生成式人工智能治理行动框架：基于AIGC事故报道文本的内容分析

Generative Artificial Intelligence Governance Action Framework: Content Analysis Based on AIGC Incident Report Texts

朱禹[1] 陈关泽[2] 陆泳溶[3] 樊伟[4]
ZHU Yu CHEN Guanze LU Yongrong FAN Wei

（1.南京大学信息管理学院，南京，210023；2.香港中文大学计算机科学与工程学系，香港，999077；3.四川大学匹兹堡学院，成都，610207；4.四川大学图书馆，成都，610065 / 1. School of Information Management, Nanjing University, Nanjing, 210023; 2. Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, 999077; 3. Pittsburgh Institute, Sichuan University, Chengdu, 610207; 4. Sichuan University Library, Chengdu, 610065）

摘要：【目的 / 意义】生成式人工智能（Generative AI）的突破性进展带来了人工智能生成内容（AIGC）的爆炸式增长，不可避免地将人们置于信息过载、信息噪声、信息安全等的负面影响之下，使得社会信息治理面临新的挑战。分析和探讨现有 AIGC 事故的特征属性，对我国生成式人工智能治理有参考借鉴作用。**【研究设计 / 方法】**基于 AI 事故数据库（AIID），以 AIGC 相关事故报道为样本进行内容分析，探析现有 AIGC 事故的类型、原因、损害对象和应对措施。**【结论 / 发现】**AIGC 事故影响客体的多元性、波及范围的广泛性、潜在危害的复杂未知性，导致任何单一行动主体的资源和能力都无法有效应对危机，需要政府、企业、社会三方行动主体形成“多元 + 协调 + 制衡”的治理参与模式，并在“情境 - 意识 - 行动”的行动框架下开展信息治理。**【创新 / 价值】**引入了 AIID 作为案例来源数据库，提供了关于现有 AIGC 事故相关细节的直观论证，并通过内容分析形成了 AIGC 事故分析三级类目框架。构建的生成式人工智能治理行动框架有助于从宏观视角促进我国生成式人工智能治理的探索和实践。

关键词：生成式人工智能；人工智能生成内容；信息治理；行动框架；内容分析

中图分类号：G203 **DOI：**10.13366/j.dik.2023.04.041

引用本文：朱禹，陈关泽，陆泳溶，等．生成式人工智能治理行动框架：基于 AIGC 事故报道文本的内容分析 [J]. 图书情报知识，2023，40（4）：41-51.（Zhu Yu, Chen Guanze, Lu Yongrong, et al. Generative Artificial Intelligence Governance Action Framework:Content Analysis Based on AIGC Incident Report Texts[J]. Documentation, Information & Knowledge, 2023,40（4）：41-51.）

Abstract: [Purpose/Significance] The breakthrough of Generative Artificial Intelligence（Generative AI）has led to the explosive growth of Artificial Intelligence Generated Content（AIGC）, which inevitably cause people to be negatively affected by information overload, information noise, information security, and other related issues, making social information governance face new challenges. This paper aims to analyze and discuss the characteristic attributes of AIGC incidents, so as to provide a reference for Generative AI governance in China. **[Design/Methodology]** Based on AI Incident Database（AIID）and taking AIGC-related incident reports as samples for content analysis, the types, causes, damage objects and countermeasures of existing AIGC incidents were discussed. **[Findings/Conclusion]** The diversity of the objects affected by AIGC incidents, the wide distribution of the scope, and the complex unknown of the potential harm, result in the resources and capabilities of any single actor not being able to effectively deal with the crisis. It is necessary for actors of government, enterprises and society to form a governance participation model of "diversity + coordination + checks and balances", and to carry out information governance under the action framework of "context-consciousness-action". **[Originality/Value]** This paper introduces AIID as a case source database, provides an intuitive demonstration of the relevant details of existing AIGC incidents and forms a three-level category framework for AIGC incident analysis through content analysis.The action frame of Generative AI governance formed in this study is helpful to promote the exploration and practice of Generative AI governance from a macro perspective.

Keywords: Generative artificial intelligence; Artificial Intelligence Generated Content（AIGC）; Information governance; Action framework; Content analysis

1 引言

随着人工智能算法、算力和算据[1]的不断发展，生成式人工智能（generative artificial intelligence）取得突破性进展，使得人工智能生成内容（AI Generated Content, AIGC）得以实现。以GPT4为代表的自然语言大模型展现出强大的自然语言加工、荟萃、整合和生成能力[2]，能利用自身强大的自然语言处理能力生成自然流畅的文本。同时，以生成式对抗网络（Generative Adversarial Networks, GANs）模型、Diffusion扩散化模型为代表的生成式人工智能在图像、音频、视频领域表现出优异的内容合成、修复、预测和生成性能[3]，引发了当前网络信息环境下AIGC的爆炸式增长。尽管这一发展趋势为传媒、电商、影视、娱乐、教育和工业设计

[通讯作者] 朱禹（ORCID:0000-0002-2548-828X），硕士研究生，研究方向：信息资源管理，Email:13350057427@163.com.（Correspondence should be addressed to ZHU Yu, Email: 13350057427@163.com, ORCID:0000-0002-2548-828X）

[作者简介] 陈关泽（ORCID:0009-0008-6974-4864），硕士研究生，研究方向：信息资源管理、可解释人工智能，Email:1625205886@qq.com；陆泳溶（ORCID:0009-0004-7629-3281），本科生，研究方向：人工智能，Email:rita111585@icloud.com；樊伟（ORCID:0009-0002-6278-910X），硕士，馆员，研究方向：信息资源管理，Email:535723984@qq.com。

等行业带来了极大的效率提升，但也不可避免地将人类置于信息过载、信息噪声、信息安全等的负面影响之下，使得社会信息治理面临新的挑战。

当前，各国政府、企业、行业机构正努力就人工智能的健康可持续发展制定指南或规划。例如，我国国务院2017年印发的《新一代人工智能发展规划》[4]指出要加强人工智能相关法律、伦理和社会问题研究，建立保障人工智能健康发展的法律法规、伦理道德框架与人工智能安全监管和评估体系；欧盟2021年发布的《人工智能法案》草案提出了AI风险预防的保障机制和实施路径[5]，但欧洲议会对该法案所规定的AI监管基本原则仍存在较大争议[6]；2023年5月，OpenAI公司CEO Sam Altman在美国国会的人工智能监管听证会上表示需要建立一个新的立法和监管体系以应对AI的潜在风险[7]。随着ChatGPT发布，生成式人工智能治理成为关注热点。2023年4月，我国国家互联网信息办公室发布了《生成式人工智能服务管理办法（征求意见稿）》，旨在促进生成式人工智能技术的健康发展和规范应用[8]；同年5月，法国国家信息自由委员会（CNIL）发布了尊重个人隐私的人工智能治理行动计划，旨在促进和指导隐私友好型的生成式人工智能发展[9]。但是，研究者普遍认为AI治理尚处在起步阶段，很少有明确的法规、条例和标准可以用来指导AI技术的有序开发与利用[10]。

目前，AI治理指南多是从AIGC的上位概念，即从AI的全局应用角度出发来制定，存在过于理想化、脱离实际问题的情况[11]。与此同时，伴随生成式人工智能而产生的虚假新闻[12]、隐私侵犯[13]等负面影响却是实际已经发生的，并给社会信息治理带来了日益严重的威胁。为此，本文对人工智能事故数据库（AI Incident Database, AIID）提供的全球AIGC事故案例[14]进行全面的内容分析，总结分析AIGC事故的特征及构成机制，以期为我国生成式人工智能治理提供方法参考与借鉴。

2 相关研究

生成式人工智能是指基于算法、模型和规则生成文本、图片、声音、视频、代码等内容的技术[8]。国内学术界对生成式人工智能潜在风险的研究主要集中在民商法和出版领域，从法律性质、权利归属等层面讨论了生成式人工智能对传统著作权与版权认知的冲击[15-17]。2022年底，相关研究得到了公共管理、信息资源管理学界的关注，开始系统介绍生成式人工智能的演进历史、应用场景和潜在问题。例如，李白杨等[18]重点梳理了AIGC发展的基础条件，认为AIGC可能在深度伪造、知识产权和网络舆论引导等方面存在的合规安全问题；蒋华林[19]认为生成式人工智能推动了知识生产方式和学术研究范式变革，将影响科研成果创新性评价、科研成果权属认定，带来学术伦理风险，冲击人才评价基本标准，影响人才评价客观性；蒲清平和向往[20]认为生成式人工智能的潜在问题分为法律危害、思想危害、社会危害三类，具体表现在知识产权、信息窃取、诈骗、意识形态安全、独立思考能力、价值观塑造、劳动者失业和极端事件发生等方面；陆小华[21]分析指出生成式人工智能会对一个国家和民族文化传承、认知形成以及意识形态产生影响，并且这种影响目前还难以准确估量；中国信息通信研究院[22]白皮书指出，当前AIGC在关键技术、企业核心能力和相关法律法规等方面还不够完善，导致社会公平、责任、安全等争议日益增多。

从国外研究来看，Khoo等[23]认为生成式人工智能的深度伪造（deepfake）在视觉质量、多样性和逼真度方面取得了巨大进步，但同时也在色情视频和政治方面产生了更多难以识别的虚假信息[24]，为知识产权保护带来了更大的挑战。Partadiredja等[25]通过问答游戏的调查，发现人们难以区分AI生成的内容与人工制作的内容，引发了民众对信息伦理的担忧。欧洲知识产权局在报告中指出“新技术（AI）对知识产权而言是一把双刃剑”，生成式人工智能工具被用于恶意文档抓取、规避图像和语音识别、恶意信息污染等犯罪行为，其滥用助长了网络信息环境的深度合成和隐私侵犯[26]。当前，生成式人工智能的局限性包括：对大量的高质量训练数据的依赖性、无法处理长尾问题、通用性有限、对特定应用场景的依赖性和人工智能开发者的固有偏见。随着人工智能越来越多地收集和处理我们身边的数据，很有可能会对社会基本人权产生深远的影响。

综上所述，国内外研究者和机构都意识到生成式人工智能潜在的社会安全威胁，并在积极探索其信息治理方案。随着《生成式人工智能服务管理办法（征求意见稿）》《网络信息内容生态治理规定》《网络数据安全管理条例（征求意见稿）》《互联网信息服务深度

合成管理规定（征求意见稿）》等政策法规的相继出台，我国正积极开展生成式人工智能的治理实践，但对生成式人工智能服务责任划归、作品权属法律认定、企业社会责任承担等风险治理的认识仍处在初期阶段。

3 研究设计

3.1 数据来源

本文主要采用内容分析法，对近年来世界各地报道的AIGC事故进行案例研究，样本数据来自AI事故数据库（AIID）[14]。对于特定事故，AIID关联了与之相关的多渠道、多时间、多角度的报道文章，旨在帮助研究者从现有的AI事故中学习，以便研究者能够预防或减轻不良后果。数据库中记录并索引了由用户提交的现实中应用人工智能所造成的危害性事件信息，且事件信息经过了专门委员会的讨论和审查，从而保证了数据库样本的质量。

3.2 选取原则

本文选取了截止到2023年2月28日前AIID记录的472个事故数据。通过人工阅读样本标题和内容，筛选和标注与AI生成文本、图像、视频和音频相关的样本共85个，剔除非负面报道后保留61个有效样本，共载有360篇报道文章①。

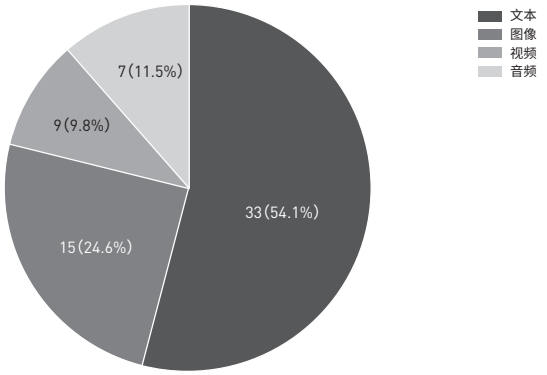

图1 AIGC事故内容类型
Fig.1 The Content Types of AIGC Incidents

3.3 样本描述性统计

如图1所示，AI生成内容的主要类型是文本信息，占比54.1%；其次是图像，占比24.6%；视频和声音占10%左右。

如图2所示，2022年发生的AIGC事故最多（N=23，37.7%），而2015年、2016年、2019年和2021年四个年份的事故较少（少于3个）。AIGC事故报道数量在时间上呈现出较明显的波动，数量在2017年和2022年出现波峰，分别为105个和103个。出现这一现象的主要原因是由于2017年Transformer模型的提出和2022年人工智能对话机器人ChatGPT的发布。革命性技术的出现与应用导致了社会对其关注的增多和现实问题的暴露。

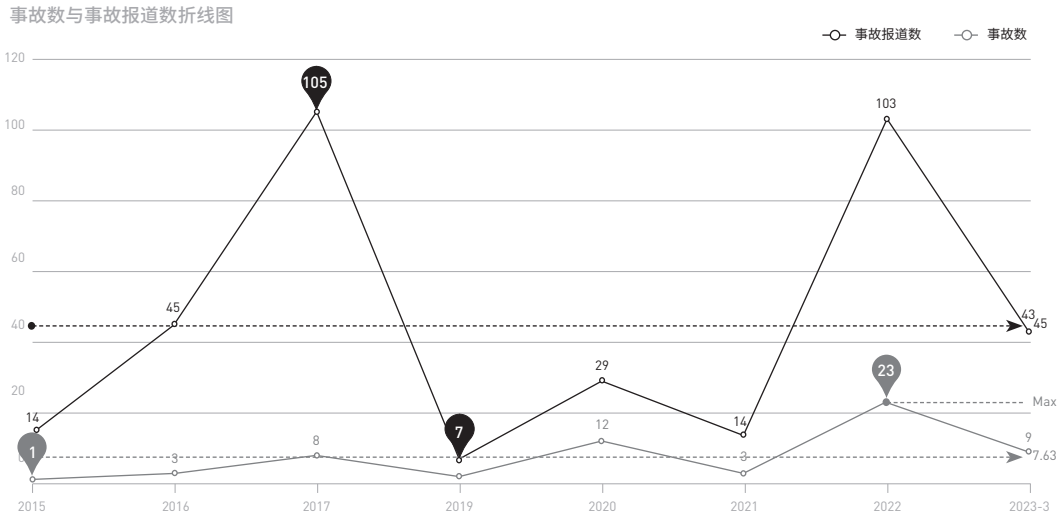

图 2 事故数与事故报道数时间趋势
Fig.2 Time Trends of the Number of Incidents and the Number of Reported Incidents

① 数据已上传至百度网盘。下载链接：https://pan.baidu.com/s/1IgdXBGUkgkqOwXGbb9iVJQ?pwd=1111。

如图3所示，AIGC事故的报道主要集中在美国（N=249）和中国（N=50），两者之和达总报道数的84%。美国和中国是大多数人工智能公司所在的国家，如谷歌（美国）、微软（美国）和百度（中国）。对报道的国家或地区分布进行统计后发现，AIGC事故是全球性的，并不局限于某个特定国家或地区，这意味着世界各地均或多或少地面临生成式人工智能引发的社会风险。

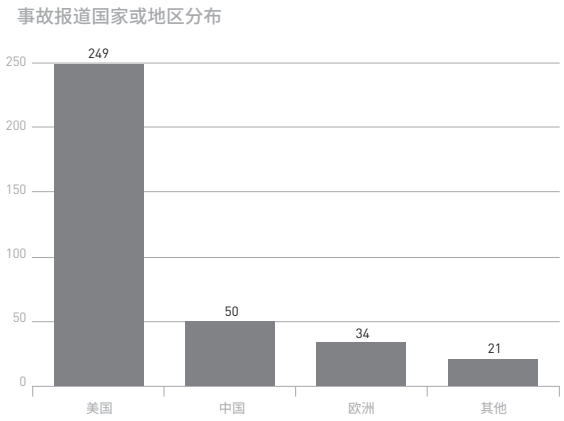

图3 事故报道数国家或地区分布

Fig.3 The Number of Incidents Reported Distributed by Country or Region

3.4 编码说明

本文选取了AIGC事故新闻报道的全文本作为编码材料的范围，并使用内容分析方法对其进行分析[27]。需要特别说明的是，由于同一事故关联了多渠道、多时间、多角度的报道，因此最终被编码的事故数量将大于样本数量。鉴于AIGC事故类型及属性的特殊性，本文没有使用AIID所提供的CSET（The Center for Security and Emerging Technology）分类法，而是不加预设地开展编码工作，逐步形成编码项目的框架。该过程由三位编码员分别对61例AIGC事故进行独立编码，以定期开会的形式讨论相关概念和编码要求，并细化编码框架。具体来说，第一轮编码选择了总样本的15%，即对9个案例进行开放式预编码，之后利用卡片分类法对内涵相近的节点向上归纳形成上一级类目，逐步形成编码框架。后续又经历了两轮编码和讨论，最终形成AIGC事故分析三级类目框架，由4个一级类目、9个二级类目、34个三级类目和296个参考点组成（如表1所示）。

表1 编码框架及结果

Table 1 Coding Framework and Results

一级类目	二级类目	三级类目	节点数	三级类目占一级类目比例
事故类型	恶意信息	色情骚扰	7	9.59%
		敲诈勒索	13	17.81%
		暴力犯罪	7	9.59%
		侵权	15	20.55%
		歧视	18	24.66%
	虚假信息	事实错误	13	17.81%
事故原因	客观	法律不完善	7	9.21%
		信息源质量低	12	15.79%
		技术不成熟	21	27.63%
	主观	恶意算法	6	7.89%
		恶意使用	29	38.16%
		违法收集数据	1	1.32%
损害对象	集体	一般（无特定属性）	4	8.16%
		企业	7	14.29%
		民族	11	22.45%
		政党	1	2.04%
		国家	1	2.04%
	个人	一般（无特定属性）	1	2.04%
		特定年龄	3	6.12%
		特定职业	12	24.49%
		特定性别	9	18.37%
应对措施	政府	修订法律	2	2.04%
		加强监管	12	12.25%
		惩罚措施	1	1.02%
	企业	主动下线	10	10.20%
		倾听用户反馈	8	8.16%
		控制用户行为	17	17.35%
		开源审查	3	3.06%
		优化算法模型	17	17.35%
		优化信息源	11	11.23%
		加强可用测试	5	5.10%
	个人	提起法律诉讼	5	5.10%
		积极反馈问题	1	1.02%
		提高信息素养	6	6.12%

4 事故内容分析

4.1 事故类型

按照事故中生成内容的危害程度，得出恶意信息事

故（N=60，82.2%）和虚假信息事故（N=13，17.8%）两类事故类型。数量上，以危害较严重的恶意事故为主。

4.1.1 恶意信息事故

恶意信息事故指旨在破坏网络信息环境，或者故意以非法手段损害第三方权益的信息事故。恶意信息事故是一种强危害性的事故类型，通常表现为色情骚扰、敲诈勒索、暴力犯罪、侵权及各类歧视。例如，样本库中一位儿童要求美国亚马逊公司的智能家居语音助理Alexa播放其喜欢的歌曲时，Alexa播放了一个色情电台，并说出了许多色情术语，这对缺乏判断力的儿童造成了错误引导和身心损害①；又如一例事故中，用户绕过了ChatGPT预先设置的安全过滤机制，让ChatGPT提供了偷车技巧和制作炸弹的方法，这可能会严重危害社会公众的人身和财产安全②。

4.1.2 虚假信息事故

虚假信息事故是指网络信息生产者受某种动机驱使，通过夸大、增加错误内容或淡化、分散、隐藏重要事实等手段，有目的地在网络环境中散布与客观事实不相符合的信息而导致的信息事故[28]。与恶意信息事故相比，虚假信息事故存在危害隐秘性、随机偶发性的特征。其包括冒充真人用户发帖③、组合存在偏见或色情的图像④、伪造国家元首发言视频⑤、撰写不准确的新闻报道⑥等，不一而足。

4.2 事故原因

总结事故原因，可以将其按主观原因（N=36，47.4%）和客观原因（N=40，52.6%）划分，由主客观条件导致的事故原因分布大体均衡。

4.2.1 主观原因

导致AIGC事故发生的原因众多，主观的恶意使用最甚。尤其是伴随着元宇宙的到来，人们可以利用人工智能技术生成各种虚拟形象或数字身份，不法分子可以盗用或冒充他人身份且难以被识别和预防。具体表现为以下三个方面。

（1）恶意使用（N=29，38.2%）。例如，一伙骗子恶意使用AI生成真人照片建立虚构的波士顿律师事务所，从而欺骗客户并扰乱市场秩序，给民众造成了经济损失和人格侵犯等损害⑦。此外，生成式人工智能技术为网络犯罪降低了门槛，如网络黑客利用ChatGPT辅助编写恶意代码、制作勒索软件来损害计算机系统⑧，这拓展了黑客攻击的技术工具，使得网络攻击呈现出扩大化和低门槛化趋势。上述案例均显示出生成式人工智能滥用存在的潜在风险，深度合成诈骗、色情、诽谤、假冒身份、网络攻击等新型违法犯罪行为正在涌现。

（2）恶意算法（N=6，7.9%）。生成式人工智能正被用于警务和医疗等领域的决策制定过程。但是，一些生成式人工智能算法开发者可能存在天然的不公正立场，算法中可能隐含针对特定种族、性别、职业的歧视或其他恶意观点，这意味着现有的不公正或歧视可能会被放大，从而产生危险的结果。例如，一家医院使用的人工智能算法建议黑人病人比白人病人得到更少的医疗护理；警务机构在利用AI评估罪犯是否可能再次犯罪时，黑人被错误分类的可能性是白人的两倍；亚马逊公司的招聘机器人也存在歧视女性申请者的问题⑨。

（3）违法收集数据（N=1，1.3%）。例如，韩国ScatterLab公司的AI聊天机器人李路达（Lee Luda）因在未经用户同意的情况下违规收集用户聊天记录、使用用户个人信息而面临集体诉讼⑩。

4.2.2 客观原因

（1）技术不成熟（N=21，27.6%）。算法的不成熟导致了AIGC的偏见和不公平性。例如，美国Facebook（Meta）公司研发的AI将视频中的黑人标记为“灵长类动物”而引发黑人民权抗议⑪，这揭示了生成式人工智能技术的不成熟性，可能会导致社会种族矛盾的进一步激化。美国Google公司的AI辅助写作工具由于机器学习算法缺陷，导致种族主义、性别歧视或其它排他性内容被推荐给用户用于文本创作⑫。这与当前生成式人工智能存在的可解释性弱、可泛化边界未知、逻辑推理能力有限等技术缺陷有关，它又可能进一步导致人们对算法信任度的降低。

（2）信息源质量低（N=12，15.8%）。信息源质量低是引发AIGC事故的又一重要因素。亚马逊、腾讯、OpenAI等公司承认当前可供生成式人工智能大模型训练的高质量、易标注的数据集有限，由于固有缺陷、加工粗糙、审核不严等情况而导致训练数据存在偏见、色情、暴力、不均衡等质量问题。

① 参见事故55，https://incidentdatabase.ai/cite/55/
② 参见事故420，https://incidentdatabase.ai/cite/420/
③ 参见事故120，https://incidentdatabase.ai/cite/120/
④ 参见事故179，https://incidentdatabase.ai/cite/179/
⑤ 参见事故201，https://incidentdatabase.ai/cite/201/
⑥ 参见事故455，https://incidentdatabase.ai/cite/455/
⑦ 参见事故236，https://incidentdatabase.ai/cite/236/
⑧ 参见事故443，https://incidentdatabase.ai/cite/443/
⑨ 参见事故420，https://incidentdatabase.ai/cite/420/
⑩ 参见事故106，https://incidentdatabase.ai/cite/106/
⑪ 参见事故113，https://incidentdatabase.ai/cite/113/
⑫ 参见事故177，https://incidentdatabase.ai/cite/177/

（3）法律不完善（N=7，9.2%）。由于相应法律的缺失、监管的缺位难以在当前技术环境下保护个人信息的自决权，小部分公司可以不受限制地恶意使用AI工具进而产生个人信息泄露风险，使得公民的个人信息权利受到侵犯。尽管本文研究样本中这部分事故原因比例较小，但也需引起重视。

4.3 损害对象

本文将损害对象分为集体和个人两大类。具有不同性质或特点的个人和集体分布总比例相近，受害面广且并未呈现出明显的倾向特征。

4.3.1 集体

集体层面，按照发生频率分别是民族（N=11，22.4%）、企业（N=7，14.3%）、一般集体（N=4，8.2%）、政党（N=1，2%）、国家（N=1，2%）。特定民族和企业是受事故损害的主要集体，如警察使用的AI模型高估了黑人、西班牙裔和亚裔人的罪行风险，造成了执法层面的种族歧视[①]；犯罪分子使用人工智能语音生成技术骗取了阿拉伯联合酋长国银行3500万美元，给企业造成巨额经济损失[②]。其次，社会性AIGC事故对整个社会公信力也会造成严重损害，如利用深度伪造技术影响美国选举舆论走势，从而引发美国民众对国家民主选举、政策制定，乃至社会完整性的严重质疑[③]；中国图灵公司的BabyQ和微软公司的小冰两款聊天对话机器人生成了含有不符合社会主义核心价值观、否定社会主义制度、不客观表达喜恶的内容[④]，其产生的虚假信息容易对用户造成错误价值观、政治观的引导，可能扰乱经济和社会秩序。此外，生成式人工智能还可能被用于国家及网络舆论战争，如俄罗斯一个团队利用人工智能生成了虚假账户、虚构人物和品牌并顺利通过了西方社交媒体审查，发布大量关于西方背叛乌克兰和乌克兰是一个失败国家的说法和所谓证据[⑤]。生成式人工智能技术成为国家间散布虚假信息、进行舆论战争的另一种手段。

4.3.2 个人

个人层面，特定职业往往容易遭受侵害（N=12，24.49%），包括记者、学生、病人、艺术家、明星、科研人员、程序员、政治家等。例如，AI聊天机器人直接剽窃人类作家的作品来进行创作[⑥]；艺术家的作品未经授权便被用于训练AI模型，造成了作品知识产权侵权的风险[⑦]；AI故意生成的美国前总统奥巴马图片呈现较低的分辨率，旨在对其进行贬低或抹黑[⑧]；GitHub Copilot在未经用户授权的情况下，直接收集和使用GitHub中的开源代码，这一行为引发了部分用户的抗议[⑨]。此外，事故损害的对象还针对特定的性别（N=9，18.37%），例如，AI模型会生成带有偏见或性化的女性图片，向AI输入用户的童年照片后却生成了女性用户裸体图片[⑩]。

4.4 应对措施

按照采取应对措施的行动参与方进行分类，得到企业（N=71，72.5%），政府（N=15，15.3%），个人（N=12，12.2%）三类主体。

4.4.1 企业措施

企业采取的应对措施较为多样，最主要的措施为优化算法模型（N=17，17.3%）和控制用户行为（N=17，17.3%）。例如，Meta正在开发一项AIGC检测技术，旨在为其产品BlenderBot3创建安全措施，以解决模型生成的偏见或冒犯性言论问题[⑪]；OpenAI通过限制用户与AI模型交互的频率（限制用户每分钟提交的API请求数量）来减少模型的大规模滥用[⑫]。其次，优化信息源（N=11，11.2%）和主动下线（N=10，10.2%）也是企业常见的应对措施。例如，OpenAI通过过滤模型训练集中的色情及暴力图片来训练DALL•E 2.2；腾讯公司在其发布的一款聊天机器人发表了不当言论后，立即下线该聊天机器人并删除了其所有数据[⑬]，以主动应对问题。此外，企业也通过倾听用户反馈、进行严格的可用性测试、开源审查的方式来应对事故。比如，Microsoft、OpenAI和Meta会根据用户反馈来有针对性地改进算法模型[⑭]；OpenAI在网络上公开其分类器算法代码以寻求更多有用反馈来改进算法[⑮]。

4.4.2 政府措施

政府往往采取行政手段，用加强监管（N=12，12.2%）的方式来应对事故。例如，韩国数据保护监管机构对一家AI初创公司处以巨额罚款，原因是该公司

① 参见事故154，https://incidentdatabase.ai/cite/154/
② 参见事故147，https://incidentdatabase.ai/cite/147/
③ 参见事故201，https://incidentdatabase.ai/cite/201/
④ 参见事故66，https://incidentdatabase.ai/cite/66/
⑤ 参见事故205，https://incidentdatabase.ai/cite/205/
⑥ 参见事故457，https://incidentdatabase.ai/cite/457/
⑦ 参见事故421，https://incidentdatabase.ai/cite/421/
⑧ 参见事故165，https://incidentdatabase.ai/cite/165/
⑨ 参见事故240，https://incidentdatabase.ai/cite/240/
⑩ 参见事故423，https://incidentdatabase.ai/cite/423/
⑪ 参见事故278，https://incidentdatabase.ai/cite/278/
⑫ 参见事故179，https://incidentdatabase.ai/cite/179/
⑬ 参见事故66，https://incidentdatabase.ai/cite/66/
⑭ 参见事故477和事故420，https://incidentdatabase.ai/cite/477/，https://incidentdatabase.ai/cite/420/
⑮ 参见事故466，https://incidentdatabase.ai/cite/466/

在开发女性聊天机器人的过程中不慎泄露了大量用户个人信息[①]。仅有少数地方政府对相关法律进行了修改（N=2，2.0%）。例如，美国加州的立法官员通过修订法律来打击非自愿图像的创建行为[②]。

4.4.3 个人措施

面对利益受损，个人主要采取提起法律诉讼（N=5，5.1%）或者提高个人信息素养（N=6，6.1%）的方式来应对。比如，艺术家联名对微软的Stability AI 提起AI艺术生成的版权诉讼[③]；个人在不同的APP时使用不同的密码并在打开陌生链接时保持谨慎，以防止黑客恶意窃取个人信息[④]。此外，报道中还提及用户以积极反馈问题的方式来应对事故。例如，用户主动标记聊天机器人响应的不当内容，以帮助俄罗斯科技公司Yandex发布的聊天机器人Alice避免此类响应[⑤]。用户的积极反馈在改进和优化生成式人工智能产品方面发挥了重要作用。

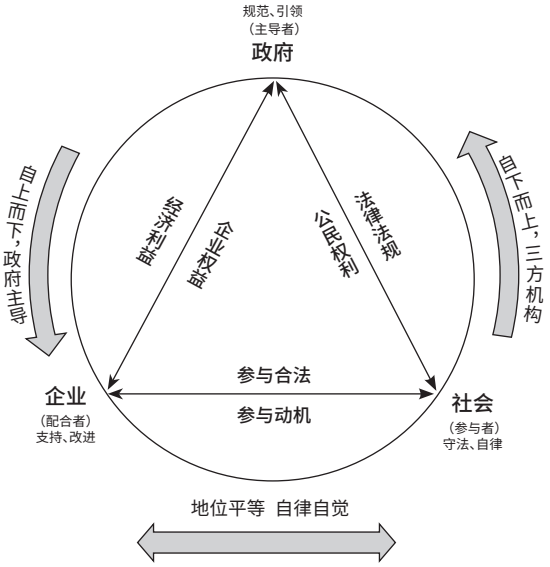

图4 “多元 + 协调 + 制衡”的生成式人工智能治理参与模式
Fig.4 A Participation Model of Generative AI Governance of "Diversity + Coordination + Checks and Balances"

5 AIGC 信息治理行动框架

5.1 “多元 + 协调 + 制衡”的参与模式

上文的分析提供了既有事故中受损害对象的直接证据。虚假信息事故和恶意信息事故的损害对象大到集体，小到个人，包括了国家、政党、民族、企业和特定社会个体。案例事实证明，其影响客体的多元性、波及范围的广泛性、潜在危害的复杂未知性导致任何单一行动主体的资源和能力都无法有效应对危机。因此，必须在政府的统一指挥领导下，与社会组织、企业、群众等行动主体形成协同治理的参与格局[29]。在生成式人工智能治理过程中，多元的行动参与主体在组织性质、利益偏好、目标导向等方面存在差异，同时由于相互间权力、地位和资源的不平等，决定了各行动主体具有不同的行动逻辑。因此，本文将现有样本中的“损害对象”进一步聚类，厘清各方信息治理行动参与逻辑，构建了一个由政府、企业和社会三方构成的“多元+协调+制衡”的生成式人工智能治理参与模式（如图4所示）。

从“多元”维度来看，政府是生成式人工智能治理的主导者，应积极探索制定相应法律法规对生成式人工智能研发者、提供者和使用者行为进行规范，并在重点应用领域进行总体性部署和前瞻性引领，在构筑我国生成式人工智能发展先发优势的同时，促进其健康发展和规范应用。企业是生成式人工智能治理的配合者，包括但不限于单个科技企业和多企业联合的行业协会在内的各主体应积极在技术、人力、资源和管理等方面为政府开展生成式人工智能治理提供强力支持。企业应秉持科技向善理念，负责任地发展和应用生成式人工智能技术，确保生成式人工智能软件、工具、算力和算据资源朝着安全可控、治理有效的方向发展。社会是由一个个公民个体构成的整体，每一个公民都是生成式人工智能治理的参与者，应充分守法、自律，并努力提升个人及全社会的数字素养能力。

从“协调”维度来看，上述三方主体需要在信息治理过程中积极沟通、协同，建立起“自上而下、政府主导”“自下而上、三方机构”“地位平等、自律自觉”的治理协同机制。在该过程中，应当平衡各方利益、诉求，政府和企业间要协调社会经济发展利益和企业合法权益的关系。政府和社会间既要保证监管法规的有效落实，又要保证公民合法、合规地使用生成式人工智能的权利。企业和社会间要协调双方参与治理的动机及合法性。

从“制衡”维度来看，要充分有效地发挥“多元+

① 参见事故106，https://incidentdatabase.ai/cite/106/
② 参见事故480，https://incidentdatabase.ai/cite/480/
③ 参见事故421，https://incidentdatabase.ai/cite/421/
④ 参见事故205，https://incidentdatabase.ai/cite/205/
⑤ 参见事故58，https://incidentdatabase.ai/cite/58/

协调+制衡”治理模式的作用，在管理与治理、惩治与救济、激励与执法之间寻求最佳的平衡[30]。管理与治理的制衡需要确保政府在规范生成式人工智能研发、提供和使用等方面深化“放管服”，制定生成式人工智能市场准入负面清单或其它界限清晰合理的法律法规，不过度管理而致使治理失灵；惩治与救济的制衡是指对违规行为的惩罚和对受害者的救济之间的相互平衡，在建立有效惩治机制的同时设立相应的社会救济机制，为因生成式人工智能引发的损害提供申诉途径或其它补救措施，切实保障受害者的权益；激励与执法的制衡需要政府在通过激励措施鼓励企业和研发者进行创新的同时，保持对违规行为的执法能力和权威。制衡机制的引入可以确保三方参与主体利益和权力的制衡，进而实现生成式人工智能的可持续发展和社会共赢。

5.2 “情境-意识-行动”的行动框架

通过进一步深入分析AIGC事故的类型和原因（即生成式人工智能治理的情境）、事故的受损对象（即生成式人工智能治理的主体与客体）、现有的应对措施（即生成式人工智能治理的意识、行动），并将“多元+协调+制衡”的治理模式融入到生成式人工智能治理的全过程之中，本文基于杨华峰提出的国家安全治理的模型假设[31]构建了“情境-意识-行动”的生成式人工智能治理行动框架（如图5所示）。

其中，“情境”子系统是由政府、企业、社会三方的市场供需情境构成的环境系统，通过各方现实地位决定的供给条件和未来需求勾勒出生成式人工智能治理环境。同时，值得注意的是，由于生成式人工智能风险的复杂性、不确定性，一旦时间、空间、议题等情境发生变化，则需要通过供给侧与需求侧予以反馈和调整。“意识”子系统是三方参与主体对生成式人工智能治理情境的知觉、理解、识别和预测，即充分发挥主观能动性调查、了解市场供需情境后，形成的生成式人工智能治理框架中的主体意识和价值观念。“行动”子系统是指在行动框架中的具体治理安排及治理全过程评价，是对意识的呈现和对情境的反馈。上述三个子系统共同构成了生成式人工智能治理行动框架的循环系统。

“情境-意识-行动”是一种交互构建的关系，由参

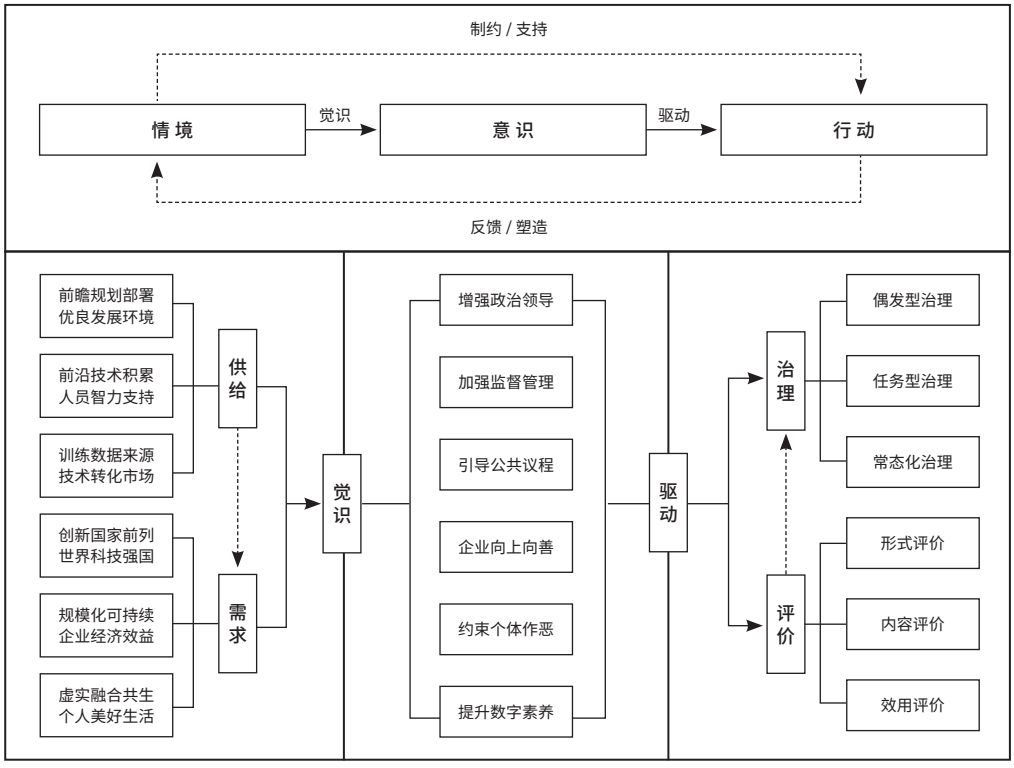

图5 生成式人工智能治理行动框架
Fig.5 Action Framework of Generative AI Governance

与三方供需关系的客观状态构建的“情境”形成了治理中的“意识”观念和理念，并在该“意识”的引导和驱动下制定和实施具体“行动”安排。“行动”在“意识”的引导和规范下，也将反馈和塑造更高水平的生成式人工智能治理情境。实际上，在这种交互构建关系中，意识与情境、行动与意识也存在反馈互动的逻辑链条，只是这种反馈在行动的反馈塑造过程中得以体现，故未重复强调。

“情境”本身是一个生成式人工智能治理的理想预设，包含了治理主体对生成式人工智能健康有序发展的愿景以及治理客体对治理绩效的期待等，是生成式人工智能治理得当、社会信息安全的行动起点。相应地，“情境”也对可能的治理“意识”和“行动”构成制约或支持。因此，“情境”可以被视为意识形成和行动发生的生态系统。“意识”是对生成式人工智能治理的知觉、理解状态，是情境感知与行动安排间的中枢与关键。觉识促使意识的形成，驱动体现意识的能动性。它既是对生成式人工智能治理情境这一客观存在的现实反映，也是认识主体即三方参与者的主观精神状态。意识决定着生成式人工智能治理观念、理念的形成与演变轨迹，并能动地预测、驱动出具体治理行动安排。“行动”是由治理和评价共同构成的具体治理实践。根据治理行动动力模式、运行特质的不同，可以将其分为具有随机性的偶发型治理、问题导向型的任务型治理和日常进行的常态化治理。此外，要注意对生成式人工智能治理质量的评价，在较长的时间、空间和议题维度上对生成式人工智能治理进行综合的评价和分析，以评促治，推动整个行动框架的反馈和再塑造。

以此框架开展生成式人工智能治理，应当形成何为生成式人工智能治理、何谓治理参与主体、有何治理行动方法的治理思路，并由此确定“多元+协调+制衡”参与模式中各主体的位置、权利、义务及其运行逻辑假设等议题。

6 结论与展望

本文通过案例分析，提供了关于现有AIGC事故相关细节的直观论证，有助于生成式人工智能治理的三方参与主体意识到潜在危害及其后果。本文并未在内容分析中直接给出AIGC事故的具体风险点位，而是聚焦事故的本质属性，按危害程度将其二分类为恶意信息事故、虚假信息事故，并按照事件全生命周期将问题的起因、经过、结果拆分并归类为事故原因、损害对象及应对措施三大方面。具体来说，贡献有三：一是引入了AI Incident Database作为案例来源数据库，其事故报道来源主要集中在欧美发达国家及地区，既是以“他山之石”供国内进行生成式人工智能治理参考，还有助于为构建符合中国国情的AI伦理数据库（AIE Database）提供参考借鉴；二是形成AIGC事故分析三级类目框架，以备后续AIGC事故分析参酌；三是简要探讨了一个包含“多元+协调+制衡”参与模式的“情境-意识-行动”生成式人工智能治理行动框架。

本文也存在样本局限、方法局限和理论构建局限。首先，案例分析主要基于现有的AI事故数据库，其覆盖面、规模和种类有限。后续可能需要通过新闻和社交媒体收集更多AIGC事故，以便进行更深入的分析。其次，本文以人工的方式进行内容编码，存在一定主观因素的影响。将来可以尝试使用NLP等文本分析工具进行更多维度的分析。最后，关于生成式人工智能治理行动框架还限于初步探讨，虽然在一定程度上显示出与国家管理思路的匹配性[8]，但仍需要大量工作有重点、分步骤地继续完善行动框架的理论和操作细节。

近期，国务院正预备提请全国人大常委会审议人工智能法草案[32]，但究竟采用何种态度和措施对生成式人工智能进行服务管理仍存在诸多未知和争议，希望本文能够以宏观视角为我国提前谋划生成式人工智能治理提供拙见。

作者贡献说明

朱禹：提出研究思路，设计研究方案，内容编码，论文起草与修改；
陈关泽：采集、清洗和分析数据，内容编码，论文起草与修改；
陆泳溶：内容编码；
樊伟：提供论文修改建议。

支撑数据

支撑数据由作者自存储，E-mail:13350057427@163.com。

1. 朱禹，陈关泽 . AIGC Incident Report Texts.zip. AIGC 事故报道文本数据 .

2. 朱禹，陈关泽，陆泳溶 . Manual coding.mx22. 样本编码数据 .

参考文献

[1] 张钹 , 朱军 , 苏航 . 迈向第三代人工智能 [J]. 中国科学 : 信息科学 ,2020,50(9):1281-1302.(Zhang Bo,Zhu Jun,Su Hang. Towards the Third Generation of Artificial Intelligence[J]. Scientia Sinica(Informationis) ,2020,50(9):1281-1302.)

[2] 陆伟 , 刘家伟 , 马永强 , 等 .ChatGPT 为代表的大模型对信息资源管理的影响 [J]. 图书情报知识 ,2023,40(2):6-9,70.(Lu Wei, Liu Jiawei, Ma Yongqiang, et al. The Influence of Large Language Models Represented by ChatGPT on Information Resource Management [J]. Documentation, Information & Knowledge,2023,40(2):6-9,70.)

[3] Borji A. Pros and Cons of GAN Evaluation Measures[J]. Computer Vision and Image Understanding,2019,179:41-65.

[4] 国务院 . 国务院关于印发新一代人工智能发展规划的通知 [EB/OL]. [2023-03-23]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.(State Council of the People's Republic of China.Notice of the State Council on the Issuance of a New Generation of Artificial Intelligence Development Plan[EB/OL].[2023-03-23]. http://www.gov.cn /zhengce/content/2017-07/20/content_5211996.htm.)

[5] 方旭 , 卫燕 , 张颖 , 等 . 欧盟人工智能法案综述 [J]. 计算机时代 ,2022(5):92-95.(Fang Xu,Wei Yan,Zhang Ying,et al. Overview of European Union Artificial Intelligence Act[J]. Computer Era,2022(5):92-95.)

[6] 康逸 . 欧盟《人工智能法案》遭遇绊脚石 [N]. 经济参考报 ,2023-02-21(5) .(Kang Yi. EU Artificial Intelligence Act Encounters Stumbling Block [N]. Economic Information Daily,2023-02-21(5) .)

[7] United States Congress. Daily Digest/Senate Committee Meetings; Congressional Record Vol. 169, No. 82 [EB/OL]. [2023-06-06]. https://www.congress.gov/congressional-record/volume-169/issue-82/daily-digest/article/D460-1?q=%7B%22search%22%3A%5B%22OpenAI%22%5D%7D&s=2&r=1.

[8] 国家互联网信息办公室 . 国家互联网信息办公室关于《生成式人工智能服务管理办法（征求意见稿）》公开征求意见的通知 [EB/OL]. [2023-04-17]. http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm.(State Internet Information Office. Notice of the State Internet Information Office on Soliciting Public Comments on the Administrative Measures for Generative Artificial Intelligence Services(Draft) [EB/OL]. [2023-04-17]. http://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm.)

[9] Commission Nationale de l'Informatique et des Libertés. Artificial Intelligence: The Action Plan of the CNIL [EB/OL]. [2023-06-06]. https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil.

[10] Viljanen M, Parviainen H. AI Applications and Regulation: Mapping the Regulatory Strata[J]. Frontiers in Computer Science,2022,3:779957.

[11] Wei M Y, Zhou Z X. AI Ethics Issues in Real World: Evidence from AI Incident Database[EB/OL].[2023-04-17]. https://arxiv.org/abs/2206.07635.

[12] 吴红旗 , 朱文文 . 互联网语境下虚假新闻的特点及规制 [J]. 新闻爱好者 ,2022(8):65-67.(Wu Hongqi,Zhu Wenwen. Characteristics and Regulation of False News in Internet Context[J]. Journalism Lover,2022(8):65-67.)

[13] 蔡立媛 , 李晓 . 人工智能广告侵犯隐私的风险与防御 [J]. 青年记者 ,2020(18):93-94.(Cai Liyuan,Li Xiao. Risk and Defense of Artificial Intelligence Advertising Invading Privacy[J]. Youth Journalist,2020(18):93-94.)

[14] McGregor S.Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database[J].Proceedings of the AAAI Conference on Artificial Intelligence,2021,35(17):15458-15463.

[15] 于雯雯 . 再论人工智能生成内容在著作权法上的权益归属 [J]. 中国社会科学院大学学报 ,2022,42(2):89-100,146-147.(Yu Wenwen. On the Ownership of Artificial Intelligence-Generated Content in Copyright Law [J]. Journal of University of Chinese Academy of Social Sciences, 2022,42(2):89-100,146-147.)

[16] 丛立先 , 李泳霖 . 聊天机器人生成内容的版权风险及其治理——以 ChatGPT 的应用场景为视角 [J]. 中国出版 ,2023(5):16-21.(Cong Lixian, Li Yonglin. Copyright Risk and Governance of Chatbot Generated Content—from the Perspective of ChatGPT Application Scenario [J]. China Publishing Journal,2023(5):16-21.)

[17] 吴汉东 . 人工智能生成作品的著作权法之问 [J]. 中外法学 ,2020,32(3):653-673.(Wu Handong. On the Copyright Law of Works Generated by Artificial Intelligence[J]. Peking University Law Journal,2020,32(3):653-673.)

[18] 李白杨 , 白云 , 詹希旎 , 等 . 人工智能生成内容（AIGC）的技术特征与形态演进 [J]. 图书情报知识 ,2023,40(1):66-74.(Li Baiyang, Bai Yun, Zhan Xini, et al. Technical Characteristics and Morphological Evolution of Artificial Intelligent-generated Content [J]. Documentation, Information & Knowledge,2023,40(1):66-74.)

[19] 蒋华林 . 人工智能聊天机器人对科研成果与人才评价的影响研究——基于 ChatGPT、Microsoft Bing 视角分析 [J]. 重庆大学学报（社会科学版）,2023,29(2):97-110.(Jiang Hualin. The Impact of Artificial Intelligence Chatbot on Scientific Research Achievements and Talent Evaluation: Based on ChatGPT and Microsoft Bing [J]. Journal of Chongqing University(Social Sciences Edition) ,2023,29(2):97-110.)

[20] 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/OL].(2023-04-13) [2023-04-17]. 重庆大学学报（社会科学版）. http://kns.cnki.net/kcms/detail/50.1023.C.20230412.1004.002.html.(Pu Qingping, Xiang Wang. Opportunities and Challenges Aroused ByChatGPT as Generative AI and Strategy for Response[J/OL]. Journal of Chongqing University(Social Sciences Edition) (2023-04-13) [2023-04-17].http://kns.cnki.net/kcms/detail/50.1023.C.20230412.1004.002.html.)

[21] 陆小华 .ChatGPT 等智能内容生成与新闻出版业面临的智能变革 [J]. 中国出版 ,2023(5):8-15.(Lu Xiaohua. ChatGPT and Other Intelligent Content Generation and the Intelligent Transformation Faced by the News and Publishing Industry[J].China Publishing Journal,2023(5):8-15.)

[22] 中国信息通信研究院 . 人工智能生成内容（AIGC）白皮书（2022年）[R/OL]. 北京：中国信息通信研究院 (2022-09-02) [2023-03-23].http://www.caict.ac.cn/kxyj/qwfb/bps/202209/P020220902534520798735.pdf.(China Academy of Information and Communications Technology. Artificial Intelligence Generated Content(AIGC) White Paper(2022) [R/OL]. Beijing: China Academy of Information and Communications Technology, (2022-09-02) [2023-03-23].http://www.caict.ac.cn/kxyj/qwfb/bps/202209/P020220902534520798735.pdf.)

[23] Khoo B,Phan R C W,Lim C H. Deepfake Attribution: On the Source Identification of Artificially Generated Images[J]. Wiley Interdisciplinary Reviews:Data Mining and Knowledge Discovery,2022,12(3):e1438.

[24] Chesney B, Citron D. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security[J]. California Law Review,2019,107(6): 1753-1820.

[25] Partadiredja R A, Serrano C E, Ljubenkov D. AI or Human: the Socio-ethical Implications of AI-generated Media Content[C]//2020 13th CMI Conference on Cybersecurity and Privacy(CMI) -Digital Transformation-Potentials and Challenges(51275) .Piscataway ,NJ:IEEE, 2020: 1-6.

[26] European Union Intellectual Property Office. Study on the Impact of Artificial Intelligence on the Infringement and Enforcement of Copyright and Designs[EB/OL].[2023-04-13].https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2022_Impact_AI_on_the_Infringement_and_Enforcement_CR_Designs/2022_Impact_AI_on_the_Infringement_and_Enforcement_CR_Designs_FullR_en.pdf.

[27] Hsieh H F,Shannon S E. Three Approaches to Qualitative Content Analysis[J]. Qualitative Health Research,2005,15(9):1277-1288.

[28] 中国大百科全书出版社 . 虚假信息 [EB/OL]. [2023-04-12]. https://www.zgbk.com/ecph/words?SiteID=1&ID=26078&Type=bkzyb&SubID=60755.(Encyclopedia of China Publishing House. False Information[EB/OL]. [2023-04-12]. https://www.zgbk.com/ecph/words?SiteID=1&ID=26078&Type=bkzyb&SubID=60755.)

[29] 张玉磊 . 跨界公共危机治理行动者网络：一种新的解释框架 [J]. 上海大学学报 (社会科学版) ,2023,40(1):74-92.(Zhang Yulei. Actor Network of Trans-Boundary Public Crisis Governance: A New Explanatory Framework[J]. Journal of Shanghai University(Social Sciences Edition) , 2023,40(1):74-92.)

[30] 邓晔 , 宋莉娜 . 网络有害信息治理模式比较研究 [J]. 湖南社会科学 ,2018(4):106-112.(Deng Ye,Song Lina. Comparative Study on Governance Models of Harmful Information on the Internet[J]. Social Sciences in Hunan,2018(4):106-112.)

[31] 杨华锋 . “情境—意识—行动”框架下国家安全治理的模型假设 [J]. 国际安全研究 ,2022,40(6):61-85(Yang Huafeng. Model Hypothesis of National Security Governance under the Framework of "Situation-Consciousness-Action"[J]. Journal of International Security Studies,2022,40(6):61-85.)

[32] 国务院办公厅 . 国务院办公厅关于印发国务院2023年度立法工作计划的通知 [EB/OL]. [2023-06-12]. https://www.gov.cn/zhengce/content/202306/content_6884925.htm.(General Office of the State Council. Notice of The General Office of the State Council on Printing and Distributing The State Council's 2023 Legislative Work Plan [EB/OL]. [2023-06-12]. https://www.gov.cn/zhengce/content/202306/content_6884925.htm.)

(收稿日期:2023-04-18)

上接第091页

[59] 北京大学前沿交叉学科研究院 . 生物医学跨学科研究中心 [EB/OL].[2023-02-16].http://www.aais.pku.edu.cn/yjzx/show.php?id=157.(Peking University Frontier Interdisciplinary Research Institute. Biomedical Interdisciplinary Research Center[EB/OL] [2023-02-16]. http://www.aais.pku.edu.cn/yjzx/show.php?id=157.)

[60] 北京大学前沿交叉学科研究院 . 学院简介 [EB/OL].[2023-02-16].http://www.aais.pku.edu.cn/about/.(Academy for Advanced Interdisciplinary Studies, Peking University. Profile[EB/OL]. [2013-02-16] http://www.aais.pku.edu.cn/about/.)

[61] 武汉大学 . 多学科交叉研究院 [EB/OL].[2023-02-16]. http://www.whu.edu.cn/kxyj/dxkjcyjy.htm.(Wuhan University. Multidisciplinary Research Institute[EB/OL] [2023-02-16]. http://www.whu.edu.cn/kxyj/dxkjcyjy.htm.)

[62] Tash W R. Evaluating Research Centers and Institutes for Success: A Manual and Guide with Case Studies[M]. Fredericksburg, VA: WT & Associates, 2006:230.

[63] Stahler G J, Tash W R. Centers and Institutes in the Research University - Issues, Problems, and Prospects[J].Journal of Higher Education, 1994,65(5):540-54.

[64] 国务院 . 关于印发积极牵头组织国际大科学计划和大科学工程方案的通知 [EB/OL].[2023-02-16].http://www.gov.cn/zhengce/content/2018-03/28/content_5278056.htm.(The State Council. Notice on Printing and Distributing the Plan for Actively Leading and Organizing International Large Science Projects and Large Science Projects[EB/OL] [2023-02-16]. http://www.gov.cn/zhengce/content/2018-03/28/content_5278056.htm.)

[65] Guns R, Rousseau R. Recommending Research Collaborations Using Link Prediction and Random Forest Classifiers[J].Scientometrics, 2014,101(2):1461-1473.

[66] 潘玉腾 . 高校实施有组织科研的问题解构与路径建构 [J]. 中国高等教育 ,2022(Z3):12-14.(Pan Yuteng. The Deconstruction and Path Construction of the Implementation of Organized Research in Universities[J].China Higher Education,2022(Z3):12-14.)

[67] 黄淑芳 . 基于跨学科合作的团队异质性与高校原始性创新绩效的关系研究 [D]. 杭州：浙江大学 ,2016.(Huang Shufang. Study on the Relationship Between Interdisciplinary Team Heterogeneity and Original Innovation Performance[D]. Hangzhou: Zhejiang University, 2016.)