《学术界》(月刊)
总第244期,2018.9
ACADEMICS
No.9 Sep. 2018

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

○ 陈　伟,熊　波
(西南政法大学　法学院,重庆　401120)

〔摘　要〕"人机交互性"作为人工智能技术的显著特质,是基于智能技术的现实和理论逻辑层面的双重考量,其旨在揭示人工智能犯罪在算法歧视数据的形成和输入、危害行为操作以及法益侵害结果的发展过程中形成的人机一体化模式。人工智能犯罪行为的发生与结果发展的进程具有隐蔽性、间隔性,结果的形态固定也并非如同网络犯罪一样具备瞬时性。因而,刑法立法应当重新构造一种"科技犯罪"上位概念,取代网络犯罪这一"大杂烩"体系,以此形成"计算机系统犯罪—信息网络犯罪—人工智能犯罪"三位一体的科技犯罪规制模式。目前,学界存在的"涉人工智能体刑事责任独立化"与"人工智能刑事归责既定化"两种观点均背离现实的发展境况与科技社会的技术性法律共治理念。人工智能刑事风险应当是"人工"之下智能化行为与结果的支配表现。在此基础上,依据刑法类型化思维,抽象人工智能技术发展在不同阶段的同质犯罪行为,继而相应地设置人工智能犯罪的立法规制模式。

〔关键词〕人工智能犯罪;信息网络犯罪;人机互动性;刑事风险
DOI:10.3969/j.issn.1002-1698.2018.09.008

目前,"涉人工智能体刑事责任独立化"[1]与"人工智能刑事归责既定化"[2]已然形成两派争锋相对的鲜明观点。人工智能技术引发的刑事风险具有内生

作者简介:陈伟,法学博士、博士后,西南政法大学法学院教授、博士生导师;熊波,博士研究生,西南政法大学法学院助理研究员,西南政法大学青少年犯罪研究中心研究人员。

〔*〕本文系国家社会科学基金西部项目"刑罚退出机制的价值确立与实践运行研究"(17XFX009)、重庆市研究生科研创新项目"人工智能刑事风险治理问题研究"(CYS18182)、西南政法大学研究生科研创新项目"科技刑法技术性立法机制研究"(2018XZXS-137)以及西南政法大学人工智能法律研究院科研创新博士项目"人工智能刑事风险的立法规制模式研究"(2018-RGZN-XS-BS-03)的阶段性成果。

性、共生性,其中,风险体系的一部分源于人类对高品质生活的需要和渴求;另一部分源于科技本身掺杂着诸多不确定性和异质性。在这一背景下,“涉人工智能体刑事责任独立化”的支持学者极易忽视现实发展的客观需求,将主观臆造的虚幻风险带入刑事归责理论和刑事立法体系之中,造成司法处断与实践依据的整体崩塌;“人工智能刑事归责既定化”的支持学者对现行刑事立法过于自信,从而面对科技犯罪类型的刑事立法混杂化和智能技术迅猛发展背景下刑事立法滞后化,却一直岿然而伺、无动于衷。人工智能刑事风险应当始终被定位为一种“人工”之下智能化创造的产物类型,对其进行刑事归责。依据客观实际发展,人工智能刑事风险主要存留于智能技术的“研发监管、制造销售、使用管理”三个阶段。

反思现行刑事立法对待科技犯罪的问题治理,则一直存在着科技风险混杂化的尴尬处境。目前学者探究仍立足于网络犯罪整体趋势,予以对待人工智能机器犯罪。[3]然而,网络犯罪仅为科技犯罪的一种类型划分,互联网时代,无论是 PC 互联网还是移动互联网,大家更多关注软件层面的行为危害。因而,网络犯罪立法更多为软件技术犯罪。在既有的刑事归责理论范围内,科技刑法治理应当将目标从网络技术犯罪类型的局部性治理,转向并聚焦于“人工”支配下的智能科技刑事风险阶段化的类型治理。在理清网络技术犯罪和智能科技犯罪之间的关联和区别的前提下,依据人工智能刑事风险阶段的类型化分析,适时调整刑事立法体系的发展方向。

一、人工智能风险的现实逻辑:“人工”之下的智能化

虽然在许多任务上,人工智能都取得了匹敌甚至超越人类的结果。但将人工智能风险归为某种机器或者产品的物质本身,其瓶颈还是非常明显的。问题的根源就在于人工智能仅是“人工”之下现实环境或事物的智能化分析和运用,也是“人机交互性”特质的反映。人脑思维支配和控制的可操作性和涉人工智能体物质经验的脱离性,致使人工智能刑事风险的归责基础,更多地应当将重心偏向于如何疏通“人工”行为对风险的支配和控制性理论研究,而并非仅将目光游离于脱离编制和设计程序内的行为与智能机器的关系配置方面。

(一)人类思维模式的全面输出

无论是强人工智能还是弱人工智能,人工智能运作的基本机制仍在于算法系统,而算法系统依赖于大量现有数据输入和自主决策分析得出。在输出和分析过程中,相应地,人工智能机器的行为举动,在某种程度上反映出数据性质的好坏与外界环境的瞬时变化。刑事责难可能性依据在于行为人对严重社会危害的行为与结果是否有主观明知或者预见可能,[4]能否将人工智能机器的行为危害视为一种刑事风险,进而,将其归为自然人刑事责任的承担依据。其实,这仍取决于科技危害结果或是行为发展是否受到人类自由意志行为支配和控制。当然,弱人工智能机器由于行为的具体操作和后续步骤的推进,完全是来源于人类

指令和代码的程序性输入。在这一方面,可以说并不存在刑事责任划分的分歧。而对于强人工智能机器的“编制和设计程序外的独立危害行为”的性质评价,则取决于深度学习过程中数据输出的来源评价和自主决策系统的数据获取能力。由此可见,数据的性质评价和客观具体情境的变化,关乎着人工智能技术发展在各个阶段危害行为归责方面的具体认知。

而笔者认为,无论是算法系统中数据的输入过程,或是自主决策系统中人工智能技术的深度学习过程,都包含着人类思维模式的全面输出,一种是行为本源性法律风险的刑事责任追溯;另一种则为间接型深度学习数据分析过程中刑事责任评价的方向性指引。

首先,人工智能刑事风险中行为浅层面的现象型风险并非刑事法所评价的危害对象。在人工智能技术发展的高强度表面上,算法系统的数据输入过程与危害行为操作、危害结果发生存在一定的时间间隔,导致无法探究算法系统数据输入过程这一行为本身与后续机器危害行为发生和结果发展,是否具备直接而具体的因果关系,但这并不阻断前行为的数据安全性确保义务这一间接性因果关系源。譬如,在智能机器研发过程,应当确保基本行为指令的数据系统安全性。如果因行为与结果发生的时间间隔,而否定研发过程中的行为故意或过失,直接将使用环节过程中脱离研发过程中的数据安全性确保义务的机器危害行为,视为一种单独的机器人行为刑事风险,那么依据行为发生的时间远近,来选择性确定的刑事归责事实,则是一种现象型的事实风险。管中窥豹的表层现象并非是刑事责任追究的法律风险依据。在后现代化智能工业时代,即使机器能够脱离行为人的全程不间断的操作范围,但是人类还是需要将操作指令输入信息系统,并辅之以基本的监控和维护。[5]

其次,自主决策系统深度学习过程的数据获取能力本质上亦可还原为一种人类思维模式全面输出的过程。强人工智能的显著特性在于深度学习过程中机器对环境、事物变化等数据信息的及时获取能力,将这一能力与人脑神经网结合起来,则形成人机交互的多层神经网络,以此突破和超越人类思维模式。[6]然而,神经网络对于环境和事物变化的捕捉过程,并非是任何类型数据的全面性操作,而是依据使用环境的不同,选择性地进行相应数据类型的择取。无论是立足科技发展层面的现实境况,还是国家政策层面的前瞻性指引,人工智能产品始终被定位为一种服务并满足于人类社会生活、工作等方面的综合化需求的工具。

因而,在智能技术产品的研发过程中,“网络社会赖以存在的技术注定应该是用于满足各类社会需要的工具”。[7]对此,相关生产者和技术研发人员必然会依据不同使用环境过程中的需求类型,来相应地设计智能机器。而选择性深度学习功能的数据分析就来源于智能机器具体使用和管理阶段中数据模块的输出。据此,就危害行为刑事风险的源头定位,仍可将“人工”之下的智能化刑事风险,依据研发监管、使用管理阶段的数据分析,直接或间接性评价为一种人类思维模式的全面输出过程。

(二)人类行为场域类型的扩展

人工智能犯罪不同于信息网络犯罪,单独将弱人工智能作为一种犯罪工具进行杀人、盗窃或者伤害行为,并非是展现智能犯罪特质性的一面。信息网络犯罪和人工智能犯罪的显著差异就在于前者可以突破地域范围的局限性,广泛地实施特定或者不特定的严重暴力行为或者恐怖犯罪行为等社会影响恶劣事件;而后者的行为操作依据的背景是虚拟空间,[8]并且,人工智能犯罪这一潜在危害并非如同人类亲自实施的实行行为,一种存在可随时支配、控制的刑事风险行为。本质缘由在于信息网络犯罪和人工智能犯罪行为载体的空间类型不同,信息网络犯罪的空间类型虽然具有虚拟性,但是毕竟空间的虚拟性桎梏了行为发展至实体场所的衍生性。[9]譬如,网络谣言对社会秩序的危害性发展不同于寻衅滋事罪中的"随意殴打、追逐、拦截他人"的行为危害。前者可在网络空间和实体场所空间随意扩散;但是后者的行为发展仅限于现实的公共场所类型。[10]而人工智能犯罪主体实施寻衅滋事罪的构成要件,完全可实现智能载体的虚拟操作和危害结果的同步性。但该点仅表明人工智能犯罪创造了一个全新的人类行为场所类型,其行为操作并不同于网络空间的全局虚拟性,亦不同于现实场所类型事物的整体可视性。因而,这并非能够成为涉人工智能体的刑事责任单独化评价依据。

之所以对比信息网络犯罪和人工智能犯罪背景下行为载体的差异,就在于两者的区别和联系,为"人工"之下智能化犯罪行为的操作环境的认定奠定了现实基础。此外,信息网络犯罪亦不同于计算机系统犯罪,其纯粹将互联网作为虚拟空间危害结果实现的一种工具。而人工智能犯罪既能将智能载体作为一种犯罪工具,亦能将智能载体作为一种犯罪对象。[11]人工智能犯罪能够通过算法系统的破坏,直接实现数据自主分析过程的紊乱,从而导致因"算法歧视"[12]造成的犯罪行为。就此而言,涉人工智能体的科技创新,其实是人类行为场域类型化的一种扩展,从计算机系统犯罪到信息网络犯罪,最后再到人工智能犯罪。追本溯源,三者显著的区别便在于行为发生的场域背景不同。人工智能犯罪中,研发、使用过程中的罪过主观预断,包含了"自然人利用隐蔽的智能机器犯罪(如行为与结果的高度间隔性)来掩盖先前行为的场域类型扩展"的故意和过失心理,而这一情境显著地影响着未来智能犯罪刑事治理的决策与行为。

(三)行为危害进程的前期预断

不同于前述人类思维模式在人工智能技术研发和产品使用阶段的全面输出,数据输入为人工智能机器危害行为的罪责评价提供的客观依据,则是从前行为的理解层面进行阐释,表明智能科技犯罪是"人工"之下的智能化危害行为。但是,对于在基本数据输出正确的情况或是不存在任何"算法歧视"的处境下,是否亦可以将危害结果的发生,归责于自然人主体"人工"之下的智能化支配作用力?笔者认为,这是肯定的。

其一,对于自主决策系统未出现任何紊乱情况下,所发生的人为无法控制危

害结果的产生、发展到最终定型的系列过程,我们应当借助客观归责理论,将该种结果视为一种可允许性风险。[13]因为,“当今中国社会呈现出‘私人风险’与‘公共风险’交织的局面,但公共风险尤为显著,这既有现代工业化的内生性原因,也有外在人为原因”。[14]对于智能新兴科技引发的危害结果,刑法无法承担起苛责的责任要求,全面管控人工智能风险类型。唯有符合社会集中利益最大化的情况下,对人工智能科技风险发生,进行技术研发和使用阶段行为标准的深度提升,以防止客观的智能风险无端扩散。

其二,在满足自主决策分析进程完美无瑕的情况下,对于符合过失或者故意罪过心理的危害行为,应当将其评价为一种滥用科技产品和技术的犯罪行为。就深度学习系统正常的机器人而言,面对后续危害行为的发生,人工智能系统“不具备人类的核心认知功能。它们没有自我意识,它们不能理性思考自己的行动目标并对其进行调整,它们不会因为实现目标的憧憬而感到兴奋,它们并不能真正理解自己所做的事情”。[15]而这一切均可归为一种“人工”智能滥用科技产品和技术的结果预断行为。诸如,在研发、销售、使用和管理阶段,自然人未严格按照研发技术的标准程序进行制造、未及时对展柜上的智能机器进行检测和维修、未履行智能机器的操作步骤和及时提交瑕疵产品进行检测、未执行技术的严格评定和不合格产品的查封行为,最终因上述行为过错导致一系列严重危害社会和国民的结果。即使是在深度学习过程中机器自主无意识实施的危害行为,亦应当将刑事可罚性归为相应阶段的自然人。因此,相应阶段“人工”支配和控制下过失和故意的罪过心态评价,仍可在算法系统正常的情况下予以进行。

二、风险治理的理论逻辑:智能风险的责任来源

人工智能犯罪与信息网络犯罪之迥异不仅在于行为发生的载体背景或者场所类型,在社会危害性的类型评价方面,人工智能犯罪亦不同于信息网络犯罪。人工智能犯罪由于其行为模式可以在虚拟和现实共同体中得以实现危害行为的全程性操作,所以,我们应当探寻一套适合智能风险科学化治理的立法模式,以此应对新兴人工智能刑事风险的来临。而这一切均无法脱离现实智能产品运行过程的阶段化认定,亦即,分别对人工智能机器的研发、销售、使用和管理等阶段,采取类型化分析思路,认清人工智能刑事风险的客观来源。

(一)义务来源的类型化

为何将人类对人工智能机器的危害行为支配力的发展进程,划分为技术研发和监管阶段、机器的制造和销售阶段以及产品的使用和管理阶段,主要在于各阶段的义务来源决定着各罪名在科技刑法体系中的定位。采取义务来源类型化的思维,在于智能科技发展的长远性和技术规模的不确定性。因此,就目前来看,刑法立法者无法构建一套精细化的智能科技犯罪立法体系。但是至此,立法者能够做到的便是首先从义务来源的类型化思维出发,克服网络犯罪中行为义务确定模式中“概念式思维的抽象性、封闭性与断裂性而生成的具有具体性、开

放性与过渡性的思维特征”,[16]以便为智能科技立法提供背景支撑。

1.确保技术安全监管与审慎研发义务要求。技术研发和监管的安全性义务作为整体流程的萌芽期,义务的核心要求应当以类型化思维为基础。[17]智能技术研发阶段决定着初始算法系统的性质好坏问题,技术监管则在研发阶段全过程中发挥着举足轻重之功效。对于确保技术安全监管与审慎研发的义务要求,应当注重如下两个方面:(1)技术安全监管主体包括研发单位和技术、产品管理部门。其实,对于研发单位和技术、产品的管理部门来说,其承担着实施对后续脱离设计和编制程序的机器人自主决策的危害行为的保证义务,以及监督积极履行研发技术的安全义务,是其作为保证人应尽的职责。(2)义务履行的最低标准应当是“作为可能性”。之所以研究研发者和监管者的义务来源要求,是因为上述阶段作为智能产品的初始环节,如果放纵技术研发和监管的滥用行为,则意味着在可移动的涉人工智能体中潜伏着巨大的安全隐患。而对于作为可能性的具体要求,即表明现实履行的义务实现。诸如,研发者和监管者应当意识到自主决策系统的技术疏忽,会导致后续机器行为操作的失范。

2.及时排除产品制造和销售隐患的义务。智能产品的制造和销售阶段,是智能技术的运作和产品成型、流通环节。在这一环节中,如果忽视产品的瑕疵责任,导致智能产品在制造和销售环节的安全隐患,亦可将其评价为一种刑事责任的义务来源。[18]作为跨越医疗、交通、工厂、家庭等诸多特定或不特定的公共和私人场所、领域的人工智能产品运用,其制造和销售的义务来源于人工智能产品系统中隐患的风险排除义务。对待信息网络犯罪,网络犯罪预备行为因为“可能威胁重大、众多法益,而且其法益侵害危险较之传统犯罪预备具有倍增性、现实性和不可控性”,[19]因而具备了预备阶段行为风险扩大可能性排除的义务责任承担,更何况于涉及众多民生领域的人工智能犯罪行为。然而,在生产、销售普通瑕疵产品的刑事责任认定中,“国家标准和行业标准”作为刑事义务来源的转折点。如何审慎认定不同领域标准,便成为人工智能刑事风险消解的关键性要素。因此,规范智能产品的标准建立问题,可以首先从构建一套完整的国家、行业标准体系,加强不同智能产品运用领域的义务和责任等方面着手准备。

3.符合机器正常使用的标准要求义务。对于智能技术和产品的标准设置,应当包含“研发标准—制造标准—使用标准”。在智能机器的使用阶段,大数据的系统输入尤为重要。依前文所述,智能机器是依据不同行业、领域的个性化需求,满足各个阶段的高效操作而服务的。不同于信息网络犯罪,在人工智能犯罪过程中,自主决策分析和后期使用的大数据输入充当着“领头羊”的角色。由此可知,符合机器正常使用的标准义务要求的实现,关键之处在于大数据信息决策整体过程的安全性确保。具体而言:其一,输入、点击和管理等其他机器操作手续的程序性的义务要求;其二,数据集合过程中信息本源安全性的义务要求,亦即确保数据分类的合理性和“人机互动性”正当展开;其三,算法系统运行过程中正常维护的义务要求。

(二)现实结果的类型化

危害结果的类型表达一直是刑法教义炙手可热的研究话题。人工智能的刑事违法性何以归为自然人主体,重点在于危害结果的现实产生,亦即产生了一种法益侵犯的威胁和结果。客观构成要件中最为重要的构成要件之一便是结果。就结果类型的上位概念而言,智能科技犯罪中的现实结果包括实质犯和形式犯,而实质犯这一概念是指“发生法益侵害或者使得法益危险的犯罪”,由此得知,实质犯由侵害犯和危险犯两种类型构成。[20]立足智能产品技术的研发、销售、使用等三阶段的法益侵犯,人工智能犯罪的侵害犯要求法益现实地遭受侵害;危险犯则只需要发生法益侵害的智能技术危险就足够了。而与实质犯相对应的则是形式犯,亦即,只要形式上违反了智能技术管理刑事法规就能成立犯罪。[21]针对实质犯和形式犯、侵害犯和危险犯,如何将其分门别类地运用于各个智能技术阶段的犯罪立法体系之中,仍需要具体理解各阶段侵害结果类型的性质来源。

1. 抽象危害如何认定?义务履行的前提在于危害结果的可能预见性,如果某种特定危害在行为前便无法预知其即将来临,那么义务的强加便无任何意义。易言之,对于研发者和使用者造成的抽象危害,亦应当立足抽象危害结果的发生可能性。譬如,智能技术的研发者违反一般业务标准,本应当对智能医疗领域的产品技术进行相当规模或数量的评测,但因为过于轻信自身的技术研发水平,即使该项智能技术还未或者即将投入生产、制造环节,而未造成特定的危害结果,也应当将其认定为一种“瑕疵技术障碍”的抽象危害。当然,这一前提也应当建立在其既可能预见危害结果的产生,也确实未履行监督义务的基础之上,否则应当按照意外事件处理。[22]然而,瑕疵技术障碍对接到具体侵犯的法益类型或程度上是否有所限制?答案是肯定的。规范保护目的认定作为一种价值取向,其内涵和性质以及适用范围,会因为实务经验或者学识阅历的不同而产生分歧。[23]因而,宽泛设置抽象危害行为模式,必然会导致适用实体依据的普遍存在。

化解抽象危险犯认定的无所适从,使刑法立法道阻且艰。原因在于抽象危险犯和具体危险犯的具体认定存在争议。对于抽象危险认定模式的设置,有学者推崇“准抽象危险犯”的模式设置,认为借助当前 5 个“足以”型危险犯的立法形态,亦可化解抽象形态的危险认定过程的抽象化,其并不要求对“足以造成……”附加形成“具体、紧迫、现实”的危险,也可成立准抽象危险犯。将抽象危险结合具体个案和实际情况,进行刑事违法性的实质评断情形,可以涤除具体危险犯的具体个案紧迫现实危险的约束性之困境。[24]就人工智能技术研发和监管阶段的过失疏忽而言,有时此种主观心态并未可能产生现实、紧迫、具体的个案风险。但其又并不同于形式犯(抽象危险犯)的行为一经作出,则满足个罪的构成要件,成立既遂。[25]因此,准抽象犯的概念塑造,为人工智能刑事风险的前期防控奠定了理论基础。但不可否认的是,准抽象危险犯的设计理念仍需借助司法解释的具体情形认定,以初步构建情节严重的诸多类型。就这一弊端而言,笔者认为,可以从各种情形的可能现象中,抽象化几种类型特征,来设置一些具

体智能技术研发的监督过失行为或是数据指令输入造成智能系统紊乱的情形，以尽可能实现罪刑法定对法律文本与具体司法个案的衔接要求。

2. 具体实害如何认定？由于人工智能犯罪能够实现双层社会空间的法益侵害，因而，如同普通犯罪一样，其侵犯的具体实害涵射国家公共利益和个人人身、财产利益范围。但就具体内容而言，信息网络犯罪和人工智能犯罪在侵犯法益的形态界定方面存在着显著的差别。不同于计算机系统犯罪和人工智能犯罪，信息网络犯罪立法更多关注的是实害结果的认定。在《刑法修正案(九)》中，编造、故意传播虚假信息罪，拒不履行信息网络安全管理义务罪，非法利用信息网络罪，帮助信息网络犯罪活动罪等分罪设置均采取“情节严重”或者“严重后果”作为入罪标准。在法益侵害行为或者后果模式中，值得考究的是，其行为模式的设置均采取类型化表述，这一标准值得人工智能犯罪立法镜鉴。人工智能犯罪不同于信息网络犯罪，后者不存在危害行为和结果样态的阶段化认定，其具体危害不可一概而论。由于算法系统的歧视效果依随着数据输入和深度学习而衍生，因而，人工智能犯罪更多侧重于一种行为阶段化和类型化评价，其思想根源于刑事立法类型化之倡导。对于实害结果的类型化认定主要存在诸种情形：

(1)技术研发行为。技术研发行为中实害结果的认定情形主要包括业务上的过失致死伤行为。对于具体业务情形认定应当有所限定，其存在于从事高度智能危险业务的人员，因其比普通人员更具注意义务。此外，对于业务人员技术研发漏洞致死行为的罪责承担主要在于其反复持续地实施危险行为。(2)产品制造和销售行为。产品制造、销售行为的实害结果情形主要在于瑕疵产品展示和流通环节中致人死伤和严重经济损失的行为，对于产品制造和销售行为的情节认定标准应当低于上述技术研发行为，主要考虑的仍是技术操作的熟练程序和行为注意和结果预见义务的行业差距。(3)机器使用行为。机器使用行为的实害结果情形，主要在于数据的输入行为所致的严重实害。对于算法系统的紊乱这一实害如何评价，则存在着刑事立法涵射的不周延现象，传统的计算机系统能否包含“人机交互性”特征显著的智能系统仍值得学界探讨。虽然“行为人自己亲自往计算机系统中输入不当或者不完全的数据的，无疑属于计算机犯罪行为的‘使用’工具行为”。[26]但是我国刑事立法中的计算机犯罪行为认定，却主要仍在于固定设备中软件数据的破坏行为之规制。[27]

三、制度运行的既有困境：模式缺乏技术可操作性

我国刑法虽然已经初步形成较为完善的“计算机系统犯罪—信息网络犯罪”治理体系，但面对“互联网＋”向“人工智能＋”的社会转型，[28]刑事立法仍存在不足与缺憾。最主要的缘由就在于现有的立法模式混淆信息网络犯罪与人工智能犯罪，导致目前面对技术超越的人工智能刑事风险，其缺乏技术可操作性。基于人工智能刑事风险治理的双重逻辑思维，可以明确人工智能犯罪不同于信息网络犯罪，前者行为违法性或是法益侵害性的表层化现象凸显较为缓慢

或者隐蔽,其行为潜伏于多个阶段,使得人工智能刑事风险呈现出量变到质变的犯罪化过程。综合刑事立法体系来看,在人工智能犯罪既有的可能风险内,如果忽视这一变化,那么出现诸如“涉人工智能体刑事责任独立化”与“人工智能刑事归责既定化”等幻化或滞后的立法现象便不足为奇。

(一)只眷注危害结果的形态,缺乏技术差异理念的融合

现行刑法立法偏好结果型立法,即在现有的风险社会中存留哪种显著的危害结果,便转向该种结果类型的犯罪模式研究。这便是刑法立法活性化的典型表现,亦是国民安全焦虑感在刑事立法方面的映射。[29]刑事立法的时代发展历来承担着化解国民体感不安结果的艰巨使命,立足目前乃至未来较长一段时间,科技犯罪的样态流变中,其实一直存在着三种类型化的刑事风险形式,亦即,计算机系统犯罪、信息网络犯罪、人工智能犯罪。计算机系统犯罪是程序性的控制、侵入,包括内部数据的破坏或窃取;信息网络犯罪是工具型的支配和利用,将信息网络作为违法犯罪的平台和手段;而人工智能犯罪是一种兼具工具型和对象型的犯罪表现,将现实直接性的普通犯罪行为通过间接乃至直接的数据输入输出过程,达至不法的犯罪目的。其中,三者最大差异之处,便在于技术性理念的不同运用。计算机系统犯罪、信息网络犯罪是一种平台利用犯罪化以及危害结果的直接性和瞬时性发生过程,人工智能犯罪是一种数据系统犯罪化以及危害结果的隐蔽性和间接性发展历程。(见表 1)在这一显著特质性背景下,刑事立法难以透过三者危害结果的现象型风险,以寻求立法体系的区别对待,这实质上是在规范层面上忽视技术差异理念在制度体例上的交叉融合。

表 1　科技犯罪的发展阶段与类型划分

	犯罪类型	犯罪特质	行为方式	结果类型	作用评价
科技犯罪	计算机系统犯罪(数据的存储和传输阶段)	行为危害的局域性和固定性	控制、侵入系统;提供控制和侵入工具和程序;破坏系统(包括制造破坏系统的病毒等程序的)	实害犯	对象
	信息网络犯罪(弱人工智能阶段)	工具型犯罪的传播性、瞬时性	拒不履行信息网络安全管理义务;利用信息网络;帮助信息网络犯罪;编造、传播虚假信息等行为	侧重危险犯	工具
	人工智能犯罪(强人工智能阶段)	机器人犯罪(再造和使用期间:行为触及局部性,制造期间:危害效果的广泛性,持续性)	制造程序漏洞风险;滥用人工智能技术行为以及产品的科技监管失职行为;人工智能技术制造和销售行为;过失使用行为	研发监管、制造销售阶段侧重危险犯;使用和管理阶段侧重实害犯	工具与对象

刑事立法层面难以渗透技术差异理念,以致出现混淆人工智能犯罪、计算机系统犯罪和信息网络犯罪的尴尬处境,归根结底这是一种结果型立法的常态现状。其实,未深入技术创造性的集合过程,从表面上看,人工智能犯罪、计算机系统犯罪和信息网络犯罪在危害结果形态表现上毫无差异。因而,即使是在忽视技术性差异理念的际遇下,也能借助关联罪名的同种危害结果,来实现犯罪行为的刑法规制目的。诸如三者均能实现侵犯公民个人信息罪的数据权利侵害的危害结果。因为归纳信息系统产生的数据来源,均能够从人工智能系统、计算机系统和信息网络中得出结论。[30]但这种铤而走险的粗犷式、发散性刑事立法背离了刑法类型化的相对确定性,从而导致刑法类推适用的人权侵蚀现象频发。

(二)只关注行为模式的认定,缺乏行为属性的明定

《刑法修正案(九)》在创设大量的信息网络犯罪模式之际,随之而来的便是刑法类型化思维的运用,将调适理想类型、具体案件事实群形成与国民的可预测程度结合起来,[31]成为此次刑法修订的一大特色。结合具体罪名而言,诸如,非法利用信息网络罪的三种类型化非法利用行为模式;帮助信息网络犯罪活动罪的技术支持等六种具体类型化的帮助型犯罪行为。这便给国民规范自身行为操作提供了基本指引,以防止在具体相关司法解释出台前造成行动自由的不便。但是与此同时,我们更应当看到该行为模式相应的不足之处,刑法类型化适用之余,如何避免"技术归化"后行为属性界定之缺乏?信息网络犯罪的大规模设置,带来诸多譬如"信息网络安全管理义务""非法利用信息网络"以及"帮助信息网络"等具体技术性行为的理解和适用问题。因为,行为模式类型化设置并非精确性的数量例示,而是采取"具体+兜底"形式的设置特色进行的。对于兜底性的行为模式,应当类比具体行为模式的类型特征进行规范解读。

"'技术归化'是网络社会治理的必然要求和核心内容,网络社会治理法治保障体系同样建立在技术保障的基础之上,充分运用现有技术,又要放眼长远、占领未来网络社会的技术高地"。[32]因而,人工智能犯罪立法模式的借鉴应当立足现有的计算机系统犯罪、信息网络犯罪模式的既有形态进行比较研究,方可知悉智能数据系统的法治治理体系的大致方向和体系构架,进而为人工智能技术支撑提供制度保障。在此基础上,人工智能的刑法立法类型化模式的现实需求,应当被附加相应行为属性的前提要义。除此之外,还应当强调此种行为属性的界定过程应当极力避免"对象的事实性白描",[33]其是一种基于特定规范目的和思维逻辑而抽象化的同质行为,并依据技术归化的设计理念和科技犯罪的惩治需要而进行的。

(三)只注重法益类型的界定,缺乏了解结果的演进

人工智能犯罪行为新类型形态在创设一系列前所未有的刑事风险之余,也带来新一股人工智能刑法法益类型界定过程的探讨清风。将人工智能刑事风险归为一种"人工"支配之下的智能化法益侵害,是立足现实基础进行的,能够实现现有刑事归责理论和刑罚适用目的的体系自洽与和谐统一。在人工智能犯罪

行为新类型形态产生之前,对现行法律保持敏感度可以避免将刑法理想化。因为“任何源自新的刑事责任模式的归纳,都将导致与刑法理论家所推崇的规范原则相矛盾”,[34]最终造成刑法理论的整体崩塌。因而,对人工智能犯罪的法益类型界定能够充分体现理论界和立法者对人工智能刑事风险的体系定位与属性确立。

然而,人工智能犯罪法益类型的界定过程如何进行,应当归于教义学的规范解读。其实,反思现行刑法对法益的立法确认,无论是计算机系统犯罪还是信息网络犯罪,都仅关注行为发展的局部性后果,继而导致部分罪名设置存在着交叉重合之处。譬如,破坏计算机信息系统罪包含部分非法侵入计算机信息系统罪和非法获取计算机信息系统数据、非法控制计算机信息系统罪。无独有偶,在网络服务提供者触犯帮助信息网络犯罪活动罪时,也存在部分拒不履行信息网络安全管理义务罪的重叠构成要件。如此一来,便导致不同业务主体之间适用刑法的错乱和不公。在网络科技发展过程中,从计算机系统犯罪到信息网络犯罪的罪名体系设置,如果仅关注行为发展的局部性后果,从现象型的刑事风险入手,还极易将一般行政违法行为的规制范畴纳入刑事治理体系之中。再如,非法利用信息网络罪中三种行为类型模式的设置都涉及违法犯罪活动的认定,因而,便出现诸如是否意味着利用型信息网络犯罪仅一般的违法犯罪行为亦可。[35]甚至有学者认为应当对非法利用信息网络罪中的“违法犯罪”作广义的宽泛理解,并进一步指出:“采取严格解释得出的限缩结论不妥,可能导致实际的适用范围非常有限,增加司法证明的难度,不利于提前介入高度危险的网络预备行为。”[36]由此得知:在上述两种情况下,从犯罪行为发展的局部性后果入手,采取结果发展的一部分进程,仅关注预备行为或者帮助行为的两个阶段性,还易混淆刑事不法和行政不法的边界问题。但从实质的罪刑法定原则来讲,这违背法益保护的比例性原则的必然要求。

追本溯源,截取行为发展的局部性过程,是囿于现行刑法立法对传统犯罪结果型立法特色的严格遵照,从而仅关注法益类型的具体界定,以此来设置具体的计算机系统犯罪和信息网络犯罪的行为模式。刑法立法界认为两者犯罪行为所侵犯的具体法益为计算机和信息网络社会的公共管理秩序,从而以此为起点,发散性地设置与该法益牵连的相关行为,从而忽视了结果发展演变的渐进过程,导致司法实践出现罪名适用不一的情况出现。诸如,邹某、彭某利用网络平台诈骗案。[37]由于该案件涉及新旧法交叉适用阶段,从而法院利用从旧兼从轻原则,依据《刑法修正案(九)》判决两被告触犯非法利用信息网络罪。而检察院抗诉认为,两被告的行为符合非法利用信息网络罪以及诈骗罪的想象竞合犯,理应择一重罪处断。该案的罪名适用不统一在于依据结果型立法或是仅关注法益的类型界定,容易出现罪名设置的交叉现象。尤其是在人工智能犯罪中,因为该罪名既可符合普通罪名的构成要件,又关联特殊的危害行为模式,因而要避免出现以法益界定为中心的发散型、局部型刑事立法现象,可采取刑法类型化立法发展,从关注危害结果发展的多个进程出发,分别根据危害行为所处的“研发监管、制造

销售、使用管理”三个阶段类型,来研究设置相应的罪名,以防止罪名体系的部分重合。

四、刑法规制的模式转变:从“局部性”到“阶段化”

人工智能刑事风险作为一种“人工”支配下的风险类型,人工智能犯罪立法模式应当立足现有的刑事归责理论展开研讨,无需将智能时代的科技风险一味归于智能机器,其既无法实现现实意义和达到刑罚适用效果,反而为人类自主创新智能科技的疏忽寻找借口,逃脱刑事责任承担。我们有必要开始省思现行刑法是否能够契合人工智能刑事风险新类型的发展趋势,以及现有的计算机刑法和网络刑法是否能够包容“气象万千”的智能时代危害行为。至少从前文所述的刑法立法的诸多技术性理念缺失可知,采用现有的“计算机系统犯罪—信息网络犯罪”双层规制模式以化解人工智能刑事风险,远不可企及智能科技社会的发展情势。

(一)前提概要:塑造人工智能犯罪的立法新理念

计算机系统犯罪和信息网络犯罪的规制思路较少有学者探讨,多数学者忽视计算机系统犯罪和信息网络犯罪的技术差异,直接从危害结果的显性风险研究转至刑法立法方法的研究,导致现在刑法理论界认为网络犯罪包含了一切的科技性工具和手段的犯罪行为,如张明楷教授认为,网络犯罪虽然作为一种严重犯罪类型,但是其体系实际上包罗万象,既包括新型犯罪也包括与计算机数据和系统相关而实施的传统犯罪。[38]但是,准确地立足刑法教义层面,其实刑法立法在修订历程中,逐步地显现出刑法类型化的一种思维现象。在《刑法修正案(九)》中,新增的几种罪名类型均以“信息网络”为犯罪的平台和载体。从此,计算机犯罪不再能够涵盖新型的网络犯罪形态,计算机系统犯罪与信息网络犯罪形成两种鲜明的体系对比。但是,略显遗憾之处在于刑法立法体系由于缺乏对危害结果发展过程的整体认知感和技术认同感,到目前为止,网络犯罪仍然是一种上位概念,包涵众多复杂的科技犯罪。而就笔者认为,准确来说,刑事立法首先应当树立一种全新的技术犯罪新理念,亦即“科技犯罪”作为一种上位概念,取代现如今推崇的网络犯罪。详言之,在智能科技犯罪这一大背景之下,以软件技术的运用和开发程度为中心,以科技发展的时间背景为轴线,形成“计算机系统犯罪—信息网络犯罪—人工智能犯罪”三足鼎立的刑法立法模式。

人工智能犯罪技术性差异理念的融合乃至行为属性的界定都建立在对相关概念和属性的理清基础之上,如人工智能犯罪、信息网络犯罪、计算机系统犯罪三者之间的概念构筑;再如算法系统、数据歧视以及互联网平台三者之间的联系和区别。毋庸置疑,信息网络作为交流信息的平台和载体,为虚拟空间和现实空间提供资源共享之路。与此同时,犯罪行为的法益侵害也随之从虚拟空间扩散至现实场域。对此,在这一背景下,最高人民法院部分学者认为:“网络犯罪分为对象型网络犯罪(纯正网络犯罪)和工具型网络犯罪(不纯正网络犯罪)。”[39]

而相对于计算机犯罪则是指“计算机操作所实施的危害计算机信息系统(包括内存数据及程序)安全的犯罪行为”。[40]这种类型的计算机犯罪,实际上是指只能在计算机空间所实施的犯罪。由此可知,按照该种思路发展,网络犯罪便包容了一切利用计算机系统和互联网技术的危害行为和结果。而在部分学者看来,“网络因素的介入,改变了组成犯罪的‘原料’和‘元素’,导致了犯罪的构成要件要素的变异,包括犯罪对象、犯罪行为、犯罪目的和犯罪结果等方面”。[41]因而,相对于“互联网 1.0 时代”,网络犯罪是一种新类型犯罪,表现在传统犯罪构成要件的诸多方面之变异。结合网络平台的赌博、诈骗、非法经营等司法解释和《刑法修正案(九)》的新设罪名,均可证成刑事立法上的网络犯罪,严格意义上,仅为“信息网络犯罪”,一种信息共享资源交流、利用基础上的科技犯罪,一种不同于计算机系统犯罪软件、数据信息破坏基础上的攻击性行为。

大数据分析和深度学习的自主决策,为科技犯罪的体系扩充提供现实基础。人工智能犯罪不同于计算机系统犯罪和信息网络犯罪,后者无论如何,行为发生的即刻,则可预见到系统指令操作的违法性范围。而人工智能研究的核心,通常是开发各种像人类一样具有某种用思考能力的软件,配合电子计算机超高速的计算能力和超大的存储容量,支持人类完成各种任务。[42]因此,“人机交互性”作为人工智能犯罪的典型表现和独特属性,意在凸显“人机共存”的行为一体化过程。智能技术危害行为的发生和危害结果发展过程具有隐蔽性、间隔性,因而结果的形态固定并非如前者那么具备瞬时性。对于智能刑事风险的不确信和结果发展的漫长过程,要求刑事立法必须提高人工智能产品系统漏洞的前期预见可能性认识。督促相关阶段的责任人员提前排除自主决策的技术瑕疵程序,如此一来,人工智能犯罪立法的刑罚苛责便具备必要性以及现实意义。

在智能科技犯罪规制新模式下,刑法类型化思维的理解和运用议题是模式构建的前提之一。人工智能的刑事风险定位目前尚存在较多争议,人工智能技术的分类方法在科技界也存在可探讨空间。[43]但是,这并不影响刑法立法对人工智能刑事风险法益侵害性的整体把握,因为技术性理念的立法融合并不要求对智能科技发展的技术运用原理进行解释,只需在通用技术手段的基础理解下进行刑事技术立法便可。例如,刑事立法不需要表明人工智能技术发展不同阶段下犯罪行为发展过程,以及解释为何人工智能犯罪会产生危害结果发展过程的间隔性和隐蔽性,因为其本可以归属为犯罪学和刑法教义学的研究视域。人工智能犯罪的刑法立法类型化,只需明确界定人工智能犯罪在不同危害行为阶段化属性下法益侵害所致量刑配置之差异即可。

(二)规制模式:聚焦人工智能犯罪发展的全过程

随着人工智能技术研究的如火如荼,智能机器已经悄然地步入现实生活之中,并潜移默化地影响着关联管理制度的发展。随着一系列人工智能技术创新战略以及自动驾驶汽车的制度规范出台,人工智能风险已经成为一种不可撼动的科技风险新类型。

在刑事法领域,如何界定因数据权利适用不当所致的人工智能刑事风险,是人工智能犯罪规制模式的关注重点。“无论是立法者还是研究者,在讨论个人数据权利配置时,都应当注意协调多方的利益关系”,[44]而在数据权利配置应建立在审慎履行人工智能刑事风险最大化预见可能基础之上。“人机交互性”作为人工智能犯罪的独特属性,相应地,制度构建应当从数据系统的危害行为发展着手进行。基于此,利用人工智能机器所进行的传统犯罪行为并非为本文关注和探讨的重点内容,其亦非属于人工智能刑事风险要素的新类型。根据前文对人工智能犯罪结果类型的探讨,人工智能犯罪的法益侵害在于智能技术的研发和监管阶段、智能产品的制造和销售阶段以及使用和管理阶段具体化的抽象危害和具体实害。具体而言:

1. 风险创设的初始环节:技术研发阶段和监管阶段。人工智能产品的安全性取决于“深度学习”的算法质量,以及它所学习的数据集的完整程度。再深度挖掘,自主学习的算法质量则来源于智能系统对外界环境变化的感知能力和应对能力。在技术研发和监督阶段,如果法律允许放任存在系统漏洞的智能科技流入制造和使用环节,无异于间接支配和控制着严重危害肆意侵蚀国民的整体安全感。诚如,有学者认为:“对于人工智能引发的现代性的负面影响,有必要采取风险措施,即预防性行为和因应性制度”。[45]其中,预防型和因应性制度要求探寻人工智能刑事风险的最初源头。

基于此,笔者认为技术研发阶段和监管阶段,应当成为预防型人工智能刑事风险的初始环节。对于技术研发行为,由于受到生产伪劣产品罪、投放危险物质罪等罪状模式设置的法益类型、行为手段以及主体身份的限制,目前现行刑法缺乏单独的对研发人员主观意志内技术生产的可控漏洞行为的规制模式。而此类故意、过失制造技术风险的行为在未来的人工智能时代将会逐渐显现出来,目前刑法立法的设置却对此捉襟见肘。[46]因此,笔者认为可单独设置非法制造技术风险罪与过失制造科技风险罪。其中,责任主体为科技的研发者,而研发者往往作为一个团队的形式存在,因而划分团队内部的技术研发的作用力完全是有必要的。但对此,有学者也担心若分别依据团队内部技术研发人员的不同作用力,追究相应的刑事责任,会存在量刑过轻和消解业务监督、协作机制的发挥之弊端。[47]出现此类立法担忧并非庸人自扰,也情有可原。但其忽略了互相监督、协助的工作义务缺失何尝不是一种研发工作的渎职行为。关键在于刑事立法应当在模式塑造时明确智能技术研发团队的内部监督、协作的义务性要求。两类罪名的危害行为模式应当为技术漏洞的故意、过失制造行为,并可以单独设置单位犯罪。其中,系统瑕疵的结果“已经预见”不是一种抽象的“预见”,而是认为智能机器人在未来可能会危害到人类社会,并且该种“预见”是可推测到的具体内容,即预见到自己研发或使用的智能机器人,在投入使用后可能会实施严重危害社会的犯罪行为。那么则可以成立相关过失犯罪,承担刑事责任。[48]

政府监督是严密科技社会治理法治监督体系制度建立的主导。[49]对于人工

智能产品研发的监管行为,行政监管者和风险监测部门的测评者应当谨慎和严格履行智能风险监控和测试行为。否则因严重职责疏忽或者明定的放任故意,促使漏洞存在的智能技术流通进入制造、销售环节,造成国家利益、公共利益或者公民的人身和财产损失的,应当被评价为一种滥用职权或者玩忽职守行为。对于模式构建的大体方向可借鉴刑法第 408 条的环境监管失职罪和食品监管失职罪,对造成大规模智能刑事风险损害的,界定不同于玩忽职守罪和滥用职权罪的法定刑模式,提高基本法定刑,提升入罪标准。对造成公私财产遭受重大损失的、造成人身伤亡的严重后果的,设置科技监管失职罪;对于故意放任智能漏洞行为,设置放纵不符合技术标准的科技产品罪。

2. 技术风险的流通环节:机器的制造和销售阶段。在目前既有的人工智能刑事风险应对理念的探讨之中,有观点认为应当加快建构研发者和使用者的刑事义务体系。[50]但是对于智能机器的销售和制造环节的研究仍尚存空缺,“技术固然沿着自己的规律在前行,但其进化进程也受到了人类需求的直接影响”。[51]这一直接影响不仅来源于技术的研发和监督过程,还受制于技术研发过后的机器制造和销售阶段存留的智能刑事风险。人类需求作为产品制造和销售的导向标,能够调控智能技术研发的基础进程。诸如,固然智能研发技术的缜密操作和监督环节严格履行,能够提前清除人工智能刑事风险的侵蚀,但并不可否认仍然可能存在着在技术转化为生活产品需求之后,不按照智能产品的设计标准和程序去制造以及销售过程中因不定期检修,造成算法系统紊乱的现象出现,继而导致存在诸多安全隐患的不合格智能产品流通于现实生活之中。最终人类对于智能技术产品的需求与渴望并不会那么强烈,挫败国家对人工智能技术的战略支持。

有鉴于此,笔者认为可以单独设置生产、销售不符合安全标准的人工智能产品罪。考虑到人工智能产品在社会实践中已经存在取代人类活动的可移动智能机器人,因而,该种安全标准不应当以符合保障使用者的人体健康、财产健康为标准,而应当以保障不特定多数人的人身、财产安全利益保障为基础标准。除此之外,此类犯罪应当不同于刑法第 146 条生产、销售不符合安全标准的产品罪的实害犯的模式设置,而采取“准抽象危险犯”的模式,对其“足以”造成的后续产品使用和管理阶段的危害结果,无需附加具体危险犯的现实、具体、紧迫的现实危险。因为,在制造和销售过程中,瑕疵智能产品的危害结果产生具有一定的间隔性和潜伏期,对于明显不符合人工智能产品的国家标准和行业标准的,应当被认定为人工智能犯罪的一种抽象危害。

3. 使用环节:产品的使用和管理阶段。在人工智能产品使用和管理阶段,针对人工智能刑事风险的行为规制,应当单独设置非法滥用科技产品罪和过失使用科技产品罪。其中,责任主体分别为科技产品的使用者、管理者和再利用者;[52]危害行为模式为科技产品的不正当使用、处置与随意再利用,故意或过失造成危害结果的发生,同时亦可以单独设置单位犯罪。在初始系统无任何风险漏洞的情况下,人工智能使用者、管理者和再利用者应当审慎按照合法、合规的

使用方法进行操作,如果人为故意损坏或者过失损害人工智能深度学习系统功能,最终导致发生严重的危害结果,则应当将这一结果评价为"人为续造风险"。而对于具体智能刑事风险的类型评价应当结合行为时的心态,具体适用非法滥用科技产品罪或过失使用科技产品罪。但是,人工智能产品使用和再利用阶段,刑事风险的理性评断在于审慎评价智能编程的后续失律与失范,按照行为人可能熟知的系统操作程序进行认定。对于使用过程中,智能系统自主的功能损耗,致使的严重危害结果应当予以现实排除。[53]

五、结　语

"网络社会技术治理与法律二元共治"的基本模式已经开始渐入部分学者的研究视角,[54]而刑法立法的技术性思维模式之专门性研究,目前为止仍鲜有学者涉足。人工智能时代的来临,意味着科技发展水平已经不仅仅局限于互联网技术平台的区域研究。大数据分析的集合化、模块化以及自主化,能够促使人工智能的法益侵害呈现出网络刑法治理轨道的背离趋势。人工智能"人机交互性"应当被视为人工智能犯罪体系独立化的一种论证依据,"人机交互性"表明算法歧视数据的形成和输入模式、危害行为操作过程以及法益侵害结果的发展,完全能够形成人机一体化。因而,刑法立法应当摆正人工智能犯罪立法模式的塑造方向,明确"科技犯罪"作为一种上位概念,取代网络犯罪这一"大杂烩"体系。在科技犯罪的体系背景下,采取类型化思维,以软件技术的运用和开发程度为中心,以科技发展的时间背景为轴线,可以形成"计算机系统犯罪—信息网络犯罪—人工智能犯罪"三足鼎立的刑法规制模式。相应地,国家对人工智能技术开发战略的支持,包括制度规范的顶层设计都要求我们在未来人工智能犯罪立法体系中,避免将人工智能犯罪纳入网络犯罪框架之中,以实现科技犯罪的技术性理念融合。

注释:

〔1〕如部分学者认为:"发展到一定阶段的智能机器人可能超越程序的设计和编制范围,按照自主的意识和意志实施犯罪行为,因而完全可能成为行为主体而承担刑事责任。"参见刘宪权、胡荷佳:《论人工智能时代智能机器人的刑事责任能力》,《法学》2018 年第 1 期。类似论述还可见刘宪权:《人工智能时代的"内忧""外患"与刑事责任》,《东方法学》2018 年第 1 期。

〔2〕储陈城:《人工智能可否成为刑事责任主体》,《检察日报》2018 年 4 月 19 日,第 003 版。

〔3〕叶良芳:《科技发展、治理挑战与刑法变革》,《法律科学(西北政法大学学报)》2018 年第 1 期。

〔4〕〔日〕山口厚:《刑法总论》,付立庆译,北京:中国人民大学出版社,2018 年,第 195 页。

〔5〕〔英〕芬巴尔·利夫西:《后全球化时代:世界制造与全球化的未来》,王吉美、房博博译,北京:中信出版集团,2018 年,第 101 页。

〔6〕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》,北京:中信出版集团,2017 年,第 104 页。

〔7〕〔32〕〔49〕徐汉明、张新平:《网络社会治理的法治模式》,《中国社会科学》2018 年第 2 期。

〔8〕对于网络犯罪的场所类型,有学者将网络虚拟空间和现实空间并称为一种"双层社会空间"。参

见于志刚:《“双层社会”中传统刑法的适用空间——以“两高”〈网络诽谤解释〉的发布为背景》,《法学》2013 年第 10 期;而有的学者则认为网络虚拟空间中的各种错综复杂的社会关系,本质上是现实领域各种关系的反映、延伸与表达。参见徐汉明、张新平:《网络社会治理的法治模式》,《中国社会科学》2018 年第 2 期。就笔者而言,本文更倾向于第一种观点。

〔9〕陈伟、熊波:《网络谣言型涉众事件:传播机理、罪罚及调整》,《西南民族大学学报(人文社会科学版)》2018 年第 4 期。

〔10〕同种观点还可参见郭旨龙:《“双层社会”背景下的“场域”变迁与刑法应对》,《中国人民公安大学学报(社会科学版)》2016 年第 4 期。

〔11〕此种类型定位划分,亦可详见下图表 1 中的分析。

〔12〕算法系统歧视主要在于数据分析环节的结果偏差和数据结构的部分性缺失,“因为这样的数据模拟过程经常会由于某一个环节的错误导致最终模拟结果的偏差甚至是错误,而这是与人类通常认知世界的思维规律相违背的——人类的思维过程是思辨的、批判性的,所以大数据溯因思维通常并不在数据完备的情况下进行推理”。参见刘伟伟、原建勇:《人工智能难题的大数据思维进路》,《新疆师范大学学报(哲学社会科学版)》2018 年第 2 期。

〔13〕对此,罗克辛教授认为:“从一开始,当行为人采取减小对被害人已经存在的危险,即以改善行为客体状况的方式,对一种因果过程进行修改时,风险创设及其可归责性就不存在了。”如果人类在确保人工智能技术研发和使用阶段无过错的前提下,都无法控制风险的发展过程,更何况谈及因果进程的修改问题。参见〔德〕罗克辛:《德国刑法学总论(第 1 卷)》,王世洲译,北京:法律出版社,2005 年,第 247 页。

〔14〕宋亚辉:《风险控制的部门法思路及其超越》,《中国社会科学》2017 年第 10 期。

〔15〕〔英〕卡鲁姆·蔡斯:《人工智能革命:超级智能时代的人类命运》,张尧然译,北京:机械工业出版社,2017 年,第 45 页。

〔16〕马荣春:《刑法类型化思维的概念与边界》,《政治与法律》2014 年第 1 期。

〔17〕例如,义务来源承担的主体类型、手段类型或是标准类型。

〔18〕Trevor N. White, Seth D. Baum, “Liability for Present and Future Robotics Technology”, in Patrick Lin, Ryan Jenkins, and Keith Abney eds., *Robotics Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, Cambridge: Oxford University Press, 2017, pp. 66 – 79.

〔19〕梁根林:《传统犯罪网络化:归责障碍、刑法应对与教义限缩》,《法学》2017 年第 2 期。

〔20〕〔日〕前田雅英:《刑法总论讲义》,曾文科译,北京:北京大学出版社,2018 年,第 35 页。

〔21〕当然,也有学者将形式犯认为是行为犯;而危险犯又依据“危险发生”,细分为抽象危险犯和具体危险犯。参见张明楷:《刑法学》,北京:法律出版社,2016 年,第 166 页。亦有学者将危险本身视为实害结果类型之一,如山口厚教授认为:“仅有行为就能成立,其实这样的犯罪是结果与行为同时或者几乎是同时发生的。其与其他犯罪的区别仅仅在于,其他的犯罪是结果的发生和行为之间存在时间的、场所的间隔而已。参见〔日〕山口厚:《刑法总论》,付立庆译,北京:中国人民大学出版社,2018 年,第 45 页。但是考虑到具体概念界定并不影响本文对现实结果的类型化分析,亦不是本文的分析重点,对此笔者将不单独予以详细阐述。

〔22〕〔47〕〔48〕刘宪权:《人工智能时代刑事责任与刑罚体系的重构》,《政治与法律》2018 年第 3 期。

〔23〕吴尚赟:《注意规范保护目的理论的本土化展开》,《政法论丛》2018 年第 2 期。

〔24〕陈洪兵:《准抽象危险犯概念之提倡》,《法学研究》2015 年第 5 期。

〔25〕比如,破坏交通工具罪,破坏交通设施罪,生产、销售不符合安全标准的食品罪,生产、销售不符合标准的医用器材罪以及非法采集、供应血液、制作、供应血液制品罪。

〔26〕该学者还进一步认为,德国刑法上的利用计算机系统的犯罪,如诈骗罪,包括类型有:不当编排应用程序、使用不当或不完全的数据、无权使用数据、其他对系统进行的无权操作行为。德国刑法中的计算机犯罪实际上已经偏离了我国刑事立法上的计算机系统犯罪。参见王钢:《德国判例刑法(分则)》,北京:北京大学出版社,2016 年,第 233 – 237 页。

〔27〕譬如,刑法第二百八十六条中破坏计算机信息系统罪的计算机信息系统功能“删除、修改、增加、干扰”行为。

〔28〕〔45〕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。

〔29〕陈家林:《外国刑法:基础理论与研究动向》,武汉:华中科技大学出版社,2017年,第16页。

〔30〕徐子沛:《数据之巅:大数据革命、历史、现实与未来》,北京:中信出版社,2014年,第281-282页。

〔31〕赵春玉:《罪刑法定的路径选择与方法保障——以刑法中的类型思维为中心》,《现代法学》2014年第3期。

〔33〕王志远:《论我国刑法各罪设定上的“过度类型化”》,《法学评论》2018年第2期。

〔34〕〔美〕道格拉斯·胡萨克:《刑法哲学》,姜敏译,北京:中国法制出版社,2015年,第11页。

〔35〕车浩:《刑事立法的法教义学反思——基于〈刑法修正案(九)〉的分析》,《法学》2015年第10期。

〔36〕孙道萃:《非法利用信息网络罪的适用疑难与教义学表述》,《浙江工商大学学报》2018年第1期。

〔37〕参见浙江省宁波市海曙区人民检察院甬海检公诉刑诉(2015)235号起诉书;浙江省宁波市海曙区人民法院(2015)甬海刑初字第258号刑事判决书。

〔38〕张明楷:《网络时代的刑事立法》,《法律科学(西北政法大学学报)》2017年第3期。

〔39〕最高人民法院刑事审判第三庭:《网络犯罪司法实务研究及相关司法解释理解与适用》,北京:人民法院出版社,2014年,第3页。

〔40〕于志刚、于冲:《网络犯罪的裁判经验与学理思辨》,北京:中国法制出版社,2013年,第23页。

〔41〕于志刚:《网络犯罪与中国刑法应对》,《中国社会科学》2010年第3期。

〔42〕刘韩:《人工智能简史》,北京:人民邮电出版社,2018年,第65页。

〔43〕譬如,智能科技界有的学者认为人工智能存在“弱人工智能、强人工智能和超人工智能阶段”;参见李开复:《人工智能》,文化发展出版社,2017年,第115页;See Sean Semmler, Zeeve Rose, “Artificial Intelligence: Ap-plication Today and Implications Tomorrow”, *Duke Law & Technology Review*, Vol. 16, No. 1 (Winter, 2017-2018), pp. 85-99。而又有学者认为仅有“弱人工智能和强人工智能”的技术阶段之差异,See Rosendo Abellera, Lakshman Bulusu, *Oracle Business Intelligence with Machine Learning*, California: Apress, 2018, p. 3。

〔44〕程啸:《论大数据时代的个人数据权利》,《中国社会科学》2018年第3期。

〔46〕刑法第一百二十五条的非法制造危险物质罪中的“危险物质”仅限于毒害性、放射性、传染病病原体以及与上述物质存在相当属性的物质,人工智能产品研发过程中故意、过失制造的系统漏洞无法被涵盖在其中。无独有偶,刑法第一百三十六条的危险物品肇事罪同样存在前述问题。而投放危险物质罪则由于投放行为种类的问题,致使“投放”无法包容人工智能风险的制造行为。

〔50〕刘宪权、朱彦:《人工智能时代对传统刑法理论的挑战》,《上海政法学院学报(法治论丛)》2018年第2期。

〔51〕牟怡:《传播的进化:人工智能将如何重塑人类的交流》,北京:清华大学出版社,2017年,第5页。

〔52〕此处的“再使用行为”是指,行为人明知或者应当知道人工智能产品的合理使用期限,在超过规定的使用期限范围的情况下,行为人仍然继续使用,并且不加以施加任何防范措施(如再次接受技术风险评估和产品定期检修),而放任危害结果发生的行为。

〔53〕参见〔德〕冈特·施特拉腾韦特、洛塔尔·库伦:《刑法总论I——犯罪论》,杨萌译,北京:法律出版社,2006年,第261、405页。当然对于人工智能产品的自主性功能损耗,使用者和管理者也具有排查义务,但是脱离现有的技术去认定刑事法层面的危害结果,则增加了行为人认识不法程度的危险情况。

〔54〕例如,郑智航教授认为:随着网络技术发展的日新月异,网络社会的法律治理关注技术性理念,是出于对“并行化治理、吸收化治理和多利益攸关方治理三种基本共治模式”的考虑。参见郑智航:《网络社会法律治理与技术治理的二元共治》,《中国法学》2018年第2期。

〔责任编辑:刘姝媛〕