

# 生成式人工智能生成内容标识义务研究

张继红\*

**摘要:**生成式人工智能的使用者为了获得其生成内容完整的著作权,往往利用生成式人工智能生成内容外观与人类作品外观在辨识上的困难,将其伪装成人类作品以掩盖生成式人工智能的实质性参与。这一行为严重损害了社会公众和自然人创作者的财产性利益和精神性利益,破坏了创作物市场的稳定秩序和信任关系,法律应当加以规制。民法上的诚实信用原则与经济法上的产品信息披露义务分别为生成式人工智能生成内容的标识义务提供了“自律”与“他律”的法律依据,透明化的创作物市场所蕴含的“共同善”理念则促使个人利益服务于社会共善。保证生成式人工智能相关主体履行标识义务的途径包括:在生成阶段,对生成式视觉艺术产品采取标识模式,在设计算法时保证生成内容的可检测性;在传播阶段,应引导相关主体自主披露信息,并建立信用体系以保障义务的履行。

**关键词:**生成式人工智能 标识义务 诚实信用原则 共同善

在生成式人工智能诞生之前,人们默认智力成果出自人类之手,该智力成果若具备独创性则作为作品受著作权法保护。随着以美国开放式人工智能(Open AI)公司开发的以聊天生成型预训练变换模型(ChatGPT)为代表的生成式人工智能技术日趋成熟,自然人与生成式人工智能协同造物已经进入互联网,其文本、音频、代码、图像和视频等内容的强大创建能力,结合高效且廉价的产出模式,已经对现有文字类和艺术类工作产生了颠覆性影响。在弱人工智能逐渐向强人工智能过渡的背景下,生成式人工智能生成内容的外观已难以与人类作品的外观相区分。然而,为确定其权利归属,需要区分人类作品与生成式人工智能作品,毕竟二者的独创性来源存在本质差异,前者产生于人的智力创造,后者则主要依靠算法和程序。事实上,以大语言模型和图像生成模型为代表的生成式人工智能可以独立生成全新内容,该生成内容并非生成式人工智能使用者(以下简称人工智能使用者)所创作的作品,后者理应披露生成式人工智能参与创作的过

\* 上海政法学院教授、上海全球安全治理研究院研究员  
基金项目:上海政法学院“问渠源学者”计划资助项目

程。即使是在偏好投资人利益保护和生成式人工智能技术最为发达的美国,对于生成式人工智能生成内容也持否定态度,认为其并非人类创造性产物,不受版权保护,应当与人类作品相区分。<sup>①</sup>拜登政府于2023年10月30日颁布行政命令,要求美国商务部制定规则要求人工智能生成内容都应加入水印,以创建更加安全可靠数字生态系统。<sup>②</sup>

在我国,生成式人工智能生成内容与人类作品如何区分的问题业已显现,在北京互联网法院公开审理的“人工智能生成图片著作权侵权案”<sup>③</sup>中,争议焦点就在于生成式人工智能生成的图片是否受著作权法保护以及未经许可去掉图片署名水印冒充人类作者并传播是否构成侵权行为。2023年8月15日生效的《生成式人工智能服务管理暂行办法》(以下简称《暂行办法》)第12条规定:“提供者应当按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识。”然而,《互联网信息服务深度合成管理规定》(以下简称《管理规定》)第17条仅规定深度合成技术在提供可能导致公众混淆或误认的与自然人人身属性相关的编辑服务时才需要进行显著标识。<sup>④</sup>也就是说,现有规范性文件虽对标识义务作出规定,但并未释明其背后的法律内涵,且《管理规定》中的深度合成技术并不能涵盖《暂行办法》所规定的生成式人工智能技术,单一的“标识”义务亦无法适应复杂的生成式人工智能使用环境,不能给予执法及司法活动明确的指引。现有文献要么侧重于讨论深度合成技术的风险与法律规制,<sup>⑤</sup>但深度合成技术与生成式人工智能技术存在技术和应用场景的区别;要么集中讨论生成式人工智能所涉及的数据与算法风险监管问题,<sup>⑥</sup>尚无针对生成式人工智能生成内容标识义务的深入分析。在我国生成式人工智能产品陆续推出的背景下,相关主体对生成式人工智能生成内容承担标识义务的法理依据和法律内涵是什么?如何结合场景进行具体划分和法律适用?以上问题都需要做进一步的深入研讨。

<sup>①</sup> 美国版权局在最新发布的《版权登记指南》中明确指出,包含人工智能生成内容的作品可以登记版权,但只保护人类作者贡献的部分;申请人必须在申请表上披露作品中包含的任何人工智能生成内容,并提供相关信息。See Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, <https://public-inspection.federalregister.gov/2023-05321.pdf>, 2023-07-25.

<sup>②</sup> See White House: Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>, 2023-11-05.

<sup>③</sup> 参见北京互联网法院(2023)京0491民初11279号民事判决书。

<sup>④</sup> 《互联网信息服务深度合成管理规定》第17条规定:“深度合成服务提供者提供以下深度合成服务,可能导致公众混淆或者误认的,应当在生成或者编辑的信息内容的合理位置、区域进行显著标识,向公众提示深度合成情况:(1)智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务;(2)合成人声、仿声等语音生成或者显著改变个人身份特征的编辑服务;(3)人脸生成、人脸替换、人脸操控、姿态操控等人物图像、视频生成或者显著改变个人身份特征的编辑服务;(4)沉浸式拟真场景等生成或者编辑服务;(5)其他具有生成或者显著改变信息内容功能的服务。深度合成服务提供者提供前款规定之外的深度合成服务的,应当提供显著标识功能,并提示深度合成服务使用者可以进行显著标识。”

<sup>⑤</sup> 参见万志前、陈晨:《深度合成技术应用的法律风险与协同规制》,《科技与法律》2021年第5期。

<sup>⑥</sup> 参见程乐:《生成式人工智能的法律规制——以ChatGPT为视角》,《政法论丛》2023年第4期;毕文轩:《生成式人工智能的风险规制困境及其化解:以ChatGPT的规制为视角》,《比较法研究》2023年第3期。

## 一、风险产生：生成式人工智能生成内容对人类创作秩序的负面影响

当外观类似人类作品的生成式人工智能生成内容流入作品市场,涉及人工智能使用者、生成式人工智能服务提供者(以下简称人工智能提供者)和公众3方主体。<sup>①</sup>由于生成式人工智能区别于人类主体,法律尚未明确生成式人工智能生成内容的性质,且其输出内容难以控制,若不对相关主体施加标识义务会产生多种风险,将冲击已有的人类创作秩序。

### (一)创作物财产性利益丧失的风险

人工智能使用者在生成式人工智能的人机交互过程中扮演重要角色,其基于自主意识与人工智能程序实施的交互过程行为,与生成式人工智能生成内容的创造存在直接因果关系。若将生成式人工智能视为纯粹的工具,则生成式人工智能使用者更接近于“作品”的“作者”,作者当然有权署名自己作品,且无须标识使用了何种辅助创作的工具。然而,目前生成式人工智能的使用大多已经超越了“辅助工具”的范畴,而是由生成式人工智能直接主导“创作”,此时著作权法能否涵盖该客体存在争议。我国现有立法对于生成式人工智能的生成内容是否构成“作品”暂无定论,学界对此亦形成截然相反的两种观点。<sup>②</sup>美国版权局对生成式人工智能生成内容则采取消极的保护态度,认为版权法不保护非人类意志主导的表达。正如2023年2月美国版权局撤销女艺术家卡什塔诺娃使用人工智能绘画创作的漫画风格插图《黎明的扎里亚》的版权认证声明(即“《黎明的扎里亚》登记案”)所指出的那样,“虽然额外的指令可以影响图形,但该过程不受用户控制,因为其不能预测 Midjourney(一种图像生成模型)将生成什么内容。从版权保护的角度来看,相较于艺术家使用的其他工具, Midjourney 有明显不同: Midjourney 用户对于生成的图片缺乏绝对控制,不能认为是这些图片的主脑”。<sup>③</sup>也就是说,人工智能使用者在创建内容的过程中,其本人意志对生成内容“思想”的表达没有决定性作用。

有别于物权人对有体物“占有”而存在的天然公示效果之外观,知识产权人对无体物的权利完全来源于法律的创设。《中华人民共和国著作权法》(以下简称《著作权法》)第12条规定:“在作品上署名的自然人、法人或者非法人组织为作者,且该作品上存在相应权利,但有相反证明的除外。”因此,他人可以通过署名推定作品的作者、权利存在及其权利归属。作者于作品上署名发表与物权法中的“物权公示”相类似,是作品市场意义上的“公示”。<sup>④</sup>公众亦难以通过外观识别

<sup>①</sup> 根据《生成式人工智能服务管理暂行办法》第22条的规定,生成式人工智能服务提供者,是指利用生成式人工智能技术提供生成式人工智能服务(包括通过提供可编程接口等方式提供生成式人工智能服务)的组织、个人;生成式人工智能服务使用者,是指使用生成式人工智能服务生成内容的组织、个人。本文所称“公众”,是指受生成式人工智能技术应用影响的广大民众,涵盖生成式人工智能生成内容后续使用者(排除人工智能使用者)、自然人创作者等。

<sup>②</sup> 有学者认为生成式人工智能生成内容具有可版权性。参见熊琦:《人工智能生成内容的著作权认定》,《知识产权》2017年第3期;吴汉东:《人工智能生成作品的著作权法之问》,《中外法学》2020年第3期。还有学者认为生成式人工智能生成内容不应受著作权法保护。参见刘银良:《论人工智能作品的著作权法地位》,《政治与法律》2020年第3期;王迁:《三论人工智能生成的内容在著作权法中的定位》,《法商研究》2024年第3期。

<sup>③</sup> See United States Copyright Office's Reply of Zarya of the Dawn on February 21, 2023, <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>, 2023-05-05.

<sup>④</sup> 参见柳励和:《论作品署名权的转让》,《湖南科技大学学报》(社会科学版)2009年第3期。

“人类作品”或“生成内容”。人工智能使用者只要使用人类署名,在保证该生成内容在外观上难以和人类作品相区分的前提下,即可将生成内容“伪造”成作品以获得著作权法保护。<sup>①</sup>此时,人工智能使用者有充分的理由拒绝披露人工智能参与创作的过程,以维护其对于该创造物的财产利益。

然而,这种掩盖生成式人工智能创作过程的行为实质性破坏了著作权法中署名的“公示”功能,在当前针对生成式人工智能内容法律态度尚不明朗的情况下极易减损作品流转和传播的运作效率,影响了市场秩序的可预期性。例如,若A使用生成式人工智能生成内容后以自己的名义发布,B按照“作品”许可方式支付费用并传播该“作品”,当B事后得知该“作品”系生成式人工智能生成内容时,可否请求A返还“不当”得利?客观来讲,若将生成式人工智能生成内容从“作品”行列剔除,则该受“欺诈”之被许可人的所有著作权也一并消灭,其所创造的衍生产品就成为“无本之木”,在不受著作权法保护的情况下存在落入公共领域的风险,严重损害了之前依赖该“作品”版权所产生的一系列行为的效力以及交易秩序的稳定。与此同时,著作权法的制度设计在于激励投资人进行投资,<sup>②</sup>当著作权法所设置的作者署名体系遭到这种“不诚信”行为的破坏时,投资人无法信任作者作品的真实性,也就实质性抑制了投资人对作品市场的投资意愿,著作权法的经济激励目的亦难以实现。

## (二)精神激励错位的风险

尽管经济激励机制是著作权制度运行的主要逻辑,但由作者对作品署名行为而产生的精神激励也发挥着鼓励作者创作的重要功效。在著作权法尚未诞生的中国古代,创作者以追求文章之美誉、能够“立言”作为其创作动力。<sup>③</sup>人工智能使用者在参与生成的过程中产生了一系列互动行为,如运用了精心策划的提示词,而对互动和内容的主观控制会增加作者感。相较于雇佣人类代笔者行为中的“剥夺”署名人身权,掩盖无人格的机器人的创作行为似乎更加合乎情理。也就是说,即使人工智能使用者对生成内容具有财产性利益,其依然会为了追求精神性利益而实施伪造行为。

事实上,这种将生成式人工智能生成内容伪装为人类作品的行为是对“作者—作品”固定关联秩序的一种破坏。当人类将生成式人工智能所生成内容以其名义发表,读者基于生成内容给予该虚假作者的社会评价也将发生错位。掩盖生成式人工智能创作过程的伪装者将生成式人工智能作为代笔人,获得了与其实际才能不相符的声誉与激励。其可能会在创作市场中获得更多机会,使得真正的创作者无法得到应有的肯定和回报,这将严重挫伤原创作者的创作热情,最终产生“劣币驱逐良币”的消极后果。这一现象在教育领域尤甚。生成式人工智能代写的泛滥严重影响了教育评估机制的公平性。在美国,已经有大量学生使用生成式人工智能代写作业以欺骗教师。这种故意掩盖生成式人工智能生成内容的行为,极易从根源上抑制创造性思维发展,进而

<sup>①</sup> 《伯尔尼公约》第5条明确了版权的自动保护原则。根据公约规定,享受和行使版权无须履行任何手续,也不管作品起源国是否存在有关保护手续的规定,即作品自动受到版权法保护。

<sup>②</sup> 参见熊琦:《著作权激励机制的法律构造》,中国人民大学出版社2011年版,第13页。

<sup>③</sup> 参见冯晓青:《著作权法之激励理论研究——以经济学、社会福利理论与后现代主义为视角》,《法律科学》(西北政法学院学报)2006年第6期。

破坏现代教育的根本逻辑。<sup>①</sup>正如美国约翰·尼姆教授所言：“仅仅因为有一台机器可以帮助我举起哑铃，并不意味着我的肌肉就会发达。”<sup>②</sup>

应该说，人类的代写行为一直存续于创作物市场，且部分学者认为当事人通过意思自治转让署名权无可厚非。美国学者波斯纳也认为如画家伦勃朗与政客希拉里都存在虚假署名他人作品的行为，但由于不存在受害者而不构成真正意义上的“剽窃”。<sup>③</sup>然而，掩盖生成式人工智能生成内容的伪装行为由于其特殊属性对于创作者人格评价体系的冲击显然更大，具体表现如下：（1）高广度性。相较于雇佣人类代笔者进行创作，使用生成式人工智能工具成本更低，该伪装行为更具有普遍性。（2）长时效性。鉴于人类代笔者会对其创作作品的精神利益产生期待，难以长时间内实施“捉刀代笔”之事，因此寻找一个长期且稳定地代笔者并非易事。相较而言，生成式人工智能则能长期、稳定的生成大量创作物，因而其伪装效应期间更长，所引发的大规模生成式人工智能代写效应将使创作成为一项“机械”工作。与人类代笔者相比较，生成式人工智能代写无法体验创作的过程、思考和灵感，仅仅是执行预设的算法和程序，这可能导致创作的机械化、空心化以及失去灵感的风险增加。换言之，如果生成式人工智能生成内容无法与人类作品区分开来，原本具有创造力和自主性的人类作者也会更倾向于使用相关程序，其宝贵的独立品格亦会被算法和程序潜移默化地束缚于特定风格或主题，最终引发人类创作行为的式微乃至湮没。

### （三）公众信任机制遭破坏的风险

无论是传统意义上的作品还是生成式人工智能生成内容都会流入公众视野，相关权利的许可使用和公共传播过程都与公共利益关系密切。<sup>④</sup>有学者抽象出“公众—使用权”这一具有救济权利性质的可诉性权利来凸显公众作为作品使用者的重要地位。<sup>⑤</sup>传统作品由自然人完全参与，其内容的真实性由自然人的信用背书，公众可基于作者的信誉相信该作品的客观性。而生成式人工智能实现的方式是算法和程序。“算法黑箱”现象表明机器深度学习时，其复杂的神经网络中存在不为人所直观捕捉到的隐层，因此在生成式人工智能生成内容的过程中，即使存在人工智能使用者的人为干预，其生成内容依然是不可控和不确定的。生成式人工智能还能够定制性地针对个体创建虚假信息，以输出多维度虚假信息和心理诱导的方式对单个或者大规模人群造成算法操控。<sup>⑥</sup>如果人工智能使用者不披露实际创作情况，传播了生成式人工智能生成的不真实内容，更易导致公众接触错误信息，产生负面舆情甚至扰乱公共秩序。例如，有人利用ChatGPT编造“杭州3月1日取消限行”的假新闻，其措辞、结构、语态与一般的新闻通告十分相

<sup>①</sup> 在线课程供应商 Study.com 对 1 000 多名 18 岁以上的学生进行了调查，其中有超过 89% 的学生使用 ChatGPT 来帮助完成家庭作业。See Productive Teaching Tool or Innovative Cheating?, <https://study.com/resources/perceptions-of-chatgpt-in-schools>, 2023-05-23.

<sup>②</sup> Douglas Belkin, Professors Turn to ChatGPT to Teach Students a Lesson, *The Wall Street Journal*, January 25, 2023.

<sup>③</sup> 参见[美]理查德·波斯纳：《论剽窃》，沈明译，北京大学出版社 2010 版，第 27~29 页。

<sup>④</sup> 参见杨勇：《著作权法中损害公共利益的认定研究》，《中国版权》2016 年第 5 期。

<sup>⑤</sup> 参见刘银良：《著作权法中的公众使用权》，《中国社会科学》2020 年第 10 期。

<sup>⑥</sup> 参见张欣：《生成式人工智能的算法治理挑战与治理型监管》，《现代法学》2023 年第 3 期。

似,大大增加了公众受骗的风险。<sup>①</sup>此外,生成式人工智能生成内容也存在著作权风险,即生成式人工智能在机器学习过程中与原作品“接触”后,若其生成内容与原作品存在“实质性相似”,则可能构成侵权行为。在被媒体称为“全球生成式人工智能平台侵权第一案”的“上海新创华文化发展有限公司诉 AI 公司(化名)案”<sup>②</sup>中,被告经营的网站生成式人工智能绘画能够生成与原告案涉“奥特曼”形象构成实质性相似的图片,广州互联网法院认定其侵犯了原告对作品所享有的复制权和改编权。而且,倘若人工智能使用者伪造自然人对作品独创性的真实背书,当原作者基于著作权人身份追究责任时,人工智能使用者会选择将该创作物“显名”,由此将侵权责任转移至人工智能提供者。是故,人工智能使用者的伪装行为使得该创作物的真实属性处于不稳定状态,不仅人工智能提供者的权益会遭受损失,生成式人工智能生成内容本身的价值也难以体现。

信任既是社会交往与理性发展的结果,反过来又可以促进社会交往与理性发展。<sup>③</sup>在创作物与公众的信息利用关系中,原本存在着一种“基于信任的公地”。公地作为一项资源或财产有许多拥有者,他们中的每一个人都有使用权,但没有权利阻止其他人使用,从而造成资源过度使用和枯竭。如果生成式人工智能生成内容的伪装行为屡屡发生且难以被控制,更多无法鉴别真实性和权利稳定性的生成式人工智能生成内容就会源源不断地流入创作物市场,这不仅会破坏自然人真实背书秩序,还会严重减损公众对创作物合法性的信任,其对信息的选择与利用也会趋于谨慎和保守,最终引发创作市场的“信任公地悲剧”。这显然与生成式人工智能的发展初衷相悖离。

## 二、标识义务证成:创作物市场的秩序维护

人工智能时代的创作物市场中包含人类作品与人工智能生成内容。规制生成式人工智能生成内容伪装行为的核心在于对实际控制者(包括人工智能使用者和人工智能提供者)施加标识义务,以维护生成式人工智能“复合式”创作物的市场秩序。

### (一)诚实信用原则的适用

诚实信用起源于罗马法中的拉丁语“bona fides”,根据词源学,其有“善信”之义,我国一般翻译为“诚信”。<sup>④</sup>在民事领域,立法者不仅应鼓励当事人的自主性,还要使民事活动受到一定的合理限制。诚实信用原则作为民事领域的一项基本准则,其核心要义是“在法律上的特别关联中,使行为人须顾及相关人利益”。<sup>⑤</sup>生成式人工智能生成内容的创建与传播涉及多方主体,《中华人民共和国民法典》(以下简称《民法典》)第7条概括性的诚实信用原则条款旨在修正各方均以自身利益为中心的行为方式。根据《中华人民共和国反不正当竞争法》第2条和《暂行办法》第4条的规定,经营者在提供人工智能服务时应当遵守法律和商业道德,并适用诚实信用原则。

诚实信用原则要求民事主体在民事活动中维持双方利益平衡以及当事人利益与社会利益平

<sup>①</sup> 参见《紧急提醒!杭州取消限行“新闻稿”传疯,是 ChatGPT 写的!警方已介入调查!》, [https://www.thepaper.cn/newsDetail\\_forward\\_21961621](https://www.thepaper.cn/newsDetail_forward_21961621), 2023-02-17。

<sup>②</sup> 参见广州互联网法院(2024)粤 0192 民初 113 号民事判决书。

<sup>③</sup> 参见安晋城:《算法透明层次论》,《法学研究》2023 年第 2 期。

<sup>④</sup> 参见徐国栋:《客观诚信与主观诚信的对立统一问题——以罗马法为中心》,《中国社会科学》2001 年第 6 期。

<sup>⑤</sup> 于飞:《基本原则与概括条款的区分:我国诚实信用与公序良俗的解释论构造》,《中国法学》2021 年第 4 期。

衡,其本质是对社会公共利益和交易道德的强调。<sup>①</sup>如前所述,人工智能使用者会为了追求财产性利益和精神性利益而实施掩盖生成式人工智能创作过程的伪装行为。有学者利用博弈矩阵得出结论,只有当生成式人工智能生成内容特殊保护的力度大于著作权保护,即人工智能实际控制者坦白的收益高于隐瞒的收益时,他们才会更倾向于坦白,然而这几乎是不可能实现的。<sup>②</sup>短期来看,在人工智能使用者采取不诚信行为与人工智能提供者不作为的情况下,二者作为生成式人工智能生成内容实际控制者的收益为最优。但在社会公众与生成式人工智能实际控制者的长期博弈中,当公众对生成内容的交易(接触、使用)频率增加,则生成式人工智能实际控制者伪装行为被发现的概率也会增加。公众采取不付费或怀疑作者本体的方式以应对不诚信行为,人工智能使用者的财产性利益和精神性利益都会相应减损,从而导致创作物供给减少,最终引发“双输”结果。由是观之,在创作者产出内容与公众的单次或短期交易中的真实标识行为表面上看是利他的,但在长期交易活动中的诚信行为不仅是利己的更是互利的。与传统民事交易活动相类似,诚实信用原则可以在信息不对称的条件下规范双方当事人行为,以消除当事人对假定不真实性状况的疑虑,从而降低交易成本。<sup>③</sup>在人工智能领域适用诚实信用原则能够引导各方利益主体理性约束自身“虚假”行为,以避免由于追求短期效益而破坏创作物市场整体交易秩序的稳定。

人类有利己的自然属性,信任本质上是一个风险行为。<sup>④</sup>人工智能提供者对于掩盖生成式人工智能创作过程的行为的默许,固然可以在短期内获得人工智能使用者的青睐而使双方获得利益,但上述“不诚信”行为对于创作物整体市场秩序的破坏显然不利于促进公众对人工智能技术和产品的信任。人工智能使用者以自己名义向公众输出真假难辨的信息,进而影响了公众对于新闻媒体与政府组织的信任,毕竟公众的信任来源于过程的公开与透明。正如有学者对算法黑箱所提出的“算法透明”主张,<sup>⑤</sup>对于生成式人工智能创作过程的“黑箱”亦应被打破。相较于披露包括源代码、输入数据、输出结果在内的算法要素而言,对人工智能使用者或服务提供者施加标识义务,显然更有可行性。生成式人工智能生成内容的标识义务正是为了基于诚实信用原则以超越人类自利的本性,维持创作物市场基于信任的高效运转,从而促进人工智能产业的良性循环。

## (二) 产品信息披露义务的扩张

欧洲议会早在2020年就建议对人工智能生产者的产品责任进行规制。<sup>⑥</sup>从产品经过加工、制作并流通的属性来看,人工智能属于产品的范畴并受产品制度的调整。<sup>⑦</sup>虽然生成式人工智

<sup>①</sup> 参见徐国栋:《民法基本原则解释:诚信原则的历史、实务、法理研究》,北京大学出版社2013版,第32页。

<sup>②</sup> 参见吴雨辉:《人工智能创造物著作权保护:问题、争议及其未来可能》,《现代出版》2020年第6期。

<sup>③</sup> 参见董灵:《论合同法诚信原则的经济学基础》,《广东社会科学》2006年第5期。

<sup>④</sup> 参见张善根:《法律信任论》,中国法制出版社2018版,第39页。

<sup>⑤</sup> 参见沈伟伟:《算法透明原则的迷思——算法规制理论的批判》,《环球法律评论》2019年第6期。

<sup>⑥</sup> See European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence [2020/2014(INL)], Civil liability Regime for Artificial Intelligence, <https://beta.op.europa.eu/en/publication-detail/-/publication/1923c62a-2640-11ec-bd8e-01aa75ed71a1/language-en/format-PDF/source-202040097,2023-05-28>.

<sup>⑦</sup> 参见宁金成、李瑞升:《人工智能伦理规则的产品质量立法构建》,《河南财经政法大学学报》2021年第2期。

能生成内容可能未必能够作为著作权法意义上的“作品”，但是由于生成式人工智能生成内容系人工智能衍生之物，只要落入公众使用的范畴，就可视为具有“作品”外观的人工智能产品，人工智能提供者即可视为人工智能生成内容的产品提供者。与之相对应，以欣赏、研究和学习为目的，通过免费或者付费方式接触人工智能产品的公众即为消费者。

与形式正义不同，经济法的核心法理之一即实现实质公平，从社会整体出发，在坚持平等原则的基础上根据利益和责任的差异性进行权利与义务的分配，通过对弱势群体给予特别保护、施加经营者更重义务和责任进行矫正，以实现结果上的正义。<sup>①</sup> 由此也衍生出经济法对人工智能介入的新兴创作物市场秩序的规范。在传统的创作物市场中，著作权人与公众对作品的种类和性质的认知不存在信息差，双方处于平等的民事主体地位，公众作为“纯粹的理性之人”可与著作权人达成意思表示真实的私人自治。然而，在人工智能时代的创作物市场，生成式人工智能生成内容的提供者并没有披露生成式人工智能参与创作的内生动力，一旦披露其由生成式人工智能生成则会降低该创作物的市场竞争力。同时，生成式人工智能生成内容具有难以识别的“作品”外观，使公众“从理性、意思表示强而智的人变成弱而愚的人”，<sup>②</sup>将消费者置于严重的信息劣势之中，而其自身也没有克服信息不对称的能力。申言之，将生成式人工智能生成内容伪装成人类作品的行为的本质是人工智能实际控制者利用自身信息优势侵害相对人获取真实信息的权益。而赋予数字产品的消费者以知情权(包括产品的实际创作者、创作过程、产品信息内容的标准、信息范围等)，可以有效解决其与产品提供者间的信息不平等关系，以适应数字内容产品的科技性和复杂性。<sup>③</sup> 因此，法律应当明确公众享有知悉创作者真实身份的权利，以对抗人工智能实际控制者的伪装行为，从而修正双方的不平等地位。

根据美国法理学者霍菲尔德的权利分析理论，权利与义务具有“相关”关系，即当某一主体被赋予权利时他人就有责任履行相应义务。<sup>④</sup> 赋予消费者知情权的核心不仅在于权利得到确认，更在于义务主体应当履行相应的信息披露义务。<sup>⑤</sup> 根据《中华人民共和国消费者权益保护法》第20条的规定，经营者向消费者提供有关商品或者服务应当承担“真实、全面”披露信息的义务；根据《中华人民共和国产品质量法》第27条的规定，经营者负有标示真实产品标识的义务，亦体现了对消费者知情权的保护。上述“真实、全面”的信息披露义务理应构成生成式人工智能生成产品提供者向公众履行人工智能标识义务的重要内容。生成式人工智能生成产品存在伪装为人类作品并流入创作物市场的可能性，其产品提供者本身制造了一个破坏消费者预期的“危险源”。生成式人工智能生成产品提供者对于该产品具有源头上天然的控制能力，若其不能为消费者披露人工智能生成的必要信息，则创作物的识别成本将不可避免地转移到信息能力更弱的消费者身上。公众只有事先明确该内容由人工智能创建之事实，才能作出相对理性的选择。也就是说，只有生成式人工智能生成产品提供者采取措施履行生成式人工智能生成产品的标识义务，才最

<sup>①</sup> 参见殷继国、阳雨璇：《寻找经济法中的法理——“经济法中的法理”学术研讨会暨“法理研究行动计划”第七次例会述评》，《法制与社会发展》2019年第6期。

<sup>②</sup> 参见[日]星野英一：《私法中的人》，王闯译，中国法制出版社2004年版，第50页。

<sup>③</sup> 参见云薇笑：《我国数字内容产品消费者权益保护研究》，《兰州学刊》2022年第4期。

<sup>④</sup> See Wesley Newcomb Hohfeld, *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 26 *The Yale Law Journal*, 723-750(1917).

<sup>⑤</sup> 参见王宏：《消费者知情权研究》，山东人民出版社2015年版，第164页。



符合创作物市场交易的公平与效率原则。

### (三)创作物市场共同善之展望

在生成式人工智能生成内容进入创作物之前,作品为人类进行创作劳动所产生的智力成果,仅指向作者的个体权利。著作权作为一项法律权利,其证成理论无论从“利益论”还是从“意志论”的角度而言都具有浓厚的个人主义特征。<sup>①</sup>然而,生成式人工智能生成内容的产生并不仅源于个体,生成式人工智能提供者、算法技术设计者、人工智能使用者都是生成式人工智能生成的必要条件,而其中最重要的人工智能机器学习的训练数据来自自然人的现有创作。因此,对于生成式人工智能如果仅仅从某个单一维度进行赋权,难以维护正常的创作物市场秩序,更会导致各方主体权利与义务的失衡。英国学者拉兹提出的共同善理论批判了个人主义权利,主张将焦点从个人(权利人)维度转移到社会或集体面向。<sup>②</sup>相关主体承担生成式人工智能生成内容的标识义务,将有利于构建人工智能创作物市场的共同善。

拉兹认为:“只有当其他人的利益与权利人的利益和谐交织在一起的时候,亦即只有当有益于权利人是有益于其他人的一种方式时,以及权利人的利益要通过有益于其他人而得到促进的地方,其他人的利益对于证成权利才是重要的。”<sup>③</sup>当我们探究人工智能使用者或人工智能提供者对该生成内容是否具有何种权利或承担某项义务时,不能将个体利益与他人利益或社会整体利益人为割裂。生成式人工智能所需要的海量训练数据大多来源于公众即人类作者的创作,生成式人工智能最终生成的内容也会流向以人类为受众主体的创作物市场。现阶段,无论是国内学者观点还是国外实践,都趋于将生成式人工智能的数据源与文本挖掘纳入著作权的合理使用范畴,以促进生成式人工智能技术发展并减少使用版权材料的交易成本。<sup>④</sup>对于生成式人工智能机器学习作为合理使用的底层逻辑不仅是对人类作者的“祛魅”,也是对著作权权利的解构——从注重著作权人的个体利益“权利中心”到生成式人工智能背景下创作物共同善维度。然而,生成式人工智能机器学习对于人类作品的“艺术抢劫”已经引起了人类创作者的不满甚至抗议,人类创作者可能会采取技术手段使得某些图像无法被生成式人工智能提取有效信息,以此来抑制生成式人工智能的发展。未来生成式人工智能对人类创作者经济利益的冲击虽然无法避免,但由相关主体履行标识义务则是对人类创作者的利益保留,至少能够在一定程度上降低生成式人工智能生成内容实际控制者的伪装行为对人类创作者精神利益的冲击。人类作者在未来的创作物市场中应当有一席之地,毕竟生成式人工智能永远无法取代人类的艺术性和灵魂。以生成式人工智能生成内容标识义务为基石的创作物市场共同善旨在建立相对公平的人类创作者与生成式人工智能的竞争环境,促进双方朝着共善的方向发展。

共同善的权利观中个人的义务未必与其追求自身利益相冲突,相反,义务常常界定了一条主干道,个人可据此促进其福利。事实上,人工智能使用者也具有主动披露创作过程相关信息的动

<sup>①</sup> See William A. Edmundson, *An Introduction to Rights*, Cambridge University Press, 2004, pp.119—127.

<sup>②</sup> 转引自王籍慧:《需求如何成为权利——共同善权利观的两种论证方式及其限度》,《法学论坛》2019年第3期。

<sup>③</sup> Joseph Raz, *Rights and Individual Well-Being*, in Joseph Raz ed., *Ethics in the Public Domain: Essays in the Morality of Law and Politics*, A Clarendon Press Publication, 1994, p.51.

<sup>④</sup> 关于人工智能机器学习作为合理使用的问题,参见林秀芹:《人工智能时代著作权合理使用制度的重塑》,《法学研究》2021年第6期。

机。例如,在“《黎明的扎里亚》登记案”中,卡什塔诺娃在社交媒体上公开表示该漫画书系使用Midjourney生成。也就是说,人工智能使用者可能因为宣传或其他利益需求主动披露生成式人工智能生成内容的环节,标识义务未必不利于实际控制人本人。而对于人工智能提供者而言,也可通过披露事实作为“广告”向公众介绍其卓越的人工智能产品,以此建立竞争优势。因此,生成式人工智能生成内容实际控制者的标识义务并非完全被动,其本身亦存在披露动力。可信度和透明度对人工智能产业发展和保护创作者至关重要。公众只有知悉创作物的真实属性,才能真正消减对人工智能产品的不信任,从客观上激励和促进人工智能产业发展。在算法与人工智能介入的创作物市场,随着“数字正义”理念的提出,“数字向善”原则已逐渐超越“技术中立”,更强调社会相关主体在数字技术发展应用之中要尊重社会伦理与公民权利。<sup>①</sup>对生成式人工智能生成内容进行标识的义务仅仅限制了那些不具合理性和道德的部分“自由”,却能在符合个人利益的同时服务于创作物市场环境的共同善。

无论是民法上的诚实信用原则、经济法上的产品信息披露义务,还是人工智能背景下的共同善展望,都要求对掩盖生成式人工智能创作过程的伪装行为进行矫正和规范,生成式人工智能相关主体应承担生成式人工智能生成内容的标识义务,以此推动创作物市场透明化,这是多方主体利益分配的最优选择。

### 三、路径选择:标识义务的场景化设定

考虑到生成式人工智能生成内容的类型与创建方式不同,标识义务的设置应依据相应场景进行区分以平衡各方权益。一方面,基于不同创作物类型,对人工智能使用者和提供者设定相应的标识义务;另一方面,如果相关主体违法实施伪装行为,可基于标识义务来追究其法律责任,引导公众理性选择并保护人类作者的创造激励。

#### (一)生成阶段

在生成式人工智能生成内容阶段,人工智能程序(归责于人工智能提供者)与人工智能使用者直接参与生成内容的生成,并由人工智能使用者决定是否将生成内容予以传播以及如何署名。因此,解决伪装行为的最佳途径是依靠人工智能使用者与人工智能提供者之间的联结——用户协议——来调整。用户协议可以按照生成式人工智能是作为“工具性使用”还是“创造性主体”来确定标识义务的必要性。<sup>②</sup>如果生成式人工智能算法仅在人类的指导下进行训练和调整,或者需要人类干预才能产生内容并发挥其真正的价值,则可以认定生成式人工智能是作为“工具性使用”而使用。工具性使用下的生成式人工智能程序与传统用于作品创作或发明创造的其他工具相比,并无实质差别。例如,使用生成式人工智能帮助搭建文章提纲、润色文章语句、检查拼写错误等都可视为工具性使用,人工智能使用者无须披露生成式人工智能参与的过程。但如果生成式人工智能可以独立地生成全新的内容,并且这些内容在艺术或其他领域中被认为是有创新和

<sup>①</sup> 参见张吉豫、汪赛飞:《数字向善原则下算法推荐服务提供者的著作权注意义务》,《知识产权》2022年第11期。

<sup>②</sup> See Kalin Hristov, *Artificial Intelligence and the Copyright Dilemma*, 57 IDEA—The Journal of the Franklin Pierce Center for Intellectual Property, 435(2017).

价值的,则可以认定生成式人工智能作为“创造性主体”而使用。例如,在“菲林律师事务所案”<sup>①</sup>中,人民法院认定人类参与者对于独创性并无实质性贡献,生成式人工智能则为“创造性主体”。当生成式人工智能程序可独立生成输出内容的独创性表达要素,人工智能使用者就应当对生成式人工智能参与过程主动承担标识义务。

通过用户协议对人工智能使用者的约束依靠其本人自律,生成式人工智能生成内容的标识则更具有区分的强制性。中国信息通信研究院《人工智能生成内容(AIGC)白皮书(2022年)》建议相关企业可以针对生成内容主动进行标识。2023年8月7日,全国信息安全标准化技术委员会发布了《生成式人工智能服务内容标识方法》(TC260-PG-20233A)用以指导生成式人工智能提供者提高内容标识要求。该标准指出,在生成式人工智能服务中可通过“显式水印标识”和“隐式水印标识”对不同类型的生成式人工智能生成产品进行标识。由于生成式人工智能生成内容种类不同,可以依据不同生成内容和应用场景设定不同的标识义务。这里,建议将生成式人工智能生成内容分为生成式视觉艺术产品和非视觉艺术产品,对于前者应当通过合理的水印标识履行强制性标识义务,而对于后者则可提供任意性标识选择。生成式视觉艺术产品与视觉艺术作品外观类似,是指利用计算机程序和算法,根据一定的规则和随机性,创造出来的具有美感和创意的图形、动画或视频等内容,可以通过视觉感知途径传达创作者的思想、情感和审美观点。生成式视觉艺术产品以生成的图像作为艺术品呈现出来,其涉及原始图像的独创性剽窃概率更高且对人类作者的利益冲击更大,对公众而言其真实性更加难以识别。与此同时,生成式视觉艺术产品本身标注的难度较低,水印或其他标注信息能够被公众清晰识别,从而达到披露目的。相较而言,生成式人工智能生成非视觉艺术产品则具有可碎片化的特征,即使用者采取拷贝或剪辑的方式能够使生成式人工智能生成的文本与音乐脱离原内容,因此对于该类作品的标识较为困难,强行通过技术手段标识容易破坏生成式人工智能服务产品本身的吸引力和商业价值。

当生成式人工智能生成内容未通过可视化方式披露其属性,生成式人工智能生成内容与人类作品的区分就需要依靠技术检测手段。现有生成式人工智能生成文本检测方法大致可分为两类:黑盒检测和白盒检测。黑盒检测方法对大语言模型通常只有应用程序接口级别的访问权限,其依靠收集人类和机器的文本样本来训练分类模型;白盒检测则拥有大语言模型的所有访问权限,可以通过控制模型的生成行为或者在生成文本中加入水印来对生成内容进行追踪和检测。随着大语言模型能力的提升,其生成的文本和人类的差距会逐渐缩小,从而导致黑盒模型的检测准确性越来越低。<sup>②</sup> 相比之下,白盒检测更有应用前景,检验结果更为精准,但白盒检测依赖于程序开发者对开源模型的设计。与此同时,人工智能技术标准作为重要的“软法”治理工具,可在人工智能技术开发早期即嵌入伦理和价值要求。当生成式人工智能技术介入创作物市场并冲击人类创作者地位,附加标识义务的技术标准应成为生成式人工智能产品质量认证的核心,人工智能提供者应当约束程序设计者,从而为产业发展提供可持续的共同善路径。需要注意的是,若采取可识别水印会实质性影响产品服务,或者标识生成式人工智能生成内容从技术上难以实现,可在设计算法时确保生成式人工智能产品生成内容具有可检测性。

<sup>①</sup> 参见北京互联网法院(2018)京0491民初239号民事判决书。

<sup>②</sup> See Ruixiang Tang, Yu-Neng Chuang, Xia Hu, The Science of Detecting LLM-Generated Texts, <https://arxiv.org/pdf/2303.07205.pdf>, 2023-07-25.

## (二)传播阶段

对于传播阶段的标识义务,可结合其风险的类型分为涉财产利益场景、涉消费者利益场景以及涉作者精神利益场景3类。

1.涉财产利益场景。目前,我国相关立法并未明确生成式人工智能生成内容与“作品”的关系,这里可以适用诚信原则化解生成式人工智能伪装行为的财产性风险。(1)行为人在缔约过程中将创作物进行权益变更时,应当主动承担标识义务。在双方签订涉及创作物权益许可使用或转让合同时,创作物提供方应当遵循诚实信用原则,如实披露该创作物是否由生成式人工智能参与主体创作的过程。若行为人与他人订立合同时故意隐瞒生成式人工智能参与主体创作的过程,导致对方在违背真实意思的情况下订立合同,可认定其实施了民事欺诈行为,适用《民法典》第500条第2、3项之规定,行为人应当承担缔约过失责任。(2)行为人在完成投资者作为著作权主体的作品时应当审慎使用生成式人工智能,并向相关主体披露使用情况。《著作权法》在第11条第3款所规定的法人作品、第15条所规定的汇编作品、第17条所规定的视听作品、第18条所规定的职务作品、第19条所规定的委托作品以及邻接权(除表演者外)中将著作权分配给投资者。<sup>①</sup>在上述作品的创作过程中,因作品的财产权益并不归属于创作者本人,故若创作人怠于人为创作而采取冒用生成式人工智能生成内容的不诚信行为,则可能导致涉及该创作物的多方主体相关财产权益存在灭失风险。根据《民法典》第509条的规定,当事人应当遵循诚信原则,根据合同性质、目的等履行相关义务。创作者既然根据合同约定需履行完成提供“作品”的义务,那就理应恪守诚信原则向投资者披露是否使用生成式人工智能以及使用程度,以防范因著作权基础权利的丧失而产生的合同目的落空以及权益受损的风险。

2.涉消费者利益场景。网络平台是信息时代公众与创作物直接接触的主要媒介,也是生成式人工智能生成内容传播的工具。平台作为创作物市场的经营者,处于信息不对称中的强势一方,应当尽其所能为处于弱势地位的消费者提供全面的产品信息、履行标识义务,以维持消费者与平台之间的信赖。例如,抖音平台《关于人工智能生成内容的平台规范暨行业倡议》第4条“发布者应对人工智能生成内容进行显著标识,帮助其他用户区分虚拟与现实,特别是易混淆场景”和第9条“平台将提供统一的人工智能生成内容标识能力,帮助创作者打标,方便用户区分”都旨在承担创作物传播阶段的标识义务。这两条虽仅为倡议条款,但也表明了平台对新兴创作物市场的包容态度与对消费者知情权的尊重。需要注意的是,德国学者对126名人工智能使用者展开的实证研究发现了代写者效应,即人们在写作中使用人工智能生成文本时,通常不会透露人工智能是作者;但在当作者被给予添加人工智能的明确选项时,他们则更有可能披露人工智能行为。<sup>②</sup>因此,平台可以给创作物提供者设计合适的作者身份声明框架,如提供“存在人工智能参与创作”的选项,或给予一定空间使创作物提供者能够清楚地区分自己和人工智能的贡献。平台通过对创作物提供者来源更清晰的认定有利于在一定程度上满足人工智能使用者对于生成式人

<sup>①</sup> 参见刘铁光:《著作权主体的二元结构及其权利配置——兼评2020年〈著作权法〉修正的主体制度》,《知识产权》2021年第8期。

<sup>②</sup> See Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, Robin Welsch, The AI Ghostwriter Effect: When Users Do Not Perceive Ownership of AI-Generated Text But Self-Declare as Authors, <https://doi.org/10.48550/arXiv.2303.03283>, 2024-04-04.

工智能产品的“作者”归属感,从而鼓励人工智能使用者诚实守信而透明地披露其创作过程。<sup>①</sup>

与此同时,当消费者怀疑创作物提供者上传内容由生成式人工智能生成,可向平台提出异议并主张知情权。平台应对该内容采取技术检测手段,对违反标识义务的创作物予以暂时下架并通知相关主体,待创作物提供者提供证据予以澄清或采取标识措施后方允许其流入市场。对于反复上传生成式人工智能生成内容的失信创作物提供者,平台可以采取封禁账号、列入黑名单等惩罚性措施。

3.涉作者精神利益场景。有别于网络平台对多元创作物开放包容的态度,由于学术研究的严谨性和出于对作者精神权利的保护,教育与学术领域大多对生成式人工智能介入的成果持谨慎和保守态度。在教育方面,美国纽约市教育部门以对学生学习的负面影响为由宣布禁止学生在学校的设备和网络上使用 ChatGPT,巴黎政治学院也对 ChatGPT 实行全面封禁。在学术领域,如《自然》虽未禁止 ChatGPT 使用但要求作者必须在“方法”或“致谢”或适当的部分中披露完整的细节;<sup>②</sup>《科学》则明令禁止论文中使用 ChatGPT 生成的文本。<sup>③</sup>《暨南学报》(哲学社会科学版)等国内期刊发布了《关于使用人工智能写作工具的说明》,要求人类作者应当披露生成式人工智能的使用,否则将予以撤稿处理。另外,根据《中华人民共和国学位法》(以下简称《学位法》)第 37 条的规定,学位论文或者实践成果被认定为存在代写、伪造等学术不端行为,经学位评定委员会审议决定,学位授予单位不授予学位或者撤销学位。基于此,高校可在其“自主性校规”<sup>④</sup>中明确完成学业任务时使用生成式人工智能实施代写、伪造属于学术不端行为,并进一步细化以何种方式使用生成式人工智能撰写学位论文属于违规情形。同样地,期刊的知识产品属性决定了期刊出版单位享有“专有权”,即期刊出版单位可以自主控制和经营期刊产品。<sup>⑤</sup> 期刊出版单位有权发表声明禁止投稿人使用生成式人工智能撰写文章或要求披露其生成式人工智能创作行为。高校和出版单位亦可采取技术手段对相应文本进行检测以查明是否存在生成式人工智能生成内容的情况,以确保内容的安全性和准确性。质言之,上述对创作物内容准确性、创造性要求较高的场景下,掩盖生成式人工智能创作过程的伪装行为对作者社会评价体系和公众对创作物真实性信任度损害较大,人工智能使用者应当承担强制性标识义务。

人工智能使用者固然存在掩盖生成式人工智能生成内容的利益诱惑,但社会信任机制促成了对诚信价值观法律化的要求。如前所述,人工智能使用者对于生成式人工智能的“工具性使用”通常并无标识义务,只有使用者将生成式人工智能作为“创造性主体”独立生成该创作物而采取隐瞒行为时才有必要追究人工智能使用者相应的失信责任。在教育场景下,如果法律或学校校规已明确禁止使用生成式人工智能,事后发现学生违反规定实施了欺骗行为,学校可对其采取相应的惩戒措施以确保教育的公平性。在作品发表场景下,倘若出版单位事先明确稿件禁止使

<sup>①</sup> 2024 年 3 月 21 日,美国国会议员安娜·G.埃肖和尼尔·邓恩提出了“保护消费者免受欺骗性人工智能法案”,通过指导制定识别和标记生成式人工智能生成内容的标准,要求生成式人工智能开发人员和在线平台披露人工智能生成内容。

<sup>②</sup> See <https://www.nature.com/nature-portfolio/editorial-policies/ai>, 2023-06-12.

<sup>③</sup> See H. Holden Thorp, ChatGPT is Fun, But Not a Author, 379 Science, 313 (2023).

<sup>④</sup> 根据《中华人民共和国高等教育法》第 11 条的规定,高校具有依法自主办学的权利,学校可在不违背法律原则的情况下制定相应的“自主性校规”。

<sup>⑤</sup> 参见王国柱:《著作权法中期刊出版者权利的体系化配置》,《吉林大学社会科学学报》2020 年第 3 期。

用生成式人工智能参与创作或必须披露生成式人工智能参与过程,而事后发现存在生成式人工智能生成内容或未按规定披露生成式人工智能参与过程的情形,则可基于“学术不端”理由撤销稿件,并可将该作者列入学术失信“黑名单”。

## 四、结 语

在人工智能技术日益成熟和普及的背景下,生成式人工智能生成内容已经悄然流入创作物市场并显露出逐渐扩张之态势。机械性、非人格性的生成式人工智能技术可能创建出胜似乃至超越具备独特人格与自由意志的人类创作,挑战甚至颠覆在整体上以人类为中心的传统著作权制度的理论基础。将生成式人工智能生成内容伪装成人类作品的行为涉及多方主体利益,若不加以规制,将会损害创作物市场的共同善,导致创新的缺失甚至作品多样性的减损。诚实信用原则与产品信息披露义务支持生成式人工智能生成内容实际控制者的标识义务;创作物市场的共同善期望以程序设计为源头采取场景化的标识措施,将在保障作者、消费者和社会的合法权益的前提下实现生成式人工智能产品服务的商业价值。生成式人工智能技术应当与人类建立一种基于互信与互惠的合作伙伴关系,而不是欺骗与对抗的敌对关系,法律规则应当引导生成式人工智能“科技向善”,促进未来创作物市场的良性有序发展。

---

**Abstract:** In order to obtain complete copyright, users of generative artificial intelligence often disguise the generated content as human works to cover up the substantial participation of artificial intelligence, which seriously damages the property rights and spiritual interests of the public and natural creators, and undermines the stable order and trust relationship of the creation goods market. This behaviour should be actively dealt with. The principle of good faith in Civil Law and the obligation of product information disclosure in Economic Law respectively provide the legal basis of “self-discipline” and “heteronomy” for the identification obligation of the generated content of generative artificial intelligence. The concept of “common good” contained in the transparent creation market encourages individual interests to serve the common good of society. The approaches to ensure the implementation of the identification obligations of the relevant subjects of generative artificial intelligence include: in the generation stage, adopt the identification mode of generative visual art products, and ensure the detectability of the generated content when designing the algorithm; in the communication stage, relevant subjects should be guided to disclose information independently, and a credit system should be established to ensure the implementation of obligations.

**Key Words:** generative artificial intelligence, identification obligation, principle of good faith, common good

---

责任编辑 何 艳