

AIGC治理的研究进展与发展趋势*

邓胜利 汪璠

(武汉大学信息管理学院, 武汉 430072)

摘要: 对国内外人工智能生成内容(AIGC)的治理研究进行全面梳理,发现AIGC治理的研究热点和问题,明确未来的发展方向,为新的AI环境下AIGC治理策略的变革和提升提供指导。通过对国内外现有研究进行详细分析,从技术治理、法律治理、伦理治理和多主体协同治理等多个维度,探讨AIGC带来的新的治理要求,以及如何建立多元主体治理体系来应对。首先,AIGC的出现使信息的复杂性和隐蔽性增强,需要通过先进的AI模型和严格的质量管理制度来确保生成的内容在伦理上是可接受的,符合社会和个体的价值观和道德标准。其次,在法律方面关注版权、隐私、虚假信息以及算法歧视等多个法律问题,预见、研究并制定相关的法规。再次,在技术治理方面需要重视算法透明度和可解释性、数据保护和隐私、技术的使用场景适配、用户指导等几个关键内容。最后,通过多主体协同治理让政府、行业和媒体等多方参与,结合技术手段和法律法规,共同形成灵活、高效的治理框架。总之,揭示AIGC治理研究的现状和趋势,提出一系列应对策略,为AIGC的治理提供指导。

关键词: 人工智能生成内容;多主体协同治理;法律治理;技术治理;伦理治理

中图分类号: G203 **DOI:** 10.3772/j.issn.1673-2286.2023.11.003

引文格式: 邓胜利,汪璠. AIGC治理的研究进展与发展趋势[J]. 数字图书馆论坛, 2023(11): 20-28.

1 AIGC的发展及其问题

随着人工智能(Artificial Intelligence, AI)的迅猛发展,特别是以GPT系列模型为代表的大语言模型的出现,人工智能生成内容(Artificial Intelligence Generated Content, AIGC)在信息生成领域爆炸性增长^[1],AIGC凭借其强大的生成能力,已经与众多行业深度融合,在应用层面取得突破性的进展。

AIGC的发展也带来诸多问题。首先,AIGC可以被用来散布谣言、虚假广告,甚至进行网络欺诈和舆情操控,导致信息污染问题日益严重,甚至给社会带来极大的不稳定性和信任危机。其次,由于AIGC的生成速度和规模远远超过了传统的人力审查能力范围,有效治理AIGC变得更加困难。此外,AIGC也可能涉及敏感、不适宜或违法的信息,进一步加大了治理的复杂性^[2]。

面对AIGC带来的诸多问题,有效治理AIGC显得尤为重要。合理的AIGC治理有助于维护信息传播的真实性和可信度,减少误导性信息的传播,防范信息污染的风险。保护用户免受有害内容的侵害,提高网络空间的安全性和健康性也是其重要目标^[3]。围绕这些目标,学界进行了对AIGC治理的一系列探索,形成了大量研究成果。本文旨在系统梳理AIGC的研究进展和发展趋势,为社会各界更好地理解AIGC治理提供帮助。

2 国内外AIGC治理的研究进展

2.1 AIGC治理研究的数据获取

为梳理国内外AIGC治理领域的研究,分析其发展现状和发展趋势,对多个中外数据库进行了文献

收稿日期: 2023-07-27

*本研究得到国家自然科学基金项目“信息生态链视角下在线知识社区用户贡献行为评价及预测研究”(编号:71974149)、国家社会科学基金重大项目“人本人工智能驱动的信息服务体系重构与应用研究”(编号:22&ZD324)资助。

检索。中文来源为：中国知网、万方，检索主题中含有“AIGC治理”或“ChatGPT治理”或“人工智能生成内容”或“大语言模型治理”等的文献；外文来源为：Web of Science、Emerald、Springer，检索主题中含有“AIGC”或“AIGC Governance”或“ChatGPT Governance”或“Artificial Intelligence Generated Content Governance”等的文献。截至2023年7月10日，共获得论文289篇，经过去重、筛选，最终得到文献124篇。从文献分布来看，ChatGPT时代AIGC的治理研究涵盖了法学、管理学、计算机科学、社会学、传播学等多个学科。这些学科为研究提供了不同的视角和见解，从而有助于更深入地探讨AIGC治理研究这一领域，并设计出最有效的干预措施。

2.2 国内外AIGC治理研究的内容框架

通过对现有文献进行内容分析，发现研究人员对于AIGC的治理研究主要涉及技术治理、法律治理、伦理治理和多主体协同治理4个方面，由此构建了AIGC治理的整体框架（见图1）。首先，技术治理是基础，其提供了AIGC开发的方法、透明度增强的手段以及数据处理的技术，确保技术本身具备合规性和道德性，为整个治理体系的实施提供了技术支持^[4]。其次，法律治理在整体框架中具有关键作用，通过建立法律框架，明确了AIGC技术和应用的法律责任和义务，解决了知识产权、隐私法规、反歧视法律等法律问题，确保AIGC的合法性^[5]。再次，伦理治理进一步强调了AIGC在伦理和道德方面的合规性，关注信息的准确性、减轻偏见和歧视、隐私保护、消费者权益以及社会价值等伦理问题，确保AIGC技术的应用在伦理和道德框架内进行^[6]。最后，多主体协同治理是整个治理框架的核心，强调各主体需要协同合作，共同解决跨领域的伦理、法律和技术问题，促进AIGC的健康发展和管理，涵盖政府、行业、用户和媒体等多个利益相关者，以实现多主体协同治理的目标^[7]。

如图1所示，AIGC的技术治理、法律治理、伦理治理互相交织、互相依赖^[8]。在伦理治理和技术治理之间，伦理要求和道德规范会支撑技术解决方案的改进和创新，技术治理的决策会反映伦理；在技术治理和法律治理之间，技术治理的现状和发展趋势会挑战现有的法律框架，推动法律制度的更新和完善，法律治理也为技术治理设定了一系列规则和限制，以确保技术的合

法性和道德性。

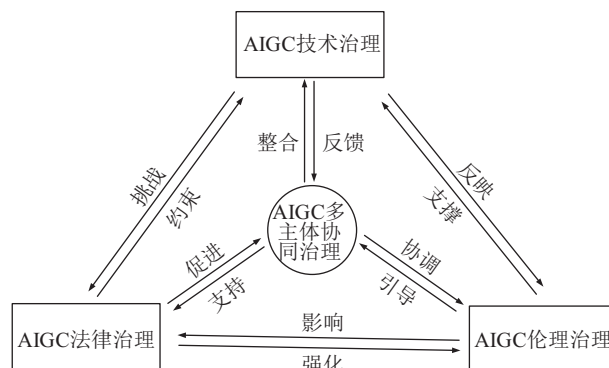


图1 AIGC治理的整体框架

同时，AIGC多主体协同治理整体能够将伦理治理、法律治理和技术治理的主体进行有机结合。在技术治理和多主体协同治理之间，技术治理需要各方的参与和反馈，以确保技术的透明性、公正性和可控性，多主体协同治理也为技术治理提供了一个使各方能够在技术治理的过程中整合协调的机制；在法律治理和多主体协同治理之间，法律治理为多主体协同治理设定了法律框架和规则，促进协同治理的实施，各方也可以通过协同治理共同解决法律制度的执行问题，支持法律治理的实施和完善；在法律治理和伦理治理之间，伦理治理中的质量、公平性和准确性等道德标准可以成为法律治理的基础，影响立法者制定法规，推动法律规定的修订和完善，法律治理也为伦理治理提供了宏观的监管框架，强化了社会的伦理观念和价值观。四者相互协作，形成了一种多元化、动态的治理模式，共同应对AIGC所带来的挑战。

3 AIGC治理研究进展

基于图1所示的研究框架，分别从AIGC的技术治理、法律治理、伦理治理和多主体协同治理4个方面来阐释AIGC治理研究的具体进展。

3.1 AIGC的技术治理

AIGC作为一项新的技术，在实践应用中会出现各种问题。因此，学界首先从技术层面对AIGC治理进行研究^[9]，围绕AI的可解释性、数据的隐私保护等^[10]提出具体的治理方式，如表1所示。

从表1可以看出，当前研究在技术治理方面涉及了多

表1 AIGC的技术治理研究

作 者	技术治理方式	内 容
朱禹等 ^[11]	数据脱敏、使用场景适配	总结了AIGC导致的各种事故类型, 提出需要使用数据脱敏和使用场景适配等技术手段解决由AIGC类技术导致的隐私泄露、偏见歧视、诈骗等多种风险
许雪晨等 ^[4]	算法透明、数据保护	行业需要建立可解释和可信赖的AIGC框架, 提高算法透明度, 避免“算法黑箱”带来的各类歧视问题
王鹏涛等 ^[12]	保障数据安全、提高AIGC技术信任、技术的场景适配	AIGC的开发者和设计者不仅要解决数据安全问题, 还要在使用前充分评估场景适配以及提高用户的技术信任
蔡子凡等 ^[13]	算法透明、用户管理	需要提高AIGC算法透明度, 增强用户管理, 以解决算法不公平等一系列问题
李宏洋 ^[14]	确定技术规范标准	对AIGC的技术原理、发展现状、应用场景等进行深入分析, 通过规范技术标准解决虚假回答、信息污染、数据隐私、路径依赖等现实问题
徐璐璐等 ^[15]	算法可解释性提高、医学与技术协同	AIGC类技术是一把双刃剑, 在正式应用到医疗场景之前, 仍需通过提高算法可解释性和增加医学与技术的协同度, 进一步增强AI模型的能力和降低AI模型的局限性
徐继敏等 ^[16]	提高可解释性、制定技术规范	大语言模型产生的AIGC具有直观性强、可解释性弱和可识别性弱的特点, 通过提高可解释性和制定技术规范能够更好地应对这些问题
Chan ^[17]	数据保护、用户管理	提出教学、治理和运营的AI生态教育政策框架, 利用数据保护和用户管理降低AIGC技术在高等教育中的负面影响
Leclerc等 ^[18]	场景适配、算法可解释性提高	讨论了AIGC在医学领域的应用, 指出其存在不可解释性等需要关注的潜在限制和风险, 并提出通过适配使用场景和增强算法可解释性, 保障其在医学领域应用的准确性
Sam ^[19]	算法可解释性提高	通过对AIGC在学术写作、患者交流等各方面的表现进行评估, 明确了在应用中需要注意的算法可解释性问题
莫祖英等 ^[20]	数据质量管理、模型审查	模型和数据来源会导致AIGC中存在数据、作品、事实、代码、翻译错误等虚假信息, 需要通过模型审查和数据质量管理来解决这些问题

项工作, 包括数据脱敏、可解释性提高、使用场景适配、算法透明、数据保护、技术场景适配、用户管理、技术规范等方面^[21], 旨在解决技术相关的问题^[22]。具体的技术治理措施如图2所示。

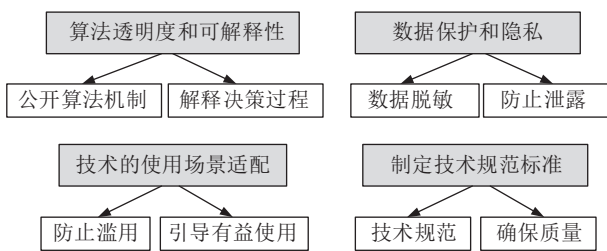


图2 AIGC技术治理的具体措施

(1) 算法透明度和可解释性: 确保AIGC技术合规治理的重要内容。通过确保算法决策过程可视化和可理解, 使外部观察者能够理解算法的工作原理, 以便更容易识别和解决偏见、歧视和误导等问题^[23]。同时, 通过公开算法机制和解释决策过程也可以防止出现“算法黑箱”问题, 即确保算法决策过程对外部观察者可见, 避免不透明结果^[24]。

(2) 数据保护和隐私: 数据脱敏和数据保护是保护用户隐私的关键步骤。数据脱敏通过删除或模糊个人身份信息, 同时维持数据的有用性, 可以防止潜在的

隐私泄露风险^[25]。此外, 数据加密、存储、传输的安全性也是数据保护的重要方面^[26]。这些措施有助于确保AIGC使用的数据受到充分的保护, 维护用户的隐私。

(3) 技术的使用场景适配: 对部分AIGC使用场景进行适配和监管^[12]。通过使用场景适配, 建立和引导有益的AIGC使用方式, 有效避免AIGC被滥用, 减少潜在的风险和问题^[18]。

(4) 制定技术规范标准: 确保AIGC技术遵循最佳实践和法规。这有助于规范AIGC技术的开发和应用, 提高信息治理水平, 减少虚假信息, 确保信息质量, 以满足伦理和法律标准^[14]。技术规范包括对算法的要求、数据处理的准则、安全性标准等, 确保AIGC技术的合规性和道德性。

3.2 AIGC的法律治理

尽管技术治理可以帮助确保AIGC技术在伦理和道德框架内运作, 但仍然存在一些限制和挑战, 这时就需要法律治理的介入和支持^[5]。围绕AIGC所产生的法律问题, 学界进行了广泛探讨。

如表2所示, AIGC的法律治理研究已经取得一系列重要进展。研究人员在探讨AIGC所引发的各类法律问题时做出了重要工作, 包括版权问题、隐私问题以及

表2 AIGC的法律治理研究

作 者	提及的法律问题	具体内容
游俊哲 ^[27]	版权问题	需要探索如何抵制对现有版权规则和科研评价体系的冲击, 这涉及AIGC的版权归属, 以及如何平衡AIGC的创新与现行的版权法律之间的关系的问题
任安麒 ^[28]	版权问题	新一代AIGC会引发出版内容同质化和侵权问题等, 应从法律角度探索新的数字出版治理方案
邹开亮等 ^[29]	隐私问题、算法歧视	加强数据监管, 应对数据偏见, 防范数据依赖造成的个人信息泄露和算法歧视
朱永新等 ^[30]	数据隐私问题	筑牢国家数字屏障, 确保个人数据隐私和国家数据安全不受威胁
张夏恒 ^[31]	隐私问题、版权问题、算法歧视	需要尽快完善AIGC和大模型的相关立法, 避免新一代AI技术发展中产生的各类重大风险
张璐 ^[32]	算法歧视	探索基于现有治理制度的改进方案, 构建AIGC的分层治理体系, 防止技术垄断导致的算法歧视问题
Cheng等 ^[5]	算法歧视和垄断	通过主题间文本分析AI伦理和法律问题, 探讨ChatGPT等生成式AI技术与人类的关系, 防范算法歧视
Lund等 ^[33]	版权问题	探讨AI技术在学术研究与出版方面的潜在影响, 尤其是自动准备学术文稿的可能性
方卿等 ^[34]	明确权利归属	从出版的角度, 关注AIGC治理中的法律问题、意识形态泄露和学术道德问题, 提出需要明确AIGC的权利归属

算法歧视问题等。

版权问题涉及了《中华人民共和国著作权法》, 该法规定了著作权人的权利, 规范了作品的使用和保护, 包括文字、美术、音乐、影视等多个领域。尤其是第十三条规定了“以独创性作品为基础, 经改编、改写、翻译、注释或者其他方式创作的作品”属于著作权保护范围。如果AIGC直接或间接涉及受版权保护的元素, 那么可能触犯该法律。

隐私问题涉及了《中华人民共和国个人信息保护法》, 该法规定了个人信息的获取、使用和贡献均需取得信息主体同意, 且个人信息控制者应当采取技术措施和其他必要措施, 保护个人信息的安全。如果AIGC的训练数据中包含个人信息, 或者本身可能涉及个人隐私, 那么可能触犯该法律。

算法歧视涉及了《中华人民共和国反垄断法》, 该法规定经营者不得滥用市场支配地位。如果企业利用AIGC的算法在市场中导致不正当竞争, 那么可能会触犯这一法律。

当前AIGC法律治理的具体措施如图3所示。

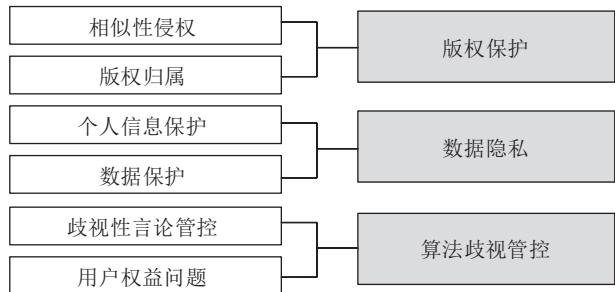


图3 AIGC法律治理的具体措施

(1) 版权问题: AIGC可能涉及版权问题, 因为它的训练数据可能来自已有版权的文本或图片, 并生成与已存在和被版权保护的作品高度相似的内容。研究者提出需要探索如何抵制AIGC对现有版权规则和科研评价体系的冲击, 并指出可能需要新的版权法规^[27]。此外, 针对AIGC自动生成学术文稿的潜在影响, 特别是在学术研究与出版方面, 需要进一步的法律研究^[28]。

(2) 隐私问题: AIGC的使用可能涉及隐私问题, 尤其是在处理大量用户数据时。建议加强数据监管以应对隐私问题, 以及防范数据偏见带来的风险^[29]。另外, 筑牢国家数字屏障, 明确AIGC应用的透明性, 也是一种法律治理手段。

(3) 算法歧视问题: AIGC的算法可能存在歧视性, 对不同群体或个体进行不公平对待^[32]。研究者提出构建AIGC的分层治理体系, 以解决算法歧视问题, 并将法律制定与伦理分析相结合, 以更好地应对这一挑战^[33]。

3.3 AIGC的伦理治理

AIGC的伦理治理是一个日益重要的研究领域, 旨在确保AIGC在质量、公平性、准确性和道德性等方面能够达到社会和个体的期望, 从而维护信息生态的健康和社会的稳定。在这方面, 现有研究也提出各种治理方案, 如表3所示。

从表3可以看出, AIGC的伦理治理主要包括伦理风险评估和审查、伦理教育和管理、伦理沟通与明晰,

表3 AIGC的伦理治理研究

作 者	措 施	伦理问题探讨
支振锋 ^[35]	注重伦理道德，平衡安全和发展	在最大化AIGC的创新与应用价值的同时，通过识别、管理和减轻信息风险，保证网络信息内容生态健康，关注伦理问题，确保生成的内容在道德和伦理上是可接受的，实现技术进步与安全保障之间的平衡
陈智等 ^[36]	兼顾伦理，可供赋能	应理解和解构ChatGPT的“可供性”，在处理伦理问题时考虑伦理原则，以兼顾社会的道德期望。通过多方伦理沟通，正确应对算法偏差、感知极化、情感流失、信息集权和社会恐慌等多重风险的问题
喻国明 ^[37]	符合伦理准则的信息推举范式	通过引入“信息助推”范式的内容标记技术，对深度合成产品进行规范化标注并充分考虑伦理问题，讨论如何有效减少认知失序现象、保障信息的客观真实性和意识形态正确性的问题
栾轶玫等 ^[38]	加强伦理风险控制	在重大主题报道中，通过分析AIGC的意识形态、加强伦理风险控制并确保处理内容不违反伦理原则，探讨如何在报道中遵守伦理原则的问题
Paul等 ^[39]	多角度伦理考量	讨论了使用AIGC的多维益处和潜在风险，强调了对偏见、信息不准确、隐私问题等潜在问题进行治理的必要性
Gradon ^[40]	风险评估，滥用防范	评估和应对由犯罪分子滥用AIGC技术进行网络钓鱼活动或传播虚假信息带来的问题，同时探讨和理解技术开发者对其产品的责任。伦理治理在风险评估和滥用防范中发挥关键作用，以确保技术使用是合乎伦理的
Liao等 ^[6]	推动伦理与技术协同，提升伦理透明度	在人机协同中，伦理原则应与技术协同，以确保医疗领域的伦理规范得到遵守。探讨在医疗领域AIGC技术作为辅助工具的优势和挑战，以及其与人类专业知识协同，推动智能医疗发展的潜力。此外，通过伦理明晰，能确保在医疗领域中AIGC技术的使用和决策过程是清晰、明确和透明的

具体举措如图4所示。

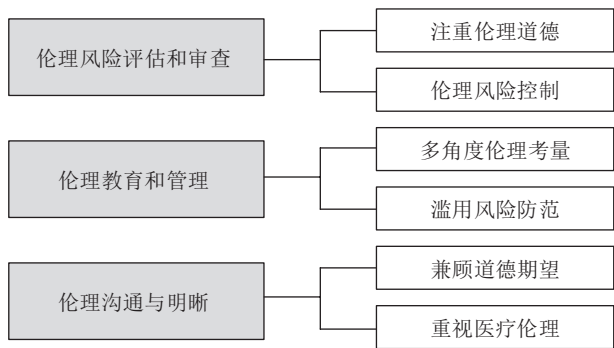


图4 AIGC的伦理治理具体措施

（1）伦理风险评估和审查：研究者和从业者需要进行伦理风险评估和审查，以确保AIGC技术在各领域遵守伦理规范^[41]。这包括加强伦理风险控制，确保内容处理不违反伦理原则，以及评估由犯罪分子滥用AIGC技术带来的问题。这些措施有助于促进AIGC技术的道德实践^[42]。

（2）伦理教育和管理：伦理教育和管理是确保各方能够全面理解和应对AIGC伦理问题的关键^[43]。必须讨论AIGC的多维益处和潜在风险，强调对偏见、信息不准确、隐私问题等潜在问题治理的必要性。通过伦理教育和管理，可以建立伦理意识，培养伦理决策能力，确保AIGC技术的合法和道德应用^[44]。

（3）伦理沟通与明晰：伦理沟通是指与不同群体

之间就伦理问题进行交流、协商和解决，以达成共识和一致的行动。多方需要通过充分沟通，理解和解构AIGC技术的“可供性”，以充分考虑伦理原则，并达到社会的道德期望^[36]。伦理明晰则是指在进行伦理决策时，确保信息的公开和透明。通过伦理明晰，可以确保AIGC技术在公平、道德和可接受的框架内运作。特别是在医疗领域，伦理沟通和明晰对于确保技术的使用符合伦理规范和法律法规，保障患者的权益和安全至关重要。

AIGC的伦理治理需要解决伦理原则、内容道德、风险评估和滥用防范等多个方面的问题。为解决这些问题，需要制定新的伦理原则、加强监管、推动伦理规范的实施以及提供伦理指导。

3.4 AIGC的多主体协同治理

AIGC涉及复杂的技术、法律和伦理等问题，对社会各方产生广泛影响。尽管技术、法律和伦理治理各有作用，但由于AIGC的多样性和广泛应用，单一治理领域难以全面应对挑战。多主体协同治理可促进各利益相关者的合作，确保治理全面有效^[45]。这种合作有助于协调技术、法律和伦理工作，共同制定标准，确保AIGC安全、可靠、可持续发展，促进AI与社会和谐发展^[46]。相关研究如表4所示。

表4 AIGC的多主体协同治理研究

作 者	治理主体	治理策略
赵精武 ^[47]	政府、行业、用户等	兼顾伦理、技术和法律, 探讨在不单独对AIGC管理进行立法的情况下, 如何构建相关安全标准和评估机制, 实现对AIGC的治理
舒洪水等 ^[48]	政府、行业	AIGC产生的虚假信息具有数据量大、传播速度极快的特点。 政府需加强综合立法、行业需加强平台治理
罗昕 ^[49]	政府、行业、媒体、社会团体	AIGC会引发新的传播革命, 并为国家、社会、个人和行业带来多种潜在风险, 应当通过更加灵活高效的多元协同治理模式去规范AIGC产业
刘艳红 ^[50]	政府、行业、知识产权机构、消费者	针对AIGC的运行机理, 发现AIGC的三大安全风险。需要分别从数据风险、算法偏见风险和知识产权风险角度对AIGC进行治理, 减少负面影响
Dwivedi等 ^[8]	学术界、政府、行业、非政府组织等	通过对知识、透明度和伦理等进行深入研究, 解决偏见问题、厘清适用场景、明确文本准确性评估方法以及伦理和法律问题等, 指出AIGC在多个领域的应用机会和挑战, 呼吁进行多主体的协同治理
Lin ^[51]	政府、行业、学术界、媒体、社会团体等	明确AIGC治理过程中多方主体的角色和责任, 以确保AIGC技术在提供有效支持的同时, 维护公正、安全、准确的信息传播, 凸显了多主体协同治理的重要性

从表4可以看出, 多主体协同治理涉及政府、行业、学术界、媒体、社会团体等多个主体的合作, 以确保AIGC的发展和应用在各方的监督合作下遵循伦理、法律和社会标准^[51]。AIGC的多主体协同治理措施如图5所示。

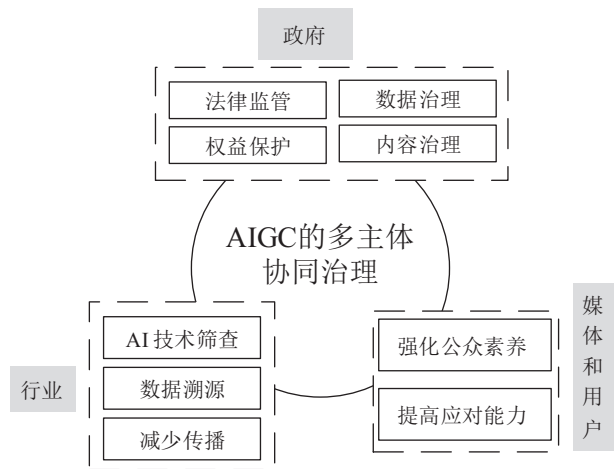


图5 AIGC的多主体协同治理措施

(1) 政府角度的协同治理。政府应牵头制定全面的法律法规, 明确AIGC技术的合规要求, 包括数据隐私、知识产权、伦理原则等。这些法规应当为行业、媒体和用户提供明确的法律框架, 以规范AIGC的发展和应用^[48]。此外, 应设立专门的监管机构, 负责监督AIGC技术的合规性和伦理性。这些机构可以与行业、媒体以及用户代表协同合作, 确保AIGC技术不受滥用, 同时促进创新和发展^[49]。

(2) 行业角度的协同治理。行业组织应共同制定AIGC技术的行业标准, 包括数据处理、算法设计、内

容生成等方面的准则, 并通过行业合作总结最佳实践和经验^[50]。这有助于确保AIGC技术在行业内遵循一致的伦理和技术标准, 从而应对伦理和技术挑战。

(3) 用户和媒体角度的协同治理。用户应当参与AIGC技术治理的过程, 他们的反馈和意见应被纳入考虑。同时, 需要进行用户教育, 帮助他们更好地理解AIGC技术的使用方式和潜在风险^[8]。此外, 建立用户举报和反馈机制, 使用户和媒体可以报告不当内容或行为, 这将有助于及时发现和应对违规行为。

4 AIGC治理研究的发展趋势

本文分析了AIGC治理的相关研究, 归纳出AIGC治理研究主要聚焦于技术治理、法律治理、伦理治理和多主体协同治理4个核心领域。对这4个方向的综合分析揭示了AIGC治理研究的一些基本特征。

(1) 多主体和跨学科性质。AIGC治理研究既需要政府、学术界、技术行业 and 用户等多主体的共同参与, 也需要对技术、法律、伦理、数据科学等多个领域进行深入研究和整合。技术研究者能在设计更安全、更公平、更透明的AIGC系统方面提供支持; 法律研究者能帮助创建和执行合规的政策和法规; 伦理研究者则能够探讨AIGC可能引发的道德和社会问题。

(2) 算法透明性与可解释性。为了加深用户对AIGC系统的理解和信任, 研究人员注重提高AIGC的透明性和可解释性。这包括让用户理解AIGC的工作原理、决策过程, 以及对内容的具体解释。研发人员需要开发新的算法和工具, 以提供更直观、更易理解的解释, 并对复杂的决策过程进行可视化。

(3) 数据隐私与安全。AIGC系统通常需要处理大量的用户数据,因此必须确保数据的安全。这涉及如何在满足AIGC系统的数据需求的同时,尽可能地保护用户的个人信息,避免数据滥用和泄露。研究者探索了匿名化、去标识化等新的数据处理方法,以增强数据安全。同时,也逐步建立了严格的数据治理框架,对数据的收集、使用和存储进行有效管理。

尽管AIGC治理的研究已取得一定进展,但还是面临一些新的挑战。这些挑战在本质上与AIGC的核心特性以及其领域交叉性质有关,需要从监管层面、研究层面和应用层面强化AIGC治理研究。

(1) 在监管层面,AIGC治理将迎来国际协调时期。国际社会需要积极地合作,制定全球性法律框架和规范,以应对AIGC治理的挑战。这将包括制订解决虚假信息、隐私保护和知识产权等问题的国际标准,以确保全球一致性和协同应对。此外,随着对数据隐私的担忧不断增加,未来更多国家会采取法律措施来保护个人数据隐私,包括更严格的数据处理规定和跨境数据流动的监管机制。AIGC伦理法规也将逐渐成为关注焦点,强调算法透明度、公平性和公正性,以引导AIGC技术的发展,确保其服务社会的公益属性。

(2) 在研究层面,多模态和多语言的AIGC将被更多关注,以满足不同文化和语境的需求。可解释性研究将成为一个持续关注的焦点,研究将集中于开发可解释的AIGC模型,以便用户和监管者理解其决策过程。研究人员也将继续探索AIGC的伦理问题,包括权责归属、道德决策和社会影响,这将有助于指导未来AIGC的合理使用和开发。

(3) 在应用层面,产业界需要建立更严格的自律机制,以确保AIGC技术的合规性和道德性,这可能包括制定行业标准和自我监管规则。企业将竞相开发可信的AIGC应用,通过提高可解释性和透明性来获得用户的信任。跨界合作将推动AIGC技术更广泛地应用于各个领域,同时确保其受到充分监管和合规制约。总的来说,AIGC治理将更加国际化,法律法规将更加严格,伦理和可解释性将成为研究的关键点,而行业界将继续追求可信度和创新。这些趋势将共同推动AIGC治理迈向更加成熟和健康的发展方向。

5 结论

AIGC作为一种新型的信息生成方式,其快速发展

对现有的技术治理提出了新的挑战和要求。本文通过文献综述的方法,对AIGC治理研究现状进行了系统的梳理和分析,发现现有研究重点关注了伦理治理、法律治理、技术治理和多主体协同治理这4个方向。通过分析,发现AIGC的治理需要具备灵活性和前瞻性,以应对信息生成方式的特性,如信息量大、技术性高、内容多样且具有迷惑性和隐蔽性。同时,治理模式必须适应这种新的技术环境,通过政府主导、行业自治和媒体宣传的方式,形成一个更为有效的治理体系。

总的来说,本文对AIGC的治理现状进行了全面的分析和总结,揭示了当前面临的挑战和可能的解决方案。这些研究有助于更好地理解AIGC的特性及其对社会、法律和技术产生的影响。这些研究结果对于政策制定者、行业实践者和学术研究者来说都具有重要的参考价值,有助于制定更有效的AIGC治理策略,确保AIGC发展的安全可控。

参考文献

- [1] WU T Y, HE S Z, LIU J P, et al. A brief overview of ChatGPT: the history, status quo and potential future development[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10 (5): 1122-1136.
- [2] PETERS M A, JACKSON L, PAPASTEPHANOU M, et al. AI and the future of humanity: ChatGPT-4, philosophy and education—critical responses[J]. Educational Philosophy and Theory, 2023: 1-35.
- [3] HELBERGER N, DIAKOPOULOS N. ChatGPT and the AI act[J]. Internet Policy Review, 2023, 12 (1): 28.
- [4] 许雪晨,田侃,李文军. 新一代人工智能技术(AIGC): 发展演进、产业机遇及前景展望[J]. 产业经济评论, 2023 (4): 5-22.
- [5] CHENG L, LIU X L. From principles to practices: the intertextual interaction between AI ethical and legal discourses[J]. International Journal of Legal Discourse, 2023, 8 (1): 31-52.
- [6] LIAO Z Q, WANG J, SHI Z Z, et al. Revolutionary potential of ChatGPT in constructing intelligent clinical decision support systems[J]. Annals of Biomedical Engineering, 2023: 1-5.
- [7] 陈兵. 促进生成式人工智能规范发展的法治考量及实践架构: 兼评《生成式人工智能服务管理暂行办法》相关条款[J]. 中国应用法学, 2023 (4): 108-125.
- [8] DWIVEDI Y K, KSHETRI N, HUGHES L, et al. Opinion

- paper: "So what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy[J]. *International Journal of Information Management*, 2023, 71: 102642.
- [9] 张辉, 刘鹏, 姜钧译, 等. ChatGPT: 从技术创新到范式革命[J/OL]. *科学学研究*: 1-15[2023-11-04]. <https://doi.org/10.16192/j.cnki.1003-2053.20230626.001>.
- [10] 张夏恒. 类ChatGPT人工智能技术嵌入数字政府治理: 价值、风险及其防控[J]. *电子政务*, 2023 (4): 45-56.
- [11] 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.
- [12] 王鹏涛, 徐润婕. AIGC介入知识生产下学术出版信任机制的重构研究[J]. *图书情报知识*, 2023, 40 (5): 87-96.
- [13] 蔡子凡, 蔚海燕. 人工智能生成内容(AIGC)的演进历程及其图书馆智慧服务应用场景[J]. *图书馆杂志*, 2023, 42 (4): 34-43, 135.
- [14] 李宏洋. ChatGPT的危与机: AIGC对传播的影响研判以及对策建议[J]. *新闻研究导刊*, 2023, 14 (7): 1-3.
- [15] 徐璐璐, 洪赟, 叶鹰. ChatGPT及GPT类技术的医学信息学应用前景探讨[J]. *情报理论与实践*, 2023, 46 (6): 38-42.
- [16] 徐继敏, 严若冰. 论大语言模型材料的证据属性: 以ChatGPT和文心一言为例[J/OL]. *四川师范大学学报(社会科学版)*: 1-15[2023-11-04]. <https://doi.org/10.13734/j.cnki.1000-5315.2023.0316>.
- [17] CHAN C K Y. A comprehensive AI policy education framework for university teaching and learning[J]. *International Journal of Educational Technology in Higher Education*, 2023, 20 (1): 38.
- [18] LECLER A, DURON L, SOYER P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT[J]. *Diagnostic and Interventional Imaging*, 2023, 104 (6): 269-274.
- [19] SAM S. Early applications of ChatGPT in medical practice, education and research[J]. *Clinical Medicine*, 2023, 23 (3): 278-279.
- [20] 莫祖英, 盘大清, 刘欢, 等. 信息质量视角下AIGC虚假信息问题及根源分析[J]. *图书情报知识*, 2023, 40 (4): 32-40.
- [21] 商建刚. 生成式人工智能风险治理元规则研究[J]. *东方法学*, 2023 (3): 4-17.
- [22] 游腾飞, 曾森. 生成式人工智能时代的网络空间治理[N]. *中国社会科学报*, 2023-09-14 (8).
- [23] CARVALHO I, IVANOV S. ChatGPT for tourism: applications, benefits and risks[J/OL]. *Tourism Review*: 1-2[2023-05-06]. <https://doi.org/10.1108/TR-02-2023-0088>.
- [24] 滕姗姗, 胡奇. 人工智能技术背景下体育传媒产业的数智化变革与挑战[J]. *宁夏社会科学*, 2023 (3): 190-199.
- [25] 斜晓东. 论生成式人工智能的数据安全风险及回应型治理[J]. *东方法学*, 2023 (5): 106-116.
- [26] 樊博. ChatGPT的风险初识及治理对策[J]. *学海*, 2023 (2): 58-63.
- [27] 游俊哲. ChatGPT类生成式人工智能在科研场景中的应用风险与控制措施[J]. *情报理论与实践*, 2023, 46 (6): 24-32.
- [28] 任安麒. 数字出版领域智能语言模型的应用、风险与治理: 基于ChatGPT技术特征的分析[J]. *出版科学*, 2023, 31 (3): 94-102.
- [29] 邹开亮, 刘祖兵. ChatGPT的伦理风险与中国因应制度安排[J]. *海南大学学报(人文社会科学版)*, 2023, 41 (4): 74-84.
- [30] 朱永新, 杨帆. ChatGPT/生成式人工智能与教育创新: 机遇、挑战以及未来[J]. *华东师范大学学报(教育科学版)*, 2023, 41 (7): 1-14.
- [31] 张夏恒. ChatGPT的逻辑解构、影响研判及政策建议[J]. *新疆师范大学学报(哲学社会科学版)*, 2023, 44 (5): 113-123.
- [32] 张璐. 通用人工智能风险治理与监管初探: ChatGPT引发的问题与挑战[J]. *电子政务*, 2023 (9): 14-24.
- [33] LUND B D, WANG T, MANNURU N R, et al. ChatGPT and a new academic reality: artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing[J]. *Journal of the Association for Information Science and Technology*, 2023, 74 (5): 570-581.
- [34] 方卿, 丁靖佳. 人工智能生成内容(AIGC)的三个出版议题[J]. *出版科学*, 2023, 31 (2): 5-10.
- [35] 支振锋. 生成式人工智能大模型的信息内容治理[J]. *政法论坛*, 2023, 41 (4): 34-48.
- [36] 陈智, 陈昊. 可供性视角下ChatGPT的赋能效用及其风险透视[J]. *情报杂志*, 2023, 42 (7): 131-139.
- [37] 喻国明. 生成式内容生产崛起环境下社会协同治理的一项重要举措: 试论全过程式AIGC标识的重要性与必要性[J]. *青年记者*, 2023 (11): 74-76.
- [38] 栾轶玫, 鲁妮. AIGC在重大主题报道中的应用及风控研究[J]. *现代视听*, 2023 (2): 5-9.
- [39] PAUL J, UENO A, DENNIS C. ChatGPT and consumers: benefits, pitfalls and future research agenda[J]. *International Journal of Consumer Studies*, 2023, 47 (4): 1213-1225.

- [40] GRADON K T. Electric sheep on the pastures of disinformation and targeted phishing campaigns: the security implications of ChatGPT[J]. IEEE Security & Privacy, 2023, 21 (3) : 58-61.
- [41] 解学芳, 高嘉琪. AIGC模式赋能数字文化创新的逻辑与善治: 基于ChatGPT热潮的思考[J]. 江海学刊, 2023 (3) : 86-95.
- [42] 马乐存, 詹希旎, 朱齐宇, 等. AIGC驱动的GLAM数智融合创新发展研究[J]. 农业图书情报学报, 2023, 35 (5) : 4-15.
- [43] 王建磊, 江浩. AIGC与融媒生产: 赋能、互补与边界[J]. 视听界, 2023 (3) : 21-25.
- [44] 谢梅, 王世龙. ChatGPT出圈后人工智能生成内容的风险类型及其治理[J]. 新闻界, 2023 (8) : 51-60.
- [45] 王芳, 朱文韵. ChatGPT的风险分析及国际治理趋势[J]. 竞争情报, 2023, 19 (3) : 40-42.
- [46] 张欣. 生成式人工智能的算法治理挑战与治理型监管[J]. 现代法学, 2023, 45 (3) : 108-123.
- [47] 赵精武. 生成式人工智能应用风险治理的理论误区与路径转向[J]. 荆楚法学, 2023 (3) : 47-58.
- [48] 舒洪水, 彭鹏. ChatGPT场景下虚假信息的法律风险与对策[J]. 新疆师范大学学报(哲学社会科学版), 2023, 44 (5) : 124-129.
- [49] 罗昕. 聊天机器人的网络传播生态风险及其治理: 以ChatGPT为例[J]. 青年记者, 2023 (7) : 91-94.
- [50] 刘艳红. 生成式人工智能的三大安全风险及法律规制: 以ChatGPT为例[J]. 东方法学, 2023 (4) : 29-43.
- [51] LIN X. Exploring the role of ChatGPT as a facilitator for motivating self-directed learning among adult learners[J/OL]. Adult Learning: 1[2023-05-06]. <https://journals.sagepub.com/doi/10.1177/10451595231184928>.

作者简介

邓胜利, 男, 博士, 教授, 研究方向: 用户信息行为。

汪璠, 男, 博士研究生, 通信作者, 研究方向: 生成式人工智能、虚假信息, E-mail: 1161252028@qq.com。

Research Progress and Development Trend of AIGC Governance

DENG ShengLi WANG Fan

(School of Information Management, Wuhan University, Wuhan 430072, P. R. China)

Abstract: In this paper, we comprehensively analyze the governance research of artificial intelligence generated content (AIGC) at home and abroad, discover the research hotspots and problems of AIGC governance, clarify the future direction of development, and provide guidance for the change and enhancement of AIGC governance strategies in the new AI environment. Through a detailed analysis of existing studies at home and abroad, we explore the new governance requirements brought by AIGC and how to establish a multi-subject governance system to cope with them from multiple dimensions, such as technological governance, legal governance, ethical governance, and collaborative governance. First, the emergence of AIGC reinforces the complexity and hidden nature of information, which requires advanced AI models and a strict quality management system to ensure that the generated content is ethically acceptable and conforms to the values and moral standards of society and individuals. Second, on the legal side, we focus on multiple legal issues such as copyright, privacy, disinformation, and algorithmic discrimination, and anticipate, study, and formulate relevant regulations in advance. Third, in terms of technology governance, it is necessary to pay attention to several key aspects such as algorithm transparency and interpretability, data protection and privacy, restrictions on the use scenarios of the technology, and user guidance. Finally, multi-party collaborative governance allows the government, industry, and media to participate, combining technical means and laws and regulations to jointly form a flexible and efficient governance framework. In conclusion, this paper reveals the current situation and trend of governance research in the AIGC era, and proposes a series of response strategies to provide guidance for the future governance of AIGC.

Keywords: AIGC; Collaborative Multi-Party Governance; Legal Governance; Technological Governance; Ethical Governance

(责任编辑: 王玮)