大数据与人工智能视域下数据安全保障的创新路径探究

高嘉阳 姜 楠 李 可
北京计算机技术及应用研究所 北京 100854

摘 要：大数据海量、异构、高速的特性，让传统数据安全技术难以应对新型安全风险，而人工智能与大数据的深度融合，为数据安全保障提供了全新路径。本文立足数智融合底层逻辑，从风险感知、加密隐私、风控应急、合规治理四大维度，探究数据安全保障创新思路，分析当前技术、管理层面存在的短板，并针对性提出优化对策，旨在构建智能化、全闭环的数据安全防护体系，提升新时代数据安全保障能力。

关键词：大数据；人工智能；数据安全保障；新思路

引言：数字时代下，大数据技术全面赋能各行业发展，数据成为核心生产要素，但海量数据流转也催生了泄露、篡改、非法访问等诸多安全隐患。传统静态、被动的数据安全防护模式，已无法适配复杂动态的大数据安全场景。人工智能技术的迭代升级，为破解数据安全治理难题提供了技术支撑。基于此，本文结合数智融合技术优势，探索数据安全保障新思路、新路径，助力完善智能化数据安全防护体系。

1 大数据与人工智能及数据安全核心理论概述

1.1 大数据核心特征与安全属性

（1）大数据具备四大核心特征，海量性体现为数据规模呈指数级增长，突破传统数据存储与处理边界；多样性涵盖结构化、半结构化与非结构化多类型数据，适配多元业务场景；高速性强调数据产生、传输与处理的实时性，要求高效数据流转能力；价值性具备低密度、高潜值特点，海量冗余数据中蕴含核心商业与安全价值。（2）大数据场景下，数据安全核心需求聚焦三点。保密性要求严控数据访问权限，杜绝数据泄露与非法窃取；完整性保障数据采集、传输、存储全流程不被篡改、损坏，维持数据真实有效；可用性确保合法用户可按需、稳定调取使用数据，规避系统故障、攻击导致的数据失效问题[1]。（3）传统数据安全技术存在显著适配局限。传统加密、防火墙、访问控制技术多适配静态、小体量结构化数据，难以应对大数据海量、高速、异构的特性，存在处理效率低、冗余防护不足、动态适配性差等问题，无法精准抵御大数据场景下的新型安全风险。

1.2 人工智能核心技术体系

（1）机器学习依托算法让机器自主学习数据规律，无需人工预设程序即可完成数据分类、预测；深度学习作为机器学习核心分支，依托多层神经网络模拟人脑认知，通过分层特征提取，实现复杂数据的深度解析与精准识别，是智能技术落地的核心基础。（2）智能算法为数据处理提供核心逻辑，特征识别可自主挖掘数据关键特征、剔除无效信息；自主决策技术依托数据分析结果，脱离人工干预完成智能判断与指令输出，广泛应用于各类自动化、智能化业务场景，提升数据处理与处置效率。（3）生成式AI可基于海量数据自主生成合规内容，适配内容创作、数据补全等场景；智能风控技术依托实时数据分析与风险研判，精准识别异常操作与安全隐患，相较于传统风控，具备实时性强、识别精度高、自适应能力突出的应用优势。

1.3 大数据与人工智能融合赋能数据安全的底层逻辑

（1）大数据为AI安全模型提供海量、多元的训练样本，丰富的数据样本可优化算法参数，提升AI模型对安全风险的识别精度与泛化能力，解决传统AI模型样本单一、预判偏差大的问题。（2）AI技术有效破解大数据安全处理难题，依托智能算法高效筛选、分析海量冗余数据，快速定位数据异常、泄露、篡改等风险，弥补传统技术人工筛查效率低、漏判率高的短板。（3）数智融合重构数据安全运行机制，形成“数据供给-智能研判-动态防护”的闭环体系，实现数据安全从被动防御、事后处置，向主动预判、实时防护、动态优化的智能化转型，全面提升数据安全保障能力。

2 大数据与人工智能赋能数据安全保障的创新思路与技术路径

2.1 基于智能算法的动态数据风险感知新思路

（1）依托深度学习实现数据异常行为实时识别。传统数据风险识别多依赖固定规则，难以应对复杂多变的异常访问与操作行为。深度学习可通过卷积神经网络、循环神经网络等模型，深度挖掘用户访问轨迹、数据传输、操作日志等多维数据特征，自主学习正常行为基线，精准识别越权访问、批量爬取、异常传输等隐蔽风

险行为，实现毫秒级实时监测，弥补传统识别技术滞后性强、误判率高的缺陷。（2）通过机器学习构建动态风险分级评估体系。借助机器学习分类与回归算法，整合数据敏感度、访问主体权限、操作场景、风险频次等多维度指标，摒弃静态统一的风险判定标准。系统可根据实时数据运行状态动态更新评估权重，自动将数据风险划分为高、中、低等级，针对不同风险等级匹配差异化防护策略，让风险评估更贴合实际业务场景，提升防护精准度[2]。（3）实现海量数据场景下被动防御向主动预判的模式升级。传统数据安全防护以事后处置、被动拦截为主，无法应对常态化、新型化的数据安全风险。依托智能算法对海量历史风险数据、实时运行数据的关联分析，可提前挖掘潜在风险隐患，预判攻击趋势与数据泄露风险，将安全防护前置，完成从风险发生后处置到风险萌芽前预警的模式升级，筑牢事前防护防线。

2.2 数智融合的数据加密与隐私保护新思路

（1）智能自适应加密算法适配差异化数据安全需求。不同类型、不同场景的数据安全防护等级存在显著差异，传统固定加密算法灵活性不足。智能自适应加密技术可依托AI算法自动识别数据类型、敏感等级与应用场景，针对普通公开数据、一般业务数据、核心涉密数据，自动匹配轻量化、中高强度、高强度加密方式，兼顾数据安全防护强度与数据传输、使用效率。（2）基于AI的差分隐私、联邦学习隐私保护技术创新。结合人工智能技术优化差分隐私算法，可智能调整噪声添加量，在抵御隐私窃取攻击的同时，最大限度保留数据可用价值；依托联邦学习技术，可实现多节点数据“数据不出域、模型共训练”，通过AI统筹调度各节点模型参数交互，有效解决多主体数据共享过程中的隐私泄露难题，平衡数据共享与隐私安全[3]。（3）动态密钥智能管理与数据脱敏精准优化方案。借助AI算法实现密钥的自动生成、动态更新、智能轮换与全程管控，杜绝静态密钥泄露、破解风险。同时，智能化脱敏技术可精准识别手机号、身份证、交易信息等敏感字段，依据数据用途实现脱敏规则动态适配，完成精准脱敏、局部脱敏，避免过度脱敏导致的数据失效问题。

2.3 智能化数据安全风控与应急响应新思路

（1）搭建大数据驱动的AI智能风控模型。以全量数据流转数据、安全日志、风险案例为基础，训练专属智能风控模型，整合数据采集、存储、传输、使用、销毁全流程风险因子，实现对数据流转全链路的动态风控。模型可自主迭代优化，持续适配新型攻击手段与新型数据风险，提升风控体系的适应性与稳定性。（2）实现网络攻击、数据泄露等风险的智能溯源与研判。面对病毒入侵、非法窃取、数据篡改等安全事件，AI技术可快速关联分析海量日志数据，精准定位攻击源头、攻击路径、影响范围与泄露数据体量，自动研判风险等级、扩散趋势与危害程度，为后续处置提供精准的数据支撑，解决传统人工溯源效率低、定位不准的问题[4]。（3）构建自动化、智能化数据安全应急处置体系。依托数智技术搭建标准化应急处置流程，针对不同类型、不同等级的安全风险，预设智能处置策略。风险触发后，系统可自动执行拦截攻击、隔离风险节点、冻结异常访问、修复数据漏洞等操作，同时自动生成处置报告，大幅缩短应急响应时长，降低安全事件造成的损失。

2.4 基于AI的数据安全合规治理新思路

（1）智能筛查适配数据安全法律法规合规要求。将《数据安全法》《个人信息保护法》等法规条款数字化、模型化，通过AI智能筛查技术，自动检测数据处理、用户信息收集、数据共享等操作的合规性，及时识别违规收集、超范围使用数据等问题，实现合规风险常态化自查。（2）搭建海量数据分级分类的智能化自动管理模式。针对海量异构数据人工分级分类效率低下、标准不统一的问题，依托AI特征识别算法，自动区分数据类型、判定敏感等级，完成海量数据的智能分级分类归档管理，统一分类标准、提升管理效率，为差异化合规防护提供基础支撑。（3）构建数据全生命周期智能化合规监管体系。依托数智融合技术，搭建覆盖数据采集、存储、流转、使用、销毁全流程的智能监管体系，实现合规状态实时监测、违规行为自动预警、合规数据动态留存，形成全链条、常态化、智能化的合规监管闭环，实现数据安全治理从人工合规向智能合规的转型。

3 数智融合下数据安全保障的现存问题与优化对策

3.1 技术层面存在的核心问题

（1）AI安全模型普遍存在算法偏差与误判风险。现有智能安全模型多依托传统历史数据训练，样本覆盖范围有限，面对新型网络攻击与隐蔽异常操作时适配性较差，易出现风险漏判、正常行为误判等问题，大幅降低智能风控可靠性。（2）海量高维数据处理效率存在明显短板。大数据场景下异构、高维、流式数据体量庞大，传统处理架构算力调度不合理，数据筛选与运算效率低下，无法实现风险实时监测与研判，严重制约数据安全动态保障效果。（3）智能安全技术多存在场景兼容性短板。多数智能防护系统为单一业务场景研发，通用性较弱，难以适配企业多业务、跨平台的异构数据体系，极易出现防护断层、适配故障等问题，无法实现全域安全

防护。

3.2 管理与体系层面存在的问题

（1）数智化数据安全制度规范体系尚不健全。现有安全管理制度多适配传统数据防护模式，针对AI风控、数据智能脱敏、联邦学习等新型数智技术的应用标准、管理规范缺失，存在治理盲区与合规漏洞。（2）复合型数据安全人才储备严重不足。数智融合数据安全涉及大数据、人工智能、网络安全、合规管理等多领域知识，行业内单一技能人才居多，兼具算法运维、安全防护、合规治理能力的复合型人才缺口较大。（3）行业智能化安全防护落地执行力不足。多数企业存在“重建设、轻运营”问题，智能防护系统部署后缺乏常态化迭代与运维优化，防护策略与业务场景脱节，导致智能化防护技术难以发挥实际效用。

3.3 技术优化创新对策

（1）优化AI算法模型，降低智能风控误判率。通过扩充多元、实时、均衡的训练样本数据集，修正算法固有偏差，引入动态迭代学习机制，让模型持续适配新型风险场景。同时融合多算法协同研判模式，优化风险判定逻辑，有效减少漏判、误判问题，提升智能识别与风控的精准度。（2）搭建轻量化、高效化大数据智能安全处理架构。重构数据处理架构，优化算力调度机制，对海量数据进行分层、分级、分流处理，筛选核心风险监测数据，精简无效数据运算。通过轻量化算法适配实时数据处理场景，提升高维海量数据的处理速度与效率，保障安全防护的实时性与稳定性[5]。（3）研发多场景适配的模块化智能安全防护系统。采用模块化设计思路，拆分风险感知、加密防护、脱敏处理、应急响应等核心功能模块，依据不同业务场景、数据类型灵活组合适配。统一系统接口标准，解决跨平台、多场景兼容问题，实现智能安全技术的全域适配与高效应用。

3.4 管理与保障体系优化策略

（1）完善数智融合数据安全治理制度与行业标准。结合《数据安全法》等法律法规，贴合数智技术应用场景，细化智能防护、数据共享、算法应用等环节的管理制度与行业标准。明确各主体安全责任与操作规范，填补数智化数据安全治理的制度空白，实现规范化、标准化治理。（2）构建复合型数据安全人才培养与引进体系。高校增设数智数据安全交叉学科，完善人才培养体系；企业搭建内部培训机制，开展算法技术、安全运维、合规管理专项培训。同时加大行业高端人才引进力度，搭建人才激励机制，补齐复合型人才短板。（3）建立技术、管理、监管一体化的数据安全保障机制。整合智能技术防护、常态化内部管理、全方位外部监管资源，构建一体化闭环保障体系。依托智能监管平台实现动态监测，搭配标准化管理制度与常态化督导机制，实现技术落地、日常管理、合规监管的深度融合，全面提升数据安全保障能力。

结束语

大数据与人工智能的融合，彻底革新了传统数据安全防护模式，实现了数据安全从被动处置向主动防护的转型。本文梳理了数智融合赋能数据安全的创新技术路径，剖析了技术适配、管理体系、人才储备等现存问题，并提出全方位优化策略。未来，需持续迭代智能安全技术、完善治理体系，深化数智技术在数据安全领域的应用，筑牢全域、智能、长效的数据安全屏障。

参考文献

[1]梁华国.区块链技术在网络数据安全共享中的应用[J].智库时代,2022,19(39):115-119.

[2]蒋忠均.人工智能驱动的数据安全治理框架与动态防御机制研究[J].中国宽带,2025,21(7):55-57.

[3]吴沈括,石嘉黎.数据安全视域下的人工智能风险应对研究[J].西北工业大学学报(社会科学版),2023,12(2):95-103.

[4]杨芳,韩钰.简议人工智能背景下的隐私与数据安全保护[J].佳木斯职业学院学报,2024,28(7):426-429.

[5]缪文升.人工智能时代个人信息数据安全问题的法律规制[J].广西社会科学,2022,10(9):101-106.