《检索增强生成系统安全技术规范》
标准编制说明

标准起草小组

1.标准范围

本标准规定了检索增强生成系统的安全架构、数据安全、模型算法安全、应用安全及基础设施安全等要求。

本标准适用于检索增强生成系统的设计、开发、部署、运营及安全评估，可为服务提供者、技术支持者及第三方检测机构提供安全技术指引与合规依据。

2.工作简况

3.标准编制原则和确定标准主要内容

本标准依据《标准化工作导则 第1部分：标准化文件的结构和起草规则》（GB/T 1.1-2020）编制。本标准规定检索增强生成（RAG）系统的安全架构、数据安全、模型算法安全、应用安全及基础设施安全等要求。适用于RAG系统的设计、开发、部署、运营及安全评估，可为服务提供者、技术支持者及第三方检测机构提供安全技术指引与合规依据。

本标准核心内容包括：1.安全架构：描述RAG系统工作流程，确立包含数据、模型、应用的分层安全架构。2.数据安全：规范源数据完整性校验、敏感信息脱敏、向量存储加密、检索权限隔离及数据生命周期管理。3.模型与算法安全：明确系统提示词保护、防注入攻击、生成内容合规过滤、幻觉检测及模型鲁棒性要求。4.应用与接口安全：规定身份认证、RBAC权限控制、API速率限制、异常流量检测及全链路审计日志。5.基础设施与运维安全：提出网络隔离、环境加固、供应链组件漏洞扫描及知识库投毒应急响应机制。

4.主要试验(或验证)的分析、综述报告

无。

5.标准在起草过程中遇到的问题及解决办法；重大分歧意见的处理经过和依据；有无重要技术问题需要说明

在本文件的修订过程中，无重大分歧意见和技术问题。

6.与国外标准的关系：包括：采用国际标准和国外先进标准的程度，与国外标准主要技术内容的差异

该项目没有对应的国际标准或国外先进标准。

7.修订标准时，说明与标准前一版本的重大技术变化，并列出所涉及的新、旧版本的有关章条(可引用标准前言的内容)；废止/代替现行有关标准的建议

不涉及。

8.说明标准与其他标准或文件的关系(可引用标准前言的内容)，特别是与有关的现行法律、法规和强制性国家标准的关系

符合现行法律、法规要求。

9.标准作为强制性标准或推荐性标准的建议

建议本文件作为推荐性标准。

10.贯彻国家标准的要求和措施建议(包括组织措施、技术措施、过渡办法等内容)；标准发布后，对国内外业界可能产生的影响

建议本文件作为推荐性标准发布实施。

本标准旨在规范检索增强生成（RAG）系统的全生命周期安全技术要求，涵盖数据接入、向量存储、检索过程、模型生成等关键环节，明确安全基线与防护机制。

随着RAG架构广泛应用，其引入的外部知识库与检索环节显著扩大了攻击面，面临知识库投毒、提示词注入、向量数据反推及检索权限隔离失效等特有风险，现有大模型安全标准无法完全覆盖，制定专用规范势在必行。本标准的必要性在于落实安全合规要求，解决敏感数据泄露与生成内容不可控难题，为监管评估提供技术依据。

通过规范数据分类分级、检索权限隔离、生成内容合规过滤及应急熔断机制，指导建设者构建可信系统。实施本标准有助于提升行业整体安全防护水位，降低幻觉与有害信息传播风险，增强用户对RAG服务的信任，促进检索增强生成技术生态的安全、可控与健康发展。

11.标准是否涉及知识产权的情况说明；如标准中含有自主知识产权，说明产品研发程度、产业化基础及进程

本文件未涉及。

12.其他应予说明的事项

本文件未涉及。