引用格式： 易磊，杨忠.生成式人工智能数据安全风险评估的标准化研究 [J]. 标准科学，2026 (2):6-15+32.
YI Lei, YANG Zhong. Research on the Standardization of Data Security Risk Assessment for Generative Artificial Intelligence [J]. Standard Science, 2026 (2):6-15+32.

生成式人工智能数据安全风险评估的标准化研究

易 磊[1]　杨 忠[2*]

（1.湘潭大学 知识产权学院；2.湖南警察学院 侦查系）

摘　要：【目的】应对生成式人工智能引发的数据安全风险，亟须构建标准化的数据安全风险评估体系。【方法】通过现状分析和规范分析，梳理我国数据安全及生成式人工智能相关规范性文件，阐明标准化生成式人工智能数据安全风险评估的逻辑与路径。【结果】生成式人工智能数据安全风险评估具有标准化的必要性，可基于“法律—行业—技术”三类触发条件，构建以“数据—数据处理活动”为主轴的要素框架，并从数据处理活动、数据安全管理制度、数据安全技术措施和个人信息保护4个维度展开数据安全风险评估内容设计。【结论】有助于防范生成式人工智能数据安全风险，为相关标准的制定与实施提供支撑。

关键词：生成式人工智能；数据安全风险评估；数据处理；标准化

DOI编码：10.3969/j.issn.1674-5698.2026.02.001

Research on the Standardization of Data Security Risk Assessment for Generative Artificial Intelligence

YI Lei[1]　YANG Zhong[2*]

(1. School of Intellectual Property, Xiangtan University; 2 Department of Investigation, Hunan Police Academy)

Abstract: [Objective] To address the data security risks posed by Generative Artificial Intelligence (GenAI) technology, it is necessary to establish a standardized data security risk assessment framework for GenAI. [Methods] Through situational and normative analysis, the paper reviews China's data security and GenAI regulatory documents and clarifies the logic and pathways of standardized data security risk assessment for GenAI. [Results] The study demonstrates the necessity of standardizing GenAI data security risk assessments, proposes initiation criteria grounded in legal-industry-technical frameworks, and establishes the assessment around “data—data processing” axis, covering four dimensions including processing activities, security management systems, technical measures, and personal information protection. [Conclusion] The study contributes to prevent data security risks in GenAI, providing support for the development and implementation of relevant standards.

Keywords: generative artificial intelligence; data security risk assessment; data processing; standardization

---

基金项目：本文受湖南省教育厅科学研究重点项目“我国重要数据安全风险评估制度研究”（项目编号：22A0091）资助。

作者简介：易磊，博士，讲师，研究方向为数字法学、知识产权。

杨忠，通信作者，博士，讲师，高级工程师，研究方向为数字法学、数据标准化。

0 引 言

数据安全风险是当下各国发展人工智能技术必须面对的核心议题。欧盟率先通过全球范围首部《人工智能法案》，将人工智能系统分为不可接受风险、高风险、有限风险、最低风险等类别[1]，由此，风险评估成为欧盟生成式人工智能监管的重要工具。在此基础上，欧洲数据保护委员会（EDPB）在2025年6月分别推出《人工智能安全与数据保护中的法律与合规》《涉个人数据的安全人工智能系统基础》，为专业人员遵守数据保护、隐私和人工智能相关法规提供指导[2]。与之相应，美国国家标准与技术研究院（NIST）在2023年牵头发布《人工智能风险管理框架》（AI RMF 1.0），其要求将包括人工智能系统问责、有害偏见、人工智能系统的影响、产品安全、责任以及安全等作为人工智能评估的主要内容[3]。我国当前已深度参与人工智能技术发展，并将在一定程度上影响全球治理格局，因此，应更加积极参与国际人工智能治理进程，推动数据安全风险评估的规则制定与实践发展。

当前，标准化正成为数据基础制度建设中支撑数据治理、推动制度创新的重要工具[4]。一方面，《中华人民共和国数据安全法》第十七条提出“国家推进数据开发利用技术和数据安全标准体系建设”，体现出建设数据安全标准的必要性[5]；另一方面，2024年9月，国家发展改革委、国家数据局等部门联合印发的《国家数据标准体系建设指南》提出“到2026年底，基本建成国家数据标准体系”[6]。然而，目前针对生成式人工智能数据安全风险评估标准化的研究匮乏，专门研究不多。鉴于此，本文从必要性、触发条件、评估要素和具体内容4个方面展开研究，以期为生成式人工智能中数据安全风险评估标准化建设提供理论支撑和实践指导，助力生成式人工智能的高质量发展。

1 标准化生成式人工智能数据安全风险评估的必要性

1.1 生成式人工智能中复杂数据安全风险的协同治理需要

基于风险来源和数据生命周期的视角，结合数据安全属性，生成式人工智能数据安全风险如表1所示。对生成式人工智能开发者而言，生成式人工智能数据安全风险贯穿于数据采集、存储、标注、训练、输出、销毁的完整生命周期，其可能给个人信息隐私权益、企业知识产权、意识形态安全、国家数据安全等带来风险，并对开发者自身造成法律责任风险[7]；对行政规制机构而言，由于生成式人工智能的独特架构与应用模式使其数据安全风险具有特殊复杂性，采用事后回应与集中监管模式难以进行精准治理，且其在实践中还面临人手不足、资源有限及权限划分不清等现实挑战。但为完成监管任务，行政规制机构须明确生成式人工智能数据在生命周期中涉及数据安全风险事实、风险管理情况等清单，方能依据此清单认可开发者决策的合法性。

表1 生成式人工智能中的数据安全风险

风险来源	数据生命周期	具体风险表现
输入阶段	收集、预处理	数据合法性、数据质量、数据隐私、数据知识产权
存储阶段	存储	数据泄露、遭受攻击、数据跨境
处理阶段	运算	算法黑箱、模型中毒、数据跨境
内容输出	生成、推理	结果失真、内容歧视、内容侵权、犯罪工具
销毁阶段	删除	非法留存

推动生成式人工智能数据安全风险治理发展的突破口，是实现相应的数据安全风险评估标准化。标准是现代社会降低风险的有效规制方式，其已成为风险规制的主要工具[8]。如GB/T 45577—2025《数据安全技术 数据安全风险评估方法》所

述，数据安全风险评估可指导数据处理者、第三方评估机构开展数据安全风险评估，供有关主管监管部门在实施数据安全检查评估时参考[9]。从这个意义上看，有别于行政规制机构主导的监督检查，标准化的人工智能数据安全风险评估不仅为人工智能开发者提供一种以数据为核心、以发现和防控数据安全风险为主要目标的合规手段，也为行政机关提供操作性更强的监管工具，有助于弥补其在生成式人工智能场景下存在的专业知识不足、监管资源有限的问题，使其能够以评估报告为抓手，对数据安全风险进行量化审查和动态跟踪。

1.2 支撑基于风险规制的生成式人工智能数据安全风险监管

风险规制是指规制干预以有明确目标为前提，基于风险评估并依照风险的高低和优先次序合比例地配置资源[10]。为在数据安全与技术创新间实现动态平衡并兼顾监管公平，我国数据安全治理法律制度采纳基于“风险规制”（risk-based regulation）的治理方法。根据《中华人民共和国数据安全法》第二十一条，国家确立数据分类分级保护制度。生成式人工智能数据安全治理应按风险所涉领域划分监管职责[5]。与之相应，国家网信办等七部门联合发布的《生成式人工智能服务管理暂行办法》提出，对生成式人工智能服务实行包容审慎和分类分级监管。提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应当按照国家有关规定开展安全评估，并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续[11]。

标准化的生成式人工智能数据安全风险评估构成了人工智能数据安全风险分级分类规制的关键工具。其核心要义在于，如果行政规制机构能识别不同环节、不同领域中生成式人工智能数据安全风险的程度差异，就可以设定和实施相应的监管工具，使监管措施与对应的风险程度匹配。因此，在数据分类分级保护制度、个人信息分类分级保护制度和网络安全等级保护制度基础上，根据生成式人工智能中数据安全风险在经济社会发展中的重要程度、一旦被非法处理可能导致的危害程度，可将生成式人工智能数据安全风险分为高风险、中风险、低风险3种类型，并在此基础上实施差异化监管。对涉及核心数据与重要数据处理、直接影响个人决策，或应用于关键信息基础设施、政府治理、司法决策等场景的生成式人工智能，应当被归入高风险类别并置于相关行政部门的重点监管清单之中，须开展风险评估并及时将相应风险评估报告反馈给相应行政部门。对中低风险的人工智能，则应采取包容审慎的监管策略，鼓励先行先试。

1.3 健全生成式人工智能数据安全风险评估的技术规范指引

通过法律法规、行政规章等硬性规则来规制生成式人工智能中数据安全风险仍存在不足。尽管《中华人民共和国数据安全法》《中华人民共和国网络安全法》《中华人民共和国个人信息保护法》等法律已勾勒出我国数据安全治理的框架，但这些法律属于基础性立法，对生成式人工智能中数据安全风险只能进行原则性、宏观性的规范。有些部门规章，如《生成式人工智能服务管理暂行办法》对数据来源合规、个人信息保护等提出要求，但缺少关于生成式人工智能数据安全风险评估的具体指引。究其原因，当前数据治理法律规则主要以人为中心，针对的是人主导下数据收集、预处理、存储、销毁等生命周期环节中的行为要求和相应法律责任，但生成式人工智能的技术智能性可能使数据处理的控制从人主导转变为人机共同主导，甚至出现机占主导的情况[12]。例如，生成式人工智能在训练过程中可能使用他人的个人信息，而其自主学习与数据处理的特性及处理数据规模的海量性，容易导致个人信息保护中的知情同意原则、自动化决策等规则难以落实。因此，对于法律法规中所规定的诸如“数据安全风险评估”“分类分级”等专业化要求，行政规制机构在监管生成式人工智能数据安全风险过程中，需借助技术标准的细密化特性，方可以准确、详尽、务实的方式践

行法律所赋予的管理职责。

我国现有的数据安全风险评估标准和人工智能标准为标准化生成式人工智能数据安全风险评估提供了研制基础。在数据安全风险评估标准方面，我国已构建起以GB/T 45577—2025《数据安全技术 数据安全风险评估方法》为核心牵引，以YD/T 3956—2024《电信领域数据安全风险评估规范》、YD/T 6415—2025《工业领域数据安全风险评估规范》等行业标准为重要支撑的标准框架。截至2025年7月，全国网络安全标准化技术委员会在数据安全和个人信息保护领域已制定发布42项国家标准[13]，在人工智能技术领域也已发布GB/T 45652—2025《网络安全技术 生成式人工智能预训练和优化训练数据安全规范》、GB/T 45654—2025《网络安全技术 生成式人工智能服务安全基本要求》、GB/T 45674—2025《网络安全技术 生成式人工智能数据标注安全规范》、GB/T 45958—2025《网络安全技术 人工智能计算平台安全框架》等标准。这些标准基本覆盖生成式人工智能中数据采集、标注、训练和服务等环节，但在数据安全的全流程覆盖、特有风险应对、标准间衔接及落地实施方面仍存在不足，尤其缺乏专门面向生成式人工智能中数据安全风险评估的标准。此外，在制定出台生成式人工智能相关的数据安全标准后，还需要对其实施情况加以监督和评估，从而降低生成式人工智能中的数据安全风险，并以“风险规制”的方式实现差异化监管。

2 生成式人工智能数据安全风险评估的触发条件

数据安全风险评估以《中华人民共和国数据安全法》为根本依据，其功能在于识别和揭示潜在的数据安全风险，并以此回应和满足合规性监管的核心要求[14]。因此，明确生成式人工智能数据安全风险评估的触发条件，既是界定其启动边界和实现精准监管的重要内容，也关系到生成式人工智能开发者的合规成本，更决定实现数据安全风险前置、隔离与缓释的成效。

2.1 法规规制背景下的刚性触发条件

基于风险规制的思想，在国家法律法规框架下，数据安全风险评估并非一项普适性的法定义务。但是在某些特殊情况下，法律法规对特定主体提出强制性的数据安全风险评估要求，典型示例如表2所示。

总体来看，在涉及核心数据、重要数据及个人信息的处理活动中，或在人工智能技术应用于关键信息基础设施，以及具有舆论引导功能与社会动员能力的场景下，应当开展生成式人工智能数据

表2 开展数据安全风险评估的法定情形（不完全统计）

法律法规	义务主体	触发事项	触发依据
《中华人民共和国数据安全法》	重要数据处理者	定期安全风险评估	第三十条
《中华人民共和国个人信息保护法》	个人信息处理者	个人信息保护影响评估	第五十五条
《关键信息基础设施安全保护条例》	关键信息基础设施运营者	定期风险评估	第十七条
《网络数据安全管理条例》	重要数据处理者	数据安全风险评估	第三十一条
《中国人民银行业务领域数据安全管理办法》	数据处理者	业务数据加工算法风险评估	第三十四条
《人工智能气象应用服务办法》	气象信息服务活动提供者	算法安全、网络安全、数据安全、信息发布审核等	第十七条
《工业和信息化领域数据安全风险评估实施细则（试行）》	重要数据、核心数据处理者	数据安全风险评估	第六条

安全风险评估。这类触发条件具有“周期性—前置性—强制性”的特征，通过持续的强制评估形成可控的风险监测闭环，体现出维护个人信息权益、社会秩序、公共利益和国家安全的强烈要求。对于仅影响自身的生成式人工智能项目，是否启动风险评估往往属于组织自主决定范畴。上述法规所规定的触发条件为生成式人工智能数据安全风险评估建立了强制性的刚性底线。

2.2 行业治理逻辑下的半刚性触发条件

生成式人工智能的应用与特定行业、场景紧密结合，而不同行业的数据敏感性和业务模式差异会导致风险评估触发条件有所不同。其形成的法律基础在于《中华人民共和国标准化法》规定，国务院标准化行政主管部门统一管理全国标准化工作。国务院有关行政主管部门分工管理本部门、本行业的标准化工作[15]，我国数据安全风险监管也采取分行业监管架构。例如，《中华人民共和国数据安全法》第六条第1款明确规定“各地区、各部门对本地区、本部门工作中收集和产生的数据及数据安全负责”[5]；《网络数据安全管理条例》第三十三条规定，重要数据的处理者应当每年度对其网络数据处理活动开展风险评估，并向省级以上有关主管部门报送风险评估报告，有关主管部门应当及时通报同级网信部门、公安机关[16]。

在行业标准方面，GB/T 45577—2025《数据安全技术 数据安全风险评估方法》将需要开展数据安全风险评估的情形细化为5类，并将数据范围、处理活动或外部环境发生重大变化等列为重新启动评估的条件。针对关键信息基础设施，GA/T 2182—2024《信息安全技术 关键信息基础设施安全测评要求》指出，应按照GB/T 20984—2022《信息安全技术 信息安全风险评估方法》等风险评估标准对关键业务链开展安全风险分析。在具体行业中，JT/T 1547—2025《交通运输数据安全风险评估指南》要求，交通运输行业的重要数据处理者、大型平台运营者、赴境外上市的数据处理者在数据出境、安全事件发生或系统架构重大调整前均需启动评估。YD/T 4560—2023《5G数据安全评估规范》则强调，5G或电信业务在新业务上线前、数据承载环境发生重大变化时，以及涉及接口开放、数据共享或跨境传输前，应及时启动评估。行业治理逻辑通过“差异化合规”将上述触发条件融入行业规范，使制度既保持统一性，又能根据行业特征进行适配。由此可见，若生成式人工智能研发者处于电信、交通、金融等重点领域，其数据安全风险评估的触发条件将直接受到行业监管文件的约束，并在生成式人工智能研发者自身差异化合规下体现出半刚性特征。

2.3 技术风险逻辑下的场景化触发条件

生成式人工智能与传统信息系统最大的区别在于对大规模训练数据、模型迭代与开放接口的依赖和数据处理链条的迭代性、复杂性与开放性，所以其数据安全风险评估启动条件还应考虑特定技术场景。例如，在训练与优化训练阶段，当新增或替换大规模训练数据、引入敏感或跨境数据前，应评估其合法性与合规性，以防范因数据来源问题导致的模型偏差、数据泄露或数据知识产权争议；在数据标注与众包环节，由于参与人员成分复杂、边界松散，极易引发数据泄露与再利用风险，应当通过事前评估建立必要的控制机制；在运行与推理阶段，当用户输入、模型输出、日志及反馈被用于再训练或微调时，由于涉及数据用途变更与迁移，应启动风险评估；在接口开放与生态嵌入环节，API开放、第三方插件嵌入往往引入新的数据流动路径与责任主体，应通过上线前的风险评估予以评价。这些场景并非孤立存在，而是生成式人工智能特有技术生态的必然产物。国际上，ISO/IEC 23894:2023《信息技术 人工智能 风险管理指南》和美国《人工智能风险管理框架》已将训练数据、接口调用、跨境流动等场景视为高风险触发点，欧盟《人工智能法案》则通过对“高风险系统”的合规性评估将这些要求制度化。积极借鉴这些国际规则，有助于我国建立与国际接轨的情景化启动条件逻辑。

2.4 风险动态逻辑视角下的触发条件

前述法律—行业—技术的触发逻辑仅停留于

静态规范层面，难以应对生成式人工智能中数据安全风险的动态演化特征，有必要引入风险动态逻辑，使触发条件能够随着风险的生成、放大、传导与暴露过程实现动态适配和灵活触发。其一，在风险生成阶段，当新增或替换大规模训练数据、引入跨境或敏感数据时，应事前审查其合法性、敏感性与跨境属性，防止风险在源头扩散。其二，在风险放大阶段，当模型迭代、参数微调或第三方插件嵌入时，应在更新或扩展前设定触发条件，避免风险累积和放大。其三，在风险传导阶段，当发生跨平台调用或大规模数据共享时，应强制开展跨域风险评估，防止局部风险演化为系统性风险。其四，在风险暴露阶段，当发生数据泄露、合规投诉或重大舆情事件时，应立即重启专项评估，阻止风险外溢并明确责任边界。从动态视角看，生成式人工智能数据安全风险评估的触发条件不再依赖固定的制度层级，而是嵌入风险生命周期之中，强调在风险生成、放大、传导和暴露的关键节点及时介入。综合而言，“法律—行业—技术”制度逻辑为生成式人工智能数据安全风险评估确立刚性的底线要求、差异化合规与场景适配的基本框架，而风险动态逻辑则从生命周期角度补充过程性触发机制。前者基于规范的权威性，后者强调风险治理的前兆识别、动态响应与再触发机制。

3 生成式人工智能数据安全风险评估的要素及其关系

3.1 评估要素及其关系的总体逻辑

确定生成式人工智能数据安全风险评估的触发条件后，需明确应围绕哪些对象和因素展开评估，即评估要素及其关系。这与触发条件回答何时需开展评估不同，评估要素关注的是评估的内部结构，回答评估启动后应当围绕哪些对象和因素展开，是风险评估体系内部运行逻辑。依据GB/T 45577—2025《数据安全技术 数据安全风险评估方法》的界定，数据安全风险评估应以数据及其处理活动为核心，通过识别业务与信息系统、数据资源、处理过程和安全措施等要素，全面掌握整体状况，发现潜在隐患，并提出相应的管理与技术防护建议。基于风险规制的生成式人工智能数据安全风险的监管要求，需要构建以“数据—数据处理活动”为主线的分析坐标[17]，并将业务、开发者、训练与推理平台、风险源以及安全措施纳入其中，如图1所示。为在实践中形成可操作的生成式人工智能数据安全风险评估体系，还有必要厘清要素

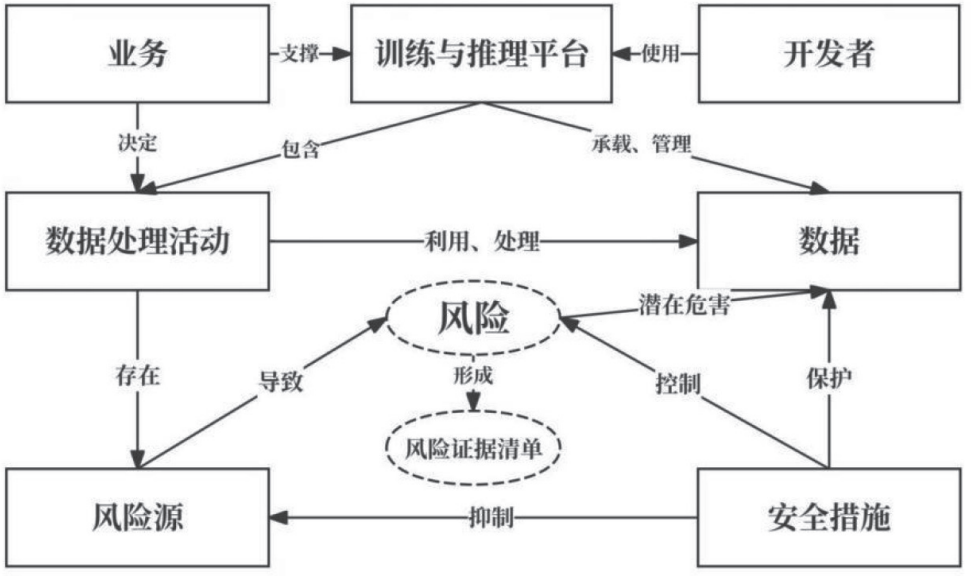

图1 生成式人工智能数据安全风险评估要素及其关系

及其互相关系，以界定评估对象的边界与作用链条，建立起可追溯、可核验、可操作的映射关系。

3.2 评估要素的阐释

数据是法律保护对象，数据处理活动是作用于客体的技术过程，两者自然构成数据风险评估的主轴。在生成式人工智能场景下，数据的外延较传统环境有显著拓展，不仅包括训练语料、标注数据和元数据，还涉及模型中间产物、提示词与用户输入、模型输出及运行日志与观测信息。理论上，任何评估结论都可回溯至某类数据在某一具体活动中如何被处理、基于何种授权、处于何种状态，并据此检验数据的完整性、保密性和可用性及处理合理性。围绕“数据—数据处理活动”的中心主轴，还需引入五类扩展要素，以明确评估的外延与边界。其一，业务用于界定处理的目的、范围与必要性，成为合法合规判断的起点。例如，《网络数据安全管理条例》第三十一条规定，重要数据处理者履行法定职责或者法定义务时可豁免风险评估义务[16]。其二，生成式人工智能开发者作为责任主体，需在授权管理、职责分离、第三方合作与审计追责方面承担组织性约束责任。例如，在委托、转委托等复杂的数据处理活动中，决定数据处理目的和方式的主体才为法律意义上的数据处理者。其三，训练与推理平台作为承载环境，具有数据管道、模型资产与版本管理、推理服务、身份与访问控制、日志与备份等能力，直接影响处理活动的可控性与取证可得性。除平台中的核心数据、重要数据、个人信息处理受到特殊法律要求外，若依赖海外云平台，还可能触发《网络安全审查办法》《数据出境安全评估办法》等规范对特定类型和规模的数据出境提出的风险评估义务。其四，风险源作为不确定性的触发点，既可能表现为攻击、数据投毒和供应链污染等外部威胁，也可能源于模型偏差、推理漂移和参数泄露等内生性缺陷，在风险的识别与评估中具有基础性地位。其五，安全措施构成治理抓手，涵盖分级分域、目的限定、访问控制与加密隔离、日志脱敏与最小留存等措施，从而形成可审计、可追溯的证据链条。

3.3 评估要素之间的关系

界定评估要素后需揭示评估要素之间的结构和相互作用关系，以形成完整的评估逻辑链。首先，主轴内部呈现双向关系。一方面，数据在多个处理活动中被收集、加工、共享与删除；另一方面，处理活动又以不同方式和强度改变着数据的机密性、完整性、可用性与处理合理性。其次，业务通过合法性依据、目的限定与范围边界，对具体数据类别、处理环节和留存时限形成约束，从而将规范性要求逐层分解为可执行的评估单元。再次，生成式人工智能开发者在组织层面承接这一约束，并将其落实到技术环节，将授权管理、职责分离和第三方治理要求映射到数据和活动的每一个节点，形成可验证且可问责的责任落实机制。同时，训练与推理平台提供承载与编排功能，其身份鉴别与访问控制、网络与分区隔离、接口治理与缓存策略、镜像与依赖管理等机制决定着风险暴露面，并通过配置业务与变更记录支撑风险证据清单的建立。风险源在这一结构中构成致因因素，既可能表现为外部攻击、技术脆弱和管理失当，也可能源自生成式人工智能特有威胁，如数据投毒、成员与属性推断、模型反向推理、提示注入和参数窃取等。这些活动直接作用于数据处理活动，也可能改变数据本身状态，进而触发风险的形成。风险一旦出现，需要通过建立风险证据清单来固化可追溯、可验证的材料，并进一步通过控制机制落实整改措施，从而避免风险在“输入—运算—输出—再学习”的流程中累积。最后，安全措施与风险源及主轴要素形成对偶的控制关系。组织、技术与合规三类控制共同构成预防、检测、响应与恢复的闭环，并与风险证据清单相结合，形成风险识别—控制—验证—改进的循环体系，以确保风险控制效果。

4 生成式人工智能数据安全风险评估的内容

4.1 评估内容框架

生成式人工智能数据安全风险评估不仅是风

险的识别、分析和评价，更是对规范性要求的系统回应。为此，抽象的要素关系需要通过制度化和方法化的安排，以转化为具体可评价的内容，进而将要素细化落实到可操作具体环节的同时，回应监管要求。结合国家数据安全与个人信息保护制度体系，参考电信、工业、交通等行业数据安全风险评估标准化经验，在制度逻辑与行业经验的共同作用下，可将生成式人工智能数据安全风险评估的内容划分为4个核心维度，即数据处理活动、数据安全管理、数据安全技术和个人信息保护，其与生成式人工智能数据安全风险评估要素之间的对应关系如图2所示。其中，数据处理活动评估维度作为数据—数据处理活动的主轴直接展开，聚焦于数据在生命周期各环节中的完整性、保密性和可用性；数据安全管理评估维度对应于开发者作为责任主体所承担的组织性约束，并落实业务要求，强调管理制度在风险防控中的统领作用；数据安全技术评估维度结合“平台条件”与“风险源”的作用机制，突出技术规制在风险识别、隔离与缓释中的地位；个人信息保护评估维度则因其高度敏感性和易受侵害性，需在评估体系中单独加以强调。

4.2 数据处理活动评估维度

生成式人工智能使用的数据来源复杂且规模庞大，既包括公开的互联网数据，也涉及行业专有数据和用户交互数据。在这一环节，评估的重点在于数据处理的合法性与必要性。一是数据采集的合法性。例如，《中华人民共和国数据安全法》第三十二条规定，“任何组织、个人收集数据，应当采取合法、正当的方式”；又如，GB/T 45652—2025《网络安全技术 生成式人工智能预训练和优化训练数据安全规范》进一步强调，应避免通过爬虫等方式获取数据。与电信行业评估规范中对采集目的、范围进行正当性和必要性审查的要求相比，生成式人工智能由于数据获取规模巨大且来源复杂，更需要明确数据收集的边界，以确保收集行为合法且必要。二是数据使用与加工的正当性。传统行业强调数据在业务范围内的合规使用，而生成式人工智能则要进一步审视数据在预训练、微调及标注环节中的合理性，尤其是数据标注直接决定训练数据以及生成内容的质量与安全。如果未经严格审查的数据被纳入训练集，可能导致模型输出带有偏见甚至安全风险。因此，可借鉴工业领域的数据风险评估经验，对数据处理各环节进行安全评估，以确保全流程合规。三是数据跨境流动。随着算力平台的全球布局和数据供应链的跨境化，生成式人工智能面临的数据跨境流动问题尤为突出，风险评估需对此予以特别关注，确保跨境环节符合法律要求和安全审查标准。四是数据销毁机制。与传统行业中“删除即销毁”的理念不同，生成式人工智能训练使用过的数据可能固化于模型参数之中，单纯删除原始数据未必能够彻底消除风险。因此，评估应关注数据销毁机制的

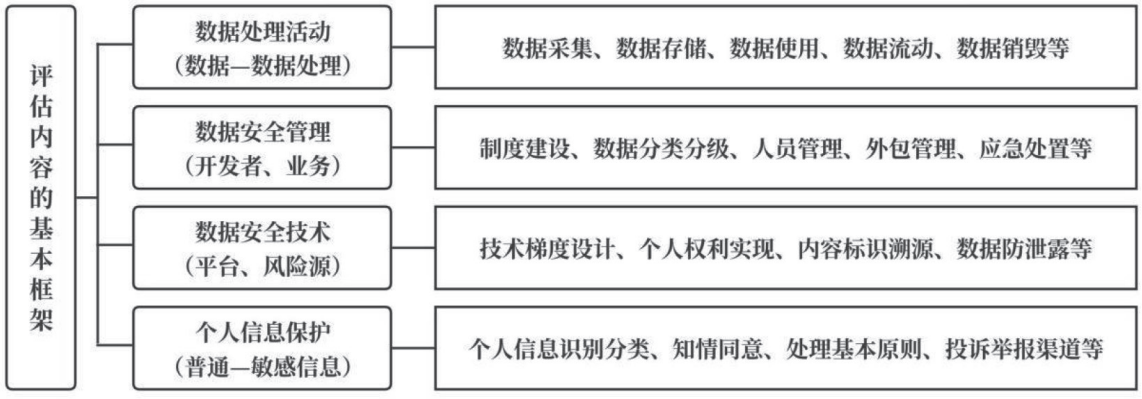

图2 评估内容与评估要素的对应关系

设计与实施，保障个人信息主体合法诉求的实现。

4.3 数据安全管理评估维度

安全管理在统筹制定人工智能数据管理策略、提升员工人工智能安全的使用能力、完善人工智能相关内外部合规规则等方面具有关键作用，是组织应对生成式人工智能安全风险的首要抓手[18]。首先，数据安全管理体现的是制度建设与组织落实，是实现和进行风险评估的基础。《中华人民共和国数据安全法》规定，开展数据处理活动应当依照法律、法规的规定，建立健全全流程数据安全管理制度。与之相应，工业和信息化领域的相关管理办法强调数据分类分级、岗位分工和责任落实。其次，数据分类分级是数据安全管理的关键手段，且高风险的数据安全工作应明确数据安全负责人和管理机构。生成式人工智能的特殊性在于训练数据往往来源不明且数量庞大，如何在模型构建前识别和隔离重要数据与核心数据，是评估不可或缺的内容。这一问题关乎风险识别的前置性，也直接影响后续的责任划分与监管部门的衔接。再次，内部人员和外包业务的管理体现生成式人工智能风险的组织特征。传统行业的评估规范均强调对从业人员的培训与管理，而在生成式人工智能中，大量数据标注和平台运维工作往往依赖第三方，因此供应链管理尤为重要。风险评估需要确认外部合作方是否具备相应的资质与责任约束，以避免形成治理真空。最后，应急响应与追责机制决定风险处置的有效性。由于生成式人工智能中数据安全风险的复杂性，重大风险事件往往超出单一部门的处置能力范围。因此，风险评估不仅要检视企业是否具备完善的应急预案，更要关注企业是否形成跨部门的联动响应机制。

4.4 数据安全技术评估维度

技术维度的考察并非着眼于具体的加密算法或接口设计，而是从数据分类分级制度逻辑上审视技术规制措施的合理性与必要性。亦即，数据安全技术须按照高、中、低3种风险等级形成梯度化设计。例如，确保高风险级别数据不可用于模型训练或推理；又如，保障个人信息主体删除权、复制权的可实现性，个人数据处理目的实现后删除个人数据。将数据分类分级保护的概念应用于实践，这在技术层面上表现为不可使用性、隐私保护、可删除性等技术特征[19]。此外，不同于访问控制、日志审计和加密机制是传统行业中风险评估的核心内容，针对生成式人工智能的技术评估还必须响应大模型带来的新型风险。在此基础上，还应特别关注数据防泄露、数据脱敏与监测预警机制的落实，这些技术直接关系到敏感信息的最小化暴露与风险的前置发现。例如，如何防止训练数据被“反向推理”还原，如何识别模型是否遭受数据投毒或对抗攻击，如何在内容生成环节实现溯源与追责，这些都成为制度层面必须考虑的问题[20]。同时，内容标识与溯源在生成式人工智能风险治理中具有独特意义。《人工智能生成合成内容标识方法》规定了显式与隐式的内容标识制度，其价值不仅在于技术实现，更在于为合规监管提供制度工具。风险评估应当在此确认企业是否落实了生成内容标识与溯源义务，并评估这些措施在防范深度伪造、虚假信息传播中的有效性。此外，技术规制还涉及前沿性的制度设计。差分隐私、联邦学习等技术虽然专业性较强，但其制度意义在于体现对数据最小暴露和多主体参与的合规考量。

4.5 个人信息保护评估维度

个人信息保护是生成式人工智能风险评估的重点领域，我国已出台《个人信息保护合规审计管理办法》《互联网信息服务算法推荐管理规定》《生成式人工智能服务管理暂行办法》等政策文件对此作出专门规定。首先，个人信息的识别与分类是保护的前提。不同于传统行业中相对清晰的敏感信息范围，生成式人工智能存在“衍生敏感信息”问题，即模型可能通过推理生成涉及个人身份、轨迹或隐私的内容。风险评估必须确认企业是否建立个人信息识别和分类机制，并在此基础上规范开展个人信息处理，以防范此类隐性侵权。其次，知情同意规则的落实是风险评估的重要指标。具有交互、自主学习等特征的生成式人

工智能须遵守知情同意法律规则，故而风险评估应检视服务提供者是否在制度上建立可感知、可追溯的用户告知体系。再次，个人信息保护原则实现的可操作性。除保障个人信息主体删除、访问、更正等权利外，还应当评估是否遵守合法、正当、必要等原则，建立便捷有效投诉的举报渠道，确保个人信息主体能够在发现侵权时得到及时响应和救济。例如，风险评估应当指出，处理个人信息应具有特定、明确、合理目的等内容，落实大模型场景下个人信息处理目的实现后的个人信息删除权实现，如是否通过模型遗忘机制响应法律要求。最后，跨境传输是个人信息保护中最复杂的环节。在交通运输和电信行业，数据出境前均需进行专项评估。而在生成式人工智能中，算力与研发的全球化部署，使数据跨境风险更加突出。风险评估应确认企业是否履行跨境传输的合规义务，并从学理上强调监管的可控性和国家数据主权的实现。

5 结 语

风险评估在保障技术创新嵌入法律规制框架中具有不可或缺的重要意义。标准化生成式人工智能数据安全风险评估不仅有助于降低生成式人工智能中数据安全风险治理的成本，也为其分级分类监管提供了可验证的检验单元。未来，我国一方面需细化生成式人工智能数据安全风险评估的技术方案，另一方面需推动生成式人工智能监管机构在风险评估报告的审查、反馈机制等方面的能力建设，为提升生成式人工智能数据安全风险防范能力和人工智能技术高质量发展提供有力支撑。

参考文献

[1] 张辛鑫,冀瑜.论欧盟《人工智能法案》中的标准化模式及对我国启示[J].标准科学,2024(5):39–46.

[2] European Federation of Data Protection Officers. EDPB publishes final version of guidelines on data transfers to third country authorities and SPE training material on AI and data protection[EB/OL]. (2025–06–05) [2025–09–02].https://www.efdpo.eu/edpb–publishes–final–version–of–guidelines–on–data–transfers–to–third–country–authorities–and–spe–training–material–on–ai–and–data–protection/.

[3] NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0) [EB/OL]. (2025–01) [2025–09–02].https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100–1.pdf.

[4] 王淼,徐丹丹,张明.标准化在构建数据基础制度中的难点与对策研究[J].标准科学,2025(8):47–51.

[5] 全国人民代表大会.中华人民共和国数据安全法[Z].

[6] 中国政府网.国家发展改革委等部门关于印发《国家数据标准体系建设指南》的通知[EB/OL].(2024–09–25)[205–09–15].https://www.gov.cn/zhengce/zhengceku/202410/content_6978809.htm.

[7] 张宁.生成式人工智能数据安全风险及其防控体系研究[J].公安学研究,2025,8(2):83–104.

[8] 陈伟.作为规范的技术标准及其与法律的关系[J].法学研究,2022,44(5):84–100.

[9] 数据安全技术 数据安全风险评估方法: GB/T 45577—2025[S].

[10] 刘鹏,张崙楠,王力.基于风险的政府监管: 理论发展与实践应用[J].中国行政管理,2024,40(3):111–124.

[11] 国家网信办、国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局.生成式人工智能服务管理暂行办法[Z].

[12] 徐伟,何野.生成式人工智能数据安全风险的治理体系及优化路径:基于38份政策文本的扎根分析[J].电子政务,2024(10):42–58.

[13] 全国网络安全标准化技术委员会秘书处.数据安全国家标准体系(2025版)(征求意见稿)[EB/OL]. (2025–08–14)[2025–09–02]. https://www.tc260.org.cn/upload/2025–08–14/1755186869496024633.pdf.

（下转第32页）

参考文献

[1] 何海艳,周国华,郑立宁.深度不确定性下重大工程创新团队的协同创新行为演化分析：以川藏铁路为例[J].运筹与管理,2022,31(10):139–146.

[2] 游贯宗,罗春林,江玮璠,等. 供应链竞合结构下的外包合作与品牌溢出策略研究 [J/OL]. 系统科学与数学, 1–20[2025–09–16]. https://link.cnki.net/urlid/11.2019.o1.20250708.1554.074.

[3] 孙嘉轶,韩玉杰,滕春贤.考虑突破式技术创新阶段和技术距离的制造商动态竞合策略研究[J]. 控制与决策, 2025,40(12):3631–3644.

[4] 张芳,蔡建峰.基于政府支持的军民合作技术创新演化博弈研究[J].运筹与管理,2021,30(2):8–15.

[5] 刘亚婕,董锋.政府参与下新能源汽车企业间协同创新的竞合策略研究 [J]. 研究与发展管理, 2022, 34 (5): 136–148.

[6] 徐建中,孙颖,孙晓光.基于演化博弈的军民融合产业合作创新行为及稳定性分析[J].工业工程与管理,2021,26(1):139–147

[7] 姜红,盖金龙,陈晨.生命周期视角下技术标准联盟企业竞合关系研究[J].科学学与科学技术管理,2022,43(9):89–107.

[8] 郭润萍,尹昊博,龚蓉.资源视角下数字创业企业竞合战略对价值创造作用机理的多案例研究[J].管理学报,2022,19(11):1588–1597.

[9] 孙凯,郭稳.竞合视角下高技术企业创新联盟稳定性研究[J].中国管理科学,2021,29(3):219–229.

[10] 何建佳,蒋雪琳,徐福缘.基于供需网企业合作博弈模型的演化路径分析[J]. 运筹与管理, 2018, 27(9): 79–86.

（上接第15页）

[14] 王伟洁.数据安全风险评估共性问题及对策[J].保密科学技术,2023(6):11–14.

[15] 全国人民代表大会.中华人民共和国标准化法[Z].

[16] 国务院.网络数据安全管理条例[Z].

[17] 于晶,田乐.数据安全风险评估方法浅析[EB/OL].(2021–12–22) [2025–09–02].https://mp.weixin.qq.com/s/sNm1ZnIoc1p5VszZBwdj9A.

[18] HEIN D K, PERSSON J, NIELSEN P A. Assessing the Security risks of generative ai in software development[J]. Journal of Systems and Software [EB/OL]. (2025–04–07)[2025–09–02]. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5208111&utm_source=chatgpt.com.

[19] LI Y M, SHAO S, HE Y,et al. Rethinking data protection in the (generative) artificial intelligence era[J]. (2025–07–19)[2025–09–02]. https://doi.org/10.48550/arXiv.2507.03034.

[20] 张建文,孙依梦.论生成式人工智能数据训练的合法性基础[J].成都理工大学学报(社会科学版),2025,33(5):24–33.