能产品的创造者，并通过程序给人工智能产品设定行为目的和行为边界。此处的人工智能产品可以是弱人工智能产品，也可以是没有超出设计和编制程序，不以独立意志指导自己行为的强人工智能产品。此时的人工智能产品