Journalism and Communications 新闻传播科学, 2026, 14(9), 27-34
Published Online September 2026 in Hans. https://www.hanspub.org/journal/jc
https://doi.org/10.12677/jc.2026.149233

Hans汉斯

生成式AI参与下的模因生成机制、内容扩散与价值反思

密梦琳

北京印刷学院新闻传播学院，北京

收稿日期：2026年8月2日；录用日期：2026年8月28日；发布日期：2026年9月8日

摘　要

生成式人工智能参与生成的网络模因在互联网中广泛传播，如“AI山海经”“我的刀盾”梗图以及各种AI二创影视剧片段等，并对信息生态和价值观产生显著影响。生成式人工智能的深度介入，推动着模因创作扩散的人机关系的演变。本文以文生图、文生视频工具产出的AI视觉模因为研究对象，本文系统讨论了AI视觉模因的生成机制、扩散机制及其引发的价值反思与治理路径。在生成层面，AI扮演了三种角色：作为模因符号被用户娱乐化解构；作为原创者依赖算法与数据生成海量原生内容；作为变异者通过结构性与语义性变异实现模因的多元迭代。在扩散层面，AI的扩散主体呈现主体泛化的特征，扩散路径突破线性链式结构，呈现非线性渗透的特征。同时本文基于AI视觉模因的“病毒式传播”进行了价值反思，并指出主体性偏移、文化多样性侵蚀、信息失序等问题和挑战。针对上述问题，本文从技术、用户、政府三个不同主体层面提出协同治理路径。本文希望能够通过梳理AI视觉模因的生产机制，引发人们对技术与人类关系的重新思考和反思，促进生成式人工智能的健康发展，更好地为人所用。

关键词

生成式人工智能，AI视觉模因，人机协同，价值反思，治理路径

Meme Generation, Diffusion, and Value Reconsideration under Generative AI: Mechanisms and Implications

Menglin Mi

School of Journalism and Communication, Beijing Institute of Graphic Communication, Beijing

Received: August 2, 2026; accepted: August 28, 2026; published: September 8, 2026

文章引用: 密梦琳. 生成式 AI 参与下的模因生成机制、内容扩散与价值反思[J]. 新闻传播科学, 2026, 14(9): 27-34.
DOI: 10.12677/jc.2026.149233

Abstract

Generative AI-generated network memes have spread widely on the internet, such as the "AI Classic of Mountains and Seas" and "My Sword and Shield" memes, as well as various AI-generated film and television clips, significantly impacting the information ecosystem and values. The deep involvement of generative AI is driving the evolution of the human-machine relationship in meme creation and dissemination. This paper focuses on AI visual memes produced by tools like Wensheng Image and Wensheng Video, systematically discussing their generation and diffusion mechanisms, and the resulting value reflections and governance paths. At the generation level, AI plays three roles: as a meme symbol deconstructed and entertaining by users; as a creator relying on algorithms and data to generate massive amounts of original content; and as a mutant achieving diverse iterations of memes through structural and semantic variations. At the diffusion level, the main body of AI diffusion exhibits a generalized characteristic, and the diffusion path breaks through the linear chain structure, exhibiting non-linear penetration characteristics. This paper also reflects on the value of the "viral spread" of AI visual memes, pointing out problems and challenges such as subjectivity shift, cultural diversity erosion, and information disorder. To address these issues, this paper proposes collaborative governance paths from three different subject levels: technology, users, and government. This article aims to stimulate a rethinking and reflection on the relationship between technology and humanity by analyzing the production mechanism of AI visual memes, thereby promoting the healthy development of generative artificial intelligence and enabling it to be better utilized by humans.

Keywords

Generative Artificial Intelligence, AI-Mediated Memes, Human-Machine Collaboration, Value Reconsideration, Governance Pathways

1. 研究缘起与研究方法

1.1. 研究背景

随着生成式人工智能技术的不断革新，人工智能已经从工具逐渐变成互联网内容生产的重要参与主体。生成式 AI 参与生产的内容在互联网中实现“病毒式传播”，成为影响人们认知的 AI 视觉模因，例如各种融合生物形象的合成兽“AI 山海经”、“雪山救狐狸”AI 短剧、“AI 古人说话”等，并引发大量二次创作。这些内容不仅体现了人工智能技术的创造能力，也反映出模因生产和传播机制的深刻变革。

1.2. 研究方法

本文主要采用参与式网络观察法与案例分析法。在 2025 年 6 月至 2026 年 6 月期间，研究者在抖音、小红书、哔哩哔哩三个平台，以“AI 绘画”“AI 视频”“AI 梗图”等为关键词进行持续追踪观察，累计收集 AI 生成视觉模因样本 1200 余条。同时，对“AI 山海经”“企鹅舞”等 6 个典型案例进行深度分析，以了解提示词调整、生成结果筛选和内容再加工等环节。在此基础上，本文通过案例分析归纳视觉型 AI 视觉模因的生成方式、变异方式和扩散特征。

1.3. 模因演化与人机关系变迁

“模因”(meme)最早是由理查德道金斯在其所著的《自私的基因》中，仿照达尔文进化论“基因”一词创造出来的，强调文化信息可以像“基因”一样复制和传播。模因作为文化复制因子，有着“基因”相似的遗传、选择、变异机制[1]。尼尔波兹曼的媒介生态学认为媒介技术并非中性的，而是作为一种环境结构对人、文化、社会等产生影响。媒介技术的每一次更迭都推动着模因生产和传播范式的深刻改变，并带动人与技术、人与机器关系的变迁。

在口语、文字广播等传统媒介的时代，人类是模因创作的唯一主体，其生产和传播是由人类主导的，模因主要依赖人与人之间的模仿和传播完成复制，文化生产主体完全由人类构成。进入互联网时代后，数字技术赋予用户便捷的复制与编辑能力，网络模因逐渐发展为表情包、短视频、网络热梗等多种形式。此时机器主要承担传播工具和技术平台功能，但是其生产和扩散仍然是由人类主导的文化选择，是人类文化之间的竞争。

生成式人工智能出现后，人机关系进一步发生变化。机器不仅仅承担传播和编辑功能，而是直接参与内容生产和意义建构。用户提供提示词与人工智能一起完成内容创作，大模型依据训练数据和算法逻辑生成文本、图像和视频。AI 视觉模因的意义和文化价值不是由人类完全决定的——模型的训练数据、生成概率、内在价值观差异都深深塑造了 AI 视觉模因的语义偏向。在这一过程中，模因生产逐渐由“人机协作”转向“人机协同”，网络文化也进入算法参与的新阶段。

2. 生成式 AI 参与模因生产的生成机制

有学者根据人工智能在模因生成中的介入方式将生成式 AI 视觉模因分为三种类型，本部分同样基于介入方式对生成式人工智能参与下的模因生成进行讨论，进行分工和角色的划分，揭示生成式 AI 在模因生成阶段的作用机制。在 AI 视觉模因的生产中，人工智能既是被用户戏谑和解构的文化符号，也是内容生成的重要主体，更是推动模因持续变异的重要力量。

2.1. 作为模因符号的人工智能：AI 自身的文化梗化

模因本质上是一种可复制和传播的文化符号。随着生成式 AI 深度介入人们的生活，AI 本身也成为网络文化解构和再创造的对象。用户通过对人工智能产品形象、交互方式和技术特征进行娱乐化解读，使 AI 从技术工具转变为网络文化符号。这种 AI 视觉模因可以分为两类，一类是 AI 产品拟人化形象的梗化。抖音平台“豆包”相关话题下涌现大量对豆包拟人头像的恶搞二创，“豆包家族”系列梗图获得数百万播放。用户在评论中表示“把 AI 当人玩才有意思”，对 AI 进行戏谑和娱乐化解读。二类是 AI 与人类交互行为的梗化。AI 经常出现答非所问等技术缺陷，用户利用这一缺陷制造热梗。如对豆包进行“调教”、向 ChatGPT 提问“你知道自己是 AI 吗”、ChatGPT 高频句式“我会稳稳地接住你”被制作为系列表情包。在研究中发现，当向豆包连续追问逻辑矛盾问题时，AI 会出现明显混乱反应，这种“AI 翻车”正是用户乐于制作成视频传播的内容。用户通过对 AI 技术缺陷的反向解读与二次创作，打破对 AI 的技术崇拜，将 AI 工具本身转化为大众娱乐文化符号，该行为本质是普通用户借助通俗创作完成对算法技术局限的公开讨论与生活化表达。

2.2. 作为模因原创者的人工智能：算法驱动下的原生内容创作

从生成式人工智能的技术原理上看，生成式人工智能通过生成式预训练模型学习海量的数据和语料，掌握数据集知识的潜在规律，进而以字词接龙的形式自动生成新的“原创”内容[2]。生成式 AI 能在极短的时间内生成满足人类需求和符合网络传播趋势的海量原创内容和模因复合体，这为模因的规模化复制

和传播提供了基础。虽然 AI 在模因生成阶段具有巨大的效能优势，AI 生成模因的质量仍然受到某些因素的制约。首先，用户与 AI 对话的能力会直接影响 AI 生成内容的质量和水平。当用户越熟悉网络传播逻辑时，AI 视觉模因的网感会越强，传播潜力会更大。研究中发现，在“AI 山海经”传播过程中，大量创作者会在评论区公开提示词模板，并分享如何调整关键词、镜头语言和角色设定。许多热门作品实际上经历了多轮提示词优化和内容筛选，AI 视觉模因的原创生产并非机器单独完成，而是人机协同创作的结果。AI 视觉模因生成更重要的是大模型算法的“选择”，是大模型对用户数据学习后智能选择的结果[3]。由于算法的运行机制具有不可见性和不可解释性，所以这种算法偏见对内容的影响是难以察觉的，如政治思想、社会结构性偏见、用户偏好偏见等都会通过不同的方式嵌入到智能模因的生产中。AI 看似具备独立生成网络梗、创意符号的自主创作能力，实则是人类交互能力、训练数据范式与算法内在偏见共同约束下的概率性文本再生产。

2.3. 作为模因变异者的人工智能：既有内容的多元迭代

模因能够持续传播的重要原因在于其不断发生变异。传统网络模因的变异主要依赖用户的二次创作，而生成式人工智能则极大拓展了模因变异的速度和空间。一方面，AI 能够对既有文化内容进行重组和再创造。例如，将经典影视作品、历史人物和网络热梗进行跨时空组合，形成新的文化表达形式。“甄嬛传 AI 恶搞视频”“古人回答我系列”等内容均体现了这一特征。另一方面，AI 还能够对生成式 AI 视觉模因进行再造和变异。两个以上的模因可以组成模因复合体，其形成过程通常伴随着新的模因加入到旧的模因或模因复合体中。旧模因与新加入的模因重新组合，或是各自舍弃部分模因后组合，都可以形成新的模因[4]。在抖音平台中，同一“企鹅”表情包会在不同创作者手中经历多轮再创作，有的利用 AI 工具改变角色外形，有的则赋予其新的故事背景和人格设定。在生成式 AI 介入后，模因依托人机协同实现持续演化，迭代次数与变异空间极大拓展，大量形态各异的变体相互融合，逐步形成多元共生的模因复合体。

3. 生成式 AI 参与模因扩散机制

3.1. 扩散主体：人机协同的复合格局

传统模因传播主要依赖人与人之间的模仿和转发，而生成式人工智能参与后，模因扩散主体逐渐由单一人类主体转向人类主体与“机器主体”联动的格局。

在 AI 视觉模因传播过程中，普通用户、意见领袖、人工智能系统和平台算法共同构成传播网络。一方面，AI 可自主生成模因变体持续供给传播素材，为圈层扩散提供内容储备；另一方面，人类依靠社会经验、文化感知筛选具备传播潜力的内容。平台算法作为隐性行动者，依靠推荐机制完成流量分配，有研究表明，平台通过热榜与编辑工具的双重干预，显著影响模因的传播效率和范围[5]。“企鹅舞”AI 视频其原始视频经 AI 生成后，在网络传播中又衍生出多种变体，如“colder than ice”“不潮不用花钱”“企鹅新疆舞”“明月几时有”“辛巴舞”等十几种企鹅舞视频，其中只有“colder than ice”企鹅舞得到大范围传播，最高点赞量达 430 万，这是“企鹅舞”模因传播中，AI、人类、算法共同选择和作用的结果。如果说传统模因和网络模因的扩散是人类文化之间的竞争，AI 视觉模因的扩散则是人–机器、机器–机器的新型文化竞争关系。

3.2. 扩散路径：非线性渗透和辐射状扩散网络

从扩散路径来看，传统互联网模因的扩散路径主要是依赖于社交关系的线性传播，通常沿着“创作者–转发者–再转发者”的路径逐层扩散，传播路径受到社交关系链的限制。AI 视觉模因则突破了线性

链式结构，呈现出非线性渗透、去中心化的新特征。这里的非线性渗透指的是，在算法推荐与用户主动分享的双重驱动下，AI 视觉模因的扩散不再依赖于既有的社交关系链条逐层传递，而是通过平台流量池的跨圈层推荐、多中心的同时引爆以及模因变体的杂交，形成一种多源发、多路径的辐射状扩散格局。AI 视觉模因的传播经过了三个主要环节：一是算法分发渗透，平台推荐、话题标签和弱关系网络使内容能够越过原有社交圈层进入新的受众群体；二是同步变异，内容抵达受众后可直接借助生成式 AI 工具完成图像、视频重构，在传播过程中同步产出模因新变体；三是多分支扩散，各类差异化模因变体独立形成传播分支，多分支并行扩散、相互交叉融合，最终形成去中心化辐射传播网络。传播与生产同步进行，构成“生成–变异–扩散–再生成”的循环机制。借助拉图尔的行动者网络理论，可以将 AI 视觉模因的扩散网络视为一个由人类用户、算法推荐系统、AI 生成工具共同构成的动态异质网络，其中少数高影响力的节点即可触发大规模传播。以“AI 山海经”为例，其传播并非单一路径扩散，而是不断产生新的角色、故事和视觉风格。大量变体同时传播，最终形成辐射状传播网络。与传统模因相比，AI 视觉模因具有更强的裂变能力和流动性。

4. 生成式 AI 参与网络模因生产的价值反思

4.1. 创作主体弱化风险：人类主动性减弱与浅层认知固化

生成式人工智能介入模因生产，人类创作者面临的不只是工具替代，更是一种深层危机–主体性偏移和认知退化。AI 视觉模因引发了创作门槛的系统性消解，使人类从“创作者”变为“提示词输入者”[6]。当人们习惯于用 AI 快速生成模因，其独立思考和认知能力会因“用进废退”而退化，这种退化不仅是技能的退化，更是创作主体性的隐蔽性让渡。在小红书 AI 创作帖子中，大量用户更关注“热门提示词模板”而非原创表达。一些创作者通过复制他人的提示词即可快速获得相似作品，而对内容本身的文化意义和表达价值关注有限。批量同质化、浅层化 AI 视觉模因持续涌入网络内容生态，长期接触简易化、情绪单一的视觉内容易造成受众认知浅层固化。同时，大批量标准化生成内容持续占据流量分发池，挤压具备独特文化表达、完整创意设计的原创视觉内容曝光机会，形成网络内容生态同质化的“马太效应”。

4.2. 文化多样性侵蚀：视觉同质化与语义窄化

一直以来，模因都被看作是文化进化的桥梁，是社会文化进化的驱动力[7]，然而，在生成式 AI 参与模因生产的背景下，这一观点的解释力似乎在逐渐下降。模因是一种社会文化信息体，包含着丰富的社会文化信息和文化价值[8]，但 AI 视觉模因包含的文化多样性正在被侵蚀。首先，AI 视觉模因不仅将原本模因所承载的多元文化内涵简化成单一视觉符号，使文化符号陷入“空洞”[9] [10]。其次，当前的图像生成大模型虽然能模仿不同的视觉风格，但是整体上仍偏向于某种可辨识算法驱动的“数字美学”——高饱和度、标签化、精致化、“光滑感”[11]，一种“AI 脸”式的视觉同质化变得不可避免，如“A4 腰”的动漫美女、充满红色元素的“中国家庭”梗图等。

其次，更隐蔽的是语义同质化。传统模因的活力在于它在文化竞争中“盗猎”出不同于主流文化的话语[12]，而生成式 AI 的内在价值观偏向和技术限制，使其在模因生成阶段就系统化过滤了非主流的表达。不仅如此，算法推荐机制会进一步固化这种语义同质化：算法奖励会引发强烈即时情绪的模因，而这些情绪化的语义表达往往以牺牲意义的复杂性为代价。当模因被持续编码为二元对立的符号框架，其承载的丰富文化信息则会变为单一的情绪对抗。

4.3. 伦理困境：信息失序与责任归属模糊化

AI 视觉模因的无限繁殖和“自生性狂欢”与互联网的复杂性叠加，网络空间充斥鲍德里亚提出的超

真实拟像内容，真实影像与 AI 合成视觉素材边界模糊，受众难以直观分辨内容真伪，进而诱发网络信息辨别失序的结构性危机，即“验证性流行病”[13]。AI 视觉模因在算法推荐下实现病毒式扩散，并且由于当前 AI 审核检测技术的局限，使“验证性就行病”的扩散速度远超事实核查机制的响应能力。更深远的影响在于这种“验证性流行病”会不断瓦解人们对社会的信任，陷入“眼见未必为实”的困境。社交媒体中的部分用户会将 AI 视频、AI 音乐误认为真实音像资料，并在评论区围绕其真实性展开讨论。AI 视觉模因已经不仅仅是娱乐文化现象，其传播过程也可能影响公众的信息判断。

责任归属的不确定性是当前生成式 AI 参与内容生产的一大挑战。布鲁诺·拉图尔指出，技术行动者网络中代理的分布使简单的因果归责模型失效。AI 视觉模因的生产过程则涉及提示词输入者、模型开发者、训练数据提供者以及平台算法等多个行动主体。内容一旦产生侵权、歧视或误导性影响，很难依据传统责任体系进行明确归责。版权问题同样突出——著作权制度尚未明确 AI 生成内容归属，而 AI 视觉模因既非数据简单拼贴，亦非纯粹人类意图表达，而是概率机制下两者的结合，现有法律框架难以有效回应。

5. 生成式 AI 参与网络模因生产的治理路径

5.1. 技术：生成式 AI 和社交平台的共治责任

作为 AI 视觉模因生产的重要基础设施，生成式人工智能开发主体和社交媒体平台应承担重要治理责任。AI 模因的非线性渗透使得模因能够在短时间衍生大量差异化分支，多轮二次改编会逐步覆盖、清除内置水印与元数据其溯源线路极易被切断；同时，现有人工复核、机器识别速度跟不上变异速度，平台审核和监管呈现“滞后性”。针对这一问题，首先要提升 AI 内容的可追溯性和透明度。针对 AI 视觉模因所产生的真实性危机和版权争议，应该从预训练模型、上下文学习、反馈强化学习等不同阶段嵌入内容溯源规则[14]，从根本上提高 AIGC 的可解释性。如构建人工智能版权侵权避风港规则[15]、构建人工智能内容溯源的内容标识模块、动态溯源模块、隐私保护模块等[16]。针对所产生的认知和价值观问题，可以进一步深化生成式人工智能大模型的价值观对齐，综合评判模型行为的伦理合规性与道德准则，减少有害输出，如例如 Anthropic 团队的“有用和无害的人类偏好数据集”[17]、西北民族大学的指令微调“VDA 价值观对齐数据集”[18]。这些规则和数据对齐实践为缓解生成式人工智能所制造的挑战提供了可行路径。

对于社交媒体平台而言，其治理重点在于传播环节。平台应建立更加精细化的 AI 内容标识制度，避免简单采用“AI 生成”与“非 AI 生成”的二元划分，而应根据生成程度、修改比例和传播场景进行分类标识。同时，平台还应优化推荐算法，适当降低低质量重复内容的曝光权重，提高原创性和文化价值较高内容的传播机会，从而缓解算法驱动下的内容趋同现象。例如 TikTok 推出 AI 生成内容调节功能，用户将可以通过自主控制“For You”推荐流中看到 AI 生成内容的比例。无论是生成式人工智能还是社交媒体，技术拥有方都应该积极承担社会责任，使技术能够更好地服务于人类。

5.2. 用户：智能传播下的媒介素养升级

当下多数 AI 视觉模因以娱乐、恶搞、戏仿为核心，大量使用者随意复制网络提示词模板，恶意 AI 伪造历史画面、公众人物恶搞、弱势群体污名化模因；大量用户将 AI 恶搞、虚假合成模因当作真实影像转发，无意识助推虚假信息、歧视内容扩散。在智能传播时代，面对新的人机关系，人类的媒介素养需要从单一“信息辨别”升级为包含 AI 认知、伦理意识及人机协同能力的复合素养。一方面，人们要对 AI 建立正确的认知。明确 AI“概率生成”的本质，对 AI 进行批判性应用，坚持工具理性和价值理性的有机结合。另一方面，人们也应提高人机互动、人机协同的能力，产出更具文化价值的 AI 视觉模因。这

种人机协同能力包括为机器提供更好资源的能力、指挥和调教能力、辨识和判断力、想象力[19]。最后，用户需意识到自身在与 AI 互动中的道德责任，提示词的设计本身即是一种带有伦理含义的行为，即便最终内容由 AI 产出，提示者的意图也应被纳入伦理审视的范畴。用户既是 AI 视觉模因的消费者，也是转发者和共同创作者，这一素养的目标，不是对技术的防御，而是在人机共生的文化环境下，重建理性审视与创造性表达的自觉。

5.3. 政府：规制治理和创新包容

作为拥有法定权威的核心行动者，政府主要通过立法、司法与行政手段进行边界工作，其核心张力在于如何在维护秩序与鼓励创新之间寻求平衡[10]。在立法上，明确 AIGC 内容的责任主体划分，尝试启动 AIGC 著作权停止侵害请求权限制机制和构建责任保险机制[20]。另一方面，倡导分类分级管理和探索“创意豁免”机制，为那些具有创造潜力的 AI 视觉模因保留合法空间[21]。如刘祖兵提到的“场景化分类分级审查制度”[22]，差异化的监管避免“一刀切”对创新的限制。在监管上，监管上，应构建 AI 内容从生产到传播的全链条监管。国家已提供了监管的框架，但要谨慎划定监管的“边界”，警惕过度监管给模因的创新带来负面效应。只有实现刚性规制和弹性空间的动态平衡，政府才能在治理 AI 视觉模因生产中的伦理失范与信息失序的同时，为技术赋能文化创新保留一定的制度空间。

6. 结论

生成式人工智能深度参与网络内容和文化的生产，人机关系也从“人机协作”走向“人机协同”。未来随着人工智能技术持续演进，人机关系将进一步发展，网络模因的生产与传播也将呈现新的发展形态。如何在技术创新与文化价值之间实现协调发展，仍将是智能传播研究的重要议题。

参考文献

[1] Dawkins, R. and Davis, N. (2017) The Selfish Gene. Macat Library.

[2] 张洪忠, 任吴炯. 大模型对互联网生态影响及其发展趋势[J]. 中国网信, 2023(6): 37-41.

[3] 任吴炯. 文化的进化: AI 大模型应用的模因演变分析[J]. 中国编辑, 2025(4): 43-49.

[4] 肖志恒, 王永祥. 模因与文化[J]. 山西师大学报(社会科学版), 2013, 40(S1): 50-51.

[5] 苏凡博, 王鑫铁, 朱慧, 等. 规律共识网络中的平台干预效应研究——基于模因传播的建模与仿真分析[J]. 新闻与传播研究, 2025, 32(4): 18-34+126.

[6] 卜芾津. 生成式人工智能驱动的模因演化、机制变革与伦理反思——基于“AI 山海经”案例的考察[J]. 天府新论, 2026(2): 40-48+154.

[7] 何自然, 何雪林. 模因论与社会语用[J]. 现代外语, 2003(2): 200-209.

[8] 王天华, 杨宏. 模因论对社会文化进化的解释力[J]. 哈尔滨工业大学学报(社会科学版), 2006(6): 128-130.

[9] 张诗娟, 漆捷. 遮蔽与解蔽: 符号消费语境下文化主体性审思[J]. 理论导刊, 2026(2): 108-117.

[10] 林丽青. 边界竞技场: AI 魔改视频中的文化失序、法律博弈与协同治理范式建构[J]. 科技传播, 2026, 18(6): 128-134.

[11] 殷乐, 申哲. 算法社会的数字美学与青年的自我抵抗[J]. 青年记者, 2024(4): 102-107.

[12] [美]约翰·费斯克. 理解大众文化[M]. 王晓珏, 宋伟杰, 译. 北京: 中央编译出版社, 2001.

[13] 胡思洋, 陈功. 智能传播视域下网络信息失序的结构性危机及其发生机制和规制理路[J]. 东南传播, 2026(2): 154-157.

[14] 张熙, 杨小汕, 徐常胜. ChatGPT 及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(5): 743-750.

[15] 黄玉烨, 杨依楠. 论生成式人工智能版权侵权“双阶”避风港规则的构建[J]. 知识产权, 2024(11): 37-58.

[16] 李文媛, 王杨, 李德有. 基于区块链的 AI 生成内容溯源框架设计研究[J/OL]. 微电子学与计算机: 1-12. https://link.cnki.net/urlid/61.1123.tn.20251119.1558.002, 2026-06-05.

[17] Bai, Y.T., Jones, A., Ndousse, K., *et al*. (2022) Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. https://www.researchgate.net/publication/359918097_Training_a_Helpful_and_Harmless_Assistant_with_Reinforcement_Learning_from_Human_Feedback

[18] 李英杰, 严琦栋, 吴文社, 等. VAD: 面向大语言模型指令微调的价值观对齐数据集[J]. 中国科学数据, 2026, 11(1): 181-190.

[19] 彭兰. 智能素养: 智能传播时代媒介素养的升级方向[J]. 山西大学学报(哲学社会科学版), 2023, 46(5): 101-109.

[20] 刘云开. 人工智能生成内容的著作权侵权风险与侵权责任分配[J]. 西安交通大学学报(社会科学版), 2024, 44(6): 166-177.

[21] 于文. AI“魔改”怎么看[J]. 中国报业, 2025(10): 5.

[22] 刘祖兵. 论 Sora 的伦理风险与我国治理因应——也谈我国参与人工智能算法伦理全球治理的基本路径[J]. 河海大学学报(哲学社会科学版), 2024, 26(5): 99-112.