生成式人工智能给信息安全领域带来的挑战与应对策略探究

吕思宇，白伟光，韩雨希
北京航天测控技术有限公司，北京 102209
DOI: 10.61369/TACS.2025110035

摘　　要： 随着生成式人工智能技术的迅猛发展，其以多元的服务形态、革新的交互方式和复杂的服务模式，在各行业领域实现广泛渗透，推动社会生产生活方式发生深刻变革。基于此，本文剖析生成式人工智能的核心特点，系统探究其给信息安全领域带来的具体挑战，并提出针对性的应对策略，旨在为化解生成式人工智能背景下的信息安全风险、推动技术与信息安全领域协同健康发展提供理论参考与实践指引。

关 键 词： 生成式人工智能；信息安全；挑战；应对策略

Exploration of Challenges and Countermeasures Brought by Generative Artificial Intelligence to the Field of Information Security

Lv Siyu, Bai Weiguang, Han Yuxi
Beijing Aerospace Measurement and Control Technology Co., Ltd., Beijing 102209

Abstract: With the rapid development of generative artificial intelligence technology, it has achieved widespread penetration in various industries and fields through diverse service forms, innovative interaction methods, and complex service models, driving profound changes in social production and lifestyle. Based on this, this paper analyzes the core characteristics of generative artificial intelligence, systematically explores the specific challenges it brings to the field of information security, and proposes targeted countermeasures. It aims to provide theoretical references and practical guidance for addressing information security risks in the context of generative artificial intelligence and promoting the coordinated and healthy development of technology and the information security field.

Keywords: generative artificial intelligence; information security; challenges; countermeasures

引言

生成式人工智能能够基于海量数据进行学习训练，自主生成文本、图像、音频、视频等多样化内容，极大地提升了生产效率，拓展了人类创新的边界。但与此同时，技术的快速迭代与广泛应用也打破了传统信息安全的防护格局，带来了一系列全新的安全风险与挑战。在数字经济时代，信息已成为核心生产要素，信息安全直接关系到个人权益、企业发展和国家战略安全。生成式人工智能的出现，使得信息生成的门槛大幅降低，信息传播的速度和范围显著扩大，传统的信息鉴别、风险防控手段难以有效适配。在此背景下，深入探究生成式人工智能给信息安全领域带来的挑战，构建科学有效的应对策略，具有重要意义[1]。

一、生成式人工智能的特点

（一）服务形态多元

生成式人工智能的服务形态可以突破传统技术固有的单一模式，在不同的应用场景中能够产生丰富的信息资源，涉及各个领域的内容生产工作，如在文字内容方面能够在很短的时间内产出新闻稿、学术论文、宣传文案以及法律文书等多种类型的文字，并且这些文字具有清晰的逻辑性和精美的语言风格。让人难以分辨是人还是机器创造出来的；在图像方面，输入简单的指令，就可以生成逼真的风景、人物或者产品图片，还可以模仿某个画家的风格进行作画；在音频视频方面，合成出精准的语音，并能生成故事脚本及高品质的音视频资源，完成了“从剧本到电影”的快速转化。

（二）交互方式革新

生成式人工智能彻底革新了人机交互方式，打破了传统人工智能技术中指令式、单一化的交互模式，实现了自然语言驱动的沉浸式、对话式交互。传统人工智能应用往往需要用户输入特定格式的指令或参数，交互过程较为烦琐，对用户的技术门槛要求较高[2]；而生成式人工智能支持用户以日常化的自然语言进行提问、下达指令，系统能够精准理解用户的意图，通过多轮对话不

断优化回应内容，实现与用户的深度互动。

（三）服务模式复杂

生成式人工智能的服务模式呈现出高度复杂的特点，主要体现在技术架构、数据来源、盈利模式等多个方面。在技术架构上，生成式人工智能通常采用大规模预训练模型，需要海量的计算资源和数据进行训练优化，涉及模型训练、微调、部署等多个环节，各环节之间相互关联、相互影响，形成了复杂的技术体系。而在数据源端环节，生成式人工智能模型的训练就需要从互联网、企业数据库、公共数据中心等不同来源获得海量数据，数据类型多样而分散，但对数据合法性和隐私性保护难以实现[3]。各个环节环环相扣又相互影响，构成了一个复杂的技术体系。

二、生成式人工智能给信息安全领域带来的挑战

（一）个人信息泄露

生成式人工智能的发展离不开海量数据的支撑，而这些数据中往往包含大量个人敏感信息，这使得个人信息泄露风险大幅提升，成为生成式人工智能带来的首要信息安全挑战。一方面，在模型训练阶段，生成式人工智能企业需要收集大量来自互联网、用户授权的个人数据，包括姓名、身份证号、联系方式、消费记录、健康信息、社交关系等。另一方面，在交互使用阶段，用户在与生成式人工智能系统进行对话交互时，可能会不自觉地泄露个人敏感信息[4]。而生成式人工智能系统如果缺乏完善的信息过滤和安全防护机制，这些敏感信息可能会被存储、滥用，甚至被第三方窃取。

（二）虚假信息传播

在强大的生成式人工智能生成能力加持下，虚假信息的生产门槛变得非常低，覆盖范围也变得极为广泛，严重扭曲了社会舆论场与信息传播空间，成为继隐私泄露后的人身安全又一大隐患。以往制作谣言需要一定的技术含量以及耗费较多的人力物力，但借助 AI 可以迅速生成高度仿真的文字、图像、语音或者视频等虚假信息，并且成本极低，人人都能轻松实现。

（三）算法安全实施

生成式人工智能的核心是算法模型，而算法本身存在的黑盒性、训练材料的不均衡性和在使用过程中暴露出来的不足之处使得算法安全问题变得越来越严重，成为新出现的信息安全隐患。生成式 AI 算法具有较高程度的黑盒特性，难以完全被人类理解和描述其判断过程及其产出的内容之间的关系逻辑。这就导致其在生成内容过程中可能出现不可预测的偏差，同时溯源较为困难，给算法安全管控带来极大挑战。

三、生成式人工智能应用于信息安全领域的应对策略

（一）加大监管治理力度，完善相关机制保障

在生成式人工智能技术应用过程中，应加大监督管理治理力度，构建起科学完善的监管体系，完善相关法律法规和工作机制，这样来有效管控技术发展和应用。第一，完善生成式人工智能相关法律法规。根据生成式人工智能的特点，应修改和完善现行的法律规范，明确新型人工智能公司的权利和义务，在收集信息、训练模型、生成内容、提供服务等环节中做出行为的规定[5]。第二，构建多部门协同监管机制。生成式人工智能的监管涉及网信办、工信部、公安部、市场监管总局等多个职能机构，应当明确各部门之间的责任划分并且建立多部门间的信息互通机制、联合执法法规制度以及案件移送程序来达到良好的监管目的。此外还可以引入行业组织机构或第三方测评机构进行监管活动，形成以政府部门管理方式、企业自律行为准则以及社会大众监督三位一体的综合监测体系[6]。第三，建立生成式人工智能安全评估和备案制度。对于使用生成式人工智能的企业而言，在其模型投入运营前应进行全链路安全测试，重点关注数据安全、算法安全以及内容的安全性等问题，并在通过相关测试后方可投入运营；落实生成式人工智能服务备案管理制度，企业应向监管部门报送自身业务范围、架构体系、数据来源及安全保障措施等内容的信息，便于监管部门对企业的动态进行监管，及时发现并处置风险点。

（二）设置统一标准规范，推进市场监督管理

统一的标准规范能够为人工智能发展和防范信息安全风险提供重要依据。因此，应制定出生成式 AI 在数据收集、模型训练等方面的标准，以推动市场监督管理的规范化进行。第一，制定数据安全标准。明确生成式人工智能训练数据范围、合法性原则以及数据安全规则等内容，并规范数据收集、存储、流通、应用行为[7]。第二，制定算法安全标准。针对生成式人工智能技术算法的黑箱特性，需要制定相关的人工智能易懂性、公平性和安全性标准，希望企业增加人工智能技术透明度并进行可追溯管理。同时，还需要制定相关的算法防护测评规范来引导算法抗破坏测试、漏洞发现等工作，从而提升了算法模型的安全性及可靠性。第三，制定内容生成标准。要明确界定生成式人工智能所产生内容的合规性、真实性、健康度要求并制定相关规则解决谣言识别、不良信息过滤等问题。同时也要加大对生成式人工智能服务的监督力度，严格防范虚假宣传、侵权盗版、传播违法和不良信息等行为，维护市场的稳定性和顾客的利益[8]。

（三）稳步推进技术革新，发挥人工智能优势

技术革新能够提升 AI 信息安全风险管控效果，相关部门应加大技术研发力度，使得安全防护技术进行创新应用，构建起智能化的信息安全防护体系，切实发挥人工智能技术的应用优势。第一，加强生成式人工智能安全防护技术研发。在数据安全、联邦学习、对抗样本、漏洞挖掘等方面对大模型进行保护性研究工作，提高生成式 AI 全流程安全性。第二，发展虚假信息识别与溯源技术。致力于构建以 AI+ 大数据 + 区块链为基础的技术平台，提升对虚假新闻的识别能力以及效率，并能够从文字、图像、音频、视频中分析其中的虚假信息。利用区块链技术记录传播源和过程，这样就完成了谣言传播者的溯源。同时也将这些新技术应用到了商业化落地，为政府部门、媒体企业、普通消费者等提供辟谣服务[9]。第三，发挥人工智能在信息安全防护中的优势。运用新型人工智能提升信息安全防护的智能化程度，例如开发基于智能的网络安全监测平台，对网络攻击事件进行实时监控，并对安

全风险进行预示；使用新型人工智能生成仿真攻击信息用于安全防御系统的测试及优化等。

（四）注重行业人才培养，提供技术应用支持

人才是应对生成式人工智能信息安全挑战的重要支撑，应加强对行业人才的培养，培养出了解生成式人工智能技术和信息安全知识的复合型人才，为技术应用和安全防护提供人才保障。第一，完善人才培养体系。高等院校以及职业院校应该根据市场需求设置相应的专业，增加生成式 AI、信息安全、信息保密等相关的课程培养复合型人才。同时加强校企合作，建立实训基地，让学弟能够参与实际的研究项目并从事网络安保工作。从而增加他们的实践经验，并可通过开展培训、技能竞赛等方式对现有从业者进行相关领域技能培养。第二，加强人才引进和激励。为吸引国内国外优秀的生成 AI 和网络安全人才参与进来，应提供相应优惠政策和人才引进机制，并对从事科学技术和安全保障方面做出突出贡献的人员给予表彰和奖励。同时，营造良好的人才环境，为他们提供广阔的事业平台和良好的工作条件，留住关键人才[10]。第三，推动行业交流与合作。举办行业峰会、学术论坛等，搭建人才交流平台，推动生成式人工智能、信息安全等领域的人才交流与协作；同时加强对外合作，引进国外先进技术和管理理念，提升我国相关领域人才的国际视野及专业水平。

四、结束语

综上所述，生成式人工智能正以前所未有的速度重塑信息技术版图。生成式 AI 与信息安全的博弈，本质是创新与秩序的辩证统一，既需正视其带来的数据泄露、深度伪造、模型劫持等系统性风险，更要构建涵盖技术防御、管理规范与法律治理的立体应对体系。随着生成式人工智能技术的不断发展，信息安全领域的挑战也将不断演变，应持续关注技术发展动态，及时调整应对策略，不断完善信息安全防护体系，推动生成式人工智能技术在保障信息安全的前提下实现健康有序发展，发挥其积极作用。

参考文献

[1] 赵洪波，李子超 . 生成式人工智能赋能体育教学高质量发展的价值、挑战与路径 [J]. 体育科技文献通报 ,2024,32(12):153-156.DOI:10.19379/j.cnki.issn.1005-0256.2024.12.037.

[2] 徐伟，李文敏 . 生成式人工智能的依法治理——基于法规文本的扎根理论运用 [J]. 岭南学刊 ,2025,(01):107-117.DOI:10.13977/j.cnki.lnxk.2025.01.010.

[3] 蔡智权，张娟 . 生成式 AI 嵌入数字政府的价值审思与路径展望 [J]. 哈尔滨工业大学学报 (社会科学版),2024,26(06):151-160.DOI:10.16822/j.cnki.hitskb.2024.06.015.

[4] 林韶 . 生成式 AI 应用背景下个人信息安全的法律挑战与治理——以网络化开放创新范式为视角 [J]. 西部法学评论 ,2024,(04):10-22.

[5] 高海燕，张优良 . 英美著名高校的生成式人工智能政策及其启示 [J]. 现代教育技术 ,2024,34(07):42-50.

[6] 宋晓红，龚程程 . 人工智能背景下个人信息保护的公益诉讼制度基础和创新路径 [C]// 上海市法学会 .《智慧法治》集刊 2024 年第 1 卷——2024 年世界人工智能大会法治论坛文集 . 上海市徐汇区人民检察院 ;,2024:283-291.DOI:10.26914/c.cnkihy.2024.009985.

[7] 潘建红，祝玲玲 . 生成式人工智能赋能高校思政课的风险生成及规避 [J]. 思想政治教育研究 ,2024,40(03):94-100.DOI:10.15938/j.cnki.iper.2024.03.014.

[8] 王新，王宣亿，黄晓亮，等 .ChatGPT 类生成式人工智能的法律风险及治理 (笔谈)[J]. 天津师范大学学报 (社会科学版),2024,(03):21-38.DOI:10.20247/j.issn1671-1106.2024.03.016.

[9] 曹乐廷 . 面向生成式人工智能信息安全问题的协同监管机制研究 [D]. 南京工业大学 ,2024.DOI:10.27238/d.cnki.gnjhu.2024.000097.

[10] 王大志，张挺 . 风险、困境与对策：生成式人工智能带来的个人信息安全挑战与法律规制 [J]. 昆明理工大学学报 (社会科学版),2023,23(05):8-17.DOI:10.16112/j.cnki.53-1160/c.2023.05.212.

