重庆大学学报(社会科学版) 2025年第31卷第4期
JOURNAL OF CHONGQING UNIVERSITY(Social Science Edition) Vol. 31 No. 4 2025

Doi:10. 11835/j. issn. 1008-5831. fx. 2023. 09. 001

欢迎按以下格式引用:朱荣荣. 生成式人工智能对个人信息保护的挑战及应对[J]. 重庆大学学报(社会科学版),2025(4):222-235. Doi:10. 11835/j. issn. 1008-5831. fx. 2023. 09. 001.

Citation Format:ZHU Rongrong. The challenge and response to personal information protection for generative artificial intelligence[J]. Journal of Chongqing University (Social Science Edition),2025(4):222-235. Doi:10. 11835/j. issn. 1008-5831. fx. 2023. 09. 001.

生成式人工智能对个人信息保护的挑战及应对

朱荣荣

(中国矿业大学 人文与艺术学院,江苏 徐州 221116)

摘要:以ChatGPT、DeepSeek为代表的生成式人工智能是指能够根据用户指令生成文字、图片、视频等相应内容的人工智能。个人信息是生成式人工智能的基础,生成式人工智能在模型训练、模型生成,以及模型优化等各个阶段均需要处理大量的个人信息,同时也对传统的个人信息保护规则带来了一定的冲击。在信息收集阶段,生成式人工智能可能虚化知情同意规则,侵犯信息主体的隐私权。在信息利用阶段,生成式人工智能可能冲击目的限制原则、公开透明原则等基本的个人信息处理规则,提高个人信息泄露的风险。在信息生成阶段,生成式人工智能可能产生虚假信息以及歧视性信息。在生成式人工智能变革式发展背景下,亟须审视个人信息保护的基本理念,寻求其在生成式人工智能领域的应然价值取向。通过考察比较法以及我国个人信息保护理念的发展脉络可知,个人信息保护或个人信息利用的单极性思维难以适应数字社会的现实需要,而个人信息保护与个人信息利用的动态平衡则是平衡各方主体利益的理想路径。生成式人工智能可以作为基础模型被广泛应用于教育、金融、科技等诸多领域,鉴此,应当协调推进个人信息保护与生成式人工智能发展的平衡兼顾。个人信息与信息主体密切相关,个人信息一旦被泄露或滥用可能使信息主体面临较高的风险,因此需要构建事前风险预防与事后损害赔偿的协同救济机制,在实现个人信息全生命周期保护的基础上,促进生成式人工智能的良性发展。就风险预防机制而言,需要在风险识别基础上完善去识别化措施,并赋予信息主体限制处理权、算法解释权,全方位遏制潜在的风险。责任主体的确定是损害赔偿的基础,应由生成式人工智能服务提供者与使用者证明其与个人信息侵权损害不存在因果关系,否则需要承担连带赔偿责任。在归责原则方面,可以根据被侵害的对象为个人一般信息或个人敏感信息,分别适用过错推定责任原则或无过错责任原则。

基金项目:2024年度江苏省社会科学基金青年项目"数字风险社会预防性侵权责任研究"(24FXC010);2024年度江苏省高校哲学社会科学研究一般项目"解释论视野下个人信息动态平衡保护的体系构造研究"(2024SJYB0779);中央高校基本科研业务费项目"健康数据商业化利用的民法规则优化研究"(2024SK16)

作者简介:朱荣荣,中国矿业大学人文与艺术学院,Email:zhurongrong1305@163.com。

为了更好地救济信息主体所受的损害，除了采取财产损害赔偿与精神损害赔偿等传统的补偿性赔偿外，还应当引入惩罚性赔偿，最大程度保障信息主体的受损权益。

关键词：生成式人工智能；ChatGPT；DeepSeek；个人信息；风险预防；侵权责任

中图分类号：D923　　**文献标志码**：A　　**文章编号**：1008-5831(2025)04-0222-14

一、问题的提出

2022年11月30日，美国人工智能实验室Open AI推出的聊天机器人ChatGPT引发社会广泛关注，其强大的内容生成能力标志着生成式人工智能技术的迭代式发展。ChatGPT可以根据用户输入的指令完成文本编辑、聊天对话、视频制作、代码编写等一系列任务，且完成质量较高，因而被评为2023年十大技术突破之一①。2024年9月，Open AI正式发布了推理模型o1，该模型可以通过一系列的推理步骤进行自我对话，并在此过程中不断纠正自己，能够解决复杂的科学、编码以及数学模型问题，更是在一项权威测试中击败了博士学者②。2025年初，杭州深度求索人工智能基础技术研究有限公司（DeepSeek）发布的DeepSeek-R1模型，以其高性能、低成本、完全开源的特性，掀起了新一轮智能应用大规模拓展的浪潮[1]。一般认为，人工智能可以进一步划分为分析式人工智能（Analytical Artificial Intelligence）与生成式人工智能（Generative Artificial Intelligence），分析式人工智能是指从大量数据中寻找隐藏模式并形成一定预测的人工智能，而生成式人工智能则指通过学习海量的数据进而生成新内容的人工智能[2]。数据是生成式人工智能的基础，生成式人工智能在模型训练、内容生成、模型优化等阶段均离不开数据的支撑。然而，生成式人工智能大规模的信息收集以及不透明的处理规则给个人信息保护带来了挑战。2025年1月，《纽约时报》将Open AI告上联邦法院，指控ChatGPT在未经授权或支付费用的情况下擅自使用其大量的文章进行模型训练，侵犯了其版权，要求Open AI赔偿数十亿美元，同时销毁ChatGPT的数据集③。在生成式人工智能变革式发展的背景下，国家互联网信息办公室联合教育部、科学技术部等其他部委于2023年7月13日公布了《生成式人工智能服务管理暂行办法》（以下简称《暂行办法》），然而，《暂行办法》关于个人信息保护的条文较少，且缺乏针对生成式人工智能技术特性的专门性规定。职之是故，如何在保护个人信息的基础上统筹推进生成式人工智能的创新发展，是当下必须回答的时代命题。

二、生成式人工智能的技术逻辑及其对个人信息保护的挑战

（一）生成式人工智能的技术逻辑

生成式人工智能虽然属于新兴事物，但其技术理论却具有较为悠久的历史。从技术发展史角度看，生成式人工智能起源于20世纪40年代的“控制论”[3]，20世纪60年代问世的聊天机器人Eliza则是生成式人工智能最早的原型，标志着人与计算机之间的自然语言对话成为可能[4]。立法层面，我国《暂行办法》首次对生成式人工智能的概念予以明确，《暂行办法》第22条规定，生成式人工智能

① See 10 Breakthrough Technologies.2023,https://www.technologyreview.com/2023/01/09/1066394/10-breakthrough-technologies-2023/?truid=&utm_source=the_algorithm&utm_medium=email&utm_campaign=the_algorithm.unpaid.engagement&utm_conten.

② See Nicola Jones,‘In awe’: scientists impressed by latest ChatGPT model o1, https://www.nature.com/articles/d41586-024-03169-9.

③ See Bobby Allyn,′The New York Times′ takes Open AI to court.ChatGPT′s future could be on the line,https://www.npr.org/2025/01/14/nx-s1-5258952/new-york-times-openai-microsoft.

技术是指具有文本、图片、音频、视频等内容生成能力的技术。据此可知,生成式人工智能具有多模态性,不仅可以处理文本任务,还能够生成图片、音视频等其他形式的内容。

现阶段,Dall-E2、Stable Diffusion、ERNIE Bot、ChatGPT、DeepSeek等生成式人工智能不断涌现,本文以用户基数较大的ChatGPT、DeepSeek为分析模型,旨在明晰生成式人工智能的技术逻辑。相较于此前Siri、小冰等功能单一且对话生硬的聊天机器人,以ChatGPT、DeepSeek为代表的生成式人工智能可以"记住"与用户之前的聊天内容,从而实现连续多轮对话。通常来说,生成式人工智能的运行模式可以概述为以下三个阶段,即用户输入指令→算法整合分析→生成相应内容。具体来说,用户在指定的对话框内输入指令,然后系统快速对该指令进行分析,如果经分析认为用户的指令违反其内置的伦理审查机制,则会拒绝回答用户的请求。相反,如果经审查用户的指令并不违反其审查机制,则会在大规模的语料库中搜索与之相关的语词,并预测下一个单词出现的频率,然后将搜集到的语词整合优化,最后输出符合人类自然语言表达习惯的文本、图像、视频等内容。

以ChatGPT的技术架构为例,ChatGPT是Chat Generative Pre-trained Transformer的简化版称谓,其中,Transformer架构是ChatGPT的关键性技术。Transformer模型诞生于2017年谷歌研究团队发表的《Attention is all you need》一文,该文指出Transformer是一种基于自注意力机制的序列转换模型,其采用多头自注意力取代了编码器—解码器架构中最常用的循环层,因而Transformer的训练速度明显快于基于循环层或卷积层的体系架构④。Transformer模型的采用使得ChatGPT能够快速学习数据之间的关系,借由人工智能训练师对模型输入与输出的大量调试,使得模型初步掌握自然语言的语法规则。为了优化模型表达,使其生成内容更加符合人类社会的认知观念,ChatGPT在GPT-3.5模型基础上采用了基于人类反馈的强化学习(Reinforcement Learning from Human Feedback),在语言模型的训练、打分模型的训练以及语言模型的优化阶段,采用人类提问机器回答与机器提问人类回答的方式,不断提高模型的生成质量[5]。

(二)生成式人工智能对个人信息保护的挑战

数据是生成式人工智能的基石,生成式人工智能在模型训练、数据处理以及内容输出等方面均离不开大数据的支持。虽然DeepSeek通过优化算法一定程度上节省了算力需求,但数据仍是生成式人工智能技术创新发展的基础性要素,通常来说,数据的"量"与生成式人工智能的"质"之间呈现正相关关系,训练数据规模越大,则生成式人工智能的生成内容质量就越高[6]。然而,由于技术的不成熟性与立法的相对滞后性,生成式人工智能在收集、利用、生成数据的过程中,给传统的个人信息保护规则带来了挑战。

1. 生成式人工智能收集信息可能带来的挑战

知情同意原则作为个人信息收集的基本原则,为各国(地区)立法所承认。然而,根据Open AI官方网站公布的隐私政策可知,其在收集个人信息过程中并没有严格贯彻知情同意规则。在告知阶段,信息处理者应当明确告知信息处理的目的、方式等具体事项,确保信息主体充分知情。然而,Open AI声称其可能将"收集的个人信息用于改进服务、开展研究、防止犯罪活动等",这些表述具有较大的模糊性与不确定性,无法给信息主体提供明确的行为预期。在同意阶段,信息处理者只有取得信息主体自主的、真实的同意才能收集个人信息,如果其在个人信息处理过程中更改信息处理方

④ See Ashish Vaswani,Ashish Vaswani,Ashish Vaswani,et al.Attention is all you need,https://arxiv.org/pdf/1706.03762v5.pdf.

式的，应当重新取得信息主体的同意。虽然Open AI宣称其在训练基础模型时所使用的数据来源于已公开的数据与获得授权许可的数据，但其至今没有对外公示数据的具体来源，数据来源的正当性存在疑问⑤。此外，Open AI在隐私政策中声明“当用户创建账户使用ChatGPT时，Open AI有权收集用户的姓名、联系方式、账户凭据等个人信息”，DeepSeek隐私政策也表明，其将收集用户与DeepSeek进行交互过程中所输入的文本、图片、文件等内容，严重违背了知情同意原则的基本内涵。

除了虚化知情同意规则，生成式人工智能在收集个人信息过程中还可能侵犯隐私权。《中华人民共和国民法典》（以下简称《民法典》）第1032条第2款规定，隐私的判断不再依据公开与否，而更强调权利主体的主观意愿，如果权利主体不愿其私密信息被他人知晓，即使是已经公开的私密信息仍属于隐私范畴。生成式人工智能采用生成式预训练转换模型，可以在没有人工监督的情形下主动抓取互联网数据，加之缺乏必要的筛选与过滤机制，导致收集的数据中可能包含私密信息，再者，由于个人信息内在的勾连性，即使生成式人工智能收集的都是非私密信息，但这些海量数据聚合重组之后也可能成为私密信息，从而威胁公众的隐私安全[7]。

2. 生成式人工智能利用信息可能带来的挑战

在生成式人工智能利用个人信息过程中，由于“算法黑箱”以及其他不确定性因素的存在，使目的限制原则以及公开透明原则难以得到贯彻，海量的数据处理提高了个人信息泄露的风险。

目的限制原则是个人信息利用应当遵循的基本原则，旨在防止信息处理者恣意处理个人信息[8]。根据Open AI的隐私政策可知，其可能“将收集到的信息用来分析用户的行为和特征”“研发新的服务”等，超越了信息收集时的处理目的，可能使信息主体的合理预期落空。公开透明原则要求信息处理者公开披露个人信息处理的具体细节，然而生成式人工智能所依赖的深度学习技术是一个“黑箱”，在生成式人工智能系统输入的数据和输出的结果之间，存在着人们无法洞悉的“隐层”[9]。生成式人工智能具有涌现性，随着算法复杂性以及数据规模的提升，生成式人工智能可能会输出意料之外的内容，即使对于生成式人工智能设计师来说，也可能难以回答为什么会输出这样的内容，这违背了《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7条以及第24条所要求的公开透明原则。

生成式人工智能储存大量的个人信息，一旦这些信息被泄露或者被不法使用，可能威胁信息主体的人身权、财产权。2023年3月，Open AI公开承认因开源数据库的错误，导致部分用户可能看到其他用户的聊天内容以及信用卡相关信息。DeepSeek同样存在数据泄露问题，有研究发现，DeepSeek可公开访问的ClickHouse数据库允许用户完全控制数据库操作，而这些数据库中包含用户的聊天记录、密钥以及其他高度敏感的个人信息⑥。个人信息不依附于信息主体而存在，因而其不法泄露一般不会立刻给信息主体造成财产损害或精神损害，当信息主体事后请求损害赔偿时可能因时间久远而无法提供有力的证据，不利于保护信息主体所受之损害。

3. 生成式人工智能生成信息可能带来的挑战

在生成内容方面，由于生成式人工智能无法实现真正意义上的推理，其可能生成看似合理但虚

⑤ 参见：李昀锴．ChatGPT内容商业使用的法律风险及应对[Z/OL].(2023-02-13)[2025-06-30]. https://mp. weixin. qq. com/s/8fzvmnyhEbIwWVTAVm8WWA.

⑥ See Gal Nagli.Wiz Research Uncovers Exposed DeepSeek Database Leaking Sensitive Information,Including Chat History,https://www.wiz.io/blog/wiz-research-uncovers-exposed-deepseek-database-leak.

假的内容,而在人工智能逻辑包装下的信息难以被人类辨别真伪[10]。不同于传统的搜索引擎,生成式人工智能输出的内容具有唯一性,并不提供多来源信息参考,使得虚假或误导性信息难以被用户发现。例如,曾有用户使用ChatGPT编辑了一则“杭州将在3月1日取消机动车尾号限行”的新闻,大量公众没有识别出其为虚假消息,造成了不良的社会影响。

除了虚假信息,生成式人工智能还可能输出歧视性信息,挑战法律上的平等原则。生成式人工智能的初始模型需要海量的数据支持,这些数据不可避免地存在人类的认知偏见,由于生成式人工智能并不具有学习能力,其通过从大型数据库中提取相关信息,并运用一定的排列组合形成特定的文本[11]。因此,如果数据库中的数据存在偏差,那么输出的内容必然存在偏差。此外,生成式人工智能系统可能嵌入设计者的显性或隐性偏见,而这种技术盲点难以从算法中剔除[12],致使生成信息存在价值偏见。

三、生成式人工智能背景下个人信息保护理念的确定

现代社会,迭代更新的高科技技术不断挑战传统的个人信息保护模式,究其原因,在于个体主义与静态思维的个人信息保护方式难以适应科技的迅猛发展[13]。在ChatGPT、DeepSeek等国内外生成式人工智能技术迅猛发展的背景下,亟须确立生成式人工智能中个人信息保护的基本理念,缓和生成式人工智能创新发展与个人信息法律保护之间的张力。

(一)个人信息保护理念的考查

1. 比较法上个人信息保护理念的历史嬗变

比较法上,不同国家(地区)对于个人信息保护的立场存在较大的差异,大致可以分为美国模式与欧盟模式。就美国而言,早期学者提倡个人信息的自由流动,波斯纳通过综合分析成本与效益,认为相较于个人信息保护,个人信息的利用更符合经济效率原则[14]。还有学者认为,个人信息只有聚合在一起形成庞大的数据池才有价值,普通人可能永远无法真正地从其单一的个人信息中获得经济价值[15]。然而,个人信息天然地蕴含人格利益,片面强调个人信息的利用可能阻碍人格的自由发展。对此,有学者认为,个人信息保护应在流动的而非静态的情况下予以确定,并且需要综合考虑社会环境、个人信息的经济价值等因素[16],从而促进个人信息保护与个人信息利用的协调发展[17]。

就欧盟而言,1950年通过的《保护人权与基本自由公约》第8条明确规定,人人有权享有使自己的私人和家庭生活以及通信自由得到尊重的权利,这种人格尊严与人格自由至上的理念深刻影响了个人信息保护立法。1995年,欧洲议会和欧盟理事会通过的《关于个人信息处理和流通过程中对个人信息保护的指令》构建了以“知情—同意”为核心的权利体系来保障信息主体对个人信息的全面控制与支配,诸如知情权、访问权、拒绝权等。信息技术的快速发展使社会对于个人信息的利用需求日益增高,2018年生效的《一般数据保护条例》顺应社会发展现状,明确规定条例之宗旨不仅在于保护自然人的合法权益,也在于促进个人数据自由流动,改变了传统只注重信息主体利益维护的单向性的立法思维,是个人信息保护理念的突破式发展。

2. 我国个人信息保护理念的发展脉络

《民法典》通过立法的形式明确了个人信息具有独立的法律地位,这一时期,社会的整体论调是强化个人信息保护,对此又存在“权利说”与“利益说”两种不同的观点。“权利说”认为,为了全面保障个人信息不受非法侵犯,应将《民法典》第111条解释为我国立法确立了“个人信息权”[18]。“利益

说”则认为个人信息的内涵及外延均存在极大的不确定性，难以达到权利客体所要求的具体特定且界限分明的品质，应将个人信息定位于受法律保护的利益[19]。可见，无论是“权利说”还是“利益说”，均强调保障信息主体对于个人信息的支配与控制，忽视了个人信息内含的经济价值对整个社会的重要性。

随着数字社会的到来，过于强调个体权益的保护可能阻碍社会的整体发展，人们开始重新审视“个人控制论”的合理性。有学者认为，“个人控制论”将信息主体预设为具有完整理性、完整意志力的主体，然而在个人信息处理过程中，信息的不对称性、风险的不确定性以及技术的复杂性等诸多因素的存在，导致信息主体难以做出理性的选择[20]。从社会发展的角度看，个人信息是推进我国数字经济发展与数字社会构建的重要驱动力，在这一特殊的时代背景下，应当更强调个人信息的社会化利用[21]。

面对“信息保护论”与“信息利用论”的争执不休，有学者提出信息保护与信息利用并不存在非此即彼的竞争关系，而是相辅相成，应当促进两者的动态平衡[22]。然而，“动态平衡论”也遭到了部分学者的质疑。有学者认为，信息保护与信息利用是一对矛盾体，个人信息权益保护必然影响个人信息合理利用，反之亦然[23]。还有学者直言，所谓信息保护与信息利用的“动态平衡”只能是理想状态，无法在实践中具体落实[24]。

（二）生成式人工智能中个人信息保护的基本理念：兼顾科技发展与权益保护

1. 规范基础

《个人信息保护法》第1条明确规定，其立法宗旨是“保护个人信息权益，规范个人信息处理活动，促进个人信息合理利用”。然而，该条并未言明信息保护与信息利用究竟具有同等的地位还是存在主次顺序，引发了学界激烈的探讨。龙卫球认为，个人信息权益保护与个人信息合理利用存在主次关系，前者是首要立法目的，而后者是次要立法目的[25]。高富平也认为，个人信息权益保护是个人信息合理利用的前提，只有个人信息权益得到了有效的保护，才有个人信息合理利用的空间[26]。许可则认为，《个人信息保护法》确立了个人信息保护与个人信息利用的平衡机制，其不仅要实现信息主体与私人处理者的利益平衡，还要促进信息主体与国家机关处理者的利益平衡[27]。

笔者认为，虽然从条文表述上看，《个人信息保护法》将“保护个人信息权益”置于“个人信息合理利用”之前，但不能就此推断出个人信息保护优先于个人信息合理利用。实践中，信息保护与信息利用不存在预先、固定不变的序位关系，应当优先保护何种价值需要根据个案具体情境加以确定。譬如，《个人信息保护法》第13条规定，如果为了履行法定职责或者保护公共利益，则可以不经信息主体同意而利用个人信息。可见，在此情形下，个人信息利用优先于个人信息保护。就此而言，《个人信息保护法》为统筹兼顾个人信息保护与生成式人工智能发展提供了坚实的规范基础。

2. 实践需求

生成式人工智能具有多模态性、生成性等特征，能够生成风格各异、内容丰富的图片、诗歌、视频等，可以被广泛应用于教育、政务服务、科技等领域。自DeepSeek-R1发布以来，仅仅7天就实现了用户量过亿，成为现象级的应用产品，华为、阿里、百度、腾讯、京东等多家大型平台企业相继接入DeepSeek大模型⑦。在ChatGPT、DeepSeek等生成式人工智能全面赋能社会生产发展的现实背景

⑦ 参见：郭倩.微信测试接入DeepSeek生态圈持续扩大[N].经济参考报,2025-02-17(02).

下,亟须在保护信息主体权益的基础上鼓励数据要素价值的释放,不能因此而阻碍人工智能大模型产业的创新发展。

2023年3月13日,第十四届全国人民代表大会第一次会议通过的《关于2022年国民经济和社会发展计划执行情况与2023年国民经济和社会发展计划的决议》明确提出,要加快人工智能、大数据等新型基础设施建设,有序推进基础设施智能升级。可以预见,未来生成式人工智能将赋能各行各业,促进我国经济社会的全面发展。在生成式人工智能场景下,个人信息的真正价值在于流动性,如果过度保护信息主体利益,将会阻碍生成式人工智能的发展[28]。因此,应重点规制生成式人工智能在应用场景下可能给个人带来的风险,而非阻止生成式人工智能进入市场⑧。

四、生成式人工智能中个人信息风险预防机制的构建

(一)风险预防的基本内涵

自1986年贝克提出"风险社会"一词以来,"风险"成为诸多领域的研究热点。从某种意义上说,风险具有普世重要性,"风险评估""风险效益分析""风险管理"已经成为政府战略或者金融战略的焦点,而理论层面关于"风险"的论断更是宏大[29]。关于"风险"的内涵,贝克认为风险是指某种危险和不安全感的方式,与财富的具体可感相比,风险具有某种非现实性[30]。吉登斯则认为,风险是指在与将来可能性关系中被评价的危险程度[31]。基于此,笔者认为可从主观与客观两个面向来认识"风险",就客观方面来说,"风险"指损害可能性的提高;就主观方面来说,"风险"指权利主体合理期待的落空。

在生成式人工智能应用过程中,个人信息一旦被泄露或滥用可能使信息主体面临较高的风险,为此,需要采取一定的措施预防风险的现实发生。罗伯特·阿多诺认为,风险预防原则是科技时代法律的新标准,其构成要件包括风险的不确定性、风险评估、严重或不可挽回的损害、措施的相称性以及举证责任倒置[32]。具体来说,风险的不确定性,指风险是否发生不是完全地确定,其发生的概率介于0到100之间,如果风险确定不会发生,则没有风险预防的必要性。相反,如果风险百分之百发生,则采取任何的预防措施都是于事无补,此时可通过损害救济保护受害人的合法权益。在风险识别基础上,需要准确评估风险的大小,并按照一定的标准将风险划分为不同的等级,据此提供与之相适应的规制措施。严重或不可挽回的损害是风险预防的立足点,如果生成式人工智能可能造成的损害微乎其微,或者损害具有可补救性,则会在一定程度上弱化采取风险预防措施的正当性。措施的相称性要求拟采取的预防措施必须与生成式人工智能可能造成的风险相匹配,不得以风险预防之名阻碍生成式人工智能的良性发展。举证责任倒置将风险的证明责任配置给生成式人工智能服务提供者,由其证明个人信息处理行为不会给信息主体造成超出社会一般人可以承受的风险,防止信息主体因信息不对称或相关知识的匮乏陷于举证不能的窘迫境地。

(二)引入风险预防的可行性

面对社会发展衍生的技术风险,法律通常从两个面向着手,一种是压制性的(repressive),即在事故发生后启动,主要关注责任归属以及损失赔偿数额;另一种则是预防性的(preventive),即事先采取措施预防特定技术活动可能产生的风险,从而避免事故的发生[33]。一直以来,损害赔偿所具有

⑧ See Lilian Edwards.Regulating AI in Europe:Four problems and four solutions,https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf.

的填补受害人损害的积极作用都不容置疑，但传统侵权法只保护客观发生的现实损害，而不保护尚未现实发生的损害，即使其发生的可能性极大也不例外。在损害赔偿的落实过程中，除了具体的损害赔偿数额外，还存在诉讼成本、证明成本等其他的额外支出，极大地消耗了社会的总体利益。对此，有学者指出，相较于事前的风险预防，事后的损害救济更多的是一种无可奈何的措施，如果能够在风险预防与损害救济之间选择的话，受害人与立法者可能都会毫不犹豫地选择后者[34]。

在生成式人工智能背景下，个人信息权益侵害呈现新的特征，使传统的损害赔偿难以应对复杂的现实实践。生成式人工智能可以在没有人工监督的情形下主动收集互联网数据，可能导致收集的数据中包含大量的敏感信息或私密信息，给信息主体权益造成了极大的风险。根据我国现行的证明标准，信息主体在诉请损害赔偿时必须证明其遭受了现实的损害以及具体的损害数额，否则需要承受举证不能的不利后果。ChatGPT作为高科技的产物，其内在的技术规则难以为一般人所知晓，在此背景下，信息主体很难证明生成式人工智能给其造成了具体的损害。风险预防是前瞻性的解决方案，其关注于未来损害是否发生[35]，并要求损害的实际发生，权利人只需要证明损害的发生具有高度的可能性时，即有权请求侵害人采取一定的措施，从而合理规避生成式人工智能引发的风险。此外，风险预防在我国具有一定的规范基础，例如，《个人信息保护法》第11条明确规定“国家建立健全个人信息保护制度，预防和惩治侵害个人信息权益的行为”。《中华人民共和国数据安全法》（以下简称《数据安全法》）第22条也要求“建立集中统一、高效权威的数据安全风险评估、报告、信息共享、监测预警机制”。因此，将风险预防引入生成式人工智能领域并不会突破法律体系的稳定性。

（三）风险预防的规范构造

现阶段，《个人信息保护法》《数据安全法》等诸多规范都体现了风险预防的基本理念，遗憾的是，尚未制定风险预防的具体规则，不利于对司法实践发挥统合作用。对此，可以从信息主体与信息处理者两个层面分别构建风险预防规则，促进风险预防的贯彻落实。

1. 信息主体层面：丰富个人信息权内容

目前，《民法典》《个人信息保护法》等相关法律赋予了信息主体较为丰富的权利，诸如知情决定权、查阅复制权、更正补充权、删除权等。须明确的是，这些“权利”并非传统意义上的民事权利，其更多地体现为工具性价值，并不具有独立的意义。为了防止生成式人工智能恣意侵犯信息主体的合法权益，除上述权利外，还应当赋予信息主体限制处理权、算法解释权等新型权利。

就限制处理权来说，虽然《个人信息保护法》第44条规定“个人有权限制或者拒绝他人对其个人信息进行处理”，但这只是“知情决定权”的具体权能或者权利内容，无法从中得出《个人信息保护法》将限制处理权作为独立的权利类型[36]。限制处理权是指在特定情形下，信息主体得以请求信息处理者以一定方式暂时停止或永久停止信息处理行为的权利[37]。实际生活中，当个人信息的正确性或完整性处于不确定状态且双方均难以证明时，限制处理权能够在保护信息主体免受不正确信息侵害的同时兼顾信息处理者的利益，避免系争事实不清状态下损及双方的权益，弥补更正权与删除权的不周延性，符合权利精细化发展之趋势。

算法解释权是指权利人认为算法自动化决策影响其合法权益时，可以要求算法设计者解释算法结果的决策过程、运行原理等事项[38]。由于生成式人工智能技术发展的不完善性，算法黑箱难以得到消解，因而生成内容的可解释性往往较低。鉴此，应当赋予信息主体算法解释权，从而要求生成式人工智能服务提供者披露算法的运行机制，实现算法的透明化。需注意的是，生成式人工智能

的输出内容由算法规则、训练数据等多方面因素决定,因此,生成式人工智能服务者除了披露算法运行的源代码和算法运行的机制外,还需要披露算法运行过程的控制情况[39]。

2. 生成式人工智能服务提供者层面:完善去识别化措施

比较法上,大多数国家(地区)的立法一致认为个人信息的核心要素在于可识别性,但可识别性不是固定不变的,其可能被消除或减弱而成为"匿名化信息"(Anonymization)、"去标识化信息"(De-Identification)或"假名化信息"(Pseudonymization)。目前,我国立法尚未规定"假名化信息",因而去识别化措施主要包括去标识化与匿名化。

《个人信息保护法》第73条规定,匿名化信息与去标识化信息的规范内涵存在差异。具体来说,在识别能力方面,匿名化信息确定地、终局地不具有识别个人的可能性。不同于此,去标识化信息只是降低而非彻底消除可识别性,因此在一定的技术条件下,去标识化信息与其他信息结合仍有可能将特定主体识别出来。在能否复原方面,匿名化将个人信息的识别元素永久地、彻底地删除,这意味着匿名化信息不具有可逆性,其永远都不可能恢复识别性。相比于匿名化信息,去标识化信息的匿名程度不够彻底,其仅仅使他人无法凭借该信息直接关联到特定个人,但若结合其他额外信息或采取一定的技术措施,去标识化信息仍然可以被复原。

在具体措施上,生成式人工智能服务提供者在处理个人信息时,可以采用假名、加密、哈希函数等技术手段将个人信息中的直接标识符或准标识符删除或变换,去除标识符与信息主体之间的关联性。此外,生成式人工智能服务提供者应当将去标识化后的信息与可用于识别个人信息的标识符分别存储,从而降低信息与特定自然人的关联程度,避免其他主体根据相关属性识别出信息主体。就匿名化措施而言,有学者认为匿名化是一个相对的概念,其范围可能会根据具体的应用情况而有所不同,完美的匿名化概念只是理论上的想象[40]。不可否认,个人信息天然地具有勾连性,大数据分析技术的进步提高了信息聚合的速度,使匿名化信息存在被识别的风险。理论上说,只要投入足够的人力、物力与时间,凭借匿名化信息也能重新识别出个人信息甚至个人敏感信息[41]。然而,若复原个人信息所花费的成本与时间过于高昂,以至于与所欲达到的目的明显不成比例时,可能在一定程度上会妨碍人们对于匿名化信息的再分析。

五、生成式人工智能中个人信息侵权责任的规则展开

(一)责任主体的确定

根据《暂行办法》第9条、第22条等相关条文可知,《暂行办法》构建了以"服务提供者"为主,"服务使用者"为辅的责任体系,且生成式人工智能服务提供者的信息责任包括网络信息内容生产者责任与个人信息处理者责任。须明确的是,生成式人工智能不应当成为个人信息侵权责任的主体,从技术角度来说,类ChatGPT生成式人工智能的生成是一种"从有到有"的生成,只是信息的整合优化而非知识的生成,其并未从根本上动摇人机之间的主客关系[42]。从实践角度来说,责任主体需要对可归责于自己的损害承担法律责任,而生成式人工智能并没有独立的财产,无法成为追责的对象。

生成式人工智能作为人工智能时代新型的数字基础设施,不仅可以为使用者提供智能问答、代码生成等直接性服务,还可以作为基础模型赋能于金融、医疗、自动驾驶等一系列下游产业中[43]。生成式人工智能的通用性使得个人信息的责任主体呈现垂直式分布,申言之,生成式人工智能的基础模型既可以作为信息内容生产者为用户提供直接性的信息服务,也可以作为技术支持者将基础

模型提供给其他平台进行更专业化的研发利用。目前，社会公众对于生成式人工智能的认知较为片面，很难提供证据证明具体的侵权行为人，一定程度上阻塞了信息主体的救济渠道。为此，应当实行举证责任倒置，由生成式人工智能服务提供者与使用者举证证明其不存在侵害行为，对于无法提供确切证据证明的，可以借鉴共同危险理论，将生成式人工智能服务提供者与使用者视为一个整体，如果受害人有证据证明侵害人利用生成式人工智能侵害其合法权益，即可推定生成式人工智能服务提供者及使用者与受害人的损害存在因果关系[44]，从而要求生成式人工智能服务提供者与使用者承担连带责任。

（二）二元归责原则的确立

《个人信息保护法》第69条规定，我国个人信息侵权归责原则采取的是一元的过错推定责任原则。然而，个人信息不仅蕴含人格利益，还具有经济性利益，一元的过错推定责任原则过于强调信息主体利益的保护，忽视了个人信息的有效利用对于社会发展的重要性，不符合个人信息利益多元化的客观现实，也违背了《个人信息保护法》的立法宗旨。

法律意义上的归责指，从法律价值层面来判断某人对于损害的发生是否应当承担法律责任[45]。可见，归责不是单纯的事实判断，而是多元价值评价的结果，其本质在于合理平衡侵害人与受害人的利益冲突。生成式人工智能中个人信息侵权归责原则的确定需要合理协调信息主体利益与生成式人工智能服务提供者的利益。目前，我国规范层面对于个人信息采取了类型化的规制思路，存在公开信息与非公开信息，私密信息与非私密信息、个人敏感信息与个人一般信息等不同的分类模式。可见，现行规范关于个人信息的类型化标准较为繁杂，缺乏体系之间的协调与融贯，导致同一信息可能归属于不同范畴，影响了受损权益的有效救济。

生成式人工智能可以主动抓取网络公开信息，并根据收集的公开信息形成用户的数据画像，因此，如果根据收集的信息是否具有公开性来设计差异化的保护制度，其合理性有待商榷。此外，对于个人而言，私密信息与非私密信息的分类标准较为模糊，无法为司法裁判提供明确的指引。在现行规范体系下，以个人敏感信息与个人一般信息为基点来构建差异化的归责原则不仅契合个人信息保护的本质属性，也能够实现弹性化的价值评价。由于个人敏感信息关系信息主体的核心利益，不当处理个人敏感信息可能侵害信息主体的人格尊严、人身自由等重要权益。与之不同，个人一般信息则更多地承载社会交往功能，是信息主体参与社会生活的重要媒介。从价值衡量的角度来说，相较于个人一般信息，法律应当对于侵害个人敏感信息的不法行为施加更严重的侵权责任。鉴此，对于侵害个人一般信息的应当适用过错推定责任原则，而对于侵害个人敏感信息的则适用无过错责任原则。

（三）补偿性侵权责任与惩罚性侵权责任的联动

1. 个人信息是人格利益与财产利益的综合载体⑨，生成式人工智能侵害个人信息可能同时造成财产损害与非财产损害。就财产损害赔偿来说，《民法典》第1182条、《个人信息保护法》第69条规定，应当根据受害人损失、侵害人获益、协议赔偿以及酌定赔偿予以确定。关于损失赔偿的范围，侵权法以填补受害人因侵害行为所遭受的损失为指导思想，因此损失赔偿范围应当奉行完全赔偿原

⑨ 个人信息蕴含人格利益与财产利益得到了我国诸多学者的认可。参见：程啸《论我国民法典中个人信息权益的性质》（《政治与法律》，2020年第8期2-14页）；吕炳斌《个人信息权作为民事权利之证成：以知识产权为参照》（《中国法学》，2019年第4期44-65页）；张素华《个人信息商业运用的法律保护》（《苏州大学学报》，2005年第2期36-39页）。

则,这也是公平正义的内在要求。关于获益返还的范围,如果要求侵害人将合法经营利益一并返还给受害人,可能导致受害人不当获利,亦使获益返还滑入惩罚性赔偿的尴尬境地。就此而言,在确定获益返还的范围时,应当将受害人的正当性获益排除出去。关于酌定赔偿的范围,需要根据个案情况综合考察侵害手段、侵害情节、侵害范围、侵权人的获益、侵权人的主观过错程度、当地经济发展水平以及侵权人的责任承担能力,同时辅之以公平原则及诚实信用原则确定赔偿数额。

关于精神损害赔偿,如果生成式人工智能服务提供者违背信息主体的意愿而不法处理个人信息,并给信息主体造成恐惧、不安、痛苦、绝望等不良情绪时,即可认为信息主体遭受了事实上的精神损害。然而,信息主体遭受的过于遥远的、轻微的精神损害一般不予赔偿,但若侵害人主观上存在故意或重大过失的,则不应再强求损害后果的严重性。在确定可予赔偿的精神损害之后,尚需对精神损害进行金钱评价,金钱评价是精神损害与损害救济之间的桥梁与纽带,但金钱评价不是简单的算术公式的运用,需要在具体个案中充分考量侵权人的过错程度、侵权行为的具体情节、侵权行为造成的后果、侵权人的获利情况等诸种因素予以确定。

2. 惩罚性侵权责任

惩罚性赔偿起源于英美国家,与补偿性赔偿旨在填补受害人损害不同,惩罚性赔偿要求加害人赔偿的数额远远高于受害人的实际损害,从而惩罚加害人的侵害行为[46]。对于ChatGPT等生成式人工智能来说,其对于个人信息的不法处理可能严重扰乱人们的正常生活以及社会秩序的稳健运行,传统的补偿性赔偿难以阻遏侵害行为的发生。惩罚性赔偿能够消除行为人继续实施侵权行为的动力,并对潜在的加害人形成一定的警示作用,进而减少生成式人工智能中个人信息侵权发生的概率。

根据《中华人民共和国消费者权益保护法》第55条、《中华人民共和国食品安全法》第148条、《民法典》第1185、1207、1232条等相关条款可知,惩罚性赔偿的适用需要同时满足主观要件与客观要件。就主观要件来说,惩罚性赔偿要求侵权人主观上具有较强的可非难性,超过了一般社会公众的包容度。就客观要件来说,惩罚性赔偿要求侵害行为情节严重,造成受害人死亡或健康受损等严重损害后果,轻微的或一般的损害无法请求惩罚性赔偿。笔者认为,生成式人工智能具有较强的技术性,且其运行机理尚未对外公布,一般公众难以举证证明侵害人主观上“明知”或“故意”。对此,应当适度降低生成式人工智能中惩罚性赔偿的适用门槛,只要受害人有证据证明侵权人主观上具有轻过失即可,无须存在严重的、实际的损害后果。

惩罚性赔偿的落脚点在于赔偿数额的确定,关于生成式人工智能中个人信息侵权惩罚性赔偿数额的确定,笔者认为应当设置一个最高赔偿限额,由法官在最高赔偿限额内依据个案具体情境综合各项因素得出妥当的赔偿数额,防止法官自由裁量权的滥用,也能一定程度保障法的安定性与可预期性。在具体计算惩罚性赔偿数额时,可以受害人实际损失、侵害人不法获益或者个人信息授权许可费为计算基数,惩罚性赔偿的数额为计算基数的三倍。如果经由上述计算方式得出的数额不足500元的,则按500元计,防止信息主体因赔偿数额与诉讼成本明显不成比例而不去积极诉请惩罚性赔偿。

结语

如同历史上任何一次的技术进步,以ChatGPT、DeepSeek为代表的生成式人工智能无疑将会引

发新一轮的产业革命。比尔·盖茨声称，生成式人工智能是继互联网、智能手机之后最具革命性的技术⑩。生成式人工智能的运用可以提高生产效率，促进社会生产力的发展，但也可能引发算法歧视、数据安全、不公平竞争等一系列法律风险，冲击既有的法律秩序。数据是生成式人工智能的基础性要素，然而，生成式人工智能在收集、利用、生成信息过程中对于传统的个人信息保护规则带来了较大的挑战。

进入21世纪以来，层出不穷的科技产品渗透社会生活的方方面面，在此背景下，以个体权益保护为由而禁锢科技的发展似乎不是明智的选择。为了有效应对生成式人工智能等新兴科技可能带来的法律风险，需要适时转变传统的法律规制理念，立足于社会发展大局，统筹兼顾科技发展与个体权益的保护，构建合理的法律因应制度，促进生成式人工智能健康有序的发展，从而造福整个社会。

参考文献：

[1] 魏钰明. 大模型应用拓展下的智能社会复杂性及其治理[J]. 电子政务，2025(3)：2-6.

[2] 於兴中，郑戈，丁晓东. 生成式人工智能与法律的六大议题：以ChatGPT为例[J]. 中国法律评论，2023(2)：1-20.

[3] 蒲清平，向往. 生成式人工智能：ChatGPT的变革影响、风险挑战及应对策略[J]. 重庆大学学报(社会科学版)，2023(3)：102-114.

[4] WEIZENBAUM J. ELIZA：A computer program for the study of natural language communication between man and machine[J]. Communications of the ACM，1966，9(1)：36-45.

[5] 朱光辉，王喜文. ChatGPT的运行模式、关键技术及未来图景[J]. 新疆师范大学学报(哲学社会科学版)，2023(4)：113-122.

[6] 张亮，陈希聪. 生成式人工智能背景下的跨境数据安全规制：基于DeepSeek、ChatGPT等主流AI的思考[J]. 湖北大学学报(哲学社会科学版)，2025(2)：120-128.

[7] 郭春镇. 生成式AI的融贯性法律治理：以生成式预训练模型(GPT)为例[J]. 现代法学，2023(3)：88-107.

[8] 朱荣荣. 个人信息保护“目的限制原则”的反思与重构：以《个人信息保护法》第6条为中心[J]. 财经法学，2022(1)：18-31.

[9] 徐凤. 人工智能算法黑箱的法律规制：以智能投顾为例展开[J]. 东方法学，2019(6)：78-86.

[10] 郑永和，丁雨楠，郑一，等. ChatGPT类人工智能催生的多领域变革与挑战(笔谈)[J]. 天津师范大学学报(社会科学版)，2023(3)：49-63.

[11] BOZKURT A. Generative artificial intelligence (AI) powered conversational educational agents：The inevitable paradigm shift[J]. Asian Journal of Distance Education，2023，18(1)：198-204.

[12] BYRNE M D. Generative Artificial Intelligence and ChatGPT[J]. Journal of Perianesthesia Nursing，2023，38(3)：519-522.

[13] 丁晓东. 大数据与人工智能时代的个人信息立法：论新科技对信息隐私的挑战[J]. 北京航空航天大学学报(社会科学版)，2020(3)：8-16，71.

[14] POSNER R A. The Right of Privacy[J]. Georgia Law Review，1978，12(3)：393-422.

[15] JEROME J W. Buying and selling privacy：Big data's different burdens and benefits[J]. SSRN Electronic Journal，2013，66(47)：47-54.

⑩ See Bill Gates.The Age of AI has begun Artificial intelligence is as revolutionary as mobile phones and the Internet,https://www.gatesnotes.com/The-Age-of-AI-Has-Begun.

[16] BURDON M. Contextualizing the Tensions and Weaknesses of Information Privacy and Data Breach Notification Laws[J]. Santa Clara Computer&High Technology Law Journal,2010,27(1):63-130.

[17] CHIK W B,PANG J K Y. The Meaning and Scope of Personal Data under the Singapore Data Protection Act[J]. Singapore Academy of Law Journal,2014,26(2):354-397.

[18] 杨立新. 个人信息:法益抑或民事权利:对《民法总则》第111条规定的"个人信息"之解读[J]. 法学论坛,2018(1):34-45.

[19] 郑晓剑. 个人信息的民法定位及保护模式[J]. 法学,2021(3):116-130.

[20] 张涛. 探寻个人信息保护的风险控制路径之维[J]. 法学,2022(6):57-71.

[21] 商希雪. 超越私权属性的个人信息共享:基于《欧盟一般数据保护条例》正当利益条款的分析[J]. 法商研究,2020(2):57-70.

[22] 申卫星. 论个人信息保护与利用的平衡[J]. 中国法律(中英文版),2021(5):24-29,103-109.

[23] 宋伟卫. 处理已公开个人信息的刑法边界[J]. 吉林大学社会科学学报,2022(6):71-82,232-233.

[24] 储陈城. 大数据时代个人信息保护与利用的刑法立场转换:基于比较法视野的考察[J]. 中国刑事法杂志,2019(5):48-62.

[25] 龙卫球. 中华人民共和国个人信息保护法释义[M]. 北京:中国法制出版社,2021:5.

[26] 高富平. 基于规范目的的个人信息治理规则[J]. 中国应用法学,2022(6):111-127.

[27] 许可. 诚信原则:个人信息保护与利用平衡的信任路径[J]. 中外法学,2022(5):1143-1162.

[28] 李振林,潘鑫媛. 生成式人工智能背景下数据安全的刑法保护困境与应对:以ChatGPT为视角的展开[J]. 犯罪研究,2023(2):25-33.

[29] 珍妮·斯蒂尔. 风险与法律理论[M]. 韩永强,译. 北京:中国政法大学出版社,2012:5-6.

[30] 乌尔里希·贝克. 风险社会[M]. 何博闻,译. 南京:译林出版社,2004:35.

[31] 安东尼·吉登斯. 失控的世界:全球化如何重塑我们的生活[M]. 周红云,译. 南昌:江西人民出版社,2001:18.

[32] ANDORNO R. The precautionary principle: A new legal standard for a technological age[J]. Journal of International Biotechnology Law,2004,1(1):11-19.

[33] SEILER H. Harmonised Risk Based Regulation:A legal Viewpoint[J]. Safety Science,2002,40(1/2/3/4):31-49.

[34] 孙政伟. 大数据时代侵权法功能定位的历史转型[J]. 民商法论丛,2020(1):125-146.

[35] 何国强. 风险社会下侵权法的功能变迁与制度建构[J]. 政治与法律,2019(7):93-104.

[36] 崔聪聪. 数据限制处理权的法理基础与制度建构[J]. 比较法研究,2022(5):75-88.

[37] 齐爱民. 大数据时代个人信息保护法国际比较研究[M]. 北京:法律出版社,2015:245.

[38] 张凌寒. 商业自动化决策的算法解释权研究[J]. 法律科学(西北政法大学学报),2018(3):65-74.

[39] 王叶刚. 个人信息处理者算法自动化决策致害的民事责任:以《个人信息保护法》第24条为中心[J]. 中国人民大学学报,2022(6):47-59.

[40] PRINS C. Making our body identify for us:Legal implications of biometric technologies[J]. Computer Law&Security Report,1998,14(3):159-165.

[41] 京东法律研究院. 欧盟数据宪章:《一般数据保护条例》GDPR评述及实务指引[M]. 北京:法律出版社,2018:8.

[42] 肖峰. 何种生成? 能否创造:ChatGPT的附魅与祛魅[N]. 中国社会科学报,2023-03-06(5).

[43] 张凌寒. 生成式人工智能的法律定位与分层治理[J]. 现代法学,2023(4):126-141.

[44] 阮神裕. 民法典视角下个人信息的侵权法保护:以事实不确定性及其解决为中心[J]. 法学家,2020(4):29-39,192.

[45] 孔祥俊. 论侵权行为的归责原则[J]. 中国法学,1992(5):70-77.

[46] 张新宝,李倩. 惩罚性赔偿的立法选择[J]. 清华法学,2009(4):5-20.

The challenge and response to personal information protection for generative artificial intelligence

ZHU Rongrong

(*School of Humanities and Arts*, *China University of Mining and Technology*, *Xuzhou* 221116, *P. R. China*)

Abstract: Generative artificial intelligence, represented by ChatGPT and DeepSeek, refers to artificial intelligence that can generate text, pictures, videos, and other corresponding content according to user instructions. Personal information is the basis of generative artificial intelligence. Generative artificial intelligence needs to deal with a large amount of personal information in all stages of model training, model generation and model optimization, which also brings a certain impact on the traditional personal information protection rules. In the stage of information collection, generative artificial intelligence may blur the rules of informed consent and infringe the privacy of information subjects. In the stage of information utilization, generative artificial intelligence may impact the basic personal information processing rules such as the principle of purpose limitation and the principle of openness and transparency, and increase the risk of personal information disclosure. In the stage of information generation, generative artificial intelligence may generate false information and discriminatory information. In the context of the transformational development of generative artificial intelligence, it is urgent to examine the basic concept of personal information protection and seek its value orientation in the field of generative artificial intelligence. By examining the development of comparative law and the concept of personal information protection in China, we can see that the unipolar thinking of personal information protection or personal information utilization is difficult to adapt to the practical needs of the digital society, and the dynamic balance between personal information protection and personal information utilization is the ideal path to properly balance the interests of various subjects. Generative artificial intelligence can be widely used as a basic model in many fields such as education, finance, science, and technology. In view of this, a balance between personal information protection and generative artificial intelligence development should be coordinated and promoted. Personal information is closely related to the information subject. Once personal information is disclosed or abused, the information subject may suffer higher risks. Therefore, it is necessary to build a collaborative relief mechanism of risk prevention and damage compensation, to promote the benign development of generative artificial intelligence based on whole life cycle protection of personal information. As far as the risk prevention mechanism is concerned, it is necessary to improve the de-identification measures based on risk identification, and grant the information subject the right to limit processing and the right to interpret algorithms, to curb the potential risks in an all-round way. Determination of the subject of liability is the basis of damage compensation. It should be proved by the service provider of generative artificial intelligence and the user that there is no causal relationship between them and the damage, otherwise they need to bear joint and several liability for compensation. In terms of the principle of imputation, the principle of presumptive fault liability or no-fault liability can be applied respectively according to the fact that the infringed object is personal general information or personal sensitive information. In order to better remedy the damage suffered by the information subject, in addition to the traditional compensatory compensation such as property damage compensation and mental damage compensation, punitive compensation should also be introduced to protect the damaged rights and interests of the information subject to the greatest extent.

Key words: generative artificial intelligence; ChatGPT; DeepSeek; personal information; risk prevention; tort liability

(责任编辑　刘　琦)