

生成式人工智能所涉特定犯罪的认定路径

王德政

(西南石油大学 法学院,四川 成都 610500)

摘要:当生成式人工智能服务使用者利用生成式人工智能生成的信息实施犯罪行为,生成式人工智能服务提供者为该犯罪的行为主体。《生成式人工智能服务管理暂行办法》对提供者规定了法定义务,提供者能履行但不履行该义务以致使用者的行为构成犯罪,必须检验提供者的行为是否构成不纯正不作为犯。只有具备形式和实质性归责条件才能将使用者行为实现的特定构成要件要素归责于提供者。提供者的提供行为一般可构成帮助信息网络犯罪活动罪。

关键词:生成式人工智能;ChatGPT;DeepSeek;提供者;使用者

中图分类号:D914

文献标志码:A

文章编号:1672-7835(2025)03-0095-10

一、问题的提出

当前,全球正在进行以人工智能为标志的第四次工业革命^①。人工智能可划分为决策型人工智能和生成式人工智能^②。就生成式人工智能而言,目前较为前沿的进展是,美国人工智能公司 OpenAI 于 2022 年 11 月推出了 ChatGPT,迄今其最先进的版本是 GPT-4.0。除了“文心一言”“通义千问”“混元助手”“360 智脑”以外,当前我国自主研发并最受关注的生成式人工智能产品为 DeepSeek。DeepSeek 的推理能力可与 ChatGPT 媲美,但与之不同的是,其采取开源模型权重和推理代码,其使用者更为广泛。

面对生成式人工智能雨后春笋般的勃兴,我国在规范层面迅速作出了回应。2023 年 7 月 10 日,国家网信办等部门出台了《生成式人工智能服务管理暂行办法》(以下简称《办法》)。2025 年的《政府工作报告》指出,应持续推进“人工智能+”行动,大力发展智能网联新能源汽车。生成式人工智能除了引发行行政管理方面的问题以外,若生成式人工智能生成的信息违反刑法,则将触及犯罪。有学者将生成式人工智能涉及的犯罪归纳为侵犯知识产权类犯罪、煽动类犯罪等^③。但单纯地对上述犯罪的罪名种类逐一进行归纳并非问题的核心,也欠缺必要性。这是因为,生成式人工智能所涉犯罪的底层逻辑在于,使用者利用其生成的信息实施相关犯罪,这些犯罪均无法脱离信息本身而独立被实施。在《中华人民共和国刑法》(以下简称《刑法》)设定的 483 个罪名中,任何可通过信息传播而实施的犯罪均可通过生成式人工智能实施,包括编造、传播虚假信息罪,编造、传播虚假恐怖信息罪,侵犯公民个人信息罪等,该信息传播媒介性即为这些特定罪名之间的特定关联性。既然上述罪名均可成为生成式人工智能涉及的犯罪,单纯地归纳罪名可能产生挂一漏万的弊端。可见,罪名种类的确定应当结合生成式人工智能具体应用场景予以展开,既无必要也不可能事先被归纳完备。因此,在对生成式人工智能涉及的刑法问题进行研究时,应当回避对具体罪名的单纯归纳这一欠缺必要性的既定选题和研究范式,转而重视一个更为重大的问题——对这些犯罪在宏观上成立与否提出统一的认定路径。该新路径最终落实到具体罪名种

收稿日期:2025-01-15

基金项目:司法部法治建设与法学理论研究部级科研项目(21SFB4036)

作者简介:王德政(1985—),男,四川通江人,博士,讲师,主要从事数字法学研究。

①刘宪权:《生成式人工智能对数据法益刑法保护体系的影响》,《中国刑事法杂志》2023 年第 4 期。

②胡裕岭,姚浩亮:《生成式人工智能的数据治理与“囚笼”——基于行业主体自我监管的回应型研究》,《征信》2024 年第 2 期。

③刘宪权:《ChatGPT 等生成式人工智能的刑事责任问题研究》,《现代法学》2023 年第 4 期。

类的确定,能有效解决上述归纳法所涉罪名认定问题。

二、行为主体的界定

生成式人工智能所涉犯罪首先涉及行为主体的界定。根据刑法通说,行为人的行为构成犯罪必须通过某罪犯罪论体系的检验。犯罪论体系分为构成要件该当性、违法性和有责性三个阶层。构成要件要素包含行为主体、实行行为、行为客体等。生成式人工智能所涉犯罪无论属于何种实行行为,该实行行为的实施必有相应的行为主体。但由于生成式人工智能本身具备特殊性,其涉及的犯罪不同于普通的自然人和单位实施的犯罪行为,如何界定其行为主体存在理论争议。一般认为,生成式人工智能的行为主体涉及技术研发者、技术持有者、数据提供者、数据所有者、技术使用者、技术监管者^①。在这些主体中何种主体属于生成式人工智能所涉犯罪的行为主体应予以逐一讨论。

(一) 相关人员主体性的确定

技术研发者无法独立成为生成式人工智能所涉犯罪的行为主体,理由如下:一方面,技术研发者仅为生成式人工智能企业所聘用的众多技术人员中的部分成员,所起的作用是为该企业研发生成式人工智能产品提供部分技术支撑。因此,技术研发者对生成式人工智能产品的研发和应用仅起到了部分作用,而非使该产品成功上市并受到应用的全部作用,无法独立对生成式人工智能所涉犯罪的不法内涵提供完整的来源,以及无法在相关犯罪的因果流程上独立提供支撑。另一方面,技术研发者属于生成式人工智能产品负责研发核心技术的成员,掌握对国家意义重大的前沿科学技术,如果成为行为主体而承担刑事责任,将间接地损害我国科技、经济和综合国力。

数据提供者和数据所有者亦无法成为生成式人工智能所涉犯罪的行为主体,理由如下:其一,生成式人工智能固然以数据、算法、算力为核心要素^②,生成信息的过程无法脱离数据而独立运作,但该过程在技术上呈现复杂的流程,分为预训练、微调、强化学习三个阶段^③,数据仅为上述流程中的一环而非全部。其二,数据经历从原始面貌到最后生成这一过程,本身很可能已“面目全非”而无法单独被视为后者本身。如果将作为原始材料的数据视为生成的信息,将因不符合客观性而导致数据提供者和数据所有者承担本不应承担的刑事责任,从而违反刑法中的罪责自负原则。其三,生成的信息可能涉及海量数据,而海量数据可能涉及众多提供者和所有者,如果将该主体一概视为行为主体以致其承担刑事责任,将导致刑事责任极度扩散而违反刑法的最后手段原则,同时将因为无视各个数据提供者和所有者的数据在生成的信息中所体现的不同作用,导致违反刑法中的罪刑相当原则。技术使用者作为注册生成式人工智能为其服务的人,利用所生成的信息实施犯罪行为,当然可成为人工智能所涉特定犯罪的行为主体,但其行为是否构成相关犯罪可根据我国《刑法》和刑法理论进行判断,本身并非疑难理论问题,因而不在于本文的讨论范围之内。至于技术持有者和技术监管者是否可以成为生成式人工智能所涉犯罪的行为主体,首先需要界定二者的具体含义。人工智能企业负责生成式人工智能产品的投资、组织研发和销售(或提供),作为企业对被开发的技术拥有知识产权,自然属于技术持有者,由此与《办法》中的生成式人工智能服务提供者(以下简称提供者)同义。由于《办法》第十四条规定了提供者负有信息内容监管和违法活动阻止义务,提供者与技术监管者同义。可见,由于技术持有者和技术监管者均为提供者,提供者可否成为行为主体成为关键所在。

提供者属于生成式人工智能所涉特定犯罪的行为主体,理由如下:一方面,在使用者利用生成的信息实施犯罪行为的场合,被害人的法益被侵害呈现一条因果链:提供者为用户提供生成式人工智能产品→生成式人工智能产品在人机交互中生成信息→使用者利用该信息实施犯罪行为→被害人的法益被侵害。可见,提供者作为因果链的初始开启者,经历了生成式人工智能产品和使用者共同参与其中的因

①袁曾:《生成式人工智能责任规制的法律问题研究》,《法学杂志》2023年第4期。

②陈兵、董思琰:《生成式人工智能的算法风险及治理基点》,《学习与实践》2023年第10期。

③郑曦、朱溯蓉:《生成式人工智能的法律风险与规制》,《长白学刊》2023年第6期。

果流程,最终导致法益被侵害。法益是指刑法所保护的对社会而言明显有价值的生活利益^①,刑法的任务在于保护法益。一切刑法规范中蕴含积极的价值判断均以生活利益为基础,这些利益对公民而言是不可或缺的,由此必然受到国家强制力以刑罚所进行的保护^②。因此,刑法分则中的每个罪名都有其特定的保护法益。而行为人实施的行为之所以构成犯罪,正是因其侵害了法益。法益在法秩序尤其是刑法中的重要角色决定了作为侵害法益“始作俑者”的提供者应当回溯成为行为主体。另一方面,尽管生成式人工智能产品在犯罪实施的过程中所起的作用是生成信息,但由于本身无法成为行为主体而独立承担刑事责任(理由详见下文),无法以此为由将提供者排除在行为主体以外。同时,尽管使用者在信息生成前的人机交互以及信息生成后的犯罪实施中均扮演着一个参与型角色,但无法据此否定提供者提供生成式人工智能产品的前期作用,二者的作用彼此独立而无法互相否定。

(二)生成式人工智能主体性的排除

对于生成式人工智能本身可否成为行为主体,我国法学界存在肯定说和否定说两种理论。肯定说赞同生成式人工智能可成为行为主体而承担刑事责任^③。立论根据在于,生成式人工智能是同时受自然法则和道德法则支配的物种,从而有自由意志^④。但否定说持相反立场,认为生成式人工智能仅为智能化程度更高的弱人工智能,没有独立意识,更无法承担责任能力,属于难以脱离人类独立运行的主体,因而无法成为犯罪主体^⑤。

毫无疑问,生成式人工智能涉及犯罪的场景完全不同于自然人在具备认识的前提及在意志的操控下独立自主地支配身体实施犯罪行为。尤其是生成式人工智能在身体方面与人类存在显著差异,将前者等同于后者而置于行为主体的地位,除非在我国《刑法》中如同单位犯罪那样有明确的法律依据,否则必须在理论上予以澄清,前者何以在“精神”或“意识”范畴上类似于后者,并且这种类似性能否达到将二者在刑法上等同视之的程度。对此涉及两个疑问需要解答。

第一,生成式人工智能在认识、意志、理性、情感、伦理方面是否等同于人类?首先,可以肯定的是,尽管生成式人工智能在人机对话中可能使人类对话者感觉似乎与“人”在进行沟通,但这种感觉并不等同于生成式人工智能确实具有人类独有的认识、意志、理性等,必须有充分的客观证据予以证明。但迄今为止,除了上述感觉以外,并无客观证据可予以证明。尽管人工智能基于深度学习具备一种自我完善的能力,但该能力不足以证明其具备人类的所有思维特质,而“无法独立地理解和建构世界”^⑥。其次,人类具有的认识、意志等思维特质决定了其从事活动具有主动性,但生成式人工智能生成信息对此恰好具有被动性。如果人类不启动信息生成及对话程序,生成式人工智能则无法独立自主地决定与人类进行对话及控制其进程,否则将被人类认为是计算机感染了病毒程序以致生成式人工智能程序失控。最后,生成式人工智能生成的信息不仅对投喂的数据具备来源上的依附性,还对这些数据具有极大程度的组合性,其创新性程度亦明显低于人类的思考结果。生成式人工智能是一种以对话为形式的撒播技术,文本的输出有赖于被投喂的数据集^⑦。而人类的大脑即使没有持续的数据输入,也能通过思考独立获得结果,并且该思考结果并非完全对大脑中已获知信息的组合,而可能具有超强的创新性,这一点在文学和艺术创作、科学研究等方面体现得淋漓尽致,并在最新的科研发现中获得证实——如果生成式人工智能使用合成数据(其自身生成的数据)进行模型训练,将导致模型崩溃^⑧。由此可知,尽管生成式人工

①Jürgen Baumann, Ulrich Weber, Wolfgang Mitsch. *Strafrecht Allgemeiner Teil Lehrbuch*. Bielefeld: Verlag Ernst und Werner Gieseking, 2016, p. 5.

②Hans-Heinrich Jescheck, Thomas Weigend. *Lehrbuch des Strafrechts Allgemeiner Teil*. Berlin: Duncker & Humblot, 1996, p. 7.

③刘宪权:《生成式人工智能对数据法益刑法保护体系的影响》,《中国刑事法杂志》2023年第4期。

④刘宪权:《生成式人工智能的发展与刑事责任能力的生成》,《法学论坛》2024年第2期。

⑤侯跃伟:《生成式人工智能的刑事风险与前瞻治理》,《河北法学》2024年第2期。

⑥郭研,沙涛:《人工智能刑事责任能力之否定》,《学术交流》2022年第1期。

⑦张文祥,沈天健,孙熙遥:《从失序到再序:生成式人工智能下的信息秩序变局与治理》,《新闻界》2023年第10期。

⑧Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, et al. “AI Models Collapse when Trained on Recursively Generated Data”, *Nature*, 2024(631): 755-759.

智能的智能程度相较以前的人工智能产品大为提高,但仍不具备真正比拟人类的思维特质。

第二,生成式人工智能是否具备辨认、控制自己行为的能力?肯定说对该问题的解答存在以下弊端:一方面,导致犯罪论体系中不同的阶层在内容和检验顺序上产生错乱。具体而言,犯罪论体系的第三个阶层是有责性,其包含刑事责任能力等要素^①。刑事责任能力的核心是行为人辨认、控制自己行为的能力,年龄和精神均可影响该能力存在与否及其程度高低。换言之,辨认、控制自己行为的能力属于罪责范畴,而如前文所述,行为主体属于构成要件该当性范畴,通过肯定生成式人工智能具备辨认、控制自己行为的能力来认定其可成为行为主体,明显是以罪责替代构成要件的审查,造成二者在内容上发生混淆,并颠覆了三阶层的审查顺序。另一方面,在事实层面,生成式人工智能并不因其较强的深度学习能力而具备辨认和控制自己行为的能力。辨认和控制自己行为的能力在刑法理论中只能在自然人中存在,其底层逻辑在于只有自然人才具有人类特有的认识和意志。生成式人工智能仅为取材于数据、由算法主导、由算力支撑从而欠缺精神或意识的程序,对客观世界并无最基本的认识能力。此外,控制自己行为的能力属于刑事责任能力的核心要素,而刑事责任能力属于罪责的要素而与自由意志紧密相连^②,但作为生成式人工智能核心技术的深度学习不等同于自由意志,所以没有控制能力可言^③。

通过上述论证可知,生成式人工智能所涉犯罪的行为主体既非生成式人工智能本身亦非其他自然人,而是提供者和使用者,但在本文语境中特指提供者。

三、犯罪认定路径

当使用者利用生成式人工智能生成的信息实施犯罪行为,提供者作为所涉犯罪的行为主体,其行为是否构成犯罪,如果构成又为何种犯罪,在刑法理论上有着深入研究的空间。

(一) 犯罪类型的初步界定

在此首先涉及在犯罪类型上对提供者的实行行为进行界定。犯罪类型属于在刑法上对实行行为进行的分类,可依据实行行为违反的规范种类划分为作为犯和不作为犯。该划分在人工智能、互联网所涉犯罪中尤为关键。这是因为,人工智能和互联网的提供者并非使用者这一直接实施犯罪行为的直接主体而是间接主体,其实施的是提供服务行为,不同于使用者直接实施的犯罪行为本身,该提供服务行为应被认定作为犯(如帮助信息网络犯罪活动罪)抑或不作为犯系关键所在。同时,不作为犯可划分为纯正不作为犯和不纯正不作为犯,前者在我国《刑法》分则中有具体罪名(如拒不履行信息网络安全管理义务罪),但后者欠缺法条依据而主要依靠刑法理论解决认定问题,因而前者相较于后者在理论上的讨论难度较低。综合考量理论难度和行文逻辑两方面,可先讨论提供者的行为是否构成不纯正不作为犯,再讨论含作为犯和纯正不作为犯在内的相关犯罪是否成立。

根据刑法通说,不纯正不作为犯构成的罪名与作为犯一致,但区别在于构成要件^④。不纯正不作为犯的客观构成要件中有保证人地位这一特殊要素^⑤。保证人地位的存在决定了行为人具有作为义务,继而决定其行为可否构成不纯正不作为犯。提供者的行为是否构成不纯正不作为犯,取决于其是否具备防止使用者利用生成的信息实施犯罪行为的保证人地位和作为义务。我国传统刑法理论对保证人地位采用了作为义务的根据这一提法,认为其包含法律明文规定的义务、职务和业务上的义务、法律行为导致的义务、先行行为导致的义务^⑥。由于提供者并非消防员、医护人员等在职务上承担作为义务的人,故难以认为其具备职务和业务上的作为义务,同时其与使用者签订的使用协议中也不可能自我规定防止后者实施犯罪的义务,故不具备法律行为导致的义务。此外,如果不限定先行行为的范围将导致不

①Hans Kudlich. *Strafrecht Allgemeiner Teil*. München: Verlag C. H. Beck, 2007, p. 121.

②Hellmuth Mayer. *Strafrecht Allgemeiner Teil*. Stuttgart: W. Kohlhammer GmbH, 1967, p. 107.

③彭景理:《论人工智能时代智能机器人的刑事责任能力》,《大连理工大学学报(社会科学版)》2020年第2期。

④Günther Stratenwerth, Lothar Kuhlen. *Strafrecht Allgemeiner Teil*. München: Vahlen, 2011, p. 276.

⑤Rudolf Rengier. *Strafrecht Allgemeiner Teil*. München: C. H. Beck, 2016, p. 387.

⑥陈兴良:《本体刑法学》,商务印书馆2001年版,第257页。

作为犯的范围扩张,因而刑法通说对其进行引起法益侵害紧密危险和本身违反义务两个限定^①,提供者的提供行为由于难以被认定为制造了使用者侵害法益的紧密危险(使用者利用生成的信息实施犯罪行为系其自主所为),本身作为合法销售行为亦未违反“禁止对使用者销售生成式人工智能产品”的义务,故无法被视为先行行为。因此,判断提供者是否具备作为义务只能从法律明文规定的义务中寻求突破口。

在《办法》出台以前,我国现行法律体系对生成信息的提供者未设定任何监管义务,因而难以认为其具备法律明文规定义务这一作为义务。但《办法》第十四条第 1 款对提供者规定了信息内容监管义务,要求其在该情形下采取停止生成、停止传输、消除等整改措施,第 2 款对提供者规定了违法活动阻止义务,要求其在该情形下采取警示、限制功能、暂停、终止提供服务等处置措施。这针对生成式人工智能所涉犯罪的两大重要环节——生成式人工智能产品生成违法信息、使用者实施违法活动,在义务范围上实现了无死角的全覆盖。在技术层面,提供者确实有能力对信息内容采取删除等措施^②,这与《办法》所设定的监管义务相呼应。有学者认为该监管义务是《办法》的规制重点^③,但该观点忽略了使用者实施违法活动这一更为重大的环节。这是因为,即使生成式人工智能生成了违法信息,但该信息本身仅诱发了法益侵害的危险,如果不被人用以实施违法行为,则无法现实化为对国家、社会和个人法益的侵害。因此,违法活动阻止义务的设定在生成式人工智能所涉犯罪的防范中作用更为关键。

提供者的上述法定义务可否被视为法律明文规定的义务这一作为义务值得讨论。一方面,该义务中的“法律”被我国刑法通说认为包含宪法、法律、行政法规、条例、规章等^④。《办法》由于是国务院多个部委联合制定而在法律渊源上属于部门规章。因此,该义务就形式层面而言可被视为作为义务。另一方面,该义务在法益和危险源这一实质层面亦可被视为作为义务。该义务之所以要求提供者对生成的信息内容进行监管,是为了以事前防范的方式消除该信息中蕴含的侵害法益危险,以防止其转化为实害;之所以要求提供者防止使用者利用信息内容实施违法活动,是为了以事中阻止的方式阻断正在进行的法益侵害。因此,该义务由于可以保护法益免受侵害而在法益层面可视为作为义务。立法者之所以未在事前阶段为提供者设定“禁止提供生成式人工智能服务”的不作为义务(尽管该义务的设定可以彻底防范法益侵害),不仅因为提供行为本身具备合法性,还因为可以通过“事前许可、事中阻止”的制度设计实现国家支持、推动人工智能产业的研发和应用这一良好愿景。

此外,该义务的设定在危险源的监控这一层面与作为义务的根据具备同质性。在我国刑法学中,作为义务的根据在概念的内涵上对应着德国刑法学中的保证人地位。在此需要参考德国刑法学中关于保证人地位的理论。德国刑法学界早先从形式上将保证人地位划分为法律、合同、先行行为、紧密生活共同体^⑤,但阿明·考夫曼采取一种从功能和实质上进行区分的方法^⑥,将其划分为两大类型:一是监控型保证人地位。该类型涉及对特定危险的管控,保证人义务的对象是监控危险以防止其现实化为结果。二是保护型保证人地位。该类型涉及对特定法益的保护,使其免除被侵害的危险^⑦。前者具体包含血缘(家庭)关系、紧密共同体关系、自愿(事实上)承担保护或帮助义务、公务员或法人机关地位,后者具体包括违反义务的先行行为、对危险源的监控义务、投放市场的危险产品、管护义务^⑧。该二分说为当代德国刑法通说所赞同^⑨。

① Urs Kindhäuser. *Strafgesetzbuch Lehr- und Praxiskommentar*. Baden-Baden: Nomos, 2010, p. 141.

② 徐伟:《论生成式人工智能服务提供者的法律地位及其责任——以 ChatGPT 为例》,《法律科学(西北政法大學學報)》2023 年第 4 期。

③ 漆晨航:《生成式人工智能的虚假信息风险特征及其治理路径》,《情报理论与实践》2024 年第 3 期。

④ 高铭喧,马克昌:《刑法学》,北京大学出版社、高等教育出版社 2017 年版,第 68 页。

⑤ Volker Krey, Robert Esser. *Deutsches Strafrecht Allgemeiner Teil*. Stuttgart: Kohlhammer, 2016, p. 489.

⑥ Reinhart Maurach, Karl Heinz Gössel, Heinz Zipf. *Strafrecht Allgemeiner Teil*. Heidelberg: C. F. Müller, 2014, p. 324.

⑦ Urs Kindhäuser. *Strafrecht Allgemeiner Teil*. Baden-Baden: Nomos, 2015, p. 313.

⑧ Bernd Heinrich. *Strafrecht Allgemeiner Teil*. Stuttgart: Kohlhammer, 2016, p. 387.

⑨ Günther Jakobs. *Strafrecht Allgemeiner Teil*. Berlin: Walter de Gruyter, 1991, p. 799.

二分说中对危险源的监控这一保证人地位与提供者的法定义务在本质上是一致的。从本质上讲,作为与不作为的区别表现为是否行为人有犯罪能量的投入^①,因而不纯正不作为犯的行为人并未积极地制造法益侵害的危险^②,而是在具备保证人地位的前提下能阻止但不阻止结果发生。在结果发生的过程中,可能选择置身事外的行为人面临法益侵害的危险,该危险的引起必然存在着一个来源,此即危险源。可见,法益和危险源必然存在于行为人实施不作为的现实过程中。在生成式人工智能所涉犯罪的场合,使用者利用生成式人工智能生成违法信息,再利用该违法信息实施犯罪行为以致法益被侵害,在此过程中存在两种危险源:一是直接危险源——生成式人工智能本身,二是间接危险源——使用者。立法者要求提供者对危险源进行监控必然意味着:一方面,要对生成式人工智能本身进行监控,具体而言,要对可能被直接用于实施犯罪行为的信息内容进行监管并根据具体情况采取整改措施;另一方面,要对使用者进行监控,具体而言,要对使用者利用生成的信息内容实施犯罪行为实施监控并根据具体情况采取处置措施。由此可见,提供者法定义务的根据实质上属于对危险源的监控。

综上,当使用者利用生成的信息内容实施犯罪行为时,提供者应当对使用者的犯罪行为进行监控并采取处置措施,并且该义务属于不纯正不作为犯中的作为义务,如果提供者能履行但不履行该义务以致使用者的行为构成犯罪,提供者的行为可构成相关犯罪的不纯正不作为犯。在此应注意,由于履行作为义务的能力属于不纯正不作为犯之犯罪论体系中的重要因素,必须在个案中审查提供者实际上能否履行该义务。该不纯正不作为犯属于正犯还是帮助犯应根据个案中的具体情况予以判断。在区分正犯与共犯的理论中,犯罪支配说系刑法通说^③。该说可用于不纯正不作为犯领域,其判断标准为:如果不作为者支配了作为者所实施的不作为,其行为系正犯,相反则为帮助犯^④。在生成式人工智能所涉犯罪的场合,这种支配性具体体现为:如果提供者阻止使用者实施犯罪行为,则使用者的犯罪行为无法既遂。在个案中,如果提供者能采取处置措施阻止使用者实施犯罪行为以致其无法既遂,其不作为系不纯正不作为犯的正犯,否则为帮助犯。

上述结论仅属于在犯罪类型意义上对提供者所实施行为的初步界定,还需要进一步解决的问题是:其一,如何将使用者实现的构成要件在形式上归责于提供者?其二,在上述归责过程中,考虑到生成式人工智能产业在我国科技发展中的重大性,是否需要附加实质性归责条件?下文将逐一为上述问题提供解决方案。

(二)不纯正不作为犯的归责方法

1.既定理论的漏洞

从表面上看,根据刑法通说,只要行为人有能力履行其承担的作为义务而不履行,造成犯罪结果发生的,该不作为则构成不纯正不作为犯。上述论断看似简单明了,但在生成式人工智能场合存在以下理论漏洞。

第一,构成要件的任务在于对不法行为进行类型化,而在构成要件中分类的任务主要由实行行为承担^⑤。但提供者实施的不作为并不等同于在实行行为种类上可以类型化的侵犯公民个人信息行为、侮辱行为等,如何对不作为本身确定实行行为的种类?刑法通说对此采取肯定的做法,在实行行为的种类上直接以使用者的行为来确定提供者的不作为^⑥。但问题在于,既然提供者的不作为与使用者的作为在犯罪类型上并不一致,何以将二者等同视之?刑法通说对此并未提出理论根据。就理论根据而言,这本质上属于归责理论在不纯正不作为犯中的运用。换言之,不纯正不作为犯的可罚性决定了作为义务

①Klaus Hoffmann-Holland. *Strafrecht Allgemeiner Teil*. Frankfurt: Mohr Siebeck, 2015, p. 280.

②Helmut Frister. *Strafrecht Allgemeiner Teil*. München: C. H. Beck, 2015, p. 293.

③Claus Roxin. *Täterschaft und Tatherrschaft*. Berlin: De Gruyter, 2006, p. 458.

④Thomas Hillenkamp. 32 *Probleme aus dem Strafrecht Allgemeiner Teil*. München: Luchterhand, 2006, p. 179.

⑤Heinrich Wilhelm Laufhütte, Ruth Rissing-van Saan, Klaus Tiedemann. *Strafgesetzbuch Leipziger Kommentar Erster Band*. Berlin: De Gruyter, 2007, p. 729.

⑥Edmund Mezger. *Strafrecht*. Berlin: Verlag von Duncker & Humblot, 1949, p. 137.

的设定是为了保护法益^①,所以提供者被设定法定义务的终极目的在于保护法益,但使用者通过实施犯罪行为侵害了法益,在法益侵害这一联结点可将使用者实施的行为归责于提供者,提供者的不作为在实行行为的种类上遂等同于使用者实施的行为^②。

第二,即使可将使用者实施的行为归责于提供者,但使用者的行为所实现的其他构成要件要素是否也可以归责于提供者?前文已述,构成要件包含实行行为等多种要素,何种构成要件要素可归责于提供者应逐一检讨。该问题相较于前述问题更为复杂,并且在生成式人工智能所涉犯罪的场合更具实务意义,但刑法通说对此欠缺解决方案。比如,使用者使用 ChatGPT 生成淫秽视频后将其出售给他人以获取经济利益,提供者对此放任不管。由于传播淫秽物品牟利罪属于目的犯^③,其构成要件中有牟利的目的这一主观构成要件要素,但作为提供者的 OpenAI 公司即使对使用者实施的传播淫秽视频行为,在事实上没有也不可能存在通过传播淫秽视频牟利的目的,该目的仅存在于使用者的主观意识中,在此情形下,能否将 OpenAI 公司的不作为认定为传播淫秽物品牟利罪?

第三,将构成要件的归责过程具体化和细致化以后,如果据此在形式上直接认定提供者的不作为成立不纯正不作为犯,而在入罪过程中不附加实质性归责条件,将导致提供者由于其行为轻易地入罪而承担刑事责任,最终将变相地产生压制、阻碍生成式人工智能产业发展的不利后果。比如,只有 1 个用户利用 ChatGPT 生成淫秽视频并无偿公开传播,OpenAI 公司对此放任不管。如果直接认定 OpenAI 公司负责人的行为构成传播淫秽物品罪,国内外生成式人工智能企业负责人极可能都会受到震慑,这对生成式人工智能的研发显然将造成不利的影响。正是因为生成式人工智能的研发和应用对全社会具有极大的正面效应,在对提供者的归责过程中则需要附加实质性归责条件,通过提升入罪难度来限缩刑罚的范围,以免对生成式人工智能产业的发展造成萎缩效果,但如何确定实质性归责条件的具体内容值得研讨。

2. 形式归责

就形式层面而言,有必要逐一讨论使用者的行为实现的构成要件中何种要素可归责于提供者。前文已述,构成要件包含主观构成要件和客观构成要件。前者包含故意、过失、动机、目的,后者包含实行行为、行为主体(含特殊身份)、行为客体、结果、因果关系、行为情状(时间、地点等)^④。主观构成要件中的故意、过失须根据提供者本身是否具备该要素而独立被检验。这是因为,故意、过失本身影响的是行为无价值(规范违反、法敌对意思)而非结果无价值(法益侵害)^⑤,提供者被设定法定义务的终极目的在于保护法益而非单纯地导正使用者的内心事实和操作规程,因此,将无法体现结果无价值的故意、过失归责于提供者将违反法定义务的设定目的。客观构成要件中的实行行为、行为客体、结果、因果关系和行为情状由于均与法益侵害密切相关而无须独立被检验,在提供者具备法定义务的前提下,应将上述要素直接归责于提供者。在此应注意,由于提供者的法定义务一般是在使用者着手实施犯罪行为以后才实际上被履行,这种时间上的滞后性存在一定的消极性,因而使用者应当积极主动地履行不作为义务(不实施犯罪行为的义务),以不作为义务的积极性消解法定义务的消极性对法益侵害的不利影响。值得进一步讨论的归责要素仅为行为主体的特殊身份、动机、目的。

行为主体的特殊身份蕴含特殊的行为无价值并由此决定了某罪的特殊不法内涵,所以应当独立被检验。如果提供者欠缺某罪的特殊身份,则无法将使用者的特殊身份归责于他。目的、动机同样具备特殊的行为无价值,因而在表面上看应当独立被检验,但当提供者不具备某罪的目的、动机时,可能由于其行为无法构成该罪而形成法益保护的漏洞。因此,折中的解决方案在于,如果提供者本身欠缺某罪的动机、目的,但明知使用者具备该动机、目的,则应将该动机、目的归责于提供者;如果提供者未认识到使用

①Thomas Frischer. *Strafgesetzbuch*. München: C. H. BECK, 2018, p. 97.

②Volker Haas. "Die Beteiligung durch Unterlassen", *Zeitschrift für internationale Strafrechtsdogmatik*, 2011, 573(5): 391-397.

③Edmund Mezger. *Moderne Wege der Strafrechtsdogmatik*. Berlin: Verlag Duncker & Humblot, 1950, p. 23.

④Hermann Blei. *Strafrecht Allgemeiner Teil*. München: C. H. Beck'sche Verlagsbuchhandlung, 1977, p. 94.

⑤陈兴良,周光权:《刑法学的现代展开》,中国人民大学出版社 2006 年版,第 272 页。

者具备某罪的动机、目的,则无法将该动机、目的归责于提供者。

3. 实质归责

前文已述,我国对生成式人工智能产业发展采取鼓励、支持的态度,仅在刑法教义学层面进行形式归责将造成入罪门槛降低,这就必须通过确立实质性归责条件来提升入罪门槛。通过将无法进行实质归责的案件认定为民事案件即进行“刑民交叉”案件的分流处理,避免刑事责任泛化而干涉生成式人工智能的合理应用,从而起到推动该产业顺利发展的效果。因此,在个案中如果具备实质性归责条件,才能将使用者的行为实现的构成要件归责于提供者,提供者的行为方可构成犯罪。

实质性入罪条件的确定应采取综合考察说,不仅应考虑被使用者用以实施犯罪行为的生成式人工智能产品本身对我国社会是否具有较大的正面作用,还应考察利用生成的信息实施犯罪行为的使用者在整体用户数量中所占的比例大小。

一方面,正面作用大小的判断包含两点:一是考察被使用者利用实施犯罪的生成式人工智能产品对当前我国社会中某个或多个行业或领域是否已发挥或很可能发挥较大的效益。如果已实际发挥或将来很有可能发挥较大的效益,比如某种生成式人工智能产品在科学研究或语言翻译方面作用较大,则无法进行归责,相反则可以进行归责;二是被使用者利用实施犯罪行为的生成式人工智能产品可否轻易地被在我国可使用的其他产品替代。如果该产品无法轻易地被我国可使用的其他人工智能产品(如 DeepSeek)替代,说明其在我国社会中扮演的功能和角色极其重要而无可取代,这就无法进行归责,相反则可以进行归责。

另一方面,利用生成的信息实施犯罪行为的使用者在整体用户中所占的数量比例可以充分反映出该生成式人工智能产品对法益的侵害程度及对我国社会的危害程度,如果比例越小,法益侵害性和社会危害性程度则越低,进行归责的必要性亦越小,相反则越大,三者之间呈正比关系。但如何确定归责所需的具体数量比例应进一步推敲。对此,有学者在判断帮助信息网络犯罪活动罪中的明知是否成立时提出超过半数规则说^①。该说在形式上的数量比例、实质上的危害性、处罚的必要性三者之间构建起一道桥梁,打通了科技与法学、实践与理论、现象与规范之间的藩篱,在生成式人工智能刑法规制领域具有极大的合理性。因而可借鉴该说确立以下归责条件:如果利用生成的信息实施犯罪行为的使用者在整体用户中所占的数量比例超过 50%,证明该生成式人工智能产品对我国社会的危害程度较大,应进行归责,否则不应进行归责。

在确定正面作用及数量比例大小时,承办相关案件的司法实践工作者应充分听取生成式人工智能领域的专家团队(应与本案没有利害关系)独立出具的专业分析意见,该意见的作出应立基于案发时该领域最权威的科学技术知识,客观、具体、富于可操作性并应遵守少数服从多数的决定规则,并以此作为是否进行实质归责的重要参考。

综上,如果使用者利用生成式人工智能生成的信息实施犯罪行为,提供者能履行但不履行法定义务以致使用者实施犯罪行为,在具备实质性归责条件的前提下才能将使用者的行为实现的特定构成要件要素在形式上归责于提供者,提供者的行为才能构成相关犯罪的不纯正不作为犯。比如,如果 OpenAI 公司将我国公民个人信息作为训练数据投喂给 ChatGPT 进行训练,有 90 人利用 ChatGPT 生成含该信息(具体为 2 000 条财产信息)的内容后,将该内容发布在网络上传播,但 OpenAI 公司明知有人实施该犯罪行为却未在技术上阻止。此时,可将侵犯公民个人信息的实行行为(违反国家规定,向他人提供公民个人信息)、行为客体(2 000 条财产信息)等客观构成要件要素归责于 OpenAI 公司,同时 OpenAI 公司独立地具备故意,这就在形式上完成了归责。在进行实质归责时,一方面,即使可以得出 ChatGPT 对当前我国社会有较大效益的结论,但只要专家团队认定其可被 DeepSeek 替代,则无法认定其对我国社会具有较大的正面作用;另一方面,由于 90 人在 ChatGPT 的整体用户中所占数量比例极低,因而进行归责的必要性极低。因此,OpenAI 公司对上述犯罪放任不管的行为由于无法通过实质归责而

^①刘杰:《ChatGPT 类生成式人工智能产品提供者之刑事责任》,《法治研究》2024 年第 2 期。

不构成侵犯公民个人信息罪。

(三) 相关犯罪的成立及犯罪竞合判断

提供者的提供行为是否成立帮助信息网络犯罪活动罪(以下简称帮信罪),其不作为是否成立拒不履行信息网络安全管理义务罪值得讨论。

1. 帮助信息网络犯罪活动罪

帮信罪是《刑法修正案(九)》增设的新罪名,其构成要件为:明知他人利用信息网络实施犯罪行为而提供互联网接入、服务器托管、网络存储、通讯传输、广告推广、支付结算等技术支持,情节严重。可见,该罪的行为主体是一般主体,包括生成式人工智能所涉犯罪的行为主体——提供者。《关于办理非法利用信息网络、帮助信息网络犯罪活动等刑事案件适用法律若干问题的解释》(以下简称《解释》)第十一、十二条对“明知”“情节严重”进行了具体规定,尤其是将“明知”的对象泛化解释为“他人”^①。

在生成式人工智能的场合,如果要认定提供者的提供行为认定为帮信罪,需要论证以下方面:首先,如何认定提供者明知其提供行为将导致使用者利用生成的信息实施犯罪行为?由于该罪中的“明知”已被具体解释为“他人可能实施犯罪行为”,这并不要求认识到使用者的具体身份,但作为深谙生成式人工智能技术特点的提供者,对于生成式人工智能产品生成违法的信息内容、该信息内容被使用者用于实施犯罪行为的可能性不可能欠缺预知。因此,在该罪的“明知”内涵被《解释》泛化的背景下,只要提供者实施提供行为,则推定其具备帮信罪的明知。其次,如何认定提供者实施了提供技术支持这一实行行为?显而易见,生成式人工智能的大规模应用发生的时间点在帮信罪被设立以后,立法者并未在刑法条文和司法解释中以“信息生成”的表述为此罪填补漏洞。但考虑到该罪的法条中对技术支持的种类有“等”字描述,在当前生成式人工智能成为最新型信息网络服务的背景下,可利用这一兜底用语将生成式人工智能服务涵摄入技术支持中,使该法条的适用在日新月异的科技迭代中具备持续的生命力。因此,提供者的提供行为应被认定为实施了提供技术支持这一实行行为。最后,认定情节严重必须适用《解释》第十二条的规定。由于提供者一般是生成式人工智能企业,其提供人工智能服务次数较多并且使用者人数众多,该条第1款第1项“为3个以上对象”、第2项“违法所得1万元以上”、第6项“被帮助者实施的犯罪行为造成严重后果”和第7项兜底条款均可轻易地被适用。因此,提供者的提供行为一般具备“情节严重”的情形。通过上述分析可知,提供者的提供行为一般可构成帮信罪。

如果提供者的不作为构成相关犯罪的不纯正不作为犯,其提供行为同时构成帮信罪,二者之间的犯罪竞合(罪数)关系值得推敲。由于提供者实施的提供行为属于作为,其与事后实施的不作为属于行为复数,由于不作为的不法内涵程度低于作为犯并且二者侵害的法益相同^②,应将提供者的整体行为认定为帮信罪。但在例外情形下,如果提供者的提供行为因不具备情节严重情形而不构成帮信罪,提供者的不作为仅构成相关犯罪的不纯正不作为犯。

2. 拒不履行信息网络安全管理义务罪

拒不履行信息网络安全管理义务罪也是《刑法修正案(九)》设立的新罪名。

该罪的构成要件为,网络服务提供者不履行法律、行政法规所规定的信息网络安全义务,经监管部门责令改正但拒不改正,致使发生违法信息大量传播等结果。《解释》第一至六条对该罪行为主体、实行行为、结果等构成要件要素进行了具体的规定。在生成式人工智能的场合,将提供者的不作为认定为该罪的障碍并不在于行为主体方面。这是因为,提供者可被解释为网络服务提供者。《解释》第一条将提供信息发布服务的单位和个人界定为网络服务提供者,尽管从表面上看,提供者并未直接对使用者发布信息,而是通过给后者提供生成式人工智能服务,由生成式人工智能直接生成信息,但生成式人工智能的研发和应用原本属于立法者在制定《解释》时未曾预料而非刻意忽略的情形,并且《解释》的制定时间离当前并不久远,因而不可罔顾目前我国最新的科技发展而机械、呆板地对“发布信息”进行解释。

^①喻海松:《帮助信息网络犯罪活动罪的司法限定与具体展开》,《国家检察官学院学报》2022年第6期。

^②柯耀程:《刑法竞合论》,中国人民大学出版社2008年版,第144页。

应采取客观解释的方法^①,将“发布信息”适当且有限地作扩大解释以包含“生成信息”,从而将提供者解释为网络服务提供者。

但问题在于,提供者的作为义务不属于该罪的信息网络安全义务。由于信息网络安全义务的来源被刑法明文规定为法律和行政法规,基于对罪刑法定原则的遵守,《办法》作为部门规章对提供者规定的法定义务由此无法成为信息网络安全义务。即使从危险源的监控这一实质性保证人地位推导出提供者具有信息内容监管和违法活动阻止义务这一作为义务,但该义务属于不纯正不作为犯而非纯正不作为犯的作为义务,同时并非当前我国法律和行政法规规定的义务,无法等同于该罪的信息网络安全义务。除非将来我国出台相关法律、行政法规,将提供者法定义务的规范来源提升至该法律渊源层级,否则提供者自始至终都不具备该罪明文规定的作为义务。因此,提供者的不作为因不符合该罪的构成要件而不构成此罪。

结论

当前,全球范围内以人工智能为标志的第四次工业革命方兴未艾,我国亦积极参与其中,旨在以新科技的力量赋能国家的综合实力并造福全体公民。但从自动驾驶汽车到生成式人工智能,不断迭代的科技浪潮在各学科领域引发了诸多难题,在刑法领域亦是如此。例如,当使用者利用生成式人工智能生成的信息实施犯罪行为,何为该犯罪的行为主体,行为主体实施的行为是否构成犯罪,如果构成又为何种犯罪?面对生成式人工智能这一人类历史上前所未有的新科技催生的上述史无前例的新问题,我国刑法应当摆脱传统理论的窠臼^②,敏锐地进行应对并提出符合我国法律体系和社会环境的新型解决方案,方可在法治层面持续推进“人工智能+”行动。

The Identification Method of the Relative Specific Crimes Involving Generative Artificial Intelligence

WANG Dezheng

(School of Law, Southwest Petroleum University, Chengdu 610500, China)

Abstract: When a user of generative artificial intelligence service uses the information generated by generative artificial intelligence to commit a criminal conduct, the user of generative artificial intelligence service are the subjects of the crime. *Interim Measures for the Administration of Generative Artificial Intelligence Services* stipulates the legal obligation for the provider, the provider can fulfill but does not fulfill the obligation so that the user's conduct constitutes a crime, the provider's conduct must be tested for the establishment of offense of non-typical omission. It is only when the formal and substantive imputation conditions are met that the specific elements of constitutive component realized by the user's conduct can be imputed to the provider. The provider's conduct of provision may generally constitute the crime of helping information network crime.

Key words: generative artificial intelligence; ChatGPT; DeepSeek; provider; user

(责任校对 张伟平)

①林山田:《刑法通论》,北京大学出版社2012年版,第85页。

②刘宪权:《生成式人工智能生成内容刑法保护的新路径》,《法学》2024年第5期。