

数据向善 联合无碍

腾讯隐私计算白皮书 [2021]

参与人员

顾问

司晓 蒋杰 徐炎 江波 刘煜宏

腾讯研究院

李南 翟尤 王京婕 管洪博 秦天雄
宋扬 赵子飞 王融 王天元

腾讯数据平台部

程勇 郝海琪 陶阳宇 陈鹏

腾讯安全

刘站奇 王海波 姜军军

腾讯云区块链

李力 邵兵 刘江 李佳 苏庆慧 敖萌 蒋昊

腾讯数据隐私保护部

张亚男 王小夏 黄晓林

腾讯安全平台部

张博 何林书 胡珀

腾讯广告

江毅 邹正勇 郭俊 徐威

| 序言

卢山

腾讯高级执行副总裁



伴随着数字技术的创新应用，人类社会进入全新数字时代。数据的融合应用驱动各行各业走向数字化、网络化和智能化，正在深刻地改变人类的生产和生活方式。

与此同时，数字社会的治理成为我们需要面对的重要议题，数据安全、个人隐私保护等问题受到社会广泛关注，也是我们要心怀敬畏审慎对待的课题。隐私计算的兴起，为人们提供了在数据安全合规、融合应用过程中寻求发展和安全之间平衡点的技术路径和解决思路，其正成为未来数字治理的有效路径之一。但是，数字治理的探索是一项系统性工程，仅仅依靠单一技术无法满足当前和未来复杂的治理需求，需要技术、法律等多种手段相结合的综合治理，才能更好适应数字社会发展需要。

腾讯作为一家科技公司，以“用户为本、科技向善”为愿景使命，对技术迭代、应用创新充满敬畏之心，始终相信守正创新方能行稳致远，希望和业界携手探索新技术、新应用、新模式创新和治理，积极探索数字世界，守护数据安全，助力人们生产效率和生活品质的持续提升。

编写说明

隐私计算作为在数据融合应用过程中保障数据安全合规的关键技术路径,其商业模式、应用场景、技术变革、产业趋势、法律问题等正成为当前政、产、学、研、用等各界关注的热点。在此背景下,腾讯多个部门联合撰写《腾讯隐私计算白皮书 2021》,旨在与业界共同探讨、推动隐私计算技术产业的发展,寻求在数字治理中发展和安全的平衡点。

白皮书主要分为五个部分。第一部分阐述了隐私计算的发展背景、基本概念和主要作用。第二部分主要分析了隐私计算的技术体系,重点对联邦学习、可信计算、安全多方计算以及区块链和隐私计算融合发展进行了探讨。第三部分主要描述了隐私计算当前应用的重点行业和场景。第四部分重点探讨了在法律视角下隐私计算在数据安全合规方面的作用和痛点。第五部分重点从技术、应用、法律等视角对隐私计算的发展进行了展望。

隐私计算的发展总体还处于起步阶段,并随着产学研用各界的研究,以及政策环境、用户需求等变化加速演进。当前我们对隐私计算认识也处于探索阶段,未来将根据腾讯及合作伙伴的实践以及来自各界的反馈意见,在持续深入研究的基础上适时修订。

目录

参与人员	02
序言	03
编写说明	04
一、隐私计算成为释放数据融合价值的助推器	06
(一) 数据融合应用需求迫切, 兼顾发展与安全合规成为行业命题	06
(二) 隐私计算应运而生, 成为数据协作过程中保护多方数据权益的技术解	08
二、隐私计算三大流派交织演进, 和区块链融合成为主流方向	10
(一) 联邦学习助力实现多方联合机器学习	10
(二) 安全多方计算提供更加安全的联合数据分析能力	12
(三) 可信计算助力隐私计算服务安全高效运行	13
(四) 隐私计算融合区块链提升数据协作全流程保护能力	16
三、数据协作需求推动隐私计算应用从金融、医疗等向其他行业延伸	18
(一) 金融	18
(二) 医疗	19
(三) 政务	20
(四) 广告	21
四、隐私计算助力数据安全合规的价值凸显, 但仍存在较大提升空间	22
(一) 隐私计算有望成为数据协作过程中数据合规和隐私保护的技术工具	22
(二) 隐私计算的推广应用仍存在合规痛点	24
五、技术演进、应用拓展和法律完善将加速隐私计算商业化进程	26
(一) 效率、性能提升和技术融合将成为隐私计算产品化的主要方向	26
(二) 隐私计算应用场景不断拓展, 有望重塑数据使用模式	26
(三) 隐私计算将通过助力法规政策落地促进数据融合应用	27
参考文献	28

一、隐私计算成为释放 数据融合价值的助推器

(一) 数据融合应用需求迫切, 兼顾发展与安全合规成为行业命题

数据驱动数字经济蓬勃发展, 数据安全合规成为焦点议题。伴随着云计算、大数据、人工智能等新一代信息技术的落地应用, 数据作为战略性和基础性资源, 不但是连接虚拟空间和实体空间的纽带, 也是数字经济体系中技术创新、需求挖掘、效率提升的重要动能。但数据在不断创造价值的同时, 其安全保护、合规应用等问题也成为政、产、学、研、用等各界关注的焦点。**一是**数据发挥价值需要融合应用。数据跨层级、跨地域、跨系统、跨部门、跨业务的融合应用才能推动新模式、新应用、新业态的不断涌现, 加速数字经济创新发展。**二是**数据可复制、可传输等特性期待多元创新的安全合规手段。数据的应用会涉及政府、社会、企业、个人等多方主体权益, 关系到国家安全、经济运行、社会治理、个人权益等多主体, 需要创新安全管理模式。**三是**数据的价值发挥和安全合规需要寻求动态平衡点。数据治理体系搭建需要兼顾发展和安全的平衡, 既要保护数据主体的权益, 也要实现公共利益和社会福利的最大化。

多方主体的数据协作成为趋势, 数据安全合规风险亟需消除。当前, 全球数据总量呈现指数性增长态势, 但从现阶段数据的从属来看, 海量数据散落于不同的组织机构和信息系统中, 即使是同一区域、产业和企业, 也仍存在“数据孤岛”问题。多方的数据协作已经成为医疗、工业、零售、金融、政务等领域挖掘数据价值的重要路径, 聚合态体系中的多方数据进行联合建模分析也是当下放大数据价值的必然选择, 但数据安全和合规仍是多方主体数据协作过程中的痛点问题。**一方面**缺乏能够兼顾安全合规和数据协作的合作机制与技术路径, 无法消除数据主体之间对商业秘密泄露风险、商业利益分配等方面的信任鸿沟, 传统的数据保护方案往往适用于单一的信息系统或者有可能降低数据可用性, 导致无法满足现有的智慧医疗、智慧金融、数字政府等涉及跨系统的业务形态。**另一方面**黑灰产、隐私保护等问题也为不同主体的数据协作带来挑战。由于黑灰产的存在, 不但加大了企业的数据保护成本, 也扩大了数据泄露的风险。此外, 由于企业的数据也会包含用户个人信息, 在协作过程中如何有效进行个人信息保护也是数据价值挖掘的难点。

数据法律体系日益完善,推动企业加速构建数据应用安全合规体系。

近年来,数据保护成为全球关注的焦点,**一方面,数据保护法律法规体系逐渐清晰完善。**领先国家纷纷出台数据保护相关法律,欧盟早在2018年5月就出台《一般数据保护条例》,旨在加强对欧盟境内居民的个人数据和隐私保护。2020年,我国的《民法典》、《出口管制法》、《数据安全法(草案)》、《个人信息保护法(草案)》出台或公布,不断填补我国数据安全方面的空白,韩国在2020年1月份通过了新修订的《个人信息保护法》、《信用信息法》、《信息通信网法》三部法律,随后又对《个人信息保护法执行令》的相关内容也进行了修订。美国继2020年1月《加州消费者隐私法案》生效后,2021年,弗吉尼亚州州长拉尔夫签署了《消费者数据保护法》,这使得弗吉尼亚州成为美国第二个制定全面隐私立法的州。**另一方面,各国关于数据保护的监管执行日趋严格。**根据跨国律师事务所DLA Piper公布的《通用数据保护条例》(GDPR) 罚款和数据违规报告,2018年5月25日GDPR实施后,数据保护当局已经执行了2.725亿欧元的罚款,涉及欧盟27个成员国以及英国、挪威、冰岛和列支敦士登。其中,2020年1月28日以来执行的罚款数额为1.585亿欧元。

(二) 隐私计算应运而生, 成为数据协作过程中保护多方数据权益的技术解

1、隐私计算基本概念和现状

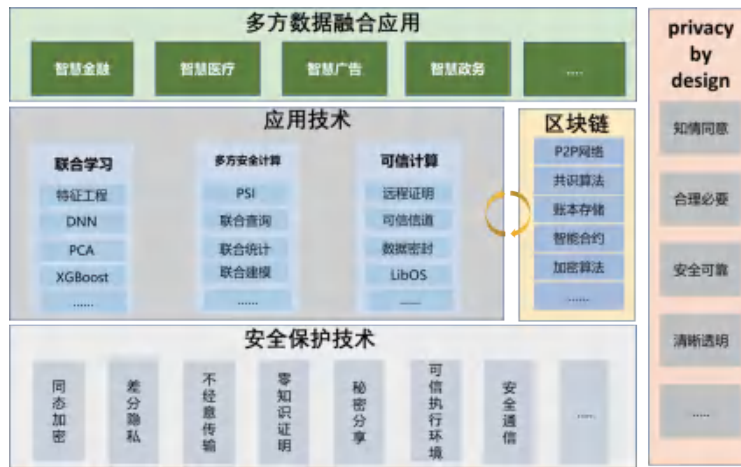


图1 隐私计算体系架构

隐私计算 (Privacy Computing) 是一种由两个或多个参与方联合计算的技术和系统, 参与方在不泄露各自数据的前提下通过协作对他们的数据进行联合机器学习和联合分析。隐私计算的参与方既可以是同一机构的不同部门, 也可以是不同的机构。在隐私计算框架下, 参与方的数据明文不出本地, 在保护数据安全的同时实现多源数据跨域合作, 可以破解数据保护与融合应用难题。常见的实现隐私计算的技术路径包括联邦学习、安全多方计算、可信计算等, 此外区块链也是隐私计算的重要补充。

隐私计算也是当前数据保护领域各界关注的热点。在学术界, 近年来有关隐私计算的学术会议和论文呈现爆发式增长, 例如, 中国计算机学会多次组织隐私计算技术研讨会, 在国际顶级学术会议上 (如NeurIPS, ICML, AAAI, IJCAI等) 也多次出现有关隐私计算技术的专题研讨会, 每年出现的与隐私计算相关的学术论文也呈指数增长 (平均每年都超过一千篇)。**产业界** 愈加关注隐私计算技术和产品, 由中国信息通信研究院牵头成立的“隐私计算联盟”有六十多家成员单位, 包括大型互联网公司、金融机构、初创型科技公司等企业。各企业单位都争相投入隐私计算研发和产品化工作, 有多家公司都推出了自己的隐私计算平台产品, 并开始进行隐私计算在金融、医疗等领域的商用落地。**政府部门和监管机构** 也非常重视隐私计算技术的发展, 一方面希望能够通过隐私计算技术推进安全的数据协同应用、推动数据经济发展, 另一方面也积极制定规范和指导意见, 促进隐私计算技术及产业健康发展, 推动合法、合规的数据协同应用。

2、隐私计算的主要作用

对于个人消费者而言,隐私计算应用有助于保障个人信息安全。个人消费者在享受数字经济便利与发展红利的同时,个人信息也被采集和广泛应用,同时也面临着信息泄露风险,而隐私计算在很多场景的应用,可以提升对个人信息的保护水平,降低个人信息在应用过程中泄露的风险。例如欧洲的MELLODDY项目中,多家药企正在探索借助隐私计算和区块链来进行的基于AI的药物研发,不同于所有数据收集到一个集中的位置进行训练,借助隐私计算只需在本地设备上训练AI模型,然后将这些学习结果传输回一个全局模型,而数据不需要离开任何特定的设备,并且可以通过区块链保持对数据的可控性,实现对诊断数据、健康数据等的保护。

对于企业和机构而言,隐私计算是数据协作过程中履行数据保护义务的关键路径。一方面,在企业内借助隐私计算,能够切实保护企业在采集、存储、分析等过程中的关键信息、商业秘密等数据,既能保护企业自身的利益,还能践行企业的数据保护责任。另一方面,隐私计算能够促进企业的跨界数据合作,由于隐私计算能够实现数据可用不可见,能够帮助不同企业和机构与产业链上下游的主体进行联合分析,打造数据融合应用,同时在数据协作的过程中履行数据安全和合规义务,实现生态系统内的数据融合,推动企业自身、产业层面的数据价值最大化。

对于政府而言,隐私计算是实现数据价值和社会福利最大化的重要支撑。一是借助隐私计算能够在政府数据开放过程中,在采集、存储、协作等方面提升数据安全和隐私保护水平,在保障数据安全的同时增强全社会的数据协作,通过数据的应用最大化社会福利。二是借助隐私计算推动数据要素赋能产业升级,例如北京国际大数据交易所上线北京数据交易系统,基于区块链和隐私计算技术支持的全链条交易服务体系,将为市场参与者提供数据清洗、供需撮合、法律咨询、价值评估等一系列专业化服务。

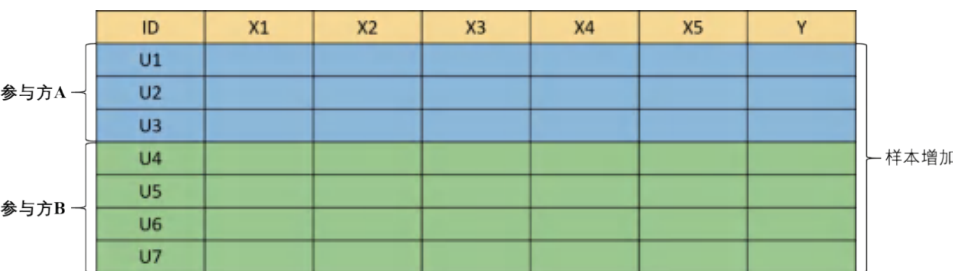
二、隐私计算三大流派交织演进，和区块链融合成为主流方向

隐私计算伴随着密码技术、硬件技术的发展加速商业化，其技术路径也处于高速的演进和变化状态，其中联邦学习、多方安全计算和可信计算是当前主流技术路径，也是当下产品化的主要方向。此外区块链与隐私计算的融合应用也成为业界的共识，两者相辅相成。

（一）联邦学习助力实现多方联合机器学习

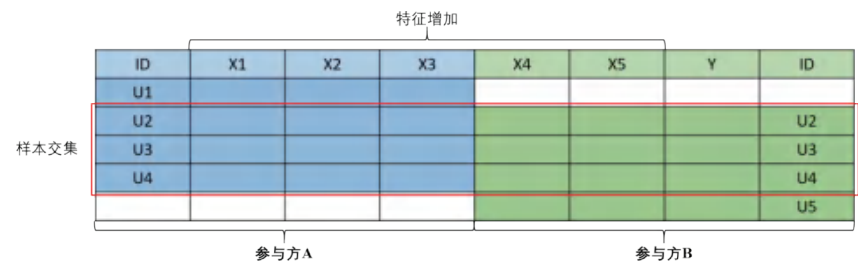
1、基本概念

联邦学习是一种分布式机器学习技术和系统，包括两个或多个参与方，这些参与方通过安全的算法协议进行联合机器学习，可以在各方数据不出本地的情况下联合多方数据源建模和提供模型推理与预测服务。在联邦学习框架下，各参与方只交换密文形式的中间计算结果或转化结果，不交换数据，保证各方数据不露出。联邦学习可以通过同态加密、差分隐私、秘密分享等提高数据协作过程中的安全性。根据联邦学习各参与方拥有的数据的情况，可以将联邦学习分为两类，即横向联邦学习和纵向联邦学习。



	ID	X1	X2	X3	X4	X5	Y
参与方A	U1						
	U2						
	U3						
参与方B	U4						
	U5						
	U6						
	U7						

图2 横向联邦学习中数据合作示例



	ID	X1	X2	X3	X4	X5	Y	ID
参与方A	U1							
	U2							U2
	U3							U3
	U4							U4
参与方B								U5

图3 纵向联邦学习中特征合作示例

如图2所示，在横向联邦学习中，参与方在各方数据的“数量”这个维度上进行合作，解决单个参与方的训练数据不足的问题。如图3所示，在纵向联邦学习中，参与方在数据的“特征”和“标签”这两个维度上进行合作，解决单个参与方的数据特征过少或者没有标签的问题。纵向联邦学习需要计算参与方共同拥有的样本ID，可以通过多方安全计算中的隐私集合求交技术实现。

2、技术趋势

联邦学习在深度学习领域的探索成为未来焦点。联邦学习在机器学习领域的应用已经比较成熟,如支持联邦逻辑回归、联邦XGBoost等模型,而在深度学习领域的应用还处于探索阶段。**一方面,联邦学习需要支持更加多样化的深度学习模型,**如广告领域常用的双塔模型、点击率预估模型、自然语言处理模型等,尤其支持多方联邦神经网络模型的训练,并提供高效的、安全的、无损的联邦模型训练协议,从而实现基于深度学习的联合建模。**另一方面,联邦学习需要支持海量数据的深度学习模型训练,**在计算机视觉、自然语言处理、广告等领域需要通过海量数据来训练深度学习模型,但受限于目前联邦学习的技术缺陷,需要通过增加联合计算的并行度,优化多方对接的接口等方式实现对海量数据处理的支持。

联邦学习与其他隐私计算技术深度融合,加速向平台化演进。一方面,单一的隐私保护技术不能满足对联邦学习的安全性、效率、性能的要求、以及应对多样化的应用场景,联邦学习将与安全多方计算、区块链、可信计算等技术进行深入融合,并通过使用硬件加速技术,进一步提高联邦学习系统的安全性和交付效率,保证联邦训练的模型与集中训练的模型有相同的性能。**另一方面,通过技术融合,联邦学习产品将会向通用型平台化发展,**丰富服务模式,满足多样化的用户需求。按需提供数据安全保护服务和全栈的联合建模和联合分析功能,将成为联邦学习产品适应多样化业务场景的演进方向。

(二) 安全多方计算提供更加安全的联合数据分析能力

1、基本概念

安全多方计算是一种在参与方不共享各自数据且没有可信第三方的情况下安全地计算约定函数的技术和系统。通过安全的算法和协议,参与方将明文形式的数据加密后或转化后再提供给其他方,任一参与方都无法接触到其他方的明文形式的数据,从而保证各方数据的安全。安全多方计算的基本安全算子包括同态加密、秘密分享、混淆电路、不经意传输、零知识证明、同态承诺等。解决特定应用问题的安全多方计算协议包括隐私集合求交、隐私信息检索及隐私统计分析等。

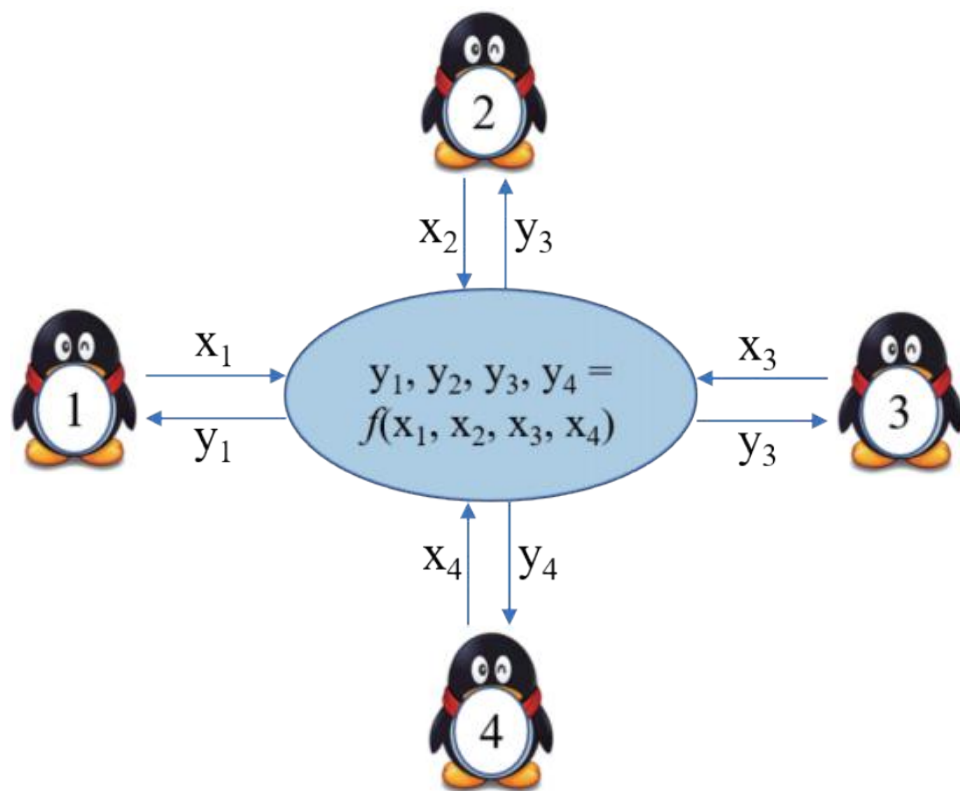


图4 安全多方计算应用示例

2、技术趋势

安全多方计算与其他隐私计算技术融合应用成为主流趋势。由于安全多方计算需要消耗大量的计算和通信资源，目前应用更加适用于小规模数据量，并且应用主要是聚焦相对简单的统计、查询等类型的计算，而基于安全多方计算的联合建模框架只能支持相对简单的机器学习模型，如逻辑回归模型等。其主流的应用主要以安全技术的形式融合在其他隐私计算解决方案中，例如与联邦学习的结合，在样本对齐阶段通过隐私集合求交来实现参与方公共样本ID的发现；在联邦模型训练阶段，可以通过同态加密、秘密分享等技术来实现对中间技术结果或转化结果的保护。

安全多方计算产品的计算和通信效率提升呈现两大路径。安全多方计算需要用到相对复杂的密码学运算，其计算和通信开销会超过实际应用能承受的范围，导致无法实现在大规模数据上的应用。提升其计算和通信效率是当下技术演进的主流方向，主要呈现两大技术路径。**一是**聚焦减少算法的计算量和安全协议的消息交互量，通过压缩算法、采样、抽样等方式减少计算和通信开销，从而实现计算和通信效率的提升。**二是**通过新的密码学技术和设计新的算法协议，结合硬件加速技术（如GPU、FPGA、ASIC加速）和专有算法实现硬件来加速计算量较大的环节和步骤，进一步实现计算效率的提升。

（三）可信计算助力隐私计算服务安全高效运行

1、基本概念

可信计算指借助硬件CPU芯片实现可信执行环境（TEE），从而构建一个受保护的“飞地”（Enclave），对于应用程序来说，它的Enclave 是一个安全的内容容器，用于存放应用程序的敏感数据与代码，并保证它们的机密性与完整性。以Intel SGX为例，Enclave的内存区域是由CPU默认加密的，且只能被同一个Enclave中的代码所访问，即便是外部高权限实体（VMM、BIOS、SMM）也无法访问。目前，TEE的实现也包括ARM平台的TrustZone、AMD下的SEV等，但在隐私计算领域，以Intel SGX的应用较为成熟。可信计算（TEE）是基于硬件和密码学原理的隐私计算方案，相比于纯软件解决方案，具有较高的通用性、易用性和较优的性能。其缺点是需要引入可信方，即信任芯片厂商。此外由于CPU相关实现属于TCB，侧信道攻击也成为不可忽视的攻击向量，需要关注相关漏洞和研究进展。

在可信计算过程中,TEE保证的可信功能主要包括:

远程证明:使用TEE进行隐私计算的必备步骤——当一项计算任务存在多方协作时,比如参与方A需要将数据(一般是加密的中间态数值)传递给参与方B,那么就需要检验B的程序的确是在TEE中运行的。在这一情况下,B需要能够提供“证明”,来证实自己的确是符合参与方A预期的运行状态,这一运行状态除了TEE环境信息以外,也进一步包括对于所运行程序代码相关信息的核验。

可信信道:在A成功验证B传递过来的证明之后,A也验证了B的身份和计算环境,便可以建立一条安全的可信信道(如基于B的证书建立加密信道),用于后续的数据传输会话。

数据密封:TEE本身支持的一种密钥映射机制,以Intel SGX为例,在使用数据密封功能时,由CPU指令对既定的入参进行计算,生成相应的密封密钥。数据密封机制保证了Enclave对数据的密封(加密)和解封(解密)过程,只能于同一Enclave内进行,而密封数据的存放,可以落盘于Enclave外,从而实现可信存储。

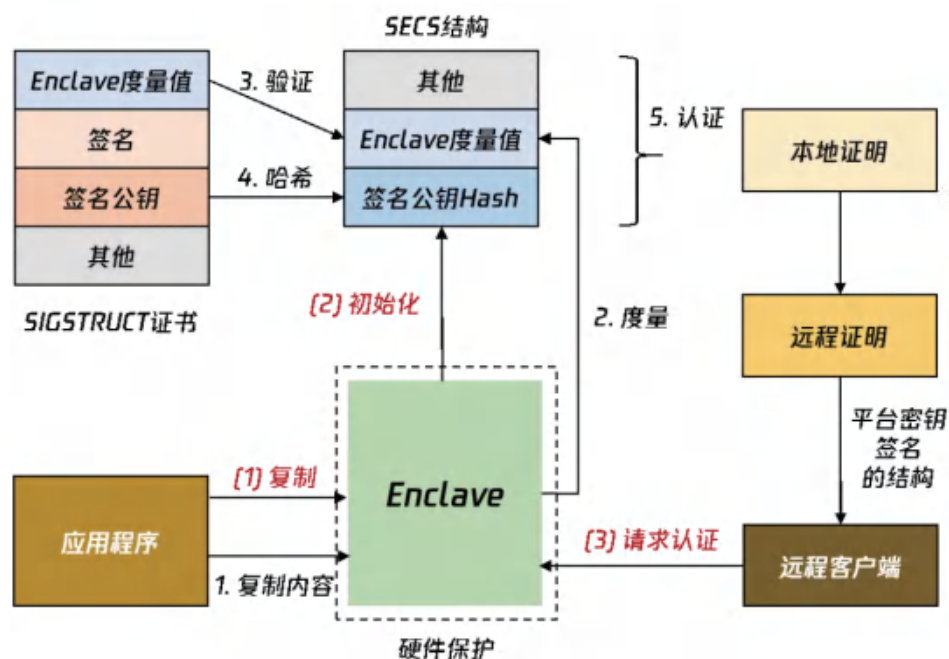


图5 可信计算应用实例图

2、技术趋势

平台化和容器化是未来可信计算与云平台融合的关键路径。可信计算开发和部署成本较高,为了满足多样化业务的需求,向使用者提供简易和低成本的服务,平台化和容器化正成为各大厂商的主流选择。例如通过将远程/本地证明、可信信道的建立、数据密封在内的可信功能整合在TEE基础平台,实现对于隐私计算任务的无差别工作流程正成为业界的主流路径。并且在此基础上实现TEE功能下工作负载的容器化,并交由Kubernetes进行统一管理与调度,达到可信工作节点“自来水”式的横向拓展和并行加速效果也是当前各方探索的路径。

可信计算的易用性提升是产品化应用的重要方向。为了更好地将平台功能应用于实际业务,易用性是建设可信计算基础应用平台所需兼顾的另一关键要素。基于原生SDK的开发存在学习门槛,很多实际业务应用依赖特定的库文件(如TensorFlow),此时基于SDK进行开发会非常繁琐。在TEE研究领域,已经出现了诸如库操作系统LibOS、程序自动分割等易用性适配方式。以SGX为例,LibOS实施方案中,比较典型的包括Graphene、SCONE、Occlum等。在使用相应LibOS的情况下,业务代码可以无需重构,直接通过LibOS在Enclave内部运行,这大大方便了业务应用的接入。

底层硬件架构的灵活切换是未来丰富可信计算应用场景的重要条件。实现对于底层硬件架构的灵活切换,是完善TEE隐私计算能力的又一关键点,也将是TEE领域下一阶段具有重要意义的工作。例如,将现有的TEE隐私计算平台进行硬件架构适配,通过硬件抽象层使其同时兼容Intel SGX、ARM TrustZone、AMD SEV等多重CPU架构,在此基础上,持续拓展可信计算的计算场景,将TEE功能由CPU延伸到GPU、FPGA等计算平台,从而满足用户在不同运算场景下的安全需求。

(四) 隐私计算融合区块链提升数据协作全流程保护能力

随着技术的不断发展, 区块链从一种防篡改、可追溯、共享的分布式账本管理技术, 转变为分布式的网络数据管理技术, 利用密码学技术和分布式共识协议保证网络传输与访问安全, 实现数据多方维护、交叉验证、全网一致、不易篡改。隐私计算虽然实现了在多方协作计算过程中对于输入数据的隐私保护, 但是原始数据、计算过程和结果均面临着可验证性问题。而区块链因其共享账本、智能合约、共识机制等技术特性, 可以实现原始数据的链上存证核验、计算过程关键数据和环节的上链存证回溯, 确保计算过程的可验证性。因此将区块链技术对计算的可信证明应用到隐私计算中, 可以在保护数据隐私的同时增强隐私计算过程的可验证性。

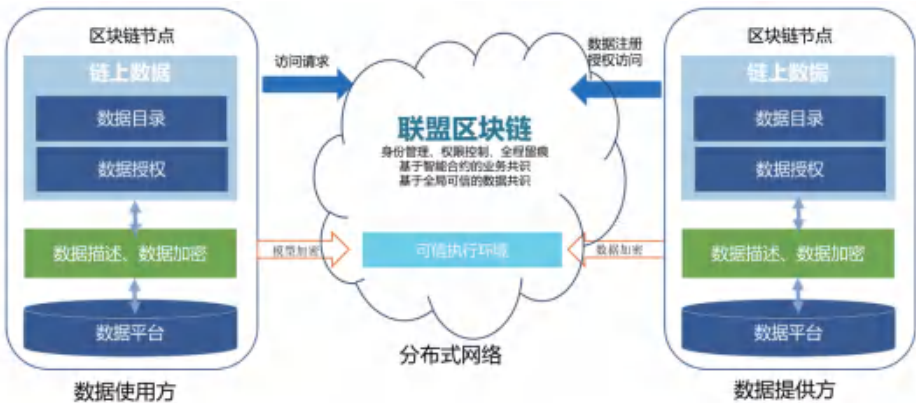


图6 区块链与隐私计算融合示例

区块链将成为隐私计算产品中必不可少的选项,在保证数据可信的基础上,实现数据安全、合规、合理的有效使用。主要体现在以下三个方面:

区块链可以保障隐私计算任务数据端到端的隐私性。通过区块链加密算法技术,用户无法获取网络中的交易信息,验证节点只能验证交易的有效性而无法获取具体的交易信息,从而保证交易数据隐私,并且可按用户、业务、交易对象等不同层次实现数据和账户的隐私保护设置,最大程度上保护数据的隐私性。

区块链可以保障隐私计算中数据全生命周期的安全性。区块链技术采用分布式数据存储方式,所有区块链上的节点都存储着一份完整的数据,任何单个节点想修改这些数据,其他节点都可以用自己保存的备份来证伪,从而保证数据不被随便地篡改或者是被删除。此外,区块链中所使用的非对称加密、哈希加密技术能够有效保障数据安全,防止泄露。

区块链可以保障隐私计算过程的可追溯性。数据申请、授权、计算结果全过程链上进行记录与存储,链上记录的信息可通过其它参与方对数据进行签名确认的方式,进一步提高数据可信度,同时可通过对哈希值的验证匹配,实现信息篡改的快速识别。基于链上数据的记录与认证,可通过智能合约,实现按照唯一标识对链上相关数据进行关联,构建数据的可追溯性。

区块链与隐私计算结合,使原始数据在无需归集与共享的情况下,可实现多节点间的协同计算和数据隐私保护。同时,能够解决大数据模式下存在的数据过度采集、数据隐私保护,以及数据储存单点泄露等问题。区块链确保计算过程和数据可信,隐私计算实现数据可用而不可见,两者相互结合,相辅相成,实现更广泛的数据协同。

三、数据协作需求推动隐私计算应用 从金融、医疗等向其他行业延伸

(一)金融

隐私计算助力银行联合建模,提升反欺诈模型水平。消费贷近年来兴起,随之而来的信贷欺诈也越来越严重,恶意骗贷、仿冒他人骗贷、团伙欺诈等欺诈行为对银行等相关信贷机构造成了严重的损失。传统上,银行都是基于历史还款信息、征信数据和第三方的通用征信分来做贷前反欺诈,仍存在数据维度缺乏、数据量较少等情况,需要融合多方数据联合建模才能构建更加精准的反欺诈模型,但这一过程中隐私保护和数据安全是不可忽视的重要环节,联邦学习可以有效解决合作中数据隐私与特征变量融合矛盾,在双方或多方合作中线上保障特征变量交换时的信息安全。例如某银行应用腾讯隐私计算产品,融合多方的黑灰产行为等特征,模型的KS提升30%以上,每年阻止数亿资金的风险贷款申请。

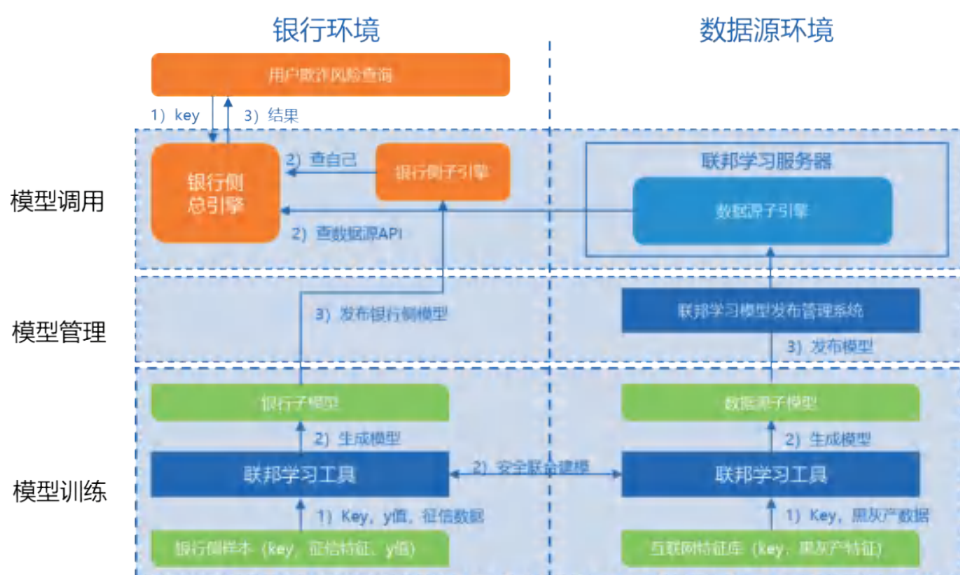


图7 隐私计算在金融反欺诈场景应用示例

(二) 医疗

隐私计算有效助力医学影像识别、疾病筛查、AI辅助诊疗、智能问诊咨询等。医疗数据通常存储于不同的机构中,且单个医疗机构拥有的带标签的数据的规模和特征维度都有限。由于病人隐私和数据保护的考虑,医疗数据无法在多个医疗机构之间直接共享或集中整合。数据整合问题制约了AI技术(如计算机视觉和目标检测等)在医疗领域的发展和应用。为了解决这个问题,医疗机构开始采用基于隐私计算的数据合作方案,多个医疗机构在不需要共享原始数据的情况下就可以进行联合建模和联合数据分析,有效推动了AI技术在医疗领域的应用多地。例如多家医疗机构可以通过横向联邦学习联合构建目标检测模型,用于辅助通过医疗图像的疾病检查(如肺部X光片检查等)。基于横向联邦学习的解决方案在各医疗机构的数据不出域的前提下,利用多家医疗机构的数据联合训练一个目标检测模型,使得有效训练数据显著增加,多方联邦训练的模型的性能比单个医疗机构训练的模型的性能提升30%以上。

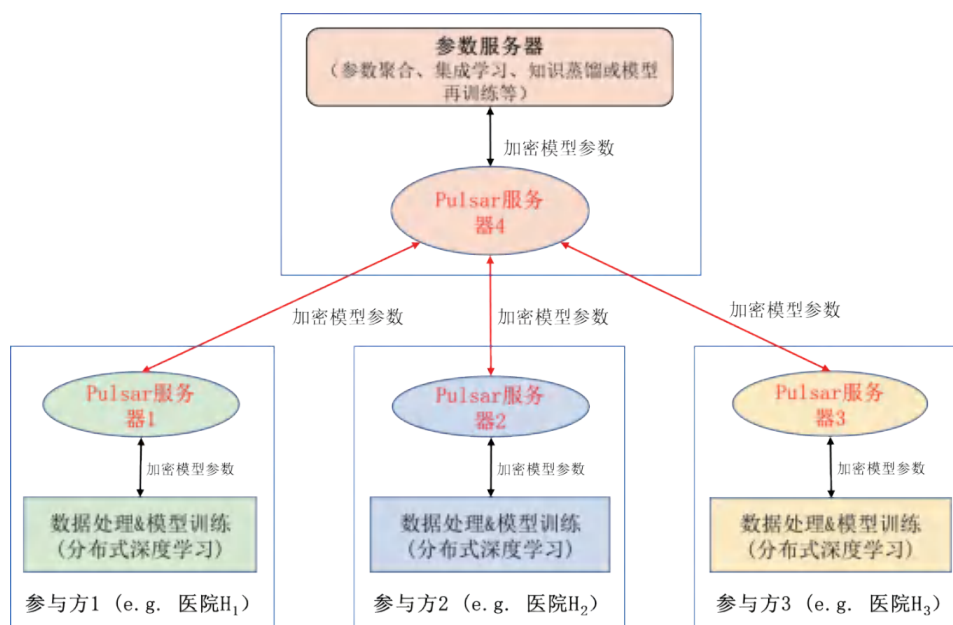


图8 隐私计算在医疗领域疾病检查应用示例

(三) 政务

基于隐私计算助力政府数据开放, 实现精准施策。政府部门汇集了大量的如交通、社保、税务、医疗和教育等高价值数据, 推进政务数据开放共享, 有助于促进社会经济的发展 and 提升政府的治理和服务水平, 尤其是在政策实施过程中, 通过政府数据与多方数据的融合, 能够实现基于数据驱动的精准施策。但因涉及到个人信息保护等问题, 以往的政务数据开放, 还是处在以统计形式为主的信息公开这个层次, 可用性大大减弱。借助隐私计算, 可以提升政务数据的含金量, 实现隐私保护下的高质量数据协作。另外, 通过隐私计算平台, 可以促进政务和企业的数​​据协作, 实现政企数据融合应用。例如在某地, 通过腾讯安全提供的联邦学习平台, 实现了政务、银行、企业的三方的协作建模, 在疫情期间对小微企业进行了精准画像, 模型的AUC提升了40%, 实现了企业综合评估、银行授信和政府贴息全闭环, 大大降低了信息不对称的成本, 提升了资金流转的效率, 促进了产业政策精准落地。

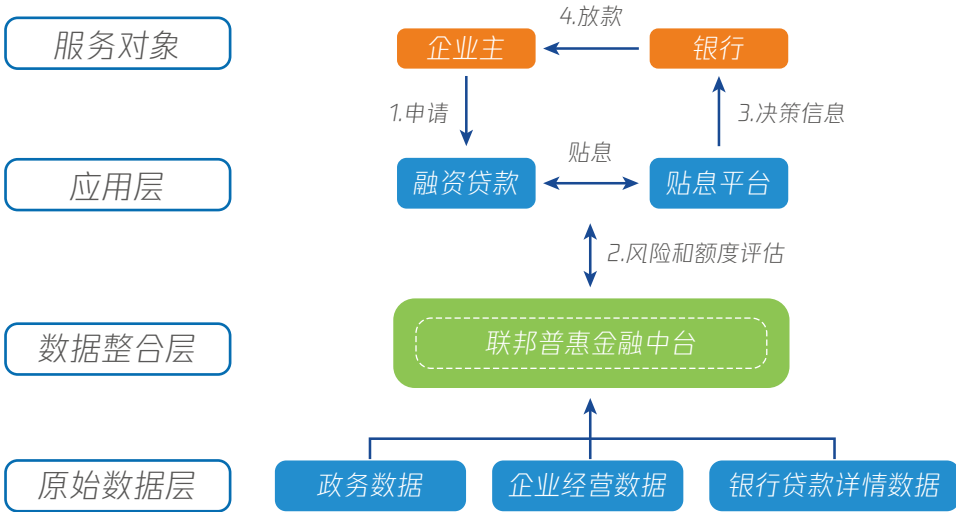


图9 隐私计算在政务领域的精准施策应用示例

(四) 广告

联邦学习助力广告程序化交易联合建模,提升广告主投放效果和用户体验。在广告场景中,流量方和广告主侧各拥有一部分链路数据,比如流量方拥有流量相关点击行为和基础画像,而广告主侧拥有深度转化链路数据如付费,后者属于广告主核心资产,不能完全同步给流量方,但是双方都有需求优化广告投放效果,以提升成本控制和起量效果。借助联邦学习可以在保护合作双方各自数据安全的前提下,联合训练、建模、优化模型效果。在这样的背景下,通过广告主和流量主的联邦建模,融合双方的数据优势,在游戏、金融、教育、电商行业的广告应用案例中能够取得显著效果提升,如某电商ADX模式中,ROI能够取得了10%以上的增长。

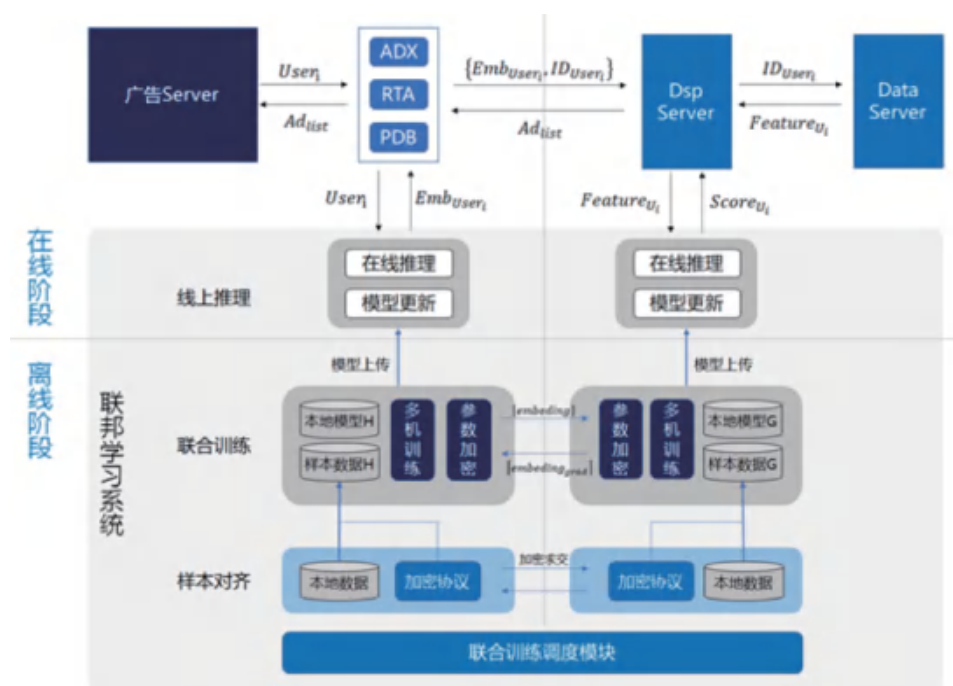


图10 隐私计算在广告领域的精准投放应用示例

四、隐私计算助力数据安全合规的价值凸显，但仍存在较大提升空间

隐私计算为数据安全制度落地提供了有力的技术支撑，面对数据安全合规，企业和机构可充分运用隐私计算技术+传统合规措施结合的方式，解决当前企业面临的数据合规难题和痛点。但由于隐私计算技术正处于快速迭代和发展的阶段，目前仍在实现用户授权同意、数据存储安全、信息主体权利保障等关键合规要求的有效性上存在争议，这些争议在一定程度上限制了隐私计算的推广应用。未来，隐私计算的发展和应用对法律合规的支撑价值凸显，但仍存在较大提升空间。

（一）隐私计算有望成为数据协作过程中数据合规和隐私保护的技术工具

隐私计算，旨在通过技术保障数据协作过程中的数据安全。对企业履行数据合规义务具有积极作用，具体体现在以下三个方面。

隐私计算在无需转移数据物理存储服务器的情况下实现数据建模分析，从而减少数据协作过程中风险。联邦学习、安全多方计算等隐私计算技术秉承“数据可用不可见，数据不动模型动”的理念，不流通原始数据，只回传数据的计算模型，并以此实现数据价值出库。同时，像全同态加密等以密码学为基础的隐私保护计算技术，还可以通过保证加密算法的强度、加密密钥的长度和密钥管理的安全性来实现数据匿名化。所以理论上，在隐私计算技术的助力下，无需转移数据物理存储服务器的情况下，数据合作方之间即可实现基于双方或多方数据的建模分析，而不需要将双方或多方数据共享、存储到某一方服务器处，从而减少数据协作过程中的风险。

隐私计算可从技术层面满足数据最小化、完整性和机密性原则要求。数据最小化、完整性和机密性均是《通用数据保护条例》(GDPR) 关于个人数据处理的重要原则,在我国《民法典》与《个人信息保护法(草案)》中也有所体现,是国际社会公认的个人信息的收集处理的基本要求。数据最小化要求“对个人数据的处理数量以满足业务需要的最小数量为限”,数据完整性和机密性则要求“避免数据被非法处理、篡改、毁损或者不当泄露”。传统的数据融合方式往往需要先将尽可能多的多源数据集中至一个数据中心,然后再训练模型。如此,传统的数据融合方式不仅存在数据过度采集的可能,而且面临数据传输与储存阶段的双重安全风险,相关参与方不仅需要阻拦外部的数据攻击,还要防范内部的数据违规使用,需要为之付出极大的成本。而采用隐私计算技术,尤其是隐私计算和区块链等技术结合形成的整体解决方案,对数据真实性、准确性进行记录,如数据被篡改、可进行精准定位和追溯,防止数据被篡改,也能够有效防止数据被内外部无权限人员随意访问、修改、导出等,保障数据的完整性和机密性,与当前国内外数据保护相关立法目的和原则高度契合。

隐私计算可证明、记载企业是否履行数据安全保障义务。一般而言,企业可围绕以下三个方面证明已履行法定的数据安全保障义务:(1) 已制定周密的数据安全管理制度;(2) 已执行严格的国际规范与标准;(3) 已采取有效的数据安全保障措施等。随着技术的成熟,隐私计算技术(包括基于芯片的可信执行环境TEE、基于密码学的安全多方计算MPC、同态加密、源自人工智能的联邦学习等)在保护数据安全方面的优势获得了行业的广泛认可,部分企业的隐私计算技术还顺利通过行业安全评估。基于此,在没有相反证据的情况下,采用隐私计算技术,可以清晰地记录企业已履行法定的数据安全保障、防止数据泄露的义务,在发生数据泄露的情形下,可及时提出相应证据证明,数据在哪个环节遭到泄露、是哪个主体泄露了数据,从而避免因难以查清泄露原因和主体,而导致企业可能需承担举证责任倒置义务,但又实际无法证明数据不是由该企业泄露的,从而承担相应法律责任。

(二) 隐私计算的推广应用仍存在合规痛点

从技术层面而言,隐私计算实现的数据保护功能与国内外数据保护相关立法精神高度契合,具有广阔的发展前景。但在全球数据合规监管日趋严格的大背景下,隐私计算仍具有较大提升空间。

采用隐私计算,仍需明确用户授权同意机制。隐私计算是解决数据流通环节用户授权的有效工具。根据我国《网络安全法》及《民法典》的规定,数据处理者在处理数据时应公开收集、使用规则,并经用户同意。从理论上而言,数据合作方通过隐私计算技术实现数据分析与建模,不需实际流转数据,且处理过程中的数据都进行了匿名化处理,或不需要获得用户授权同意。但实践中,在原始数据采集阶段,数据合作各方仍需获得用户授权同意。此外,由于个人信息的匿名化标准尚存争议,因此做好告知同意的授权管理,对强化企业数据合规仍具有重要意义。

此外,即便隐私计算的参与方可以对外公开或提供的是数据模型而非原始数据为由,规避协作环节的用户授权,但其在本地服务器中建模的行为也可能面临用户授权与否的拷问。即使企业在采集数据时通过隐私政策取得了用户对本地建模行为的授权,但该授权仍需保持在与数据实际处理目的直接或合理关联的范围内。因此,在借助隐私计算技术解决用户授权问题时,也需关注数据处理目的本身的合法性。

隐私计算技术本身具有中立性,在判断是否合规的问题上,关键还在于数据处理目的本身是否合法。对应地,建议隐私计算技术的研发、应用各参与方应当遵循科技向善原则,引导隐私计算技术在良性数据生态环境中落地。

隐私计算应用过程中也需重视数据安全风险。以联邦学习为例,尽管其无需参与者直接共享原始数据,但模型更新仍然会泄露参与者训练数据的相关信息,攻击者可以采用推理攻击判断具体的数据点或数据属性是否被用于训练,或采用逆向学习的方法还原原始数据。如果有切实的证据证明经过隐私计算的数据结果具有可逆性且已被泄露,那么它便不再属于法律规定的“经过处理无法识别特定个人且不能复原”的数据,企业未经授权或授权不充分的共享与转让行为将很可能被认定为对个人信息主体权益的侵犯。因此,企业可能需从模型隐私、输入隐私、训练数据隐私、输出隐私四方面保障数据的安全。

隐私计算应用过程中个人信息主体权利请求的实现仍需进一步探索。当下,我国数据立法整体倾向于加强对个人信息主体权益的保护,这是隐私计算合规无法回避的重点问题。以差分隐私为例,“差分隐私对个人信息保护与数据利用的平衡,部分取决于隐私计算这一参数的自行设定,和 δ 概率容忍度的主观评价”,这将不可避免地导致外界对隐私计算逻辑的质疑,虽然数据处理者可尝试用通俗的表达明示隐私计算的基本逻辑,但披露的程度则又是另一个复杂的问题。此外,个人信息主体的权利请求在隐私计算下如何得以保障,也是一个需要重点攻克的难题。

隐私计算参与各方权利义务的边界有待进一步明确。隐私计算涉及多方主体:(1)个人信息主体;(2)数据持有方,即为隐私计算提供数据的个人或组织;(3)计算方,即为隐私计算提供算力的个人或组织;(4)结果方,即接收隐私计算结果的个人或组织。各方之间的法律关系尚未厘清,数据收集处理的商业合作将处于不合理的高风险状态,如个人信息主体是否基于对原始数据的所有权而对经隐私计算的数据模型享有权益,发生数据泄露且溯源取证困难时,后三者间应如何进行责任划分,这些都将影响隐私计算商业模式的发展。现阶段,隐私计算参与各方宜通过协议方式,约定彼此的数据安全权利和义务边界,以便在发生争议时,明确各自的责任范围。

五、技术演进、应用拓展和法律完善 将加速隐私计算商业化进程

（一）效率、性能提升和技术融合将成为隐私计算产品化的主要方向

隐私计算效率和性能提升是未来规模化推广的重要前提。隐私计算虽然已经开始在不同行业初步应用,但是受限于计算复杂度、多方交互效率、模型性能等问题,大部分的应用场景均聚焦于少量数据的支持,对海量数据场景的支持能力还有待提升。但随着当前大数据产业的迅速发展,支持更大规模的数据合作和联合计算需求将越加迫切,通过优化算法和协议设计、与云平台的融合应用、软硬件协同设计等方式提升计算、交互效率将是当下和未来隐私计算发展需要重要方向,效率、性能、成本等综合能力将是各类主体在隐私计算产业竞争的重要抓手。

隐私计算多种技术路径深度融合,通用型隐私计算平台有望成为未来主要产品形态。为了完善和增强现有隐私计算解决方案,多种技术路径融合是必然的发展趋势。不同技术路径可以形成优势互补,多种安全技术机制深度融合,能够提供多层级的、按需的安全解决方案,从而适应多种应用场景。隐私计算产品会从单一的技术类型转向使用多种技术的方案,形成通用型隐私计算平台,以便能够提供多种安全技术路线和安全等级,具备从数据采集、接入、存储到建模、分析、应用的全流程隐私保护能力,整体推动数据合作和大数据产业发展。

（二）隐私计算应用场景不断拓展,有望重塑数据使用模式

隐私计算应用场景将从金融、医疗等领域向其他行业加速拓展。在日益严格的数据监管趋势和多方数据协作的迫切需求下,隐私计算将会成为大数据产业发展过程中数据协作基本解决方案,根据Gartner预测,到2025年,将有一半的大型企业会通过隐私计算赋能多方数据合作场景中的数据融合应用。目前虽然隐私计算的场景主要聚焦金融、医疗等领域,但随着其产品化、商业化的进程的加速,以及用户对隐私计算的接受度的提高,隐私计算也正往交通、教育、工业等领域延伸,并且将形成跨机构、跨企业、跨行业的多类应用场景,有望在更多行业进行拓展应用。

隐私计算将加速基于数据协作的业务模式创新。一方面隐私计算能够规避数据协作过程中传统数据收集、传输、交易等过程中带来的安全风险,解决网络连接费用昂贵、传输速度缓慢、传输安全性低等问题,为业务的发展提供更多的自由空间。另一方面隐私计算通过安全机制和技术手段联通多方数据源,重新定义各数据协作方的合作方式,可以解决以往数据主体的协作困境,从而实现业务形态、应用场景、商业模式等方面的创新。

(三) 隐私计算将通过助力法规政策落地促进数据融合应用

通过技术与制度配套推进的方式实现数据保护将是隐私计算发展的有效路径。隐私计算虽然从技术层面实现了隐私保护与数据协作之间的动态平衡,对桥接数据孤岛、释放数据价值具有不可替代的作用。但需要强调的是,技术固然是实现合规的关键手段,但是合理、科学的制度也是数据保护过程中必不可少的一环。对于隐私计算而言,在接受法律制度规制的同时,配合法律、政策、标准等相关制度共同实现数据保护将是其产品化和商业化的前提。

隐私计算需动态适应法规政策的变化与完善,进一步落实合规要求。2019年10月,《中共中央关于坚持和完善中国特色社会主义制度 推进国家治理体系和治理能力现代化若干重大问题的决定》首次将数据增列为生产要素,数据要素与资本、劳动、土地等要素一样,也应当建立健全由市场评价贡献、按贡献决定报酬的机制。2020年3月,《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》再次强调,“加快培育数据要素市场”。在此背景下,若将来法律明确了数据及其衍生品(如隐私计算的模型)的权属及法律地位,规定了相应的定价方式与交易制度,那么隐私计算等技术的发展方向也将改变。比如,信息处理者是否可在未经充分授权的情况下,将合法采集的无法识别特定主体且无法复原的数据进行自由地共享转让?虽然目前该问题的答案暂时是肯定的,但如果数据立法对个人信息权益进行更加严格的保护,认为数据衍生品的所有权也属于个人,其流通必须经过个人的明确授权,那么问题的答案也将随之变化,隐私计算合规方案也将随之调整。

参考文献

(1) 著作类

[1] 杨强, 刘洋, 程勇等:《联邦学习》, 电子工业出版社2020年版。

(2) 论文期刊类

[1] 杨庚, 王周生:《联邦学习中的隐私保护研究进展》,

载《南京邮电大学学报(自然科学版)》2020年第40卷第5期。

[2] 李凤华, 李晖, 牛犇等:《隐私计算——概念、计算框架及其未来发展趋势》,

载《工程(英文)》2019年第5卷第6期。

(3) 网址及其他

[1]《联邦学习能否解决金融数据整合难题?》, 载威科先行网,

[https://law.wkinfo.com.cn/professional-articles/detail/NjAwM-](https://law.wkinfo.com.cn/professional-articles/detail/NjAwM-DAwNzM2OTA%3D?searchId=54188083a980472b9f8c811f3cc13619&index=1&q=%E8%81%94%E9%82%A6%E5%AD%A6%E4%B9%A0%E8%83%BD%E5%90%A6%E8%A7%A3%E5%86%B3%E9%87%91%E8%9E%8D%E6%95%B0%E6%8D%AE%E6%95%B4%E5%90%88%E9%9A%BE%E9%A2%98%EF%BC%9F&module=)

[DAwNzM2OTA%3D?searchId=54188083a980472b9f8c811f3cc13619&index=1&q=%E8%81%94%E9%82%A6%E5%AD%A6%E4%B9%A0%E8%83%BD%E5%90%A6%E8%A7%A3%E5%86%B3%E9%87%91%E8%9E%8D%E6%95%B0%E6%8D%AE%E6%95%B4%E5%90%88%E9%9A%BE%E9%A2%98%EF%BC%9F&module=.](https://law.wkinfo.com.cn/professional-articles/detail/NjAwM-DAwNzM2OTA%3D?searchId=54188083a980472b9f8c811f3cc13619&index=1&q=%E8%81%94%E9%82%A6%E5%AD%A6%E4%B9%A0%E8%83%BD%E5%90%A6%E8%A7%A3%E5%86%B3%E9%87%91%E8%9E%8D%E6%95%B0%E6%8D%AE%E6%95%B4%E5%90%88%E9%9A%BE%E9%A2%98%EF%BC%9F&module=)

[2]《中国信通院“卓信大数据”计划-联邦学习安全评估第一期顺利完成》, 载中国信通院CAICT公众号, <https://mp.weixin.qq.com/s/ZeJqQGZpUj9B8AcZM54BaA>.

[3]《规则的激荡与新生——2020年数据治理年度报告》, 载腾讯研究院公众号, <https://mp.weixin.qq.com/s/LTi33dYJSC7NSUn4NaL96A>.

[4]《隐私计算工具的<个人信息保护法>评价(一)——差分隐私》, 载公安三所网络安全法律研究中心公众号, https://mp.weixin.qq.com/s/tV-EwPsgClFx47_8nvVAJg.

[5]《隐私保护计算技术研究报告(2020年)》, 载中国信通院, <http://www.caict.ac.cn/kxyj/qwfb/ztbg/202011/P020201110408006418997.pdf>

[6]《隐私计算——实现数据价值释放的突破口》, 载中国信通院CAICT公众号, <https://mp.weixin.qq.com/s/2jL3x-m-IX5Cc8ZRSxpdVw>

[7]《Gartner Top Strategic Technology Trends for 2021》, 载Gartner网, <https://www.gartner.com/smarterwithgartner/gartner-top-strategic-technology-trends-for-2021/>

[8]《隐私计算产品测试及行业发展观察》, 载隐私计算联盟公众号, <https://mp.weixin.qq.com/s/472o6CBMYbh2TPvX3rWv9g>

[9]《Federated Machine Learning: Concept and Applications》, 载Cornell University, <https://arxiv.org/abs/1902.0488>

- [10]《A Pragmatic Introduction to Secure Multi-Party Computation》,
<https://securecomputation.org/>
- [11]《区块链白皮书(2020年)》,载中国信通院, <http://www.caict.ac.cn/english/research/whitepapers/202101/P020210127494158921362.pdf>
- [12]《隐私计算产品测试及行业发展观察》,载隐私计算联盟公众号,
<https://mp.weixin.qq.com/s/472o6CBMYbh2TPvX3rWv9g>
- [13]《隐私计算应用发展现状与趋势》,载隐私计算联盟公众号,
<https://mp.weixin.qq.com/s/ETVL2QGL4Nxl9NT84TZqZQ>
- [14]《Gartner Top Strategic Technology Trends for 2021》,载Gartner网,
<https://www.gartner.com/smarterwithgartner/gartner-top-strategic-technology-trends-for-2021/>
- [15]《深潜数据蓝海——2021隐私计算行业研究报告》,载金融科技微洞察公众号,
<https://mp.weixin.qq.com/s/hiCbnCS2iiJSrj4lawHSow>
- [16]《腾讯云区块链TBaaS产品白皮书(2018年)》,载腾讯云,
<https://main.qcloudimg.com/raw/565be73-decf6badd55779613908a3319/%E8%85%BE%E8%AE%AF%E4%BA%91%E5%8C%BA%E5%9D%97%E9%93%BETBaaS%E4%BA%A7%E5%93%81%E7%99%BD%E7%9A%AE%E4%B9%A6.pdf>
- [17]《Intel SGX Explained》, <https://eprint.iacr.org/2016/086.pdf>
- [18]《Innovative Technology for CPU Based Attestation and Sealing》,
载英特尔官网, <https://software.intel.com/content/www/us/en/develop/articles/innovative-technology-for-cpu-based-attestation-and-sealing.html>
- [19]《Certificate authority》,载维基百科,
https://en.wikipedia.org/wiki/Certificate_authority
- [20]《AI药物发现的数据共享模式探索:十大顶尖药企参加的MELLODDY项目》,
载知乎, <https://zhuanlan.zhihu.com/p/348495604>
- [21]《Graphene-SGX: A Practical Library OS for Unmodified Applications on SGX》,
载USENIX, <https://www.usenix.org/system/files/conference/atc17/atc17-tsai.pdf>
- [22]《Occlum: Secure and Efficient Multitasking Inside a Single Enclave of Intel SGX》,载arXiv, <https://arxiv.org/pdf/2001.07450.pdf>
- [23] 中国人民银行《多方安全计算金融应用技术规范》(JR/T 0196—2020),载中国金融电子化公司, <https://www.cfsc.org/jinbiaowei/2929436/2978019/index.html>
- [24] 中国人民银行《金融业数据能力建设指引》(JR/T 0218—2021),载中国金融电子化公司, <https://www.cfsc.org/jinbiaowei/2929436/2978882/index.html>
- [25]《28.1万起数据泄露:GDPR生效后,欧洲已有这些处罚》,载南方都市报,
<https://www.163.com/dy/article/G0VPUN1405129QAF.html>