分数据集中多数类的样本,而上采样平衡数据集的方式是向数据集中添加一些少数类的样本^[6]。主要包括随机下采样、随机上采样和SMOTE算法等实现方式。但采样技术存在一些局限,下采样会导致数据集中一部分数据减少,失去数据集全面性,而上采样会出现模型过拟合现象。

(2) 生成模型

主流的数据生成模型有变分自编码器和生成对抗网络^[7]。可以生成无限逼近真实样本的伪造样本,也不会造成数据样本丢失、类之间数据重叠等情况,提升生成样本的质量,避免了训练样本不足而导致的过拟合问题。

(3) 代价敏感

代价敏感不同于扩充少数类来达到数据平衡的目的,其更注重的是少数类的错分类,给予错分类更高的错分代价和相应的权重,提升分类器对少数类样本的关注度和检测率^[8]。

5 入侵检测算法

传统机器学习入侵检测方法^[9]主要有以下几种:逻辑回归,属于一种线性算法,选择合适的阈值获得计算事件概率并解决分类问题。朴素贝叶斯,该方法使用概率统计的知识对各个数据样本分类,基于贝叶斯定理,要求测量的特征之间是相互独立的,但是属性独立性条件一般很难满足,仅适用于较小的数据集;决策树,该方法采用树形结构,对不同的分支中的信息分层排序。树中每个非叶子节点代表对数据集中的某一个特征的测试条件,每个分支表示一个测试的结果,每个叶子节点表示一个决策结果,即类别;支持向量机,用于回归和分类,该方法不适用于大型数据集;人工神经网络,该网络由大量的神经元相互连接构成。包括输入层和输出层。

6 结束语

互联网的日益飞速发展,网络使用率不断增加,入侵活动也在增加,网络环境面临着越来越多的安全问题。入侵检测能够识别特定类型的攻击并作出防御反应来保障信息社会的安全。面对日益多样化、复杂化、智能化的网络攻击手段,以及入侵检测数据规模大,维度高,类别不平衡的现实状况,严重影响着传统的检测算法的性能,在未来的研究中,数据高维不平衡问题将一直存在,也是入侵检测中存在的一大难点。因此,如何高效地处理大规模的动态数据流的分类问题依旧需要继续探索。

参考文献:

- [1]陈仲磊,伊鹏,陈祥,等.基于系统调用的入侵检测技术研究[J].网络安全技术与应用,2022(03):1-6.
- [2]李波.基于机器学习的入侵检测防火墙设计与实现[D].电子科技大学,2021.
- [3]Jimmy Shum, Heidar A.Malki. Network intrusion detection system using neural networks[C]. in Proceedings of Fourth International Conference on Natural Computation, 2008: 242-246.
- [4]莫少慧,古兰拜尔·吐尔洪,买日旦·吾守尔.高维混合入侵检测数据的单类分类算法研究[J].计算机应用与软件,2022,

39(04):300-310.

- [5]李英深.基于PCA的特征子集选取及网络入侵检测算法研究[D].天津理工大学,2021.
- [6]刘景美,高源伯.自适应分箱特征选择的快速网络入侵检测系统[J].西安电子科技大学学报,2021,48(01):176-182.
- [7]胡希文,彭艳兵.基于ONE-ESVM的入侵检测系统[J].电子设计工程,2021,29(20):86-91.
- [8]崔鑫,徐华,朱亮.面向不平衡数据的多分类集成算法[J].计算机工程与应用:1-10[2021-11-12].http://kns.cnki.net/kcms/detail/11.2127.TP.20201216.1453.006.html.
- [9]Anderson J P. Computer security threat monitoring and surveillance[J]. Technical Report, James P. Anderson Company, 1980.
- [10]Denning D E. An intrusion-detection model[J]. IEEE Transactions on software engineering, 1987(2):222-232.
- [11]蹇诗婕,卢志刚,牡丹,等.网络入侵检测技术综述[J].信息安全学报,2020,5(04):96-122.
- [12]Krishnaveni S, Sivamohan S, Sridhar S S, et al. Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing[J]. Cluster Computing, 2021, 24(3):1761-1779.
- [13]Salo F, Nassif A B, Essex A. Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection[J]. Computer Networks, 2019(148):164-175.
- [14]Lv L, Wang W, Zhang Z, et al. A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine[J]. Knowledge-based systems, 2020(195):105648.
- [15]Zhang H, Huang L, Wu C Q, et al. An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset[J]. Computer Networks, 2020(177):107315.
- [16]Tan X, Su S, Huang Z, et al. Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm[J]. Sensors, 2019, 19(1):203.
- [17]Lee J H, Park K H. GAN-based imbalanced data intrusion detection system[J]. Personal and Ubiquitous Computing, 2021, 25(1):121-128.
- [18]Mulyanto M, Faisal M, Prakosa S W, et al. Effectiveness of focal loss for minority classification in network intrusion detection systems[J]. Symmetry, 2020, 13(1):4.
- [19]Verma A, Ranga V. Machine learning based intrusion detection systems for IoT applications[J]. Wireless Personal Communications, 2020, 111(4):2287-2310.

语言生物信息安全视域下语音数据的跨境流动治理

◆俞杰 陈美华

(东南大学外国语学院 江苏 211189)

摘要:在当前数字经济全球化的背景下,随着网络技术的高速发展,数据跨境流动现象愈发普遍。本研究关注承载语言生物信息的语音数据的跨境流动,梳理其核心要义及驱动因素,分析因数据跨境流动而产生的国家安全、个人隐私安全和商业技术安全挑战。从承载个体和群体语言生物信息、采集流动的便利性、语言信息的复杂性等维度阐明跨境语音数据流动治理的必要性。为进一步保障语言生物信息安全,践行总体国家安全观,提出语音数据跨境流动治理可通过健全监管制度,推动分级分类管理;强化国际合作,完善数据治理冲突调解机制;引导企业建立数据合规化工作规范,平衡本地化存储与合法跨境流动等安全策略,为国家生物信息安全政策制定、网络安全管理体系建设提供参照。

关键词:语言生物信息;语音数据;跨境流动;网络安全

基金项目:江苏省社会科学基金重点项目:中国语言生物信息安全体系建构研究(20YYA002)

2020年2月14日,习近平总书记在中央全面深化改革委员会第十二次会议上强调,要从保护人民健康、保障国家安全、维护国家长治久安的高度,把生物安全纳入国家安全体系,系统规划国家生物安全风险防控和治理体系建设,全面提高国家生物安全治理能力^[1]。生物安全是指国家为防范由生物因子、生物技术问题而引起的生物性危害,以保障人民的生命健康和财产安全。自2020年初全球爆发新冠疫情,人类更加重视生物安全的保障工作,在此背景下,将生物安全纳入国家安全策略具有重大的现实意义。

语言生物信息安全包含语言生物信息与信息安全两个层面。语言与指纹、人脸一样,属于生物信息的范畴。人类个体及群体语言可以表现出识别性强的特征信息,如个体声纹身份鉴定、群体方言识别鉴定等。在当今信息时代,语音声纹识别、指纹识别、人脸识别等生物信息识别技术已经在我国社会经济生活中大规模应用,然而指纹窃取、人脸数据及语音隐私泄露等问题逐渐显现,给生物信息安全提出了新的挑战。其中语音声纹识别作为最为普遍的生物信息识别技术之一,因智能终端成本降低、语音数据采集便捷和云服务器上传处理等因素,语音生物信息安全的重要性进一步凸显。

早在2000年,Spengler就指出,生物信息学是一个顺应时代潮流而产生的集生物、数学、计算机、工程学等于一身的学科,人类对其认识尚处在局限阶段^[2]。Laurie等专门论述了司法鉴定中运用生物信息的伦理问题^[3]。Kim等从司法、财经和伦理角度对医疗机构对病人的个人生物信息保护策略进行研究,并提出了相关信息安全的评估手法和工具^[4]。McCartney等提出了司法鉴定中使用生物信息的主要原则,并提出了作为证据的生物信息数据库的操作和管理规范^[5]。陈宗波提出,随着我国网络对公民个人数字化生物信息利用频率的迅猛发展,生物信息作为国家层面的战略资源,呈现出越来越高的使用价值,但生物信息滥用的问题日益突出,国家生物信息安全保护法的进程必须加快^[6]。郑方等系统综述当前语音声纹识别中的技术问题,以及可能出现的例如“假冒闯入”等隐私泄露隐患,对语音声纹识别技术的应用前景进行了展望^[7]。

综合以上分析,现有研究集中于生物信息应用的法律和伦理问题,关注生物信息存储及使用的法律规制,研究对象以人脸识别、指纹识别等常规生物信息为主,但针对语言生物信息的研究则不多见。目前较少研究基于语言生物信息安全的视角,进一步探讨语言大数据,特别是承载个体、群体语言生物特征等重要信息的语音数据的存储、使用和加工。因此,在当前数字经济全球化的背景下,随着网络技术的高速发展,数据跨境流动现象愈发普遍。本研究重点关注承载语言生物信息的语音数据的跨境流动,梳理其背后动因及所带来的安全挑战,阐明跨境语音数据治理的必要性,并为践行总体国家安全观,进一步保障语言生物信息安全,针对性地提出语音数据跨境流动治理的安全策略,进而为国家生物信息安全政策制定、网络安全管理体系建设提供有益参照。

1 跨境数据流动

1.1 概念厘定及驱动因素

随着数字经济与网络技术的迅速发展,全球数据总量呈现指数化增长,与此同时大量的数据基于互联网等多种传输介质,开始脱离国界束缚,并逐步在全球范围内流动和交互。跨境数据流动主要涉及“数据”与“跨境流动”两方面的核心要素。

一是数据具有能被识别、可读取特征,同时包含特定的信息^[8]。从数据类别角度,基于所承载内容信息的差异,跨境数据流动既包含例如个人生物信息等个人数据,还包括例如商业数据、技术数据等非个人数据。尽管目前跨境数据流动主要以个人数据为主,且非个人数据经匿名化处理后,敏感度较低,但是依旧涉及部分关键信息基础设施上所产生的公共数据及政府数据,因此跨境数据流动治理的重要性不仅包含对个人隐私的保护,同时已经上升到了经济乃至国家安全层面。二是数据流动出现了跨境现象。狭义的跨境是指数据在国家边境或边界内外的机构之间进行传输,以跨境电商为例,消费者在网络购物过程中,通常会涉及例如购物偏好、商品检索信息、物流方式选择、个人支付信息、税务信息等一系列敏感数据的境外传输。广义的跨境则可以从数据在不同法域之间的流动来理解,即数据跨越司法管辖域或条约规制域的传输,例如数据在欧盟国家的传输流动尽管跨越物理国境,但仍然处于欧盟数据保护条例的管辖范围。此外,“流动”不仅指数据的单向或多向传输,也包含域外对本国数据的储存、使用和加工。因此,结合以上特征,跨境数据流动可以理解为以电子或其他可被识别方式记录的特定信息跨越司法管辖域或条约规制域的读取、储存和处理活动^[9]。

跨境数据流动的驱动因素可分为客观因素与主观因素。一方面数据天然地具有流动性,数据的价值通过流动得以实现^[10],这是跨境数据流动最主要的客观因素。在当前经济贸易全球化趋势不可逆的背景下,无论是与数据直接相关联的互联网产业的蓬勃发展,还是知识与信息的迅速传播,都依赖于数据的自由流动,而经济贸易的全球化则进一步凸显了数据跨境流动的必要性 and 现实性。此外,数据的存储、使用和加工在很大程度上依赖于例如云计算等高端数字技术和数据存储基础设施,技术的迭代更新弱化了数据存储地理位置的约束^[15]。因此通过数据服务外包以获得先进技术支持,租用境外大型数据基础设施以降低商业成本得以实现,这些举措都带动了数据的跨境传输。另一方面,数据的价值基础性所带来的经济效益是跨境流动的主观因素。数据在全球层面的交互、共享可以最大化实现数据的资源要素价值,有助于降低成本、增加总产出。同时数据作为数字经济时代的“石油”,具有极大的商业价值^[11],相关市场主体可以将数据直接转化为企业的经营性资产。以TikTok(抖音海外版)为例,在用户使用软件的过程中,其日常行为模式和浏览偏好可以为用户画像、精准广告投放、商业营销提供极具价值的参考信息。

1.2 跨境数据流动的安全挑战

在上述驱动因素的影响下,跨境数据流动为经济产业发展带来新的增长和机会,数据传输规模不断扩大,有效促进了企业的全球运营与商业拓展,提升了各国的经济效率和社会福祉^[12]。但是数据跨境流动关乎经济利益与风险控制之间的动态平衡,不可避免地带来对国家、个人隐私安全、商业技术安全等方面的风险关切。

(1) 国家安全挑战

自2013年“棱镜门”事件披露美国国家安全部门通过入侵他国网络系统,非法监视他国公民通信数据、获取他国敏感部门数据并进行跨境转移,数据的跨境流动带来了对国家安全的挑战。近年来,不仅个人信息泄露频发,中国国家互联网中心发现我国境内上万个网站内容遭到篡改、网站被植入后门、网站所存储用户信息传输境外,对中国国家安全与社会稳定造成极大风险隐患。除此之外,部分基于商业目的的跨国企业数据跨境传输同样因影响国家安全而引发关注。例如华为、中兴等中国知名企业在海外推广布局新一代5G信息传输技术时,受到以美国、加拿大为代表的西方国家严格审查,并最终威胁国家安全为由,禁止他国企业参与本国5G网络建设。一些社交软件例如TikTok和Wechat也同样受到美国、印度等国家的禁止使用,理由是这些软件可能会将国家(联邦)工作人员的敏感信息在未经允许的情况下,跨境传输至中国境内。

总体而言,数据跨境传输所带来的国家安全挑战主要有两个方面。一是因发展中国家和发达国家在经济、科技水平上不对等,在现如今数字经济的时代背景下,数据跨境传输可能会导致数据的“单向”传输现象,即从发展中国家不断积聚到部分发达国家,由此导致“数据霸权”的产生,加剧经济发展的失衡状态^[13]。另一方面,以5G为代表的先进信息技术、部分社交软件中敏感数据的跨境传输,使得基于大数据的群体行为模式分析得以实现,如果对数据类型和内容不加以限制,数据跨境流动必然导致部分国家的重要国情信息被大数据手段识别,成为国家安全的潜在隐患。

(2) 个人隐私安全挑战

跨境数据流动的监管治理不可避免地带来对个人隐私安全的重要关切,数据跨境流动与个人信息保护相伴而生。以搜索引擎为例,如使用国外互联网公司提供的搜索引擎,那么用户IP地址所包含的地理位置及身份信息、搜索记录、使用偏好和广告点击情况都会成为数据收集来源。同样全球日益普及的跨境电商交易中,用户所购买的商品信息、收货地址信息、个人取件联系方式会以数字化的形式为国外服务提供商所获取。如果境外服务提供商未能按照法律监管要求采取充分的个人数据保护措施,那么个人隐私就会存在被滥用的可能性。由此产生的信息泄露将会带来跨境网络诈骗、网络钓鱼等个人隐私安全风险。

随着以AI人脸识别技术、VR虚拟现实技术为典型代表的信息技术的发展,个人隐私数据也不再局限于传统的个人身份信息。以VR虚拟现实技术为例,为了带来更好的沉浸式体验,VR产品会对用户的个人生物信息进行记录和识别,除了用户在最初注册时主动提供的信息外,部分VR产品在用户与系统的交互中会收集如生理状况、行为习惯以及性格特征等隐私信息^[14]。这些隐私信息一旦跨境泄露或非法滥用,极易对人身财产、个人名誉、身心健康造成威胁。同时,数据在跨境传输之后会受到数据所在地条约规制域的多重限制,后期境外电子数据的司法取证往往会出现单方跨境取证难、耗费时间成本高、

法律条款限制多等问题^[15]。

(3) 商业技术安全挑战

在商业领域,人工智能、大数据、云计算等高科技改变了当今以互联网为代表的高新技术企业的竞争态势。从今日头条 AI 算法实现个性化内容推荐、谷歌 Alpha Go 击败职业棋手,到先进工业制造、智慧零售、智慧城市的出现,这些技术的实现都依赖于大数据^[16]。数据在商业技术中显现出越来越重要的作用,与之相关的数据获取、利用也呈现出竞争态势。跨境数据流动给商业技术安全所带来的挑战主要是基于大数据所形成的商业产品在复制、传播过程中的产权保护问题。尽管针对商业产品已有较为完善的知识产权保护机制,但是产品研发设计所基于的大数据集本身却仍旧缺少明确的法律规制^{[13][9]},由此而产生数据盗用、网络爬虫、企业间不正当数据竞争等问题。尽管跨境数据流动有效促进了商业产品的落地与推广,但是在数据获取、利用环节伴随产生的不正当数据资源竞争,无形中给商业技术安全带来了挑战。

2 跨境语音数据流动治理的必要性

语音是一种生物行为特征,语音数据具体指人通过发音器官所发出的、带有一定实际含义的声音特征信息,既包含说话人语音的个性信息和发音的内容信息,也有声音周围的环境噪声信息、信道信息等^{[7][45]},例如视频中的人声、使用语音输入法时上传至服务器的语音片段。现有研究多关注广义的数据跨境流动,但较少研究基于语言生物信息安全的视角,进一步探讨语言数据,特别是承载个体、群体语言生物特征等重要信息的语音数据的跨境流动治理,并提出针对性地提出语音数据跨境流动安全策略。保障国家安全、维护个人隐私信息权益及商业技术安全、促进数字经济高质量发展离不开健全的跨境数据流动治理体系,跨境语音数据流动治理的必要性主要有以下三个方面。

(1) 语音数据所承载的个体语言生物信息是个人信息安全的重要组成部分

以声纹为代表的个体语言生物信息主要指可将语言识别为个体生物信息的语言类特征,具体提取参数包括频谱、倒频谱、共振峰、反射系数等与个人发声部位解剖学机械结构相关的参数,这些特征参数是实现个体声纹识别、说话者身份判定的重要组成部分。

在语音数据中,个体声纹信息具有特定性和稳定性特征,每个人在说话过程中所蕴含的语音特征和发音习惯几乎是独一无二的^[17],因此语音数据属于个人敏感隐私信息,是个人信息安全的重要组成部分^[18]。在实际应用场景中,早在17世纪语音中所包含的声纹信息就已经被用来作为身份司法认证的手段,随着信息技术的迅速发展,个体声纹识别技术更是在银行金融(如信用卡持卡人身份信息识别)、刑侦破案(如电话勒索案件中基于声纹辨别技术锁定嫌疑人)、网络支付(如在支付交易中加入声纹确认以避免网银盗刷)、声纹锁控(如微信声纹动态口令、智能声控门锁)等领域广泛应用。但正是应用场景的多样化,无形之中因语音数据传输而造成个体语言生物信息泄露的风险急剧上升,以跨国电信诈骗为例,犯罪分子一般集团化作案,借助高科技手段,利用非法途径所获取的声纹信息,从最原始的声音模仿,到计算机语音合成、声音转换、录音重放手段,实现语音假冒闯入,对受害者的亲属、朋友实施电信诈骗,甚至直接破解受害者的声纹锁支付认证。

(2) 语音数据所承载的群体语言生物信息是构成国家安全要素的重要基石

群体语言生物信息主要是指地域或社会阶层所能反映出的语言类特征,具体提取特征包括地域方言特征、国家语种特征、受教育和经济水平影响所使用不同词汇特征、社会知识背景特征等,这些特征是群体识别的重要元素。

语音数据不同于文字信息,是社会群体日常生活中语言使用最直接的体现。通过大规模群体语音数据收集、建库,再基于语料库分析系统对方言语音特征、声纹特征、发音习惯进行识别,可甄别出关键群体的身份信息。从语言表达中甚至可以推断出社会群体的语言使用态度,预判其情感走向和行动方向^[19],针对性制定舆论引导方案,可为国家安全建构过程中的情报搜集和研判提供语言佐证。以我国澳门言语社区为例,基于大规模语音数据,并从语言类别(如英语、葡语、粤方言、普通话)、粤方言口语变体(如懒音现象)^[20]、英语使用场景等多维度分析,可对澳门言语社区中的语言使用、语言态度进行清晰“画像”,进而可对社会群体的内在身份建构作出推断。因此,语音数据不仅是个人信息安全的重要组成部分,同时大规模的语音数据所承载的群体语言生物信息更是构成国家安全要素的重要基石,减少重要群体语音数据的大规模跨境流动对于保障国家安全具有重要作用。

(3) 语音数据采集流动的便利性及语言信息的复杂性导致识别难、跨境监管难

语音数据一般获取方便,数据采集装置成本低廉,除专门配备的外置麦克风外,在硬件方面可直接使用智能手机、智能手表等日常通讯设备进行录音;在软件方面,例如微信、QQ、imessage、各种语音输入法的语音识别功能都可以便捷提取。此外,相较于文字数据,语音数据往往包含更为丰富的语言信息,例如语气、情感等,相比文字,人们在输入语音信息时,不比输入文字信息谨慎,更容易在不经意间泄露重要信息。以语音情感识别为例,在跨境电商、在线教学、医疗卫生等领域,语音大数据中所包含的情感信息,往往会揭示个人的心理感受,服务提供商在获取语音数据后通过建库分析能够获取潜在的身份信息,进而诱发诸多风险。此外,在监管方面,一般语音数据的识别是通过将语音转录成文字,再通过文字关键词对敏感内容实现监管,但是语音识别过程中往往会因为非标准普通话、方言语音变体,导致识别率较低。

3 跨境语音数据流动治理的安全策略

(1) 健全语音数据跨境流动监管制度,推动跨境数据分级分类管理

目前世界上已经有70多个国家和组织针对数据保护制定了专门规章制度^[21],例如欧盟制定了《通用数据保护条例》,俄罗斯制定了《个人数据保护法》,美国通过了《隐私法》《数据安全法案》等。我国近年来已密集出台数据和信息安全相关的法律法规(见表1),《数据出境安全评估办法》也已于2022年9月1日起正式施行,有关跨境数据传输的规定不断充实完善。

表1 数据和网络安全相关法律法规

法律法规	发布或正式施行时间
《网络安全法》	2017年6月1日施行
《数据安全法》	2021年9月1日施行
《关键信息基础设施安全保护条例》	2021年9月1日施行
《个人信息保护法》	2021年11月1日施行
《数据出境安全评估办法》(征求意见稿)	2021年10月29日发布
《网络数据安全条例》(征求意见稿)	2021年11月14日发布

但是,现有规定没有对跨境流动数据作出具体的分类,在法律条层面多为概括性的指导,未配备例如“个人数据”和“重要数据”的界定、跨境传输评估机制和评估方法等明确细化的实施细则,在法律实际操作性上还有待加强。因此,厘清各法律规范中有关跨境数据流动条目的相互关系和适用范围,有必要成立专门的数据监管机构,在对数据跨境流动实施监管与安全评估过程中,就例如语音数据等跨境数据类型采取明确适当的分级分类管理模式。同时在跨境数据分级分类的基础上,根据保密程度与个人数据敏感程度不同,进一步细化评估标准、评估规范和评估程序,既达到跨境数据分级分类的初衷,同时也减少可能的额外行政监管负担。基于跨境数据的分级分类管理有利于维护国家安全、个人信息安全、商业技术安全的目标实现,同时也能够有效平衡解决经济效率和数据主权问题,对于涉密数据、政府数据、特种行业数据进行重点行政监管,对于一般数据则可灵活采用自由流动、有条件出境等管理措施,兼顾有关商业服务数据的跨境流动、对于个人数据的保护和维护国家数据主权的规制目标。

(2) 强化跨境语音数据流动治理国际合作,完善数据治理冲突调解机制

除了健全语音数据跨境流动监管制度,推动跨境数据分级分类管理,在国际合作层面,强化与现有跨境数据流动治理国际规则的配套衔接,推动完善数据治理冲突调解机制同样至关重要。

目前在国际合作层面,我国还未与他国或区域组织实现有效的双边或多边合作^{[11][9]},呈现出治理参与度低、数据话语权弱、影响力不足的问题,这与我国作为数据大国的身份存在不匹配现象。以华为、中兴拓展海外5G市场,却受到以美国为首的多个国家以所谓威胁国家安全为由遭到封禁为例,这其中固然有大国政治博弈的因素,但是由此也体现出我国在数据治理国际合作领域中处于被动地位,国内企业数字化业务的拓展仍面临严峻挑战,有必要在国家层面采取调解协助措施,反制外国政府的无理干预。因此,在语音数据跨境流动治理实践中,一方面我国应积极参与双边或多边对话,有效发挥我国在“一带一路”沿线国家的影响力,通过高水平双边或多边协议,明确与第三国管辖权或规则冲突的解决方案,形成语音数据跨境治理的双边或多边协作机制^[22]。另一方面,应积极参与跨境语音数据流动治理国际规则制定,通过对现有法律条款的合理解释和运用,突破我国在欧洲和亚太地区的数据流动壁垒,争取在对话过程中获得更多国家对跨境

语音数据流动治理中国方案的支持,提升中国方案影响力。

(3) 引导企业建立数据合规化工作规范,平衡数据本地化存储与合法跨境流动

目前我国数据跨境流动采取的是“数据本地化原则+安全评估”治理模式^[23]。在《网络安全法》第三十七条中有明确的数据本地化表述,即“关键信息基础设施的运营者在中华人民共和国境内运营中收集和产生的个人信息和重要数据应当在境内存储。因业务需要,确需向境外提供的,应当按照国家网信部门会同国务院有关部门制定的办法进行安全评估”。但是该条规定仅是原则性表述,数据跨境合规程序具体如何施行较为模糊,在企业层面仍面临着合规压力。

因此可以积极引导企业建立相关数据合规部门,主要负责企业跨境数据合规事宜,自觉实现语音数据跨境合规。以美国苹果公司为例,iCloud 数据中心落地中国贵州的关键原因就是在中国市场合规运营的需要,例如语音数据等中国用户信息会完全在中国境内存储。作为拥有上亿用户的关键信息基础设施的运营者,苹果公司为推动数据流动合规化工作起到了良好的示范性。此外,在国家专门数据监管机构的统筹下,可以引入第三方认证评估机构,对企业数据获取、加工、流动等环节进行考核评级,探索建立“白名单”制度,平衡语音数据本地化存储与合法跨境流动,进而提升数据流动审查的工作实效性。

4 结语

随着数字经济与网络技术的迅速发展,数据的跨境流动呈指数化增长。基于语言生物信息安全的重要性,语音数据承载着以声纹为代表的个体语言生物信息,以及地域方言特征、国家语种特征、受教育和经济水平影响所使用不同词汇特征、社会知识背景特征等群体语言生物信息,有必要加强语音数据跨境流动治理。

为进一步保障语言生物信息安全,践行总体国家安全观,跨境语音数据治理应健全监管制度,推动数据分级分类管理,细化评估标准、评估规范和评估程序。在国家层面应积极参与双边或多边对话,强化跨境语音数据流动治理国际合作,完善数据治理冲突调解机制,争取在对话过程中获得更多国家对跨境数据流动治理中国方案的支持,提升中国方案影响力。同时引导企业建立数据合规化工作规范,平衡语音数据本地化存储与合法跨境流动,以一系列安全策略的有效施行,在语音数据跨境流动方面推动中国语言生物信息安全、网络安全管理体系建设。

参考文献:

[1] 习近平主持召开中央全面深化改革委员会第十二次会议强调 完善重大疫情防控体制机制 健全国家公共卫生应急管理体系[J].中国纪检监察, 2020(4): 2.

[2] Spengler, S. J. Bioinformatics in the information age[J]. Science, 2000(287): 1221-1223.

[3] Laurie, G., Hepple, B., Cooke, G. Forensic Use of Bioinformation: Ethical Issues[M]. Nuffield Council on Bioethics. 2007.

[4] Kim, S. S., Lee, Y. H., Kim, et al. Privacy protection

for personal health device communication and healthcare building applications[J]. Journal of Applied Mathematics, 2014: 1-5.

[5] McCartney, C., Williams, R., Wilson, T. The Future of Forensic Bioinformation[M]. Nuffield Foundation, 2010.

[6] 陈宗波.我国生物信息安全法律规制[J].社会科学家, 2019(1): 96-103.

[7] 郑方,李蓝天,张慧,等.声纹识别技术及其应用现状[J].信息安全研究, 2016, 2(1): 44-57.

[8] 孙方江.跨境数据流动:数字经济下的全球博弈与中国选择[J].西南金融, 2021(1): 3-13.

[9] 蔡翠红,郭威.中美跨境数据流动政策比较分析[J].太平洋学报, 2022, 30(3): 28-40.

[10] 吴沈括.数据跨境流动与数据主权研究[J].新疆师范大学学报(哲学社会科学版), 2016, 37(5): 112-119.

[11] 张奕欣,王一楠,陈继鑫,等.从数据跨境流动的域外规制看中国对策[J].重庆邮电大学学报(社会科学版), 2022, 34(2): 51-62.

[12] 惠志斌,张衡.面向数据经济的跨境数据流动管理研究[J].社会科学, 2016(8): 13-22.

[13] 马其家,李晓楠.论我国数据跨境流动监管规则的构建[J].法治研究, 2021(1): 91-101.

[14] 张欢,文铭,汪友海.从 Facebook 数据滥用事件谈个人信息法律保护[J].重庆邮电大学学报(社会科学版), 2018, 30(6): 56-63.

[15] 卜学民.论数据本地化模式的反思与制度构建[J].情报理论与实践, 2021, 44(12): 80-87+79.

[16] 田小军,曹建峰,朱开鑫.企业间数据竞争规则研究[J].竞争政策研究, 2019(4): 5-19.

[17] 刘晓晨,潘孝勤,曹金璇,等.声纹识别和语音识别技术在公安领域的应用[J].网络安全技术与应用, 2021(4): 153-155.

[18] 宋杰.声纹识别对抗攻击对公安领域的影响及对策[J].网络安全技术与应用, 2022(4): 152-154.

[19] 郭继荣,杨亮.国内语言安全研究述评[J].情报杂志, 2021, 40(6): 95-101+58.

[20] 阎喜.言语社区语言安全研究——以澳门为例[J].中国语言战略, 2020, 7(2): 98-104.

[21] 胡炜.跨境数据流动的国际法挑战及中国应对[J].社会科学家, 2017(11): 107-112.

[22] 陈兵,杨鏊林.因应数字平台集团数据跨境流动安全管理挑战[J].统一战线学研究, 2022, 6(3): 91-107.

[23] Liu, J. China's data localization[J]. Chinese Journal of Communication, 2020, 13(1): 84-103.

大数据背景下计算机网络安全防范研究

◆陈阳

(苏州大学计算机科学与技术学院 江苏 215006)

摘要:大数据时代,计算机信息技术飞速发展,海量信息交互密集,人们对计算机网络的发展和数据安全开始有了新的认知和要求,如何安全上网和保障网络数据安全逐渐成为社会关注的重点话题。在此情况下,加大计算机网络安全防范措施和提高网络安全的管理力度是发展信息技术、促进社会和谐发展的必然措施。基于此,本文结合大数据时代背景,重点分析了大数据时代背景下计算机网络安全方面存在的问题,并针对性研究探讨了大数据时代背景下计算机网络安全防范的有效策略,以有效净化网络环境,提高计算机网络安全等级,最终保障计算机网络用户的信息安全,促进社会的和谐发展。

关键词:大数据;计算机;网络安全;安全防范;防范策略

大数据时代背景推动着计算机网络和信息技术实现质的飞跃,但同时大数据的发展也加剧了计算机网络的复杂性和不可控性,给一

些黑客和违法者可乘之机,导致计算机用户在使用过程中面临着严峻的安全问题和网络安全风险,如信息泄露风险、网络诈骗风险等,同