制药领域质量管理中生成式人工智能的应用实践与虚假信息防范探索

Exploration of Application Practice and False Information Prevention of Generative Artificial Intelligence in Quality Management of the Pharmaceutical Field

吴 斌，陈剑芬，李俊丽，宋雪洁

（润尔眼科药物（广州）有限公司，广东广州 510530）

WU Bin, CHEN Jianfen, LI Junli, SONG Xuejie

(*Ocusun Ophthalmic Pharmaceutical*(*Guangzhou*) *Co., Ltd., Guangzhou 510530*)

摘要：生成式人工智能（GAI）正深刻影响并重塑制药行业质量管理工作范式。GAI在质量报告自动化生成、风险预测与评估等场景能显著提升工作效率，但在“幻觉”特性的干扰下可能导致平均约37%的结论偏差或错误，存在药品安全事件与合规风险的隐患。文章从“构建多源数据交叉体系、建立标准GAI提示词、实施人机协同工作流、制定配套管理制度”方面，提出“四维防范机制”，防范GAI虚假信息，从而降低相关风险，并为制药企业智能化转型提供了兼顾技术创新与风险防控的系统方案思路，以期为业内提供一定的参考。

关键词：人工智能；质量管理；虚假信息

中图分类号：R95 **文献标志码：**A **文章编号：**1001-8255(2026)01-0135-05

引用：吴 斌, 陈剑芬, 李俊丽, 等. 制药领域质量管理中生成式人工智能的应用实践与虚假信息防范探索[J]. 中国医药工业杂志, 2026, **57**(1): 135-139. DOI：10.16522/j.cnki.cjph.2026.01.016

生成式人工智能（generative artificial intelligence，GAI）已开始影响传统的制药行业，并重塑制药领域质量管理范式[1]。GAI技术在传统制药行业的快速渗透，为药品生产流程优化和质量控制提供了全新的视角[2-3]。尤其是自动化生成报告，显著提升了质量管理的工作效率，为质量报告自动化、风险预测模型构建提供了新的技术路径。

然而，技术红利背后也隐藏着严峻的挑战。GAI的“幻觉（hallucination）”特性[指由模型生成的、与输入源内容存在语义偏离（不忠实）或逻辑荒谬的无效输出]，可能引入虚假信息[4-6]。例如，虚构或杜撰未报道数据、错误索引法规条款或杜撰法规条文等，导致生成不符合《中华人民共和国药品管理法》《药品生产质量管理规范》相关要求的错误结论。此类风险很有可能引发严重的药品安全事件或严重的违规行为。更严峻的是，GAI的“黑箱”特性（指从输入到输出的决策过程缺乏透明性，导致难以追溯生成逻辑）[7]与GMP“数据完整性”原则冲突[8]。当前研究多聚焦技术效率提升，忽视了对虚假信息生成机制的系统性防范，因此制约了GAI在药品质量管理中的规模化应用。

本研究拟构建“技术应用-风险防控”双轨制分析框架。一方面，用2022—2024年某企业的实际质量活动案例，与GAI应用进行对比，分析其在质量报告自动化、风险评估及预测等实践价值；另一方面，从构建多源数据交叉体系、建立标准GAI提示词、实施人机协同工作流、制定配套管理制度共4个方面提出“四维防范机制”，防范GAI生成虚假信息，从而保障药品全生命周期质量安全。

1 GAI在制药质量管理中的应用实践

GAI在制药质量管理中的应用实践，主要体现在提升质量报告自动化的工作效率和优化风险评估及预测2个方面。以下将对比GAI[“DeepSeek-R1”，中国深度求索（DeepSeek）公司]生成的质量报告与传统人工编制的质量报告。

1.1 对照组

采用2022—2024年某企业已完成并经质量管理

收稿日期：2025-06-04

作者简介：吴 斌，主管药师，从事药品质量管理与控制的研究。

部门批准的质量管理活动报告为对照组。其中包括30份质量文件（如变更管理规程、某设备或某检验方法操作规程、某职务/部门岗位职责、物料验收/贮存/发放/退货管理规程等）、20份变更文件（如场地变更、标准变更、内包材变更、组织架构变更等）、10份设施设备确认方案（如洁净厂房确认方案、某灌装设备确认方案、某灭菌设备确认方案等）、10份风险评估文件（包括如工艺验证前/清洁验证前/设备设施验证确认前的风险评估、某产品质量属性和工艺参数风险评估、其他事件的风险评估等），共计70份。

1.2 试验组

向“DeepSeek-R1”输入对照组质量活动的提示词，将生成的质量报告作为试验组。

1.3 提示词设置

本研究中所有类别质量活动的提示词设置为固定格式，“任务+生成主体+细节（可选）+形式（可选）”，取首次生成报告为试验组，提示词设置如下。

①质量文件。任务为编制“某文件名称”。主体结构包括“目的、范围、职责、定义、程序（或正文）、附件、引用参考、修订历史等（根据企业对不同类型文件的编制要求输入）”。各章节及正文段落、文字排版、表格设计、流程图/公式等细节，均遵循企业文件的编制要求。形式符合我国相关法律、法规及指南的要求。

②确认与验证。任务为编制“某确认/验证方案”。主体结构包括“概述、确认目的、确认范围、职责、法规和指南依据、风险评估、确认/验证项目设计、可接受标准、偏差处理、变更处理等（根据不同类型的确认与验证需要输入）”。验证与确认前的风险评估、项目设计、标准等具体要求，根据不同的验证/确认类型的需要输入。形式符合我国相关法律、法规及指南的要求。

③风险评估。任务为编制“某事件的风险评估”。主体结构包括“概述/事件背景、评估目的、评估工具、评估标准、风险识别/分析/等级、风险控制及行动项建议、风险结论、引用参考等（根据不同风险事件的需要输入）”。细节涵盖风险工具选择、风险识别/分析/控制、行动项等具体要求（根据风险事件的实际情况需要输入）。形式符合我国相关法律、法规及指南的要求。

④变更。编制“某事件的变更”。主体结构包括“变更申请、变更理由、变更内容（变更前后对比）、变更等级判定、变更影响评估（包括对生产、质量、注册、物料、设施设备、安全及其他方面等的影响评估，如某方面不涉及可以评估无影响）、变更行动计划、变更效果实施评估、变更结论/关闭等（按企业对不同类别变更流程的要求输入）”。变更等级判定、影响评估、行动项等的具体要求，应根据变更情况的需要进行输入。形式符合我国相关法律、法规及指南的要求。

1.4 计算方法

通过“文件编制平均耗时［指GAI显示的生成时间+人工复核时间（固定为10 min)］”“格式及完整性错误率（指格式缺陷项数占总检查项数的百分比）”“法规指南等引用/使用错误率（指错误引用数占总引用数的百分比）”“决策或建议错误率（指错误建议数占总建议数的百分比）”这4个指标考察GAI的工作效率和风险评估及预测情况。其中“决策或建议错误率”指标仅选取对照组中有建议及行动项或结论的报告进行比较，包括“风险评估文件”“变更文件”共计30份。通过上述4个指标的数据来论证GAI在制药质量管理中的实践价值与风险边界，并衡量其在质量管理中应用的可行性。

1.5 GAI实践应用情况

上述4个指标的统计结果如表1所示。

由表1可知，GAI生成报告较传统人工编制（按

表1　GAI实践应用情况

Tab.1　Practical Applications of GAI

提示词	每份文件编制平均耗时/h	格式及完整性错误率/%	法规指南等引用或使用错误率/%	决策或建议错误率/%
质量文件	0.17	3.50	14.85	—
确认与验证	0.17	5.60	29.26	—
风险评估	0.17	3.29	18.42	29.03
变更	0.17	7.00	57.89	36.52[1)]
平均	0.17	4.84	30.10	36.81

注：[1)] 此项是变更等级判定错误率（55.00%）和变更行动项错误率（17.50%）的平均值。

经验平均约每份 3 ～ 5 h）在工作时间上具有显著优势，大幅提升了质量文件编制的工作效率；在格式及完整性上，与人工编制水平相当，能够节省大量、烦琐的文档编制及版面排版工作。但在法规指南等引用 / 使用、决策或建议的错误率上存在显著缺陷，平均为 30.1%、36.81%。

GAI 生成报告的错误主要表现在以下 2 个方面。①法规指南引用 / 使用错误。例如，在生成“变更管理规程文件”时，撰写的章节包括了引用“ISO 14001：2015”，经查询该标准为“环境管理体系标准”；在生成“某药品有效期延长变更报告”时，变更申请章节撰写依据是《药品注册管理办法》，该依据本身无错误，但错误引用第 33 条，该条款内容为“药物临床试验登记和信息公示的要求”。另外，还有“杜撰标准或数据”“杜撰法规法条”等情形。例如，在生成“某灭菌柜确认方案”时，气密性确认章节中的“真空泄漏率”接受标准为“≤ 0.5% /h”，但未查询到该标准出处，业内通常遵循的《药品 GMP 指南：无菌制剂》中建议的接受标准为“≤ 1.3 kPa/min”；在生成“厂区虫鼠控制风险评估报告”时，引用我国 GMP 附录中的“厂房与设备”，以及 WHO 发布的“制药企业虫害防治指南”，经查询，我国 GMP 并无“厂房与设备”附录，且 WHO 未发布过上述虫害防治指南。②决策或建议错误。主要表现在“不适宜的行动项”。例如，在生成“VHP 灭菌报表新增部分信息记录”的风险评估报告时，建议的行动项包括“重新进行 VHP 灭菌确认”和“重新进行工艺验证”，但研究人员在提示词中已说明此次所增加的报表信息为“时间记录、初始消毒剂质量记录”，并不涉及灭菌程序变动的信息增加，因此该行动项是不适宜的。另外，对于风险分析复杂程度高、法规依赖强的变更等级判定上，错误率尤为突出，高达 55.00%。无法有效识别出变更本身的复杂性，包括是否存在叠加变更。例如，对“处方变更”“制剂原料变更”“制剂生产场地变更”等复杂且可能存在多种叠加状态的变更情形判定错误。研究人员认为，GAI 现阶段可能还无法完全识别多种情况并进行综合分析，仅能对单一情况进行分析并套单一法条进行判定。

综上所述，本研究认为在现阶段制药领域质量活动管理上，直接引用或完全依赖 GAI 所生成的质量报告存在较大风险，可能引发严重的药品安全事件或合规风险，需谨慎考虑。

2 人工智能虚假信息防范探索

有报道[6]指出，“幻觉”现象的主要成因在于数据模糊，即便训练数据准确无误，且经过人为反复反馈调整，也无法完全避免这一问题。纯粹通过技术手段消除虚假信息并不乐观[5]。

因此，本研究认为现阶段可以通过后期人工干预来降低虚假信息。从“构建多源数据交叉体系”“建立标准 GAI 提示词”“实施人机协同工作流”“制定配套管理制度”4 个方面，提出“四维防范机制”，防止 GAI 虚假信息。

2.1 构建多源数据交叉体系

向 GAI 接入实时且真实的数据，降低虚构和杜撰数据概率。包括内部数据和外部数据 2 个层面。(1) 内部数据。在考量质量管理应用人工智能的场景时，对接企业已有的集成化系统数据库［例如环境监测系统（environmental management system，EMS）、建筑设备管理系统（building management system，BMS）、实验室信息管理系统（laboratory information management system，LIMS）］，实现环境监测数据、生产工艺数据、检验检测数据（如厂房设施 / 库区等的温湿度数据、洁净区在线监测的环境数据、设备运行参数数据、检验分析数据等）与 GAI 分析结果的实时比对。(2) 外部数据。向 GAI 数据库里嵌入以下内容：①实时的法律法规、技术指南、指导原则、行业共识等；②权威的文献报道（包括与产品相关的工艺技术报道、标准 / 检测方法的报道、异常情形报道、不良反应报道等）；③可信的新闻媒体报道（包括行业动态、与产品相关的社会事件报道等）。

通过接入或主动嵌入内部数据库、外部数据，建立真实数据溯源链。在确保原始数据不可篡改的前提下，降低 GAI 检索到错误 / 虚假信息概率。例如，在年度回顾应用中，GAI 通过与历史真实数据或实

时监测数据全面关联，可提高数据准确性。

2.2 建立标准“GAI 提示词”

GAI 提示词是指用户输入给 GAI 模型的指令、问题或描述性文本，用于引导模型生成符合预期的结果，是人类与 GAI 交互的核心媒介，直接影响生成内容的质量和方向[9]。研究表明，同样的活动或事件，输入不同的提示词（语言表述方式或语句顺序不同）、不同交互次数、不同交互顺序，甚至不同的人在不同的设备上输入相同的提示词等，GAI 生成的报告、建议或结论大多会存在一定的差异，且在多次反复交互后还会产生与原本提示词偏离巨大的错误结论。

通过查阅相关文献[10-11]可知，提示词的标准化应用在其他领域已有类似研究，相关结果均强调了 GAI 提示词设计的重要性，提示词标准化能够显著提升 GAI 生成内容的准确性与专业性等。因此，对于不同的质量活动，应分别建立提示词的标准化格式，包括关键信息表述的标准化、交互要求等，但该项工作可能还需长期探索。

2.3 实施人机协同工作流

现阶段完全依赖 GAI 开展质量管理工作仍存在较大风险，需要通过后期人工介入进行干预，建立有效的人机协同机制是降低虚假信息风险的有效手段。例如建立“AI 初筛、小组复审、QA 批准”的三级审核流程，具体如下。

① AI 初筛。由交互者按照标准“GAI 提示词”，多次交互并经过甄别及初审后选择最可行的一版提交复审。②小组复审。该小组由跨部门的各领域专家或技术骨干组成，对提交的 GAI 文本，从文本结构、逻辑连贯性、引用法规正确率、数据来源正确性、原因分析可信度等多个维度进行打分（如设计一个“人工复核关键点检查表”），当分值低于预设标准时，则考虑终止人工智能生成，转为人工审核或调查，反之则通过。③ QA 批准。由质量保证部门终审批准 AI 文本。

2.4 制定配套管理制度

某受访调查（逾 2 300 人）显示，仅 10%受访者表示所在组织正式颁布了 GAI 应用的管理政策[12]。目前，应建立合适的 GAI 配套管理文件，有效推动 GAI 应用的规范化、标准化和可操作化，从而为 GAI 的开发、部署和应用划定清晰的行为边界与责任框架，在开发和使用的源头上降低其生成虚假信息的风险。本研究认为，GAI 管理相关文件的制定方向如表 2 所示。

3 讨论

本研究通过构建“技术应用 - 风险防控”双轨制分析框架，论证了 GAI 在制药质量管理中的实践价值与风险边界，并得出以下结论。①效率革新层面，GAI 在质量报告自动化生成、风险预测与评估等方面，展现出显著优势和技术红利，可使企业质量管理综合成本显著降低。②风险防控层面，GAI 仍存在“幻觉”特性等显著缺陷，导致生成虚假信息或错误结论，现阶段完全依赖其进行质量管理活动仍需谨慎。

本研究探索性地提出“四维防范机制”，即通过后期人工干预模式降低虚假信息的风险，但仍需在不同环境中不断改进。

另外，本研究在系统论证方面也还存在以下不

表 2　GAI 管理文件

Tab.2　GAI Management Documents

文件名称	主要内容
GAI 管理规程	企业或单位内部人工智能管理和应用的总体原则和要求规定，应当符合当下法律法规的要求
GAI 操作及维护规程	人工智能具体应用软件的详细操作步骤、更新、维护、退役等要求
GAI 提示词标准化操作规程	经过长期的探索与训练，筛选出适配不同质量活动的交互提示词，并推动提示词交互的标准化
GAI 多源数据接入管理规程	建立与其他集成系统（EMS、BMS、LIMS、SCADA 等）的对接标准，规定区块链数据溯源技术要求，设定数据异常波动阈值及复核流程等
GAI 法规知识库维护规程	建立 NMPA/FDA/ICH 等法规更新同步机制（如规定固定的同步或更新周期等），确定知识图谱节点维护标准（条款适用场景、修订版本标注），匹配准确率校准方法等
GAI 人机协同三级审核工作指引	明确“AI 初筛 - 小组复核 -QA 终审”各环节职责，要求所有 GAI 输出必须标注法规条款编号及原始文件链接
GAI 数据真实性争议处理程序	建立传感器数据与 GAI 结论冲突处理流程、专家委员会介入标准、数据篡改追溯机制等

足。①未对比不同提示词；②未对比不同企业或不同药品剂型报告，如无菌 / 非无菌产品、化药 / 生物制品等；③未对比不同的大语言模型，仅测试了 DeepSeek-R1，缺少横向对比研究数据。因此，本研究结论存在一定的局限性，这将是此类研究需要完善的方向。

随着 GAI 技术的不断发展，与传统制药领域进行高度融合是必然的，但目前仍需不断探索。建议从以下 3 个方向进行融合。①技术融合方面。探索文本、图谱与影像数据的融合分析，以应对复杂质量事件（如偏差调查、年度回顾）。②风险防控深化方面。探索开发端层面的虚假信息来源，减少 GAI 因“黑箱”特性导致决策逻辑不可追溯的技术问题[13]，实现符合“数据完整性”的原则要求。③法规建设方面。如数据泄露、隐私侵犯、算法偏见和深度伪造等安全问题[14]，又如著作权问题[15]，均亟须业内不断探索后适时出台配套的法律法规、技术指南等，以保障 GAI 技术健康可持续发展。

参考文献：

[1] 中华人民共和国中央人民政府．国务院关于印发《中国制造2025》的通知(国发〔2015〕28号)[EB/OL]．(2015-05-08)[2025-03-20]．https://www．gov．cn/gongbao/content/2015/content_2873744．htm．

[2] 中国日报网．安永发布《智启新质生产力：生成式人工智能在医疗医药领域的潜在应用》白皮书[EB/OL]．[2025-03-20]．https://hain．chinadaily．com．cn/a/202404/14/WS661bad53a3109f7860dd9588．html．

[3] 张学博, 马征远, 袁银池, 等．人工智能驱动下的制药产业变革[J]．竞争情报, 2024, **20**(5): 28-36.

[4] JI Z, LEE N, FRIESKE R, *et al*．Survey of hallucination in natural language generation [J]．*ACM Computing Surveys*, 2023, **55**(12): 1-38.

[5] 中国互联网络信息中心．生成式人工智能应用发展报告(2024)[EB/OL]．(2024-12-02)[2025-03-10]．https://www．cnnic．cn/n4/2024/1216/c88-11196．html．

[6] 科技日报．生成式 AI“幻觉”困境如何破解 [EB/OL]．[2025-03-20]．http://edu-app．people．cn/n1/2025/0131/c1006-40411161．html．

[7] MARCUS G．The next decade in AI: four steps towards robust artificial intelligence [EB/OL]．(2020-02-17)[2025-03-20]．https://arxiv．org/pdf/2020．06177．

[8] 国家药品监督管理局．国家药监局关于发布药品记录与数据管理要求（试行）的公告 [EB/OL]．(2020-07-01)[2025-03-20]．https://www．nmpa．gov．cn/xxgk/fgwj/xzhgfxwj/20200701110301645．html．

[9] LIU P, YUAN W, FU J, *et al*．Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing [J]．*ACM Computing Surveys*, 2021, **55**(9): 1-35.

[10] 林 颖, 魏梦婷．作为后勤媒介的AI提示词：大语言模型何以形塑人类语言实践[J]．新闻与写作, 2024, (12): 77-88.

[11] 宋筱璇, 徐 丹, 张紫蕴, 等．提示词重构类型对AI生成内容效果的影响[J]．图书馆论坛, 2025, **45**(4): 129-138.

[12] ISACA．Generative AI 2023: an ISACA pulse Poll[EB/OL]．(2023-11-16)[2025-03-20]．https://www．isaca．org．

[13] 王 明, 陈建兵．生成式人工智能的责任机制：溯源、挑战与构建路径[J]．北京工业大学学报(社会科学版), 2025, **25**(3): 111-122.

[14] 马 佳．法治视角下的人工智能安全体系建设思考——评《人工智能安全：原理剖析与实践》[J]．安全与环境学报, 2025, **25**(3): 1234.

[15] 祝建军．生成式人工智能著作权侵权问题研究 [J]．知识产权，2025, (2): 44-60.