隐私科技白皮书

CSA GCR cloud security alliance®
GREATER CHINA REGION

致 谢

云安全联盟大中华区（简称：CSA GCR）隐私科技工作组在 2021 年 7 月成立。由高轶峰、徐震天 担任工作组联席组长，工作组专家来自安永、优衣库、平安科技、安恒、美创、数安行科技、极氪汽车、腾讯、e 签宝、宇链科技、优刻得、竟安科技、Oppo、阿里巴巴、世平、360 数科、观安信息、爱加密、工行、安华、大华等二十多个单位。

本白皮书由 CSA 大中华区 隐私科技 工作组专家撰写，感谢以下专家的贡献：

联席组长：高轶峰、徐震天

贡献者名单

原创作者：谢江、沈赟、滕海明、郭伟、何永德、聂桂兵、蔡毅、朱垒

审核专家：欧建军、顾伟、王安宇、贺志生、郭鹏程、姚凯

研究协调员：麦尔维娅

贡献单位：上海观安信息技术股份有限公司、杭州宇链科技有限公司、优刻得科技股份有限公司、杭州安恒信息技术股份有限公司、OPPO、北森云计算有限公司

（以上排名不分先后）

关于研究工作组的更多介绍，请在 CSA 大中华区官网（https://c-csa.cn/research/） 上查看。

如本白皮书有不妥当之处，敬请读者联系 CSA GCR 秘书处给与雅正! 联系邮箱：info@c-csa.cn；云安全联盟 CSA 公众号：

序 言

首先祝贺《隐私科技白皮书》的发布，这本白皮书由 CSA 隐私科技工作组编写，CSA 大中华区专家组评审。

科技的迅猛发展为社会生活带来了极大的便捷性，但随之而来的是海量个人信息的收集与处理，这为数据保护与个人隐私权益的保护带来了巨大挑战。近年来，国家层面相继发布了多部个人信息保护与网络安全、数据安全相关的法律法规，保障国家安全、公共利益和个人隐私权益。如何在满足法律合规要求、保障个人安全性、保护个人隐私权益的同时，促进个人信息的有序流动与使用。

本书从隐私合规、数据安全、数据可用的维度出发，开创性的提出了“隐私科技”的概念，详细描述了其定义、发展历程、技术以及应用场景，分析了全球以及中国的隐私科技产业环境，同时深入浅出的描绘了隐私科技的发展趋势，值得大家参考。

李雨航 Yale Li

CSA 大中华区主席兼研究院院长

目录

1. 隐私保护风险、合规及趋势大环境分析

1.1 隐私保护风险分析

在当今时代，互联网、大数据、人工智能等科技的迅猛发展为大众生活带来了便捷与高效，于此同时伴随而至的是对海量数据与个人信息的处理。2019 年 10 月，党的十九届四中全会决议通过《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》将数据列为生产要素。为落实党中央的重大决策部署，2020 年 4 月，中共中央、国务院出台《关于构建更加完善的要素市场化配置体制机制的意见》，明确要求推进政府数据开放共享、提升数据资源价值、加强数据资源整合与安全保护，加快培育数据要素市场。除此之外，全球其他国家或地区近年来相继出台法规政策以在国际上争夺数据主权。可以看到，数据作为一种新型生产要素，其价值及影响力不言而喻。

然而，数据的开发利用、价值挖掘、跨境流动等生产活动给数据安全或个人信息保护带来了巨大的挑战，可能伴随着危害国家安全、公共利益或个人隐私权益等一系列风险。在科技飞速发展的同时，全球各当局逐渐重视数据安全与个人信息保护。据不完全统计，目前已有 140 个国家和地区制定了与个人信息保护相关的法律或规定，亦有多个国家正在起草制定相关法律法规。特别是 2018 年生效的欧盟《通用数据保护条例》（GDPR）作为代表，对数据处理活动提出了较为严苛的合规要求，且自该法案生效后相关执法机构持续不断开展执法活动。

早在 2012 年，我国发布《全国人大常委会关于加强网络信息保护的决定》，从国家层面确认了个人信息保护的重要性与决心。党的十九届五中全会上，习近平总书记对保障国家安全、加强个人信息保护提出了明确要求。目前，在数据安全与个人信息保护领域我国已完成顶层制度设计，即以《网络安全法》、《数据安全法》、《个人信息保护法》为一体的网络与数据保护综合规范体系，并辅以相关配套法规规范、标准文件、行业要求等，从数据收集、使用、对外提供、公开披露、删除等全生命周期建立了多方位立体的数据合规体系。此外，从目前新法出台的频繁程度以及监管持续不断的通报处罚行动可知，我国对于数据安全与个人信息保护相关事宜的重视程度几近顶峰，相关要求不断提高，监管持续加码，企业在开展数据处理活动时面临巨大的合规压力与合规成本。

面对如此严峻的合规监管态势，同时日益增长的信息化时代智能便利的需求，如何释放数据要素价值、真正落地合规要求，成为了当前信息社会关切的重点。隐私科技的出现，一方面通过技术手段，帮助组织实现数据安全与个人信息保护的合规要求，如开展数据分类分级、个人信息保护影响评估、管理用户授权与响应等；另一方面，在数据使用与流通过程中，通过隐私计算技术对数据进行处理，使数据结合算法等技术手段，在数据“可用不可见”的前提下实现商业或公益目的，充分释放数据要素价值。

1.2 中国隐私保护合规与标准概述

我国早在 2012 年就在国家层面开始关注网络数据保护，同年出台了《规范互联网信息服务市场秩序若干规定》、《全国人民代表大会常务委员会关于加强网络信息保护的决定》文件，明确了“合法、正当、必要”原则，后续相关立法或标准文件制定中均予以沿用。2016 年，我国颁布了《网络安全法》，全面规定了网络空间中有关个人信息的安全与保护制度。2021 年，我国相继颁布《数据安全法》、《个人信息保护法》。自此，我国数据安全与个人信息保护领域的“三驾马车”已形成。在数据安全与个人信息保护领域的规则演进过程，我国呈现出了区别于其他国家和地区的鲜明特色，即“标准先行”的特点。在 2017 年，全国信息安全标准委员会发布 GB/T 35273《信息安全技术　个人信息安全规范》，从个人信息全生命周期角度提出个人信息处理应遵循的原则和安全要求。而后，全国信息安全标准委员会又陆续发布《信息安全技术 个人信息告知同意指南》（征求意见稿）、《信息安全技术 移动互联网应用（App）收集个人信息基本规范》（征求意见稿）、《信息安全技术 网络数据处理安全规范》（征求意见稿）、GB/T 39335-2020《信息安全技术　个人信息安全影响评估指南》等，从多维度提出了合规要求，但多数标准文件目前仍为征求意见稿。此外，《民法典》、《消费者权益保护法》、《电子商务法》等法律中也对个人信息保护提出相关规定与要求，相关部门就特定行业或特殊类型个人信息也发布了个人信息保护相关规定，如《电信和互联网用户个人信息保护规定》、《人口健康信息管理办法（试行）》、《个人信用信息基础数据库管理暂行办法》、《征信业务管理办法》、《汽车数据安全管理若干规定（试行）》、《儿童个人信息网络保护规定》等。

我国隐私合规领域法律文件、合规要求、涉及的相关国家标准与隐私科技应用的映射关系如表 1-1 所示：

表 1-1 隐私合规法律法规清单

法律文件	合规要求	相关规定/国家标准	隐私科技应用
《数据安全法》	数据分类分级制度	《工业数据分类分级指南（试行）》（工信厅信发〔2020〕6 号） JR/T0158-2018《证券期货业数据分类分级指引（试行）》 JR/T0197-2020《金融数据安全 数据安全分级指南》 YD/T 3813-2020《基础电信企业数据分类分级方法》 GB/T 38667-2020 信息技术 大数据 数据分类指南 YD/T 2781-2014 电信和互联网服务-用户个人信息保护-定义及分类 YD/T 2782-2014 电信和互联网服务-用户个人信息保护-分级指南	数据识别与分类分级工具
《个人信息保护法》	1.处理个人信息的，应当获得个人信息主体同意或具体其他合法事由。 2.个人信息处理者应当建立便捷的个人行使权利的申请受理和处理机制。	《信息安全技术 个人信息告知同意指南》（征求意见稿）	隐私设计工具（包括用户同意管理工具、用户权利响应平台等）
《个人信息保护法》	1.个人信息处理者应事前进行个人信息保护影响评估，并对处理情况进行记录： （一）处理敏感个人信息； （二）利用个人信息进行自动化决策； （三）委托处理个人信息、向其他个人信息处理者提供个人信息、公开个人信息； （四）向境外提供个人信息； （五）其他对个人权益有重大影响的个人信息处理活动。 2、个人信息保护影响评估应当包括下列内容： （一）个人信息的处理目的、处理方式等是否合法、正当、必要； （二）对个人权益的影响及安全风险； （三）所采取的保护措施是否合法、有效并与风险程度相适应。 个人信息保护影响评估报告和处理情况记录应当至少保存三年。	GB/T 39335-2021《信息安全技术个人信息安全影响评估指南》	隐私保护影响评估工具、隐私设计开发工具
《个人信息保护法》	采取相应的加密、去标识化等安全技术措施保障个人信息安全。	GB/T 37964-2019《信息安全技术 个人信息去标识化指南》 《信息安全技术 个人信息去标识化效果分级评估规范》（征求意见稿） DB31/T 1311-2021《数据去标识化共享指南》	数据去标识化/匿名化、同态加密、联邦学习等隐私计算技术

2. 隐私科技概述

隐私科技是一系列技术与解决方案的集合，它包含了如隐私计算，隐私增强技术，数据安全技术，数据及隐私合规科技等诸多技术领域范畴。隐私科技通过数字化手段解决组织在隐私保护工作中面临的痛点，在提升数据流通与共享能力的基础上确保数据安全与个人隐私得到有效的保护。

从应用场景出发，隐私科技主要解决以下三方面的问题：

● 隐私合规—作为隐私保护最强的驱动力，合法合规的处理数据是绝大多数组织所面临的最大痛点。随着数据量的爆发性增长以及数据处理场景的多元化，传统的人工方式识别隐私合规风险的方式已无法满足需要，需要使用自动化/智能化手段为组织展示隐私数据在组织内部的全貌，进而识别合规风险。

● 数据安全 – 海量的数据往往能够产生巨大的价值，这种巨大的诱惑也导致了各类数据盗用、泄漏事件不断发生，造成了社会各界对隐私数据安全性的担忧。各类法律法规的出台，也迫使数据所有者不断寻找可靠的方法保护数据的安全性。

● 数据可用 – 一方面，虽然全球数据总量处在指数性增长的过程中，但绝大多数仍分布在不同企业及信息系统当中，“数据孤岛”问题明显；另一方面，数据作为基础性资源，其所产生的效能也是持续推动数字经济体系发展的重要支撑。这就意味着不同组织间数据协作进而最大化挖取数据价值，已成为不可逆的趋势。如何在兼顾合规安全的前提下，打通不同组织/企业之间的数据壁垒，实现数据的“流通”与“共享”并挖掘其最大的价值，已成为数字经济发展的重要课题和推动力。

● 从技术角度出发，隐私科技涵盖了众多隐私计算技术，主要包含以安全多方计算为代表的基于密码学的技术、以联邦学习为代表的基于人工智能与算法的技术，以及以可信执行环境的为代表的基于硬件环境的技术。

● 从产品角度出发，主流产品主要包括两大类：一类是以数据可视化工具、隐私合规影响评估工具等为代表的管理类工具，主要解决组织的隐私合规问题；另一类是以可信计算/联邦学习框架等为代表的技术类产品与服务，主要解决组织的数据安全与可用的问题。

隐私科技全景图如 2-1 所示：

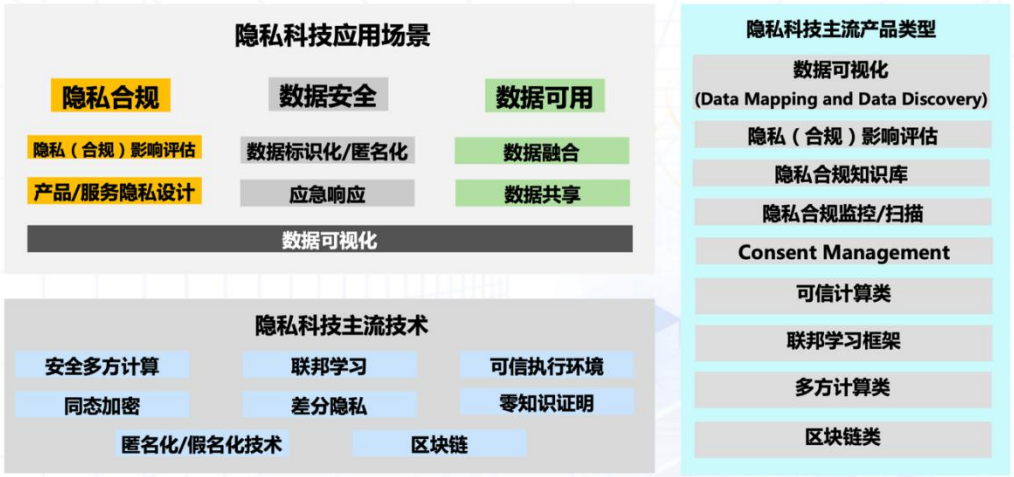

图 2-1 隐私科技全景图

3. 隐私科技技术

3.1 隐私合规技术

3.1.1 隐私合规影响评估

作为隐私保护最强的驱动力，合法、合规的收集和使用数据是所有政府组织与企业在隐私保护领域的最高优先级工作事项。现阶段企业组织面临来自不同监管机构的多重监管要求，上至近些年颁布的《网络安全法》、《数据安全法》与《个人信息保护法》这三部法律，下至针对特定行业或应用的要求，如移动应用软件的个人信息收集和处理要求，目的、颗粒度与监管的方式手段均有较大差异，企事业组织的合规成本显著增加。

在多头监管、多重监管的压力下，企事业组织基于传统的人工流程开展的“法条对标”工作，不但耗费大量资源，而且及时性和有效性也无法得到充分保障；一旦在某一环节有所遗漏，极易引发各类隐私合规问题进而遭受监管的严厉处罚。

隐私合规影响评估，指通过工具化的手段有效识别个人信息在收集、使用、存储、转移、销毁（统称“个人信息处理”）等各个环节的合规风险点，并指导数据处理者通过业务流程和技术手段规避风险，最大限度降低个人信息处理风险。

隐私合规影响评估技术高度依赖于系统工具，通过一种或多种系统工具帮助数据处理者将个人信息处理的业务场景与所有适用的法律法规对标，识别合规差距并提出改进建议。

一般而言，隐私合规影响评估工具包含如下功能模块：

● 具有较强实时性的法律法规标准库，并且该知识库可通过用户视角以清单、问卷的形式展示

● 供数据处理者输入业务流程信息的功能

● 评估结果与风险展示功能

● 改进建议展示功能

隐私合规影响评估工具的核心能力在于能够根据用户输入的结构化与非结构化信息，有效识别出合规差距并展示改进建议，因此其核心竞争力在于实时性较强的法律法规标准库、以及后端的判断与计算引擎。

此类工具在全球范围内已有了广泛的推广，截止 2021 年已有超过 300 家不同厂商为全球范围内数以万计企业提供服务，头部厂商的估值已超过 50 亿美金。在我国，无论是产品的成熟度，还是厂商/用户的重视程度，均与全球领先水平有着较大差距。但随着《个人信息保护法》及配套细则的逐步落实，在可以预见的未来，此类产品也会在国内得到快速推广和应用。

3.1.2 隐私设计

隐私设计又称 Privacy by Design, 是将隐私合规要求融入到产品、服务流程设计中的工作方法。任何一款需要收集、使用个人信息的产品或系统，在需求阶段就应该将隐私合规要求融入到产品设计之中，以确保该产品在功能层面能够满足适用的隐私合规要求。

隐私设计一般从数据生命周期出发，结合产品/系统的具体功能，将数据在收集、存储、使用、转移/传输、加工、提供、公开、销毁各阶段的隐私合规要求以系统需求的形式展现出来，并最终体现在产品功能之中。

由于不同行业的业务特性差别较大，因此隐私设计很难通过一个集成化的工具覆盖不同业务生态，目前全球范围内应用范围较广的隐私设计产品主要包括以下几类：

● Consent Manager 工具（知情同意管理工具） - 在产品功能层面收集、追踪、管理“用户同意”；

● Data Subject Request 工具（数据主体权利管理工具） - 管理并满足用户行使的隐私权利，如知情权、撤回同意权、个人信息删除权等；

● 匿名化/假名化 – 通过技术手段处理个人信息，使其在匿名化或假名化的状态下存储、使用和传输。

上述 Consent Manager 工具（知情同意管理工具）和 Data Subject Request 工具（数据主体权利管理工具）在传统基于 PC 浏览器的各类网站上已有了较大规模的应用和较好的实践，但仍然缺乏移动互联网下的应用场景：大多数的中国互联网移动应用程序（如 App/小程序），目前还依赖于应用程序本身的功能满足上述需求。

关于匿名化/假名化的实现，目前主流操作通过各类隐私计算技术实现此类需求。

3.2 隐私计算技术

3.2.1 多方安全计算

安全多方计算（Secure Multi-Party Computation，MPC）起源于 1982 年姚期智教授的百万富翁问题，也就是如何在不暴露各自财富的前提下比较谁更富有。安全多方计算用于解决在一组互不信任的参与方之间保护隐私的协同计算问题，能确保输入的独立性、计算的正确性、去中心化，同时不将输入值泄露给参与计算的其他成员。安全多方计算主要解决在无可信第三方的情况下，如何安全地计算一个约定函数的问题。安全多方计算是电子选举、电子投票、门限签名以及电子拍卖等诸多应用场景实施的密码学基础。

安全要点

如下图 3-1 所示：MPC 不是一个单一技术，是由一些列技术组成的协议栈，可以分为支撑技术层和具体构造 MPC 协议层两层。其中，支撑技术层主要提供用来构建 MPC 基础技术的实现协议，包含常用的加密解密、hash 函数、密钥交换、同态加密（Homomorphic Encryption，HE）等，同时还包含 MPC 中的基础工具：秘密分享（Secret Sharing，SS）、不经意传输协议(Oblivious Transfer，OT)等。构造的 MPC 协议分为两类，专用算法和通用框架。专用算法是为解决特定问题而构造的特殊 MPC 协议，由于是针对性构造并进行了优化，因此专用算法的效率会比基于混淆电路（Garbled Circuit，GC）的通用框架高很多，包含四则运算、比较运算、隐私集合求交集和隐私数据查询等；通用框架指可以满足大部分计算逻辑的通用 MPC 协议，基于混淆电路实现，可将计算逻辑编译成电路，然后混淆执行，并且支持大部分计算逻辑。

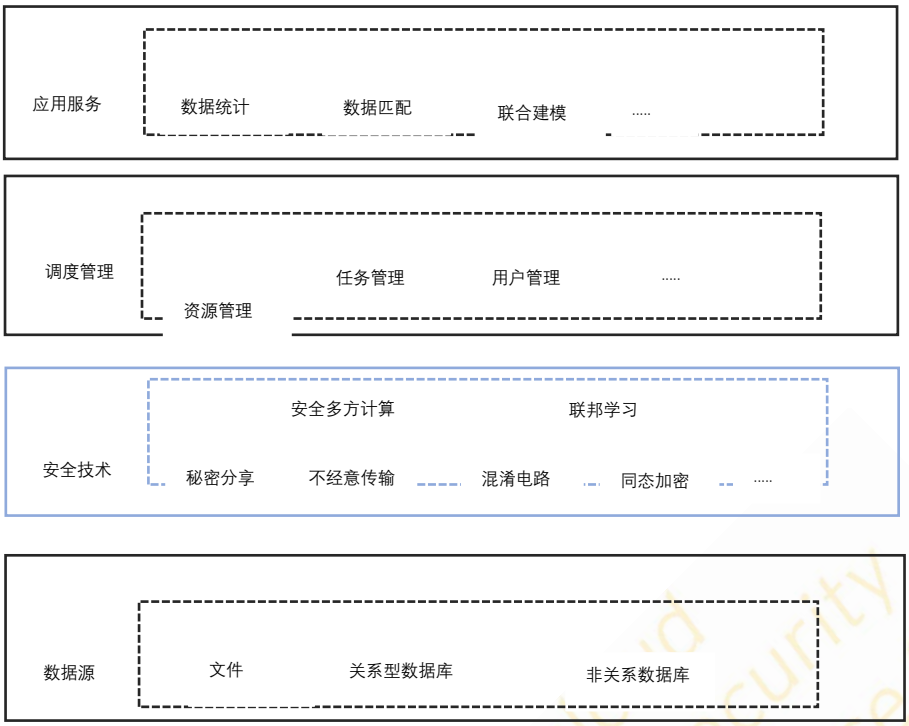

图 3-1 安全多方计算技术体系框架

安全多方计算技术体系

多方安全计算技术体系中，最重要的支撑技术有同态加密、混淆电路、可搜索加密、秘密分享、零知识证明这五类。

同态加密（Homomorphic Encryption，HE)是一种加密函数，对明文先进行加法和乘法运算再加密，即在无需解密的情况下直接对加密数据执行计算，与加密后对密文进行相应的运算，结果是等价的。由于这个良好的性质，人们可以委托第三方处理数据而不泄露信息。根据支持密文运算的程度，同态加密可分为全同态加密和部分同态加密。部分同态加密仅支持有限的密文计算深度，仅支持同种加密运算即同乘或同加。全同态加密可同时满足加同态和乘同态性质，是可以进行任意多次加和乘运算的加密函数。

混淆电路 （Garbled Circuit，GC）又称姚氏电路(Yao’s GC)，是由姚期智教授于 1986 年针对百万富翁问题提出的解决方案。它是安全两方计算协议，参与方在不知道他人数据的前提下，通过使用私有数据共同计算一个用逻辑电路表示的函数，由于任何安全计算函数都可转换成对应布尔电路的形式，相较其他的安全计算方法，具有较高的通用性。

可搜索加密（Searchable Encryption，SE）是一种支持用户在密文上进行关键字查找的密码学原语，即在加密状态下实现搜索功能。

秘密分享(Secret Sharing，SS)的思想是将数据以适当的方式拆分成多个无意义的数，拆分后的每一个数（？）由不同的参与者管理，单个参与者或者少数几个参与者无法恢复原始数据，只有若干个参与者一同协作才能恢复原始数据。通过拆分原始数据，将秘密分散到一群参与者中，能有效地防止系统外敌人的攻击和系统内用户的背叛。基于秘密分享的多方安全计算可支持加减乘除及多项式运算。

零知识证明(Zero—Knowledge Proof)是一种涉及两方或者多方的协议，证实者使验证者确信证实者知道秘密值但不会向验证者泄漏任何有关秘密值的信息，采用交互式零知识证明方法验证访问者的身份。

3.2.2 联邦学习

联邦计算/联邦学习背景

联邦学习（Federated Learning）是一种新兴的人工智能基础技术，在 2016 年由谷歌最先提出，原本用于解决安卓手机终端用户在本地更新模型的问题，其设计目标是在保障大数据交换时的信息安全、保护终端数据和个人数据隐私、保证合法合规的前提下，在多参与方或多计算结点之间开展高效的机器学习。

技术要点

一个典型的联邦学习过程如图 3-2 所示：

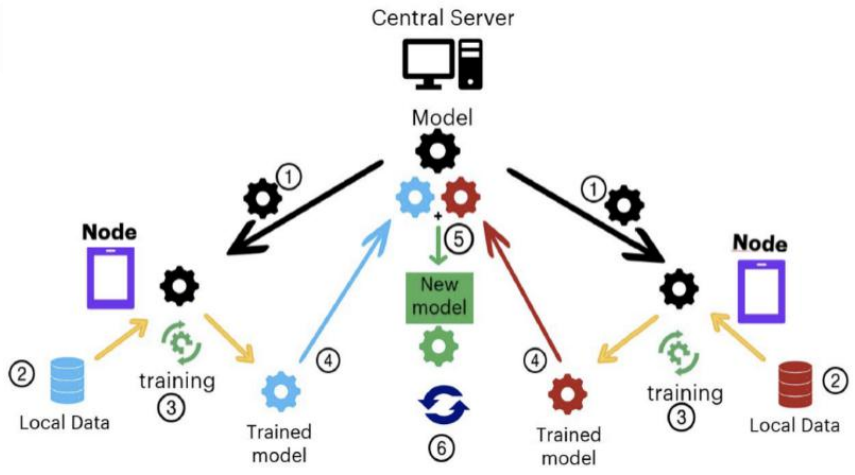

图 3-2:联邦学习过程示意图

1. 中心服务发送最新的模型参数到各个节点(Node)
2. 各节点收集本地数据 （Local Data）
3. 各节点基于最新的模型参数在本地训练（Training）模型
4. 更新模型参数，返回给全局模型(Trained Model)
5. 中心服务汇总各模型的更新并重新训练全局模型，得到新模型(New Model)
6. 重复步骤 1

从数据安全和隐私保护的角度看，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，保证各方数据不出本地节点。同时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三类（见图 3-3），即横向联邦学习、纵向联邦学习和迁移联邦学习。

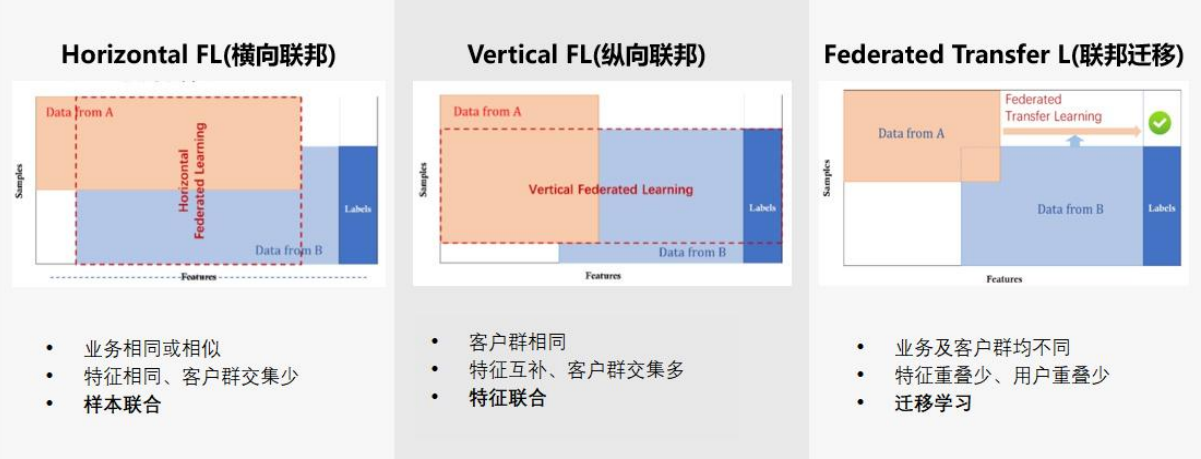

图 3-3 三类联邦学习示意图

横向联邦学习：在两个数据集的用户特征重叠较多，而用户重叠较少的情况下，把数据集按照横向（即用户维度）切分，并取出双方用户特征相同而用户不完全相同的那部分数据进行训练。这种方法叫做横向联邦学习。

纵向联邦学习：在两个数据集的用户重叠较多而用户特征重叠较少的情况下，把数据集按照纵向（即特征维度）切分，并取出双方用户相同而用户特征不完全相同的那部分数据进行训练。这种方法叫做纵向联邦学习。

联邦迁移学习：在两个数据集的用户与用户特征重叠都较少的情况下，不对数据进行切分，而利用迁移学习克服数据或标签不足的情况。这种方法叫做联邦迁移学习。

技术优势

联邦学习过程中原始数据隔离，数据不会泄露到外部，满足机器学习过程中用户隐私保护和数据安全的需求；能够保证模型质量无损，不会出现负迁移，保证联邦模型比割裂的独立模型效果好；参与者地位对等，能够实现公平合作；能够保证参与各方在保持独立性的情况下，进行信息与模型参数的加密交换，并同时获得成长。

由于法规或商业机密等原因，很多行业的数据不能直接聚合用于训练机器学习模型，这些行业有金融、医疗、政务、教育、智慧城市、边缘计算、物联网、区块链以及第 5 代（5G）移动网络等。联邦学习作为能够在满足隐私、安全、合规的前提下，使用分散于多方的数据构建共享和定制化模型的机器学习建模机制，在诸多领域都有广阔的应用前景。

3.2.3 可信计算

可信计算的基本逻辑是通过一个不可篡改和伪造的信任根，和一套可信验证机制建立起来的一条可传递的信任链。利用这条信任链，不仅可以在本地设备内部验证固件和 OS 内核的真实性完整性，即安全启动（Secure Boot）。同时还可以实现对远程设备（如手机、汽车）身份的可信验证。由于可信计算的根基是不可篡改和伪造的信任根，因此可信计算离不开硬件的支持配合。

在 PC 时代，由 AMD、惠普、IBM、英特尔和微软组成的 TCG 组织提出的 TPM 几乎是可信计算的代名词。TPM 是一颗硬件安全芯片，可以用来存储和校验 BIOS/CMOS 的密码，从而在 PC 上承担起信任根的作用，后来我国又在 TPM1.2 的基础上推出了 TCM 芯片。

技术要点

这里以 ARM 的 Trustzone 技术（信任域）为例，介绍一下可信计算在隐私保护领域的作用。

ARM 在 v7 架构中就已经有了 Security Extension 安全扩展，也就是 TrustZone（信任域），其目的也很简单，即在不增加硬件成本的情况下为设备提供一个可信执行环境，以此支持 TEE 技术

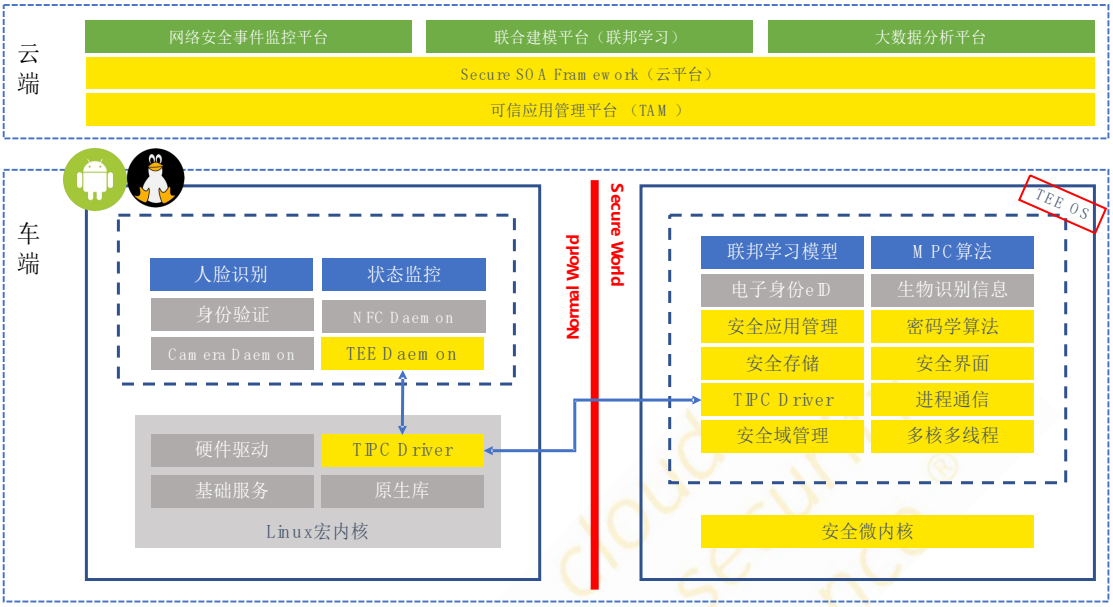

图 3-4 ARM CPU 结构示例

如图 3-4 所示，在同一颗 ARM CPU 上存在着两套操作系统，一套是左侧的 Normal World（正常场景）中运行的 Android 或 Linux，另一套是右侧的 Secure World（受信场景）中运行的 TEE OS，负责信任根管理、验证及权限管理，以及 MPC 算法或 FL 模型等关键业务逻辑的存储和运行。Normal World 与 Secure World 之间使用安全的 API 通信，当 Android 或 Linux 中某些应用程序要运行关键业务逻辑，或访问用户隐私数据时，必须通过这些安全 API 的验证才可以获得访问权限。

值得注意的是，可信执行环境与联邦学习(FL)、安全多方计算(MPC)是融合关系，而非替代关系，从上图也可以看出，可信执行环境可以在保护隐私数据的同时，对保护 FL 模型和 MPC 算法，这样一方面可以防止攻击者通过对算法和模型的研究，逆向分析得到隐私元数据，同时还可以保护商业 IP。

另外在云端架构中，通过 TAM（Trusted Application Management）管理每个终端 TEE 系统，包括 TA（Trusted Application）升级、双向可信验证、安全服务访问、安全数据传输等。利用 TAM 对接终端 TEE 获取处理过的隐私数据后，后台的网络安全事件监控平台、大数据分析平台以及联合建模平台将可以安全合规地分析处理数据。同时这套体系也可以用于管理隐私数据生命周期，如用户希望删除之前注册过的个人数据

时，就可以通过 TAM 精确的识别终端设备并对设备进行不可逆的数据删除，而后台由于并不存储原始的隐私数据，因此也不需要进行相关的删除操作。

3.2.4 同态加密

同态加密（Homomorphic Encryption, HE）是指满足密文同态运算性质的加密算法，即数据经过同态加密之后，对密文进行特定的计算，得到的密文计算结果在进行对应的同态解密后的明文等同于对明文数据直接进行相同的计算，实现数据的“可算不可见”

技术要点

如果一种同态加密算法支持对密文进行任意形式的计算，则称为全同态加密（Fully Homomorphic Encryption, FHE）；如果支持对密文进行部分形式的计算，例如仅支持加法、仅支持乘法或支持有限次加法和乘法，则称其为半同态加密或部分同态加密，英文简称为 SWHE（Somewhat Homomorphic Encryption）或 PHE（Partially Homomorphic Encryption）。一般而言，由于任意计算均可通过加法和乘法构造，若加密算法同时满足加法同态性和乘法同态性，则可称其满足全同态性。

技术优势

同态加密的这个“可算不可见”的特性对于保护信息的安全具有重要意义，利用同态加密技术可以先对多个密文进行计算之后再解密，不必对每一个密文解密而花费高昂的计算代价；利用同态加密技术可以实现无密钥方对密文的计算，密文计算无须经过密钥方，既可以减少通信代价，又可以转移计算任务，由此可平衡各方的计算代价；利用同态加密技术可以实现让解密方只能获知最后的结果，而无法获得每一个密文的消息，从而可以提高信息的安全性。

应用领域

同态加密的概念最初提出用于解决云计算等外包计算中的数据机密性保护问题，防止云计算服务提供商获取敏感明文数据，实现“先计算后解密”等价于传统的“先解密后计算”。随着区块链、隐私计算等新兴领域的发展及其对隐私保护的更高要求，同态加密的应用边界拓展到了更为丰富的领域。

3.2.5 区块链技术

技术要点

区块链（Blockchain）是一种按照时间顺序将若干数据区块相连的、不可篡改、不可伪造、全程留下痕迹、交易可以追溯的分布式共享账本。通过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，是解决多方协作和多方信任问题的有力工具。通过共识机制在参与方之间建立信任基础，实现点对点的价值传递。通过智能合约实现链上数据真实性验证和审计。通过协作机制、激励机制的设计和共识，促进数据的开放共享和价值协作。

技术优势

将区块链技术与隐私计算结合，已成为业内厂商的共识。隐私计算实现了在隐私计算过程中对于输入输入数据的隐私保护，但身份认证、数据来源可信以及计算结果追溯及验证均存在问题，与区块链技术相结合，利用区块链的分布式账本、智能合约等技术可以实现参与计算的原始数据链上存证、计算过程关键步骤的上链存证回溯，确保整个计算过程的可验证性。

早期的区块链通过大规模的复制传递计算信任，但吞吐量有限，隐私和安全性也不完善。将区块链与链下隐私计算方案结合，区块链专注链上业务逻辑的可信执行与数据权属凭证的流通，将密集的数据计算业务放在链下，通过链下隐私计算-可信计算网络进行大规模数据运算和数据价值流通业务，将提高链的吞吐量，保证工作任务的完整性，及隐私数据的机密性，充分发挥“数据可用不可见”，促进数据的价值流通。

通过区块链将各方愿意共享的数据通过文件上链智能合约进行上链存证，存证的数据主要包含文件的哈希、发布者等相关元数据信息，便于使用时在链上对数据进行溯源和交叉验证，进而提升隐私计算的活动监测和监管审计能力；文件通过上链智能合约进行上链，上链后监听合约的执行，形成记录分享数据元信息的数据市场；数据使用者通过浏览检索数据元数据市场找到目标元数据，通过授权使用智能合约进行申请授权使用，在申请时使用分布式身份，进行身份校验。通过对参与计算的各方进行数字身份管理，链上记录参与者行为，提高恶意参与者的作恶成本。数据使用者再获得使用授权后，在隐私计算平台上进行相关隐私计算的操作，通过平台提供的隐私计算方法，获得隐私计算结果。

将密集的数据计算放在链下，保持主链的性能，通过预言机打通链上链下数据的连接。链上轻量保存上链数据的相关元数据信息，如数据的产生者、文件哈希、文件相关元数据，链下进行数据的密集计算。不同机构接入区块链网络后，将参与多方计算的敏感数据集哈希上链，同时发布数据集的相关元数据信息(如数据内容，数据集格式和数据价值等)；另外数据集将通过可信信道加密后保存到可信计算服务器上；其次，数据使用方加入区块链网络后，通过区块链浏览器查看链上发布的相关元数据信息，选择目标数据，申请授权使用；

在整个隐私计算过程中数据“可用不可见”，且全流程审计上链，方便日后追溯审查。

3.3 其它隐私增强技术

3.3.1 差分隐私

差分隐私（Differential Privacy，DP)是 Dwork 等人在 2006 年针对数据库隐私问题提出的一种严格的、可量化的隐私定义和技术。DP 是密码学中的一种手段，旨在提供一种当从统计数据库查询时，最大化数据查询的准确性，同时最大限度减少识别记录的机会。

技术要点

差分隐私技术的基本原理是：在计算结果中添加噪声（如适用于数值型输出的拉普拉斯噪声和适用于非数值型输出的指数噪声），使得修改数据集中单条记录不会对统计结果造成显著的影响，从而保证攻击者在拥有背景知识的情况下也无法推断出该记录对应的敏感信息。差分隐私具有两个重要的优点：一是提出背景知识无关的隐私保护模型，实现攻击者背景知识最大化假设；二是为隐私保护水平提供严格定义和量化评估方法。

技术优势

（1） 严格分离

差分隐私在实际运行过程中，会严格限制攻击者获得的背景内容，也就是说，假设攻击者知道原数据中的大部分信息，哪怕只是不知道一条信息，即使在这样理想化的攻击形态下，利用差分隐私，依然可以保证整个数据的安全，这在过去是不敢想象的。

（2） 开发效率高

差分隐私拥有严谨的统计学模型，极大地方便了数学工具的使用以及定量分析和证明。同时，由于利用差分隐私技术构建模型相对比较便捷，在实际开发中，也能获得更高的开发效率，这也是差分隐私能在离开实验室后快速得到运用的原因之一。

应用领域

差分隐私可以被应用于推荐系统、网络踪迹分析、运输信息保护、搜索日志保护等领域。一个比较常见的应用领域便是对于用户行为的统计，特别在一些拥有大量用户日常数据的企业中，如果希望保证用户隐私行为的安全性，则需要采用差分隐私技术确保攻击者无法通过观察计算结果而获取准确的个体信息。

但差分隐私也不是万能的，虽然相比于传统技术，在实现同样复杂度和效果的情况下，差分隐私的效率会比较高，但由于过于强调背景知识的假设，需要在查询结果中加入大量的随机化，导致数据的可用性急剧下降。特别对于那些复杂的查询，有时候随机化结果几乎掩盖了真实结果。

3.3.2 动态数据屏蔽

动态数据屏蔽是利用功能性虚构数据(如字符或其他数据)代机密数据，从而实现隐藏敏感数据的技术。数据屏蔽的主要目的是在企业与第三方共享数据的情况下保护公司业务敏感或机密的数据，以及员工和客户的个人隐私信息。

技术要点

动态数据屏蔽通过限制用户访问权限，从而防止敏感数据暴露给未授权用户。在配置动态数据屏蔽过程中，不需要第二个数据源动态存储被屏蔽的数据，原始的敏感数据依旧被保存在原始数据库中，根据用户不同的访问权限，数据字段被系统授予相应的权限。因此，当授权用户访问数据时，永远只能查看其权限之内的真实数据，其他数据内容会被实时打乱屏蔽。动态数据屏蔽主要用于基于角色安全的数据库或应用程序，为了防止屏蔽数据被回写到数据库的复杂性，动态数据屏蔽只能应用于只读内容，如报告或客户服务查询功能等。

动态数据屏蔽可以通过不同的方式实现，主要实现方式包括数据库和 web 代理。当用户想要查询存储敏感数据的数据库时，所有用户的 SQL 查询指令都通过数据库代

理检查并传输到用户想要访问的数据库对象，SQL 查询指令会在传到数据库之前被代理根据数据库中权限配置范围修改，从而最终将屏蔽的数据返回给访问用户。

动态数据屏蔽技术被视为应用程序与数据库之间的保护盾。通过拦截发送到数据库的请求并依据特定的规则对请求的数据进行屏蔽，加密，隐藏等处理，确保数据库中敏感数据的安全。该技术作用于数据库协议层面，应用该技术时无需变更数据库，也不必修改应用程序源代码。

技术优势

（1）便于使用和变更管理

该技术仅需要编写一次屏蔽策略，便可应用于数据库中的数千列数据和对象。同时，屏蔽策略内容也易于变更，无需对应用程序源代码进行变更

（2）高效的数据管理和授权治理

屏蔽策略易于对管理进行职责分离，并且支持集中式或分散式管理模型。策略的决定权掌握在安全隐私的管理者手中，而非数据库对象的所有者。安全隐私管理者可以根据相关要求管理屏蔽。动态数据屏蔽技术还可根据个人用户的角色或自定义权限限制数据访问，支持由安全隐私管理者实施的数据治理，并可以防止具有特权的用户查看不必要的数据。

（3）提高保护个人和敏感数据成本效益

动态数据屏蔽可以有效地保护个人数据和敏感数据，从而提升应对内部和外部数据泄露威胁的能力。动态数据屏蔽易于实施即时的敏感数据匿名化，无需投入过多人力及时间成本。

应用场景

动态数据屏蔽可用于任何需要进行数据共享和检索的场景和领域当中。当组织需要与未获得敏感权限的用户或系统共享包含敏感信息的数据时，都可以利用该技术保护数据的安全隐私。该技术的使用需求，多数来源于公司政策和隐私相关法规要求（如 GDPR 或 HIPAA），并且普遍存在于金融机构、医疗组织、企业、政府部门等大型组织当中。

3.3.3 云访问安全代理

安全产品伴随着信息化发展而演进。过去，企业能够掌控基础设施、应用软件、数据存储，安全也是围绕这些建立的，但是随着 SaaS 来临，云服务的大量使用，让企业安全负责人无法准确定位各种 ShadowIT，用户、设备、数据在云中的活动情况，也就无法精准控制用户使用、访问云应用的权限。为解决如何保障数据在第三方系统中的安全问题，云访问安全代理（Cloud Access Security Broker，CASB）应运而生。

技术要点

云访问安全代理是位于云服务消费者和云服务提供商（CSP）之间的内部或基于云的策略实施点，用于监视与云相关的活动，应用的安全性、合规性和治理规则与基于云的资源的使用相关。CASB 允许企业将应用于本地基础设施的相同类型的控制扩展到云中，并可以组合不同类型的策略实施，例如：

- 用户凭据身份验证，因此仅向批准的云服务提供访问权限
- 通过加密、令牌化或其他方式保护数据，因此敏感信息不会暴露在云服务或 CSP 中
- 云服务活动监视，以便记录、标记和分析用户和实体的行为，以了解异常使用模式或受损的凭据
- 数据防丢失（DLP），因此敏感信息不能离开组织的网络，以及恶意软件检测和修复，使敏感信息无法离开组织的网络

因此，CASB 的目的是提高企业安全、安全地利用云服务的能力。CASB 可以被认为是一个“安全节点”，通过它可以控制对企业云服务的访问。作为企业安全基础设施的一个组件，它补充而不是取代企业和 web 应用程序防火墙、IDaaS（身份即服务）和安全 web 网关（SWG）等技术。

技术优势

CASB 最初专注于发现影子 IT，即 IT 部门许可范围之外的个人或业务单位使用的未知服务，但企业随后意识到，解决此问题的方法更多地指向受控支持，而不是删除这些服务，CASB 开始提供跨越可见性、数据安全、威胁防护和合规这四大支柱的功能集。

- 可见性：可以让企业负责人知道所有员工在网络中使用云服务是否安全。利用 CASB 可以通过整合视图展示企业云服务的资源使用情况，了解员工通过什么设备在哪

里访问了云服务，执行了什么操作，使用了什么数据；包括对云服务进行风险评估（计分、打分）。

● 数据安全：以 DLP 为核心，辅以加密、令牌化、DCAP、EDRM 能力。与精细化的访问控制相结合，防止出乎预期的数据使用发生，比如企业敏感数据共享到外部，或将数据从企业资源转移到到个人资源，亦或在不恰当的时期发布了不恰当的言论，也支持云上数据的发现、分类、分级，用户对敏感数据的访问行为监测，以及特权账号的权限扩大等等。

● 威胁防护：通过自适应访问控制能力防止出乎预期的设备、用户，也包括出乎预期的版本的应用与受保护的云服务交互，能识别和响应恶意会话或不想其发生的会话，响应动作包含：允许、限制、触发多种级别告警、立即发起额外的认证要求、结束会话等等。

● 合规：通过合规模板检测、审计企业使用云服务过程中不合规的场景，同时通过自身的安全功能帮助企业满足合规要求。

应用场景

CASB 可以作为代理和或 API 代理部署。

在代理模式下部署的 CASB 通常侧重于安全性，并且可能配置为数据访问路径中的反向或正向代理，在云服务消费者和 CSP 之间。反向代理 CASB 不需要在端点上安装代理，因此可以通过避免配置更改、证书安装等更好地为非托管（如 BYOD）设备工作。

在 API 模式下部署的 CASB 专注于通过 SaaS（以及越来越多的 IaaS 和 PaaS）服务提供的 API 管理 SaaS 应用程序，包括静态数据检查、日志遥测、策略控制和其他管理功能。

3.3.4 格式保留加密

格式保留加密（Format-preserving Encryption，FPE）的发明是为了应对互联网早期诞生的大型数据库的加密需求。多数互联网早期的大型商用应用系统都是基于数据库的应用系统，如金融、社保、电子政务、电子商务等，如果数据库中存储的大量用户敏感信息(如银行卡号、身份证号、用户名和密码等)被窃，将对公司造成致命的破坏。

传统的加密方式如 ciphertext 会扩展数据，并且改变数据长度和类型。

对于大量存放数据并且有加密需求的早期数据库而言，应用传统加密方式需要改变数据库结构，造成数据库维护成本大大提高，格式保留加密技术提供了在不改动现有软件系统的代码和现有数据库结构的同时提高数据库的安全性的解决方案：它简化了加密流程，使加密过后的数据与原始数据保持长度和结构一致。

技术要点

2002 年，Black 和 Rogaway 提出了三种 FPE 构建方法：Prefix，Cycle-walking 和 Generalized-Feistel.

1）Prefix 方法

一种创建格式保留加密的简单方法是为每一位整数{0,….,N-1}分配一个伪随机权重，然后根据权重对其排列。加密原理是为每个整数分配一个分组密码。Black 和 Rogaway 把这个方法叫做“Prefix Cipher”并且证明了它和传统的分组加密方式一样可靠。

但是，此种方法仅对少量的整数值有用，面对大量数据，过于繁琐的加密过程会让实际操作不可行。

2）Cycle-walking 方法

使用现有的分组密码 (AES 或 3DES 等)对输入值循环处理，直至加密过后的数据长度输出值落在应用程序/数据库可接受的输出范围内。

3）Generalized-Feistel 方法

使用 Feistel 网络制作格式保留加密技术算法。Feistel 网络需要使用每轮子密钥的伪随机值，而 AES 算法则可以用作输出这些伪随机值。完成此操作后，如果使用足够的轮数，则可以产生的良好运行的 Feistel 结构。使用 AES 和 Feistel 网络实现 FPE 算法的一种方法是使用所需的 AES 输出比特数，使 Feistel 网络的左半部分或右半部分长度相等。例如，如果需要 24 位值作为子键，则可以使用 AES 输出的最低 24 位作为该值。

技术优势

对受保护数据的分析：保存格式的受保护数据元素，如信用卡号码、身份证信息等，仍然可以用作索引键进行跨数据库的统计研究。使用格式保留加密技术的情况下，对算法的输入将创建相同格式的密文。这种加密保留了数据的长度和结构，从而使研究者能够从受保护的数据集收集有价值的信息。通过使用格式保留加密技术实现

的安全分析的另一个好处是，在不损害安全性和隐私的情况下，扩展了分析师对数据的访问的范畴，以及数据集的潜在货币化。

跨应用程序数据流保存：格式保留加密技术允许受保护的数据在应用程序之间流动，而不需要更改这些应用程序以接受受保护的数据，这对于传统加密方法来说是不可行的，因为应用程序需要特定长度和格式的数据。

使用受保护的数据而不需要解密：格式保留加密技术允许使用处于受保护状态的数据。例如，信用卡号的“前 6 位”数字用于收费路由，而手机号的“后 4 位”用于客户验证。如果这些都是不需要的，那么数据流中的许多应用程序将不需要访问整个数据元素，就可以执行业务功能，且不需要对应用程序进行任何更改，也不需要执行任何解密。这样的部分加密可以方便一些功能，如排序和某些搜索用例，如“以 XXX 开始”，“以 XXX 结束”等，而不需要对加密的数据进行任何解密。

测试数据管理——格式保留加密技术可以 被用于以其不可逆变量的形式混淆/删除生产数据来填充测试数据库，实现基于生产数量、可变性等的真实测试条件。

应用场景

● 保护产生于互联网早期的设备和软件。这些设备或软件无法采用传统的加密方式，因为传统的加密方式会改变数据的长度或格式。尤其对于数据库类的应用对于数据格式的变化尤其敏感，这时就需要格式保留加密技术加密这些数据。

● 保护敏感数据，如银行卡信息，银行账户信息，社保卡号，和各种个人信息。这些信息主要存储在零售，医保，金融类数据库和应用中。使用格式保留加密技术加密过的数据仍然可以被作为指数用作统计研究，意味着对加密数据的研究可以在数据加密状态下进行，确保了研究过程中数据的安全性。

3.3.5 匿名化/假名化技术

匿名化技术（Anonymization）可以实现个人信息记录的匿名，数据匿名化是模糊化信息的过程，使这些信息无法用于识别个人。匿名化可减少意外泄露个人可识别信息的风险，并且，如果确实发生数据泄露，则被窃取的信息将对攻击者毫无用处。从合规性来看，匿名数据被认为是非个人数据，因为它无法识别个人。

匿名化的最终目标是消除个人可识别信息暴露个人的机会，同时保留信息对企业的价值。匿名化数据与破坏数据使其无法提供有意义的业务洞察力之间只存在细微的界限。

假名化技术是指用生成的新字符，即假名，取代原来的直接标识符，使得在不借助额外信息情况下无法识别出个人信息主体。

与匿名化措施相区别，假名化技术也是保障措施的一种，与匿名化不同的是假名化具有再识别的可能性，且 GDPR 规定假名化之后的信息仍属于可识别信息。就数据利用率而言，由于匿名化数据不属于可识别数据，因此利用率高于假名化数据。

技术要点

目前主流的匿名化方法主要是通过泛化(Generalization)和抑制(Suppression)操作实现。该技术不同于一般的扭曲、扰乱和随机化等方法，能保持发布前后数据的真实性和一致性。

泛化通常将准标识符（英文全称，QID）的属性用更抽象、概括的值或区间代替。

泛化分为全局泛化和局部泛化两类。全局泛化也称为域泛化，是将 QID 属性值从底层开始同时向上泛化，一层一层泛化，直至满足隐私保护要求时才停止。局部泛化也称为值泛化，是指将 QID 属性值从底层向上泛化，但可以泛化到不同层次。单元泛化及多维泛化是典型的局部泛化。单元泛化只对某个属性的一部分值泛化。局部泛化可以对多个属性的值同时泛化。

抑制又称为隐藏，即抑制（隐藏）某些数据。具体的实现方法是将 QID 属性值从数据集中直接删除或者用如“*”等不确定的值代替原来的属性值。采取这样的方式可以直接减少需要进行泛化的数据，从而降低泛化所带来的数据损失，保证相关统计特性达到相对比较好的匿名效果，保证数据在发布前后的一致性、真实性。抑制可分为 3 种方式：记录抑制、值抑制及单元抑制。其中，记录抑制是指将数据表中的某条记录进行抑制处理；值抑制是指将数据表中某个属性的值进行抑制处理；而单元抑制是指将表中某个属性的部分值进行抑制处理。

常用的假名生成技术有如下几种：1）带密钥加密（Encryption with secret key）；2）哈希函数（Hash Function）；3）带密钥的哈希函数（Keyed-hash function with stored key）；4）令牌化（Tokenization）等。带密钥的哈希函数其实是加盐（Add salt）哈希的一种情形。所谓加盐，是指一种增强哈希函数安全性的常用技术手段，即在进行哈希加密前在原标识符的特定位置（通常是头部或者尾部）

加上一串字符（盐值，Salt value）。对于盐值的选择，通常有固定字符串或一次性随机字符串等。带密钥的哈希函数指的就是通过在标识符中加入一串密钥（Key）（密钥单独保密存储），这里的密钥就是盐值，比如对标识符手机号码进行加加盐哈希处理，即对“Key+手机号码”进行哈希处理得到假名。这样在攻击者不知道盐值的情况下，可以极大的提升彩虹表破解的难度。

概念辨析

需辨别的是，匿名化(Anonymization)、假名化(Pseudonymization)两个概念有些联系，但不尽相同，却常常被混为一谈。

假名化(Pseudonymization)：将身份属性的值重新命名，如将数据库的名字属性值通过一个姓名表映射，通常这个过程是可逆。该方法可以基本完好保存个人数据的属性，但重识别风险非常高。一般需要通过法规、协议等约束不合规行为保证隐私的安全性。

技术优势

匿名化的泛化技术的优点是不引入错误数据，方法简单，泛化后的数据适用性强，对数据的使用不需要很强的专业知识。其缺点是预定义泛化树没有统一标准，信息损失大，对不同类型数据的信息损失度量标准不同。

抑制技术的优点表现为泛化技术使用前可减少信息损失，缺点是不适合复杂场景，发布数据量太少，会降低数据的真实性和可用性。

3.4 隐私科技发展路径

隐私科技也是当前数据保护领域各界关注的热点。在学术界，近年来有关隐私科技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技术的专题研讨会，每年出现的与隐私科技相关的学术论文也呈指数增长（平均每年都超过一千篇）。产业界包括许多大型互联网公司、金融机构、隐私科技公司等企业越来越开始关注隐私科技技术和产品。各企业单位都争相投入隐私科技研发和产品化工作，有多家公司都推出了自己的隐私科技平台产品，并开始进行隐私科技在金融、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科技技术的发展，一方面希望能够通过隐私科技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，促进隐私科技技术及产业健康发展，推动合法、合规的数据协同应用。

隐私科技的发展与行业对于隐私保护的需求趋势也是密不可分的。数据最小化、数据分级分类及数据匿名化都是近一两年来随着相关法律法规出台所带来的热点合规需求，如何从技术层面满足这些需求也促进了隐私科技的发展。

3.4.1 数据最小化面对的风险控制和合规满足需求

数据最小化原则几乎是世界各国个人信息保护立法中共通的原则。数据最小化原则可悲理解为要求企业和公共机构收集和使用个人信息时，以实现产品和服务目的为标准，在功能可实现的前提下在最小范围内收集数据。无论是国外的合规类要求如GDPR 还是国内的法律如《个人信息保护法》、《数据安全法》、《网络安全法》、《民法典》等，都把个人数据最小化原则作为个人信息保护的基本的原则之一。数据最小化原则之所以成为大多数国家个人信息保护立法的基本原则之一，其主要原因在于个人信息处理者存在扩大收集个人信息的倾向，在大数据时代尤为突出，需从法律上加以限制，使其保持克制。使数据收集行为受到数据处理目的的限制。隐私科技相关技术可以减少不必要的数据披露和传播，一定程度上体现了数据最小化的思想。

数据分类分级面对的风险控制和合规满足需求

数据分级分类是开展数据安全治理的起始点，也是数据精细化管理控制的重要手段，在数据立法中也被反复强调。引入数据分类分级这一基础性数据安全管理方法，综合考虑数据属性、特点、数量、质量、敏感度等因素，对数据资源进行分类分级，梳理出非敏感、低风险等级、权属相对明确的数据资源，以要素形式优先进入数据交易市场，同时明确在市场交易过程中应配备的安全保护措施，可以在最大限度释放数据价值的同时，又兼顾数据安全和个人隐私的保护。

数据匿名化面对的风险控制和合规满足需求

个人信息概念的界定是个人信息保护立法的核心问题和逻辑起点，关系到法律保护对象的范围。一旦个人信息进行了匿名化处理，就不再具有个人信息属性。我国的《民法典》明确了个人信息是识别特定自然人的各种信息，而《个人信息保护法》中规定的个人信息为与已识别或可识别的自然人有关。过窄的个人信息定义无法实现对个人信息的充分保护，过宽的定义又容易阻碍数据要素的流通和促进数据要素市场培育。

根据《个人信息保护法》，个人信息不包括匿名化处理后的信息。匿名化是指个人信息经过处理无法识别的特定自然人且不能复原的过程。

3.4.2 隐私科技产业发展现状

信息技术、移动通信技术等的紧密结合与快速发展，以及智能终端软硬件的不断升级与换代，促进了互联网、移动互联网、云计算、大数据、物联网等方面的技术发展，同时催生了以 Amazon/淘宝为代表的电商、以 Facebook/微信为代表的社交、以 Uber/滴滴为代表的出行等各种新型服务模式，大幅度提升了人们的生活品质。

随着信息技术的快速发展和个性化服务的不断演进，海量用户个人信息数据的频繁跨境、跨系统、跨生态圈交互已成为常态，加剧了隐私信息在不同信息系统中有意/无意留存，随之而来的隐私信息保护短板效应、隐私侵犯追踪溯源难等问题越来越严重，现有的隐私保护方案已不能提供体系化的保护。世界各国纷纷颁布具体措施保护隐私信息。

我国在隐私科技方面，首先在政策层面提出了利用隐私计算解决相关问题，例如，2019 年 8 月，人民银行《金融科技发展规划 2019-2021》、工信部《工业大数据发展指导意见（征求意见稿)》、发改委《关于加快构建中国一体化大数据中心协同创新体系的指导》等，都在政策层面上提出了对隐私科技产业和技术发展的指导意见。

新技术、新服务模式的产生与快速发展促使海量用户个人信息跨系统、跨生态圈甚至跨境交互成为常态，用户个人信息在采集、存储、处理、发布（含交换）、销毁等全生命周期各个环节中不可避免地会在不同信息系统中留存，导致信息的所有权、管理权与使用权分离，严重威胁了用户的知情权、删除权/被遗忘权、延伸授权。另一方面，缺少有效的监测技术支撑，导致隐私侵犯溯源取证困难。

据智研咨询统计，自 2012 年以来，隐私计算初创企业数量呈增长趋势，2020 年，隐私计算初创企业 71 家，较比上年增加 18 家，同比增长 33.96%。隐私计算企业背景多样:有互联网龙头企业、网络安全及大数据公司、初创型科技企业、行业数据高度聚合企业。

随着资本关注度的提升，隐私计算产业融资事件基本呈现逐年递增，至 2020 年底，隐私计算融资事件数量 10 起，较 2019 年增加 3 起。

随着隐私科技、隐私计算的不断发展和科技企业的研究投入的不断加大。越来越多的隐私科技专利技术不断涌现。截至 2021 年 3 月 19 日，在我国目前有三家科技企业专利数量已经达到了 200 件以上，分别是蚂蚁集团、阿里巴巴、中国平安。蚂蚁集团全球隐私技术专利数量累计 740 件，排名第一；第二是阿里巴巴，隐私技术专利数 299 件；第三是中国平安，隐私技术专利数量 282 件，其中微众银行、腾讯科技、华为、国家电网挤进前十。在这近 26 年中，我国的隐私计算相关专利申请量的增长发生了根本变化。在 1995 年我国隐私计算专利申请量只有 1 项；2005 年我国隐私计算专利申请量 14 项，10 年间增加 13 项；2015 年我国隐私计算专利申请量 215 项，较 2005 年增加 201 项；至 2020 年隐私计算专利申请量达到 1535 项。

3.5 中国隐私科技数据安全合规与保护现状

为应对全球隐私保护及数据安全合规要求，我国的隐私保护遵循者必须考虑实施完善的内部隐私及数据安全内部控制体系。面对愈加复杂的外部合规监管要求以及用户诉求，企业逐步意识到构建体系化的隐私及数据安全内部控制框架并确保持续有效执行内部控制体系的重要性，并相信这是提升自身产品在海外市场竞争力的重要手段。为了使充分且持续有效的安全防控能力及前瞻且全面高效的安全合规能力成为服务商企业的核心竞争力，企业必须从以下几个方面向客户、向市场、向服务的各个利益相关方充分展示和证明自己的隐私及数据安全保护的内部控制体系和能力：

图 3-5 我国隐私及数据安全保护内部控制体系情况

3.5.1 在内部隐私保护政策和组织架构层面

在企业内部隐私保护政策和组织架构层面，企业关注自身隐私合规政策是否清晰定义，并包含一系列应遵循的隐私保护管理要求，如数据保护官的职责、保留数据处理活动记录、定义数据主体可以行使的权益、企业应遵循的告知义务等内容。隐私合规政策在内部发布前须通过公司 DPO 或相关责任部门（如隐私保护办公室、法务部门等）的审阅，确保政策内容符合业务及业务所在国家和地区的法律及监管要求。为保障企业内部隐私合规政策的贯彻落实，企业会构建包含制度发布、隐私培训、宣导及定期考核等在内的多种内部沟通渠道与方式，确保员工充分知晓自身须遵守的隐私保护控制要求并持续有效执行相关控制流程。

3.5.2 在隐私保护风险管理方面

企业在隐私保护风险管理方面须具备一套隐私风险识别和评估机制，对个人信息安全和隐私保护相关的风险重要性等级进行定义，并明确在风险评估过程中应考虑的全部要求。此外，企业会评估自身可以承受的风险，并基于风险承受能力的评估结果明确在遇到不同级别风险时应采取的决策及响应处理流程。同时，须建立责任人机制，确保当风险识别后，由指定责任人负责跟进处理，确保风险得以被及时响应。

3.5.3 在隐私设计管理方面

在隐私涉及管理方面，企业须具备系统化的隐私需求评估机制，确保在产品服务需求设计与评估阶段将隐私合规要求纳入评审范围，并在产品功能正式上线前进行隐私合规评审，确保产品功能可以符合公司对于隐私保护的要求。同时企业会构建和使用系统化的集中式管理平台，对设计阶段识别的数据处理场景及数据处理、流转记录进行保存并执行事后定期审计工作，确保这些产品设计阶段的隐私要求得以实现。对于建立的隐私需求评估机制及流程要求，公司应同步开展多样化的培训宣导活动，确保员工能够深入贯彻隐私设计的理念。

3.5.4 在隐私数据处理合法性评估层面

在隐私数据处理合法性评估层面，企业会建立数据处理合法性的评估标准并在产品需求评审过程中设置强制的审核卡口，确保涉及收集、使用或披露个人数据的需求得以及时评估，以符合隐私合规政策。企业会确保针对涉及个性化推荐或营销的需求，在消息发送前获取数据主体的独立确认，同时给予数据主体拒绝接收此类消息的权利。企业会对于数据处理合法性的评估标准须经过法务团队或隐私保护专家的评估确认。

3.5.5 在数据主体权益响应处理层面

在数据主体权益响应处理层面，企业会建立多样化的请求接收方式、响应处理流程及响应时限要求，确保数据主体可以随时行使自己的权益（如访问权、纠正权、拒绝权、限制处理权及可携带权等），并可以在指定响应期间内予以满足。为实现处理来源于真实数据主体的请求，企业会建立强有效的身份校验机制，准确识别数据主体的身份。为确保可以及时实现数据主体所提请求，企业会部署一系列的技术手段予以支持，如数据打标、数据删除或匿名化、限制数据处理等。此外，为保障数据主体请求的响应处理过程可被追溯，企业会构建信息系统以记录自接受请求、响应处理、至反馈的全过程。

3.5.6 在合作方管理方面

在合作方管理方面，企业会建立并实施合作方隐私管理及准入审核要求，并通过签订数据处理协议等方式明确企业与合作方在隐私数据管理方面各自应承担的责任和义务。对于现有涉及隐私数据交互的合作方，企业会建立系统化的数据流转及处理管理机制，确保准确记录现有合作方涉及数据收集的类型、处理目的、存储位置、已采取的安全保护措施及是否涉及数据跨境等维度的信息，，并对合作方未按照数据处理协议的行为进行识别。

3.5.7 在跨境数据传输管理方面

在跨境数据传输管理方面，企业会具备强制性的系统审核卡口，确保对于涉及数据跨境的处理需求可以被及时识别并对其是否符合受影响两地现有的跨境数据保护要求予以确认。企业会同时构建技术性的监测手段，实现对异常数据跨境传输行为的识别并采取措施予以响应处理。此外，企业会建立定期评审措施，针对自身业务模式和

业务形态的变化及时调整数据处理协议的条款要求，确保企业在作为数据处理者或数据控制者参与数据跨境传输时采取的管理措施可以持续满足相关隐私法案的要求。

3.5.8 在数据处理安全性及合规性方面

在数据处理安全性及合规性方面，企业会从信息安全风险的角度出发，识别所提供的产品在处理个人信息的过程中产生的涉及数据处理安全性及合规性的潜在安全风险。对于识别出的风险，企业会结合自身的风险承受能力、风险重要性水平及应遵循的法律法规要求制定数据处理策略及详细的安全应对措施。

3.5.9 在隐私数据泄露事件响应处理方面

在隐私数据泄露事件响应处理方面，企业会具备体系化的隐私数据泄露事件的应对机制，包括对相关服务合作伙伴的管理要求。

3.5.10 在隐私审计监督方面

在隐私审计监督方面，企业会建立完善的隐私保护审计机制，明确审计范围的评定标准，确定审计过程中所识别的问题的重要性水平、问题责任人及问题响应跟进处理机制。企业会至少每年执行一次隐私保护审计，对隐私保护内部控制体系的运行有效性进行检查，有条件的情况下可以构建信息系统，统筹管理审计过程中执行的审计程序、获取的证据资料及相关底稿。

目前，我国企业可以通过建立并运行完善上述有效的隐私保护及数据安全内部控制体系，更好地实现隐私及数据安全保护体系。通过增强企业内部运营及安全人员的管理能力，加强多方合作，可以实现多元化的监督管控，保障企业服务安全合规能力的持续稳定输出。

3.6 产业环境概述

从“十三五”期间的云计算、大数据、人工智能，到近年快 速发展的区块链、物联网、量子计算，各界对数据安全性和隐私 性的重视提高到了前所未有的高度，这都对数据安全及数据隐私保护都提出了更多更高的要求，隐私计算前景与市场潜力巨

大。 但产业目前处于初期探索阶段，从技术、企业主体、行业应用到市场模式仍有较大发展空间。Gartner 将隐私计算作为 2021 年重要战略科技趋势。隐私计算已成为近两年的热点，并将在接下来的几年持续保持热度。现阶段，产业已涌现出一批创业型企业，该领域也成为投融资机构的关注焦点之一。此外，随着创新成果的转化，相关技术专利申请与标准化建设也在持续推进。

3.6.1 政策支持

我国以政策手段促进技术创新发展，利用规划指明发展方向， 防止监管遏制科技进步。在数字经济迅速发展的背景下，隐私计算技术的关键作用正在逐渐显现，发展规划等各项相关推进政策也将不断向行业化、地方化方向细分发展，自 2019 年起多行业及各地方规划提出研究利用隐私计算解决相关问题（如下表）。从行业角度看，近两年隐私计算政策侧重于金融科技、工业大数据、 区块链三个领域。在地方层面，2020 年，湖南、山东、上海等地区都对隐私计算做出了更为细致的规划。可以肯定的是，在未来 5 年时间，我国关于隐私计算的相关政策和立法的落实、执行及深化将进一步推动行业发展需求。除了个人数据及隐私保护，国家在数据安全、数据要素流通等方面的一系列举措，更将从数据监管和国家利益的高度确立隐私计算的巨大价值和重要地位。

表 3-1: 隐私计算政策相关发文【11】

政策名称	地区	发布时间
《金融科技（fintech)发展规划（2019-2021 年）》	国家（人民银行）	2019.8.26
《工业大数据发展指导意见（征求意见稿）》	国家（工信部）	2019.9.4
《关于加快构建全国一体化大数据中心协同创新体系的指导意见》	国家（国家发改委）	2020.12.23
《湖南省区块链发展总体规划（2020-2025 年）》	湖南省	2020.10.27
《上海公共数据开放管理办法草案》	上海市	2019.4.29
《成都市金融科技发展规划（2020-2022 年》	四川成都市	2020.5.9

3.6.2 金融保障

自 2016 年起，隐私计算领域融资 28 起。随着资本关注度的 提升，融资事件基本呈现逐年递增的情况。从融资事件来看，微众银行及矩阵元于 2016 年最早获得 A 轮投资。微众银行、翼方健数、数篷科技、趣链科技、DataExa、星环科技、京东数

科、同盾科技融资金额超 1 亿人民币。其中融资金额最高达到 17.8 亿人民 币。从融资轮次来看，92%的融资事件处于 B 轮及 B 轮之前，其 中天使轮及 A 轮占比达到 80%。 隐私计算目前处于起步阶段，大多数企业组建小规模团队在内部试验或初步形成产品，尚未形成成熟的商业模式，因此融资基本集中在 B 轮之前。对于投资机构而言，现阶段重点关注隐私计算企业的核心技术竞争力及团队优势。

3.6.3 标准建设

目前，金标委、信安标委、中国通信标准化协会、IEEE、ITU-T、ISO 等各个标准化组织，针对数据共享流通应用的技术如联邦学习、 多方安全计算、可信执行环境等各个技术领域，已经开展了相关领域的技术标准研制工作。国际标准方面，ISO/IEC JTC1 的 SC27 信息安全分技术委员会在隐私框架方面已经制定了诸多标准，如隐私框架与架构（ISO/IEC 29100、ISO/IEC 29101）等，该部分标准是隐私计算的隐私保护基础标准。此外，IEEE 也于 2019 年开始了多方安全计算、共享学习的国际标准制定工作，并陆续开展了 基于可信执行环境的安全计算、联邦学习技术框架与应用指南等标准制定工作。国际电信联盟标准化部 ITU-T 分别在 SG16 和 SG17 开始制定共享学习和多方安全计算的国际标准。国家标准方面，目前针对多方安全计算、可信执行环境、联邦学习已有 3 项测试 标准，此外另有联邦学习参考框架、基础架构与应用两项团体标 准。 标准化工作需要科学的顶层设计，隐私计算等软件和信息技术类标准通常从基础支撑、技术方向、产品工具、规范管理及行业应用等方面构建高质量、体系化的标准体系。总体而言，我国隐私计算领域现有标准从隐私计算技术类型及标准内容方面缺乏大量针对性的专用标准，尚未形成具有指导作用的标准体系，需要通过标准化的途径规范认知，促成行业共识，推进隐私计算产业健康发展。2021 年 3 月，工信部发布的年度标准工作要点，明确提出将围绕包括网络和数据安全在内的安全生产领域编制强制性国家标准体系建设指南。随着隐私计算技术发展、应用落地、监管收紧，标准化建设工作需求将越来越迫切。下一步，隐私计算标准化工作将集中以下方面：一是促进不同厂商及技术 之间互联互通；二是各细分场景的隐私计算安全分级，如原始数据的计算性和隐私性、计算过程的安全性、结果信息反推原始数据的安全性等。

3.6.4 技术产品市场

2019 年，Gartner 首次将隐私计算列为处于启动期的关键技术 [15]。2020 年，Gartner 又将隐私计算列为 2021 年企业机构九大重要战略科技之一，并预测隐私计算将迅速得到落地应用，预计到 2025 年应用范围将覆盖全球一半的大型企业机构 [2]。近两年来，伴随着技术的不断成熟，国内外隐私计算产业化的步伐明显加快。可以预见，未来几年将是技术产品加速迭代，应用场景快速升级，产业生态逐步成熟的重要阶段。

a) **国外隐私计算技术研究创新活跃,但商业化进展稍缓**

从技术发展的历程来看，谷歌、Intel 等国际领军企业开创了隐私计算产业的时代潮流 [但从整体发展路径来看，相比国内企业，国际科技企业在学术研究和开源生态的建设上更为活跃；相比之下，商业化的产品形态较为局限，产业生态也尚未形成火热竞争或垄断格局。

微软研究院自 2011 年开始大规模推进多方安全计算的研究，从两方安全计算入手，逐渐拓展至三方计算和不存在交互行为的多方计算。但微软前期的 MPC 研究存在两个瓶颈，一是加密协议只针对一些简单的分析功能有效，如聚类分析、线性回归等；二是计算的执行必须运行在低水平的与或门电路中，执行过程麻烦而低效。2018 年，微软印度研究院推出了 EzPC 项目，希望克服上述两个问题。作为一个高效、可扩展的 MPC 协议，EzPC 是一个加密成本感知编译器，使用算术和布尔电路的组合，通过高级语言执行计算，支持神经网络训练和预测等复杂的算法。

谷歌是联邦学习的引路人，自 2017 年 4 月，谷歌便提出了联邦学习的概念，并于 2019 年发布论文给出了可扩展大规模移动端联邦系统的描述，用于改进谷歌输入法的自动关联与推荐。但与此同时，2019 年 8 月，谷歌又开源了名为 Private Join and Compute 的新型多方安全计算开源库，结合了隐私求交和同态加密两种基本的加密技术，帮助各组织和隐私数据集协同工作，针对个别项目还使用随机密钥进行高度加密，提高隐私性。

Intel 的 SGX 和 ARM 的 TrustZone 处于 TEE 硬件的垄断地位 [9,11]。基于 ARM TrustZone 实现的可信执行环境是一种硬件隔离安全机制，以物理方式将系统划分为安全和非安全组件，确保在正常操作下的软件无法直接访问安全区域的数据；而基于 Intel SGX 实现的可信执行环境是一种算力和内存隔离的安全机制，使用特殊指令 和软件将应用程序代码放入一个 Enclave 中执行，Enclave 可在虚拟机监控器、主操作系统和驱动程序均被恶意代码攻陷的情况下，仍然对其内的代码和内存数据提供安全

保护。TrustZone 在 2008 年推出，而 SGX 最早在 2013 年推出，二者都是随着移动手机的大发展而繁荣起来，目前市场上可信执行环境的商业化落地都是基于 TrustZone 或 SGX 的解决方案。

除上述提到的科技巨头外，国外互联网、AI、区块链领域的相关企业和机构也快速布局了隐私计算，Facebook 将基于 PyToych 的隐私计算机器学习框架 CrypTen 进行开源；AI 公司 Zama 开源了基于全同态加密的软件库 Concrete；麻省理工学院创办的区块链公司 Enigma 推出了基于多方安全计算的新加密系统，并与 Intel 合作研发基于 SGX 的可信执行环境；创业公司 Sharemind、Privitar 致力于搭建自研的多方安全计算平台。但从应用场景来看，目前的主要应用大多局限于将 MPC 技术应用于分布式密钥管理领域，如美国的 Unbound Tech 和丹麦的 Sepior。

从技术路径上看，各国际企业相对更关注基于可信执行环境的隐私计算。2019 年成立的 Linus 基金会旗下的机密计算联盟（Confidential Computing Consortium）便聚焦于此，关注基于可信硬件和云服务生态的数据安全，该联盟的创始会员包括阿里巴巴、腾讯、ARM、谷歌、Intel、微软、百度、华为等世界级企业，2020 年 AMD、英伟达、埃森哲、R3 等新一批知名企业也陆续加入。

b） 国内隐私计算技术产品蓬勃发展,形成一定优势

我国的隐私计算技术产业化在 2018 年后开始进入快速启动阶段，形成了互联网大厂、大数据公司、运营商、金融机构和金融科技企业、隐私计算初创企业为代表的五大类市场主要参与者。

阿里巴巴、百度、腾讯、京东、蚂蚁等各互联网巨头凭借自己在技术领域的积累，自 2019 年开始纷纷推出了各自的隐私计算产品，形成了跨业务、多团队、强支撑的发展态势，集团内部不同业务根据自身的业务特点和需求，选择一种或多种技术方案融合的方式进行开发；作为大规模数据资源拥有者的电信运营商为拓展业务形态，不仅三家运营商均在集团层面开始了隐私计算技术的选型与应用，天翼支付、电信云等子公司还自建平台服务于内部或其他机构的数据流通业务；金融机构是数据流通与安全应用最主要的需求者，国有银行的研究院或是事业部也均开始了隐私计算技术的研究工作，新心数科、神谱科技、平安科技、百融云创、度小满等金融科技类企业也将传统的数据建模、数据分析等业务拓展到基于联邦学习平台等的隐私计算服务中；同盾科技、星环科技、Talking Data、京信数科等代表性的大数据技术厂商也快速布局基于隐私计算的数据流通产品或平台。

以上企业的商业化路径大多是既要服务于企业自身运营需求，也可作为服务方为政务、金融等领域提供技术支撑。区别于此，如富数科技、华控清交、矩阵元、翼方健数、数牍科技、锘崴科技、光之树科技、零知识科技等一批专注于隐私计算产品化的初创企业也不断涌现。作为促进数据流通的关键技术，在国内大数据产业稳步发展、数据要素市场化配置加快推进的背景下，我国隐私计算技术产品日渐成熟，各领域应用场景加速落地,产业快速发展。面对隐私计算技术领域的国际竞争，我国已初具竞争优势，有望占据有利地位。

从技术路线上来看，多方安全计算的复杂度高、开发难度大,以华控清交、富数科技、矩阵元等为代表的隐私计算初创企业多致力于此，专注于打造以底层多方安全计算技术为基础的数据流通基础设施；可信执行环境对于硬件的局限及国外芯片的强依赖，使得其在国内的产品选型相对较少，较集中于百度、阿里巴巴等互联网大厂和冲量在线、隔镜科技等初创企业，但目前已出现冲量在线与兆芯在国产化硬件研发上的合作探索；对于联邦学习，由于机器学习类应用需求的突出，且有较成熟的开源社区为基础，开发难度相对轻松，因而运营商、金融科技公司等业务需求方大多专注在基于联邦学习的隐私计算产品化中。

2020 年，为提升行业认知，推进隐私计算技术与应用的融合，在工业和信息化部相关司局的指导和支持下，中国信息通信研究院云计算与大数据研究所牵头成立公益性合作平台“隐私计算联盟”（Privacy Preserving Computing Alliance），目前联盟已有包含技术厂商、政府单位、运营商和金融机构等在内的 50 余家成员单位。

c） 隐私计算的开源生态逐渐显现

开源社区的知识共享和多方协同有利于加快技术升级迭代和商业化项目落地的效率。对比传统的大数据技术工具，开源已成为生态中的绝对主流。作为保障数据合作与安全的重要基础，隐私计算有望进一步拥抱开源。近两年，很多大厂不断提供开源资源，目前国内外科技巨头在隐私计算领域的开源项目情况如图 3-6 所示。

序号	项目名	开源时间	机构	技术路径
1	PySyft	2017 年 7 月	OpenMined	多方安全计算、联邦学习
2	TF-Encrypted	2018 年 3 月	DropoutLabs、Openmined、阿里巴巴	多方安全计算
3	Asylo	2018 年 5 月	谷歌	可信执行环境
4	MesaTEE	2018 年 9 月	百度	可信执行环境
5	FATE	2019 年 2 月	微众银行	联邦学习
6	TF-Federated	2019 年 8 月	谷歌	联邦学习
7	Private Join & Compute	2019 年 8 月	谷歌	多方安全计算
8	PaddleFL	2019 年 9 月	百度	联邦学习
9	CrypTen	2019 年 10 月	Facebook	多方安全计算
10	Fedlearner	2020 年 1 月	字节跳动	联邦学习
11	Rosetta	2020 年 8 月	矩阵元	多方安全计算
12	KubeTEE	2020 年 9 月	蚂蚁集团	可信执行环境

图 3-6 代表性隐私计算开源项目

从目前国内外影响力较强的隐私计算开源项目来看，联邦学习主要有 PySyft、TF-Federated 和 FATE。PySyft 和 TF-Federated 目前仅支持试验环境。PySyft 是开源社区 OpenMined 开源的隐私计算框架，主要针对实现基于隐私计算的深度学习。PySyft 将联邦学习、多方安全计算以及差分隐私、远程执行等技术结合在一个编程模型中并集成到不同的深度学习框架中，如 PyTorch、Keras 或 TensorFlow；谷歌基于 TensorFlow 开源的 TF-Federated，则主要针对类似谷歌输入法案例的横向联邦学习。

FATE 是国内联邦学习商业化产品的主要贡献力量，由微众银行于 2019 年 2 月开源。FATE 提供了一种基于数据隐私保护的分布式安全计算框架，为机器学习、深度学习和迁移学习算法提供高性能的安全计算支持，支持同态加密、SecretShare 等多种多方安全计算协议，简单易用。目前，社区内已有超 370 家企业、164 所高校合作。

d) 配套标准体系日渐完善

技术体系的发展壮大需要配套标准指引的支撑。《多方安全计算技术框架》和《基于 TEE 的安全计算》两项国际标准分别于 2019 年 4 月和 2020 年 9 月在电气电子工程师学会（IEEE）立项，但现有标准的内容主要给出了通用性的技术框架，尚没有深入到应用中的细节。

中国信息通信研究院依托中国通信标准化协会大数据技术标准推进委员会于 2018—2020 年分别牵头制定了《基于多方安全计算的数据流通产品》《基于联邦学习的数据流通产品》《基于可信执行环境的数据计算平台》《区块链辅助的隐私计算技

术工具》4 项隐私计算技术产品功能上的系列标准。随着技术的火热发展，这些标准正在快速迭代、不断完善，针对不同产品的性能和安全性标准也正在加速制定中。

与此同时，中国信息通信研究院云计算与大数据研究所还依据已有标准积极开展技术产品的标准化评测，帮助市场建立起对于市场产品的客观评价体系，助力行业行稳致远。自 2019 年下半年开始启动的隐私计算技术产品评测已完成 3 批共 40 余次的产品评测。透过每一批评测中产品数量的快速增长，也可见证国内隐私计算产业的火热发展。

4. 隐私科技行业应用场景分析

4.1 金融行业-互联网信贷

4.1.1 业务背景及痛点

小微企业因自身经营风险大、财务制度不健全等问题，一直存在着融资难、融资贵的问题。近年来，为了解决上述问题，我国金融机构开始积极与互联网机构展开深度合作，进行了许多有益的业务探索。互联网机构依托自身优势，收集了多维度的小微企业数据。通过双方的合作，可以帮助金融机构获取更加完善的客户画像，识别有效需求，优化风险管控，并通过互联网渠道为小微企业提供便捷的线上融资服务，助力小微企业发展。

对于金融机构来说，在与互联网机构的合作中，由于客户数据往往掌握在互联网机构手中，而且监管机构三令五申，要求银行等金融机构禁止外包授信审查、风险控制等核心风控业务。因此，在合规的前提下与互联网机构展开数据合作，对银行等金融机构来说是展开互联网信贷业务必不可缺的一环。

对于互联网机构而言，随着互联网贷款的规模快速增长，其对用户数据的需求越发饥渴。一些机构打着“大数据”的旗号，通过“爬虫”技术涉嫌违规违法收集个人信息，盗取、滥用、买卖、泄露个人信息，侵犯消费者的个人隐私，造成了不良的社会影响。因此，监管机构积极加强对金融消费者的权益保护，严禁金融机构与涉嫌数据违规违法的企业展开合作。

因此，在个人隐私保护愈加重要的背景下，在互联网贷款领域如何实现数据安全有效融合，成为行业升级转型面临的重大挑战。

4.1.2 解决方案

本场景下，业务的主要需求是融合银行和互联网机构的数据，进行联合建模。模型一般分为训练和预测两个阶段。训练阶段对实时性要求较低，但是计算精确度要求加高；而在预测阶段，为了提升用户体验，所以对实时性要求较高。另一方面，需要融合的数据，主要包含用户的可识别信息、财产信息，往往涉及到大量敏感信息，因此对数据隐私保护要求很高，所以需要采用对个人金融信息保护较高的技术来实现。比如，可将联邦学习用于联合建模，并结合多方安全计算（MPC）进行安全加强。下文主要以该方案为例来展开相关解决方案。

在结合 MPC 的联邦学习风控合作模式下，参与机构在业务流程中的角色可分为任务发起方、调度方、数据提供方、多方计算引擎和结果获取方五种角色。其中，多方计算引擎的各密文计算节点应由互相独立的机构组成，或引入第三方可信机构；本地建模、密文数据输入、结果解密输出节点可以根据参与者动态调整；任务调度方、结果获取方根据合作模式确定，一般来说由金融机构既资金提供方担任。

在此模式下，数据融合过程可以分为五个环节。一是任务发起和调度环节，例如由银行发起对该信贷申请人的风险评估。二是本地明文计算与中间结果的计算因子输入环节。任务调度器启动本地模型，并将模型计算的中间信息进行数学转换，通过密码转换后的安全信息提交至多方计算引擎。三是中间结果密文融合计算环节。多方计算引擎将接收到的中间结果分发至各计算节点，分别各自计算与交互后完成中间结果的融合计算，再经解密后得到明文，返回到各参与方参与下一轮本地明文计算。经过多次迭代后，当建模结果的价值损失符合预期后，停止建模过程。四是计算结果解密输出环节。输出节点对建模结果的输出因子进行处理，得到明文结果。五是将计算结果反馈给风控系统。

该方案能够使银行和互联网机构之间的合作更加精确、安全与便捷，对用户、机构和监管皆有积极意义。

对用户来说，首先，可以解决个人征信信息被银行和互联网机构反复查询的问题；其次，该方案能使不同机构之间避免交换明文数据，极大程度上保障个人隐私安全。

对于机构来说，一是可以减少现有风控合作模式的价值损失，提高计算精度和数据使用率。二是提升融合效率，缩短审批流程，提高审批效率。

对于监管来说，本方案中金融机构和互联网机构的原始明文数据都不需要出本地，由此可以真正承担起对借款人数据保护的主体责任。同时，针对小微企业的融资难题，此方案可以促进银行加大小微信贷投放，帮助银行服务好小微企业并有效防范自身信贷风险。

4.2 医疗大健康行业

4.2.1 医疗数据共享现状及问题

随着医改、医疗大数据、医疗人工智能时代的到来，通过可信的数据共享来消除信息孤岛，已经成为各界的共识。传统的数据共享模式是通过统一的数据仓库或大数据平台集中采集、处理、存储并应用数据。但是在多家医疗机构之间需要共享数据，采用中心化的共享模式就会带来一系列问题。

多方不协调

传统的数据共享解决方案需要各个医疗机构将各自数据集中汇聚到统一的数据中心，但数据的主导权、管理权、运营权、使用权、共享权等，时常会带来较多的争议与不满，导致推动有较大阻碍【3】。

数据泄漏隐患

数据共享不可避免会产生医疗机构有数据泄露、数据共享后难以管控的风险，如何在充分保障数据安全的前提下，实现数据共享是一个必须尽快解决的难题。

数据确权难

数据在共享及流通过程中很容易被复制。如果不能对数据确权，明确数据的产生者、使用者、管理者及受益者，将无法很好实现数据的精准授权，严重阻碍数据的共享及流通。

无激励机制

传统的数据集中方式很难量化每个医院或者个人数据贡献的实际贡献大小，因此没有很好的共享激励机制。参与方无论共享的数据是多是少，数据质量是好是坏，获得的收益是一样的。如果没有合理的激励机制，每个参与方对自己的数据都会倾向于除了要求的数据，其他尽可能少共享或干脆不共享。

4.2.2 解决方案

利用区块链、隐私计算等隐私安全技术，构建公平和高效的医疗数据分享和流动安全方案，实现医疗数据流转的发布、授权、分享的线上化、可追溯及可审计，形成透明及可信的医疗数据共享和流动业务流程。利用联盟链、区块链密码学、区块链分布式账本、区块链智能合约、多方安全计算等隐私安全技术，提升数据确权能力和完善数据授权共享机制。建立准入机制，权限与逻辑控制体系等，促进数据实体交换和价值交换。实现透明的数据共享流动路径和责任明晰机制，确认数据安全责任的边界，提升各方数据共享的积极性。

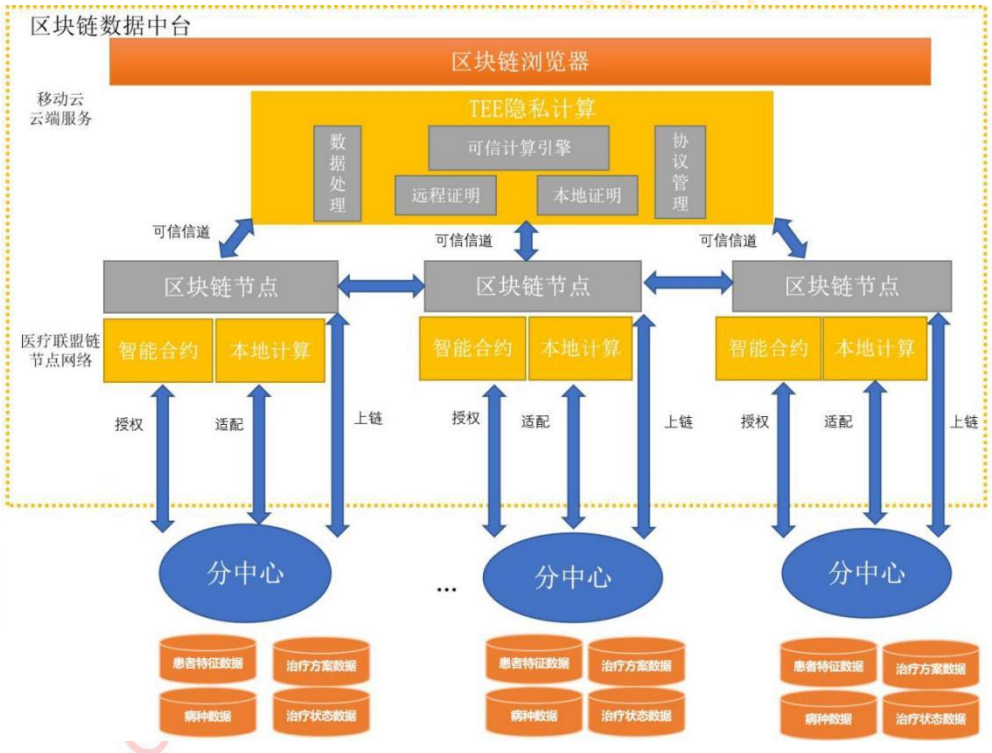

图 4-1 区块链数据中台示意图

如图 4-1 所示，用区块链、隐私计算等隐私安全技术，建立健康医疗数据共享和流动中隐私安全的技术方案：（1）区块链建设与准入机制设立。建立医疗数据区块链，在医疗机构、监管部门、科研机构、保险机构等建立区块链节点，部署联盟链。建立准入机制，做好与外部数据平台对接准备，建立权限与数据逻辑控制体系。（2）数据可信共享交换。医疗数据元数据相关信息通过本地的区块链节点上链存证，形成联盟链内的数据交易市场。通过区块链智能合约方式进行数据细粒度权限管控；利用

区块链分布式账本实现数据的价值激励，借助隐私计算中可信执行环境技术，达到数据“可用而不可见”，实现隐私数据的价值交换。（3）数据全流程监管。通过联盟区块链 CA 证书、电子签名等技术，完整存证每笔数据交易对应的数据产生者、使用者、管理者、受益者身份及行为，实现完整的全过程回溯，促进数据可信共享。

4.3 政府机构

4.3.1 政务数据开放背景介绍

在中国共产党十九届四中全会上，中央首次公开提出“健全劳动、资本、土地、知识、技术、管理和数据等生产要素按贡献参与分配的机制。”这反映了随着经济活动数字化转型加快，数据对提高生产效率的乘数作用凸显，成为最具时代特征新生产要素的重要变化。数据作为基础性和战略性资源，是提效降本、需求挖掘、技术创新的核心生产要素。如何实现数据的安全开放共享，释放数据价值，构建大数据支撑的科学治理与决策，带动行业数字化转型，促进经济发展，是每个地方政府急需解决的问题。

公共数据安全开放，释放政务数据价值，首先是善政科学治理要求。需要构建大数据支撑的政府决策，推动政务数据与社会数据的安全共享交换体系。实现互联网+政务服务，依托政务数据的安全有序开放提供支持。其次是振兴数字产业需求。5G、人工智能等数字经济新产业，急需海量数据，对政务数据开放有着迫切需求。第三是带动行业数字化转型的需要。也就是需要通过进一步开放政务数据和行业数据，推动大数据融合、促进行业转型。第四是创新经济发展需要：通过数据为引，促进商业合作，推动数据招商引资。推动实体经济和数字经济融合发展。

4.3.2 政务数据开放痛点

我国大数据的巨大商业价值还未被充分挖掘，具有价值的数据大部分集中在政府内部、大型国企以及互联网企业中。而分散的数据无法挖掘出大数据的巨大价值，不利于数据经济蓬勃发展，亟待政府和企业有序、安全开放数据。

一方面政务不敢共享开放，《网安法》、《数据安全法》、《个人信息保护法》等法规出台后，政府部门对共享开放数据敏感度提升，对于开放方式和方法更为谨慎。同时政务数据也难于共享开放，缺乏技术能力，严重制约了大数据作为基础性战略资源的开发应用和价值释放。

另一方面企业需要政务开放数据，但因缺乏相互信任，有需求无渠道。政企之间的合作关系需要信用支持，缺少信任机制、第三方监管和中立平台，审核严格周期漫长，合作很难顺利开展，企业有大量需求，但开放过程存在很大安全风险。跨领域数据孤岛严重且缺少开放技术路线。当前大多数地市政务数据缺少安全可靠的开放技术手段，形成数据孤岛，企业在获得许可后依旧可能遇到通过何种技术安全有序的使用数据的问题。

4.3.3 智能政务开放应用案例

探索公共信息资源开放，有利于促进资源的社会化利用。2019 年 7 月，全国首个大数据安全开放平台在厦门正式上线试运行，通过安全屋平台实现对厦门公共数据资源开放过程的合规、有序管理，提供安全的数据融合应用计算服务。开放数据涵盖 22 个领域主题、20 个行业分类、39 个政府部门的信用服务、交通运输、市场监管、生态环境、地理空间、生活服务等数据内容。同时归集了 44 个国家部委、31 个省市地方政府以及第三方机构 500 多亿条、涵盖 3400 多万家企业的信用信息，以及厦门本地已归集的 71 个部门、3.7 万项信用信息目录、5.3 亿条信用数据，已有政府、高校、企事业单位、科研机构等 114 家大数据生态合作伙伴入驻。

厦门大数据安全开放平台，能够支持政府数据、企业自有数据和第三方数据的接入。允许需求方使用数据，而不直接“拥有”数据，实现“数据可用不可见”的所有权和使用权的分离，实现数据物理隔离逻辑集中的管理模式，通过多种隐私计算技术及权限管控保证数据共享、数据流通、融合计算全流程的数据使用安全、高效、可控

图 4-2 安全屋技术示意图

安全屋支持精细化管理数据安全等级，对数据、表、字段级别可设置不同的安全级别，只有高于该级别的用户可拥有查看数据目录、使用数据的权限。

流程方面，产品使用具有完善的身份管理功能，支持多种鉴权、认证形式；在数据发布，数据授权，结果使用等都需要经过审核才能进入下一环节；所有的操作都写入日志并同步存入区块链，保证记录的审计合规、安全。

产品支持探索环境及真实环境，探索环境可以看到部分脱敏后的样例数据，便于算法模型和应用的调试与实现。真实环境下使用真实数据，通过多种技术保护数据的隐私和安全，包括：

通过基本的数据脱敏算法从原始数据中生成一部分脱敏后的数据，或者使用数据生成算法合成与普通的敏感用户数据无差异的合成数据。在部分场景下，这种脱敏后或者人工合成的数据已达到相关数据隐私保护的要求，可以直接作为样例数据提供给算法提供方或者数据使用方。

采用差分隐私技术（Differential Privacy），对数据进行进一步的匿名化处理，以便不再与其它的用户信息联系起来，从而完全保障样例数据的安全性。

使用同态加密，解决高敏感数据安全性、以及算力、数据的分离。

使用区块链，所有的操作日志记录在区块链上，确保审计与监管的安全可信，支持区块链驱动的数据使用流程。

使用联邦学习及安全多方计算，在企业自有数据、第三方数据或政府共享数据都需要保护且不能离开本地节点的场景下，进行数据使用和机器学习建模。基于厦门市大数据开放平台，推动产业实际应用例子包括:

(1) 全国信易贷平台

向信易贷平台开放经企业主体授权、脱敏后的市监、发改、自规、人社等相关数据服务，帮助金融机构评估企业信用，为信用良好企业提供信用贷款，破解企业融资难、融资贵痛点。

厦门市承建的全国信易贷示范平台，是国家发改委牵头推进的信用服务中小企业融资的国家重要基础设施，平台累计注册企业 52758 家、累计授信 37823 笔、已为信用良好的中小企业获得纯信用贷款超过 800 亿元，未来两年内将布局全国 100 多个城市、服务全国 1000 万家中小企业、撬动 1 万亿的信用贷款规模。

(2) 中国人工智能大赛平台

继第一届人工智能大赛成功举办后，第二届(2020 年)人工智能大赛包括多媒体信息技术识别大赛、语言与知识技术竞赛两项赛事，吸引了依图、网易、阿里等人工智能技术领先企业，还有中国科学技术大学、中科院计算所等知名高校和科研单位参与。最后，188 支队伍报名参加 277 个比赛项目，项目涉及人工智能多媒体领域多个前沿技术方向，贴近实际应用场景、综合性强。大赛采用了“厦门专有云+安全屋”主题作为在线竞赛环境，厦门为大赛搭建了全新云平台，提供计算、存储、网络基础资源等算力环境，既便于参赛选手远程参加比赛，又保证数据安全不泄漏。

参赛团队比赛思路紧跟应用学术界的最新进展，大赛涌现出了许多新颖、前沿的技术创新。通过该竞赛，摸清了国内人工智能行业的技术成熟度现状，为今后的政策引导、市场引领等提供有效的参考依据。大赛可以逐步成为引领人工智能发展方向、与行业应用深度融合的平台，并同步 10 个项目落地厦门 ，总投资 40 亿元，涵盖人工智能基础、技术、应用及业务合作各领域。

(3) 厦门白鹭信用分

向白鹭分系统开放经主体授权、脱敏后的个人身份识别、个人信用信息等数据服务，结合白鹭分系统提供的白鹭分算法模型计算个人白鹭信用分，提供信用借书、信用停车等便民服务。

(4) 连锁店商业选址

向第三方商业服务选址专业机构开放人口、交通、规划、兴趣点、房价、教育等空间数据、融合第三方机构自有数据（如手机信令）、选址模型算法（综合评估客流、年龄、消费、关注门店分布等因素），为连锁企业提供科学选址服务。

(5) 算法演练

企业、科研机构的人工智能产品应用需要有丰富、海量的数据训练。在现实中，数据提供方缺乏数据开放安全感，不敢过多提供数据，数据需求方无法获得足够的数据进行相关模型的训练。

通过引进“数据安全屋”技术，从网络、数据、业务多层次建立数据安全保障机制，打造不同场景下开放数据完全隔离的私有环境。对于开放靶场训练数据，企业可以通过演练专题模型、共享训练数据、解决人工智能落地问题。

4.4 零售与快速消费品行业

4.4.1 业务痛点

与传统的门店/卖场模式相比，零售与快速消费品等直接面向消费者的企业（Direct to Consumer,简称‘D2C’）在数字化时代开发了大量的线上业务渠道/业务：进驻天猫、京东、拼多多等电商平台开设品牌店、通过 App&小程序等移动应用渠道开设自营官网、“线上下单、线下取货”的“O2O”业务等。大量的线上业务渠道协助 D2C 企业收集了海量的数据，这也是企业进行数据分析并开展用户画像、精准营销的数据基础。

由于隐私合规与数据泄漏方面的顾虑，绝大多数企业还是基于“一方数据”进行用户画像和精准营销。“一方数据”的缺点在于根据其分析得出的用户画像往往是割裂的，数据标签也相对较为单一，只有通过整合不同类型 D2C 企业的数据，才能构建较为立体的用户画像，从而提升营销的精确度。

4.4.2 解决方案

通过联邦学习，可以协助企业在不交换原始数据的情况下加强其数据分析模型的计算能力：不同企业可对营销分析模型的标签、特征、梯度等进行加密后的交互，提升其分析模型的能力，再反哺至本地环境对其数据进行分析，从而增加标签维度、增加特征的精确性，从而提升用户画像的精确性和营销的精确度。

通过多方安全计算与差分隐私等技术，企业可将一方数据与三方数据进行有限度的安全求交（持有隐私数据集的多方计算其数据交集）、联合训练（如与广告服务商的转化链路数据进行联合建模），利用三方数据对一方数据进行更为精准的人群包细分，识别特定用户群并制定对应的营销与广告投放策略，提升营销精度与广告投放效果。

4.5 汽车行业

4.5.1 监管要求和业务痛点

2021 年 4 月份以来，工信部和国家互联网信息办公室相继出台了《智能网联汽车生产企业及产品准入管理指南(试行)》(征求意见稿)、《信息安全技术 网联汽车 采集数 据的安全要求》标准草案、《汽车数据安全管理若干规定》(征求意见稿)等相关文件，以《网络安全法》《数据安全法》和《个人数据保护法》为基础，围绕着智能网联汽车的数据治理工作提出了针对性的指导意见。同时汽车标准化技术委员会、中关村车载信息服务产业应用联盟等相关机构也先后立项或发布了多个智能汽车数据相关的国家或团体标准编。自此智能网联汽车及车联网数据业务进入规范发展快车道，

从合规层面杜绝了个别企业违规收集、使用，甚至变卖个人数据的行为。基于上述政策监管的变化，如何有效解决隐私数据合规与车联网应用对“人、车、路”大数据强烈需求之间的矛盾问题，成为车联网业务将来能否高速发展的燃眉之急。在这样的背景下隐私科技如在智能网联汽车行业落地，通过“数据可用不可见”的技术手段使车联网业务继续得以发展，也成为了行业内近一段时期以及未来一段时间内的热点话题。

4.5.2 解决方案

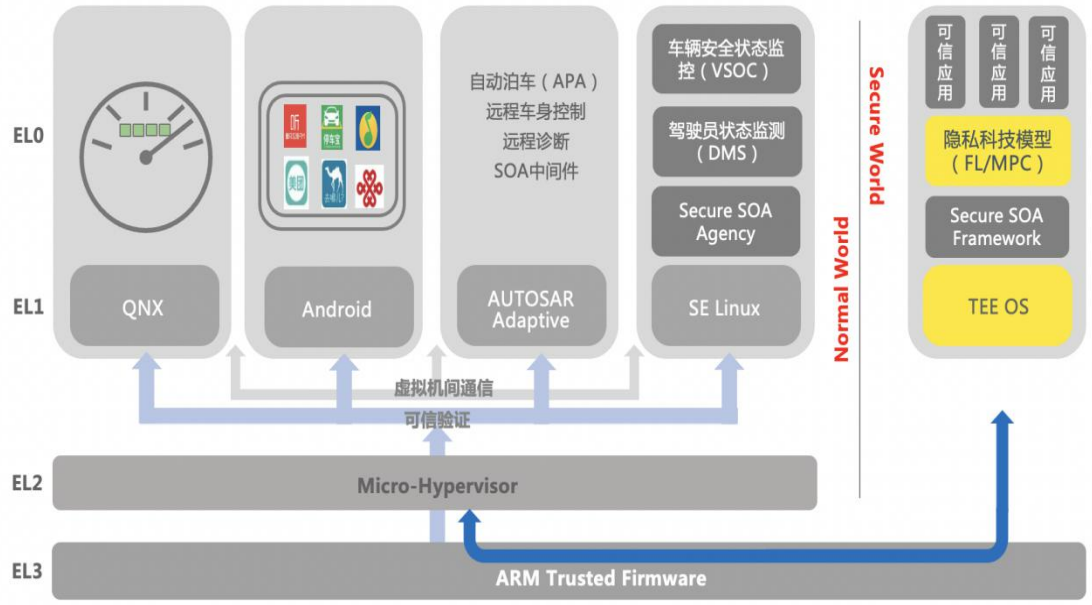

图 4-3 车端隐私科技架构示意图

如图 4-3 所展示的，在智能汽车集中式的电子电气架构中，在一颗高性能芯片上构建仪表盘、车机、ADAS、以及 Service OS（如图中 SE Linux）将成为方向，并与 SOA（面向服务的架构）配合以实现“软件定义汽车”的愿景。如前面 3.2.3 章节中所介绍的，在 ARM 芯片中以硬件隔离的方式构建了一个 Normal World（即 Android/Linux 的运行空间）和一个 Secure World（即 TEE OS 的运行空间）。

通过车端芯片 TEE 的隔离保护，联邦学习、多方安全计算、个人隐私数据等关键数据可以受到高级别的硬件保护，并通过 Secure SOA 的 API 向 Normal World 的 Android/Linux 等操作系统提供隐私计算服务。这个过程就如同手机的人脸/指纹解锁过程可以在手机芯片的 TEE 中完成一样。

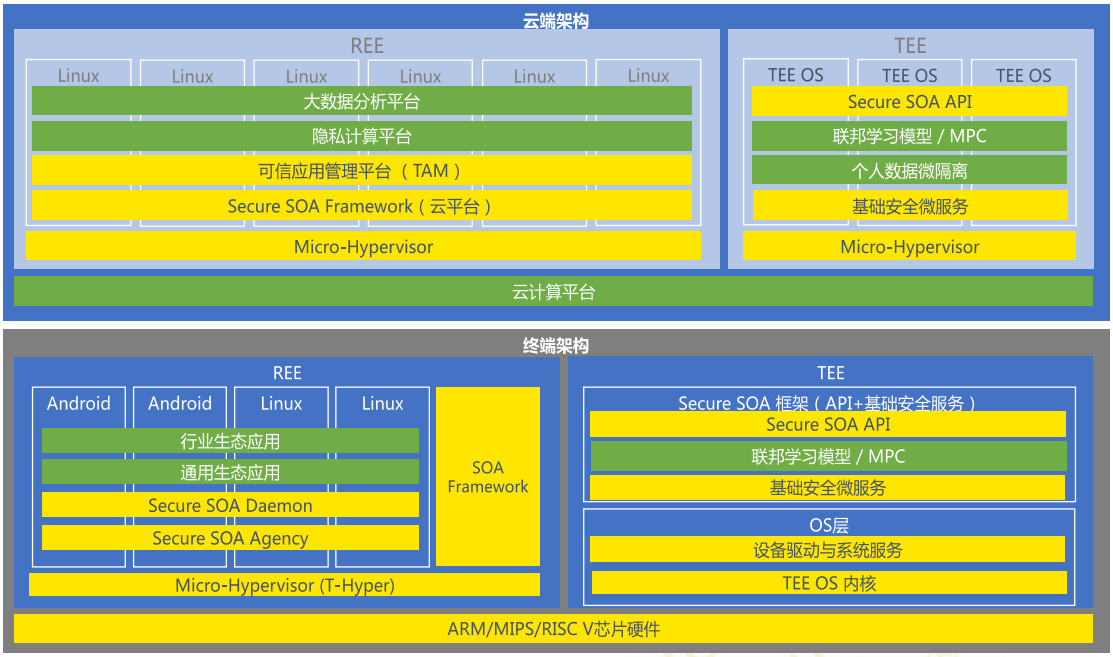

图 4-4 车联网整体架构示意图

如图 4-4 所示，车联网隐私科技体系由云端和车端两部分构成，并且均使用 TEE 对联邦学习模型和多方安全计算算法及运算过程进行保护。其中在云端的通过利用 ARM 最新的 CCA（机密计算）架构，可以实现在 TEE 中运行多个虚拟机，使不同虚拟机对应不同的车端系统，从而确保每个用户/车辆的数据都受到细粒度的访问控制权限保护。

4.6 电信运营商

4.6.1 未脱敏数据存在的安全风险

数据是运营商的重要资产，其中包含大量的运营商客户隐私数据。如果未脱敏的客户隐私数据被泄露或数据脱敏不当，将会带来巨大的业务安全风险。 未脱敏的隐私数据被泄露所涉及的过程包括：未脱敏数据被主动共享、未脱敏数据的开放查询、未脱敏数据库被攻击，等等。针对运营商行业的特点，未脱敏的数据面临如下安全风险：

未脱敏数据内部流转过程中的安全风险

数据存储于计算机终端后，由于业务需要，数据会经过业务系统或者内部网络进行交互传输。在该过程中，可能存在数据网络窃取，误操作导致错发等问题，如果流转中的数据未经脱敏处理，会使企业数据面临安全威胁。

未脱敏数据离网及外发过程中的安全风险

数据离开企业内部环境后，无法得到有效控制，任何接触数据的人员均可进行传播，如果数据没有经过脱敏处理，存在着二次泄漏的风险，这也往往是运营商数据泄漏事件的罪魁祸首。

各类应用系统中数据的安全风险

在运营商的各类应用系统中，存储着大量的行业核心资料，如客户信息、经营分析数据、财务报表和邮件资料等，在保证各应用系统的安全的同时，还应密切注意系统中数据的安全。数据脱敏是保障应用系统中数据安全的有效手段，经过脱敏处理后的数据，即使被泄露，也能大大降低数据的安全风险，这也是各运营商数据安全的重要问题。

各类重要数据生命周期的管控风险

目前，电信运营商已部署了 4A 统一安全管理平台，加强身份认证管理体系，但是 4A 平台的体系仅限于各业务系统的访问权限管控，无法细化到具体的数据体，也就无法对数据的生命周期进行全面、细致的管控，从而带来数据安全风险。显然，数据脱敏能有效弥补 4A 统一 安全管理平台在管控方面的不足，从而降低各类重要数据生命周期的安全风险。 另一方面，在上述各过程中，如果采用了不当的数据脱敏方法，脱敏后的数据仍有可能会被非法还原成原始数据，从而造成敏感信息泄露，带来巨大的安全风险。因此，在数据脱敏过程中，应优先选择不可逆的数据脱敏方法，并尽量避免对数据脱敏方法的不当使用，保证数据脱敏的有效性。

4.6.2 大数据脱敏解决方案

脱敏数据

案例中，需要脱敏的数据主要是针对电信用户个人敏感信息，比如： 1、个人属性类字段，包括用户身份信息，比如姓名、身份证号（其他证件号码）、家庭住址、位置、工作地、工作单位等，以及能映射到用户个体的相关信息，包括手机号码、联系人、用户 ID、IMEI、IMSI、服务密码和银行账号等。 2、用户通信（短信、语音）时的对端号码、位置等通信属性。 3、集团客户名称、集团客户关键联系人及其联系方式等集团属性。 4、通过 FTP 传输的结果数据文件内容。 5、经营分析系统对外提供的涉及敏感信息的数据。

脱敏场景

需要脱敏的场景主要是针对以下电信运营商应用场景，比如： 1、进入大数据平台后的各个环节做好脱敏处理，脱敏方法主要是加密。 2、在营业/客服前台展示、外部接口调用、投诉处理后台查询等过程，以及开发测试环境中进行数据脱敏。 3、在经分系统对外提供数据并且数据设计敏感信息的情况下需要进行脱敏。 4、对通过 FTP 传输的结果数据文件内容进行加密。

脱敏实现

大数据平台内的数据资产的应用，尤其是大数据分析系统其对数据的读取和计算都是发生在大数据平台之内，数据并不需要传输到大数据平台外部。在这种应用场景下，在数据传输通道的脱敏方式不适合这类应用。此外，对于大数据处理来讲，将数据导出后再进行脱敏，不能满足大数据脱敏的高性能要求。因此，利用大数据平台的自身能力实现高性能的大数据脱敏。大数据脱敏系统功能架构如下图所示：

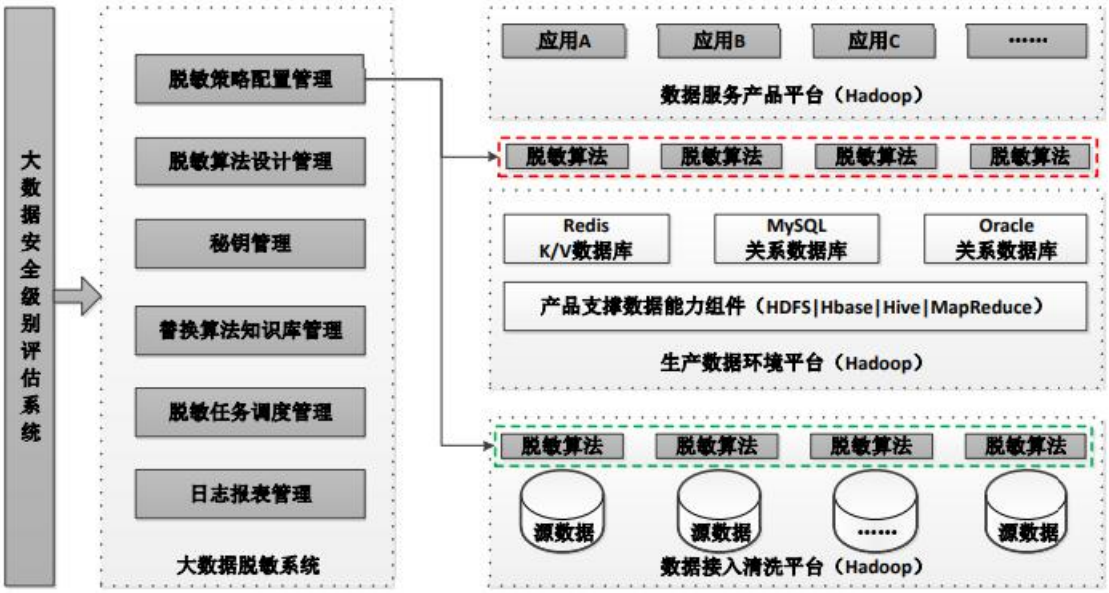

图 4-5:大数据脱敏系统架构示意图

第一步：大数据脱敏系统从大数据安全级别评估系统得到大数据资产的安全级别数据， 根据不同的安全级别以及应用的数据要求，制定不同的脱敏策略。 第二步：通过大数据脱敏系统配置好响应的脱敏策略，并将数据脱敏设置为脱敏任务， 一旦满足触发条件，立即进行脱敏。触发条件可以是时间、某一数据处理过程的调用。 第三步：触发脱敏条件后，大数据脱敏系统将脱敏算法的执行算法包下发到大数据平台，在大数据平台各节点实现大数据的并行脱敏。

5. 隐私科技未来发展趋势展望

5.1 隐私科技相关的法律与政策生态将持续完善与优化

从隐私保护立法的整体情况出发，近年来已出台了多部相关法律与相关要求，但我国隐私相关法律体系尚在不断完善的过程中。聚焦在隐私科技领域，虽然部分国家部委（如发改委）与行业主管机构（如银保监）已出台了一系列的政策，但仍缺乏国家层面整体性的产业政策与指导意见。

可以预见在不久的将来，法律与政策层面将持续完善。特别是针对数据交互，数据流通，数据共享，数据变现等方面的监督与指导意见，针对隐私科技领域相关技术的标准与认证体系，将对用户方选择产品有着极大的指导与促进作用。

5.2 通用性及行业性隐私科技解决方案并行

目前主流的隐私增强的底层技术如联邦学习、安全多方计算等，都存在一定的局限性。对参与各方特别是用户方的数据治理成熟度及其本身计算能力要求较高，无形中提高了适用性与技术应用的门槛与成本，目前主要还是集中在金融、医疗、政务等少数行业的试点应用阶段。国内隐私科技产品目前还处于发展的早期阶段，且呈现较为明显的“两级分化”状态，即大厂与少数特定行业需求明确且成熟度较高、需求已提升到需要”硬核“的隐私科技技术解决常规管理手段及技术无法解决的合规问题。而大多数行业仍然处在为满足基本合规要求、“通过 Excel 解决问题”的阶段，这部分需求未得到有效满足。

除了隐私科技技术本身的持续演进和优化以外，需降低隐私增强技术的使用门槛与使用成本，进而扩充用户群体，形成更加普适性的合规科技赋能。此外， 需进一步扩充及丰富应用场景，更加契合大多数行业的需求。在满足大多数用户基本需求的层面，可参考国外比较成熟的产品，结合国内用户的痛点与需求，打造中国版的一站式隐私合规科技产品。在相对高阶的“隐私计算”层面，探索不同的商业模式，如建立跨行业的底层技术与数据平台结合垂直行业的应用平台等，定位于打造未来数据融通交互的底层基础架构。

5.3 隐私科技赛道将进一步细分且明确定位，形成隐私保护合规新生态

在隐私科技蓬勃发展的短暂历程中，各家对隐私科技都有着不同的定位，是硬核技术？是数据合规的基础架构？是监管科技？是法律科技？抑或是合规科技？这表明了隐私科技覆盖了很大的范畴。隐私科技不同的定位将次生出诸多细分的赛道和领域，并且不同的细分赛道和领域之间将相互作用和协同，最终形成隐私保护合规的新生态链。

在撰写本白皮书的期间，围绕着隐私计算，隐私科技涌现出大量的研究与调研的报告和白皮书，我们正处在一个隐私科技急速发展的时期。可以预见在未来的几年内，市场上将涌现出大量的隐私科技新技术，新赛道和新厂商。随着隐私合规落地实施与运行的不断深入，隐私科技市场的趋势也会随之不断变化，并在这一过程中不断迭代，演进，持续赋能隐私合规。

6. 附录—参考文献

【1】 https://www.freebuf.com/articles/database/244536.html

【2】 《隐私计算与区块链技术融合研究报告(2021)》，中国信息通信研究院，2021

【3】 https://t.cj.sina.com.cn/articles/view/6932851545/19d3aeb5900100ppeq?sudaref=cn.bing.com&display=0&retcode=0

【4】 https://www.cnblogs.com/hzcya1995/p/13312525.html

【5】 https://database.51cto.com/art/202012/635979.htm

【6】 http://www.360doc.com/content/21/0516/19/70074794_977472415.shtml

【7】 http://news.hexun.com/2021-03-25/203282463.html

【8】 https://www.fx361.com/page/2019/0524/5145367.shtml

【9】 https://baijiahao.baidu.com/s?id=1711051969147151296&wfr=spider&for=pc

【10】https://www.chyxx.com/industry/202106/957171.html

【11】https://baijiahao.baidu.com/s?id=1713484731584124782&wfr=spider&for=pc

【12】《中国移动大数据安全脱敏实施指南》