文章编号：2095-2163(2024)01-0224-04 中图分类号：TP181；G203 文献标志码：A

AIGC 驱动下虚拟数字人的社会安全风险及其治理策略研究

徐英健
（中国人民公安大学 治安学院，北京 100038）

摘 要：以 ChatGPT 为代表的人工智能生成内容（AI Generated Content，AIGC）技术的突破，解决了虚拟数字人在诸多方面的应用难题，使得虚拟数字人呈现出井喷发展态势，也给元宇宙相关产业链带来全新的想象空间。与此同时，一系列社会安全风险也随之而来，数据和个人信息安全风险、潜在滥用和虚假信息传播风险、社会影响和道德伦理风险、法律和监管等问题逐渐浮现，给社会秩序和个人权益带来了挑战。基于此，应该加强技术层面的风险治理能力、加强公众教育和社会参与、完善相关法律法规和行业规范以及构建多元化主体参与的监管体系，来积极应对并化解风险。

关键词：人工智能；AIGC；互联网；社会安全风险治理

Research on social security risks and governance strategy of virtual digital humans driven by AIGC

XU Yingjian
（School of Rublic Order，People's Public Security University of China，Beijing 100038，China）

Abstract：The breakthrough of AIGC （AI Generated Content） technology represented by ChatGPT has solved the application problems of virtual digital people in many aspects，making virtual digital people show a blowout development trend，and bringing new imagination space to the industry chain related to the metaverse. At the same time，a series of social security risks have followed，such as data and personal information security risks，potential abuse and false information dissemination risks，social impact and moral and ethical risks，legal and regulatory issues gradually emerge，which bring challenges to social order and personal rights and interests. Based on this，it is necessary to strengthen the risk governance ability at the technical level，strengthen public education and social participation，improve relevant laws and regulations and industry norms，and build a regulatory system with diversified subject participation to actively cope with and defuse risks.

Key words：artificial intelligence；AIGC；internet；governance of social security risks

0 引 言

随着信息技术的快速发展和数字化社会的崛起，元宇宙作为一个虚拟的、多维度的数字世界，正逐渐成为人们日常生活的一部分。在元宇宙中，人们可以通过虚拟现实技术与数字化的环境进行互动、创造和体验丰富多样的虚拟场景和虚拟角色。其中，虚拟数字人作为元宇宙中的虚拟实体[1]，正逐渐成为社交媒体、虚拟现实和在线平台中的重要参与者。据 IDC（国际数据公司）发布的《2022 年中国 AI 数字人市场现状与机会分析》报告预测：到 2026 年中国 AI 数字人市场规模将达 102.4 亿元人民币。《虚拟人深度产业报告》指出：中国虚拟数字人的市场规模到 2030 年将达到 2 700 亿元。可以说，近年来虚拟数字人发展备受关注且市场规模呈现高速增长的态势。当前，AIGC 正通过其强大的内容生成能力，不断推动数字人产业的创新和发展，未来，数字人产业一定会迎来更加广阔的发展空间和机遇[2]。技术的进步往往是一把双刃剑，人们在看到其对社会生产力带来的巨大推动之外，也需未雨绸缪，充分关注其对社会的多元影响，尤其是加强对其风险的认知以及治理策略的研究。

1 AIGC 对虚拟数字人的全面赋能

AIGC 全称为 AI Generated Content，即生成式 AI 或者人工智能生成内容，是继 PGC 和 UGC 之后的一种新型内容创作方式[3]。AIGC 被认为是元宇宙和 web3.0 的底层基础设施之一，目前已广泛应用于游戏开发、数据分析、计算机图形学、艺术创作等多个领域。AIGC[4] 技术在自然语言、图像、声音等多模态处

作者简介：徐英健（1998-），男，硕士研究生，主要研究方向：治安管理。

收稿日期：2023-12-04

理方面的快速突破，正在不断驱动虚拟数字人产业升级，推动实现元宇宙场景下虚拟数字人更加智能化的人机交互与信息共享。AIGC 既直接可以生成数字人，也可以为数字人提供与外界交互的信息内容，进而增加数字人在互动需求场景的适配性[5]。AIGC 对虚拟数字人的赋能主要体现在以下几个方面：

1.1　生成模式

AIGC 推动了数字人生成模式的不断创新。借助 AIGC 不仅降低了教字人的制作成本、缩短了制作周期，其生成的数字人形象更加逼真。同时 AIGC 还可以驱动调整数字人的口型、微表情、声音，建立起与文本的映射关系，使得数字人从“形似”走向“神似”。当前，基于 AIGC 技术有 3 种主流的数字人生成模式——照片自动生成形象、视频自动生成形象以及 GLB 模型自动生成形象。例如，2023 年 4 月，腾讯发布的名为“腾讯智影” AI 智能创作助手，只需根据少量图片、音频和视频素材，即可快速制作一个数字人分身，并根据要求直接生成视频内容。

1.2　交互能力

AIGC 技术赋予数字人更强的交互能力，实现多模态交互。在 AI 技术助力下，虚拟数字人的感知能力、思维能力、语言能力都得到显者增强，在思想、语言、行为上和人类更加相似。随着 AI 模型及应用向多模态的升级迭代，ChatGPT、文心一言等多模态大模型愈发成熟，使得数字人能够将深度学习模型、神经网络渲染、自然语言处理等技术进行有机结合，使之具有感知、表达等无需人工干预的自动交互能力。

1.3　内容生成

AIGC 技术为数字人的内容生成提供了强大的支持，使得数字人在语音、文字和图像等方面具备了更加丰富和多样化的表现能力。比如，通过自然语言生成技术，虚拟数字人能够根据特定情境和需求自动生成文本内容，这使得数字人可以回答问题、讲述故事、提供信息等，为用户提供更加个性化和智能化的交流和服务。这种能力使得虚拟数字人在虚拟助手、教育培训等领域具有更广泛的应用。随着 AIGC 技术的不断创新和应用，虚拟数字人将继续发展，并在各个领域展现出更加出色的表现和应用前景。

2　AI 数字人的社会安全风险分析

AIGC 驱动的虚拟数字人（简称 AI 数字人），是基于深度学习、机器学习和自然语言处理等技术构建而成的，具备一定程度的智能水平，能够模拟和学习人类的行为和情感反应的虚拟数字人。作为元宇宙社会中的虚拟实体，AI 数字人已经开始在各个领域中发挥着越来越重要的作用。随着其在社会生活中广泛应用的同时，数字人所带来的各个方面的社会安全风险也应受到足够的重视。

2.1　数据和个人信息安全风险

AI 数字人的运行依赖于大量的用户数据，包括个人信息、行为数据、交流内容等。这些数据不仅是数字人提供个性化服务的基础，也是其持续学习和优化的重要资源。然而，这也意味着用户的隐私和个人信息可能面临着被滥用或泄露的风险[6]。首先，AI 数字人在收集和处理用户数据时，可能会涉及到用户的敏感信息，如身份信息、偏好习惯、社交关系等。如果这些信息被未经授权的第三方获取，可能会对用户的隐私权造成侵犯。例如，用户的行为数据和交流内容可能会被用于推送广告或进行其他商业活动，而用户可能并不知情。其次，AI 数字人的数据存储和处理过程中，也可能存在信息安全风险。如果数据存储和处理系统的安全防护措施不足，可能会被黑客攻击，导致用户数据的泄露。此外，如果 AI 数字人的开发者和运营者没有采取有效的数据管理措施，也可能导致用户数据的误用或滥用。

2.2　潜在滥用和虚假信息传播风险

AI 数字人潜在滥用和虚假信息传播风险涉及非法活动利用、用户隐私侵犯和虚假信息传播等多方面内容。首先，AI 数字人的潜在滥用风险主要体现在两个方面。一是 AI 数字人可能被用于进行非法活动。由于 AI 数字人具有强大的数据处理和学习能力，一些不法分子可能会利用这些特性，通过 AI 数字人进行网络攻击、诈骗等非法活动。例如，他们可能会利用 AI 数字人收集的用户数据，进行身份盗窃或者金融诈骗。二是 AI 数字人可能被用于侵犯用户隐私。AI 数字人需要收集和处理大量的用户数据，包括个人信息、行为数据等。如果这些数据被滥用，可能会对用户的隐私权造成严重侵犯。其次，AI 数字人的虚假信息传播风险也不容忽视。一方面，由于 AI 数字人具有强大的信息处理和传播能力，一些不法分子可能会利用这些特性，通过 AI 数字人进行虚假信息的传播。例如，他们可能会利用 AI 数字人发布虚假新闻、误导性广告等，对公众造成误导。另一方面，AI 数字人的回答是基于预先训练的模型和算法生成的，但由于数据源的质量、算法模型的局限性等问题，可能会从网络中学习到虚假信息，并将这些信息传播出去。这不仅可能对公

众造成误导,还可能对社会稳定造成威胁。如果生成和传播具有偏见和歧视的内容,甚至会造成并加剧社会的对立和冲突。

2.3 社会影响和道德伦理风险

AI 数字人的迅速发展对社会产生了深远的影响,也带来了一定的风险。首先,AI 数字人的出现改变了人们的社会交往模式。人们可以通过与 AI 数字人进行交流和互动来满足社交需求,这可能导致人们更倾向于与虚拟实体进行交流,而忽视真实社交关系,从而影响人际关系的质量和社会交往的方式。其次,AI 数字人的出现可能加剧数字鸿沟,使得那些无法接触和使用 AI 数字人的人群处于社会交往的不利地位,进一步加剧社会不平等现象。最后 AI 数字人的出现也带来了一系列道德伦理风险。AI 数字人的行为可能受到其开发者和运营者的影响,从而导致公平和正义的问题。例如,AI 数字人可能会对某些用户或内容进行优先处理或歧视处理[7],侵犯用户的公平权和正义权。此外,AI 数字人的行为和决策可能引发道德伦理问题,例如模拟人类情感和感知,可能导致人机关系的道德和伦理问题,以及模拟人类思考和决策,可能引发人工智能的道德和伦理问题。

2.4 法律和监管问题

AI 数字人的法律和监管风险是元宇宙社会安全治理中的重要问题,涉及到虚拟实体的法律地位、责任认定和监管机制等方面。首先,AI 数字人的法律地位尚不明确,是否应被视为具有法律主体性质的实体,以及在法律体系中的地位和权利义务尚未有明确规定。这为 AI 数字人的行为和责任认定带来了困难,也为其滥用和违法行为的监管带来了挑战。其次,AI 数字人的行为和内容产出可能涉及到知识产权、肖像权、隐私权等法律问题,如何保护用户和第三方的合法权益,以及如何界定 AI 数字人与平台运营商之间的责任边界成为亟待解决的问题。此外,AI 数字人的内容可能涉及到违法和有害信息,如淫秽色情、暴力恐怖等,如何建立有效的内容审查和监管机制,以及如何保障公共安全和社会秩序成为监管的重要挑战。最后,AI 数字人的跨界性和全球性也带来了监管的困难。由于 AI 数字人的服务和活动不受地域限制,其可能涉及到多个国家和地区的法律和监管规定。例如,如何对跨国的 AI 数字人进行有效监管,如何处理不同国家和地区的法律和监管冲突,如何建立全球性的 AI 数字人法律和监管框架,这些问题在现有的法律和监管体系中尚无明确的解答。

3 AI 数字人的社会安全风险的治理策略

3.1 加强技术层面的风险治理能力

AI 数字人的发展与应用需要依赖先进的技术与算法,因此从技术层面对 AI 数字人存在的社会安全风险进行治理至关重要,通过技术手段实现对隐私泄露、算法歧视、违规内容泛滥等问题的高效治理。[8]一是要加强数据加密和安全传输、数据脱敏和匿名化等技术手段的创新,提高 AI 数字人的数据保护和隐私保护能力。同时,建立强大的数据访问和使用权限管理系统,确保只有经过授权的用户才能够访问和操作相关数据,从而减少数据泄露和滥用的风险。二是要加强 AI 数字人的漏洞修复和风险监测技术创新。AI 数字人作为一种数字化产品,可能会受到各种网络攻击和恶意操作的威胁。因此,需要采取安全漏洞扫描和修复技术、入侵检测和防御技术等措施,及时发现和应对潜在的安全风险。建立有效的安全监测和告警系统,对 AI 数字人的行为进行实时监控和分析,发现异常情况并采取相应的应对措施,防止安全威胁的扩散和危害的发生。三是要加强算法透明度和可解释性,数字人技术的核心是人工智能算法,而算法的不透明性和不可解释性可能导致决策的不公平和不可信任,因此需要采取技术手段提高算法的透明度和可解释性,包括开发可解释的 AI 算法、建立算法审计机制等。

3.2 加强公众教育和社会参与

在当前数字化社会中,AI 数字人的应用已经深入到人们的日常生活中,因此必须提升公众对 AI 数字人的认知水平和安全意识,加强社会参与 AI 数字人社会安全风险的治理。首先,应加强公众教育与知识普及,通过举办研讨会、培训课程、宣传活动等方式,向公众传递 AI 数字人社会安全风险的知识、风险防范措施和权益保障。让公众了解 AI 数字人技术的基本原理、应用领域和潜在风险以及如何保护个人隐私和数据安全。数字人伦理教育也是公众教育的重要内容,通过开展数字人伦理教育,引导公众形成正确的数字人伦理观念和行为规范,提升公众对 AI 数字人风险的防范能力。其次,要建立健全公众参与机制,建立政府与公众、企业之间的对话机制,形成多元化的公众参与渠道和监督机制,增加公众对治理工作的参与感。通过定期举行公众听证会、研讨会等活动,促进不同利益方的互动与理解,形成公共治理的共识和合力。此外,还要加强舆论引导与风险沟通,政府及时公布风险预警和相关信