生成式人工智能的规范发展：风险、监管与标准化

王 淼[1]　朱思婧[2]

（1.中国标准化研究院；2.商务部国际贸易经济合作研究院）

摘　要：生成式人工智能（AIGC）技术，以其卓越的内容生成能力，正在重塑信息生态并推动商业和产业的快速发展。本文深入分析了AIGC技术的原理、应用进展，并对其带来的隐私泄露和滥用等风险进行了全面评估。鉴于AIGC技术的全球影响力，文章重点讨论了国际和国内监管策略与标准化工作的现状，并提出了相关策略建议。这些建议旨在促进AIGC技术的规范发展，同时确保技术进步与社会利益的和谐统一，推动其在负责任和安全的基础上实现社会价值。

关键词：生成式人工智能，标准化，风险管理，国际监管

DOI编码：10.3969/j.issn.1674-5698.2024.11.010

Normative Development of Generative Artificial Intelligence: Risk, Regulation and Standardization

WANG Miao[1]　ZHU Si-qi[2]

（1. China National Institute of Standardization; 2. Chinese Academy of International Trade and Economic Cooperation）

Abstract: Generative Artificial Intelligence (AIGC) technology, with its superior content generation capabilities, is reshaping the information ecology and driving the rapid development of business and industry. This paper analyzes the principles and application progress of AIGC technology, and comprehensively evaluates the risks of privacy disclosure and abuse brought by AIGC technology. In view of the global influence of AIGC technology, this paper focuses on the current status of international and domestic regulatory strategies and standardization efforts, and proposes relevant strategies. These recommendations aim to promote the normative development of AIGC technology, while ensuring the harmony of technological progress and social interests, and promote its realization of social value on a responsible and safe basis.

Keywords: generative artificial intelligence, standardization, risk management, international regulation

基金项目：本文受中央基本科研业务费项目“技术性贸易措施影响评估方法与实证研究”（项目编号：292023Y-10407）资助。

作者简介：王淼，研究实习员，主要研究方向为技术性贸易措施。

朱思婧，商务部国际贸易经济合作研究院，硕士研究生。

0 引言

随着信息技术的突飞猛进，生成式人工智能（AIGC）技术已经成为加速商业化和产业化进程的新动力源泉。AIGC技术凭借其卓越的内容生成能力，不仅极大地拓展了信息生态的边界，而且在社会各个层面激起了广泛而深入的讨论。从ChatGPT的广泛流行到层出不穷的AIGC工具，AIGC技术无疑占据了信息技术革新的制高点，引领着一场内容创造和分发的新浪潮。然而，正如历史上每一次重大技术突破一样，AIGC技术的颠覆性同样伴随着隐私泄露、滥用等风险问题的出现。这些问题因其复杂性和全球性，已经引起了技术风险研究机构、政府和国际组织的高度重视。2023年11月1日至2日，首届全球人工智能（AI）安全峰会在英国举行，中国等28个国家和欧盟共同签署了《布莱切利宣言》（以下简称《宣言》）。《宣言》的签署表明了各国对人工智能治理问题的共同关注，不但促使各国在确定治理方向与建立监管底线等方面形成了共识，也推动了未来人工智能领域国际标准的形成。

本文致力于深入探讨AIGC技术的发展背景，全面评估其带来的风险，并探索有效的治理策略。通过深入分析AIGC技术的原理、应用进展以及潜在风险，旨在为AIGC技术的规范发展提供坚实的理论基础和实用的实践指导。只有深入理解AIGC技术的内在机制和外部影响，才能更好地把握其发展趋势，制定合理的政策和措施，确保技术进步与社会利益的和谐统一。

1 AIGC技术的原理与风险

1.1 AIGC技术原理与应用进展

生成式人工智能（AIGC）技术正以其革命性的算法模型和卓越的数据处理能力，在多个领域内实现突破性发展。AIGC技术的核心在于深度学习，它通过分析和学习大量的数据集，使算法能够创造出新颖的、与训练数据具有相似特征的内容。这种技术不仅极大地提升了内容创作的效率，还为各行各业带来了前所未有的创新机遇。

在文本创作领域，AIGC技术能够根据给定的主题或关键词，快速生成连贯、有逻辑性的文章和报告。在图像生成方面，它能够创造出逼真的图像和艺术作品，甚至在某些情况下达到以假乱真的水平。此外，在复杂的策略制定和问题解决中，AIGC技术通过模拟和预测，为决策者提供了强有力的支持。

随着AIGC技术的不断进步，其应用案例也日益丰富。从个性化新闻推荐系统到智能客服机器人，再到医疗诊断辅助工具，AIGC技术正逐步渗透到日常生活的方方面面，展现出其巨大的实用价值和发展潜力。

1.2 技术带来的多维风险审视

尽管AIGC技术展现出巨大的潜力和应用前景，但其伴随的风险同样不容忽视。隐私泄露问题尤为突出，因为AIGC技术在生成内容的过程中，可能会无意中泄露训练数据中的敏感信息，对个人隐私构成威胁。

技术的滥用问题也引起了社会的广泛关注。虚假信息的生成可能导致社会信任的瓦解，知识产权的侵犯可能损害创作者的合法权益，对特定群体的歧视可能加剧社会不公。这些风险不仅威胁到个人和社会的安全，也对AIGC技术的健康发展构成了严峻挑战。

伦理问题是AIGC技术发展中不可忽视的重要议题。随着技术的发展，如何确保AIGC工具的决策过程透明、公正，以及如何避免算法偏见，成为亟待解决的问题。国际标准化组织（ISO）、国际电工委员会（IEC）和国际电信联盟（ITU）等国际组织正在积极制定相关标准和指导原则，旨在推动AIGC技术的负责任、可信赖和安全发展。

此外，公众对AIGC技术的认知和接受程度也是影响其可持续发展的关键因素。因此，加强公众教育，提高社会对AIGC技术潜在风险的认识，以及培养负责任的使用习惯，对于构建健康、可持续的技术生态环境至关重要。

1.3 AIGC安全治理方案

1.3.1 以标准化支撑人工智能安全治理

《宣言》指出，发挥人工智能的潜力需要政府、企业、学术界和社会的共同关注、协作和信息共享。标准和资源共享可以促进政府间的能力建设、国际合作以及对人工智能风险和应对措施的共同理解，从而通过共同努力克服全球性挑战。作为人工智能安全研究合作的初步贡献，英国成立了世界上第一个人工智能安全研究所，该研究所将建立公共部门进行人工智能安全测试和研究，对可能出现的安全风险进行预判。尽管前沿人工智能公司在人工智能安全政策方面取得了重大进展，包括责任承担和风险告知方面，但是公司政策只是基线，并不能取代政府制定标准和监管的需要。标准化基准可以由可信的外部第三方（如：最近宣布成立的英国人工智能安全研究所）提供。

1.3.2 以互操作性应用人工智能安全治理

人工智能互操作性的实现需要依托标准化、共同的原则和准则通过制定人工智能可互操作的框架可以有效减轻前沿人工智能的风险，并促进广泛实现人工智能的惠益。从国内来看，互操作性的实现需要根据国情和适用的法律框架制定有针对性的方法。从国际来看，经济合作与发展组织（OECD）和全球人工智能伙伴关系（GPAI）等组织在具有互操作性的人工智能开发、应用和治理方面提供了详细依据和政策指导。标准机构，包括国际标准化组织（ISO）、国际电工委员会（IEC）、电气和电子工程师协会（IEEE）和国际电信联盟（ITU）等为应对人工智能安全风险，也开展了人工智能安全各个方面的标准化工作。

2　全球与中国关于AIGC技术的监管与标准化现状

2.1　国际社会对AIGC风险的监管策略和标准化工作

生成式人工智能（AIGC）技术的快速发展，引起了国际社会的广泛关注。AIGC技术的潜在风险，特别是隐私泄露和滥用问题，已经促使多个国家和国际组织采取行动。生命未来研究所（FLI）的公开信就是一个明显的例子，它不仅呼吁全球AI实验室暂停训练更强大的AI系统，还强调了开发AI治理系统的紧迫性。国际社会对AIGC风险也表达了关切，这封公开信得到了包括埃隆·马斯克在内的众多科技界人士的支持。此外，英国政府发布的人工智能产业监管白皮书和欧洲议会通过的《通用产品安全条例》（GPSR）进一步体现了国际社会在监管AIGC风险方面的积极努力：这些政策文件不仅为AIGC技术的监管提供了框架，也为国际合作和协调奠定了基础。

在标准化方面，2022年，英国建立人工智能标准中心，为国际标准化工作提供信息，在制定“安全可靠”的规则方面发挥作用。2023年11月26日，美国、英国、澳大利亚等18个国家的网络安全监管部门联合微软、谷歌等23个网络安全组织发布《安全AI系统开发指引》，旨在从设计阶段强化AI 系统的安全性，以防范可能的安全风险。该指南被视为全球首个AI安全标准，并为AI系统开发提供了必要的建议。国际标准化组织（ISO）、国际电工委员会（IEC）和国际电信联盟（ITU）等机构也正在积极开展工作。ISO/IEC JTC 1/SC 42分委会正在制定一系列标准，这些标准将涵盖整个AI生态系统，包括技术治理、风险管理、伦理要求等关键领域。这些标准的制定有望为AIGC技术的规范发展提供重要的指导和支持。

2.2 中国对AIGC风险的监管策略和标准化工作

中国作为一个在AIGC技术领域具有重要影响力的国家，也在积极构建自己的监管框架和推动标准化工作。2023年10月11日，我国发布《生成式人工智能服务安全基本要求》（征求意见稿），这是全国信息安全标准化技术委员会发布的国内首个专门面向生成式AI安全领域的规范意见稿。征求意见稿首次提出生成式AI服务提供者需遵循的安全基本要求，涉及语料安全、模型安全、安全措施、安全评估等方面，给出了语料及生成内容的主要安全风险共5类31种。尽管这些标准目前不具备法律效力，但相关专家认为它们很可能会被纳入未来的法律框架中，并已被科技公司如：华为、阿里巴巴和腾讯等视为具有约束力的规则。这份文件的出台不仅提供了内容审核的技术性细节，还可能预示着新的审查制度的开始，从而在全球范围

内对AI的监管方式产生指导作用。

2.3 国际与国内监管策略和标准化工作的差异与联系

尽管国际社会和中国都在积极推动AIGC技术的监管和标准化工作，但在策略和实施上存在一些差异。例如：一些欧洲国家如意大利已经采取了明确的禁限用措施，而美国则更侧重于通过公开征求意见来推进监管措施的制定。与此同时，中国在监管策略上更侧重于行业倡议和政策引导，同时在标准化工作中强调科技伦理的重要性。这些差异不仅反映了不同国家在监管理念和方法上的特点，也与各国在AIGC技术发展阶段和应用场景上的差异有关。然而，无论是国际还是国内，监管和标准化工作的最终目标都是促进AIGC技术的健康发展，保障社会安全和伦理要求得到满足。

在全球化的今天，国际合作与交流在AIGC技术的监管和标准化工作中显得尤为重要。通过分享最佳实践、协调标准制定和加强技术交流，国际社会和中国需要共同应对AIGC技术带来的挑战，推动构建一个更加安全、负责任的AI技术发展环境。

3 加快制定完善我国AI治理相关标准的策略建议

随着人工智能技术的快速发展，AI治理标准的重要性日益凸显。为了确保AI技术的健康发展，避免潜在风险，我国亟需加快制定和完善AI治理相关标准。本文提出了以下策略建议。

3.1 优化AI治理标准体系建设

持续优化AI治理标准体系的建设意味着需要建立一个全面、协调、高效的标准体系，涵盖AI技术的研发、应用、评估、监管等各个环节。标准体系的建设应以促进技术创新、保障社会安全、维护伦理道德为基本原则。为了加强重点领域标准布局，我国应识别和确定AI技术发展的关键领域，如：自动驾驶、医疗诊断、智能教育等，并针对这些领域制定相应的标准。这些标准应包括技术规范、安全要求、伦理准则等，以确保AI技术的应用既高效又安全。

3.2 开展AI评测类标准研制

我国应稳步开展AI风险、隐私、伦理、可信等方面的评测类标准研制。这些标准将为AI技术的安全性、隐私保护、伦理合规性、可信度等方面提供评估依据。在风险管理方面，应制定相应的评估标准，以识别和评估AI技术可能带来的各种风险，包括技术风险、安全风险、社会风险等。在隐私保护方面，应制定严格的数据收集、处理、存储和传输标准，以保护用户的个人隐私。在伦理方面，应制定AI伦理准则，明确AI技术在设计、研发、应用过程中应遵循的伦理原则，如：尊重人权、公平公正、透明可解释等。在可信度方面，应制定AI系统的可信度评估标准，评估AI系统的可靠性、稳定性、抗干扰能力等。此外，在前沿AI技术方面，重点关注人工智能关键技术、核心产品和应用迭代等标准化研究，重视技术赋能，为人工智能在国内和国际的安全治理方面提供标准化技术支撑。

3.3 促进AI领域的互操作性与协同发展

互操作性是确保不同系统和平台之间无缝协作的基础。在AI领域，互操作性不仅涉及数据格式和接口的标准化，还包括算法、模型和应用程序之间的兼容性。通过制定统一的标准和协议，可以降低技术整合的复杂性和成本，提高AI系统的整体协同作用和响应速度。为了实现这一目标，需要政府、企业和学术界的共同努力。政府可以制定相应的政策和激励措施，鼓励数据共享和技术创新；企业可以通过合作和竞争，推动数据的开放和互操作性技术的发展；学术界则可以通过研究和教育，培养新一代的人工智能专业人才，为数据共享和互操作性提供理论和实践的支持。

3.4 促进国际合作与交流

在制定和完善AI治理标准的过程中，我国还应积极参与国际合作与交流。通过与国际标准化组织、其他国家的监管机构和AI企业等进行交流合作，我国可以学习借鉴国际先进的标准制定经验和监管实践，同时也可以分享我国在AI治理方面的成果和经验。国际合作与交流有助于促进全球AI治理标准的协调一致，避免因标准不一致而带来的技术壁垒和贸易摩擦。此外，国际合作还可以

促进AI技术的全球创新和发展，为解决全球性问题提供技术支持。

4 结论

深入理解AIGC技术的内在机制和外部影响，把握其发展趋势，制定合理的政策和措施，才能努力做到技术进步与社会利益的和谐统一。通过对AIGC技术原理的阐释、应用进展的梳理以及潜在风险的评估，我们对这一新兴技术有了更深刻的理解。展望未来，AIGC技术的演进预计将在促进社会进步和经济发展中扮演更加关键的角色。技术的不断进步将为各行各业带来深远的影响，从而推动创新和生产力的提升。然而，技术的健康发展需要一个健全的监管框架和标准化体系作为支撑。这不仅对于缓解技术风险、保护消费者和企业利益至关重要，也是确保AIGC技术长期可持续发展的关键。

参考文献

[1] 叶海波,相梦垚.我国人工智能标准化建设现状与展望[J].质量与认证,2024(03):28–30.DOI:10.16691/j.cnki.10–1214/t.2024.03.003.

[2] 全球人工智能专家共同探讨挑战、趋势和标准化（英文）[J].China Standardization,2024(03):58.

[3] 徐慧芳,秦铭浩,王毓欣,等.基于标准必要专利的人工智能产业竞争态势研究[J].中国发明与专利,2024,21(05):48–55.

[4] 林阳荟晨,上官晓丽.欧美人工智能网络安全标准化最新动态[J].信息技术与标准化,2023(12):57–61+66.

[5] 宋恺,屈蕾蕾,杨萌科.生成式人工智能的治理策略研究[J].信息通信技术与政策,2023,49(07):83–88.

[6] 余继超.数据安全“拷问”ChatGPT类AI[N].国际金融报,2023–04–17(011).DOI:10.28403/n.cnki.nifnb.2023.000394.