生成式人工智能使用者注意义务的规范来源

刘承韪　马瑞聪

摘要：在生成式人工智能全生命周期治理要求下，对使用者行为的法律监管不可或缺。鉴于生成式人工智能具体应用场景与使用者身份的多样性差异，使用者的注意义务类型与内容应当作区分设计。原则上，使用者享有使用自由，但不排除其因合同性质而承担亲自履行义务，或因制造禁止型使用风险而被禁止使用的可能。使用者通常承担基于交往安全义务产生的内容标识义务与结果审查义务。但在商业服务与个人信息处理活动中，使用者还需履行必要的信息披露以及解释说明义务，以满足消费者与个人信息主体的特别保护需求。对于可能引发高度使用风险的专业使用者，还应当承担风险管理与预防、输入数据质量控制、系统更新与维护等义务。

关键词：生成式人工智能；深度伪造；使用者义务；注意义务；DeepSeek

DOI：10.19836/j.cnki.37-1100/c.2025.05.014

一、问题的提出

以ChatGPT、DeepSeek等为代表的人工智能模型被广泛应用于文本创作、图像生成、智能对话等多元领域。与此同时，由新兴技术应用引发的虚假信息传播、自动化偏见等系统性风险，对现行法律提出了治理挑战①。生成式人工智能技术在应用阶段的风险转化主要取决于使用者。结合我国《人工智能示范法3.0》的定义，使用者是指“依照人工智能的性能和用途对其加以利用的个人、组织”。使用者的控制能力通常表现为指令输入、参数调整与场景选择。无论是通过提示词工程引导内容生成方向，还是利用深度合成技术制造虚假信息，使用者的具体操作都直接影响技术应用的合规边界。对此，应当首要关注现行法律如何评价使用者行为。

当前我国人工智能治理呈现法律与各项规章制度相结合的多层次治理规范体系②。受规范制定主体权限与目标的影响，现行规章制度主要聚焦于人工智能的技术风险治理与服务提供者监管③，而未直接规范使用者。因此，当深度合成内容侵犯他人合法权益时，仍需依据《中华人民共和国民法典》(以下简称民法典)第1165条过错责任原则、《中华人民共和国个人信息保护法》(以下简称个保法)第13条个人信息处理合法性基础等法律规则评价使用者行为。但在面对生成式人工智能技术的复杂性与不透明性问题时，传统以过错为中心的责任认定机制已难以应对④。

在此背景下，明确生成式人工智能使用者的注意义务及其标准，具有双重实践意义。一方面，使用者的注意义务，是指使用者按照生成式人工智能的性质或用途，在使用过程中应当承担的必要的审

基金项目：北京市社科基金重点项目“北京市文化产业立法研究”(24FXA001)。

作者简介：刘承韪，中国政法大学比较法学院教授，博士生导师(北京 100088；lexway@126.com)；马瑞聪，中国政法大学比较法学院博士研究生(北京 100088；17808087450@163.com)。

① 林北征：《论生成式人工智能服务提供者的注意义务》，《法律适用》2024年第10期。

② 高志宏：《回应与超越：生成式人工智能法律规制——以〈生成式人工智能服务管理暂行办法〉为视角》，《社会科学辑刊》2024年第5期。

③ 张凌寒：《深度合成治理的逻辑更新与体系迭代——ChatGPT等生成型人工智能治理的中国路径》，《法律科学》2023年第3期。

④ 丁晓东：《从网络、个人信息到人工智能：数字时代的侵权法转型》，《法学家》2025年第1期。

慎义务。通过明晰注意义务的类型与内容，能够为使用者从事合规行为提供正向引导，避免因其自身过错而侵害他人权益。另一方面，注意义务标准是对过错责任认定的客观外化[①]。司法者不必陷入主观过错证明之窠臼，只需以社会交往中的一般行为标准为依据判断使用者应否承担过错责任。基于此，通过结合生成式人工智能具体应用场景、使用风险等级、使用者身份、与上游服务提供者义务内容衔接等核心要素，可以对使用者的注意义务体系进行分层设计。首先应考察法律是否事先禁止使用生成式人工智能；若未禁止，则需进一步区分一般使用与特殊使用两种情形，分别从生成合成内容标识、内容质量审查、必要信息提供、风险管理与预防等多个方面规范使用者注意义务。

二、生成式人工智能禁止使用的注意义务来源

使用者原则上享有使用自由，但在特定应用场景中，不排除生成式人工智能本身被禁止使用的可能。即便在体现意思自治的合同领域内，债务人借助生成式人工智能履行合同义务亦可能因违反亲自履行义务而不被允许。

（一）基于生成式人工智能使用风险的一般禁止使用义务

现如今，人工智能生成内容已经达到公众难辨真伪的深度伪造程度。为防止虚假信息泛滥引发公共信任危机，《互联网信息服务深度合成管理规定》（以下简称《深度合成管理规定》）第6条、《生成式人工智能服务管理暂行办法》（以下简称《生成式人工智能管理办法》）第4条均要求，使用者不得利用深度合成服务或生成式人工智能制作、复制、发布、传播违法、有害、虚假信息。然而，这些禁止性规则在实务中面临执行难的困境。究其原因，在于规章制定者仅从一般性的禁止违法或侵权角度出发，对生成式人工智能的使用方式及用途作出限制，而缺少结合生成式人工智能的技术特征以及应用场景的具体化标准。据此，应当采取何种标准对生成式人工智能的使用提出明确的限制条件，或者采取何种方式从根本上减少生成式人工智能被滥用于非法途径的可能，亟待讨论。

鉴于生成式人工智能在使用过程中可能引发侵犯他人人格权益的侵权风险，有学者主张应当将基于深度合成技术创建的系统一律归为高风险人工智能系统[②]。在此前提下，通过提高对系统开发者与服务提供者的技术要求，加强生成式人工智能在应用阶段的风险防治。在比较法上，欧盟《人工智能法案》采取了人工智能系统风险分级划分的差异化规则建构模式，其中第8—15条对高风险人工智能系统的设计与开发提出了一系列要求，包括但不限于建立与高风险人工智能系统有关的风险管理系统，使用符合质量标准的数据集进行开发、编制技术文件，自动记录系统事件等。如果生成式人工智能系统无法满足上述技术条件，便无法投入使用。然而，这一观点的合理性有待进一步检验。

值得关注的是，深度伪造内容只是生成式人工智能基于中立技术创造的产物，至于其被用于何处取决于使用者的个人意志，并不是系统设计与开发阶段能够决定的。即便是在欧盟《人工智能法案》语境下，所谓高风险型也仅指向人工智能系统本身的风险等级，而非系统被滥用导致的高度使用风险。然而，一个客观现实是，深度伪造内容的创作者通常“知法犯法”，即便他们完全知晓其行为违法，仍会故意为之。因此，无论从系统设计和开发层面提出多高的要求，都无法有效减少使用者的滥用行为。总之，提升人工智能系统风险等级的方案并不能用于降低生成式人工智能系统在应用阶段的使用风险。

考虑到生成式人工智能应用的广泛性，对其禁止使用情形的考察不能仅局限于规制人工智能系统本身，还应当结合具体的应用场景。现如今，生成式人工智能时常被滥用于制造、传播包含色情、暴

① 程啸：《侵权责任法》（第三版），北京：法律出版社，2021年，第301页。

② Block M. J., “A Critical Evaluation of Deepfake Regulation through the AI Act in the European Union”, *Journal of European Consumer and Market Law*, 2024, 13(4), pp. 191-192.

力等内容的深度伪造视频[①]。这些视频中的角色形象逼真,对公众具有极强的欺骗性与误导性。如不予以禁止,势必对第三方合法权益造成不可逆的损害。对此,为满足我国的人工智能治理需求,有必要对上述禁止型生成式人工智能使用风险作出更为精确的法律描述,以供未来执行。

考察域外立法例后发现,欧盟《人工智能法案》第5条第1款列举了诸多人工智能系统应当被禁止投入市场、提供服务或加以使用的情形,其中第(a)项所描述的内容与深度伪造技术的非法用途最为相似。其主要特征为:(1)被禁止的系统采用了超出个人意识的潜意识技术或故意操纵、欺骗性技术;(2)使用该类技术的目的是实质性地扭曲个人或群体的行为;(3)使用后果是显著损害该人或该群体的知情决定能力,使其作出本不会作出的决定。除去其对技术特征的描述不作参考外,其余二者所指涉的使用目的和使用后果对上述问题的解决具有启发意义。

与之相似,使用者滥用生成式人工智能的目的通常也是利用深度伪造内容欺骗他人,最终使其作出违背自己真实意思的决定。因此,在我国法律语境下,当生成式人工智能被用于创建深度伪造内容以故意操纵或欺骗特定主体,并会显著侵害该主体或其他主体的合法权益时,将会产生禁止型使用风险,使用者即负有一般性禁止使用义务。对使用者可能引发的高度使用风险进行规制并不以过度拔高系统开发与设计要求为必要,而应当从提高使用者注意义务标准的角度入手予以解决。

(二)基于债务人亲自履行合同义务的特别禁止使用义务

在合同磋商阶段,债务人的履行能力与资质是决定债权人缔约意愿的关键因素。然而,深度合成技术的出现在很大程度上改变了债务人的合同履行方式。以提供文本创作服务为例,生成式人工智能提供的文本生成服务不仅可以成为债务人创作过程中的辅助工具,甚至可以直接取代债务人的全部创作行为。尽管这为债务人履行合同义务创造了极大便利,却使债权人被迫承担了由技术缺陷导致服务瑕疵的违约风险。特别是当按照民法典有关规定,债务人应承担亲自履行合同义务时[②],生成式人工智能的介入很有可能违背债权人订立合同的初衷。对此,应首先分析生成式人工智能在合同履行关系中的法律地位,然后再讨论债务人应当在何时避免借助生成式人工智能履行合同义务。

债务人的亲自履行义务是承揽合同、委托合同等劳务给付合同的基本要求。承揽合同与委托合同的共同之处在于,合同的订立都建立在债权人对承揽人或受托人有能力处理承揽工作或委托事务的信任基础上[③]。依据民法典第772条和第923条之规定,承揽合同要求承揽人以自己的工具、技术和劳力完成主要工作,委托合同则要求受托人亲自处理委托事务。因此,在未经定作人或委托人同意的情况下,承揽人将承揽的主要工作交由第三人完成,或者受托人擅自转委托于第三人的,均构成根本违约。问题在于,当债务人借助生成式人工智能完成主要工作或委托事务时,生成式人工智能是否属于上述合同履行关系中的"第三人"?

从技术原理来看,生成式人工智能通常采用两种特殊的大模型:"Transformer"或"GANs"。以"Transformer"模型为例,其通过深度学习输入数据,产生常用词汇上的概率分布,并在文本输出时使用该概率分布选择下一个最有可能生成的词汇。以此类推,直至其认为文本无法继续生成[④]。这种"文字接龙"的背后实则隐含着对已学习数据相似性的追求,输出文本很可能构成对版权作品的使用与复制[⑤]。与之相似,以生成器和判别器为组成的"GANs"模型,本质上是由生成器负责生成数据来"欺骗"

① 王璇、宋春龙:《基于深度伪造技术的"二创"视频伦理风险及规制治理研究》,《西南民族大学学报(人文社会科学版)》2024年第5期。

② Müller/Glöge, in: Münchener Kommentar BGB, 9. Aufl., 2023, § 613 Rn. 2.

③ 黄薇主编:《中华人民共和国民法典合同编释义》,北京:法律出版社,2020年,第625、874页。

④ 张凌寒:《生成式人工智能的法律定位与分层治理》,《现代法学》2023年第4期;周学峰:《生成式人工智能侵权责任探析》,《比较法研究》2023年第4期。

⑤ 刁佳星:《生成式人工智能服务提供者版权侵权注意义务研究》,《中国出版》2024年第1期。

判别器，在不断地生成与对抗过程中，最终输出判别器认为最有可能区别于真实数据的生成数据①。由此可见，两种模型并非以生成最具有人类可理解性、最符合现实生活真实性的内容为目标，而只在一定概率上生成符合使用者要求的内容。生成式人工智能并非真正具有人类智慧，也难以成为学理上所称的“电子人”（ePerson）②。我国学者同样主张，不宜承认生成式人工智能的法律主体地位③，而应当将其作为民事法律关系中的客体对待④。综合上述分析，尽管生成式人工智能不属于上述合同履行关系中的“第三人”，但这并不意味着使用生成式人工智能会被当然允许。此时，问题便转化为，债务人将合同债务交由生成式人工智能完成时，是否还存在违反亲自履行义务的可能？

当生成式人工智能完全取代债务人自身行为时，合同履行过程已然从债务人亲自劳作，转化为正确输入提示词、调整参数、检查并改进生成文本。此时合同履行基础已经发生重大变更，债务人因此节省了大量的时间与经济成本，不免引发合同履行成本是否仍然与约定对价相当的质疑⑤。毕竟从一般理性经济人的角度出发，债权人给出的高额对价应当包含债务人亲自履行这一基本条件。倘若允许债务人以畸少于预期合同履行成本的方式完成合同义务，势必有损债权人的经济利益。此外，合同服务内容与成果质量也是债权人重点关注之处。不同于委托合同仅以劳务给付本身为标的，承揽合同是以结果为导向的债务，其不仅要求承揽人提供劳务，也必须完成定作人指定的工作成果⑥。对此，依据民法典第781条之规定，承揽人还须保证工作成果符合质量要求。倘若人工智能生成内容存在质量过低、内容错误等问题，承揽人未能予以纠正或及时避免损害发生的，应当向定作人承担违约责任。实际上，承揽人时刻面临着生成合成内容质量不一的不确定性，因为在强大的自主生成能力背后是大量训练数据的输入，而互联网上可供获取的数据来源既包含受保护的个人数据，也包含非法、有害数据。一旦训练数据存在偏差，或者出现模型过拟合，极有可能产生承揽人无法预见或避免的虚假信息，也即出现“机器幻觉”（hallucination）⑦。因此，对于追求成果质量的定作人而言，深度合成技术的不可信赖性使其通常无法接受生成式人工智能完全取代承揽人工作的事实。

需要补充的是，合同一方是否允许另一方使用生成式人工智能，还需以另一方主动披露这一事实为前提。对此，合同当事人是否在合同订立过程中便负有向相对方主动披露的义务，有赖于附随义务内容之解释。依据民法典第500条第2项的规定，当事人不得“故意隐瞒与订立合同相关的重要事实或者提供虚假情况”，其目的在于要求一方在未被询问时，仍主动将可能明显影响合同决策的情况告知另一方，也即履行信息提供义务（informationspflichten）⑧。在生成式人工智能的应用场景中，智能对话、合成人声、人脸生成等深度合成服务都有可能使人们产生对现实的混淆或误认，进而影响其在合同订立过程中的决策判断。例如，消费者接受“人工智能”商家的商品推荐服务并受此影响而订立买卖合同，其背后则隐藏着个性化推送服务的“算法黑箱”；再如，债务人以人工智能辅助绘画作为服务提供的一部分，这一事实决定了另一方如何评价合同履行的基础条件。因此，当生成式人工智能的介入足以构成影响合同订立及履行的重要事实时，一方有义务向另一方主动披露使用生成式人工智能的基本情况。此时的信息提供义务并非仅为防止另一方的人身、财产权益受损，更在于保护决策自由这一特殊利益⑨。

总而言之，在委托合同、承揽合同等特殊的劳务给付关系中，基于民法典中亲自履行合同义务的

① 彭桂兵、张凤逸：《“深度伪造”侵犯肖像权问题研究》，《湖南师范大学社会科学学报》2023年第6期。

② Riehm, Nein zur ePerson!, RDi 42 (2020), S. 46f.

③ 黎四奇：《对人工智能非法律主体地位的解析》，《政法论丛》2023年第5期。

④ 刘智慧：《民法典视域下人工智能法律规制论纲》，《学术交流》2023年第9期。

⑤ Schaub, Nutzung von künstlicher Intelligenz als Pflichtverletzung?, NJW 2145 (2023), S. 2146f.

⑥ 游冕：《〈民法典〉第919条（委托合同的定义）评注》，《南大法学》2023年第4期。

⑦ Ziwei J., et al., “Survey of Hallucination in Natural Language Generation”, *ACM Computing Surveys*, 2022, 1(1), pp. 3-4.

⑧ Bachmann, in: Münchener Kommentar BGB, 9. Aufl., 2022, § 241 Rn. 184.

⑨ Bachmann, in: Münchener Kommentar BGB, 9. Aufl., 2022, § 241 Rn. 57-58.

要求,债务人应当承担必要的禁止使用义务。这一义务要求债务人不得以生成式人工智能完全取代其合同履行行为,仅可以辅助履行的方式对生成式人工智能加以利用。至于债权人如何知晓债务人借助生成式人工智能履行合同义务这一事实,则有赖于债务人积极履行主动披露的附随义务。

三、生成式人工智能使用的一般注意义务来源

生成式人工智能的广泛应用可能带来有害信息泛滥、自动化偏见等诸多负面影响①。囿于生成式人工智能创作过程中的"黑箱"属性,人们难以从技术上追溯侵权行为发生的根本原因。在此背景下,现行规定不断提高对提供者的技术要求以加强对合成内容的监管。然而,课以提供者过重的主动审查义务并不现实,不仅会导致逐底竞争,也会抑制科技创新②。相比之下,使用者作为输出端的决策者与受益者,却尚未切实履行必要的风险防范义务。据此,使用者承担注意义务的法理基础为何,其应当如何对深度合成内容进行正确标识,并在多大程度上审查生成内容的合法性与真实性,这些问题殊值讨论。

(一)生成式人工智能使用时的注意义务证成

生成式人工智能侵权行为具有偶发性,凡是在数据训练、加工等阶段出现数据质量瑕疵情形的,都有可能生成合成有害内容③。此外,使用者自身操作不当,例如错误使用提示词等,也有可能成为有害内容产生的主要原因④。由此可见,生成式人工智能侵权事件并非都是由使用者故意利用生成式人工智能导致的,相反,使用者通常存在未能合理预见或避免有害内容产生的过失心理。此时,注意义务成为使用者是否承担侵权责任的判断依据,即因为使用者未能履行合理注意义务而造成第三人合法权益损害的,应当认定使用者有过错⑤。

我国法上行为人承担不作为侵权责任的理论依据主要借鉴自德国的"交往安全义务"(verkehrspflichten)⑥,其旨在规范行为人应当保护第三人合法利益的一般法律义务⑦。根据行为人与被监控危险源或受威胁法益之间的特定关系,交往安全义务可以分为危险源控制义务与第三人法益保护义务⑧。前者系多数学者主张提供者应当承担安全保障义务的主要依据⑨,后者则可以作为使用者承担注意义务的正当性基础。

德国法上的第三人法益保护义务源起于"运输公司案"⑩和"兽医案"⑪。两案判决的核心要点是:专业人士基于职业身份应当承担一定的风险控制责任,却未能尽到与其专业地位相匹配的注意义务,

① 支振锋:《生成式人工智能大模型的信息内容治理》,《政法论坛》2023年第4期。

② 林北征:《论生成式人工智能服务提供者的注意义务》,《法律适用》2024年第10期。

③ 高阳:《通用人工智能提供者内容审查注意义务的证成》,《东方法学》2024年第1期。

④ 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》,《比较法研究》2023年第5期。

⑤ 最高人民法院民法典贯彻实施工作领导小组主编:《中华人民共和国民法典侵权责任编理解与适用》,北京:人民法院出版社,2020年,第279页。

⑥ 冯珏:《安全保障义务与不作为侵权》,《法学研究》2009年第4期;李昊:《交易安全义务论——德国侵权行为法结构变迁的一种解读》,北京:北京大学出版社,2008年,第243页。

⑦ Wagner, in: Münchener Kommentar BGB, 9. Aufl., 2024, § 823 Rn. 482.

⑧ Wagner, in: Münchener Kommentar BGB, 9. Aufl., 2024, § 823 Rn. 500.

⑨ 杨显滨:《生成式人工智能服务提供者间接侵权责任的承担与限制》,《法学家》2024年第3期;高阳:《通用人工智能提供者内容审查注意义务的证成》,《东方法学》2024年第1期;胡巧莉:《人工智能服务提供者侵权责任要件的类型构造——以风险区分为视角》,《比较法研究》2024年第6期。

⑩ 汉堡一家滚轮运输公司只为运输车配备一名马车夫,因此当马车夫离开、货车短时间无人看管时,货物有可能被盗。RGZ 102, 38 (42 f.).

⑪ 一名兽医被要求紧急屠宰一头患有炭疽病的奶牛,但未能保护一名手上有伤口的屠夫免受感染。RGZ 102, 372 (374 f.).

导致第三人因可预见的危险失控而遭受损害。因此，若要认定由行为人对其行为引起或增加的危险承担避免第三人权益因此遭受损害的注意义务，首先需要满足一个前提条件，即行为人是客观上最具风险控制可能性的人。换言之，行为人如果需要以不合比例的经济成本为代价来避免第三人合法权益受到侵害，则不宜作为交往安全义务的承担者①。据此，在生成式人工智能应用场景下，须进一步检验使用者的风险控制能力。

我国《人工智能示范法3.0》设置了个人或家庭使用人工智能不适用本法的例外情形。此为比较法上的通行做法，例如，欧盟《人工智能法案》使用“部署者”（deployer）概念，便是为了排除个人非专业（non-professional）活动使用人工智能的情形，并在用语上与普通用户相区别②。不可否认的是，专业使用者通常掌握算法运作机理，且有能力合理预见大模型可能产生的偏见放大、信息失真等系统性风险，相较于普通用户更具风险控制能力。不仅如此，企业级使用者还通常具备部署内容过滤系统、建立内容审查机制的条件，由其承担因使用行为引起的第三人法益保护义务更具技术和经济上的合理性。然而，如果将个人使用行为或非专业活动全部排除在外，则存在豁免范围过大的嫌疑③。因为在绝大多数情况下，深度合成技术对第三人法益侵害的危险性恰恰是普通用户不当使用生成式人工智能使然。普通用户通常能够轻易获取并操控生成式人工智能，若不要求其承担一定程度上的第三人法益保护义务，对受害者而言有失公允。亦如民法典第1019条明确禁止个人利用信息技术手段伪造他人肖像。考虑到深度合成技术既可以合成肖像，也可以合成人声、文字等，有学者主张，可以将上述禁止性规定扩张适用于全部的人格权益及著作权④。因此，无论是普通用户还是专业使用者，都应当受到第三人法益保护义务的规制，也即应当在使用生成式人工智能过程中承担必要的注意义务。鉴于不同使用者在技术水平和经济能力上的明显差异，使用者的注意义务应当作分层设计：先讨论一般应用场景中，所有使用者都应当承担的注意义务类型与内容；然后再讨论商业服务、个人信息处理活动、高风险领域等具体应用场景中，专业或企业级使用者应当额外承担的注意义务。

（二）生成式人工智能使用者的内容标识义务

内容标识义务是应对生成合成内容侵权、虚假信息传播风险的必要前提⑤。因为只有使受影响的第三人知晓其接受或使用的内容来自人工智能，才有进一步审查内容合法性、提出算法解释要求、寻求权利救济的可能性。对此，我国《深度合成管理规定》《人工智能生成合成内容标识办法》（以下简称《内容标识办法》）等规定对提供者与使用者均提出了较为全面的内容标识义务。内容标识本质上也是贯彻人工智能全生命周期治理中“透明原则”的一部分⑥。在此前提下，如何衔接与协调上游提供者与下游使用者之间的义务关系，需要进行详细讨论。

首先应当明确，提供者应当为使用者自行标识生成合成内容创造必要的技术条件。依据《深度合成管理规定》第17条第2款和《内容标识办法》第8条的规定，提供者应当为使用者提供显著标识功能，并在用户服务协议中为用户说明标识方法、样式等，提示其遵守相关的标识管理要求。相比于提供者承担普遍意义上的隐式标识义务（《深度合成管理规定》第16条），使用者的内容标识义务并不具有普遍性，而仅发生在特定情形之下。

一是，使用者对构成深度伪造的生成合成内容负有内容标识义务。由于深度伪造内容与真实的人、物或事件之间具有极高的相似性，极易导致第三人混淆或误认，并引发网络诈骗、信任危机，应当

① Wagner, in: Münchener Kommentar BGB, 9. Aufl., 2024, § 823 Rn. 502-503.

② Block M. J., “A Critical Evaluation of Deepfake Regulation through the AI Act in the European Union”, *Journal of European Consumer and Market Law*, 2024,13 (4), pp. 184-192.

③ Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 117 (2024), S. 121.

④ 王利明：《生成式人工智能侵权的法律应对》，《中国应用法学》2023年第5期。

⑤ 支振锋：《生成式人工智能大模型的信息内容治理》，《政法论坛》2023年第4期。

⑥ Luise Merkle, Transparenz nach der KI-Verordnung - von der Blackbox zum Open-Book?, RDi 414 (2024), S. 417ff.

予以重点规制。然而,应该如何辨别深度伪造内容并建立一个一般性的认定标准?使用者个体差异与技术难度是其中的难点所在。例如,掌握技术原理的年轻人通常比老年人更易识别某一图像的改动迹象。在技术上,图片或视频内容中人物牙齿和眼睛部分轮廓模糊及不清晰的过渡虽然可以被视为人工合成迹象①,但这种常人难以识别的细节无法成为辨别深度伪造内容并使其免于标识的一般方法。因此,在深度伪造内容识别方法无法普及的背景下,不应当对"相似性"识别设置过高标准,只要日常使用社交媒体且具有一般见识的用户快速浏览时有可能认为某一深度伪造内容是真实内容即可。此外,为进一步解决使用者不作为的问题,我国《深度合成管理规定》第17条第1款将深度伪造内容的识别与标识义务主要交由提供者承担。具体而言,提供者应当在提供模拟自然人对话或写作、改变个人身份特征等与自然人人身属性相关的编辑服务时,承担显著标识义务。当深度合成技术的应用领域与自然人的人身利益密切联系时,由提供者采取技术措施,在使用者获取之前便对特定内容进行显著标识,有助于降低因使用者不作为而造成公众混淆或误认的风险。然而,对于涉人身属性外的其他深度伪造内容,《深度合成管理规定》第17条第2款不仅豁免了提供者的显著标识义务,且只要求使用者可以自由决定是否进行显著标识。如此规定显然不利于深度伪造内容的"公开化",并会加剧公共信任危机。因为从短期内的经济效益来看,使用者作为生成合成内容的实际控制者,采取拒不标识的不诚信行为带来的收益反而更佳②。相比之下,欧盟《人工智能法案》第50条第4款已经规定,部署者应当主动披露深度伪造内容系人为生成或操纵的,这一做法值得借鉴。据此,我国法也应当明确,当生成合成内容构成深度伪造时,使用者有义务对其进行显著标识。

二是,我国《内容标识办法》第10条特别规定,用户(使用者)在将生成合成内容(不构成深度伪造)上传至网络信息内容传播平台时,负有显著标识义务。这一规定能够从源头上降低生成合成内容未经标识而在网络上传播的可能性,以此回应因不明来源的有害内容引发网络暴力、欺诈等乱象的现实治理需求。需要注意的是,《内容标识办法》第9条同时为使用者保留了最终决定是否对生成合成内容进行显著标识的自主权。例如,当生成合成内容构成具有艺术性、创造性、讽刺性的作品时,尤其是视觉性作品,显著标识可能有碍使用者对作品进行展示或表达。此时,使用者对其作品的展览权或者以作品形式行使言论自由权的利益更值得保护。不仅如此,当使用者利用生成式人工智能创造的内容构成著作权法意义上的作品时,能否免于履行显著标识义务,一定程度上也影响着作品的可流通性与财产价值。基于上述理由,我国法允许提供者在向用户告知标识义务和使用责任后,按用户要求提供不包含显著标识的生成合成内容,同时保留日志信息不得低于六个月的做法值得肯定。

明确使用者的内容标识义务也有利于确定其侵权责任。例如在"上海新创华案"中,法院以提供者未履行内容标识义务为由,认定其存在过错并需承担侵权责任③。然而,履行内容标识义务并不足以构成侵权责任的免责事由,也不会化解因黑箱效应导致的侵权原因追溯困难。一方面,诸如互联网上传播的深度伪造色情内容,尽管对其进行了标注,但经证实的虚假内容仍会继续传播,受害者仍然受到虚假内容的持续侵害。另一方面,当生成合成内容引发对既有作品的侵权争议或著作权归属争议等著作权纠纷时,司法裁判仍然缺少必要的事实证据。归根结底,当前我国《网络安全技术 人工智能生成合成内容标识方法》仅以"人工智能生成合成内容""AIGC"等字样对接触者作出信息提示的要求并不足够。未来应当从技术层面实现一般用户可查的、更为详细的信息标注功能。例如,增加生成合成内容参与创作者身份、权属设定、具体使用条件等权利描述性信息,以及增加体现AI贡献度的"AI辅助生成"与"AI自主生成"标识等④。

① Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 117 (2024), S. 120.

② 张继红:《生成式人工智能生成内容标识义务研究》,《法商研究》2024年第4期。

③ 广州互联网法院(2024)粤0192民初113号民事判决书。

④ 陈俊凯:《人工智能生成内容信息披露机制构建研究》,《中国科技论坛》2024年第3期。

（三）生成式人工智能使用者的结果审查义务

对生成式人工智能输出内容的结果审查义务，是避免因有害内容传播而损害他人合法权益的最重要防线。由于生成合成内容不计其数，若由提供者承担对输出结果进行有害内容筛查的全部工作，既存在技术障碍，也不符合经济理性[①]。反而是使用者作为输出内容的实际控制者，更应当承担必要的结果审查义务。现行法虽然对使用者的使用行为有所规范，但对其结果审查义务着墨不多。考虑到生成式人工智能决策过程的不透明性，使用者对数据来源合法性与真实性的追查能力十分有限。因此，《深度合成管理规定》第6条和第9条仅要求使用者不得生成合成禁止性信息或虚假新闻信息，并应承担信息安全义务。可见，使用者对输出结果的审查范围限于法律禁止性内容或虚假内容，上述规定也可以成为使用者承担结果审查义务的特别法依据。然而，对于那些不具有明显违法性的或难辨真伪的生成合成内容，使用者是否以及在何种程度上履行结果审查义务，这些问题尚未解决。

从侵权法角度出发，生成式人工智能本身并不能独立承担民事责任，而是作为辅助工具用于生产和提供服务，因此如若发生侵权行为，应当由使用者对其输出结果承担最终责任。使用者履行输出结果审查的注意义务，应当采取“理性人标准”，即根据使用者所属行业、年龄层等因素，将普通个体通常具备的一般注意力作为标准[②]，要求使用者合理预见有害内容的生成及传播风险，并采取合理措施加以避免[③]。考虑到生成式人工智能系统正处于研发与试验阶段，且多数情况下向公众免费开放，因此需要在义务设置与鼓励使用之间建立平衡。一方面，不宜过分要求使用者对生成式人工智能的一切输出结果进行审查；另一方面，应当建立使用者对输出结果最低质量标准的合理预期。据此，唯有在使用者被告知生成式人工智能的具体功能和局限性时，才能期待其在某一具体应用场景中合理预见侵权危险，并履行必要情形下的结果审查义务。而这取决于提供者如何向下游使用者履行信息提供义务。

《生成式人工智能管理办法》第10条要求提供者明确其服务的人群、场景及用途，并指导使用者科学理性认识和依法使用生成式人工智能。这一规定虽然能够为提供者履行信息提供义务提供规范指引，但仍有进一步细化的空间。比较法上，欧盟《人工智能法案》第13条对高风险人工智能系统提供者的信息提供义务有所提及。尽管高风险等级划分并不适合我国的生成式人工智能治理方案，但是施以提供者必要的信息提供要求并不会过度提高合规成本、阻碍技术进步。因此，转换至我国法律语境，可以从以下几个方面强化提供者的信息提供义务：首先，提供者应当确保系统操作的透明度，使得使用者能够解释系统的输出结果并正确使用。其次，提供者应当以附带数字格式等方式，向使用者提供易于获取且可理解的使用说明。再次，使用说明中除包含提供者的身份信息外，还应包括对系统特点、能力和性能限制的必要描述，例如系统应用的预期目的、准确性水平、全部已知或可预见的侵权风险、对特定使用群体的性能表现以及输入数据的规格要求等。最后，如果存在对系统及其性能的预期修改，提供者应当在首次合格评定后便告知使用者。

四、生成式人工智能使用的特别注意义务来源

对使用者的注意义务规范还应当考虑生成式人工智能的特别应用场景。按照《中华人民共和国消费者权益保护法》（以下简称消保法）、个保法的相关规定，在商业服务、个人信息处理活动中，使用者应当向消费者、个人信息主体主动披露是否以及如何使用生成式人工智能的基本情况。问题在于，当使用行为可能对消费者、个人信息主体的权益产生重大影响时，使用者应当在何种程度上履行上述信息提供义务。与此同时，为防止具有重大权益影响性的使用行为造成侵权损害，还需讨论使用者进

① 高阳：《通用人工智能提供者内容审查注意义务的证成》，《东方法学》2024年第1期。

② 王泽鉴：《侵权行为》，北京：北京大学出版社，2016年，第299页。

③ 刘文杰：《论侵权法上过失认定中的“可预见性”》，《环球法律评论》2013年第3期。

行风险防控的注意义务类型与内容。

(一)生成式人工智能使用者双重身份下的二阶信息提供义务

生成式人工智能也可以通过智能聊天机器人等人机交互方式向公众提供内容生成服务,每一次信息输入、个性化回复及聊天记录存储等都有可能涉及个人信息处理。与此同时,其也可以被用来分析消费者行为或进行用户信用评分,并通过自动化决策提供个性化推荐、排序精选、排序过滤、调度决策等服务。因此,当使用者借助生成式人工智能提供服务时,其可能同时具备双重身份。从消费者角度来看,使用者如果长期、持续性地利用深度合成技术从事商业活动并以盈利为目的,可以构成消保法上的经营者。从个人信息主体角度来看,使用者如果使用并分析了服务相对人的个人信息,则应被视为个保法上的信息处理者。基于上述特殊身份,使用者不仅承担一般注意义务,还应根据消保法与个保法承担特别的信息提供义务。

通常情况下,消费者享有知悉服务真实情况的知情权以及公平交易权(消保法第8条与第10条)。这就要求使用者应当在初次交互时,便以通知或显著标识等方式向消费者披露生成式人工智能的辅助使用,避免其对真实情况产生混淆或误认。与此同时,消费者有权自由选择是否接受该服务。经营者应当对消费者提出的关于服务质量和使用方法等内容的询问,给出真实、明确的答复(消保法第19条),例如生成式人工智能在内容生成服务中的贡献度、系统可能出现的输出错误或质量偏差等。同样的情况下,在个人信息处理活动中,使用者应当遵循公平、透明原则,并向个人信息主体披露信息处理的目的、方式和范围(个保法第7条)。相应地,个人也享有对信息处理情况的知情权与决定权(个保法第44条)。上述规则表明,使用者在处理服务相对人的个人信息时,应当向其披露训练数据的来源、类型及处理方式,以及所使用人工智能的模型结构、决策过程等特别信息。因为在诸如文生图片或文生视频等内容生成服务中,为提高输出结果的准确性,了解模型的注意力机制、正确使用提示词、调整训练数据,对服务相对人而言至关重要。由此可见,在使用者具备经营者或个人信息处理者身份时,需要满足一定的信息提供要求,其本质上也是上游提供者义务在使用阶段的延续和补充①。需要注意的是,上述义务范围及履行程度在涉及服务相对人的合法利益保护时,并不足够。

为防止因自动化偏见、机器幻觉导致生成式人工智能决策错误而损害服务相对人的合法权益,使用者应当承担更进一步的信息提供义务。消保法第18条要求经营者在提供可能危及人身、财产安全的服务时,向消费者作出真实说明并明确警示,例如对系统能力和局限性的解释说明等。个保法第24条则要求信息处理者在自动化决策方式对个人权益产生重大影响时,依个人要求作出必要的说明,例如披露其排名与信息推送作出时的参数权重、计算公式以及预期结果等②。显然,上述信息内容已经超出了一般的信息提供范围。学理上,将个人向信息处理者提出异议,并要求其对决策过程进行解释的权利,称之为"算法解释权"③。这一基于个性化决策场景以及重大权益影响性产生的特别权利,建立在个人与处理者的沟通信任基础上④。此时,使用者对相对人的义务内容不仅是披露,还应当落脚于解释。对此,有关要求使用者完全公开复杂冗长且晦涩难懂的算法代码的观点⑤,并无现实意义。这不仅会导致披露行为与商业秘密保护需求之间的冲突⑥,也不利于实现信息处理活动的"实质透明"⑦

① 郑志峰:《人工智能使用者的立法定位及其三维规制》,《行政法学研究》2025年第1期。

② 衣俊霖:《数字孪生时代的法律与问责——通过技术标准透视算法黑箱》,《东方法学》2021年第4期。

③ 林洹民:《自动决策算法的风险识别与区分规制》,《比较法研究》2022年第2期。

④ 丁晓东:《基于信任的自动化决策:算法解释权的原理反思与制度重构》,《中国法学》2022年第1期。

⑤ Frank P., *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge: Harvard University Press, 2015, pp. 142-143.

⑥ 袁康:《可信算法的法律规制》,《东方法学》2021年第3期。

⑦ 张永忠:《论人工智能透明度原则的法治化实现》,《政法论丛》2024年第2期。

或“析理性透明”[①]。正如欧盟《人工智能法案》在其立法考量部分第27项对“透明度”的定义，人工智能系统的开发与使用应具备一定的可追溯性（traceability）和可解释性（explainability）。可解释性本质上是指人工智能系统运行具备对一般人而言可被理解的解释能力[②]。因此，使用者应当以可理解性为目标，向服务相对人提供通俗易懂的信息。然而，为合理考虑使用者履行义务的成本，不应苛求使用者普适化、一般化地履行此义务。此时，问题便转化为，生成式人工智能使用行为何时满足重大权益影响性这一前提条件。

“高度使用风险”可以成为进一步确认解释说明义务的判断标准。我国学者已提出了基于系统风险等级划分对透明度义务进行多层设计的观点[③]。其中人工智能系统涉及教育和职业培训资格、就业和员工管理及社会福利待遇等权益的，都可以被视为对个人权益产生重大影响的元素[④]。因此，诸如将生成式人工智能用于员工远程识别并构建画像的使用行为，应被视为具有高度使用风险，使用者应当由此对其承担解释说明义务。

综上所述，根据使用者所提供的服务类型和由此引发的风险等级不同，使用者的信息提供义务可被分为两个层次。第一层的信息披露旨在满足一般的透明度要求，实现相对人对信息的可获取性，为后续的解释说明创造条件；第二层的解释说明性信息发生在使用行为具有高度使用风险的情形下，侧重于实现相对人对已披露信息的有效择取与正确理解。值得关注的是，《互联网信息服务算法推荐管理规定》第12条、第16—17条以及第21条，分别提及了算法推荐服务提供者对决策规则透明度和可解释性的优化义务、对重大权益影响性的说明义务以及保护消费者公平交易权的义务。按照该规定第2条之定义，算法推荐服务提供者是指利用包括生成合成类等算法技术提供互联网信息服务的组织或个人。虽然该规定尤指生成式人工智能服务提供者，但或许可以通过扩张解释将使用者纳入其中。据此，该规定或可成为使用者承担信息提供义务的特别法基础。

（二）生成式人工智能使用者于高风险使用时的注意义务补充

生成式人工智能的高度使用风险应当作为重点监管内容，仅采取提高信息提供义务标准的解决方案并不足够。事实上，能够引发高度使用风险的使用者通常是大型企业、司法机关、医疗机构等专业性组织[⑤]，区别于普通用户，其通常具备良好的技术条件和经济能力。对此，可以从以下几个方面增加对专业使用者的义务要求，这样不仅能够促进行业自律，也有助于减少高度使用风险带来的负面影响。首先，在风险管理与预防方面，使用者必须严格按照系统使用说明操作系统，同时安排能力足够、培训到位且权限适配的人员进行监督。履行过程中，使用者不得违反其他法律义务，也不能限制自身组织资源和活动的自由。其次，在控制输入数据质量方面，当使用者对输入数据有控制权时，要保证数据与系统预期目的紧密相关且具有充分代表性，为系统准确运行和决策提供可靠支持。再次，在更新和维护人工智能系统时，使用者需依据使用说明持续监测系统运行状态，一旦发现按说明运行仍有风险，需立即通知提供者、分销商以及相关市场监督机构并暂停使用。最后，在留存日志方面，使用者在自身控制范围内要将系统自动生成的日志保留至少六个月。

需要补充的是，比较法上，欧盟《人工智能法案》亦对生成式人工智能的应用治理有所回应，但相关规则的合理性不无疑问。一方面，欧盟立法者引入通用目的人工智能模型治理专章（《人工智能法案》第五章）正是为了回应大型语言模型广泛应用带来的紧急立法需要[⑥]。依据欧盟《人工智能法案》

① 苏宇：《优化算法可解释性及透明度义务之诠释与展开》，《法律科学》2022年第1期。

② 刘文杰：《何以透明，以何透明：人工智能法透明度规则之构建》，《比较法研究》2024年第2期。

③ 张永忠：《论人工智能透明度原则的法治化实现》，《政法论丛》2024年第2期。

④ 张丽英、段佳葆：《自动化决策下的个人信息保护——以〈个人信息保护法〉第24条为中心》，《东岳论丛》2023年第10期。

⑤ 郑志峰：《人工智能使用者的立法定位及其三维规制》，《行政法学研究》2025年第1期。

⑥ Martini/Wendehorst, in: Kommentar KI-VO, 2024, § 51 Rn. 9.

第3条第63项对“通用目的人工智能模型”的定义,即“通用人工智能模型是指能够通过使用大量数据进行大规模自我监督训练,并在投入市场后能够胜任不同任务,具有显著的通用性的人工智能模型”,通用目的人工智能模型的用途实则极为广泛,生成式人工智能模型通常可以被纳入其中①。但是,不同应用场景下的模型使用风险明显存在巨大差异,立法者无法全面规范所有可能的风险情形,因此该部分规则在具体应用场景中的可适用性有待检验。另一方面,该法案第51条以浮点算力阈值量化了具有系统性风险的通用目的人工智能模型的标准,即模型用于训练的计算数量超过10^{25}时,推定其具备高影响能力。然而,这一标准的可适用性已显僵化。不仅超出该标准的模型目前仅有GPT-4和PaLM2两种②,而且随着DeepSeek的横空出世,知识蒸馏(knowledge distillation)即模型压缩技术的应用在保证较高性能的同时,也愈发减少模型的计算资源和存储需求③。

综上所述,欧盟立法经验表明,在科技迅速发展面前,法律政策往往具有滞后性。对于我国的生成式人工智能治理而言,应当尽量避免上述宽泛且僵化的立法方式,相反,应结合具体应用场景,进行差异化的规则设计。特别是将生成式人工智能应用于化学、生物、军事等高风险领域,或其他容易导致严重事故、威胁公共健康安全的场景时,应当对专业使用者施以更高标准的注意义务。

五、结语

生成式人工智能技术强大的内容生成能力既为创新生产注入活力,也催生了新型法律风险。在此背景下,从技术开发到实际应用的全生命周期治理成为一项重要命题。需要认识到,既往以上游服务提供者责任为核心的治理模式已经难以应对因下游使用者滥用而形成的现实挑战。过度加重提供者的义务负担,将会抑制技术创新。而将监管重心适当向使用者倾斜,不仅能够缓解上游压力,也有助于从应用终端遏制风险扩散,以更好地实现科技发展与风险治理的动态平衡。鉴于我国人工智能立法尚处于探索阶段,应当以民法典等现行法律为解释基础,合理构建包含使用者禁止使用义务、内容标识义务、结果审查义务、信息提供义务、高风险使用者特别义务在内的注意义务体系,以期对使用者实施必要的法律监管,并满足司法实践对规范适用的紧迫需要。

The Normative Sources of the Duty of Care for Users of Generative Artificial Intelligence

Liu Chengwei　Ma Ruicong
(College of Comparative Law, China University of Political Science and Law, Beijing 100088, P.R.China)

Abstract: The full-life cycle regulation of generative artificial intelligence (GenAI) represents a governance model adopted both domestically and internationally. Within the entire chain of GenAI, from GenAI's development and training to its application, risk governance in the application phase remains an indispensable component. At present, China's legal framework for GenAI is drafted by

① 陈亮、张翔:《欧盟生成式人工智能立法实践及镜鉴》,《法治研究》2024年第6期。

② 《谷歌PaLM 2训练所用文本数据量是初代的近5倍》,https://www.163.com/tech/article/I4UN0BU600097U7T.html,访问日期:2025年2月15日。

③ 邓建鹏、赵治松:《DeepSeek的破局与变局:论生成式人工智能的监管方向》,《新疆师范大学学报(哲学社会科学版)》2025年第4期。

low-level legislative authorities and the legislative content is fragmented. The framework's key feature is that several rules and regulations are centered on technical risk mitigation and supervision of providers. However, it lacks direct provisions governing users. Consequently, when users' improper operation of GenAI infringes upon others' legitimate interests, their conduct can only be evaluated based on the general tort liability rules stipulated in Article 1165 of the Civil Code of the People's Republic of China. Nevertheless, due to the complex, opaque, and random nature of GenAI, determining the standard of care users should bear to avoid causing harm poses significant challenges.

To reasonably define users' duty of care, it is necessary to adopt a contextualized perspective while considering the differences in economic capacity and technical conditions among different users. In principle, users are entitled to freely use GenAI, except where such use is prohibited. The tort risks that may arise from users' use of GenAI can be categorized into three tiers: prohibited risk, high risk, and general risk. Specifically, using GenAI to create deepfakes for intentional manipulation or deception, which significantly infringes upon others' personal and property rights, should be universally prohibited. Additionally, in service contracts, the conclusion of the contract is entirely based on the creditor's trust in the debtor's performance capability, thereby imposing an obligation on the debtor to perform personally. Therefore, whether the debtor resorts to GenAI to fulfill contractual obligations constitutes a material fact regarding the basis of contract performance. Without the creditor's express permission, the debtor shall not use GenAI to completely replace the contractual performance that should be undertaken.

Where the use of GenAI is permitted, the issue of reasonable usage can be explained by the theory of social intercourse safety obligation. Specifically, users shall bear the obligation to protect the legal interests of others. When AI-generated content constitutes deepfakes or is about to be uploaded to online platforms, users generally bear an obligation to provide conspicuous labeling to avoid causing unnecessary confusion and misinformation to the public. Meanwhile, to prevent the spread of harmful content from the source, users should fulfill a duty of necessary result review over the output content. However, the scope of obligations borne by users shall be limited to obviously illegal, discriminatory, or false content that users can reasonably foresee, or be confined within the scope notified by the provider. Finally, professional users may also trigger high-risk usage that exerts significant impacts on the legitimate interests of others. As professional users generally possess stronger risk prevention capabilities than ordinary users, they shall assume additional obligations such as risk management and prevention, input data quality control, and system update and maintenance.

Keywords: Generative artificial intelligence; Deepfake; Obligation of users; Duty of care; DeepSeek

[责任编辑:岳　敏　苏　捷]