

金融数据安全治理智能化技术与实践

侯鹏^{1,2}, 李智鑫^{1,2}, 张飞¹, 孙旭^{1,2}, 陈丹^{1,2}, 崔毅浩^{1,2}, 张寒冰^{1,2}, 荆一楠¹, 柴洪峰^{1,2}

(1. 复旦大学计算机科学技术学院, 上海 200438; 2. 复旦大学金融科技研究院, 上海 200438)

摘要: 在“十四五”新时期, 金融业数字化从多点突破迈入深化和高质量发展新阶段, 加强金融数据安全治理、保护金融数据安全, 已经成为国家经济安全发展的客观需要和必然要求。以数据治理措施为基础工具, 以敏感数据为核心关注点, 以保障金融数据全生命周期安全为目标的金融数据安全治理是实现金融机构数据流通、激活数据价值、促进金融数据要素市场化配置的有效措施。随着新一代信息技术在金融领域的广泛应用, 金融数据安全治理的方式也正和大数据、人工智能、云计算、区块链等技术紧密结合, 从传统的数据安全治理方式转变为智能化治理, 驱动金融数据安全治理加速向自动化、智能化、高效化、精准化方向演进。介绍了金融数据安全治理的内涵、治理对象和治理体系; 阐述了金融数据安全治理向智能化发展的核心理念和关键支撑技术, 并描绘了治理智能化的发展路线。结合金融数据安全治理需求和智能化技术特点, 从数据分级、数据溯源、内容管控、隐私保护和数据孪生等方面说明了金融数据安全治理智能化技术实践的关键方向, 并举例介绍了银行、证券和保险业当前的治理智能化应用实践。为推动我国新时期金融数据安全治理规范化和智能化发展, 实现持续金融机构和行业的健康发展, 提出了相应的政策建议。

关键词: 金融数据安全; 数据治理; 金融科技; 人工智能; 智慧金融

中图分类号: TP309; F832

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023048

Technology and practice of intelligent governance for financial data security

HOU Peng^{1,2}, LI Zhixin^{1,2}, ZHANG Fei¹, SUN Xu^{1,2}, CHEN Dan^{1,2}, CUI Yihao^{1,2},
ZHANG Hanbing^{1,2}, JIN Yinan¹, CHAI Hongfeng^{1,2}

1. School of Computer Science, Fudan University, Shanghai 200438, China

2. Institute of Financial Technology, Fudan University, Shanghai 200438, China

Abstract: In the “14th Five-Year Plan” era, the digitization of the financial industry has entered a phase of deepening and high-quality development. Strengthening Financial Data Security Governance (FDSG) and protecting financial data security have become an objective need and essential requirements for developing national economic security. FDSG utilizes data governance measures as a fundamental tool with a focus on sensitive data to ensure the comprehensive

收稿日期: 2023-01-06; 修回日期: 2023-05-30

通信作者: 荆一楠, jingyn@fudan.edu.cn

基金项目: 中国工程院战略研究与咨询项目 (2022-XY-12, 2023-33-14)

Foundation Item: Strategic Research and Consulting Program of the Chinese Academy of Engineering (2022-XY-12)

引用格式: 侯鹏, 李智鑫, 张飞, 等. 金融数据安全治理智能化技术与实践[J]. 网络与信息安全学报, 2023, 9(3): 174-187.

Citation Format: HOU P, LI Z X, ZHANG F, et al. Technology and practice of intelligent governance for financial data security[J]. Chinese Journal of Network and Information Security, 2023, 9(3): 174-187.

security of the entire lifecycle of financial data. It aims to promote data circulation among financial institutions, activate data value, and facilitate market-oriented allocation of financial data elements. FDSG is increasingly intertwined with big data, artificial intelligence, cloud computing, and blockchain technologies, transforming traditional data security governance into intelligent governance and accelerating the evolution of FDSG towards automation, intelligence, efficiency, and precision. FDSG's essence, scope, and governance framework were introduced, and the core concepts and critical supporting technologies for the intellectual development of FDSG were elaborated, outlining the roadmap for intelligent governance. With consideration of FDSG's demands and characteristics of intelligent technologies, vital directions for the practical application of intelligent technologies in FDSG were outlined, including data classification, data traceability, content control, privacy protection, and data twinning. Examples of current intelligent governance practices in banking, securities, and insurance industries were provided. Policy recommendations were proposed in the paper to promote the standardized and intelligent development of FDSG in China's new era, aiming to foster the sustainable development of financial institutions and industries and ensure a healthy and secure environment for the future.

Keywords: financial data security, data governance, fintech, artificial intelligence, smart finance

0 引言

建设数字中国是“十四五”时期的重要任务，以数据为核心的数字技术成为经济发展的新动能，正在改变技术进步和经济发展的格局。在金融领域，银行、证券、保险等金融机构在多年的业务经营和数字化发展过程中积累了海量的金融数据，且金融数据随着应用场景和应用机构的爆炸式增长，变得更加丰富和复杂多样，其价值密度高、应用价值大的特点也越来越突出。然而，金融数据具有高敏感性和强隐私性等特点，其在快速流转和使用过程中所蕴含的数据安全风险也被逐渐放大。

金融数据是金融业的核心要素，在金融数字化转型过程中，高效地协调包括政府、企业和个人在内的多个相关主体共同参与金融数据安全的建设，系统、全面、有效地建设和提升金融数据安全防护能力并切实保障金融数据价值发挥是金融机构需要关注和亟需解决的核心问题。充分的实践已经证明，金融数据治理是实现金融数据价值释放和保护数据安全的有效举措^[1]。推进金融数据安全治理、促进金融数据流通、提升金融数据价值对金融业高质量发展具有重大意义，在防止金融数据滥用、缓解金融数据市场的信息不对称、降低金融风险传染、监管跨境数据流动和提升国家金融治理水平等方面能够发挥重大作用。

当前，各金融业机构数据安全治理技术能力和

治理规范尚处于参差不齐的状态，但云计算、大数据、人工智能（AI，artificial intelligence）和区块链等技术在金融行业的普及和应用，为金融业务发展赋予了更多的科技智能化能力。金融业可以综合运用大数据、云计算、人工智能等技术具备的数据分析、知识提取、自主学习和智能决策能力，提升金融数据安全智能化治理能力，增强金融数据安全保障能力，推动新时期金融行业智能化发展^[2]。

1 金融数据安全治理

本节介绍金融数据安全治理的内涵、对象，说明金融数据安全治理的本质与范围。

1.1 金融数据安全治理的内涵

金融数据安全治理是金融行业在数字化转型过程中为了提升金融数据资产价值并降低数据安全风险而开展的管理活动，即在金融数据应用场景下，以数据资产为基础，围绕数据全生命周期环节，结合企业或者组织自身数据现状，从而建立与制度流程配套的技术和工具，包含法律规定、行业规范、组织管理、技术支撑、业务创新等在内的由多主体参与的体系化数据管理^[3]。

金融数据安全治理与传统网络安全和数据治理既有紧密的关联性，又有面向金融数据的独特性。金融数据安全治理是以金融数据安全使用为目标，以数据分类、分级为基础，以信息合理、安全流动为目标，兼顾发展与安全，保障手段不仅包括金融数据使用过程中的技术手段，而且包

括安全管理手段,并且需要实现安全产品技术和流程管理深度融合。但金融安全治理并非是要取代数据治理和传统网络安全体系,而是在数据治理和传统网络安全体系的基础上有效地实现对金融数据的安全管控。

1.2 金融数据安全治理的对象

金融数据安全治理的对象包括“域内”和“域外”两种来源的金融数据。金融数据不仅具备一般数据的特性,而且存储着大量重要信息,包括显式信息和隐式信息,如个人信息、商业机密信

息等。金融数据呈现出产生主体的多元性、运用场景的广泛性、运用价值的特殊性和数据利用的不对称性等众多特征^[4]。

金融数据安全治理对象模型如图1所示,来自金融机构内部的客户数据、业务数据、经营数据和监管数据等被视为“域内”数据,而来自金融机构之外的合作方数据、交易获得的数据等被视为“域外”数据。“域内”和“域外”数据模式涵盖了金融数据的所有来源。总之,凡是金融领域所涉及的相关数据都可以被纳入金融数据安全治理对象中^[5]。

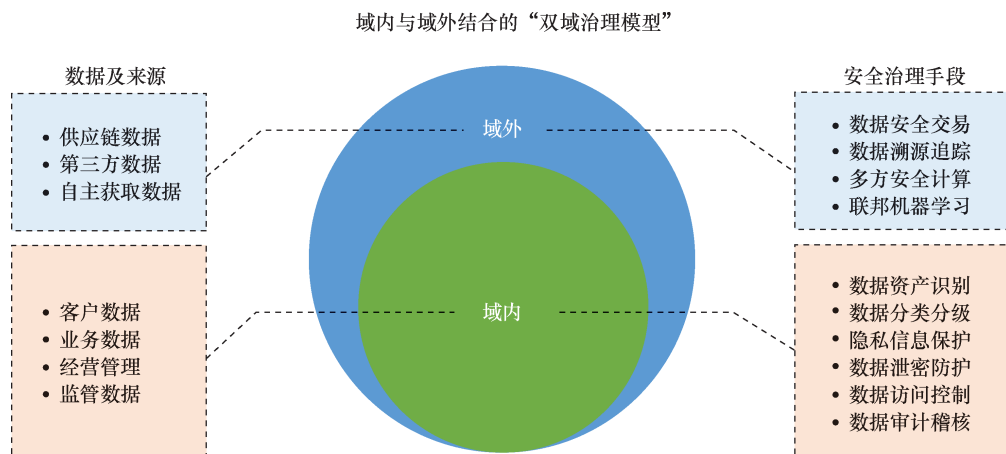


图1 金融数据安全治理对象模型
Figure 1 Financial data security governance object model

(1) “域内”数据

“域内”数据指由金融机构(银行、保险、证券等)、厂商(信息服务提供商等)、个人和政府机构(行政、监管等部门)在进行支付结算、股票投资决策、成本定价、期货期权交易、债券投资、资金拆借、货币发行、票据贴现与再贴现等业务交易及相关行为时产生的信息,不仅包含原始的业务活动生成的数据,而且包括经过数据排序、聚合或者过滤等数据处理操作转变而来的信息数据。在《金融数据安全 数据安全分级指南》中按数据来源将典型金融数据分为客户、业务、经营管理和监管4类。

(2) “域外”数据

“域外”数据主要指来自金融机构之外的其他金融数据,包括供应链数据、第三方数据和自主获取的数据等。首先,金融机构自身的供应链信息并不是全面的,无法完全支撑所有的供应链金融服务,金融机构需要和具有这些信息的电商平

台(如阿里、京东等)合作。其次,经过第三方分析整理的研究数据,包括消费者行为数据、行业分析报告、竞争与市场数据、宏观经济数据、特殊定制数据等可以被金融机构所利用,金融机构可以利用第三方的专业分析报告来制定自身风险偏好,同时为自身的大数据分析、产品开发、风险管理提供决策支持。此外,金融机构可以通过网络爬取、合作等方式获取到其他行业数据,开展与其他行业的合作发展,提升金融业务范围和质量。

2 金融数据安全治理体系

金融数据安全治理是跨技术、法律和社会的交叉领域,也是涉及技术、管理和业务的多层面活动,需要在保证金融数据安全的前提下充分释放数据价值,全方位提升金融发展质量。数据安全治理专业委员会以数据安全管控策略为核心,以管理体系为指导,以运营体系为纽带,以技术体系为支撑的治理框架构建思路,构筑形成组织、

管理、技术、运营四位一体的金融数据安全治理框架^[3]，如图2所示。

金融数据安全治理体系在国家法律法规和行业政策指引下由运营体系、组织架构、安全管理体系、安全技术体系和数据实践体系5个方面组成，涵盖了金融数据的全生命周期。该体系能够协调技术、人员、产品等各个要素，并防范多层次、多维度防范信息安全风险，包括规范的制度体系、先进的技术防护体系、成熟的安全服务机制以及安全监测预警体系，并能够持续监测评估安全态势与管控目标的差距，定期审计持续优化试点运行效果，反哺管理、技术、运营体系，不断提升金融数据安全治理效果。因此，金融数据安全治理不是仅仅依靠一方力量就可以开展的工作，从顶层设计的政府部门、制定金融规范的行业协会、开展金融活动的金融企业到个人消费者都是金融数据安全治理的参与者，多方主体需要共同发力，共同建设金融数据安全治理体系。

3 金融数据安全治理智能化

2022年，由中国人民银行发布的《金融科技发展规划（2022—2025年）》提出要积极应用多方安全计算、联邦学习、差分隐私、联盟链、图计算、数字孪生、知识图谱、大数据、云计算和边缘计算等技术聚焦金融科技应用前沿，发挥数据和技术双轮驱动作用。由此可见，国家、地方

和行业方面都积极推进金融智能创新发展，支持以人工智能和隐私计算为代表的新一代信息技术在金融风险防控、业务应用、科技监管等领域的广泛应用。金融数据安全治理的方式和技术也需要和大数据、人工智能、云计算、区块链等技术紧密结合，从传统的数据安全治理方式转变为智能化治理，推动金融数据安全治理在管理模式、思维模式和技术体系上的变革，驱动金融数据安全治理加速向自动化、智能化、高效化、精准化方向演进^[6]。

3.1 智能化关键技术

为实现金融数据安全治理的智能化，需要一系列的智能化技术支撑。如表1所示，金融数据安全治理过程中存在的一系列问题都有相应的技术解决方案，以机器学习、大数据、云计算、区块链等技术工具为支撑，为金融数据安全治理赋予数据安全治理智慧的同时，数据安全治理为上述技术在金融行业的创新发展提供了前驱动力。

（1）人工智能

人工智能关键技术包括机器学习、深度学习、人机交互、自然语言处理、机器视觉等。在金融数据安全治理方面，人工智能在数据模型管理、元数据管理、主数据管理、数据治理管理、数据质量管理、数据要素流通和数据安全等方面都可以提供更加高效的解决方案^[7]。例如，人工智能与知识图谱的结合可以建立完整、可扩展、稳定

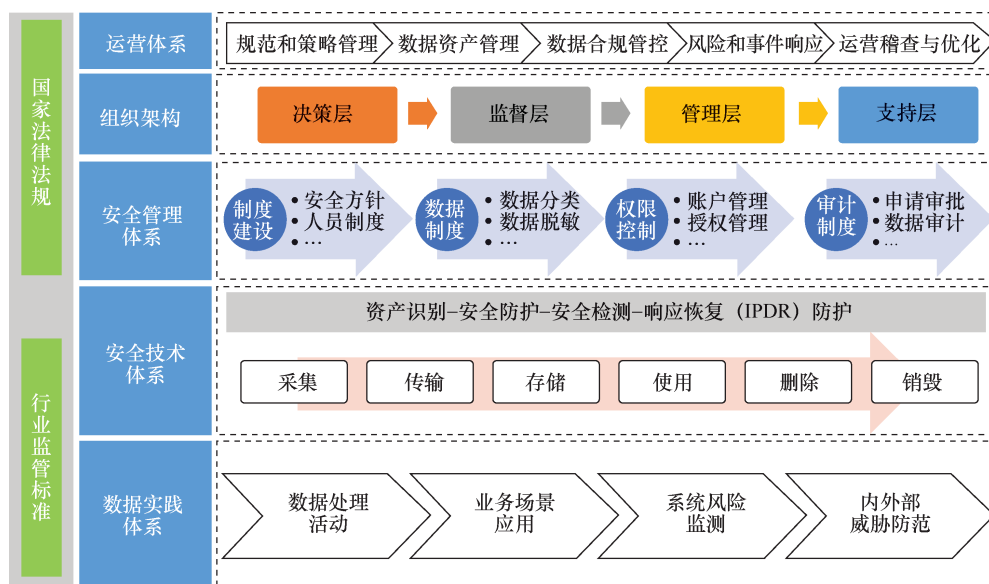


图2 金融数据安全治理框架

Figure 2 Financial data security governance framework

的数据模型，以此清楚地表达数据、企业、业务等主体时间的相关性；人工智能可以实现对非结构化数据的采集以及关键信息的提取，在予以筛选等方面帮助维护、整合元数据；机器学习、自然语言处理可以帮助建立数据匹配规则，辅助金融机构识别主数据；借助机器学习和自然语言处理对金融数据库分析建立数据标准管理体系；结合隐私计算、语义计算、认知计算等加强金融数据的安全使用和流通，促进金融数据安全保障体系的完善；通过机器学习自主识别数据质量并自动对数据进行评估和修复；联合区块链建立安全可信的跨部门、跨机构的信息交互，提高金融数据要素流通效率；通过机器学习的训练和推理完成人工智能模型的迭代优化，针对不同场景和不同业务需求开展个性化金融服务，如反欺诈、个性化推荐、数据分类等。

表 1 金融数据安全治理智能化技术的应用
Table 1 Application of intelligent technologies for financial data security governance

问题	技术
金融数据规模大，人力处理效率低，易出错	机器学习、大数据技术、云计算等
数据模型不清晰、元数据质量低	机器学习、知识图谱、大数据等
重复性工作，人工成本高，效率低	深度学习、云计算、计算机视觉等
无法为客户提供定制化金融服务	语音识别、人脸识别、人工智能等
高度依赖专业人员分析判断（如敏感信息识别、数据分类分级等）	自然语言处理、大数据技术等
数据资产管理混乱	人工智能、知识图谱、自然语言处理等
智能化组织管理	深度学习、人脸识别、数据指纹等
信息不对称、信息孤岛	隐私计算、区块链技术、可解释人工智能等
数据和系统安全防护	多方安全计算、云计算、深度学习等
风险评估与反欺诈	知识图谱、机器学习等

(2) 区块链

区块链是去中心化的、由节点参与的分布式数据库系统，具有难以伪造和全程溯源的特征。区块链自身的技术特性使其天生就具备可对数据进行组织管理的特点，区块链能够在以下几个方面为金融数据安全治理提供技术支撑：① 有助于提高金融数据质量，区块链的认证、共识、上链存证、不易篡改等特性保障了金融数据的一致性

和可追溯性，能够有效地保障金融数据的真实、完整；② 有助于保障金融数据安全，区块链融合了非对称加密认证算法、数据隐私保护匿名算法，并且具有时间戳的区块链数据提供了验证和记录机制，这样不仅能够保证数据安全，而且可以解决用户身份认证和隐私保护问题；③ 有助于实现金融数据共享，根据区块链的去中心化思想，不同的机构和部门可以建立点对点的联系，智能合约技术和全网共识机制可以建立统一的标准和流程，保障可获取金融数据的一致和可靠，此外，多方参与共识特性加快了信息公示、数据确权等进程^[8]。

(3) 隐私计算

隐私计算是密码学、机器学习和数据科学等多种技术的交叉融合^[9]，金融业对于数据安全和隐私保护高要求、严监管的现状对隐私计算在金融场景中的落地起到了极大的推动作用。隐私计算已逐渐成为金融数据安全治理的重要技术手段。隐私计算使得金融数据以可用不可见的方式被高效利用，在金融数据隐私保护、促进金融数据流通、联合提升金融数据价值等方面发挥了重要作用，如隐私计算可以有效控制金融数据的用途以及用量，保证在安全的前提下尽可能使数据价值最大化，解决“数据孤岛”问题，加速金融数据要素安全流通^[10]，助力多方合作挖掘多业务场景数据价值。此外，隐私计算可以融合区块链技术来强化在“身份认证、访问控制、计算和监管”等方面的信任机制，与大数据分析技术结合起来提高数据应用价值。

(4) 大数据

在金融数据安全治理过程中，大数据技术可用于金融数据的全生命周期。① 大数据平台将安全措施优先用于金融数据生命周期的收集阶段；② 在金融数据存储阶段，收集的数据被存储于大型的数据中心供下一阶段使用；③ 对于大数据的传输安全，可通过加密算法等密码学技术来确保数据传输中的机密性；④ 在金融数据的使用和共享环节，访问控制框架、数字水印、数据加密等技术确保各行业领域大数据共享和整合使用；⑤ 在金融数据的销毁阶段，通过数据安全删除策略、数据安全删除系统模型等数据可信删除技术来保证数据的安全销毁。

(5) 云计算

从IT运营角度看,金融机构需要经常开展算力升级、存储升级和软件升级,而依靠云服务商提供的云计算服务,可提升金融数据合规工作执行的灵活性和可扩展性,降低合规应用软件开发部署的成本。在金融数据安全治理中,云计算可有效提供以下支撑:① 可以提供强大的数据分析引擎;② 能够提供更好的数据冗余备份和海量存储能力;③ 为推动人工智能算法的应用提供高要求算力支撑;④ 可实现按需操控、安全管控、访问和共享数据。

(6) 其他技术

除上述技术之外,还有一些新兴技术如数字孪生、量子计算等,虽然目前还未达到可以进行广泛的应用,但随着技术的快速发展和革新,这些技术也将为金融数据安全治理提供有力补充。

3.2 智能化发展路线

(1) 厘清数据安全治理智能化概念、思路

确定智能数据安全的治理理念,完善智能数据安全技术的发展思路。坚持安全与发展并重的原则,即在发展的基础上,智能数据治理应始终以安全为轴心运行。智能数据安全技术与智能数据安全治理理念相互促进,治理理念可以为智能安全技术的发展提供指导,技术的进步可以进一步促进治理理念的创新。

(2) 建立数据安全治理智能化体系

金融数据安全智能治理是一个多主体、多场景、多工具的立体协同治理,应建立技术与制度并重的治理体系,协作多方共同努力。在技术与制度双重运作的前提下,首先支持人工智能数据安全评估体系的发展,并加大科技伦理体系的研究,推动技术评估平台和算法训练数据集的建设。此外,需要注重管理和运营方面的建设,实现治理的体系化和规范化。

(3) 提出数据安全治理智能化、场景化治理方案

金融数据安全治理的核心在于场景化安全。不同用户基于业务、访问途径、使用需求,会产生不同的使用场景。在保证金融数据被正常流转、使用的目标下,需要根据不同场景的特点制定相应的智能化治理方案和安全策略。场景化的金融数据安全

治理,能够及时发现金融数据面临的风险,使金融数据安全治理更具针对性,更安全地使用金融数据。

(4) 聚焦数据安全治理智能化重点领域

金融数据安全治理涉及金融数据的方方面面,是一个自上而下的复杂体系化工程,治理难度与业务类型和数据规模紧密相关。此外,智能化技术自身具有安全风险,并在金融数据安全治理过程中容易导致新型金融风险的产生,如人工智能面临着数据投毒、数据滥用等问题,区块链的安全性和可靠性还未完全成熟等。因此,金融数据安全治理的智能化要在保障金融数据安全的基础上聚焦如数据分类、数据全生命周期管理、数据共享等重点领域,并在较小范围内开展试点工作,待技术成熟后由点及面地全方位展开应用。

(5) 完善数据安全治理智能化全生命周期全流程

金融数据是金融数据安全治理的核心,金融数据的全生命周期覆盖了金融数据的采集、传输、存储、使用、应用、销毁的整个过程,代表了金融数据安全治理的一条主线。生命周期的不同阶段对金融数据的安全防护机制、安全管理制度等有不同的需求,应根据金融数据在全生命周期中的流转明确合理的治理目标,建立完善的技术解决方案、管理和运营制度;促进金融数据的高效流通和安全共享,提升金融数据价值释放潜力。

4 金融数据安全治理智能化技术实践

智能化技术全面提升金融数据安全治理质量和治理效率,为金融数据安全治理智能化转型赋能。结合智能化技术,本文提出以下技术实践提升金融数据安全智能化治理能力,增强金融数据安全保障能力。

4.1 自动敏感资产盘点与数据安全分级

金融数据来源广泛,数据类型复杂多样,不仅具备一般数据的特征,而且蕴含了众多个人敏感信息和机构的机密信息。对复杂多样的金融数据开展数据资产盘点和数据安全分级是金融机构开展金融数据安全治理的基础,有助于金融机构建立清晰的资产目录,并明确数据安全防护策略^[1]。2020年中国人民银行正式发布了《金融数据安全

数据安全分级指南》行业标准,对金融机构开展数据分类分级提出了明确的实践指导和评估要求。但是,目前金融数据资产的盘点还存在以下问题:① 数据资产定义不完善;② 资产盘点基于人工,工作效率低,准确率低。因此,自动化数据资产盘点成为迫切的需求。

智能化数据资产盘点与分类分级技术框架

如图3所示,首先应确定决策机构、牵头部门、实施部门等,建立组织保障;之后利用人工智能等技术对数据资产进行智能化盘点;综合国家、行业和自身特点,确定分类分级策略,利用AI算法,融合小样本学习及无监督学习等前沿AI技术,快速建立数据分类分级能力;最后,根据实际需求完成分类分级标签落地,并持续迭代。

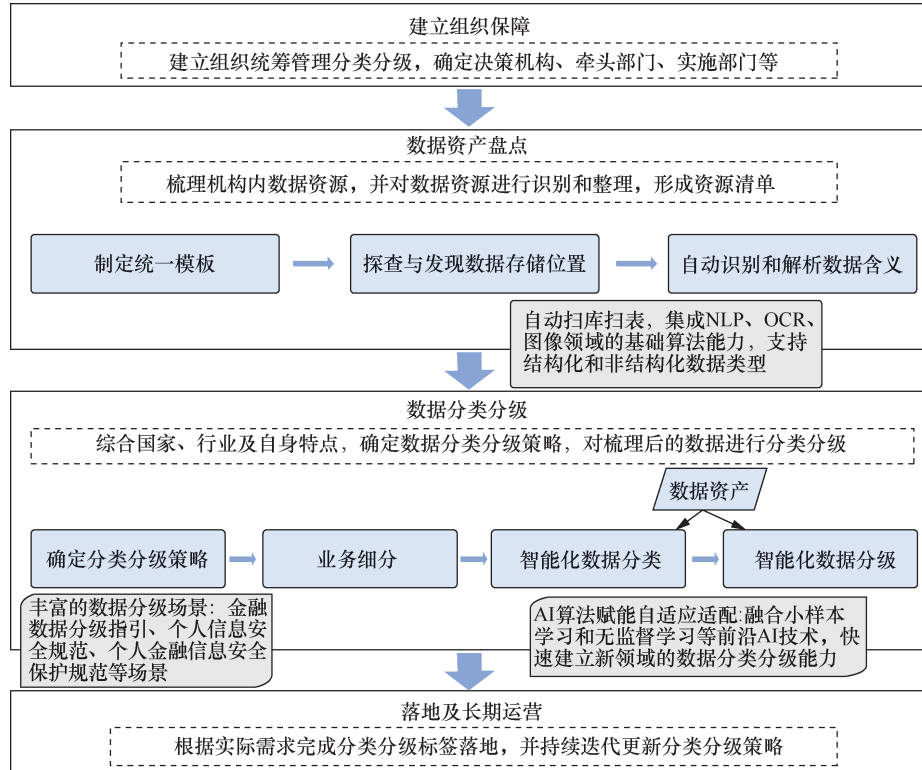


图3 智能化金融数据资产盘点与分类分级技术框架

Figure 3 Intelligent financial data asset counting and classifying technology framework

4.2 基于数据血缘的数据资产溯源管理

数据溯源是记录数据从产生到消亡或转换的整个生命周期内所发生的变化和经过处理的信息,包括源头数据的追溯与描述,过程信息的追溯、捕获或记录,影响数据状态的因素(如影响数据的实体、工具等)的追溯、描述和记录^[12]。金融领域对数据的真实性有很高的要求,在金融数据的全生命周期流转过程中,原始数据可能被非法篡改、非法窃取,也有可能发生畸变,导致数据缺失或者不可用,进而影响数据的完整性和可行性。因此,在金融数据流转过程中,需要通过原始金融数据的全生命周期进行溯源和记录,以便开展数据审计和漏洞风险弥补,有针对性地防范数据风险,是降低数据风险的重要手段之一^[13]。

数据资产溯源管理模型如图4所示,主要包括数据溯源信息描述模型、数据溯源通用表达模型、数据溯源安全管理模型和区块链溯源管理模型4个内容^[12]。具体地,包含W3模型、W7模型和更为精细的W7+R3模型^[12]在内的数据溯源信息描述模型可以应对不同的需求,是数据溯源的基础,主要用于对数据溯源信息进行获取并记录。数据溯源通用表达模型主要用于规范数据溯源过程,使得不同场景下的数据溯源遵循一致的表达方式,由此提升数据溯源的扩展性,常见的表达模型是开放溯源模型(OPM, open provenance model)和改进后的数据起源模型(PROV-DM, provenance data model)、(PRroVOC, provenance vocabulary model)^[13]。数据溯源安全管理模型的

作用是保护数据溯源信息安全，包括来自外部和内部的安全威胁。此外，具有不可伪造、全程溯源和强一致性特征的区块链技术非常适合数据资产的溯源管理，能够有效降低溯源成本，并弥补传统溯源技术的账本篡改、信息丢失等情况。

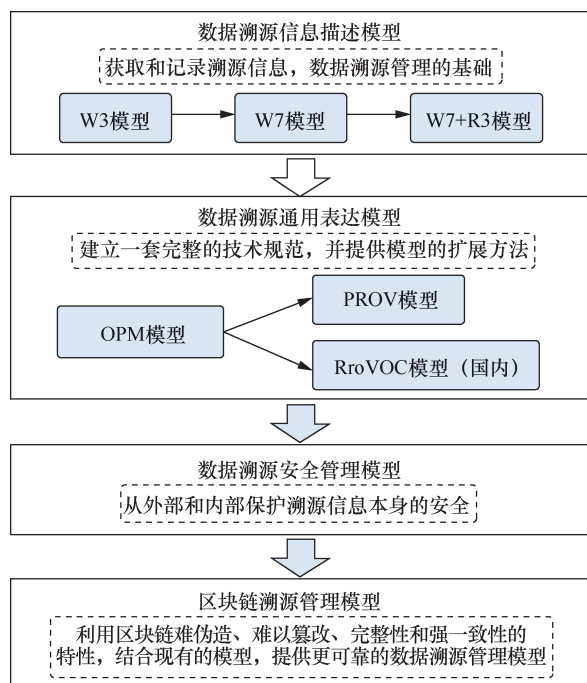


图4 数据资产溯源管理模型

Figure 4 Data asset traceability management model

4.3 基于数据内容的动态矩阵管控机制

金融行业业务多元复杂，金融数据被非授权对象获取、利用，甚至以此进行窃取、篡改、销毁或拒绝服务，数据的未经授权披露或传播等所造成的后果会对金融机构造成极其恶劣的影响，访问控制通过对数据的访问进行全量审计，对访问者的权限进行细粒度的控制，对高危行为进行实时阻断告警来防止敏感数据泄露、保证数据库的审计与防护、满足金融行业合规性要求。因此，建立高效、动态、安全的数据访问控制系统是至关重要的，访问控制技术将在金融数据安全治理领域内发挥重要作用：① 防止非法用户对金融数据和系统的访问；② 允许合法用户访问受保护的金融数据；③ 防止合法用户对金融数据进行非授权访问^[14]。

基于数据内容的访问控制技术实现机制如图5所示，主要包括5个步骤：① 访问者发起数据访问请求后由管理层进行身份认证并确定访问权限，

并进一步由控制层确定访问控制策略，同时记录全流程访问轨迹以用作审计；② 管理层执行数据加解密授权，访问者访问并加载数据加密组件并开始对数据进行处理；③ 在访问和使用数据过程中，系统监测操作的合法性，并通过关键字或关键段提取对敏感信息加以控制；④ 在与敏感数据库对比后，如果数据敏感程度较高，则控制数据访问权限，并执行数据脱敏操作，包括权限控制、数据筛选、数据变形、数据遮蔽和数据漂白等；⑤ 完成数据脱敏输出数据时要检验输出数据与用户意图获取信息的一致性或关联性，若关联性和一致性较高则输出，否则对事件审计并向管理层报备^[15]。

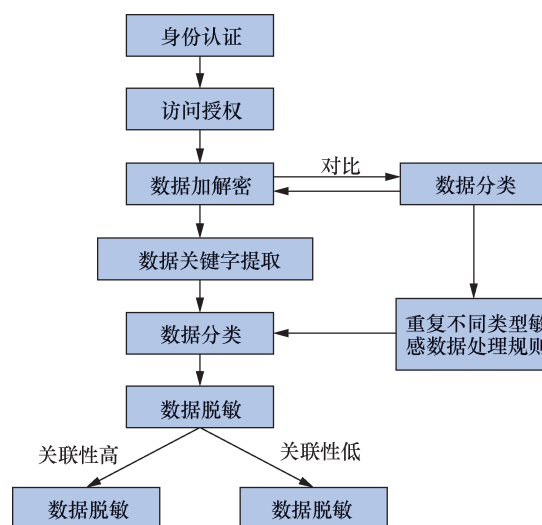


图5 基于数据内容的访问控制技术实现机制

Figure 5 Data content-based access control techniques

4.4 基于隐私计算的可信金融隐私保护

隐私计算是由两个及以上数据参与方在保障原始数据安全前提下进行联合计算的一种技术。随着金融领域数据分析、数据挖掘成为不可或缺的热门技术和业务规模的快速增长与发展需要，各金融机构对数据跨系统、跨领域及跨境交互融合从而有效释放数据资源价值的需求愈加迫切。如何借助隐私计算等技术手段保障网络信息安全、保护用户隐私数据成为各金融机构面临的重要课题。而隐私计算技术则是各机构充分践行国家大数据战略、严格满足数据监管、积极推进数据流通共享的重要解决方案。隐私计算较好地解决了数据安全与隐私保护、多方合作的技术信任、监管合规和监管穿透问题以及监管手段信息滞后问题。

目前主流的隐私计算技术可以分为三类,分别是基于硬件的可执行环境、基于统计学安全的联邦学习和差分隐私,以及基于密码学安全的密码算法和多方安全计算(如图6所示)^[16]。其中,多方安全计算是应用最为广泛的技术方案,但隐私计算不同技术之间是可以相互融合的,因此实际业务实践中多方安全计算也会结合同态加密和联邦学习等技术开展联合建模。隐私计算在金融数据安全治理领域的应用主要包括联合管理、智能风控和智能营销3个方面。① 联合管理:金融数据的管理、使用、审计等通常涉及多个部门、机构,基于隐私计算实现数据的所有权和使用权分离后,可以在保证金融数据安全可靠的前提下,实现跨部门、机构和企业的联合管理,构建多方安

全可信的数据协同模式。② 智能风控:隐私计算通过分布式学习的方式联合来自不同机构和不同平台的数据来构建全局数据风险控制模型,极大地提高了风控模型的训练效率和精度,实现由局部最优到全局最优的转化,并且能够充分利用来自不同场景、不同业务、不同人群的金融数据,提升风控模型的精准化^[17]。③ 智能营销:隐私计算由于联合了来自不同渠道的数据,能够精准识别客户意愿,构建更加精准的用户画像,实现用户需求、数据画像和产品营销之间的高质量匹配,有效提升客户体验质量,还可以通过数据共享的方式不断地优化营销策略,动态提升业务质量。此外,隐私计算可以融合区块链技术增强金融数据安全治理在身份认证、访问控制等方面的安全防护。

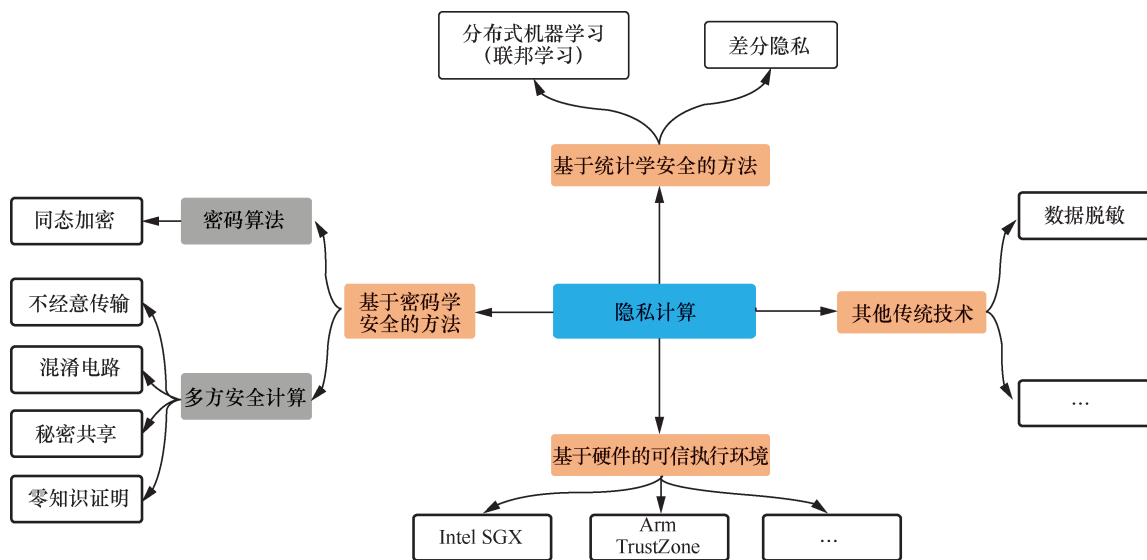


图6 隐私计算技术分类

Figure 6 Privacy computing technology classification

4.5 基于数据孪生的金融数据合成与共享

金融数据中的敏感数据一旦以真实状态泄露,或遭到恶意破坏,不仅有关金融机构的品牌声誉,还会为其客户带来一系列潜在安全威胁和商业损失,最终将严重影响金融机构自身的正常经营发展。此外,在敏感性高、隐私性强的金融领域,数据的获取极为困难,即使是在金融企业内部,各部门间的数据共享都是一件极难实现的事情,通常共享数据需要非常繁杂的审批流程,这个过程往往需要耗费大量的时间。

金融机构可以利用数字孪生技术来克服隐私和数据共享的挑战。基于人工智能的金融数据数

字孪生技术通过自动学习原始数据的特征来生成在统计上和结构上与原始数据相似的合成数据。由于数据是合成的,没有映射到任何原始数据,因此可以自由使用、共享和存储。基于数字孪生技术的合成数据可以为金融业数据的治理提供基础保障,尤其是在客户获取和高级分析、抵押贷款分析、信用决策和限额评估、风险管理和定价、欺诈和异常检测、监控和收集、ESG(environmental, social and governance)数据补全等多个场景中可以得到广泛应用^[18],可以为金融机构在数据隐私性与可用性方面提供数据安全性保证、数据快速访问和数据规模可控、保护数据公平性等多方面的保障。

此外,数据孪生技术在数据增强方面也有很大的潜力,在供给数据的时候可以结合人工智能的相关技术帮助筛选、调整和管理数据,让供给出去的数据质量更高并提高数据价值。

目前在金融数据领域中主流的金融数据孪生框架大致可以分为4个阶段(如图7所示)。**① 数据预处理**:对原始数据进行处理,从而将原始数据中一些不适合用来建模的数据利用机器学习等相关技术处理为适合一系列模型处理的输入数据,如一些离散型数据或者数据范围跨度很大的连续型数据。需要注意的是,经过转换处理之后数据还需要保留原始数据的所有特征。**② 数据建模**:对经过转换处理之后的数据进行采样得到训练数据,然后既可以利用神经网络架构中的生成对抗网络(GAN, generative adversarial network)、变分自编码器(VAE, variational auto-encoder),也可以使用统计方法中的Copula函数、贝叶斯网络等一系列方

法对数据进行建模^[19],在建模过程中也可以利用相关技术保证数据的隐私不会被泄露,最终使得模型能够获取数据分布、数据属性间的联系等一系列特征,并得到一个保护隐私的数据生成器。**③ 数据生成**:将随机噪声(如高斯噪声)作为生成器的输入,得到一系列数据特征的输出,利用和数据预处理阶段相同的方法将数据转换为原始数据的格式,从而得到一份和原始数据高度相似的合成数据。**④ 合成数据评估**:对合成的数据进行评估,主要包括数据可用性和数据隐私程度两个维度的评估。其中,数据可用性一般是在该合成数据集与对应的原始数据集上分别训练分类、聚类、回归等机器学习模型,利用两个模型的性能指标差距(如分类准确度、回归预测效果的差异性)作为衡量指标。数据隐私保护程度的度量一般采用和原始数据匹配的命中率、同原始数据之间的距离等指标进行评价,最后得到两份数据质量评估报告。

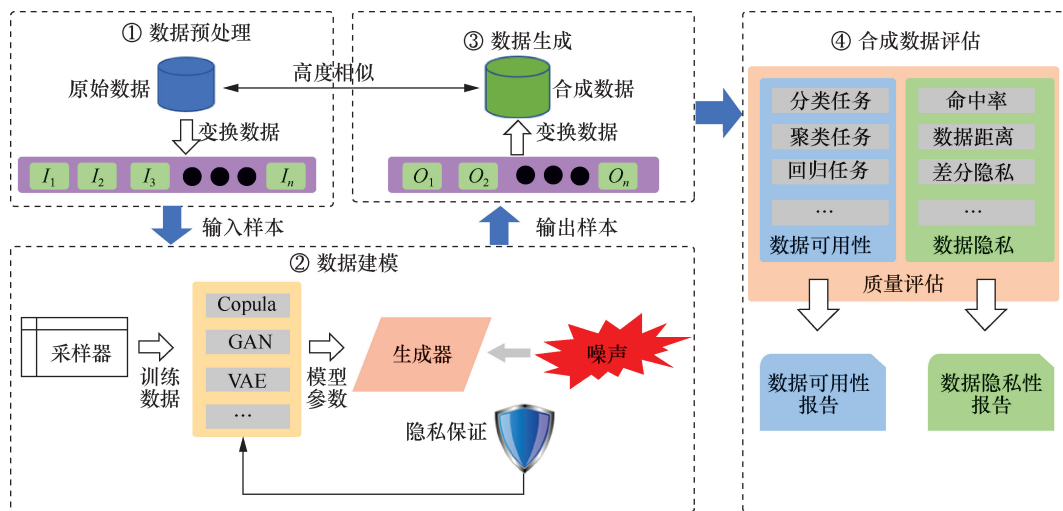


图7 金融数据孪生框架
Figure 7 Financial data twins framework

5 金融数据安全治理智能化应用实践

5.1 银行业的金融数据安全治理智能化

银行业作为数据驱动型行业,在提供金融服务的过程中,积累的海量用户数据、交易数据以及外部数据已经成为银行业的重要资产。然而,银行数据特征复杂多样,缺乏标准,元数据也缺乏管理,同时由于缺乏整体的监管数据管控体系,存在数据定义不清晰、缺乏统一的指标口径、缺乏系统支撑、时效性不强、数据可信度低等问题,

极大增加了银行数据安全治理的难度^[20]。此外,大量从外部机构引入的数据给银行的数据安全治理工作带来额外的困难。

2022年,中国银保监会发布《关于银行业保险业数字化转型的指导意见》,明确提出“健全数据治理体系。完善数据治理制度,运用科技手段推动数据治理系统化、自动化和智能化。”银行通过数据安全治理可以推动全行业数据标准化、加强数据质量管理、防范风险发生、加速业务创新、提高精细化运营和更好满足监管要求等。目前,国内众

多银行逐步开始了智能化数据治理实践。

1) 中国工商银行实施数据治理数字化转型项目和“数字工行”(D-ICBC)等措施,成为国内首家获得DCMM5级认证的银行。强化智能化治理手段,将人工智能、自然语言处理和图像识别等技术应用于数据治理场景,以数治数,有效提升数据治理效率,实现金融数据安全治理体系的优化提升。

2) 中国银行建立了集团数据治理架构体系,并重点关注金融数据安全治理结构、数据标准管理和数据服务管理4个方面的治理工作。基于机器学习、生物识别、知识图谱、人工智能和计算机视觉等技术搭建了机器学习、生物识别、语音识别、机器人和知识库5大平台,建设了AI平台累计支撑617个场景。

5.2 证券业的金融数据安全治理智能化

2020年,中国证券业协会发布《关于推进证券行业数字化转型发展的研究报告》,提出要“逐步建立完善人工智能、区块链、云计算、大数据等数字技术在证券行业的应用标准和技术规范。增强数字化治理能力,促进业务融合发展。”证券交易数据可能对行情走势或者操盘有影响,如基金公司的持仓、仓位等。业务数据具有保密的需求,用户数据一旦泄露进入黑产,可能对企业造成名誉和业务上的双重损失^[21]。目前,证券机构在推进数据治理的过程中,仍然存在数据治理内驱力不足、缺少数据治理体系规划、数据安全防护不到位、数据问责体系不完善、数据治理难提升等众多问题,传统的数据治理方式已无法满足新时期的需求。

虽然国内证券行业中人工智能在数据安全治理领域正处于起步阶段,各种智能化手段的有效性在数据安全治理上尚未得到充分体现,但是也开启了一系列智能化活动,如国泰君安、中信证券等纷纷开始采用智能化的技术提高数据治理的手段,保证数据的安全以及高效利用。

1) 国泰君安近年来以“数据化、智能化、自动”为主线构建运维服务平台,自2018年就大力推进数据治理工作,开展以配置管理数据库(CMDB, configuration management database)为核心的IT数据治理和运维指标体系建设,建设大数据平台将数据治理手段融合到平台各个环境,充分运用人工智能、机器学习和知识图谱等技术实现证券数据的全

生命周期管理和异常发现能力,并将多方安全计算与区块链进行有机结合保护数据的安全。

2) 中信证券利用大数据技术,加快海量客户的指标和标签的计算能力,通过分布式数据库等相关技术提供海量计算和实时数据处理能力,在此基础上引入自然语言处理与人工智能相关技术,应用于客户服务留痕文本分类分析、客户意见挖掘、新闻与市场热点抓取、员工情感分析和智能语音交互等场景。

5.3 保险业的金融数据安全治理智能化

保险数据安全治理的需求包括防止外部窃取、防止数据破坏及篡改、防止数据滥用等,随着保险行业信息化建设的不断推进,加之保险业面临的监管环境日益严格和复杂,保险数据的治理创新还有待加强。当前保险业的数据治理还存在以下挑战:① 保险业的数据安全环境存在明显不足,我国对于保险数据安全管理方面的专门法律制度和管理规范方面还在发展阶段,保险机构的数据安全管理意识还有待提高;② 保险数据中大量的敏感信息为保险业数据安全治理带来困难,保险运营和管理与数据安全高度相关;③ 保险业务的复杂性使得数据安全治理体系的建设难度不断加大,数字化转型过程中的新产品、新业务、新模式也带来了新风险和新的安全问题;④ 与外部机构的合作越来越多,给保险业的数据安全治理带来了新的问题,包括外部网络环境不确定风险、外部数据源的保护、安全评估等。

随着保险业数字化转型进程的加快,保险机构的竞争力来源于充分发挥保险数据要素的效用。依托于人工智能等技术的智能治理模式必将在未来发挥出显著的引领作用,在保密性、准确性和计算效率之间实现动态平衡,逐步发挥隐私计算在主力数据要素市场化配置、防范数据泄露、促进多方数据安全合规协作等方面的重要作用,构建隐私计算应用的安全可信协作网络。

1) 中国人寿财险公司深入探索与人工智能、数据深度融合的智慧金融模式,搭建了一款基于隐私计算的数据智能应用平台,基于语音识别技术、自然语言理解技术的联盟轻量级服务聚合平台,基于微服务、大数据、人工智能的CLIMB数据分析、投资和管理平台,以及融合云计算、

边缘计算、5G 等技术的国寿物联网等。

2) 中国太平洋保险在科技战略规划 ITDP2.0 中, 将“数据安全能力建设”列为子项目, 目标是通过优化策略、深化推广既有数据安全平台, 为内部数据使用提供安全保障。研究和引入隐私保护计算、数据存储免加密改造等高阶数据安全防护技术, 不断强化全周期、全场景下的数据安全防护能力, 满足各项监管要求。

6 政策建议

6.1 加强金融数据安全治理智能化发展顶层设计和机制保障

制度建设方面, 一是建议中国人民银行金融科技委员会开展针对金融数据安全治理智能化的顶层设计和规划, 完善数据安全治理智能化发展基本规则体系, 包括监管目标、基础架构和技术应用等方面; 二是重视规划引领, 建立健全发展纲要、发展框架、发展目标、重点任务和保障措施。机制保障方面, 一是建议逐步出台数据安全治理方面的规范; 二是探索制定数据安全治理智能化管理办法, 为数据采集、报送和风险智能评估等提供制度保障, 做好新兴技术与传统治理手段的有效衔接和有机融合; 三是强化标准供给, 统筹金融数据安全治理智能化国家标准、行业标准、团体标准和企业标准的制定和发布, 充分发挥标准对金融数据安全治理的引领和指导作用。

6.2 加快关键技术研发及金融数据安全共享平台建设

面对金融科技风险复杂多变、产品日新月异等形势, 应加快金融数据安全治理和安全使用的理论与技术研究。建议监管部门、金融机构、科研院所等加大金融数据安全相关基础理论、共性关键技术、社会公益性技术研发投入, 形成一批国产自主可控、国际领先的创新型金融数据安全治理系统工具、金融数据库测试基准等。平台建设方面, 建议在充分论证基础上, 有序推动金融数据仿真平台和数据安全治理智能化平台建设, 构建包括事前预防管控、事中监测预警、事后处置总结在内的全流程监管模式, 形成对金融数据生成、安全使用、监管的全面、全流程的支持。

6.3 积极利用科技手段创新金融数据安全治理新体系

针对金融数据呈现出的新特征, 积极利用金融科技手段建立智能化治理体系, 提升数据治理的安全性、前瞻性、主动性, 筑牢风险防控底线。建议金融机构: 一是通过自建或共建大数据平台, 汇聚和融合不同行业的金融数据, 联合多方数据开展金融风险分析; 二是采用复杂网络、知识图谱等技术, 深度挖掘金融机构间关联关系, 完整呈现金融风险传导路径和演变趋势; 三是利用文本挖掘、语义分析等方法, 将海量非结构化数据转化为结构化数据, 为风险分析和预警提供数据支撑, 形成创新、健康的金融数据安全治理生态。

6.4 积极防范新技术在金融领域带来的新型风险

金融数据安全治理事关金融安全, 要牢固树立安全发展观, 做好智能化技术应用风险防范。建议金融机构强化算法监管与安全管控, 建立健全智能化手段应用的更新更正机制, 加强预期目标与实际结果之间的对比验证, 提升智能化方法的安全性、可信度与可靠性; 三是加快构建与金融数据的安全适应度、匹配度和成熟度相匹配的评估认证体系, 基于信息化手段强化技术的自身优化和新型技术风险规避。建议加强金融数据安全治理的伦理建设。研究制定符合国情、与国际接轨的伦理制度规则, 健全多方参与、协同共治的伦理体系。金融管理部门和监管机构履行监管科技伦理管理主体责任, 建立伦理审查、信息披露等常态化工作机制, 提前预防、主动化解金融数据应用、治理的伦理风险, 严防技术滥用。行业组织发挥自律作用, 出台相关自律公约和行动指南, 筑牢数据安全防范自律防线。

6.5 坚持自主创新并加强“政产学研”用协同创新

当前, 在金融、医疗、工业制造以及科学研究领域, 数据治理和安全使用仍然存在掣肘。在大国战略博弈背景下, 数据的安全治理智能化是深入实施创新驱动发展战略的内在要求, 也是我国各行业数字化转型过程中长期高质量发展的必然要求, 建议加快建立健全政产学研用协同创新研究机制, 发挥各自优势, 形成合力, 增强金融数据安全治理智能化的自主创新、协同创新能力。

6.6 积极完善金融数据安全治理相关配套促进措施

政策扶持方面,建议政府部门通过专项补贴、税收优惠、研发费用加计扣除等途径,促进资金、人才和技术向金融科技领域流动。政企合作方面,支持金融结构与成熟技术提供商开展合作。数据共享方面,建议由中国人民银行金融科技委员会统筹数据开放共享工作,制定数据共享、数据脱敏相关制度和实施规范,对监管数据进行分级分类管理,推动数据要素合理流动与高效利用。生态建设方面,建议发挥行业协会作用,构建一个涵盖监管部门、金融机构、技术研发企业、检测认证机构、投融资机构和学术界的平台,推动形成金融数据安全治理智能化创新生态。

6.7 深化人才培养与国际合作

人才培养方面,建议金融机构深化与高等院校合作,充分利用高校学术资源对学生开展跨专业教育,造就一批既通晓金融业务体系又熟悉数据保护治理技术的交叉复合型人才。同时,金融机构应重视技术型人才引进,并加大对人员的技术培训,持续保持员工对新兴技术、产品和业态的学习能力,使数据安全治理始终跟上创新步伐。国际合作方面,建议一方面通过“引进来”方式,吸收借鉴国外成熟的金融科技理念、方法和技术,助力我国金融科技发展和基础设施建设;另一方面坚持“走出去”战略,推动国内外金融数据安全治理理念、步调与标准的统一。

7 结束语

金融数据安全治理是一个集国家政策、法律法规、组织机制和技术要素相结合的系统工程。大数据、云计算、区块链和人工智能等技术为金融数据安全治理赋予了新的潜力,驱动金融数据安全治理智能化。可以预见,金融数据安全治理智能化是未来金融数据安全治理的重要转变,能够最大程度地挖掘金融数据价值,助力开展高质量、精准化、自动化、高效化的金融数据安全治理,实现金融机构的智能化数字化转型和行业的健康发展。

参考文献:

- [1] 张凯. 金融数据治理的突出困境与创新策略[J]. 西南金融, 2021, (9): 15-27.

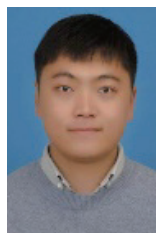
- ZHANG K. Prominent dilemma and innovation strategy of financial data governance[J]. Southwest Finance, 2021, (9): 15-27.
- [2] 鲁金彪, 胡国强, 周金赞, 等. 金融数据安全技术现状与趋势研究[J]. 中国金融电脑, 2022, (5): 39-43.
- LU J B, HU G Q, ZHOU J Z, et al. Research on the status and trends of financial data security technology [J]. Financial Computer of China, 2022, (5): 39-43.
- [3] 中关村网络安全与信息化产业联盟; 数据安全治理专业委员会. 数据安全治理白皮书 4.0[R]. China (Zhongguancun) Cyberspace Affairs Industry Alliance; Data Security Governance Committee. Data security governance white paper 4.0[R].
- [4] 周坤琳. 金融数据安全的治理困境与法律对策研究[D]. 重庆: 西南政法大学, 2021.
- ZHOU K L. Research on the governance dilemma and legal countermeasures of financial data security[D]. Chongqing: Southwest University of Political Science & Law, 2021.
- [5] 董小君, 宋玉茹. 加快推进我国金融数据治理现代化建设研[J]. 行政与法, 2022, (8): 11-21.
- DONG X J, SONG Y R. The research on accelerating the modernization of china's financial data governance[J]. Administration and Law, 2022, (8): 11-21.
- [6] 中国信通院. 人工智能数据安全白皮书(2019)[R]. China Academy of Information and Communications. Artificial Intelligence data security white paper (2019) [R].
- [7] 李雨霏. 人工智能在数据治理中的应用[J]. 信息通信技术与政策, 2019, (5): 23-27.
- LI Y F. Application of artificial intelligence in data governance [J]. Information and Communications Technology and Policy, 2019, (5): 23-27.
- [8] 胡志军, 王飞龙, 鲍时意. 区块链在金融数据治理的应用思考[J]. 金融科技时代, 2020, 28(12): 47-50.
- HU Z J, WANG F L, BAO S Y. Application of blockchain in financial data governance [J]. FinTech Time, 2020, 28(12): 47-50.
- [9] 闫树, 吕艾临. 隐私计算发展综述[J]. 信息通信技术与政策, 2021, 47(6): 1-11.
- YAN S, LYU A L. Overview of the development of privacy preserving computing[J]. Information and Communications Technology and Policy, 2021, 47(6): 1-11.
- [10] 张燕, 杨一帆, 伊人, 等. 隐私计算场景下数据质量治理探索与实践[J]. 大数据, 2022, 8(5): 55-73.
- ZHANG Y, YANG Y F, YI R, et al. Exploration and practice of data quality governance in privacy computing scenarios[J]. Big Data Research, 2022, 8(5): 55-73.
- [11] 单曙兵, 胡冰冰, 茅廷, 等. 金融数据分级标准的对比分析与实施建议[C]//第十八届中国标准化论坛. 广东, 2021.
- DAN S B, HU B B, MAO T, et al. Comparative analysis of financial data classification standards and implementation suggestions[C]//Proceedings of the 18th China Standardization Forum. Guangdong, 2021.
- [12] 王芳, 赵洪, 马嘉悦. 数据科学视角下数据溯源研究与实践进展[J]. 中国图书馆学报, 2019, 45(5): 79-100.
- WANG F, ZHAO H, MA J Y, et al. Research and practice progress of data provenance from the perspective data science[J]. Journal of Library Science in China, 2019, 45(5): 79-100.
- [13] 张学旺, 冯家琦, 殷梓杰, 等. 基于区块链的数据溯源可信查询方法[J]. 应用科学学报, 2021, 39(1): 42-54.
- ZHANG X W, FENG J Y, YIN Z J, et al. Trusted query method for data provenance based on blockchain[J]. Journal of Applied Sciences, 2021, 39(1): 42-54.
- [14] 刘武, 段海新, 洪张, 等. 基于信任的 P2P 网络安全管理体系结构研究[J]. 通信学报, 2008, (11A).
- LIU W, DUAN H X, HONG Z, et al. Study on trust based P2P network

- security management[J]. Journal on Communications, 2008, (11A).
- [15] 张佳发, 邹洪, 冯国聪, 等. 基于数据安全法范畴的数据访问控制分析方法[Z]. 2022.
- ZHANG J F, ZOU H, FENG G C, et al. Data access control analysis method based on the scope of data security law[Z]. 2022.
- [16] 王国赛, 李艺, 陈琨, 等. 隐私计算技术的金融应用思考[J]. 金融发展研究, 2022, (8): 31-37.
- WANG G S, LI Y, CHEN K, et al. Reflection about financial applications of privacy computing technology[J]. Journal of Financial Development Research, 2022, (8): 31-37.
- [17] 袁均良. 隐私计算在金融行业应用及其影响[J]. 金融科技时代, 2022, 30(10): 23-28.
- YUAN J L. The application and influence of privacy computing in financial industry[J]. FinTech Time, 2022, 30(10): 23-28.
- [18] ASSEFA S A, DERVOVIC D, MAHFOUZ M, et al. Generating synthetic data in finance: opportunities, challenges and pitfalls[C]//Proceedings of the First ACM International Conference on AI in Finance. 2020.
- [19] FAN J, LIU T, LI G, et al. Relational data synthesis using generative adversarial networks: a design space exploration[J]. arXiv preprint 2020, arXiv:2008.12763.
- [20] 宋首文. 商业银行数字化转型：数据治理体系的重构[J]. 国际金融, 2022, (11): 54-64.
- SONG S W. Digital transformation of commercial banks: reconfiguration of data governance system[J]. International Finance, 2022, (11): 54-64.
- [21] 安华金和. 证券行业数据安全治理实践与方案[R]. Beijing Dbsec Technology Co., LTD. Data security governance practices and solutions for the securities industry[R].

[作者简介]



侯鹏（1995-），男，云南玉溪人，复旦大学博士生，主要研究方向为金融科技、边缘计算。



李智鑫（1992-），男，云南昆明人，复旦大学博士生，主要研究方向为金融科技与安全，可信联邦学习。



张飞（1999-），男，山西吕梁人，复旦大学硕士生，主要研究方向为数据合成。



孙旭（1995-），女，吉林辽源人，复旦大学硕士生，主要研究方向为金融科技、知识图谱补全。



陈丹（1998-），女，广东东莞人，复旦大学硕士生，主要研究方向为金融科技、知识图谱。



崔毅浩（1997-），男，上海人，复旦大学硕士生，主要研究方向为金融科技、数据仿真。



张寒冰（1992-），男，黑龙江大庆人，博士，主要研究方向是金融科技与数据库。



荆一楠（1978-），男，江苏丹阳人，复旦大学副教授，主要研究方向为大数据分析、数据库系统。



柴洪峰（1957-），男，山西临汾人，中国工程院院士，复旦大学教授、博士生导师，主要研究方向为金融科技与安全。