生成式人工智能在制造业数字化转型中的安全挑战与韧性治理研究

向宁
乐山职业技术学院，四川 乐山 614000
DOI:10.61369/IED.2025060035

摘　　要： 随着全球制造业加速向智能化、数字化方向转型，生成式人工智能（AIGC）作为关键技术引擎，正在深度重塑产品设计、生产流程与供应链管理。然而，其在赋能效率跃升的同时，也引发了前所未有的数据安全、算法失控及系统脆弱性等风险。本文立足于制造业数字化转型的现实需求，系统剖析 AIGC 应用中的核心安全威胁，并创新性提出“技术－管理－制度”协同的韧性治理框架。研究表明，唯有通过动态风险感知、自适应防护机制与多元主体共治，方能实现 AIGC 在制造业中的安全可控与可持续发展。

关 键 词： 数字化转型；生成式人工智能（AIGC）；安全挑战；韧性治理

Research on Security Challenges and Resilient Governance of Generative Artificial Intelligence in the Digital Transformation of Manufacturing Industry

Xiang Ning
Leshan Vocational and Technical College, Leshan, Sichuan 614000

Abstract: As the global manufacturing industry accelerates its transformation towards intelligence and digitalization, Generative Artificial Intelligence (AIGC), as a core technology engine, is profoundly reshaping the product design process, production operation mode and supply chain management system of the manufacturing industry. However, while AIGC brings values such as leapfrog improvement of production efficiency and reduction of innovation costs to the manufacturing industry, it also triggers unprecedented security risks including data leakage, out-of-control algorithmic bias and system vulnerability transmission. Based on the actual needs and application scenarios of the digital transformation of the manufacturing industry, this paper systematically analyzes the types and formation mechanisms of core security threats in the full-chain application of AIGC in the manufacturing industry, and innovatively proposes a "technology protection-management optimization-institutional guarantee" trinity resilient governance framework. The research shows that only by building a dynamic risk perception system, deploying an adaptive protection mechanism, and establishing a multi-subject co-governance model involving government, enterprises and research institutions, can the safe and controllable application of AIGC in the manufacturing industry be realized, providing stable technical support for the digital transformation of the manufacturing industry and promoting its sustainable development.

Keywords: digital transformation; Generative Artificial Intelligence (AIGC); security challenges; resilient governance

引言

制造业的数字化转型已成为工业4.0时代的核心战略。在中国制造2025、德国工业4.0等国家政策推动下，以生成式人工智能（AIGC）为代表的新兴技术正快速渗透至制造全链条。例如，通过生成对抗网络（GAN）自动生成产品设计原型，利用时序模型预测设备故障，或构建数字孪生体优化工厂布局，这些应用显著提升了研发效率与资源利用率。当前研究多聚焦于 AIGC 的技术实现，对其在制造业中的安全风险系统性分析仍显不足。尤其当 AIGC 深度融入物理生产环境时，其算法黑箱特性、数据依赖性与系统耦合性可能引发级联故障[1]。因此，本研究旨在填补该领域理论空白，并为制造业构建 AIGC 安全治理体系提供实践路径。

一、AIGC 在制造业的应用深化与安全挑战

（一）技术赋能的典型场景

在智能产品设计领域，生成式模型通过学习历史工程数据，可在数小时内输出数百种符合功能需求的零部件方案，将传统设计周期压缩50% 以上。例如，航空航天企业利用扩散模型（Diffusion Models）生成轻量化舱体结构，材料利用率提升37%。在预测性维护场景中，长短期记忆网络（LSTM）结合生成式故障模拟技术，通过对设备运行数据的时序建模，提前14天预警机械失效风险，使意外停机率下降30%[2]。此外，基于 AIGC 构建的虚拟工厂仿真系统，可动态生成供应链中断、能源波动等突发事件下的生产调度方案，助力企业实现资源弹性配置。此外，基于 AIGC 构建的虚拟工厂仿真系统，可动态生成供应链中断、能源波动等突发事件下的生产调度方案，助力企业实现资源弹性配置。

（二）多维安全风险的显现

随着应用深化，AIGC 的安全挑战呈现多层次特征。在数据层面，训练数据的质量直接决定生成结果可靠性。若原始数据被植入恶意样本（如篡改的轴承振动信号），AIGC 可能输出存在隐蔽缺陷的零件设计图纸。更严峻的是，生成模型可能通过反演攻击从合成数据中还原敏感信息 [3]。某数控机床厂商曾发现，其发布的 AI 生成工艺参数可被逆向推导出核心合金配方，造成商业机密泄露。

在算法层面，模型不可解释性成为关键瓶颈。由于深度学习固有的黑箱特性，AIGC 生成决策的逻辑难以追溯。例如，某工业机器人制造商采用 AIGC 优化运动控制算法时，模型因未识别极端工况下的扭矩限制，生成错误指令导致机械臂过载损坏。同时，对抗性攻击的威胁日益凸显：攻击者只需在输入数据中添加人眼难辨的扰动，即可诱使 AIGC 生成完全失效的3D 打印模型，造成精密部件批量报废 [4]。

在系统层面，AIGC 与操作技术（OT）的深度融合放大了安全漏洞。当生成模型直接接入 PLC（可编程逻辑控制器）系统时，恶意构造的虚假传感器数据可能触发错误的生产指令。更值得警惕的是供应链 AI 系统的级联风险。根据 Gartner（2024）预测，到2025年17% 的制造业停机将源于 AI 供应链攻击。类似案例中，丰田（2022）因供应商系统遭攻击导致停产1天，损失1.3万辆产能 [5]。仿真研究表明，若 AIGC 预测系统被入侵，可能引发72小时级停产（ISA，2024），凸显加强供应链 AI 安全防护的紧迫性。

二、韧性治理框架的理论建构

传统安全防护强调刚性防御，如防火墙、访问控制等静态手段，难以应对 AIGC 引发的动态风险。韧性治理则着眼于系统在遭受冲击后维持核心功能并快速恢复的能力，其核心包含三重维度：抗逆性（抵抗冲击的强度）、适应性（动态调整的能力）与学习进化性（从危机中迭代升级）。基于此，本研究提出的三维协同治理模型。

（一）技术韧性

技术韧性要求构建动态防御与透明决策相结合的 AIGC 安全体系。在数据安全层面，需采用联邦学习（Federated Learning）框架，通过分布式模型训练实现 " 数据可用不可见 "，避免传统集中式存储导致的单点泄露风险。在算法可解释性方面，应部署 Shapley 值（SHAP）等归因分析方法，量化输入特征对生成结果的贡献度，例如：当 AIGC 生成工程设计时，可追溯关键参数（如应力系数、公差范围）的决策权重 [6]。此外，建议引入差分隐私（Differential Privacy）技术，在模型训练阶段注入可控噪声，防止通过生成结果反推敏感数据。

（二）管理韧性

主要聚焦于组织架构与流程的动态适应能力，确保在面临 AIGC 相关冲击时核心业务持续运转。要求构建覆盖 AIGC 全生命周期的韧性治理体系。首先需建立明确的跨部门职责分工和专门的应急响应团队，实施常态化安全培训，提升组织成员的风险意识和协同应对能力。其次应将韧性理念系统性地嵌入从需求分析、设计开发、安全测试到部署上线及持续监控的每一个环节，形成一个动态闭环的“识别 – 防护 – 检测 – 响应 – 恢复”管理流程 [7]。同时，应积极利用“数字孪生沙盒”等先进技术手段，在安全可控的虚拟环境中模拟各类攻击场景（如勒索软件对供应链的渗透），验证应急预案的有效性并不断优化响应策略。最终，通过鼓励从演练和真实事件中持续学习、汲取经验，迭代管理机制，塑造主动适应和不断进化的韧性文化。

（三）制度韧性

旨在构建能够有效缓冲冲击、支持灵活调整并促进持续进化的规则体系与政策环境。要求建立完善的标准规范和政策协同机制，为 AIGC 的韧性治理提供稳定的制度基础。首先需要积极借鉴和转化国际成熟标准（如 ISO 23247工业数据安全框架），结合 AIGC 特性制定具有约束力的生成内容质量追溯规范，明确权责界定和追责路径，增强系统的规则抗逆性。其次应探索创新性的“监管沙盒”模式，在风险可控的前提下允许企业测试新兴 AIGC 应用，以此识别潜在风险、验证监管规则的有效性并适时调整政策，提升监管框架的动态适应性 [8]。更重要的是，需建立常态化的跨部门、跨地域政策协调机制与动态反馈回路，确保标准与政策的持续迭代更新，能够从实践中学习经验教训，从而系统性提升制度体系的学习进化能力，最终实现安全可控与技术创新发展的动态平衡。

三、韧性治理的实践路径

（一）技术层的动态防护

技术层动态防护的核心在于构建能够主动适应威胁、抵御攻击并维持核心功能的韧性技术体系。针对敏感数据保护，需采用联邦学习等分布式架构，通过在本地节点训练模型并仅交互加密参数，实现数据的“可用不可见”，有效规避集中存储风险并保障数据主权。同时，利用合成数据技术生成符合原始数据统计特

性的仿真数据，可替代真实敏感数据进行模型训练，为化解隐私保护和模型效能之间的冲突提供关键路径[9]。在提升算法可信与鲁棒性方面，对抗训练是关键手段，通过将精心构造的对抗性样本融入模型训练过程，显著增强模型抵御恶意输入扰动和规避攻击的能力，从而大幅降低生成错误率，确保系统在复杂威胁环境下的稳定输出和持续服务能力。

（二）管理层的流程再造

管理层流程再造的核心是将韧性治理理念系统性地融入 AIGC 应用的全生命周期管理，构建事前预防、事中监控与事后响应的闭环体系。要求制造企业将 AIGC 安全治理深度整合至其组织架构与管理流程中。关键举措包括实施“安全左移”策略，即在系统开发设计阶段预先嵌入威胁建模与安全需求分析，对核心生成组件进行主动安全验证。在部署与运行阶段，必须建立覆盖全流程的实时监控机制，部署先进的异常检测技术以快速识别并拦截针对生成过程的恶意攻击行为[10]。事后则需构建高效的响应回溯机制，利用如区块链等技术实现生成结果与决策链条的不可篡改存证，确保事件可溯源、责任可界定，并基于此持续优化安全策略，形成动态提升组织韧性的管理闭环。

（三）制度层的协同进化

制度层协同进化旨在构建动态演进、兼顾规范与激励、并能促进多方协作的韧性规则体系与政策环境。其核心在于通过前瞻性的制度设计，为 AIGC 的稳健发展提供坚实的规则基础和政策支撑。在国内层面上，需建立健全明确的法律责任框架，明晰 AIGC 生成内容造成损害时的归责原则与责任边界，以增强系统的规则抗逆性。同时，应积极探索灵活的创新支持机制，如设立专项保障基金或提供政策激励，赋能企业尤其是中小企业提升安全治理能力和水平。至关重要的是，必须建立高效的跨部门、跨层级政策协调机制与动态反馈回路，确保制度规则能及时响应技术迭代和风险演变。在国际层面上，亟需推动构建跨境协同治理机制，共同应对全球性供应链风险挑战，并积极参与国际标准互认与规则协调，通过持续的制度适应性调整与协同进化，系统性地提升韧性治理的制度保障水平，最终实现安全可控与创新发展的动态平衡。

四、案例实证

为验证韧性治理框架的有效性，本研究深入调研了一家部署 AIGC 系统的高端制造企业。该企业在应用初期曾遭遇因技术缺陷（如模型过拟合）导致的严重生产质量和经济损失。在全面实施涵盖技术、管理与制度三个维度的韧性治理方案后，成效显著。在技术层面，通过部署先进的数据安全防护机制与实时异常检测预警模块，核心资产得到加固，系统生成结果的可靠性大幅提升。管理层则重构了开发与决策流程，引入多方独立验证与仲裁机制，显著增强了生成结果的审慎性与可信度。制度层创新性地探索了风险管理工具，建立了安全绩效与保障成本挂钩的激励约束机制。实施评估显示，韧性治理体系有效推动了设计效率的大幅提升，关键质量缺陷率显著下降，并在遭遇针对性攻击时展现出强大的系统容灾与快速恢复能力，关键业务中断时间被压缩至极短范围内，全面验证了该框架对提升企业 AIGC 应用韧性的实践价值。

五、结论与展望

本研究系统论证了生成式人工智能（AIGC）在制造业数字化转型中引发的多维安全挑战，并提出“技术－管理－制度”三维协同的韧性治理框架。研究表明，传统静态防御机制难以应对 AIGC 的动态风险，而韧性治理通过技术层的自适应防护、管理层的全流程闭环管控及制度层的动态规则演进，可显著提升系统抗逆性、适应性与进化性。未来研究需进一步探索量子安全加密与跨境治理协议等前沿方向，推动构建覆盖“数据－算法－系统”全局的韧性生态。同时，亟须深化政策工具创新，建立兼顾技术创新与风险防控的动态平衡机制，为制造业智能化转型提供可持续的安全基座。

参考文献

[1] 乔朋华，杜鑫，韩先锋 . 生成式人工智能如何提升制造业企业韧性？ [J/OL]. 科学学与科学技术管理，1-22[2024-09-25].

[2] 吕珏妙 . 生成式人工智能在数字经济中的创新应用与挑战 [J]. 产业创新研究，2024，(12)：16-18.

[3] 金永花，辛伟涛，王珊珊 . 人工智能赋能传统制造业升级的路径、挑战与建议 [J]. 科技智囊，2024，(04)：19-25.

[4] 唐先达，叶露露，程志鹏 . 生成式人工智能 (AIGC) 赋能金华市制造业高质量发展策略研究 [J]. 商场现代化，2024，(09)：158-160.

[5] 李志，胡赫男 . 生成式人工智能驱动制造业生产力变革的机制与路径——基于广东省的实证分析 [J]. 工业工程，2024，28 (02)：1-11.

[6] 李诗婧，赵爽 . 生成式人工智能应用于制造业的网络安全风险及对策研究 [J]. 信息通信技术与政策，2025，51 (01)：20-24.

[7] 黄仕晖，周峰，梁梦瑶 . 人工智能时代制造业企业数据安全的挑战与应对 [C]// 2024 年山东省数据技术与应用论文集 . 山东商业职业技术学院；，2024：73-83. DOI:10.26914/c.cnkihy.2024.051389.

[8] 傅文军，吕骏，屠宇飞 . 发展制造业生成式人工智能的相对视角与业务路径 [J]. 中国仪器仪表，2024，(05)：22-26.

[9] 李奕 . 生成式人工智能对制造业发展的影响 [J]. 上海质量，2024，(02)：38-43.

[10] 郑世林，陶然，杨文博 . ChatGPT 等生成式人工智能技术对产业转型升级的影响 [J]. 产业经济评论，2024，(01)：5-20.

