

## 论生成式人工智能服务提供者侵权损害赔偿

徐 伟\*

**内容提要：**生成式人工智能侵害人格权是否会导致损害，须结合具体人格权类型加以判断。基于生成的侵权内容而发生的下游损害不应纳入服务提供者应承担的损害范围；潜在的再次生成侵权内容的损害需视再次发生的几率来决定是否予以赔偿。因被侵权人的损失和侵权人的获利难以准确计算，酌定赔偿或法定赔偿是确定赔偿金的主要方式。在生成阶段，服务提供者侵权的过错程度是过失，人机交互的侵权方式使得侵权影响范围往往有限，故服务提供者应承担的损害赔偿金不宜高，且通常也不应承担精神损害赔偿。消除影响、恢复名誉的责任承担方式在生成式人工智能场景中可以表现为服务提供者向接触过侵权内容的使用者推送澄清声明或专门开辟页面来发布澄清声明，且不宜再适用赔礼道歉。相较于传统损害赔偿理论，生成式人工智能侵权的特殊之处在于改变了事实基础、冲击了价值选择和影响了治理逻辑。

**关键词：**生成式人工智能 侵权责任承担方式 损害赔偿 财产损失 精神损害

### 一、问题的界定

随着生成式人工智能技术的快速发展和广泛应用，因生成侵害他人合法权益的内容（以下简称“侵权内容”）而引发的纠纷日渐增多。我国已出现涉知识产权侵权的案件，且法院判决服务提供者构成侵权。<sup>〔1〕</sup> 涉人格权侵权判决在我国尚未出现。<sup>〔2〕</sup> 但国外已出现多起涉人格权侵

\* 徐伟，上海政法学院上海司法研究所余山学者特聘岗教授。

本文为2024年度教育部人文社会科学研究青年基金项目“人工智能开发者侵权责任研究”（24YJC820052）的阶段性研究成果。

〔1〕 参见广州互联网法院（2024）粤0192民初113号民事判决书。

〔2〕 虽然我国并未出现以“生成式人工智能”为名的侵权案件，但与人工智能相关的人格权纠纷则已有判决，例如“AI陪伴者案”“AI声音案”“AI换脸案”等。参见北京互联网法院（2020）京0491民初9526号民事判决书；北京互联网法院（2023）京0491民初12142号民事判决书；北京互联网法院服务保障新质生产力十大典型案例之八。

权纠纷。<sup>〔3〕</sup>可预见未来生成式人工智能侵权纠纷将不断出现。本文旨在探讨当服务提供者构成侵权时，其应承担的损害赔偿责任。在具体展开损害赔偿问题前，须对本文讨论的问题作出界定。

第一，在生成式人工智能侵权构成尚无共识的情况下，探讨损害赔偿问题是否可能？侵权构成是损害赔偿的前提。目前关于服务提供者何时构成侵权尚远未达成共识，例如关于人工智能侵权归责原则，学界存在采过错责任、过错推定责任、无过错责任以及根据场景分别适用不同归责原则的主张。<sup>〔4〕</sup>在“归责”问题尚未解决的情况下，探讨损害赔偿问题显得“先天不足”，因为侵权责任中的“损害”是规范性概念，即并非所有事实上的损害都能够成为法律上应予救济的损害。是否存在“损害”归根结底取决于法律在价值判断上是否要求服务提供者承担侵权责任。但基于下述考虑，探讨损害赔偿问题仍有其价值。其一，损害是侵权损害赔偿责任的构成要件之一，对损害的探讨有助于厘清服务提供者责任。至少对于不存在法律上损害的情形，服务提供者必然不构成侵权损害赔偿责任。其二，除了是否存在损害外，如何判断损害的大小，尤其是结合生成式人工智能场景的分析，有其价值。其三，如何救济损害亦是损害赔偿中需要解决的问题。据此，即便在生成式人工智能侵权尚存在分歧的背景下，对损害赔偿的探讨仍有其价值。

第二，生成式人工智能侵权涉及诸多行为人，各行为人承担的责任可能不同，如何在一般意义上探讨损害赔偿？生成式人工智能涉及的潜在责任主体包括开发者、应用提供者、部署者、销售者，乃至使用者等。不同主体承担的责任范围和方式可能有所不同。例如，根据人工智能业态可将服务提供者分为基础模型、专业模型和服务应用提供者，<sup>〔5〕</sup>不同服务提供者在责任构成和责任范围上可能应予以差别对待。鉴于本文旨在为生成式人工智能损害赔偿提供一般性的原理和规则，故本文的分析主要是从救济受害人角度探讨哪些损害应予赔偿。至于该损害应由哪个主体来承担，特定主体间的责任形态乃至赔偿限额等，非本文关注所在。据此，本文所称生成式人工智能服务提供者系广义概念，是指代生成式人工智能开发者、应用提供者、部署者等除使用者之外的其他可能承担责任的主体。当然，开发者是责任主体中的典型。

第三，对损害赔偿的分析是否要根据生成式人工智能侵权场景进行类型化区分？根据侵权场景对人工智能侵权予以类型化是侵权构成分析的常见思路。但侵权构成系以行为人为着眼点，而损害赔偿的核心规范目的在于填补受害人的损害，即主要着眼于受害人，故当前基于侵权构成而展开的类型化方式在探讨损害赔偿问题时并不完全适用，详言之如下。

其一，根据人工智能技术路线，人工智能可分为采符号主义和采联结主义两类。<sup>〔6〕</sup>相较而言，采“符号主义”的人工智能“可控”，而采“联结主义”的人工智能则“难以完全可控”。

〔3〕 See *Walters v. OpenAI, L. L. C.*, No. 1: 23-cv-03122.

〔4〕 参见《人工智能示范法 2.0（专家建议稿）》第 70 条第 1 款（采过错推定责任），《中华人民共和国人工智能法（学者建议稿）》第 85 条（区分关键人工智能和其他人工智能而分别采过错推定责任和过错责任）。采过错责任的观点参见王利明：《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。采无过错责任的观点参见徐伟：《生成式人工智能服务提供者侵权归责原则之辨》，载《法制与社会发展》2024 年第 3 期。国外的讨论参见〔瑞〕米里亚姆·布伊滕、〔比〕亚历山大·德·斯特里尔、〔德〕马丁·佩茨：《人工智能责任的法律与经济学》，张韬略、陈沪楠译，载《上海政法学院学报（法治论丛）》2024 年第 4 期。

〔5〕 参见张凌寒：《生成式人工智能的法律定位与分层治理》，载《现代法学》2023 年第 4 期。

〔6〕 参见徐英瑾：《人工智能哲学十五讲》，北京大学出版社 2021 年版，第 3-24 页。

鉴于当前的生成式人工智能多是采联结主义技术路线（尤其是人工神经网络），且唯有“难以完全控制”的人工智能才会引发侵权难题，故本文探讨的生成式人工智能以采联结主义技术为前提。

其二，根据生成式人工智能的全流程，侵权可分为发生在训练阶段、生成阶段和移除阶段三类。在训练阶段，因服务提供者是有意识且能控制训练数据的选择，故训练数据侵权问题与人工智能自主性等技术特征无关。生成阶段主要涉及人工智能生成侵权内容时的责任问题。移除阶段主要涉及在权利人通知服务提供者存在侵权内容后，服务提供者未采取必要措施时的责任问题，这其实是生成阶段侵权问题的进一步延伸。本文聚焦于探讨生成式人工智能自主性等技术特征引发的损害赔偿问题，故本文分析限于生成和移除阶段的侵权损害赔偿问题。

其三，根据人工智能应用场景、智能程度、风险程度等进行的类型化对损害赔偿问题的分析并无实质性影响，无需以此展开对损害赔偿的类型化分析。生成式人工智能的应用场景包括文生文、文生图、文生视频等诸多场景。无论何种场景，其本质上都是对民事权益的侵害（主要包括人格权和知识产权），故只需根据侵害的权益类型展开对损害赔偿的探讨，无需根据应用场景来类型化。关于智能程度，一般可分为弱人工智能、强人工智能和超人工智能。无论是弱人工智能还是强人工智能，都存在难以预见和控制的技术特征。<sup>〔7〕</sup>而该技术特征是引发损害赔偿难题的重要原因。故对损害赔偿问题的分析无须区分人工智能强弱作类型化。同理，基于风险程度可将人工智能分为禁止的风险、高风险、有限风险和低风险四类，对侵权损害赔偿问题的分析影响亦不大。

综上，本文对生成式人工智能侵权责任的探讨，限于采联结主义的生成式人工智能在生成和移除阶段侵权时，服务提供者（广义）因对受害人的侵害而承担的损害赔偿责任。

## 二、生成式人工智能服务提供者侵权中损害的认定

生成式人工智能侵害的权利主要包括人格权和知识产权两类。对损害的认定须结合具体的权利类型加以分析。

### （一）损害是否存在的认定

根据损害能否以金钱直接加以衡量，损害包括财产损害和非财产损害（精神损害）。生成式人工智能侵害知识产权时，例如生成了他人享有著作权等知识产权的内容，此时知识产权人遭受了财产损害当无疑问，例如丧失了本可授权他人使用知识产权而可获得的经济利益。损害认定的难题主要存在于侵害人格权的情形。生成式人工智能侵害人格权时，是否必然会存在损害，结合人格权类型分析如下。

第一，对于侵害生命权等物质性人格权而言，人工智能诱导他人实施了侵害生命等的行为便可认定为存在损害。因生成式人工智能具有无体性，其往往无法直接对他人的利益实施物理上的侵害，故其加害行为多表现为通过不当诱导影响行为人的观念，使行为人实施伤害自我或他

〔7〕 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。

人生命等人格利益的行为。<sup>〔8〕</sup>例如在美国的“角色 AI 案”中，受害人苏埃尔（Sewell）在与被告角色 AI 公司（Character. AI）提供的人工智能角色丹尼（Dany）进行长达约一年的聊天后，为了回到丹尼的“身边”，选择了举枪结束自己的生命。<sup>〔9〕</sup>值得一提的是，若是生成式人工智能与实体设备结合后（例如形成人形机器人）造成损害，此时一般适用产品责任规则，而非纳入生成式人工智能侵权中讨论。

第二，对于侵害隐私权而言，人工智能生成他人私密信息便可认定为存在损害。隐私包括私密空间、私密活动和私密信息，而侵害隐私权的行为主要包括刺探、侵扰、泄露和公开等方式。在生成式人工智能场景中，侵害行为主要表现为人工智能生成私密信息而导致隐私被公开。生成式人工智能的人机对话模式不会影响损害的认定，因为人机对话模式虽然导致仅使用者可以获得相关内容，但因为不特定主体都可以进行人机对话，都有机会获得相关内容，故应认定该侵权内容被公开传播。该侵权内容是否确实被多次生成过并不影响损害存在与否，而只是影响损害的大小。

第三，对于侵害名誉权而言，是否存在损害须考量诸多因素。侵害名誉权所造成的损害主要表现为社会公众对受害人的评价降低。人工智能生成不实内容是否会导致相关主体社会评价降低取决于人工智能使用者是否会对该不实内容“信以为真”。首先，从人工智能技术原理来看，使用者不应将人工智能生成的内容当真。“有研究表明，大部分语言模型只有在 25% 的时间是真实可信的。”<sup>〔10〕</sup> 据此，若人工智能使用者了解人工智能生成内容的原理，其不应“相信”人工智能生成的内容系完全真实可靠。其次，从现实状况来看，对多数使用者而言，其并不了解人工智能的技术原理，在人工智能准确率和逼真程度不断提高的情况下，使用者会越来越“信赖”人工智能生成的内容，即将生成的内容作为真实信息对待。最后，还需将人机交互过程纳入考量。若人工智能服务提供者在设计人机交互界面时，已经事先向使用者声明了内容的不可靠性（例如在生成内容的对话框中或用户协议中声明内容系人工智能自动生成且未必真实可靠），此时使用者“理应”不完全信赖生成内容的真实性。当使用者不相信或不应当相信生成的内容完全真实时，难谓相关主体的社会评价因生成的内容而降低。综上，不实内容是否会造成相关主体社会评价降低进而侵害名誉权，需综合考虑人工智能的技术原理、使用者的普遍认知、人机交互界面的设计等诸多因素后，基于使用者对生成内容为真的信赖程度来加以判断。鉴于人工智能的应用将日益普及，人类会越来越信赖人工智能，法律上宜将生成不实内容作为受害人名誉受有损害的初步证据，服务提供者若欲主张损害不存在，则负有举证责任。

第四，对于侵害个人信息权<sup>〔11〕</sup>而言，服务提供者在生成阶段侵害个人信息权的典型表现是未经个人同意而生成包含个人信息的内容。此时是否存在损害需视情况而定。

〔8〕 关于人工智能对人类的操纵性风险，参见林涓民：《人形机器人的操纵性风险及规范进路》，载《东方法学》2024年第3期。

〔9〕 See Megan Garcia v. Character Technologies, Inc., No. 6: 24-cv-01903.

〔10〕 曹建峰：《迈向可信 AI：ChatGPT 类生成式人工智能的治理挑战及应对》，载《上海政法学院学报（法治论丛）》2023年第4期，第33页。

〔11〕 个人信息是采“权利”抑或“权益”予以保护存在分歧，本文采“权利说”，故使用“个人信息权”概念。采权利说的文献参见彭诚信、刘海安主编：《个人信息保护法》，上海交通大学出版社2024年版，第12-13页；采权益说的文献参见程啸、王苑：《个人信息保护法教程》，中国人民大学出版社2023年版，第143页。

若生成的是个人并未自行公开或通过其他合法方式公开的个人信息，此时服务提供者系违法公开个人信息，侵害了个人信息权。对于违法处理个人信息是否必然导致财产损害，学界存在不同意见，<sup>〔12〕</sup>但共识是若信息处理者通过该个人信息处理行为而直接获利，则个人存在财产损害。生成式人工智能场景中，服务提供者的行为正是通过生成含个人信息的内容而获利，故此时个人存在财产损害。

若生成的是个人自行公开或通过其他合法方式公开的个人信息，此时是否存在损害与该处理行为是否合法相关，因为“凡合法行为造成的不利后果，不属于侵权责任法上‘可救济的损害’”<sup>〔13〕</sup>。根据《中华人民共和国个人信息保护法》（以下简称《个人信息保护法》）第27条，在个人没有明确拒绝时，合法处理公开个人信息的条件是在“合理的范围”处理且该处理行为并未对“个人权益有重大影响”。该条件系认定信息处理者违法性/过错的行为规范，“合理的范围”与“个人权益有重大影响”系包含关系，后者是判断前者的反面因素。对“合理范围”的判断，既不能限缩于个人信息“被公开时的用途”，也不可宽泛地解释为只要法律没有明确禁止的行为均为合理，而应结合信息的特点与属性、信息的处理目的、信息的处理方式等来综合判断。<sup>〔14〕</sup>就“信息的特点与属性”而言，处理已合法公开的敏感个人信息仍须取得个人同意，故若人工智能生成的是敏感个人信息，则系违法行为，此时个人存在损害。就“信息的处理目的”而言，包含目的明确与目的合理两方面。生成式人工智能的处理目的明确，即生成相关内容。关于目的合理，鉴于人工智能生成内容具有广泛的积极应用价值，故除非个人在公开个人信息时明确排除了将其个人信息用于人工智能，否则宜认定人工智能生成含个人信息的内容系目的合理。就“信息的处理方式”而言，人工智能生成含个人信息的内容方式较为单一，一般而言并不存在可同样实现处理目的且对个人权益影响更小的方式。综上，对合法公开的个人信息：若生成的是敏感个人信息，此时个人存在损害；若是其他个人信息，一般而言服务提供者的生成行为合法，故个人不存在可得利益等的丧失，不存在损害。

## （二）损害范围的认定

若因人工智能生成侵权内容直接造成了权利人财产利益的丧失或精神痛苦，此时存在损害当无疑问。若权利人的损害并非侵权内容直接造成，而是第三人利用侵权内容造成，以及损害是否发生尚不确定（风险性损害）时，是否应纳入服务提供者应予赔偿的损害范围？

第三人利用侵权内容造成损害的典型表现是人工智能使用者利用自动生成的内容实施加害行为导致权利人合法权益受损，例如将侵权内容转发而导致损害后果扩大等。此时，利用生成内容实施侵权行为的加害人应对其行为承担侵权责任。服务提供者是否要对加害人造成的损害承担赔偿责任？《中华人民共和国民法典》（以下简称《民法典》）第1175条规定：“损害是因第三人造成的，第三人应当承担侵权责任。”故当造成损害后果的原因是第三人的加害行为时，应由第三

〔12〕 支持意见参见张新宝主编：《〈中华人民共和国个人信息保护法〉释义》，人民出版社2021年版，第525-526页。不同意见参见李昊：《个人信息侵权责任的规范构造》，载《广东社会科学》2022年第1期。

〔13〕 张新宝：《侵权责任构成要件研究》，法律出版社2007年版，第123页。

〔14〕 参见徐伟、俞秀芝：《论合法公开个人信息处理规则的适用》，载《厦门大学法律评论》总第38辑，厦门大学出版社2024年版，第34-54页。

人承担责任，即不应将第三人侵权造成的损害纳入服务提供者损害赔偿范围。<sup>〔15〕</sup>

另一值得探讨的损害类型是风险性损害。风险性损害系近年来在个人信息侵权领域热议的一种损害类型。<sup>〔16〕</sup>关于风险性损害的含义，存在不同界定。狭义风险性损害是指个人信息泄露等导致的潜在的下游人身和财产损害，例如因个人信息被用于电信网络诈骗而遭受的损害。广义的风险性损害除了潜在的下游损害外，还包括信息主体为了预防损害而支出的合理费用、因潜在的损害而引发的精神焦虑等。<sup>〔17〕</sup>此处对广义风险性损害在生成式人工智能侵权中的适用展开分析。

第一，预防侵权而支出的合理费用应予赔偿。既有理论和实践都支持预防性费用可赔偿。生成式人工智能侵权中的预防性费用亦然。权利人根据《民法典》第1195条向服务提供者首次发送“通知”要求其采取必要措施的费用一般不宜纳入可赔偿的损害范围，因为在通知前服务提供者往往没有过错，不构成侵权。但若通知合格且服务提供者未及时采取必要措施，则后续的通知费用应纳入损害范围。

第二，潜在的下游人身和财产损害不应纳入损害范围。当下游损害是第三人直接造成时，即便该损害已确定发生，该损害系因第三人造成而应由第三人承担责任，因为第三人是介入因素而中断服务提供者与下游损害结果间的因果关系。举重以明轻，下游损害是否发生不确定时，亦不应将其纳入服务提供者应承担的损害范围。此外，鉴于潜在的损害不纳入损害范围，因潜在的损害而引发的精神焦虑也不宜纳入损害范围。这与个人信息领域不少学者主张潜在的下游损害应纳入损害范围不同。

第三，与个人信息领域不同，生成式人工智能场景中还存在另一种风险性损害，即人工智能在生成侵权内容后，存在将来再次生成侵权内容的风险。在受害人起诉时，能否将未来可能再次生成侵权内容的损害纳入赔偿范围？根据损害赔偿原理，损害要具备“确定性”。为此，传统上对损害的赔偿原则上仅限于已经发生的损害。对于尚未发生的损害，除非有证据表明该损害很可能发生，否则不会事先作出赔偿。故生成式人工智能侵权中的损害能否赔偿，取决于侵权内容再次生成的几率。

若侵权内容再次生成的几率较高，则可在尚未发生之时给予赔偿。<sup>〔18〕</sup>生成式人工智能作为尚处于快速发展过程中的新兴技术，同一内容的生成几率难以给出确定性的判断，因为会受到诸多因素的影响，包括人工智能模型设计、被侵害的权利类型、<sup>〔19〕</sup>服务提供者在知道侵权内容后

〔15〕 参见徐伟：《生成式人工智能侵权中因果关系认定的迷思与出路》，载《数字法治》2023年第3期。

〔16〕 支持风险性损害可赔偿的观点参见朱晓峰、夏爽：《论个人信息侵权中的损害》，载《财经法学》2022年第4期。反对观点参见程啸、曾俊刚：《个人信息侵权的损害赔偿》，载《云南社会科学》2023年第2期。

〔17〕 参见谢鸿飞：《个人信息泄露侵权责任构成中的“损害”——兼论风险社会中损害的观念化》，载《国家检察官学院学报》2021年第5期。

〔18〕 对尚未发生的损害予以赔偿时，是采未来损害必然会发生的“实质性风险”（substantial risk）标准，还是采未来损害发生几率较高的“客观合理可能性”（objectively reasonable likelihood）标准，存在分歧。See Patrick J. Lorio, *Access Denied: Data Breach Litigation, Article III Standing, and A Proposed Statutory Solution*, 51 *Columbia Journal of Law and Social Problems* 79, 91-103 (2017). 从我国现行法来看，似采“实质性风险”标准。此处为全面讨论问题的需要，以“客观合理可能性”标准为讨论的前提。若采“实质性风险”标准，则因侵权内容并非必然会生成而不必纳入损害范围。此外，即便采“实质性风险”标准，侵权内容再次生成的几率仍可能是影响法院酌定赔偿金的因素之一。

〔19〕 例如，著作权领域的过滤技术日臻成熟，避免侵害著作权内容生成的几率在技术层面往往高于避免侵害其他权利类型内容生成的几率。参见崔国斌：《论算法推荐的版权中立性》，载《当代法学》2024年第3期。

采取的措施，等等。故法院须借助技术人员的协助来判断致害内容再次生成的几率。<sup>[20]</sup>此外，法院也可将案涉人工智能是否曾多次生成同一侵权内容作为参考因素。若结论是再次生成的几率较高且难以通过采取必要措施来降低，则应支持受害人对未来损害的请求。若几率较低，则没有必要在再次生成前对受害人给予赔偿。据此，生成式人工智能侵权中法院须考虑未来侵权内容再次生成的几率来决定是否支持风险性损害。一种可能的质疑是，风险性损害中侵权内容是否会再次生成难以判断，因为即便在诉讼时得出侵权内容再次生成的几率较高，该可能性也会因技术的改进、人工智能的自主学习等而发生变化，故无从判断未来侵权内容再次生成的几率。这一质疑难以成立，因为损害的计算时点以庭审时为准。故生成式人工智能未来随着技术改进等原因而引发的侵权内容生成几率的变化，不影响对当前案件损害的认定。

### 三、生成式人工智能服务提供者侵权中的财产损害赔偿

在我国，财产损失与精神损害的区分标准是损害能否以金钱加以衡量。所有能以金钱加以衡量的损害都是财产损失。<sup>[21]</sup>

#### （一）人格权侵权中的财产损害赔偿

人格权侵权的财产损害赔偿规则主要规定于《民法典》第 1182 条。根据该条，生成式人工智能人格权侵权中的财产损害赔偿，首先须考虑被侵权人的损失或侵权人的获利，并由法院酌定来兜底。

第一，关于被侵权人的损失。对于侵害人格权而造成的调查取证费等合理费用，应予赔偿，当无疑问。对于其他情形，是否存在损失及损失如何计算则有待厘清。一般而言，被侵权人的损失主要指未经其同意而商业化利用其人格而造成的损失，包括许可使用费的丧失、丧失对人格权商业化利用的潜在机会、社会评价降低导致人格利益经济价值的降低，等等。（1）侵害隐私权和名誉权是生成式人工智能侵权的重要情形。隐私权、名誉权等非标表型人格权一般不允许商业化利用，自然也不存在许可使用费的丧失，但其可能导致受害人社会评价降低进而减损其肖像等其他人格利益的经济价值。（2）对于可以进行商业化利用的人格权，若权利人没有商业化利用其人格权意愿，则权利人也难以证明服务提供者对其造成了财产损失。对于普通大众而言，多没有商业化利用其人格权的计划。即便是公众人物，也可能明确拒绝某种类型的商业化利用。例如，演员约翰逊（Johansson）曾数次拒绝了 OpenAI 希望将其声音用于 GPT-4o 内置的 Sky 语音的邀请。<sup>[22]</sup>（3）能进行商业化利用且权利人有意进行商业化利用的人格权，如何计算该损失金额是实践中的难题。对于公众人物人格权商业化，尽管存在传统场景中形成的许可使用费等“市场

[20] 有学者认为，人工智能引发的风险难以事先评估。参见林涓民：《论人工智能立法的基本路径》，载《中国法学》2024年第5期。

[21] 参见徐建刚：《财产损失计算的价值基准与规范表达》，载《法学》2024年第6期。

[22] See Bobby Allyn, *Scarlett Johansson Says She Is 'Shocked, Angered' over New ChatGPT Voice*, available at <https://www.npr.org/2024/05/20/1252495087/openai-pulls-ai-voice-that-was-compared-to-scarlett-johansson-in-the-movie-her>, last visited on Nov. 20, 2024.

价”经验，但该经验在人工智能场景中难以适用，因为二者的利用场景差别较大。前者系侵权人利用肖像等标表型人格标识来推广自身业务以获利，后者是服务提供者利用提供生成服务来获利，二者的“计价”应有明显差别。对于没有“市场价”经验的多数普通大众而言，其损失更难计算。

第二，关于侵权人的获利。基于原《中华人民共和国侵权责任法》第20条与《民法典》第1182条在条文表述上的变化，宜认为在认定财产损失时被侵权人的损失和侵权人的获利是并列关系，权利人可自行选择采用何种方式来计算财产损失。据此，即便被侵权人难以证明自身损失的存在，其也可以通过证明侵权人的获利来主张赔偿。关于获利赔偿请求权的性质，存在侵权请求权说、不当得利返还请求权说、无因管理请求权说、独立请求权说等。此处采侵权请求权说展开分析。<sup>[23]</sup> 基于侵权请求权的定性，侵权人的获利并非指侵权人获利的总额，而是指与违法利用被侵权人人格权具有因果关系的获利部分，即侵权人因利用了被侵权人的人格权而在商业利益上的增量。在生成式人工智能场景中，这一“增量”难以准确计算，不仅是因为服务提供者的获利总额难以计算，而且生成的侵权内容在该总额中发挥了多大的作用也难以判断。侵权人获利的另一计算方式是采用“对价节省”，即被侵权人的人格权被商业利用时，侵权人本应支付却未支付而省下的费用。此时，可以按照拟制的许可使用费标准来确定侵权人的获利金额。但这在生成式人工智能场景中同样难以计算，因为传统人格权商业化利用场景形成的“市场价”不宜适用于生成式人工智能场景。综上，在生成式人工智能场景中，侵权人的获利尽管理论上确实存在，但在客观上难以准确计算。

第三，关于法院酌定。就酌定的标准，有意见认为，宜以损害预防类参考因素为基础进行酌定，“一般应当根据拟制的许可使用费以及行为人的获利数额确定大致的赔偿数额，然后再基于对受害人损害的填补与惩罚不法行为人等因素的考量，对赔偿数额进行一定的调整，以确定最终的赔偿数额”<sup>[24]</sup>。关于许可使用费和行为人获利，已如前述。此处就其他因素加以说明。《民法典》第998条规定：“认定行为人承担侵害除生命权、身体权和健康权外的人格权的民事责任，应当考虑行为人和受害人的职业、影响范围、过错程度，以及行为的目的、方式、后果等因素。”一般认为，本条是将动态系统论适用于人格权侵权领域，本条列举的因素不仅适用于侵权责任的构成，也适用于侵权责任的承担。<sup>[25]</sup> 故上述所列因素可作为酌定赔偿时的考量因素。

在生成式人工智能场景中，就行为人的职业而言，服务提供者从事的是人工智能服务，这是国家积极鼓励的产业，甚至是战略性产业。就侵权影响范围而言，人机对话模式意味着只有参与对话的使用者才能了解到侵权内容，即影响范围往往有限。就过错程度而言，一般而言，服务提供者不会主动追求侵权内容的生成，且受害人要举证证明服务提供者存在主动追求侵害的故意也困难重重，故服务提供者多不存在直接故意。从实践来看，服务提供者的过错多表现为违反保护

[23] 相关争议参见王叶刚：《论人格权擅自商业化利用中的获利赔偿请求权》，载《法学评论》2016年第4期。相关学说的选择，除了获利赔偿请求权与何种请求权的理念、构成要件和法律后果更契合外，最主要的差别在于理论上如何处理行为人获利远超权利人许可使用费案型，即此时是否应允许行为人保有该获利。

[24] 王叶刚：《论侵害人格权益财产损失赔偿中的法院酌定》，载《法学家》2021年第3期，第115页。

[25] 参见王利明：《民法典人格权编中动态系统论的采纳与运用》，载《法学家》2020年第4期。

性规范，例如违反《生成式人工智能服务管理暂行办法》要求的“使用具有合法来源的数据和基础模型”（第7条）、“制定符合本办法要求的清晰、具体、可操作的标注规则”（第8条）等。<sup>〔26〕</sup> 尽管服务提供者能够认识到这些义务的违反可能导致侵权内容的生成，但不应认定其构成间接故意，因为“成立故意必须认识到足以构成不法的全部客观事实，认识到行为违反规章制度，并不表明行为人一定认识到了危害结果发生，更不表明行为人希望或者放任危害结果发生”<sup>〔27〕</sup>。对服务提供者而言，其只具有认识到侵权内容有可能生成的概括预见能力，无法认识到某一特定侵权内容是否会生成，故其不满足间接故意的认识因素。据此，服务提供者在生成阶段的过错不存在故意，侵权内容的生成系过失引发。唯有在移除阶段，即服务提供者收到侵权通知后未及时采取必要措施时，其过错可能成立故意。就行为目的、方式和后果而言：服务提供者行为的直接目的系提供人工智能服务，并不存在侵害权利人的目的；通过人工智能生成内容并以人机对话方式提供给使用者本身并无不妥之处；人机对话方式的传播范围有限，故侵权内容对受害人造成的损害后果往往并不严重。综上，就法院酌定的考量因素来看，一般而言生成式人工智能场景中的酌定赔偿金不宜过高。

最后，鉴于损失金额在实践认定上的困难，在生成式人工智能侵权中是否有必要引入法定赔偿？就此，个人信息保护领域的经验可供参考。在个人信息侵权中，也存在受害人的财产损失金额难以判断问题。为此，我国台湾省“个人资料保护法”（2015年修正）第28条第3款采取了“每人每一事件新台币五百元以上二万元以下计算”的法定赔偿。在我国《个人信息保护法》起草过程中，有意见认为可按照每一起侵权事件向每位受害人赔偿500元至1000元。但考虑到侵害个人信息案件往往涉及人数众多，若作此规定，可能给个人信息处理者造成极大的负担，故立法机关没有采纳。<sup>〔28〕</sup> 尽管立法未采法定赔偿，但仍有学者认为，以担心个人信息处理者赔偿数额巨大为由否定法定赔偿制度的设立并不合理，因为法定赔偿制度的关键在于固定个人信息财产价值的范围，为法院裁判提供参考，起到补偿与预防的双重功能，且个人信息处理者并非对所有损害都予以赔偿。<sup>〔29〕</sup> 可见，法定赔偿制度是否会导致个人信息处理者承担巨额赔偿金是影响规则选择的主要考量。

在生成式人工智能场景中，这一担忧不会发生。其一，与个人信息领域常见的大规模收集或泄露个人信息不同，生成式人工智能中一般不会出现大规模生成侵害人格权内容现象，因为内容的生成有赖于使用者的主动提问。其二，即便出现了大规模生成侵权内容，此类侵权多发生于生成的是相似的侵权内容，侵害的是同一受害人的情形，而非众多受害人。其三，即便受害人数量众多，鉴于发现成本和诉讼成本的存在，真正提起诉讼的受害人比例也不会高。最终提起诉讼主张赔偿的，是对自身人格权受侵害较为重视（往往也是受侵害程度较重、损害较大）的受害人。为此类受害人提供一定金额的法定赔偿，保障其不会因损失举证困难而无法获得赔偿，既有利于

〔26〕 关于涉生成式人工智能的保护性规范，参见徐伟：《生成式人工智能服务提供者侵权过错的认定》，载《法学》2024年第7期。

〔27〕 张明楷：《刑法学》（第6版，上册），法律出版社2021年版，第346页。

〔28〕 参见程啸：《个人信息保护法理解与适用》，中国法制出版社2021年版，第521页。

〔29〕 参见王雪：《个人信息双重风险性损害判定与赔偿框架的重塑》，载《人权》2024年第5期。

保护受害人，也不会对服务提供者造成严重负担。因此，在生成式人工智能人格权侵权中增设法定赔偿规则应是可取选择。

综上，既有理论关于人格权侵权的财产损害赔偿，系以人格权商业化利用为典型事实基础而构建。生成式人工智能侵权的特殊之处在于，传统上公众人物才有机会将人格权商业化利用，在生成式人工智能场景中所有人的的人格权都可能被商业化利用。同时，传统上基于市场机制而形成的商业化利用定价经验，在生成式人工智能场景中因场景差异较大而不具有参考价值。这意味着权利人的财产损失更加难以计算，法院酌定成为实践中的主要计算方式。

## （二）知识产权侵权中的财产损害赔偿

根据《中华人民共和国著作权法》（2020年修正，以下简称《著作权法》）第54条、《中华人民共和国专利法》（2020年修正）第71条和《中华人民共和国商标法》（2019年修正）第63条，侵害知识产权的财产损害赔偿计算方式包括实际损失、侵权获利、许可使用费（或其倍数）和法定赔偿。<sup>[30]</sup>这四种方式在生成式人工智能场景中的适用，说明如下。

第一，关于实际损失。权利人的实际损失包括权利人因侵权而遭受的可得利益损失、制止侵权而支出的合理费用等。制止侵权而支出的合理费用可以赔偿并无疑问。关于可得利益损失，根据《最高人民法院关于审理著作权民事纠纷案件适用法律若干问题的解释》（2020年修正）第24条，在著作权侵权中可以根据权利人因侵权所造成复制品发行减少量或者侵权复制品销售量与权利人发行该复制品单位利润乘积计算；发行减少量难以确定的，按照侵权复制品市场销售量确定。在生成式人工智能场景中，相关作品发行量的“此消彼长”往往难以衡量，因为服务提供者未以传统的方式复制和销售作品。相关作品的生成次数能否相当于侵权复制品销售量亦存疑，因为二者在市场中的呈现方式和对消费者的价值有所不同。故权利人的实际损失多难以计算。

第二，关于侵权获利。根据《最高人民法院关于审理侵犯专利权纠纷案件应用法律若干问题的解释》第16条第1款等，侵权获利指侵权人因侵害知识产权而额外获得的利益，并非指侵权人获得的所有利益。而根据《最高人民法院关于审理商标民事纠纷案件适用法律若干问题的解释》（2020年修正）第14条，侵权获利的计算方式可通过“侵权商品销售量与该商品单位利润乘积计算；该商品单位利润无法查明的，按照注册商标商品的单位利润计算”。在生成式人工智能场景中，侵权内容的生成次数在技术层面有可能在一定程度上查明。例如，根据OpenAI用户协议，为了监督使用者滥用行为，OpenAI会保存30天用户的使用记录。<sup>[31]</sup>故可以通过查找此类保存的内容来了解侵权内容的生成次数。但“单位利润”目前尚无“市场价”可资参考，因为销售侵权商品的利润与许可生成内容的利润不同。该“单位利润”有待市场的逐渐形成。

在“单位利润”无从得知的情况下，若以服务提供者整体上的利润来计算，则面临的难题是如何确定侵权内容对服务提供者利润的贡献程度。关于侵权获利分摊的计算，至少有四种衡量方

[30] 这四种赔偿方式的适用顺序略有不同。在著作权和专利权侵权中，实际损失和侵权获利并列为首选计算方式，在商标侵权中实际损失优先于侵权获利作为赔偿的计算方式。

[31] See *How do I Turn Off Chat History and Model Training?*, available at <https://help.openai.com/en/articles/8983082-how-do-i-turn-off-chat-history-and-model-training>, last visited on Nov. 20, 2024.

法，即量化比例法、替代品比较法、消费者调查法和定性分析法。<sup>〔32〕</sup>其中，量化比例法系根据侵权内容占整个侵权产品的量化比例来确定侵权内容对侵权产品的利润的贡献比例。若采此，侵权内容占有生成内容的比例甚微。替代品比较法主要适用于专利侵权案件，系将专利产品和不含专利但可获得类似效果的产品进行比较所得的利润差额作为该专利所带来的利润。消费者调查法系调查消费者愿意为含专利的产品比不含专利但其他方面相同的产品多支付的价格作为专利带来的贡献。这两种方法都借助比较侵权产品和功能相似但不侵权产品来判断侵权在产品中的贡献。生成式人工智能场景中需比较的是允许生成侵权内容和无法生成侵权内容的人工智能。例如，侵权人并未采取著作权过滤等技术而允许人工智能生成侵害著作权的图片，而作为比较的人工智能不提供图片生成功能。但这样的比较只能得出“所有”图片的生成对服务提供者利润的贡献程度，无法得出作为原告享有权利的“某一”侵权图片对服务提供者贡献的大小。因此，替代品比较法和消费者调查法在生成式人工智能场景中也难以适用。据此，定性分析法仍是生成式人工智能场景不得不采取的方法，即根据侵权内容和其他因素对利润的贡献程度来酌定。

第三，关于许可使用费。<sup>〔33〕</sup>目前尽管知识产权人与生成式人工智能服务提供者已尝试就利用知识产权的许可使用费达成合作，但许可使用费定价等问题至今尚无共识。例如，美国《纽约时报》曾尝试与 OpenAI 公司就后者使用前者享有著作权的报道内容达成协议，但最终未能在许可费用上达成共识，进而诉诸法院，主张 OpenAI 公司构成侵权。<sup>〔34〕</sup>据此，目前采许可使用费的计算方式尚存在并无市场价可资参考的难题。

第四，关于法定赔偿。<sup>〔35〕</sup>采法定赔偿时，法院会将与损害相关的诸多因素纳入考量，综合确定赔偿金额。影响赔偿金的因素包括行为人的侵权方式、过错程度、被侵害内容的类型和知名度、侵害后果的大小、侵权持续时间等。在生成式人工智能场景中，服务提供者在知道侵权后采取的措施亦是影响赔偿金的重要因素之一。就生成式人工智能场景而言，服务提供者的侵权方式是通过人工智能自动生成侵权内容。服务提供者在生成阶段的过错系过失。人机交互模式导致侵权后果往往有限。据此，侵权的损害赔偿金一般不宜高。例如“新创华公司案”法院最终判决了一万元赔偿金（含合理开支）。<sup>〔36〕</sup>若该案中被告生成侵权内容（奥特曼图片）的次数不多，则该一万元赔偿金或许具有一定的惩罚性色彩。<sup>〔37〕</sup>但知识产权惩罚性赔偿以“情节严重”为要件，结合生成式人工智能侵权场景而言，一般不宜课以惩罚性赔偿。

综上，传统场景和生成式人工智能场景侵害知识产权的侵权行为方式存在较大差别，故既有的知识产权许可使用费市场价在生成式人工智能场景中的参考价值有限。此外，传统场景中的知

〔32〕 参见刘晓：《论知识产权损害赔偿中侵权获利的分摊方法》，载《法律科学（西北政法大学学报）》2018年第4期。

〔33〕 在人格权侵权中，人格权商业化利用的许可使用费多纳入被侵权人的损失或侵权人的获利中讨论，但在知识产权侵权中，基于现行法的规定，许可使用费与被侵权人的损失、侵权人的获利并列。

〔34〕 See *New York Times Co. v. Microsoft Corp.*, No. 23-cv-11195.

〔35〕 我国在知识产权领域采取的是“法定赔偿”，在人格权领域采取的是“酌定赔偿”。二者的差别在于，前者存在赔偿金额的上限和下限，后者则无限制。

〔36〕 参见广州互联网法院（2024）粤0192民初113号民事判决书。

〔37〕 我国知识产权侵权虽然规定了惩罚性赔偿，但法院很少直接适用惩罚性赔偿条款，而是通过具有惩罚性色彩的法定赔偿来实现惩罚功能。参见焦和平：《知识产权惩罚性赔偿与法定赔偿关系的立法选择》，载《华东政法大学学报》2020年第4期。

知识产权侵权多为行为人利用了权利人的知识产权来获利，例如制造并销售权利人享有著作权的作品等。对权利人知识产权的违法利用往往会明显提升行为人的获利。但在生成式人工智能场景中，生成的侵权内容虽然会在一定程度上提升服务提供者的获利，但“某一特定权利人”的知识产权对服务提供者获利的提升效果往往较小。这导致传统知识产权的损害赔偿计算方式在生成式人工智能场景中面临挑战。法定赔偿仍是生成式人工智能侵权中确定赔偿金的主要方式。

#### 四、生成式人工智能服务提供者侵权中的精神损害赔偿

我国现行法中并未明确规定知识产权侵权的精神损害赔偿。实务中支持知识产权人精神损害赔偿的判决亦不多见，原因在于：一方面，涉知识产权侵权的权利人多为企业等非自然人，而我国精神损害赔偿的权利人限于自然人；另一方面，在知识产权人是自然人的案件中，侵害知识产权中的人身权多通过赔礼道歉、消除影响等方式予以救济。故本文对侵害知识产权的精神损害赔偿不作探讨。此处的精神损害赔偿聚焦于侵害人格权的情形。

##### （一）精神损害赔偿构成的判断

依据《民法典》第1183条，我国精神损害赔偿的构成要件之一是“造成严重精神损害”。对于如何解释“严重精神损害”存在不同意见，主要的分歧在于是否应将严重精神损害作为限制精神损害赔偿的要件。<sup>〔38〕</sup>“严重性”要件的争论，在根本上是关于“精神损害赔偿制度究竟是让被侵权人更容易得到救济为好，还是更应防止该制度被滥用进而妨碍‘侵权人’的行动自由为妙”<sup>〔39〕</sup>。在价值选择方面，双方的分歧并没有那么大。主张取消“严重性”要件的观点，并非支持精神损害都可赔偿，轻微精神损害不予赔偿仍是共识，一般精神损害是否应予赔偿仍需视情况而定。主张保留“严重性”要件的观点，也存在降低严重性认定标准的趋势。因此，双方观点上的差异与其说是严重性要件存废之争，不如说是论证说理的路径之别。本文无意陷入该争论，而是聚焦于影响精神损害赔偿的考量因素（无论该考量因素是否置于“严重性”要件之下），探讨生成式人工智能场景是否应课以精神损害赔偿。

《最高人民法院关于确定民事侵权精神损害赔偿责任若干问题的解释》（2020年修正）第5条规定：“精神损害的赔偿数额根据以下因素确定：（一）侵权人的过错程度，但是法律另有规定的除外；（二）侵权行为的目的、方式、场合等具体情节；（三）侵权行为所造成的后果；（四）侵权人的获利情况；（五）侵权人承担责任的经济能力；（六）受理诉讼法院所在地的平均生活水平。”尽管本条规定的是精神损害赔偿数额的考量因素，但其中的前三项亦可作为判断精神损害成立与否（“严重性”要件满足与否）的因素。<sup>〔40〕</sup>据此，结合生成式人工智能场景，前三项因素分析如下。

第一，侵权人的过错程度。侵权人的过错程度越高，越可能导致严重精神损害。<sup>〔41〕</sup>在生成

〔38〕 支持意见参见程啸：《侵权责任法》（第3版），法律出版社2021年版，第230页。反对意见参见张新宝：《从司法解释到侵权责任法草案：精神损害赔偿制度的建立与完善》，载《暨南学报（哲学社会科学版）》2009年第2期。

〔39〕 李东宇：《论侵害个人信息权益的精神损害赔偿》，载《财经法学》2023年第4期，第140页。

〔40〕 参见邹海林、朱广新主编：《民法典评注：侵权责任编》（第1册），中国法制出版社2020年版，第219-220页。

〔41〕 关于侵权人的过错程度是否影响精神损害赔偿的成立，多数采肯定意见。参见张新宝：《侵权责任构成要件研究》，法律出版社2007年版，第243页；另见程啸：《侵权责任法》（第3版），法律出版社2021年版，第862页。

式人工智能场景中，侵权人的过错程度可区分生成阶段和移除阶段来分别判断。在生成阶段，若服务提供者履行了法律要求的各项预防性义务，则此时侵权内容的生成系因人工智能技术本身的难以预见性特征所致，难谓服务提供者有故意或重大过失。从实践来看，鉴于我国要求生成式人工智能在向公众提供服务前要按照国家有关规定开展安全评估，并履行算法备案手续，故实践中多数服务提供者系因过失，而非故意导致了侵权内容的生成。在移除阶段，服务提供者的过错程度取决于其在收到合格通知后采取了何种措施。若服务提供者应采取而未采取任何措施，其过错程度较高；若服务提供者及时采取了必要措施，此时难谓其存在过错。须指出的是，服务提供者负有采取必要措施的义务并不意味着其要确保侵权内容必然不会再次生成。是否履行了必要措施义务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。<sup>[42]</sup>除上述两种情形外，还存在服务提供者采取的措施能够在一定程度上避免侵权内容再次生成的情形，例如只屏蔽了部分可能生成侵权内容的提示词。当服务提供者采取的措施能在一定程度上避免侵权内容生成时，其过错程度需视采取的措施类型、实施效果等因素来综合判断。

第二，侵权行为的目的、方式、场合等具体情节。侵权行为的具体情节既体现出了侵权人的主观可非难性，也在一定程度上影响了受害人的损害程度。生成式人工智能服务提供者的目的在于提供人工智能服务，并不存在侵害权利人的故意。服务提供者系通过人机对话的方式自动生成了侵权内容，该行为方式并无不当。关于行为场合，一般而言，公开场合侵害人格权比非公开场合侵害人格权的性质严重。生成式人工智能侵权固然应认定为公开场合，但相较于传统网络侵权，生成式人工智能侵权的公开程度更低，因为人工智能与使用者间的对话内容并不向第三人公开。

第三，侵权行为所造成的后果。侵权行为造成的后果在人格权领域主要表现为信息泄露和社会评价降低等。一般而言，单纯的信息泄露本身并不会造成严重的后果，因信息泄露而造成的社会评价降低或下游损害才是造成严重后果的主要原因。生成式人工智能场景中的社会评价降低可分为两个方面。其一，人工智能本身对受害人的不当评价不宜作为社会评价降低的理由。在人工智能生成对受害人的不实评价时，人工智能的“评价”与传统社会中加害人对受害人的评价并不相同。因为就技术层面而言，人工智能只是学会了表达，却并不理解文字本身的含义。<sup>[43]</sup>而且我国现行法也并未认可人工智能具有主体地位，<sup>[44]</sup>故不宜将人工智能对受害人的评价等同于传统社会中民事主体对受害人的评价。其二，使用者对受害人的评价可能因人工智能生成侵权内容而降低，评价降低的程度与生成的内容相关。同时，人机对话模式、人工智能标识以及对生成内容未必准确的声明会降低损害后果。

综上，除了移除阶段服务提供者在知道侵权内容后未采取任何措施来避免侵权内容再次生成外，实践中多数情况下服务提供者的过错程度不高，侵权行为的具体情节也难谓恶劣，直接造成

[42] 参见徐伟：《论通知规则在生成式人工智能侵权中的适用》，载《现代法学》2024年第3期。

[43] 即使是最先进的语言模型也只是在模仿人类语言的表达方式，而没有真正理解其含义。See Emily M. Bender & Alexander Koller, *Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data*, available at <https://aclanthology.org/2020.acl-main.463.pdf>, last visited on Nov. 20, 2024.

[44] 参见韩旭至：《人工智能法律主体批判》，载《安徽大学学报（哲学社会科学版）》2019年第4期。

的后果往往并不严重，故多数情况下不宜认定为构成“严重精神损害”。因此，总体而言，生成式人工智能侵权场景中多不成立精神损害赔偿。

## （二）消除影响、恢复名誉和赔礼道歉的适用

服务提供者不成立精神损害赔偿并不意味着受害人没有遭受精神痛苦，也并不意味着服务提供者对受害人的精神痛苦不承担任何责任。在情节轻微时，受害人可要求加害人承担消除影响、恢复名誉和赔礼道歉等责任。<sup>〔45〕</sup>

在生成式人工智能侵权中，应具体采取何种方式来消除影响和恢复名誉？就规范目的而言，应结合加害行为的场景而采取最有助于消除负面影响和恢复受害人名誉的方式。例如，在报刊作为主要传媒的时代，消除影响、恢复名誉的常见方式是要求加害人在报刊上登报澄清，这是我国司法裁判主要采取的方式。进入网络时代，针对网络侵权的消除影响、恢复名誉，主要表现为要求网络服务提供者在其网站首页显著位置或者在实施加害行为的网络用户页面作出澄清。<sup>〔46〕</sup>故消除影响和恢复名誉的最佳方式是向最有可能接触到侵权内容的主体发布澄清声明。

在生成式人工智能场景中，基于人工智能与使用者间一对一的人机互动关系，最有助于消除影响和恢复名誉的方式应是向曾收到了侵权内容的使用者主动推送澄清声明。但这一方式在技术层面是否可行、成本是否可堪负荷亦须纳入考量。若生成式人工智能服务提供者能以合理的成本确定曾收到相同或相似侵权内容的使用者，则向其推送澄清声明应是消除影响、恢复名誉的最佳方式。鉴于服务提供者多会保存一定期限内的人机对话记录，故仅就技术层面而言，服务提供者应有能力判断出某一期限内接触过侵权内容的使用者并向其推送澄清声明。

若在技术层面难以精准定位或因法律障碍（例如服务提供者基于对隐私的保护而无权查阅使用者的人机对话记录）无法确定哪些使用者曾收到过相同或相似的侵权内容，则退而求其次的方式可以是在服务提供者的网站上发布相关声明来澄清。就此，可以参考电子商务领域的经验。在电子商务场景中，存在电子商务经营者利用平台销售假货等侵害知识产权现象。对此，商务部发布的《电子商务信息公示管理办法（征求意见稿）》第21条至第24条规定：电子商务平台经营者在收到知识产权权利人认为其知识产权受到侵害的通知后，经核查符合相关规定的，四十八小时内专门区域公示该通知，公示中应当包含知识产权权利人和侵权人的法定名称；电子商务平台经营者在收到平台内经营者提交不存在侵权行为的声明后，应当在四十八小时内专门区域公示该声明，公示行为应在文件名称或链接名称中加入“知识产权疑似侵权通知”字样。从企业实践来看，电子商务平台经营者也多会专门开辟页面来公示平台内经营者的违规行为。<sup>〔47〕</sup>类似的，生成式人工智能场景中，服务提供者也可通过专门开辟页面来公示人工智能生成的侵权内容，并以就真实情况作出澄清的方式来消除影响、恢复名誉。

〔45〕 尽管我国精神损害赔偿的权利人仅限于自然人，但请求消除影响、恢复名誉和赔礼道歉的权利人并不限于自然人，也并不限于人格权侵权。知识产权侵权亦可提出此类请求。故此处的讨论同时适用于人格权和知识产权。唯一的差别是，人格权领域的消除影响、恢复名誉责任承担方式，在知识产权领域多称之为消除影响，例如《著作权法》第52条等。

〔46〕 例如，在加害人在其微信朋友圈发布不实内容后，法院要求加害人在其朋友圈发布致歉内容。参见吉林省辽源市中级人民法院（2017）吉04民终97号民事判决书。

〔47〕 例如，淘宝网会将违规卖家公示。参见《违反平台规则行为公示》，载 <https://rulechannel.taobao.com/#/rulePublicity>，最后访问时间：2024年11月20日。

关于赔礼道歉，尽管赔礼道歉是否应作为一项民事责任承担方式在理论上存在分歧，<sup>[48]</sup>但司法裁判多支持赔礼道歉作为责任承担方式。《民法典》第1000条第1款规定：“行为人因侵害人格权承担消除影响、恢复名誉、赔礼道歉等民事责任的，应当与行为的具体方式和造成的影响范围相当。”就生成式人工智能场景而言，服务提供者过错多为过失，而赔礼道歉多适用于行为人故意实施加害行为的场合，<sup>[49]</sup>此外，仅使用者能接触到侵权内容导致影响范围有限。这些共同决定了在消除影响、恢复名誉等责任承担方式已能够弥补受害人损害的情况下，不宜再适用赔礼道歉。

## 五、结 语

基于上述，相较于损害赔偿一般原理，生成式人工智能侵权损害赔偿存在以下一些特点。

第一，改变了事实基础。传统损害赔偿制度立基的事实基础是：在侵害发生后，通过采取一定的措施，可以有效避免侵害的再次发生或将再次发生的几率降至较低；在此基础上，通过金钱赔偿、消除影响等来使受害人恢复至假设损害未发生的状态。这一事实基础在人工智能场景中发生了变化，即侵害发生后，有时难以通过必要的措施来有效避免侵害的再次发生。这一变化源于人工智能技术的“自主性”特征，即人类难以完全控制人工智能的举动。在此前提下，若仅赔偿已经发生的损害，则无法解决未来的侵害问题，而避免未来再次侵害的发生恰恰是生成式人工智能场景中受害人最核心的诉求。因此，对未来可能发生的损害，是否有必要以及如何在本次纠纷中加以救济，是生成式人工智能侵权面临的难题之一。更一般性而言，在人工智能带来的不确定性时代，如何回应未来“概率性”侵害问题，系理论上亟待回答的问题。

第二，冲击了价值选择。传统上，“损害赔偿之最高指导原则在于赔偿被害人所受之损害，俾于赔偿之结果，有如损害事故未曾发生者然”<sup>[50]</sup>。但为了强化对受害人的救济并抑制侵权行为，理论上逐渐倾向于加重行为人的责任，例如在权利受侵害时直接推定受害人有损害，以行为人的获利作为损失的计算方式、即便该获利可能超过了受害人的损失，等等。但在生成式人工智能场景中，加重行为人责任的价值取向受到冲击，因为生成式人工智能的技术特征和服务模式表明，不应对其课以较重的赔偿金。因此，生成式人工智能场景的侵权损害赔偿规则呈现出对服务提供者“较为有利”（赔偿金不高），而对受害人“显得不利”（无法通过高额赔偿金来促使服务提供者采取措施避免侵权内容再次生成）的局面。这与当前强化受害人救济的价值选择有所不符。对已不完全在人类控制之内的人工智能侵权而言，将侵权发生几率降低到公众能接受的范围内，接受人工智能发展与侵权的“共生”状态，或许是不得不面对的局面。

第三，影响了治理逻辑。传统侵权多侵害“某一”权利人，故损害赔偿金额可以通过衡量行为人为人侵害“该权利人”获得的利益来判断。但生成式人工智能服务提供者获利的特殊之处在于，

[48] 参见葛云松：《民法上的赔礼道歉责任及其强制执行》，载《法学研究》2011年第2期。

[49] 参见张平华：《民法典侵权责任编应处理好的三对关系》，载《财经法学》2018年第6期。

[50] 曾世雄：《损害赔偿法原理》，中国政法大学出版社2001年版，第16页。

其侵害的权利人众多但单一权利人的损害较小，即“大规模微型侵权”。故当“某一”权利人主张赔偿时，其能证明和获得的损害赔偿金额有限。此外，提起诉讼的受害人比例在整体受害人中只占一小部分。据此，传统损害赔偿理论中试图通过“获利剥夺”乃至“惩罚性赔偿”的方式来遏制侵权行为在生成式人工智能场景中难以实现其效果。此时，唯有通过公私法协同在损害赔偿之外对服务提供者课以行政责任的方式，才有可能促使服务提供者积极采取措施减少侵权。换言之，对大规模微型侵权而言，单纯的私法损害赔偿已经难以发挥预防侵权的功能。

---

**Abstract:** Whether the infringement upon personality rights by generative artificial intelligence will lead to damages should be judged according to the specific types of personality rights involved. Downstream damages based from the generated infringing content should not be included in the scope of damages to be borne by the generative artificial intelligence service providers. Whether the potential damages for regenerating infringing content should be compensated is determined by the probability of recurrence. Because it is hard to calculate the loss of the victim and the profit of the tortfeasor accurately, discretionary compensation or statutory compensation should be the main way to determine the compensation. In the generation stage, given that the fault of generative artificial intelligence service providers is negligence, and the mode of human-computer interaction limits the scope of infringement, the property damages borne by service providers should not be high in general. Also the emotional damages should not be approved. In the generative artificial intelligence scenario, the liability to eliminating adverse effects and rehabilitating reputation can be fulfilled by issuing a clarification statement to users who have accessed the infringing content or creating a dedicated page for such clarifications by service providers. Making apologies are no longer applicable. Comparing with the traditional compensation theory, the special features of generative artificial intelligence infringement lies in changing the basic factual circumstance, impacting the value choice and affecting the governance logic.

**Key Words:** generative artificial intelligence, methods of assuming tort liability, damages compensation, property damages, emotional damages

---

(责任编辑：林涸民)