

论通知规则在生成式人工智能侵权中的适用

徐伟

(上海政法学院上海司法研究所,上海 201701)

摘要:通知规则是归责条款,而非免责条款。这决定了通知规则可以直接适用于生成式人工智能侵权,而非类推适用。反通知在生成式人工智能侵权中的缺失不足以成为否定通知规则可适用性的充分理由。通知中“准确定位侵权内容的信息”可通过对话记录和截图的方式提供,且通知人可以不是权利人。传统通知规则中的“必要措施”在生成式人工智能侵权中发生了诸多变化,包括直接避免损害扩大的措施从处理已发生的侵权转变为预防侵权再次发生,警示性措施从警示实施了侵权行为的网络用户转变为警示生成式人工智能使用者,担保性措施无法适用、限制乃至停止提供服务的措施从针对网络用户转变为针对诱发致害内容生成的使用者。生成式人工智能服务提供者在收到通知后负有避免致害内容再次生成的义务,但该义务的边界须根据个案中的具体商业模式、人工智能技术发展状况等因素综合判断,且应有期限限制。在致害内容无法再次生成时,服务提供者负有告知义务。

关键词:生成式人工智能;通知规则;避风港;网络侵权;必要措施

中图分类号:DF523.1 文献标志码:A

DOI:10.3969/j.issn.1001-2397.2024.03.10 开放科学(资源服务)标识码(OSID):



随着生成式人工智能的广泛应用,人工智能生成损害他人合法权益的虚假内容(以下统称为“致害内容”)的情况时有发生,如何为受害人提供及时有效的救济成为生成式人工智能时代的难题。从制度供给来看,我国现行法应对网络侵权最具特色的制度是通知规则。由此产生的问题是,通知规则能否适用于生成式人工智能侵权?其应如何具体适用?相较于传统网络侵权,生成式人工智能侵权中的通知规则是否需要作出调整?笔者试图对此作出回答,以期实现通知规则在人工

收稿日期:2024-02-22

基金项目:上海市哲学社会科学规划课题“生成式人工智能服务提供者侵权责任研究”(2023BFX014)

作者简介:徐伟(1985—),男,浙江台州人,上海政法学院上海司法研究所余山学者特聘岗教授,法学博士。

智能时代的理论更新,并为实践提供助益。

一、通知规则在生成式人工智能侵权中的可适用性

生成式人工智能侵权是由人工智能自动生成致害内容,这与通知规则的典型适用场景,即网络用户利用网络服务实施侵权行为存在差异。故需厘清通知规则能否适用于生成式人工智能侵权。

(一)通知规则能否适用的理论争议

关于通知规则能否适用于生成式人工智能侵权,学界存在分歧。第一种观点认为可以直接适用,“在认定生成式人工智能虚假信息侵权责任时,仍然应当适用我国《民法典》第1195条所规定的‘通知规则’”,理由在于,“反对适用避风港规则实际上是将其作为一种直接侵权,一旦虚假信息构成诽谤等,人工智能产品的服务提供者就要承担侵权责任,这显然对其过于严苛”。^①第二种观点认为不应适用,因为“《正当通信法》(Communications Decency Act)第230条旨在保护第三方提供内容时的服务提供者,其并不保护参与内容提供的服务提供者。ChatGPT发挥的功能并非第三方提供内容时的中介,而是自行提供内容”。^②第三种观点认为虽然无法直接适用通知规则,但应当类推适用,“AI生成虚假信息的侵权行为应当类推适用‘通知规则’”。因为“AI生成侵权信息并非‘网络用户利用网络服务实施侵权行为’,而是AI自己生成的,因此在构成要件上不满足《民法典》第1195条所规定的通知删除规则”,但“由于AI服务提供者一方面难以对AI生成的信息进行实时审查,且根据用户的通知进行删除成本较低”,故应类推适用通知规则。^③由此可见,影响通知规则适用的主要因素包括生成式人工智能服务提供者的类型、责任轻重以及采取删除等措施的成本。

就理论层面而言,通知规则能否适用主要取决于两个方面:一是通知规则的法律性质,即通知规则属于免除网络服务提供者侵权责任的免责条款,还是属于认定其侵权责任的归责条款。这会决定通知规则的适用条件和法律效果。二是生成式人工智能服务提供者的法律地位,即其是属于网络内容服务提供者、搜索链接服务提供者,还是属于一种新型的网络服务提供者。对于不同类型的服务提供者,通知规则在适用时存在区别。据此,笔者将主要从以上两个方面展开分析。

(二)通知规则可适用的理论证成

1. 通知规则可适用的主要理由

厘清通知规则能否适用的前提是,明确我国通知规则的法律性质,即其应定性为归责条款抑或免责条款。尽管不乏通知规则是免责条款的主张,但笔者认为在我国法律体系下通知规则应定性为归责条款,理由如下:首先,从归责原则来看,我国网络侵权一般采用过错责任原则,这决定了在认定网络服务提供者侵权责任时,受害人应证明服务提供者存在过错,而向网络服务提供者发送合格通知正是认定其存在过错的方式之一,故通知规则是归责条款。这与美国版权侵权一般采无过错责任原则,其通知规则应定性为免责条款不同。其次,从实证法来看,尽管2006年的《信息网络

^① 王若冰:《论生成式人工智能侵权中服务提供者过错的认定——以“现有技术水平”为标准》,载《比较法研究》2023年第5期,第32页。

^② Hasala Ariyaratne, *ChatGPT and Intermediary Liability: Why Section 230 Does Not and Should Not Protect Generative Algorithms*, SSRN (May 16, 2023), <https://ssrn.com/abstract=4422583>.

^③ 参见王利明:《生成式人工智能侵权的法律应对》,载《中国应用法学》2023年第5期,第34页。

传播权保护条例》第22条等条款采取了免责条款的表述,但自原《中华人民共和国侵权责任法》第36条第2款采用归责条款的表述以来,我国通知规则在立法上便一直采用归责条款的表述。最后,采用归责条款的定性我国法律内在体系相符。既然网络用户实施侵权行为的场合可适用通知规则,那么根据“举轻以明重”的法理,在网络服务提供者直接实施侵权行为的场合更应赋予权利人通知对方移除的权利。^①

即使通知规则应定性为归责条款^②,也不应以该规则能避免网络服务提供者承担过重的责任使其可适用性正当化,因为能否适用通知规则与网络服务提供者所承担责任的轻重无关。认为通知规则能避免网络服务提供者承担较重责任的观点建立在该规则是免责条款的基础上,即网络服务提供者在遵守了通知规则后,不必对通知前后发生的损害承担责任。但事实上通知规则仅能影响通知后的责任,至于网络服务提供者对通知前发生的损害是否承担责任,则取决于其在通知前是否知道侵权内容的存在。因此,通知规则只是权利人用于证明网络服务提供者主观上存在过错的方式之一。网络服务提供者对通知前发生的损害是否承担责任并不影响该规则的适用。

影响通知规则适用的另一因素是网络服务提供者的类型。一般而言,网络自动接入服务提供者、网络自动传输服务提供者和网络自动存储服务提供者并不适用通知规则。但生成式人工智能服务提供者难以归入这三类服务提供者。存有争议的是,生成式人工智能服务提供者是否会因其被归入网络内容服务提供者而无法适用通知规则。笔者对此予以否定,因为生成式人工智能服务提供者不应定性为网络内容服务提供者。虽然生成式人工智能生成内容的表现形式相当于直接提供内容,且《生成式人工智能服务管理暂行办法》(以下简称《生成式人工智能办法》)第9条第1款规定“提供者应当依法承担网络信息内容生产者责任”,但在侵权法意义上将生成式人工智能服务提供者定性为网络内容服务提供者并不妥当,因为其无法完全控制具体的生成内容,即人工智能具有一定程度的“自主性”。^③而对内容的控制力是决定义务和责任的关键因素之一。故生成式人工智能服务提供者与网络内容服务提供者虽在形式上相似,但在本质上有所区别。同时,基于生成式人工智能服务提供者在控制力上的不足,其侵权规则应更多地参照搜索链接服务提供者。^④

退一步而言,即使是网络内容服务提供者,也可适用通知规则。因为通知规则属于归责条款,无论网络服务提供者是否要对通知前发生的损害承担责任,其在收到通知后,都应采取必要措施以避免损失的扩大。因此,不应以加害行为并非由网络用户直接实施来否定通知规则在生成式人工智能侵权中的适用。虽然《中华人民共和国民法典》(以下简称《民法典》)第1195条在规规定通知规则时提及了网络用户利用网络服务实施侵权行为,但不应对此作反面解释,即认为在加害行为并非由用户直接实施时便不能适用通知规则。

此外,决定通知规则适用的关键因素之一是网络服务提供者在收到通知后,其在技术层面能有效采取必要措施。在生成式人工智能侵权中,服务提供者确实有能力采取一定的措施,典型表现为

^① 参见徐伟:《通知移除制度的重新定性及其体系效应》,载《现代法学》2013年第1期,第65-66页。

^② 通知规则的归责条款定性意味着其是判断服务提供者过错的方式之一。服务提供者的注意义务可细分为三类:一是审查义务,该义务主要存在于生成阶段;二是影响侵权过错认定的公法上的义务,即“保护性规范”;三是采取必要措施义务,该义务存在于通知后的移除阶段。本文聚焦于第三类义务。

^③ 参见彭诚信、陈吉栋:《论人工智能体法律人格的考量要素》,载《当代法学》2019年第2期,第53-59页。

^④ 参见徐伟:《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》,载《法律科学》2023年第4期,第72-75页。

生成式人工智能服务提供者在其平台中自行表示其将遵守通知规则。例如,OpenAI“使用协议”中约定,“如果你认为你的知识产权受到侵犯,请发送通知到如下地址。我们会删除或断开涉嫌侵权的内容,并会终止重复侵权者的账户”。此外,《生成式人工智能办法》第14条等条款也对服务提供者发现违法内容后应采取的措施作出了规定,故服务提供者有能力且应当采取必要措施。综上所述,作为归责条款的通知规则可以直接适用于生成式人工智能侵权情形。

2. 否定通知规则适用的理由及其回应

否定通知规则可适用性的理由是,在生成式人工智能侵权中不存在反通知,这将导致利益失衡。通知规则的运作过程中可能出现错误通知,即向网络服务提供者发送侵权通知的“权利人”可能并非真实的权利人,或通知所涉内容可能不构成侵权(例如属于合理使用情形),而居中审查通知是否合格的网络服务提供者无论在能力还是意愿方面都可能作出错误判断,故为了避免错误通知乃至恶意通知对网络用户造成损害,通知规则须与反通知规则相配套,从而实现较好的利益平衡。^①由此可见,受通知影响的网络用户应有权通过发送合格的反通知来恢复被采取了必要措施的内容。在生成式人工智能侵权中,因并不存在可能发送反通知的网络用户,故有必要检视通知规则的适用是否会导致利益失衡。

就利益影响而言,若完全否定通知规则的适用,不仅将导致权利人失去重要的救济途径,由此获得的收益(因反通知得到保护的用户合法权益)也相对较少,毕竟实践中用户发送反通知的情形较少。^②两相权衡,不宜以不存在反通知为由否定通知规则在生成式人工智能侵权中的适用,而应通过制度安排尽可能减少因错误通知对他人合法权益和社会公共利益造成损害。为此,可取的途径是,提高生成式人工智能服务提供者对通知的审查标准,即对通知实行实质性审查。^③

关于网络服务提供者对通知的审查标准,一直存在形式审查和实质审查之争。前者是指只要通知在形式上符合法律的要求,网络服务提供者便应采取必要措施。后者则要求网络服务提供者对通知内容是否真实、侵权是否成立展开实质性判断。从司法实践来看,我国对网络服务提供者科以一定程度的实质审查义务。例如,《最高人民法院关于审理涉电子商务平台知识产权民事案件的指导意见》第5条第2款规定,“通知涉及专利权的,电子商务平台经营者可以要求知识产权权利人提交技术特征或者设计特征对比的说明、实用新型或者外观设计专利权评价报告等材料”。之所以要求权利人提供专利侵权对比表等材料,是为了便于网络服务提供者对通知所涉内容是否构成侵权加以审查。据此,我国传统网络侵权中存在要求网络服务提供者对通知内容采实质性审查的做法。在生成式人工智能侵权中也应采实质性审查标准,一般而言,只有在通知所涉内容具有较大可能构成侵权时,才需采取必要措施。如此可在一定程度上减轻反通知缺失所带来的负面影响。综上,反通知在生成式人工智能侵权中的缺失不足以成为否定通知规则可适用性的充分理由,故通知规则可适用于生成式人工智能侵权情形。

^① 参见张建华主编:《信息网络传播权保护条例释义》,中国法制出版社2006年版,第58页。

^② See Jennifer M. Urban & Laura Quilter, *Efficient Process or “Chilling Effects”? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 Santa Clara Computer and High Technology Law Journal 621, 679-680 (2006).

^③ 另一路径是提高合格通知的认定标准,尤其是“构成侵权的初步证据”的标准,例如,通知人只有提供人民法院的裁判文书才足以证明通知所涉内容构成侵权。但这一转变将导致通知规则演变为类似于诉讼的规则,从而使自身丧失了本应具备的简便、快速解决纠纷的特征,故本文不采此路径。

二、生成式人工智能侵权中合格通知的判断

权利人向生成式人工智能服务提供者发送侵权通知时,该通知需要满足哪些要件?根据《民法典》第1195条第1款第2句之规定,合格通知的要件有二:一是构成侵权的初步证据,二是权利人的真实身份信息。其中,前者往往通过权属信息和能够准确定位侵权内容的信息(例如网络地址)等加以证明;后者则须提供身份资料和联系方式等。在生成式人工智能侵权中,合格通知的判断与传统网络侵权在多数方面一致。此处仅探讨在生成式人工智能场景中会发生改变或存在分歧之处。

(一)准确定位侵权内容信息的提供

合格通知的要件之一是构成侵权的初步证据,为此权利人须提供能够准确定位侵权内容的信息。这在传统网络侵权中大多可通过提供网络地址来实现。但在生成式人工智能侵权中,人工智能是通过对话方式生成侵权内容,该对话内容并不存在相应的网络地址。故生成式人工智能带来的问题之一是,权利人如何提供能够准确定位侵权内容的信息?

在传统网络侵权中,搜索引擎自动生成检索词侵权具有一定的参考价值,因为其与生成式人工智能侵权的表现形式相似。在我国搜索引擎自动生成检索词纠纷中,定位侵权内容的方式是“检索词和截图”。例如,在“任某玉案”中,在某度搜索引擎输入“任某玉”,在页面“相关搜索”处显示有“陶氏教育任某玉”等。因“陶氏”口碑不佳,任某玉以“相关搜索”处的内容侵害其名誉权为由要求某度公司承担侵权责任。诉讼前任某玉曾向某度公司发送通知,通知中说明检索的关键词是“任某玉”,并提供搜索结果的网页截图。人民法院认为,因某度公司未侵害任某玉名誉权,故“某度公司未履行‘通知-删除’义务无需承担侵权责任”^①,即人民法院认为该通知是合格通知。据此可知提供检索词和截图是定位侵权内容的方式之一。生成式人工智能自动生成致害内容与搜索引擎自动生成检索词在表现形式上具有相似性,故在生成式人工智能侵权中,权利人也可通过提供“对话记录和截图”方式来定位侵权内容。生成式人工智能服务提供者在收到对话记录后,可以通过输入同样的对话来发现致害内容。

此处面临的问题是,对话内容有可能无法“再现”,即网络服务提供者输入同样的提问但未得到致害内容。此时服务提供者是否有义务主动采用相近的提问方式来确认致害内容是否会生成?该问题的核心是,查找侵权内容的成本主要应由权利人还是网络服务提供者承担。在权利人仅提供了侵权内容的线索时,网络服务提供者是否负有主动查找侵权内容的义务,传统司法裁判对此存在分歧。例如,在“花某参案”中,原告美国威州花某参总会在诉前向被告某宝公司发送了通知。一审人民法院认为,原告通知有侵权情况存在后,被告即应主动履行其监管义务,采取必要措施,而不能以商户的数量众多为由不予核查,消极等待原告提交具体链接地址。但二审人民法院认为,原告在向被告发出通知时并未将具体侵权链接告知被告,而在被告网站搜索花某参产品时出现的相关链接有27000个左右,故原告怠于履行告知义务。^②总体而言,判断权利人提供的定位信息是否充分的关键在于网络服务提供者查找侵权内容的成本是否合理。若成本合理,则网络服务提供者负有

^① 参见北京市海淀区人民法院(2015)海民初字第17417号民事判决书。

^② 参见吉林省高级人民法院(2012)吉民三涉终字第3号民事判决书。

查找义务。^①我国司法裁判中的这一认识与比较法经验一致。根据美国《数字千年版权法》第512(c)(A)(iii)条的规定,发现侵权内容的成本可否负担(burdensome)是法院判断通知中“指明侵权内容”是否得到满足的重要考量。^②

因此,权利人是否提供了足以定位侵权内容的信息,关键取决于网络服务提供者能否以合理的成本找到通知所涉内容。在生成式人工智能侵权中,若生成式人工智能服务提供者根据通知所提及的对话方式提问后,未能生成致害内容,则除一般理性的生成式人工智能服务提供者应当想到其他相关的提问以外,生成式人工智能服务提供者不负有主动尝试“再现”致害内容的义务,因为对其而言,超出理性人通常能想到的其他相关提问已经是不合理的成本负担。

(二)通知人可以不是权利人

在传统网络侵权中,侵权内容往往在网络上处于公开状态,不特定主体均可以接触到该内容,故通知人大多是权利人。生成式人工智能侵权的特点之一是,其生成的内容只有参与对话的使用者才能看到,而使用者未必是合法权益受侵害的权利人。故生成式人工智能侵权中可能出现通知人并非权利人的情形。若通知人并非受侵害的权利人,该通知是否合格?

通知人是否必须是权利人或权利人授权的主体(以下统称为权利人)?有学者认为只能是权利人,“从通知规则确立的目的考量,其就是要适当限制网络服务提供者的责任,因此,在解释上应当采用严格解释,而不能任意扩张。如果任何人都可以‘通知’,则网络服务提供者的义务过重,不利于网络事业的发展。更何况,网络侵权大多是对名誉、隐私等人格权的侵害,既然受害人不通知,表明受害人能够容忍此种损害,法律也没有必要再给受害人提供保护”。^③

这一理解存在诸多问题。首先,将通知规则的立法目的理解为“适当限制网络服务提供者的责任”在我国并不成立。在我国通知规则是归责条款,合格通知是证明网络服务提供者存在“过错”的方式之一,而非限制其责任。其次,任何人都可以通知,并不意味着会加重网络服务提供者的义务,因为网络服务提供者在收到一项通知并对通知所涉内容采取了删除等措施后,后续同类通知便不会再产生,因为后续通知无法再提供“构成侵权的初步证据”。最后,在受害人未通知时,“法律也没有必要再给受害人提供保护”这一判断与我国立法价值不符。我国立法确立的“知道规则”表明,即使权利人未投诉,网络服务提供者在知道或应当知道侵权内容时,仍负有采取必要措施的义务。综上,主张通知人必须是权利人的理由均难以成立。事实上,从司法实务来看,也有判决表明人民法院并不在意通知人是否为本人。例如,在“雷某吉案”中,被告网站的用户发表了侮辱性内容。原告委托律师向被告邮寄了删帖通知,但被告并未删帖。人民法院认为,“上诉人获知该低俗、下流、侵犯他人权利的网帖存在就应当立即予以删除,其称不能明确是否系本人通知、不能确定律师是否接受委托等都不能成为其拒不立即处理的理由”。^④

此时必须厘清的问题是,若通知人并非权利人,网络服务提供者采取必要措施的义务是源于通知规则抑或知道规则?相较而言,以通知规则进行解释更为恰当,理由如下:(1)真实身份信息在通知规则中的核心规范目的是追究错误通知人的责任,而非沟通网络服务提供者、网络用户和通知

^① 参见浙江省宁波市中级人民法院课题组:《“通知-删除”规则的区别适用》,载《人民司法(应用)》2018年第4期,第64页。

^② See *Capitol Records, Inc. v. MP3tunes, LLC*, 821 F. Supp. 2d 627 (S. D. N. Y. 2011).

^③ 王利明:《侵权责任法研究》(下卷),中国人民大学出版社2018年版,第131页。

^④ 重庆市第五中级人民法院(2015)渝五中法民初字第03251号民事判决书。

人,也不是用于证明侵权成立,故通知人是否为权利人并不影响提供真实身份信息规范目的的实现。^① 据此,通知人身份信息,而非权利人身份信息,才是合格通知的必要条件。(2)若认为网络服务提供者的义务源于知道规则,将导致因不当采取必要措施而受侵害的网络用户无从获得救济。易言之,若网络服务提供者的义务源于知道规则,则意味着其是基于自身的判断而对所涉内容采取了必要措施,受害人将难以向通知人主张全部赔偿。同时,服务提供者是基于通知的提示而采取了必要措施,其也可基于没有过错或用户协议中的免责条款而免于承担赔偿责任。这一结果将助长非权利人的通知人滥发通知的现象。因此,在通知人并非权利人时,将网络服务提供者采取必要措施的义务解释为源于合格通知是更为可取的选择。

通知人不以权利人为限这一判断在现行法上并无障碍,因为《民法典》第1195条中的“权利人”应采广义解释。《民法典》第1195条在两处使用了“权利人”一词:一是第1款关于通知的规定,二是第3款关于错误通知的规定。该条第3款第1句规定:“权利人因错误通知造成网络用户或者网络服务提供者损害的,应当承担侵权责任。”从规范目的来看,该规定是为了在通知所声称的事实与客观事实不一致时,对遭受损害的网络用户和网络服务提供者进行救济。错误通知发生的情形,不应限于权利人(狭义)误以为自己的权利受到了侵害,也应包括并无权利的主体自称其为权利人的恶意通知。后者在实践中更为常见,也是该款主要规制的对象。据此,第3款中的“权利人”应作广义解释,不仅包括确有权利的主体,也包括并无权利但自称有权的主体。

综上,尽管合格通知以提供真实身份信息为必要,但这并不意味着通知人必须是合法权益受害的主体。只要通知提供了构成侵权的初步证据,就无须限制通知人的范围。据此,在生成式人工智能侵权中,并不要求通知人必须是权利人。

三、生成式人工智能侵权中必要措施的认定

在网络服务提供者收到合格通知且审查通过后,其应采取何种必要措施?传统网络侵权中的必要措施能否适用于生成式人工智能侵权,有待检视。

(一)传统必要措施在生成式人工智能侵权中的变化

我国网络侵权中的必要措施类型经历了逐渐扩大的过程。在通知规则发展的早期,必要措施是指删除、屏蔽、断开链接等能够直接避免损害扩大的措施。但随着实践的发展,尤其是最高人民法院指导案例83号发布之后,必要措施的类型在我国逐渐多样化。^② 总体而言,我国必要措施可分为以下四类:(1)可以直接避免损害扩大的措施,如删除、屏蔽、断开链接、终止交易和服务等。这是传统典型的必要措施。(2)发挥警示功能的措施,例如转通知等。转通知是较为缓和的措施,其无法直接避免损害扩大,适用情形主要有二:一是通知所涉内容侵权与否不易判断,如专利侵权;二是网络服务提供者(例如云存储服务提供者)基于商业伦理不宜对通知所涉内容采取移除等措施。^{③④}

^① 参见徐伟:《网络侵权中“合格通知”规则检视及其完善——以〈民法典〉第一千一百九十五条第一款第2句为中心》,载《法治社会》2023年第3期,第24-25页。

^② 参见徐伟:《〈民法典〉中网络侵权制度的新发展》,载《法治研究》2020年第4期,第102-104页。

^③ 参见最高人民法院指导案例83号(2017年)。

^④ 参见北京知识产权法院(2017)京73民终1194号民事判决书。

(3)冻结账户、提供保证金等发挥担保功能的措施。这种措施以网络用户提供担保的方式来避免其内容被删除。这大多发生于通知所涉内容是否侵权不易判断,或一旦错误采取了措施将对用户利益产生重大影响的情形。(4)限制乃至停止向网络用户提供服务的措施,如临时性甚至永久性停止账户的使用等。这大多适用于故意、多次侵权者或侵权造成严重后果的情形。我国现行法中并未明确规定该规则,但司法裁判中有人民法院将重复侵权作为应采何种“必要措施”的考量因素。^①

上述传统网络侵权中的必要措施,并非都可以适用于生成式人工智能侵权。即便可以适用,其具体表现也发生了变化,对此分析如下:

第一,可以直接避免损害扩大的措施虽可适用于生成式人工智能侵权,但其具体表现与传统网络侵权有所不同。传统网络侵权中的删除等措施,其处理的对象是过去已经发生的侵权内容,其功能主要在于“停止侵害”。在生成式人工智能侵权中,将人工智能与使用者对话中已经生成的致害内容删除意义甚微,传统上网络服务提供者也并不负有此类义务。^②避免损害扩大的关键措施是避免致害内容再次生成,这才是生成式人工智能服务提供者应采取的措施,其功能主要在于“消除危险”。^③可见,传统网络侵权中以“面向过去”为主的措施^④在生成式人工智能侵权中需转变为以“面向未来”为主的措施。

第二,发挥警示功能的措施从警示网络用户转变为警示生成式人工智能使用者,这意味着警示措施的规范目的从提醒网络用户停止潜在的侵权行为转变为提醒使用者相关内容可能存在不实。传统网络侵权中的警示性措施针对的是网络用户,希望通过向可能实施了侵权行为的网络用户发送通知等警示性信息,提醒侵权用户主动停止侵权。但在生成式人工智能侵权中,并不存在直接实施侵权行为的用户^⑤,制止侵权行为也无须借助用户之手来完成。生成式人工智能侵权中的警示信息只能提供给人工智能的使用者。该警示措施表现为:当通知所涉内容再次生成时,通过该内容添加警示标识等方式来提醒使用者该内容可能存在不实。该警示义务不同于生成式人工智能服务提供者负有的一般性标识义务,后者是针对所有生成内容的普遍性义务,并非因通知人的合格通知而产生的义务。

有疑问的是,警示性措施在生成式人工智能侵权中是否有存在的必要。生成式人工智能服务提供者若要对生成的涉嫌侵权的内容添加警示标识,便意味着其需审查所有生成内容,而一般性审查义务一直为传统网络侵权规则所否定。^⑥但这不足以否定生成式人工智能侵权中的警示性措施,理由如下:(1)传统网络侵权规则在否定一般性审查义务的同时,也认可特定情况下的特殊审查义

^① 相关案例载《最高人民法院公报》2012年第1期,第38-48页。

^② 传统网络侵权中亦存在类似的情形,即行为人通过通信软件(例如微信群)发布致害内容,此时法院一般要求行为人在同一通信软件中发布道歉声明,而不会要求网络服务提供者删除行为人此前发布的致害内容。典型案例参见最高人民法院指导案例143号(2020年)。

^③ 一般认为,停止侵害针对的是正在进行的侵害行为,消除危险针对的是尚未发生的潜在侵害行为。但有学者主张对停止侵害做广义理解,既包括停止已发生的侵害行为,也包括停止未来可能发生的侵害行为。若采此种理解,则停止侵害(删除等措施)也具有“面向未来”的效果。但此种理解会导致停止侵害与消除危险(尤其是重复危险情形)的重叠,故本文未采此种理解。

^④ 传统必要措施并非面向未来的典型表现是,网络服务提供者原则上不负有避免致害内容再次出现的义务。

^⑤ 若致害内容是因使用者诱导而生成,则使用者是直接侵权者,但在具体实践中,难以确定某一致害内容是因使用者的诱导行为生成。参见《中华人民共和国人工智能法(学者建议稿)》第87条。

^⑥ 参见《最高人民法院关于审理侵害信息网络传播权民事纠纷案件适用法律若干问题的规定》(法释[2012]20号)第8条第2款。比较法经验参见 Marisa N. Sanchez, *Eu Directive on Copyright in the Digital Single Market: An Outlier in Intermediary Liability and the Death of Safe Harbor Protections*, 55 University of San Francisco Law Review 251, 268-269(2021)。

务。目前逐渐出现的一种特殊审查义务要求网络服务提供者采取技术措施对某些内容进行普遍审查,典型例证是版权作品过滤技术。随着视频指纹识别技术等版权过滤技术的成熟,产业界已能够在技术层面普遍采用版权过滤审查技术。有学者据此主张网络服务提供者应当负有采用版权过滤技术的义务。^①生成式人工智能侵权中对通知所涉致害内容的审查类似于通过版权过滤技术来审查版权作品。(2)审查技术本身的成熟度和成本是法律决定是否对网络服务提供者科以审查义务的重要考量因素之一。^②生成式人工智能侵权中的内容系由人工智能自动生成,避免特定内容的生成技术可内嵌于算法的设计中自动实现,它并不会对生成式人工智能服务提供者造成较大的成本负担。(3)在人格权侵权中,事实真伪不明是常见情形。若异议方发送通知,生成式人工智能服务提供者大多无法判断事实的真伪。此时对生成的内容进行警示标识,而非直接禁止相关内容再次生成,是更可取的选择。综上,采取警示性措施仍有必要。当然,鉴于生成式人工智能的应用类型多样,警示性措施的具体表现方式可结合相关技术特征予以确定。

第三,发挥担保功能的措施无法适用于生成式人工智能侵权。此类措施主要针对涉嫌侵权的网络用户,但在生成式人工智能侵权中不存在直接实施侵权行为的用户。

第四,限制乃至停止提供服务的措施从针对网络用户转变为针对诱发致害内容生成的使用者。传统网络侵权中限制或停止服务措施一般针对反复侵权的网络用户。在生成式人工智能侵权中,引发致害内容生成的原因除了人工智能本身外,也包括第三人的不当诱导。根据《生成式人工智能办法》第14条第2款的规定,生成式人工智能服务提供者发现使用者利用生成式人工智能服务从事违法活动的,应当依法依约采取警示、限制功能、暂停或者终止向其提供服务等处置措施。故生成式人工智能服务提供者可以对诱导致害内容生成的第三人采取限制性措施。^③该义务与避免致害内容再次生成义务的差别在于,前者针对的是人(诱导者),其试图避免出现“累犯”,影响的是某一使用者;后者针对的是内容,其试图避免同一事件再次发生,影响的是所有使用者。对第三人采取限制性措施这一义务具有双重属性,其既是行政法上的义务,也是民事义务,违反该义务可能同时产生行政责任和民事责任。

(二)避免致害内容再次生成的义务

避免致害内容再次生成是权利人发送通知的首要目的。我国《生成式人工智能办法》第14条第1款规定:“提供者发现违法内容的,应当及时采取停止生成、停止传输、消除等处置措施。”对于该义务,有两方面问题需要厘清:一是该义务的正当性;二是该义务的限制度。

避免致害内容再次生成的义务是一项面向未来的预防性义务,是否要对网络服务提供者科以此种义务无疑会引发争议。传统网络侵权中与之相似的争议体现为“通知且保持下架”(notice and staydown)规则,即网络服务提供者在收到合格通知后不仅要下架通知所涉侵权内容,而且要确保该内容不会被再次上传。一般认为,德国是采纳了通知且保持下架规则的代表性国家^④,欧盟则是采

^① 参见崔国斌:《网络服务商共同侵权制度之重塑》,载《法学研究》2013年第4期,第158页。

^② 参见李安:《智能时代版权“避风港”规则的危机与变革》,载《华中科技大学学报(社会科学版)》2021年第3期,第116页。

^③ 然而,未来生成式人工智能服务普及后,禁止某一用户使用生成式人工智能服务将对其生活产生显著影响。故承认这一规则,保有这一规则对侵权用户的威慑效果,同时对该规则设置较为严格的适用条件,较为妥当。

^④ See Jan Bernd Nordemann, *Liability for Copyright Infringements on the Internet: Host Providers (Content Providers) - The German Approach*, 2 *Journal of Intellectual Property, Information Technology and E-Commerce Law* 37, 40-43 (2011).

纳该规则的代表性国际组织。^①但也有许多国家并未接受通知且保持下架规则。

学界对通知且保持下架规则本身的合理性存在较大分歧。支持者的主要理由包括:其一,通知规则会导致版权人陷入无休止的维权,即通知后侵权内容会再次出现。其二,高质量的版权执法自动化出错较少,与此同时,版权执法自动化大大改善了版权保护现状。其三,通知后仅移除的规则将使网络服务提供者在与权利人谈判时处于有利地位,不利于促进双方合作。^②反对者的主要理由是:其一,通知且保持下架规则影响基本人权的实现。其二,该规则增加了企业的经营负担,不利于中小企业发展,也不利于创新。其三,欧盟采纳该规则是为了监管来自美国的互联网企业,因为欧洲主要的大型互联网企业都来自美国。其四,应将是否采取保持下架的选择权交由平台。例如,若通知较少,则平台更倾向于人工审查处理;若通知较多,则使用自动化过滤会更有效率。其五,通知且保持下架规则对于过滤有更高要求,可能导致相关市场的经营者集中,市场的竞争性减弱。最后,从国际竞争的角度来看,采通知且保持下架规则将导致该国在国际竞争中处于制度洼地。^③此外,也有学者主张是否应适用通知且保持下架规则取决于在技术层面能否达到较高的过滤精确度。通知且保持下架规则应适用于算法自动化处理的准确率不低于人工审查准确率的情形。^④

在我国,鉴于必要措施在概念上的包容性,采取避免致害内容再次生成的义务在解释上可被“必要措施”所涵盖。尽管学界对普遍性的通知且保持下架规则持谨慎态度^⑤,但在具体个案中,我国司法裁判已采纳了该规则。例如,在“某度网盘案”中,人民法院认为“网络服务提供者针对用户侵权所应采取的必要措施,不应仅限于在侵权发生之后断开链接一项,否则即便其做到了及时断开,往往也会使侵权用户、权利人和网络服务提供者之间陷入‘侵权-通知-断开-再侵权-再通知-再断开’的往复循环之中,难以有效制止用户的持续侵权和权利人损失的扩大”。^⑥随后,人民法院比较了采取屏蔽分享链接这一预防措施导致用户不能对其享有合法权利的内容进行正当分享造成的损害和不采取预防措施导致著作权人遭受的损害,以及屏蔽措施的技术成熟度和成本等,最终认定某度公司负有采取屏蔽措施的义务。据此,我国网络服务提供者负有采取预防性措施的义务并无立法和实践障碍,问题的关键在于综合考量诸多因素后合理确定该义务的适用条件。

根据最高人民法院指导案例 83 号裁判要点二,“‘必要措施’应遵循审慎、合理的原则,根据所侵害权利的性质、侵权的具体情形和技术条件等来加以综合确定”。据此,生成式人工智能侵权中是否要求生成式人工智能服务提供者负有避免致害内容再次生成的义务,需考量的因素如下:首先,被侵害权利的性质。一般而言,人格权较之知识产权等财产性权利应受到更严格的保护。但事实上在我国通知规则中,二者长期以来受到一体保护,故在生成式人工智能侵权中也没有必要对二

^① See Directive (EU) 2019/790 of the European Parliament and of the Council of April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, Article 17 Section 4(c). 虽然指令规定了保持下架义务,但并未提及违反该义务的法律后果,而是将其交由各成员国自行规定,故其法律后果既可能是行政处罚等行政责任,也可能是损害赔偿等民事责任。

^② See United States Copyright Office, *Section 512 of Title 17: A Report of the Register of Copyrights*, U. S. Copyright Office (May 2020), <https://www.copyright.gov/policy/section512/section-512-full-report.pdf>.

^③ See Daniel Etcovitch, *DMCA § 512 Pain Points: Music and Technology Industry Perspectives in Juxtaposition*, 30 *Harvard Journal of Law & Technology* 547, 562-563 (2017).

^④ See Martin Husovec, *The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown? Which is Superior? And Why?*, 42 *Columbia Journal of Law & the Arts* 53, 67-73 (2018).

^⑤ 参见刘文杰:《“通知—移除”抑或“通知—拦截”:算法时代的选择》,载《新闻与传播研究》2020年第12期,第35-36页。

^⑥ 北京知识产权法院(2020)京73民终155号民事判决书。

者进行区别对待。其次,从侵权的具体情形来看,生成式人工智能生成的致害内容仅参与对话的使用者可见,故较之传统的网络侵权,其侵权严重程度难谓更高。最后,从技术条件来看,生成式人工智能服务提供者能否以成本合理且准确性较高的技术来避免致害内容再次生成,是影响其是否负有避免致害内容再次生成义务的主要因素。如果服务提供者能够以某种成本可负担的技术来较为准确地避免致害内容的再次生成,其就应采取该措施。反之,若该技术成本高昂,或该技术在避免致害内容生成的同时会对诸多合法内容的生成造成负面影响(准确性较低),则不宜科以生成式人工智能服务提供者此义务。《生成式人工智能办法》制定过程中的条文变化也表明,避免致害内容再次生成在技术层面可以实现,但也存在一定的难度。^①从比较法的视角来看,欧盟《人工智能法》第9条第5款也采取了类似的规定,即对于高风险的人工智能系统,其风险管理措施应“通过充分的设计与开发,尽可能消除或者降低风险”,同时“在适当的情况下,对于无法消除的风险,采取充足的缓解和控制措施”。此外,我国提供生成式人工智能服务的企业在用户协议中也表明会根据权利人的通知而采取“停止生成、停止传输、消除等合理处置措施”。^②据此,生成式人工智能服务提供者在一般意义上确实负有避免致害内容再次生成的义务,但该义务的边界需根据个案中的具体商业模式以及人工智能技术的发展状况等因素综合判断。

此外,避免致害内容再次生成的义务应有期限限制,理由如下:其一,避免致害内容再次生成会导致相关内容无法再次出现,而相关内容是否侵权可能会随着时间推移而发生变化,例如谣言可能在事后被证明是事实等。故对该义务设置一定的期限更有利于实现各方利益的动态平衡。其二,避免致害内容再次生成对言论表达造成了一定的限制,对该义务设置期限限制有助于保持生成内容的开放性和多样性。其三,生成式人工智能会随着技术和数据集的更新而不断学习,进而自动纠错,可能不再生成致害内容。其四,与避免致害内容再次生成义务相似的通知且保持下架规则在部分国家也有一定的期限限制。^③其五,人格权侵害禁令和人身安全保护令都有期限限制。例如,人身安全保护令一般不超过六个月,人格权侵害禁令也应有期限限制。^④经人民法院裁判确认的禁令尚且有期限限制,根据“举重以明轻”的法理,因通知所产生的避免致害内容再次生成的义务也应有期限限制。至于该期限的长短,理想状态是由生成式人工智能服务提供者根据具体情况自行判断和决定,但为避免其基于自身利益考量而确定不合理的期限,也可由法律直接规定,例如,规定避免侵害人身权益内容再次生成的期限为六个月。

(三) 致害内容无法再次生成时的告知义务

在生成式人工智能侵权中,权利人在通知中提供的是对话记录,但生成内容的不确定性导致致害内容可能并未再次生成。此时生成式人工智能服务提供者是否负有告知权利人的义务?该告知义务具有双重意义:一是告知权利人该通知未能提供构成侵权的初步证据,因为生成式人工智能服务提供者未能再现对话记录;二是若权利人认为致害内容确实出现过,则告知义务发挥的功能是告知权利人致害内容已经移除,因为其在生成式人工智能服务提供者处没有再次生成。在传统网络

^① 参见《生成式人工智能服务管理办法(征求意见稿)》第15条和《生成式人工智能办法》第14条第1款。

^② 参见《文心一言用户协议》第9.1条,载文心一言官网,<https://yiyi.baidu.com/infoUser>,2024年3月30日访问。

^③ See Martin Husovec, *The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown? Which is Superior? And Why?*, 42 Columbia Journal of Law & the Arts 53, 56 (2018).

^④ 参见徐伟:《〈民法典〉人格权侵害禁令的法律适用》,载《法制与社会发展》2021年第6期,第216-217页。

侵权中,第一种情形属于通知不合格,故一般纳入不合格通知的法律后果中加以讨论;第二种情形可纳入必要措施的概念下进行讨论,但传统上网络服务提供者一般不负有在采取必要措施后向权利人告知的义务。此处将两种情形纳入必要措施中一并讨论。

对于通知不合格时网络服务提供者是否负有告知义务这一问题,理论上存在分歧。有学者认为,“对于不合格的侵权通知,网络服务提供者不予以回应,不采取措施的话,可以不承担责任。”^①不同意见则认为应告知权利人。^②从司法裁判来看,有判决直接否定了网络服务提供者的告知义务。例如,在“阿某云案”中,人民法院认为权利人的通知因内容欠缺等原因不合格,且阿某云公司不负有主动联系通知人进行核实和调查的义务,理由如下:其一,《信息网络传播权保护条例》并未规定该义务。其二,要求网络服务提供者联系和核实将导致法律对通知合格要件的规定落空。其三,要求网络服务提供者对不合格通知主动联系核实,将大幅增加其运营成本。其四,权利人在依法通知时应当本着对自身权利负责的态度准备通知书,不能因为权利人未全面尽到该义务,转而向网络服务提供者施加核实等义务。^③可见,无论是通知不合格的告知义务,还是采取必要措施后的告知义务,都难以在我国现行法层面找到依据。

但在生成式人工智能侵权中,致害内容无法再现时生成式人工智能服务提供者应负有告知义务,理由如下:

第一,相较于生成式人工智能服务提供者,无辜的权利人应受到优先保护。致害内容无法再现的原因既可能是该致害内容本就不会出现(通知人伪造对话记录),也可能是该致害内容在权利人处出现了但在生成式人工智能服务提供者处未能出现。而造成致害内容无法再现的原因在客观事实层面往往难以查明。此时法律宜推定通知人系善意,即无法再现的原因是生成内容具有不确定性。因为该推定与客观现实中的多数情况相符(可合理推测伪造对话记录是少数),不应为了避免少量不实通知而罔顾对真实权利人的保护。在推定权利人系善意通知的前提下,对话记录无法再现系由生成式人工智能的技术原因所致,而非权利人怠于查找和提供侵权初步证据所致,即权利人对致害内容无法再现并无过错。另外,生成式人工智能服务提供者对致害内容无法再现并无过错,因为这是人工智能自主自发的结果。但就技术层面而言,人工智能的自主性正是生成式人工智能服务提供者在技术上主动追求的结果。^④从侵权一般原理来看,“被告企业应为自己特殊、不寻常和危险的特征付出代价”。^⑤可见,尽管权利人和生成式人工智能服务提供者对致害内容无法再现均无过错,但相较而言权利人更为无辜,其对致害内容无法再现缺乏控制力,故不宜由权利人承担无法再现的不利后果。据此,宜要求生成式人工智能服务提供者告知权利人致害内容无法再现这一情况,以便权利人查验并决定后续行为的选择。

第二,对服务提供者科以告知义务与通知规则的规范目的相契合。通知规则的规范目的之一是促进权利人与网络服务提供者之间的合作以减少网络侵权。若网络服务提供者不予告知,将导致权利人无从知晓自己发送的通知被如何对待。同时,生成式人工智能侵权与传统网络侵权的区

① 薛军:《民法典网络侵权条款研究:以法解释论框架的重构为中心》,载《比较法研究》2020年第4期,第140页。

② 许谅亮:《网络交易平台提供商专利侵权责任》,载《科技与法律》2015年第3期,第513页。

③ 参见北京知识产权法院(2017)京73民终1194号民事判决书。

④ 参见冯珏:《自动驾驶汽车致损的民事责任》,载《中国法学》2018年第6期,第116页。

⑤ Restatement (Second) of Torts § 519 General Principle (1977).

别之一是,在传统网络侵权中权利人易于查验网络服务提供者是否采取了必要措施,但在生成式人工智能侵权中,若生成式人工智能服务提供者不主动告知,则权利人无从判断其是否采取了必要措施,因为致害内容未再次生成未必是生成式人工智能服务提供者采取了必要措施后的结果。这些都可能引发潜在的后续争议。例如,若致害内容再次生成,权利人将主张生成式人工智能服务提供者未采取必要措施,生成式人工智能服务提供者则主张权利人的通知不合格,进而引发双方在责任承担方面的争议。因此,由生成式人工智能服务提供者向权利人告知通知情况有助于减少后续争议的发生。

第三,生成式人工智能服务提供者的告知义务有助于权利人进一步查验致害内容是否会再次生成,进而更好地维护自身合法权益。与传统网络侵权不同的是,生成式人工智能服务提供者的告知义务发挥的功能之一是告知权利人相关致害内容已经自然消失。鉴于生成内容具有不确定性,若致害内容在生成式人工智能服务提供者未采取任何措施的情况下未能再次生成,此时交由权利人来查验致害内容将有助于确认该致害内容是否确实不再生成。同时,权利人在了解对话记录无法再现后,其可能会进一步以其他提问方式来检验类似的致害内容是否会生成,因为作为利益最密切相关者,其更有动力查找致害内容。

第四,比较法经验支持在一定条件下对网络服务提供者科以告知义务。尽管我国法并未直接规定网络服务提供者的告知义务,但美国法区分通知中欠缺事项的不同而对网络服务提供者提出了不同要求。根据《数字千年版权法》第512(c)(3)(B)条的规定,不合格通知被分为两类:一类是完全不合格的通知,网络服务提供者收到后可不作回应。另一类是瑕疵通知,网络服务提供者收到此类通知后,应积极联系投诉人(*complaining party*),协助其补正缺失的材料。所谓“瑕疵通知”,是指通知至少满足了合格通知的三项要件^①,其缺失的只能是投诉人或被授权人的签名、善意相信被投诉内容未经授权的声明以及愿对虚假陈述承担责任的声明。美国的做法具有参考价值,因为在侵权并不明显时,由投诉人而非网络服务提供者承担发现侵权内容的成本是通知规则的既有价值选择,若投诉人的通知已基本满足了合格通知的多数实质性要求,只是欠缺部分非实质性内容时,网络服务提供者仍对此置之不理,则对投诉人过于苛刻。参考美国的这一思路,若因人工智能本身的技术原因导致通知所涉对话记录无法再现,此时生成式人工智能服务提供者的置之不理对权利人而言过于苛刻。

综上,在致害内容无法再现时,生成式人工智能服务提供者应负有向权利人告知的义务。须指出的是,该告知义务限于致害内容无法再现的情形,并不意味着在所有通知不合格的情形中服务提供者均负有告知义务。

四、结语

通知规则在生成式人工智能侵权中的适用,既与通知规则本身相关,又受到生成式人工智能技术特征的影响。一方面,传统网络侵权中关于通知规则的诸多争议直接影响了该规则在生成式人

^① 根据《数字千年版权法》第512(c)(3)(A)条的规定,合格通知的三项要件是:被侵害的版权作品的信息、被侵害的作品内容和合理的足以定位侵权内容的信息以及合理的足以供服务提供者联系投诉人的信息。

工智能这一新场景中的适用。同时,生成式人工智能让传统通知规则中的一些搁置问题再次凸显。例如,通知规则的定性问题虽已讨论多年,但因将其定性为归责条款抑或免责条款对实践中当事人利益的影响不明显,导致该问题虽具有重要理论价值却一直未有定论。与此相类似,通知人是否只能是权利人是传统通知规则中未能引起学界关注的问题,但在生成式人工智能侵权中,这一问题的重要性进一步凸显。另一方面,生成式人工智能也在一定程度上改变了传统的通知规则,尤其是丰富了“必要措施”义务。由于生成式人工智能场景中并不存在网络用户,故传统上针对网络用户的措施都发生了转变。而传统上针对内容的措施,也从主要面向“过去”(停止侵害)转变为主要面向“未来”(消除危险)。这些转变都表明,传统通知规则需在人工智能时代进行理论更新! **ML**

On the Application of Notification Rule in the Scenario of Generative AI Infringement

XU Wei

(Shanghai Justice Research Institute, Shanghai University of Political Science and Law, Shanghai 201701, China)

Abstract: The notification rule is a liability fixation clause, not a liability exemption clause in China. This determines that the notification rule can be applied to Generative AI infringements directly, rather than by analogy. The absence of counter-notification in the scenario of Generative AI is not a sufficient reason to deny the applicability of notification rule. The “information to accurately locate the infringing content” in the notification can be provided by the conversation records and screenshots, and the complaining party could not be the right holder. In Generative AI infringement cases, there are several changes in the “necessary measures” stated in traditional notification rule: (a) measures to prevent the expansion of damage shift from focusing on the infringing content that has already occurred to preventing infringing content in the future; (b) warning measures shift from alerting tortfeasors to alerting users of Generative AI; (c) security measures are not applicable; and (d) measures restricting or even ceasing shift from against network users to users who induce the generation of infringing content. Generative AI Service Providers are obligated to avoid the generation of infringing content after receiving the notification, but the boundary of the obligation should be determined according to the specific business model, the development of artificial intelligence technology, and other factors. Also there should be a time limited duration of this obligation. Generative AI Service Providers are obligated to inform the complaining part when the infringing content is failed to be generated again.

Key words: generative artificial intelligence; notification rule; safe harbor; online infringement; necessary measure

本文责任编辑:黄 汇