论人工智能缺陷产品生产者的刑事责任

黄陈辰

摘要：在生产环节，人工智能缺陷产品生产者主要承担故意责任，即构成生产不符合安全标准的产品罪。但目前尚未形成人工智能产品安全标准体系，需要进行增补；同时，算法黑箱导致因果判断失灵，应结合刑法理论、先进技术与相关政策予以破解。在流通环节，人工智能缺陷产品生产者主要承担不作为责任，其作为义务的来源，不仅限于利用信息优势而取得的对事故发生因果进程的间接支配，还包括通过主服务器、信息网络实现的对流通领域中人工智能缺陷产品的直接掌控。另外，作为义务的内容增加了远程禁用与修复，作为可能性也得到相应的提升。在产品缺陷是由于当前技术条件下无法发现，或者是由于介入因素所导致，或者在生产者发出警示公告后，使用者不听劝阻、继续使用所导致的情况下，人工智能缺陷产品生产者的刑事责任得以阻却。

关键词：人工智能；产品刑事责任；生产不符合安全标准的产品罪；算法黑箱；作为义务

DOI：10.19836/j.cnki.37-1100/c.2020.06.006

人工智能产品①与传统产品不同，因此人工智能缺陷产品生产者②的刑事责任的判断与传统缺陷产品生产者相比亦具有明显区别。例如算法黑箱导致因果关系的认定更加困难、信息网络技术使得生产者对流通领域内的产品具有更高的支配力等。这些区别导致传统缺陷产品生产者刑事责任的相关结论无法直接适用，必须加以调整。而这也正是本文研究的核心，即人工智能缺陷产品生产者的刑事责任应如何认定与承担。

一、人工智能缺陷产品生产者在生产环节的故意责任

根据我国《产品质量法》第46条的规定，产品缺陷一般是指产品存在危及人身、他人财产安全的不合理的危险。但若有保障人体健康和人身、财产安全的国家标准、行业标准的，产品缺陷则是指不符合该标准，例如自动驾驶汽车的雷达传感器失灵，或者手术机器人的三维成像系统故障等。对于缺陷产品生产者在生产环节的故意责任，我国《刑法》第146条“生产、销售不符合安全标准的产品罪”已有规定。需要注意的是，人工智能产品虽具有智能性、自主性等特征，但在弱人工智能时代，因其尚未产生自主意识，故仍属于刑法上所称的“产品”③。因此，若人工智能产品生产者主观上存有故意，明知其所生产的人工智能产品未能达到国家标准或行业标准而仍然生产，最终造成严重后果，则其行为亦构成生产不符合安全标准的产品罪。对于缺陷产品生产者在生产环节的过失责任，我国《刑法》并未

收稿日期：2020-03-17

作者简介：黄陈辰，中国政法大学刑事司法学院博士研究生（北京 100088；592712007@qq.com）。

① 人工智能产品是指运用大数据、机器学习、云计算、语音与图像识别等人工智能技术的产品，例如具有自动驾驶功能的交通工具、手术机器人、智能金融代理等。另外，部分学者根据自主意识的有无，提出弱人工智能与强人工智能的划分。但由于目前尚未产生且在可预见的相当长时间内不会产生所谓的强人工智能，因此本文仅在弱人工智能视角下进行研究。

② 本文所述“生产者”，指的是包括研发、设计、制造等各部门在内的集合体，并不仅限于进行组装、制造的部门。

③ 虽然人工智能系统多表现为一种“无形”的软件，但本文所探讨的人工智能产品整体表现为软件与硬件的“合集”，相关系统被内置于物质载体之上，例如自动驾驶系统的物质载体是自动驾驶汽车，因此人工智能产品能够被认定为刑法所称的“产品”。参见 Lori A. Weber, “Bad Bytes: The Application of Strict Products Liability to Computer Software”, *St. John's Law Review*, 1992, 66(2), pp. 469-485.

规定。按照通说观点,作为单位的生产者不承担过失责任,但根据《关于〈中华人民共和国刑法〉第三十条的解释》,可以对组织、策划、实施该危害社会行为的人依法追究刑事责任。由此可知,人工智能缺陷产品生产者在生产环节主要承担故意责任,且与传统缺陷产品生产者的故意责任相同,但其具体认定需要注意以下两方面的问题:

(一)人工智能产品安全标准的欠缺与增补

根据我国《刑法》第146条的规定,生产不符合安全标准的产品罪的构成要件之一为"产品不符合国家标准、行业标准"。因此,相关产品的国家标准与行业标准成为判断生产者行为构罪与否的重要依据。

1.人工智能产品安全标准体系尚未形成

当前,根据我国《标准化法》《标准化法实施条例》,相关部门针对传统产品,制定了种类繁多、细致翔实的国家标准、行业标准,可以说已经形成了完备的安全标准体系。但对于人工智能产品,仅在汽车、航空、航天等人工智能技术运用较早的领域制定了少量标准,例如《QJ 2634-1994 自动驾驶仪仿真试验方法》《HB 8435-2014 民用飞机飞行控制计算机系统通用规范》等,而其他领域则仍属于真空状态。这样一来,当某人工智能缺陷产品造成严重损害时,仅在少数具有国家标准、行业标准的情况下,能够以该标准为依据判断生产者是否构成生产不符合安全标准的产品罪。而其他大多数情况,则只能根据《产品质量法》第46条的规定,以是否存在不合理的危险分析产品有无缺陷,进而判断生产者是否构成生产假冒伪劣产品罪。这样的规制模式看似也能追究所有人工智能缺陷产品生产者的刑事责任,但后罪中"不合理的危险"这一标准过于模糊,不利于缺陷产品的认定与生产者责任的判断。另外,其仅依据销售金额定罪处罚,导致生产者对其缺陷产品造成的损害后果并未承担相应的刑事责任。

2.增补针对人工智能产品的安全标准

产品安全标准等规范性文件的内容往往是针对已经发生的社会现实,故其制定具有一定的滞后性。但科技的不断发展,尤其是从弱人工智能到强人工智能再到超人工智能的进程,既不可避免,也无法叫停。若不能及时更新、完善相关规范,则会造成规制上的空白与漏洞。因此,相关部门应紧跟时代发展,保持对科技进步的敏感性与前瞻性,及时增补针对人工智能产品的国家标准、行业标准,以构建人工智能产品安全标准体系。

在具体增补时,应根据各领域内人工智能产品的发展现状,优先针对人工智能技术运用相对成熟且与人身、财产安全密切相关的人工智能产品制定国家标准、行业标准,例如自动驾驶汽车、手术机器人等,之后再逐渐补充与完善其他领域内的安全标准。另外需要注意的是,与传统产品相比,人工智能产品不仅具有物质载体,而且装配有各种人工智能技术系统,后者的缺陷同样可能造成人身、财产损失的严重后果。例如2016年5月,一辆特斯拉自动驾驶汽车在佛罗里达州高速公路上行驶时,由于自动驾驶系统未能有效地从明亮的天空背景中识别出横穿马路的白色拖车,导致发生碰撞并致乘车人死亡(以下简称"特斯拉案")①。因此,在针对人工智能产品制定安全标准时,不仅要对其物质载体进行规范,而且也应关注其智能系统。完备的人工智能产品国家标准与行业标准应将生产者的安全保障义务具体化,这有利于规范其在生产过程中的行为,敦促其确保相关产品的质量,同时也为生产者怠于履行义务时追究其刑事责任提供明确的衡量尺度和判断标准②。

(二)因果关系认定的困境与破解

1."算法黑箱"导致因果判断失灵

根据合法则性条件理论,在传统的产品刑事责任中,应以普遍的自然法则为前提,利用条件公式

① Gareth Corfield, "Tesla Death Smash Probe: Neither Driver Nor Autopilot Saw the Truck", https://www.theregister.com/AMP/2017/06/20/tesla_death_crash_accident_report_ntsb/, 访问日期:2020年1月20日。

② Alfred R. Cowger Jr., "Liability Considerations When Autonomous Vehicles Choose the Accident Victim", *Journal of High Technology Law*, 2018, 19(1), pp.1-60.

判断生产者生产缺陷产品的行为与损害后果之间的因果关系。而由于传统产品的运行原理具有可解释性等特征，相关的自然法则较为明确，因此其因果关系易于被我们发现与证明。但人工智能产品的运行所遵循的是程序与算法，我们所能看到的只有表面上指令的输入与结果的输出，而中间环节则不得而知，整个过程呈现出一种“端对端”的模式，形成我们无法洞悉的“隐层”，又被称为“黑箱”①，即作为外部观察者，我们无法确知人工智能产品内部的具体运算过程，而只能看到其最终的结果反馈，并且即使其试图向我们解释，我们也无法真正理解②。例如在“特斯拉案”中，能被我们感知到的外化现象是驾驶员开启了自动驾驶模式，而特斯拉汽车却未能及时刹车或采取其他措施避免碰撞，至于其自动驾驶系统内部具体的运算过程，即汽车为何会做出如此决策我们不得而知③。正是由于人工智能技术存在算法黑箱，导致人工智能产品的运行原理不具有可解释性，因此其缺陷与损害结果间的关系在现行知识体系下未能得到充分的论证，进而使得在这种情况下不存在明确的普遍自然法则，合法则性条件理论无法适用，传统因果判断在此失灵。

2.“理论＋技术＋政策”破解“算法黑箱”困境

机器学习是指利用大数据对计算机进行训练，使其从原始数据出发，自动学习和生成高级的认知结果。因此，在以机器学习为基础的人工智能技术中，不可避免地会产生“算法黑箱”或“算法隐层”。相关学者据此认为应承认“算法黑箱”的合理性，一切想要将其透明化的措施都是无用且无益的④。但刑事责任的判断与归属不能建立在无法确定的模糊事实之上，因此应对“算法黑箱”困境予以破解，以明确归因乃至归责的事实基础。

(1)理论层面。为解决食药品、环境等公害犯罪领域内因某些科学法则缺位而导致的归因困难问题，日本学者提出了疫学因果关系理论，即虽然某因素与相关疾病之间的关系无法在医学、流行病学等学科知识上得到证明，但根据统计的大量观察，只要认为二者之间的联系具有高度盖然性，则能够肯定因果关系的存在⑤。疫学因果关系理论并不要求存在明确的自然法则，其遵循“高度存疑即罚”的理念，有利于解决公害犯罪的归因问题，同时也为“算法黑箱”困境的破解提供了重要借鉴，即当由于“算法黑箱”的不可解释性导致无法证明人工智能产品缺陷与损害后果之间的因果关系时，若根据现有技术可知该产品缺陷引起损害后果的概率达到高度盖然性的程度，则仍可以认定二者之间存在因果关系。例如在上述“特斯拉案”中，虽然自动驾驶系统的运行过程不可知，但倘若其识别系统缺陷导致车辆在很大程度上无法辨别颜色相近的物体，则可以认定其缺陷与致人死亡后果之间的因果关系，进而追究特斯拉公司作为生产者的法律责任。

(2)技术层面。虽然在当前技术条件下，我们无法破解“算法黑箱”，但随着科技的进一步发展，相信我们能够在技术层面实现突破。另外，既然“算法黑箱”问题已为研究者们所注意，那么在人工智能技术的下一阶段研发中，即应向着具有可解释性的方向发展：其一，在人工智能产品中安装数据记录器(即俗称的“黑匣子”)，全自动、不间断、无差别地记录下人工智能产品运行过程中的全部数据信息。当发生产品事故时，可以通过调取、还原、分析这些数据实现人工智能产品的运行可视化，进而判断因果关系的存在与否。其二，搭建自动测试分析系统对“算法黑箱”进行解读，即利用算法辅助理解算法，在此方面已有学者进行了初步尝试⑥。其三，在开发、设计人工智能系统与算法之初，即在不影响

① 许可：《人工智能的算法黑箱与数据正义》，《社会科学报》2018年3月29日，第06版。

② Davide Castelvecchi, “The Black Box of AI”, *Nature*, 2016, 538, pp. 20-23.

③ Nynke E. Vellinga, “From the Testing to the Deployment of Self-driving Cars: Legal Challenges to Policymakers on the Road Ahead”, *Computer Law & Security Review*, 2017, 33(6), pp. 847-863.

④ Davide Castelvecchi, “The Black Box of AI”, *Nature*, 2016, 538, pp. 20-23.

⑤ 大塚仁：《刑法概说(总论)》，冯军译，北京：中国人民大学出版社，2003年，第167-168页。

⑥ 张吉豫：《打破人工智能算法黑箱》，见华宇元典法律人工智能研究院编：《让法律人读懂人工智能》，北京：法律出版社，2019年，第380页。

运算效率的前提下加入具有可追溯性与可解释性的模块,使其支持交互分析,实现运算过程的可还原、可理解。以手术机器人为例,有学者提出可以通过"生成模型化"(generative modeling)技术使手术机器人将自己的运算过程生成相应的图像,以便于程序员对其进行观察、分析与判断①。

(3)政策层面。技术的发展并不是随心所欲的,其必须在一定的规范框架之下进行,故各国人工智能技术发展的具体方向与样态,很大程度上取决于各国在该领域内的政策导向。因此,要破解"算法黑箱"困境,则需要在政策层面对人工智能技术的可解释性提出要求。目前,部分国家已经开始探索并做出了一定的回应,值得我们借鉴。例如,美国电气和电子工程师协会(IEEE)发布的《人工智能及自动化系统伦理设计白皮书》中,提出了人工智能及自动化系统设计的五项基本原则,其中第四项原则为"透明性原则",即要求一个特定的自主和智能系统的决策基础应该始终是可发现的,强调了人工智能算法的可发现性②;美国计算机协会(USACM)发布了关于算法透明和可责性的七项原则,其中第四项原则为"可解释性原则",鼓励使用算法决策的组织和机构对算法所遵循的程序和所做出的具体决策进行解释③。

二、人工智能缺陷产品生产者在流通环节的不作为责任

产品一经销售,即移转为消费者占有,生产者随即丧失对其的物理控制。但若该产品具有危及人身、财产安全的缺陷,生产者需要及时采取公告、召回等措施,否则将承担相应的法律责任,这一点在民法上已经明确④,在刑事责任领域也有所探讨。理论界对缺陷产品不召回行为⑤刑事责任的研究,主要集中在处理模式的选择上。目前共有三种模式,即将不召回行为与生产行为进行一体评价的故意作为犯模式、将作为义务消解于注意义务之中的过失犯模式,以及不真正不作为犯模式。其中,不真正不作为犯模式更加符合不召回行为的本质,且其能够有效填补前两者处罚范围有限与难以解释积极义务的弊端,因此属于相对合理的选择。需要注意的是,人工智能产品亦属于刑法所称的"产品",故生产者不召回人工智能缺陷产品的行为同样构成不真正不作为犯。但与传统产品相比,人工智能产品能够依托大数据进行深度学习与自主决策,智能化程度更高,因此生产者不作为责任的具体认定具有其特殊性。

(一)作为义务的来源与内容

1.作为义务的来源

生产者承担不作为责任的最重要前提是存在作为义务,在不召回缺陷产品情形中则为召回义务。理论界对召回义务的来源具有不同观点,主要有安全确保义务说、支配领域说、先行行为说等⑥。本文考虑到生产者在信息方面的优势地位、缺陷产品可能危害的对象范围、继任生产者的责任承担等因素,采取对因果进程的排他支配说,即由于生产者对损害后果发生的因果进程处于排他性支配地位,因此其具有防止损害发生的作为义务。

当然,产品一经销售即在物理上脱离生产者的控制,例如生产厂商将汽车销售之后,汽车的占有即移转给购买者,生产厂商不再对汽车具有实际的支配权。因此,在传统缺陷产品场合,只能将生产

① Eliza Strickland, "Making Medical AI Trustworthy", *IEEE Spectrum*, 2018, 55(8), pp.8-9.

② "Ethically Aligned Design", https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf. 访问日期:2020年1月20日。

③ Simson Garfinkel, Jeanna Matthews, Stuart S. Shapiro, and Jonathan M. Smith, "Toward Algorithmic Transparency and Accountability", *Communications of the ACM*, 2017, 60(9), p.5.

④ 参见《侵权责任法》第46条、《消费者权益保护法》第19条、《食品安全法》第63条等。

⑤ 为方便表述,本文将生产者不及时消除缺陷产品所造成的法益侵害危险的行为统称为"不召回行为"。

⑥ 吕英杰:《风险社会中的产品刑事责任》,《法律科学》2011年第6期。

者对因果进程的排他支配限定在对“产品处于流通领域并随时可能引发事故”这一危险状态的支配，即生产者利用自己对产品质量的信息优势，通过左右消费者对缺陷产品危险性的认知来控制其是否继续使用该缺陷产品，进而达到对损害结果发生之因果进程的支配①。产品购买即是为了使用，但绝大多数的产品缺陷都不易被消费者发现。因此，若生产者不告知其该产品存在缺陷，消费者便会继续使用，而产品事故都是在这种“无知”状态下发生的。但对于人工智能产品而言，其缺陷主要在于系统漏洞，而人工智能系统虽内置于相关产品，但仍未完全脱离研发端的总服务器，生产者可以通过该服务器对其生产的所有产品中的人工智能系统进行调试与更新。因此可以说，其对流通领域内的产品以及产品事故发生的因果进程具有更高的支配力，并且这种支配是直接的、具有物质载体的，相较于传统产品而言，排他性效果更强。例如，自动驾驶汽车的生产者能够远程对自动驾驶系统进行操作，且其技术与算法具有保密性，因此不用借助于信息的不对称，生产者可直接对因系统缺陷导致的产品事故具有排他性支配。

2.作为义务的内容

对于传统缺陷产品而言，生产者的作为义务即召回义务的内容，主要有对消费者提出警示与公告、及时召回市场上的全部缺陷产品、对缺陷产品进行修理等。但由于人工智能缺陷产品具有智能性等特征，其生产者的召回义务在上述内容的基础上还应增加以下两项：

其一，生产者可以通过主服务器禁用存在缺陷的人工智能系统。例如自动驾驶汽车的自动驾驶系统存在漏洞，则生产者可以直接在后台进行操作，禁止装配有该系统的汽车使用自动驾驶模式，而只能由人类驾驶员亲自驾驶。这样的措施与传统被动的警示、公告相比，具有主动性，生产者能够及时停止缺陷系统的使用，防止事故的继续发生。若生产者在得知系统存在缺陷之后未能及时禁用该系统，则应对扩大的损害后果承担不作为责任②。

其二，生产者可以在禁用有缺陷的人工智能系统后，通过主服务器直接远程对其进行漏洞的填补与修复，而无需将产品全部召回，仅在需要对硬件设备进行修理或更新时，再召回相应部分的缺陷产品。这样不仅能够提高修复效率，而且也极大地缩减了召回成本。例如在上述“特斯拉案”发生以后，特斯拉公司对该事故车辆上的信息进行了收集、还原与分析，利用这些分析结果，特斯拉公司升级了其自动驾驶系统并远程发送给所有装配有该系统的特斯拉汽车。现在，每辆特斯拉汽车都有一个改进的算法来更好地辨别颜色相近的物体，例如明亮天空背景下的白色拖车③。同时，基于人工智能产品预置知识的有限性以及持续更新的可行性，笔者认为，即使在未发现产品存在缺陷的情况下，若相关技术升级，生产者亦负有对人工智能系统进行更新的义务，否则应对由此产生的损害后果承担不作为责任④。

另外值得注意的一个问题是，人工智能缺陷产品生产者是否具有数据共享的作为义务？即在生产者发现某一系统存在漏洞之后，是否需要将其获取的数据信息向其他同类生产者乃至全社会公开？因为一方面，根据因果进程排他支配说的观点，生产者因其信息优势而对事故发生的因果进程具有排他性的支配力，因此存在相应的作为义务；但另一方面，对于原始设备制造商来说，数据信息意味着核心技术，数据共享不利于商业秘密的保护，同时也有碍于形成公平、创新的市场环境。对于这一问题，目前尚未有定论，但各生产商均拒绝数据共享。例如上述“特斯拉案”发生后，特斯拉公司并未公开原始数据及改进措施；又如2016年，谷歌公司的一辆自动驾驶汽车发生交通事故，事后谷歌公司没有公

① 黎宏、常康爽：《缺陷产品不召回的刑事责任》，《上海政法学院学报》2018年第5期。

② 牛天宝：《通过现有规范解决自动驾驶汽车肇事之刑事责任归属问题》，《法学杂志》2020年第3期。

③ Jesse Krompier, “Safety First: The Case for Mandatory Data Sharing as a Federal Safety Standard for Self-Driving Cars”, *Journal of Law, Technology & Policy*, 2017(2), pp.439-468.

④ Damien A. Riehl, “Car Minus Driver: Autonomous Vehicles Driving Regulation, Liability, and Policy”, *The Computer & Internet Lawyer*, 2018, 35(5), pp.1-18.

布事故数据,而是仅宣称"我们升级了软件"①。

(二)作为的可能性

要认定人工智能缺陷产品生产者的不作为责任,除具有作为义务以外,还需要具备作为的可能性,即在客观上能够有所作为。若由于不可抗力等因素导致生产者无法采取相应措施,则其不对损害后果承担不作为责任。人工智能产品的特殊属性,使得其生产者在召回措施的作为可能性上与传统产品生产者有所不同:其一,生产者能够利用无线技术对各人工智能系统进行实时监控,因此无论产品的流通范围多广,也无论是否有其他渠道获取信息,一旦系统出现故障,生产者能在第一时间发现并采取相应措施。其二,传统产品出厂销售以后,由于经手各经销商、分销商等中间环节,生产者很难了解产品的具体去向,且某些产品的市场覆盖面广,甚至远销海外,故召回的成本高、难度大,即使最基本的警示与公告等措施,也很难保证所有购买者都能得知。但基于通信手段与互联网技术的发展,对于人工智能产品,生产者只需在总服务器上进行操作,即可将产品故障信息准确送达每一位产品使用者,并且还可以对故障系统实施远程禁用与修复。因此,人工智能缺陷产品生产者具有更高程度的作为可能性。

三、人工智能缺陷产品生产者刑事责任的阻却因素

并非所有因人工智能产品缺陷导致的损害后果均应归属于生产者。在下列情形中,生产者的刑事责任受到阻却:

(一)当前技术无法发现的缺陷

人工智能的核心在于大数据、机器学习与云计算,但这三项技术目前仍处于发展阶段,尚未完全成熟,还存在许多不可知的技术盲点。因此,建基于这些技术的人工智能产品可能具有相应的缺陷,但这些缺陷是在当前技术水平下无法发现的,故生产者不具有主观明知,亦没有预见可能性。所以,由于该缺陷导致发生产品事故并造成人身、财产的损害后果,生产者不承担刑事责任②。并且,之所以允许可能存在当前技术无法发现的缺陷的人工智能产品投入使用,是基于技术发展的需要。任何创新均具有一定的风险,但考虑到其所能带来的巨大社会效益,并确保其在当前技术条件下不存在缺陷,其潜在的未知风险是被允许的。另外需要注意的是,当前无法发现的产品缺陷可能随着科技的进步而能为生产者所知。因此,虽然其无须对之前由此缺陷导致的损害后果承担责任,但从此时起,其具有防止损害进一步扩大的作为义务,若其未能及时采取相关措施,则需要对之后的损害后果承担不作为责任。

(二)介入因素导致的缺陷

导致损害后果的产品缺陷除了由生产者的生产行为引起外,还有可能归因于某些介入因素。在此情况下,生产者的刑事责任得以阻却,转而由介入因素主体承担。此处所说的介入因素有以下两类:

其一,使用者违规修改系统。以自动驾驶汽车为例,生产者在研发自动驾驶系统时,除部分个性化设置(例如自动驾驶的开启与关闭、定速巡航的具体速度等)外,均采取标准化设计模式,以确保产品性能并有利于对系统进行统一维护、更新与升级。但某些使用者在使用自动驾驶汽车时,为满足自身特殊的需求,擅自对自动驾驶系统进行修改,例如删除人工接管提示功能、修改车载高精度地图等。

① Jesse Krompier, "Safety First: The Case for Mandatory Data Sharing as a Federal Safety Standard for Self-Driving Cars", *Journal of Law, Technology & Policy*, 2017(2), pp. 439-468.

② Gary C. Robb, "A Practical Approach to Use of State of the Art Evidence in Strict Products Liability Cases", *Northwestern University Law Review*, 1982, 77(1), pp. 1-33.

若由于使用者对自动驾驶系统的修改导致发生交通事故，进而产生损害后果，则生产者不承担刑事责任①。

其二，犯罪人非法侵入系统。人工智能技术的发展在便利社会生活的同时，也为犯罪分子实施新型犯罪提供了条件。当前与人工智能有关的犯罪中，除利用人工智能技术作为犯罪手段外，还有另一种类型，即以人工智能系统为犯罪对象，犯罪人非法侵入人工智能系统并进行相应的操作，以实现犯罪目的。例如电影《速度与激情》中的“僵尸汽车”情节，即是黑客利用无线网络侵入自动驾驶汽车的操作系统并阻断人工接管功能，使汽车只能在自动驾驶模式下运行。在这种情况下，自动驾驶系统的漏洞并非由生产者的生产行为导致，而应归咎于黑客的侵入与修改行为，因此生产者无须对损害后果承担刑事责任。但需要注意的是，生产者在开发自动驾驶系统时应添加相应的安全模块以防止黑客入侵，若生产者怠于采取上述措施进而导致黑客侵入自动驾驶系统，则生产者需要承担相应的刑事责任。

（三）生产者发出警示公告后，使用者不听劝阻继续使用

在生产者发现其产品存在安全缺陷并采取相应补救措施后，例如向消费者发出警示公告、禁用相关人工智能系统、对缺陷产品实施召回，若使用者明知该产品存在缺陷而仍然继续使用，则根据自担风险的原理，由使用者承担由此带来的损害后果，生产者的责任随即得以阻却②。例如某汽车生产商在技术更新时发现，其所生产的自动驾驶汽车中的自动驾驶系统存在安全漏洞，随即通过总服务器远程禁用全部自动驾驶系统并向消费者发出警示公告。若车主某甲利用其计算机技术擅自解除该系统的禁用并继续使用自动驾驶模式，最终发生交通事故，则由某甲自己对损害后果承担责任，生产者因已尽作为义务，因此无须担责。

另外，若产品存在缺陷的同时伴有使用者的操作不当，二者共同导致事故发生，则虽不能完全阻却，但应适当减轻生产者的刑事责任。例如自动驾驶汽车的警示功能存在故障，未能在需要驾驶员接管车辆操控时做出警示提醒，但驾驶员亦未遵守该汽车对其每隔几秒需触碰一次方向盘的明确要求③，而是在车上睡觉，因此没有意识到风险并及时采取措施，最终导致事故发生。在这种情况下，生产者因产品缺陷而需要对损害后果承担刑事责任，但由于介入了驾驶员的疏忽因素，因此应适当减轻责任。

四、结语

当前，人工智能技术广泛运用于交通、医疗等领域，与人们的日常生活息息相关，若其存在缺陷，会对社会公众的人身、财产安全产生巨大的威胁。而生产者编写并控制着人工智能系统的程序与算法，这在很大程度上决定了产品缺陷的存在与否，因此为其设置合理的刑事责任，有利于促使其推进算法的更新与升级，提高人工智能产品的安全性，同时也能够在其怠于履行上述职责时，为追究其法律责任提供明确的依据④。人工智能产品虽然智能化程度更高，但就其本质而言，仍属于刑法所称的“产品”，因此传统缺陷产品生产者的刑事责任大体上能够直接适用，但在具体认定上有所不同：首先，在生产环节的故意责任中，目前尚未形成人工智能产品安全标准体系，需要进行增补。同时，算法黑

① Jeffrey K. Gurney, “Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles”, *Journal of Law, Technology & Policy*, 2013(2), pp. 247-277.

② David G. Owen, “Products Liability: User Misconduct Defenses”, *South Carolina Law Review*, 2000, 52(1), pp. 1-80.

③ David Goldstein, “Autonomous Vehicles Will Drive Themselves-But They Won't Regulate Themselves”, *Hastings Business Law Journal*, 2017, 13(2), pp. 241-256.

④ Jeffrey K. Gurney, “Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles”, *Journal of Law, Technology & Policy*, 2013(2), pp. 247-277.

箱导致因果判断失灵,应结合刑法理论、先进技术与相关政策予以破解。其次,在流通环节的不作为责任中,生产者作为义务的来源不仅限于利用信息优势而取得对事故发生因果进程的间接支配,还包括通过主服务器与信息网络而实现的对流通领域中人工智能缺陷产品的直接掌控。另外,作为义务的内容增加了远程禁用与修复,作为可能性也得到相应的提升。同时需要注意的是,当产品缺陷在当前技术条件下无法发现,或者由于介入因素导致缺陷发生,或者在生产者发出警示公告后,使用者不听劝阻继续使用的情况下,人工智能缺陷产品生产者的刑事责任得以阻却。

On Producers' Criminal Liability of Defective Products for Artificial Intelligence

Huang Chenchen

(Criminal Justice College, China University of Political Science and Law, Beijing 100088, P. R. China)

Abstract: When it comes to the production process, the producers of artificial intelligence defective products mainly bear the intentional liability, which constitutes the crime of producing products that do not conform to safety standards, but at present, the artificial intelligence product safety standard system has not been formed, which needs to be updated; at the same time, the algorithm black box leads to the failure of causal judgment, which should be solved by combining the theory of criminal law, advanced technology and relevant policies. In the inaction responsibility of the circulation link, the source of producers' act obligation is not limited to the indirect domination of the causal process of accident occurrence obtained by using the advantage of information, but also includes direct control of defective products in the field of convection through the master server and information network; in addition, as the content of obligations, remote bans and repairs have been added, and the possibility of act is also improved. In the case of product defects that cannot be found under current technical conditions, or are caused by intervention factors, or when the user does not listen to dissuasion and continues to use after a warning notice is issued by the producer, the criminal liability of producers of artificial intelligence defective products can be prevented.

Keywords: Artificial intelligence; Product criminal liability; Crime of producing products that do not conform to safety standards; Algorithm black box; Act obligation

[责任编辑:李春明]