北京金融科技产业联盟
BEIJING FINTECH INDUSTRY ALLIANCE

# 隐私保护技术在金融应用研究

## ——以个人金融信息保护和金融反欺诈为例

北京金融科技产业联盟

2023 年 12月

# 版权声明

本报告版权属于北京金融科技产业联盟，并受法律保护。转载、编摘或利用其他方式使用本报告文字或观点的，应注明来源。违反上述声明者，将被追究相关法律责任。

# 编制委员会

**编委会成员：**

何　军　　聂丽琴　　朱　钢

**编写组成员：**

| | | | | | |
|---|---|---|---|---|---|
| 黄雅琼 | 李松涛 | 李　敏 | 王自冲 | 隆　峰 | 徐　超 |
| 方　竞 | 杨天雅 | 黄　淼 | 王鹏达 | 郭　璐 | 贾金龙 |
| 陈　涛 | 黄翠婷 | 张晓蒙 | 高　靓 | 彭　晋 | 金银玉 |
| 单进勇 | 蔡超超 | 李晶晶 | 王湾湾 | 李　博 | 郑华祥 |
| 时　代 | 董婉婷 | 李克鹏 | 陈　明 | 刘站奇 | 李　力 |
| 唐仕豪 | 周蓓杰 | 张海燕 | 高强裔 | 马　鸣 | 陈嘉俊 |
| 张敬之 | 曹旭涛 | 蒋美献 | 顾逸晖 | 李秉帅 | 邵云峰 |
| 高志民 | 平庆瑞 | 张佳辰 | 郑亚松 | 于　博 | 高　扬 |
| 包　芬 | 陶建萍 | 刘敬谦 | 郭瑞峰 | 王　超 | 陈小军 |
| 范廷钰 | 张　伟 | 林　挺 | | | |

**编　审：** 黄本涛　　郭　栋　　刘宝龙

**参编单位：**

中金金融认证中心有限公司

上海富数科技有限公司

北京冲量在线科技有限公司

同盾科技有限公司

蚂蚁科技集团股份有限公司

北京数牍科技有限公司

深圳市洞见智慧科技有限公司

华控清交信息科技（北京）有限公司

深圳市腾讯计算机系统有限公司

北京市竞天公诚律师事务所

北京国家金融科技认证中心有限公司

浙商银行股份有限公司

华为技术有限公司

中钞区块链技术研究院

上海光之树科技有限公司

上海浦东发展银行股份有限公司

蓝象智联（杭州）科技有限公司

中国科学院信息工程研究所

天津科技大学

# 目 录

一、 个人金融信息保护的合法合规需求 ..... 1
（一） 国际法律法规 ..... 1
（二） 国内法律法规 ..... 4
（三） 国内行业标准 ..... 8
（四） 涉及个人金融信息安全合规的重点 ..... 14
二、 金融隐私保护技术发展概述 ..... 16
（一） 金融隐私保护技术概念 ..... 17
（二） 金融隐私保护研究必要性 ..... 18
（三） 金融隐私保护技术要求及应用场景 ..... 22
（四） 金融隐私保护技术国内外应用发展 ..... 26
三、 隐私保护技术的基本体系 ..... 31
（一） 隐私保护脱敏技术 ..... 32
（二） 隐私保护计算技术 ..... 35
（三） 隐私保护辅助技术 ..... 48
四、 隐私保护技术在个人金融信息保护中的应用 ..... 50
（一） 隐私保护计算环境 ..... 50
（二） 隐私保护脱敏技术 ..... 53
（三） 隐私保护计算技术 ..... 57
（四） 隐私保护辅助技术 ..... 71

五、 使用隐私保护技术的金融反欺诈案例分析..... 73
（一） 风险控制..... 73
（二） 反欺诈..... 84
（三） 反洗钱..... 104
六、 隐私保护技术在金融领域应用的展望..... 109
（一） 金融隐私保护技术发展方向..... 109
（二） 金融隐私保护应用发展方向..... 110
参考文献 ..... 112

北京金融科技产业联盟 BEIJING FINTECH INDUSTRY ALLIANCE

**摘要：**目前数据保护技术不断发展，但根据金融数据保护具体场景的技术选择等尚无指导方案，本课题以《个人信息保护法》《数据安全法》等法律法规和监管要求为背景，参考《个人金融信息保护技术规范》等行业规范，从合规角度，分析现有隐私保护技术在个人金融信息保护中的应用情况，针对个人金融信息保护和金融反欺诈的不同应用场景，分析各隐私保护技术的优劣势，解决大数据量、高维度、高吞吐的多层级跨机构金融数据高效安全共享问题，为涉及个人金融信息保护的系统实施方提供技术选型的参考。

# 一、个人金融信息保护的合法合规需求

近年来，金融领域与人工智能、大数据、物联网等新兴科技不断深度结合，通过数据共享创造出了大量实用的金融科技应用，为客户带来个性化的金融服务体验，创造了可观的经济效益。与此同时，客户也面临着个人数据在未经同意的情况下被滥用与共享的风险，从而造成了公民信息泄露等数据安全问题。针对隐私问题，金融机构也逐步探索相关隐私保护技术，实现“数据可用不可见”，在保证数据安全的前提下实现数据共享，进而重构与客户之间的信任关系。同时，各国政府也相继出台相关的法律法规来限制个人信息滥用，从而保护公民的个人隐私，如欧盟出台的《通用保护条例》、美国加州政府出台的《加州消费者隐私法案》以及国内政府出台的《个人信息保护法》等，通过法律来规范数据在整个生命周期中的安全使用及管理。

## （一）国际法律法规

当今世界，以欧盟和美国为代表的个人信息保护模式较为突出。前者采取统一立法的模式，制定标准、严格地规范以保护公民个人信息，后者采取分散立法的模式，往往兼具突出行业自律力量的特征。个人金融信息是指银行业金融机构在开展业务时，或通过接入中国人民银行征信系统、支付系统以及其他系统获取、加工和保存的个人信息[1]，具有高度的敏感性与私密性，往往会对公民个人的人身和财产安全

产生重大影响，因此，对于个人金融信息保护的立法也成了世界各国的立法重点。

### 1. 分散立法

在数据治理领域，美国一向推崇行业自律，各州立法都以维护信息自由流动为宗旨。在个人金融信息保护方面，美国沿用隐私法的规制框架，通过联邦和各州层面立法进行法律保护，并且同时区分银行业、保险业、证券业等不同行业类别实现分业管理，重视行业自律组织的自律管理力量。在联邦法层面，美国对个人金融信息保护的相关规定主要可见于《公平信用报告法案》（Fair Credit Reporting Act）、《金融服务现代化法案》（Financial Services Modernization Act of 1999），其中对金融机构应如何收集、使用、存储、披露、分享个人信息作出了较为明确的规定；在州立法层面，诸如加利福尼亚州发布了《加州消费者隐私法案》（California Consumer Privacy Act）、弗吉尼亚州制定了《消费者数据保护法》（Consumer Data Protection Act）、科罗拉多州颁布了《科罗拉多州隐私法案》（Colorado Privacy Act）等，其中均对个人信息安全管理提出了更为细化的要求。除立法规制外，美国对个人金融信息的保护还高度依赖于行业自律组织的约束，金融机构重视、遵守金融行业协会制定的相关行业规范。

与美国类似的还有日本。日本对于个人信息保护也采取了分散立法的模式，对于个人金融信息的保护，不仅包含《日

本个人信息保护法》，日本还制定了《金融领域个人信息保护方针》与《金融领域个人信息保护方针的安全管理措施实务指南》等专项性法律规范文件，这些文件共同构成了日本个人信息保护的法律体系。

**2. 统一立法**

针对个人信息保护与数字经济，欧盟的立法始终走在世界的前列。对于个人信息保护，欧盟确立了制定统一法，构建流畅行权机制的保护模式，即采取了以国家立法为主导，秉持个人信息的合理使用与严格保护并行的观念，通过法律规范树立统一的个人信息保护标准，同时完善对个人信息权益侵权行为的救济途径，从而构建起强有力的个人信息保护体系。1995 年，欧盟发布《个人数据保护指令》（Data Protection Directive），构成了欧盟隐私和人权法的重要组成部分，为个人金融信息保护提供了统一的标准，也为欧盟各国对个人数据的监管建设提供了立法参考。2018 年，欧盟《通用数据保护条例》（General Data Protection Regulation）正式生效，对涉及欧盟境内的所有数据处理行为都作出了更为严格的规制措施，是欧盟在数据治理与个人信息保护领域中最为重要、核心的法律。

除欧盟外，韩国也采取了为统一专门立法的模式，在个人信息及数据保护法律领域，形成了《个人信息保护法》《信息通信网利用促进及信息保护法》及《信用信息的利用及保护法》三法分立的局面。

## （二）国内法律法规

### 1. 国内法律

法律方面，2016 年 11 月 7 日十二届全国人大常委会第二十四次会议表决通过《中华人民共和国网络安全法》（以下简称“《网络安全法》”），自 2017 年 6 月 1 日起施行，强调收集的用户信息应严格保密，维护网络数据的完整性、保密性和可用性，实行网络安全等级保护制度。 2021 年 6 月 10 日十三届全国人大常委会第二十九次会议通过《中华人民共和国数据安全法》（以下简称“《数据安全法》”），自 2021 年 9 月 1 日起施行,强调数据安全与开发利用并重，确立数据分类分级管理制度，多种手段保证数据交易合法合规。2021 年 8 月 20 日十三届全国人大常委会第三十次会议表决通过《中华人民共和国个人信息保护法》(以下简称“《个人信息保护法》”），自 2021 年 11 月 1 日起施行，强调个人信息在数据流通过程中的安全合规，明确了敏感个人信息处理规则、个人信息处理者的义务等内容，国内相关法律见表 1 所示。

表 1 国内相关法律

| 法律法规 | 主要内容 | 实施时间 |
| --- | --- | --- |
| 《中华人民共和国网络安全法》 | 强调收集的用户信息应严格保密，维护网络数据的完整性、保密性和可用性，实行网络安全等级保护制度。 | 2017 年 6 月 1 日 |
| 《中华人民共和国数据安全法》 | 鼓励数据依法合理有效利用，确立数据分类分级管理规则，促进以数据为关键要素的数字经济发展。 | 2021 年 9 月 1 日 |
| 《中华人民共和国个人信息保护法》 | 强调个人信息在数据流通过程中的安全合规，明确了敏感个人信息处理规则、个人信息处理者的义务等内容。 | 2021 年 11 月 1 日 |

个人信息保护法

《个人信息保护法》作为我国个人信息保护领域中的首部纲领性法律，对个人信息处理者开展个人信息处理活动提供了重要指引。尽管《个人信息保护法》并没有专门关注个人金融信息，但其对个人金融信息在内的各类个人信息提供了周密的保护。金融机构在日常业务中接触到的自然人的大部分个人信息可能都属于敏感个人信息（如：姓名、身份证号码、银行账户、通讯方式、生物识别信息等），《个人信息保护法》中第二章中明确了敏感个人信息的处理规则，为个人金融信息的保护提供了明确的法律依据。此外，金融机构作为个人信息处理者时，一般需要极为重点地关注处理权限及合规义务两方面内容。其中，处理权限绝大部分来源于

个人信息主体的明示同意，对于个人敏感信息，根据《个人信息保护法》的规定，处理个人敏感信息需要取得信息主体的单独同意，同时还应当向个人告知处理敏感个人信息的必要性以及对个人权益的影响。

数据安全法

《数据安全法》作为数据安全领域的基本立法，为企业合法处理数据、保障数据安全等提供了较为明确的指引，也对金融机构加强数据安全合规、完善数据治理体系、夯实技术能力支持等提出了更高的要求。《数据安全法》第 21 条规定，国家建立数据分类分级保护制度。根据数据在经济社会发展中的重要程度，以及一旦遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度，对数据实行分类分级保护。数据分类分级有利于提升企业自身的信息化水平和运营能力，可以更好地将数据资产化，也是开展数据全生命周期管理的基础和前提。通常来讲，数据分类分级流程包括：数据资产梳理、数据分类、数据定级、审核标识管理、数据分类分级保护。

此外，大部分金融机构会被认定为重要数据处理者。《数据安全法》第 27 条、30 条、31 条对重要数据提出了三方面要求，一是重要数据的处理者应当明确数据安全负责人和管理机构，并落实数据安全保护责任；二是重要数据的处理者应当按照规定对其数据处理活动定期开展风险评估；三是关

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重要数据的出境安全管理。

**2. 国内法规**

（1）反洗钱、反恐怖融资与客户身份识别

2021 年 6 月 1 日，中国人民银行组织起草了《中华人民共和国反洗钱法（修订草案公开征求意见稿）》，其中明确规定金融机构应当设立反洗钱专门机构或者指定内设机构负责反洗钱工作，应当建立客户尽职调查制度和客户身份资料和交易记录保存制度。金融机构通过第三方识别客户身份的，应当对第三方的反洗钱义务能力承担责任。2021 年 8 月 1 日，中国人民银行印发《金融机构反洗钱和反恐怖融资监督管理办法》，对金融机构的信息系统建设提出了新要求，应搭建反洗钱信息系统，并应当根据风险状况、反洗钱和反恐怖融资工作需求变化及时对系统进行优化升级。

（2）《银行保险机构信息科技外包风险监管办法》

科技外包对金融机构而言是很常见的一种情形。2021 年 12 月 31 日原中国银行保险监督管理委员会印发了《银行保险机构信息科技外包风险监管办法》（银保监办发〔2021〕141 号）（以下简称“《办法》”）。《办法》首次明确了银行保险机构不得将信息科技管理责任、网络安全责任外包的原则。针对信息科技外包，《办法》的主要监管思路为分类、分级管理，并罗列了属于重要外包的情形。同时，《办法》

并不仅局限于科技外包这一种形式，而是将相关的业务活动，只要具备信息合作属性的业务合作均可以视为科技外包。

（3）《征信业务管理办法》

2021 年 9 月 1 日，中国人民银行发布《征信业务管理办法》，分别针对个人征信和企业征信，提出了细化《个人信息保护法》在金融行业的要求。针对个人征信业务，征信机构应当制定采集个人信用信息的方案，并向中国人民银行报告；针对企业征信业务，征信机构应基于合法目的开展业务，不得侵犯企业商业秘密。此外，征信机构应当对信息使用者接入征信系统的网络和系统安全、合规性管理措施进行评估，落实网络安全等级保护制度，并对征信机构提出了数据本地化的要求。

### （三）国内行业标准

我国在个人信息保护方面出台了一系列国家标准和金融行业标准，明确提出了个人信息保护的安全合规要求。在此，对相关的国家标准和金融行业标准进行梳理总结如下。

#### 1. 涉及个人信息保护的国家标准

（1）《信息安全技术 个人信息安全规范》GB/T 35273-2020

标准针对个人信息面临的安全问题，规定了开展收集、存储、使用、共享、转让、公开披露、删除等个人信息处理活动应遵循的原则和安全要求。2020 版标准在 2017 版标准

基础上，对于用户画像的使用限制、个性化展示的使用、第三方接入管理等方面做出了补充和修订。

（2）《信息安全技术个人信息安全影响评估指南》GB/T 39335-2020

指南是《信息安全技术 个人信息安全规范》的配套标准，明确了个人信息安全影响评估的基本框架、方法和流程。针对个人信息处理活动，检验其合法合规程度，判断其对个人信息主体合法权益造成损害的各种风险，以及评估用于保护个人信息主体的各项措施有效性。

**2. 涉及个人金融信息保护的行业标准**

自 2020 年以来，金融行业监管机构高度重视个人金融信息保护以及金融数据安全等方面的标准建设，先后修订、发布实施了一批标准。

（1）《个人金融信息保护技术规范》JR/T 0171-2020

标准明确给出了个人金融信息、支付敏感信息的定义、范围，标准规定了个人金融信息在采集、传输、存储、使用、删除、销毁等生命周期各环节的安全防护要求，从安全技术和安全管理两个方面，对个人金融信息保护提出了规范性要求。为金融行业个人金融信息保护工作制定了基线要求。

（2）《金融数据安全 数据安全分级指南》JR/T 0197-2020

标准给出了金融数据安全分级的目标、原则和范围，以及数据安全定级的要求、规则和定级过程。金融数据复杂多

样，对数据实施分级管理，能够进一步明确数据保护对象。统一的数据分级管理制度，能够促进数据在机构间、行业间的安全共享，有利于金融行业数据价值的挖掘与实现。

（3）《金融数据安全数据生命周期安全规范》JR/T 0223-2021

标准规定了金融数据生命周期安全原则、防护要求、组织保障要求以及信息系统运维保障要求，建立覆盖数据采集、传输、存储、使用、删除及销毁过程的安全框架。标准在满足金融业务基本需求的基础上，强化数据保护能力，保障金融数据安全流动等方面，将起到积极意义。

（4）《多方安全计算金融应用技术规范》JR/T 0196—2020

标准规定了多方安全计算技术金融应用的基础要求、安全要求、性能要求等，适用于金融机构开展相关产品设计、软件开发、技术应用等。标准有助于实现在不泄露原始数据、保障信息安全前提下推动多个主体间的数据共享与融合应用，确保数据专事专用、最小够用，杜绝数据被误用、滥用。

（5）《金融业数据能力建设指引》JR/T 0218-2021

标准对数据保护能力域提出明确要求，包括数据保护策略、数据保护管理、数据保护审计等方面。明确通过数据脱敏、加密、过滤等手段，保障数据安全和数据隐私；依据个人金融信息保护策略开展数据隐私保护。

### 3. 金融业标准中涉及个人金融信息保护的合规要求

金融业在相关领域的标准中也明确了对于个人金融信息保护的要求，具体如下。

（1）《中国金融移动支付 检测规范 第 8 部分：个人信息保护》JR/T 0098.8-2012

标准规定了移动支付个人信息保护的内部管理、组织管理、访问控制和个人信息生命周期管理等方面的基本要求以及检测细则。标准明确了个人信息和敏感个人信息的定义、个人信息生命周期管理要求。

（2）《云计算技术金融应用规范 安全技术要求》JR/T 0167-2020

标准规定了金融领域云计算技术应用的安全技术要求，其中第九章数据安全要求识别个人信息并在传输、存储过程中采取技术措施保证信息的保密性。

（3）《网上银行系统信息安全通用规范》JR/T 0068-2020

标准规定了网上银行系统安全技术要求、安全管理要求、业务运营安全要求，同时明确了网上银行系统中涉及客户支付敏感信息（动态口令令牌、短信验证码、生物特征等）的安全保护要求。

（4）《征信机构信息安全规范》JR/T 0117-2014

标准规定了不同安全保护等级征信系统的安全管理、安全技术和业务运作方面的要求，对征信系统在信息采集、加

工、保存时提出安全要求。标准明确了征信机构基于理论研究、模型设计、产品开发等目的使用个人信息的，应当以汇总统计数据或者不能识别个人身份的方式进行。征信系统采集的个人不良信息应当按照法律法规规定的期限进行保存；超过保存期限的个人不良信息，应当从征信系统中删除，或者进行去标识化处理，应采取有效措施，确保去标识化处理后，个人身份不被直接或间接识别。

（5）《金融科技创新安全通用规范》JR/T 0199-2020

标准从交易安全、服务质量、算法安全、架构安全、网络安全、业务连续性保障等方面，提出了对金融科技创新相关科技产品的基础性、通用性要求。明确了金融科技创新中个人金融信息保护标准应遵从 JR/T 0171-2020 中的要求。

（6）《金融科技创新风险监控规范》JR/T 0200-2020

标准规定了金融科技创新应用风险监控的要求，明确了风险监控的原则包括应按照隐私保护原则，开展风险监控过程中对于可能涉及的个人或金融交易隐私信息应采取有效安全保护措施，确保隐私信息安全不被泄露。明确规定了个人金融信息保护方面应监测隐私政策合规、信息收集权限和范围、异常访问管理、信息泄露监测等内容。

（7）《区块链技术金融应用 评估规则》JR/T 0193-2020

标准规定了区块链技术金融应用的具体实现要求、评估方法、判定准则等，同时明确了区块链技术应用中涉及隐私保护策略、技术、监控与审计的评估规则。

（8）《金融大数据平台总体技术要求》JR/T 0237-2021

标准规定了面向金融业的大数据平台的框架结构、功能技术要求、非功能技术要求以及接口技术要求，同时明确了金融大数据平台的安全性要求应支持个人隐私信息保护。

**4. 金融业个人金融信息保护监管趋势**

中国人民银行高度重视个人金融信息保护工作，早在2011年发文《关于银行业金融机构做好个人金融信息保护工作的通知》（银发〔2011〕17号），明确了个人金融信息定义、范围，提出了相关保护要求。随后在金融科技、金融消费者权益保护、征信管理等多个业务条线明确个人金融信息保护的要求。

在2021年底发布《金融科技发展规划(2022—2025年)》中明确重点任务之一是充分释放数据要素潜能，一方面要求做好数据安全保护，严格落实数据安全保护法律法规、标准规范，明确数据安全负责人和管理机构，综合运用声明公示、用户明示等方式，明确原始数据和衍生数据收集目的、加工方式和使用范围，确保在用户充分知情、明确授权前提下规

范开展数据收集使用，避免数据过度收集、误用、滥用。建立健全数据全生命周期安全管理长效机制和防护措施，运用匿踪查询、去标记化、可信执行环境等技术手段严防数据逆向追踪、隐私泄露、数据篡改与不当使用，依法依规保护数据主体隐私权不受侵害。另一方面，推动数据有序共享。在技术方面，积极应用多方安全计算、联邦学习、差分隐私、联盟链等技术，探索建立跨主体数据安全共享隐私计算平台，在保障原始数据不出域前提下规范开展数据共享应用，确保数据交互安全、使用合规、范围可控，实现数据可用不可见、数据不动价值动。

由此可以看出，隐私计算技术在保证数据可用不可见方面有着先天优势，监管机构将会进一步加强隐私计算技术在个人金融信息保护方面的应用，加快隐私计算技术相关金融标准的制修订，推动并规范隐私计算技术在金融行业的应用，实现数据可用不可见、数据不动价值动，强化数据作为生产要素的有效配置，促进数据的有序流动，充分释放数据要素潜能。

### （四）涉及个人金融信息安全合规的重点

《中华人民共和国个人信息保护法》的颁布与施行健全了新时代个人信息保护的相关法律体系，其强调个人信息在数据流通过程中的安全合规，明确了个人信息处理和跨境提供的规则、个人信息处理者的义务等内容。随着金融行业数字化的持续发展，金融机构往往持有大量的重要金融数据，

数据的安全与风险防范一直是国家和相关监管部门关注的重点。个人金融信息是个人信息在金融领域的细化与扩展，包括但不限于个人身份信息、账户信息、交易信息、财产信息、借贷信息等。《个人金融信息保护技术规范》作为中国人民银行发布的金融行业标准，对金融业机构加强个人金融信息安全管理，规范处理个人金融信息起到了指导与规范作用。

在银行业领域，贷款业务、理财业务、个性化营销业务都会涉及个人金融信息安全合规要点。例如在贷款业务中，商业银行会对该自然人进行信贷审批，涉及个人征信的查询、存储、加工、使用等活动。根据《征信业管理条例》和《个人金融信息保护技术规范》，首先个人征信信息采集和使用需经信息主体本人同意，其次还需要满足个人信息保护的生命周期技术要求、安全运行技术要求、安全准则、安全策略、访问控制、安全监测与风险评估、安全事件处理等要求。

在保险业领域，商业保险的售前、售中、售后行为都会涉及个人金融信息安全合规要点。原中国银行保险监督管理委员会曾公开征求意见的《保险销售行为管理办法（征求意见稿）》中明确要求保险公司、保险中介机构应当按照合法、正当、必要的原则收集处理投保人、被保险人、受益人以及保险业务活动相关当事人的个人信息，并妥善保管，防止信息泄露；未经该个人同意，保险公司、保险中介机构、保险

销售人员不得向他人提供其个人信息，法律法规规章另有规定的除外。

在证券业领域，金融产品代销、期货中间介绍等业务中都会涉及个人金融信息安全合规要点。证券业务属于信息密集型行业，在经营管理中对个人信息保护与金融数据合规方面面临前所未有的挑战。个人金融信息属于证券公司的数据资产，在建立个人信息数据台账，梳理个人信息数据资产时需要参考《个人金融信息保护技术规范》《证券期货业数据分类分级指引》《金融数据安全 数据安全分级指南》等标准和规范，对个人金融信息进行专业化、标准化的分类分级管理，高度重视投资者及相关业务个人的个人信息保护。

## 二、金融业隐私保护技术发展概述

随着大数据、云计算、人工智能、区块链等新兴数字科技与金融行业的深度融合，孕育了各种新的金融应用和服务模式，促进了新一代金融业态的繁荣与发展。在金融服务高度信息化的今天，这些新业态、新应用和服务，为了提供给用户精准化、个性化、便捷化的服务体验，共享个人金融信息对于个人和社会都会带来好处。然而，人们在享受金融服务便利的同时，对于个人金融信息不当扩散可能带来的不利影响也日益关注。个人金融信息是金融机构面向客户提供金融产品和服务时沉淀的重要基础数据，也是金融消费者的重要隐私内容。由于这些数据具有非常大的潜在商业价值，其

被非法获取、滥用甚至出售营利的风险也随之加剧。因此，如何在保护个人信息的前提下，发挥这些数据对于金融机构及其整个行业领域的商业价值，保证其在合法、合规的条件下得到充分、有效的变现及应用，成为人们广泛关注的焦点。

### （一）金融隐私保护技术概念

隐私保护技术指用于支撑隐私保护与合规的日常运营流程，且嵌入到 IT 架构和业务场景中的一系列技术解决方案，在保证个人信息全生命周期的增强保护和个人信息处理活动规范化的基础上，实现保护个人信息权益、提升数据流通、共享与开放、促进个人信息合理开发利用的目的。金融隐私保护技术则是指应用到金融领域、保护个人金融信息的隐私保护技术。据中国人民银行《个人金融信息保护技术规范》，个人金融信息包括账户信息、鉴别信息、金融交易信息、个人身份信息、财产信息、借贷信息和其他反映特定个人金融信息主体某些情况的信息。

隐私保护技术并不是一种单一的技术，它是一套包含人工智能、密码学、数据科学等众多领域交叉融合的跨科学技术体系，在不泄露原始数据的前提下对数据进行分析计算，保障数据在流通与融合的过程中实现“可用不可见”，从而为数据流通与融合业务提供新的发展路径[3]。一般认为，隐私保护包含两个目的：1）如何保证数据应用过程中不泄漏隐私；2）如何更有利于数据的应用。当前，隐私保护领域的研究工作主要集中于如何设计隐私保护原则和算法更好地达

到这两方面的平衡。基于隐私保护技术将实现数据价值的流通，促进企业数据的合法合规应用，激发数据要素价值释放，进一步培育数据要素市场。

## （二）金融隐私保护研究必要性

目前我国金融用户隐私保护形势严峻，存在大量银行数据泄露、保险数据泄露、平台金融数据泄露、网贷业务及大数据风控乱象等问题。国内金融隐私数据频频泄露，存在金融隐私保护法律不完善、金融隐私信息买卖市场需求巨大、新技术结合金融应用发展的步伐较快、企业经营对隐私信息保护不够重视等问题。

近期，“数据二十条”发布，提出坚持促进数据合规高效流通使用、赋能实体经济这一主线，以充分实现数据要素价值、促进全体人民共享数字经济发展红利为目标，是数据要素市场治理的顶层设计，具有极强的制度导向性和政策倾向性，为我国数据要素市场举旗定向。数据合规是一个以法规为依据的法律行为判断问题。金融机构采购或应用数据时，应注意数据合规相关要求。目前，从国家、地方、金融监管政策以及相关标准逐渐完善，金融领域数据要素安全流通的边界逐渐清晰，但仍有部分环节细则尚待出台。同时，“数据二十条”也拓展了数据合规的基本内涵，不仅仅是网络安全法、数据安全法和个人信息保护法的范围，而应当将视野拓宽到数据基础制度，从数据生产到销毁，强化确权、推动流通，构建安全可控的治理体系。

因此，亟需从顶层设计，统筹谋划，从法律、监管、技术防护等多个方面精准施策以确保金融产业健康有序发展。

**1．金融隐私保护是各金融主体实现数据应用的重要安全保障**

（1）针对个人金融信息主体方，保护其合法权益

个人金融信息是个人信息在金融领域围绕账户信息、鉴别信息、金融交易信息、个人身份信息、财产信息、借贷信息等方面的扩展与细化，是金融机构在提供金融产品和服务的过程中积累的重要基础数据，也是个人隐私的重要内容。存储在各金融机构的个人金融信息因为缺乏相应的加密、保密和跟踪审计措施，在数据采集、整理、加工、存储、传输、使用过程中容易被窃取、篡改和非法使用。并且，而随着互联网金融新业态、新模式的不断出现，个人金融信息的安全威胁也更加复杂。如果这些信息被非法企业和个人作为牟利的手段非法买卖和使用，将严重侵犯信息主体的合法权益。

（2）针对金融机构及其他数据应用方，平衡数据价值与安全两难问题，充分释放数据价值

当前，金融隐私防护领域没有建立起统一的技术防护架构以保护金融行业个人信息。个人金融信息被中心化存储在不同的金融机构，数据碎片化现象严重，形成了明显的“数据孤岛”。数据在完整性、一致性和规范性上缺乏统一，数据不能及时互联互通和共享，导致金融机构间无法对各方数据进行有效融合与利用，造成了各方的数据资源和价值无法

充分被发掘的现状。为了解决在数据安全隐私前提下的数据融合与价值发掘问题，隐私保护计算技术实现了数据的“可用不可见”，有助于在保护各方数据隐私、确保业务合法合规的前提下，进行多方数据联合查询、联合运算、联合建模、联合预测等业务探索，赋能金融机构业务场景，提高营销和风控等能力。

（3）针对金融监管方，维护金融市场稳定

个人金融信息面临网络安全风险和考验。近年来，由于各金融机构间隐私信息保护策略和防护能力存在较大差异，导致高价值的个人金融信息安全保护不力，使得个人金融信息及其保存机构成为黑客攻击的重点目标，金融信息泄漏成为金融稳定的重要威胁。根据中国互联网协会发布的《网民权益保护调查报告(2015)》，78.2%的网民的个人身份信息、63.4%的网民的网络金融交易记录曾被泄露过。近年来，每年发生金融隐私泄露事件大约以 35%的速度在增长，有公开报道或记录 2016 年 1093 起，2017 年 1511 起，2018 年 1967 起，2019 年 2300 余起[2]。个人金融信息一旦泄露，不但会直接侵害个人金融信息主体的合法权益、影响金融机构的正常运营，甚至可能会带来系统性金融风险。相比欧美国家，我国隐私保护体系建设起步相对较晚，加之近年来各类新技术在金融行业迅速广泛应用，由此带来的金融隐私保护问题日益凸显。因此，加强个人金融信息安全管理，将有利于金融市场监管，维护国内金融市场的稳定。

### 2. 已有金融隐私保护技术发展不成熟，给金融数据应用带来了不可控的治理风险

随着“数据二十条”的发布，隐私计算作为一项解决数据安全与数据应用两难问题的关键技术措施，成为业界技术创新的焦点。隐私计算通过密码学技术，计算过程只交互秘密分享的碎片数据或者建模的梯度计算参数不涉密信息等，可实现数据二十条中“原始数据不出域、数据可用不可见”的要求。通过多方计算、联邦建模等技术，生成模型及提供 API 接口等 SaaS 服务，可满足数据二十条要求中的“以模型、核验等产品和服务等形式向社会提供”指导思路。通过密码学手段实现不暴露原始数据情况下的模型结果的输出，可实现数据二十条中的“持有权、加工使用权、经营权”的三权分离要求。由此可见，隐私计算技术可拉通多方数据的融合，打破数据与技术孤岛，实现数据三权分离，促进数据要素安全高效流通，由此实现数据二十条中的“推进互联互通，打破数据孤岛”的愿景，具有非常大的应用潜力。

然而，数据合规不仅仅是技术问题，更是法律、制度问题。隐私计算只是一种实现数据流通与共享的一种技术载体。受限于两大问题，隐私计算无法成为数据要素流通的终极解决方案：首先，使用了隐私计算，不等于合规，隐私计算常用的安全求交、联合统计、联合建模等，所使用的数据涉及个人的，属于个人信息，仍然需要采取告知、同意等合规措施落地；其次，隐私计算使用加密方法对原始数据进行保护，

无论是采用同态加密的数据还是多方安全计算（秘密分享）的碎片数据，都可以进行解密或复原。所以，隐私计算不等于匿名化，不符合《个人信息保护法》的匿名化要求。

因此，基于隐私计算的数据流通共享技术在金融领域的应用仍然面临数据安全、个人信息保护方面的合规风险。同时也面临着隐私技术产品不合格，不能满足相关技术要求，没有通过相关机构检测的风险；隐私技术是多种技术的集合，在相应的场景没有匹配对应隐私技术方案的风险；没有按照国家要求在金融数据相关法规和要求进行合法的数据全生命周期管控、应用与治理的风险等。

### （三）金融隐私保护技术要求及应用场景

#### 1. 金融隐私保护技术要求

《个人金融信息保护技术规范》根据信息遭到未经授权的查看和未经授权的变更后所产生的影响和危害，将个人金融信息按敏感程度从高到低分为 C3、C2、C1 三个类别[4]，如表 2 所示。另外，两种或两种以上的低敏感度类别信息经过组合、关联和分析后可能产生高敏感程度的信息，同一信息在不同的服务场景中可能处于不同的类别，应依据服务场景以及该信息在其中的作用对信息的类别进行识别，并实施针对性的保护措施。

表 2 个人金融信息等级类别

| 类别 | 等级 | 描述 | 说明 |
| --- | --- | --- | --- |

| | | | |
|---|---|---|---|
| C3 | 高敏感 | 主要为用户鉴别信息 | 直接鉴别信息，如银行卡磁道、验证码、有效期、密码、支付交易密码、账户登录密码、个人生物识别。 |
| C2 | 中敏感 | 主要为可识别特定个人金融信息主体身份与金融状况的个人金融信息，以及用于金融产品与服务的关键信息。 | 账号、证件类型、证件信息、手机号、登录用户名、辅助鉴别信息、动态口令、短信验证码、密码提示等，余额、流水、理赔、照片、影像、家庭住址等。 |
| C1 | 低敏感 | 主要为机构内部的信息资产，主要指供金融业机构内部使用的个人金融信息。 | 开户时间、开户机构、支付标记，其他。 |

个人金融信息生命周期主要包括数据收集、数据传输、数据存储、数据使用、数据删除以及数据销毁等过程，其生命周期各阶段对安全隐私保护的技术要求也有所不同。在数据使用阶段由于所涉及的过程比较复杂，具有较大的不确定性，所以各细分阶段对数据隐私保护的要求也比较多。针对该情况，隐私计算作为一种保障数据使用过程中的隐私保护技术，在金融行业受到更广泛的关注，个人金融信息全生命周期安全隐私技术要求如图 1 所示。

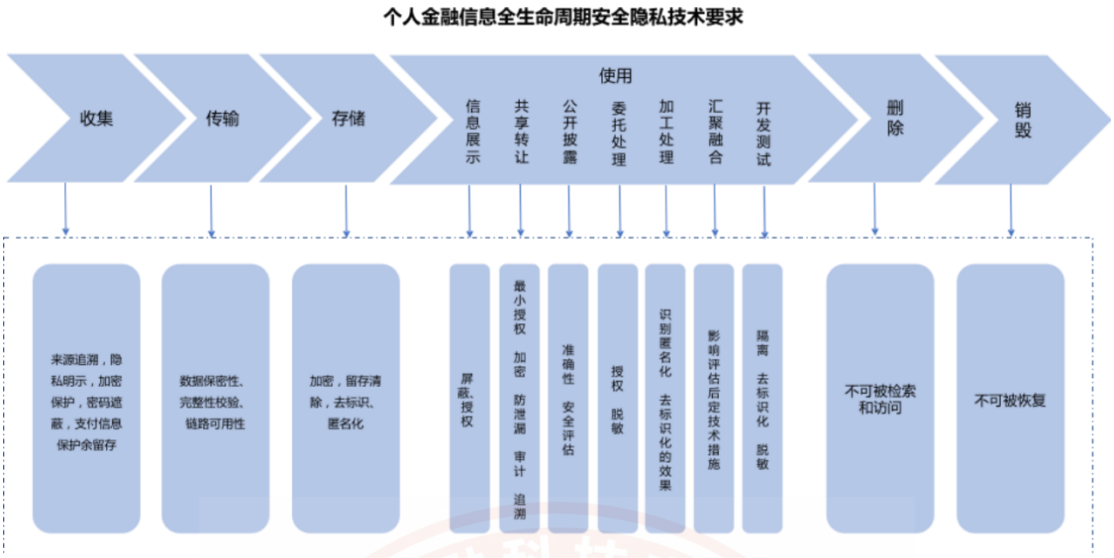


图 1 个人金融信息全生命周期安全隐私技术要求

## 2. 金融隐私保护技术应用场景

随着数字科技的不断演进，新技术与金融融合应用的领域也越来越广泛，包括人工智能、大数据、生物识别、移动互联网、区块链、物联网等几大类。数字新技术的应用，在推动金融产业创新、增加便民性、提升工作效率的同时，也给金融用户隐私保护带来了更多的风险与挑战。比如，基于人工智能和生物识别技术的人脸识别支付面临人脸仿冒的风险；各类金融 App 存在高危漏洞、被植入后门程序以及隐蔽收集用户信息的隐私风险；物联网、大数据及云计算技术同样会给金融隐私带来各类威胁，尤其是数据存储服务器常常是黑客攻击的重点目标。

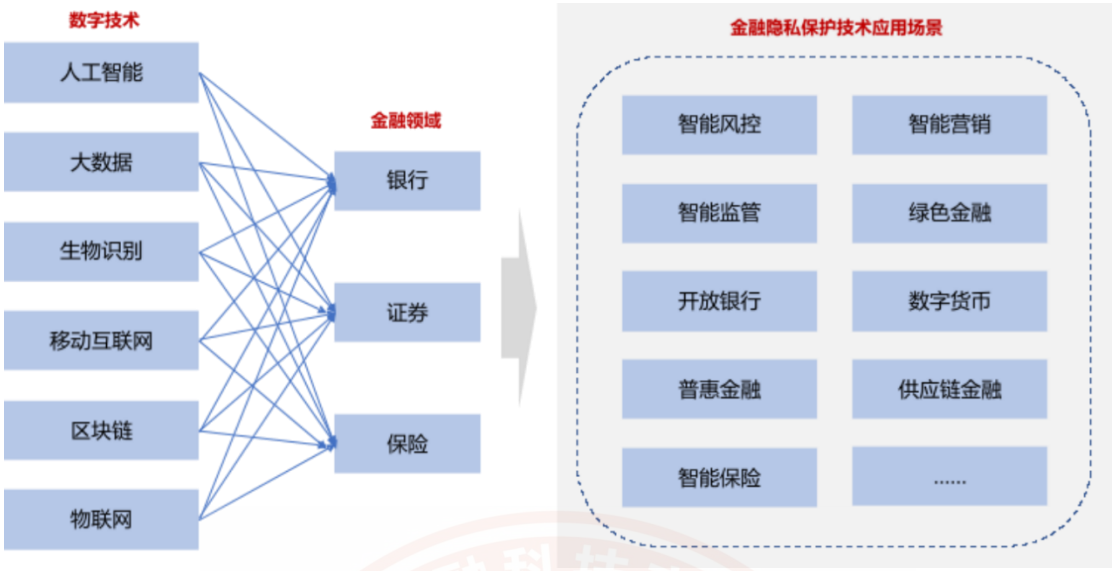


图 2 金融隐私保护技术应用场景

因此，隐私保护技术在金融领域的场景应用也主要围绕上述数字新技术在银行、证券、保险三大应用领域的融合应用场景展开，包括智能风控、智能营销、智能监管、绿色金融、开放银行、反洗钱、数字货币、普惠金融、供应链金融等场景，如图 2 所示。以金融机构联合风控和智能营销场景为例，其对隐私保护技术都存在迫切的需求。

联合风控场景下，随着金融机构消费贷快速发展，由于不需要抵押物等，个人借贷特别是消费贷的欺诈行为层出不穷，网络借贷的欺诈行为中以身份冒用最为常见。不法分子通过各种手段获取到用户的个人信息如手机号、身份证等，或者获取了学生的身份证、学籍、学历等信息，而后通过个人信息向多家金融机构网贷平台申请了多笔信贷业务。针对上述场景，一方面需要隐私保护技术实现个人信息不被泄漏，另一方面又需要对不法分子进行风控识别，通过对多家金融

机构网贷平台的联合隐私保护计算，获取到恶意诈骗的用户信息，实现对个人多头借贷场景的联合风控。

智能营销场景下，电信运营商和互联网公司拥有丰富的手机使用行为标签，如用户每月登陆理财 APP 的次数、每月理财类 APP 使用时长等。金融机构可以结合这些数据构建更精准的用户画像标签，是达到营销目标的有效手段。针对上述场景，一方面需要保证各机构合作时，不能泄漏用户的个人隐私信息；另一方面又需要利用运营商、互联网等多个数据源构建更立体的用户画像。因此，通过隐私保护计算基于多方数据构建标签和模型，实现数据安全和精准营销效果成为该场景的核心需求。

### （四）金融隐私保护技术国内外应用发展

金融隐私信息是个人信息的一种，因而它的保护也受到个人信息保护体系的影响。目前世界上个人信息保护主要有两种模式：欧盟综合性模式代表当今世界个人信息保护的最高标准；美国的分行业保护个人信息模式使得隐私保护显得格外复杂，各行业各领域个人信息的保护水平参差不齐。在各国隐私法规的强有力推动下，国内外数据安全相关技术和产品得到快速发展，逐步形成以“合规应用”为主线的安全细分领域。据 2019 年 11 月 Gartner 的一份预测报告指出，在 2023 年之前，全球 80%以上的企业将面临至少一项以隐私为重点的数据安全保护规定，并且在合规上的投入将突破 80

亿美元。由此可见，隐私保护技术及数据安全合规未来具有广阔的市场应用前景。

### 1. 国外应用发展

从隐私保护产业化发展历程来看，国外企业布局隐私保护领域相对较早，其前期更多聚焦于区块链和加密虚拟货币的应用，譬如集中于将多方安全计算应用于分布式密钥管理领域。2008 年，第一家专攻多方安全计算解决方案的技术厂商在丹麦成立，主要为商务合同、加密拍卖等场景提供安全方案。其后几年，国外大型科技公司纷纷在多方安全计算、联邦学习、可信执行环境、同态加密等隐私技术领域发力，形成技术引领。多方安全计算、联邦学习以及可信执行环境等隐私保护技术正作为高级匿名化方案，解决金融、医疗、交通领域数据在跨境，跨公司等业务场景中所面临的数据安全及隐私问题。发展至今，隐私保护相关技术越来越多地被应用于数据安全。鉴于解决个人信息安全问题更需要站在数据全生命周期治理角度，基于综合且全面的视角，因此可以利用隐私科技解决方案，通过实施隐私管理、隐私控制和以数据为中心的综合性安全能力来实现更为成熟的安全合规体系。在金融应用领域，2019 年世界经济论坛发布《下一代金融领域数据共享服务：使用隐私保护技术解锁新价值》阐述了在欧盟《通用数据保护条例》等隐私法规的限制下，金融数据如何通过隐私保护技术实现交流互通，从而使得机构、

客户和监管部门之间达成“允许数据共享、符合监管原则、保护客户隐私”的平衡关系[5]。

在具体的金融应用场景中，国外更注重于金融监管领域，其主要涉及金融反欺诈、政企联合风控和交易安全等场景。例如，英国金融行为监管局(FCA)协同科技公司 Duality 以及汇丰、花旗等银行针对金融反洗钱场景构建互通的隐私计算平台。利用同态加密技术，使金融机构间共享隐私数据，从而有效识别资金流动网络，对洗钱行为进行有效打击；丹麦科技公司 Partisia 利用安全多方计算技术，在确保数据隐私法规的前提下，联合使用银行、监管机构等隐私数据，有效识别存在资产风险的用户，协助政府对金融犯罪行为进行调查；Paypal 旗下的 Curv 针对区块链电子钱包提出安全多方计算解决方案，对非对称加密中的私钥进行秘密分享，保证交易过程中的私钥安全，避免私钥泄露带来的交易风险；Intel 与德国软件公司 Eperi 合作，基于 Intel SGX 可信硬件执行环境为银行构建基于云的数据分析平台，同时保证数据的安全和隐匿性。

隐私保护技术在国外金融领域中的实践呈现出了许多可能性，尤其是改变了公民、政府和企业之间的信任关系，重塑了数字经济体系结构。据 2019 年英国皇家学会发布的行业报告《实践中的隐私保护》，各国政府、企业及研究机构也不断展开对隐私保护技术的研究，英国阿兰·图灵研究所以及美国情报高级研究计划局(IARPA)都将隐私保护技术

作为战略重点项目[6]。面对当前不断增长的数据量，国外隐私保护技术呈现出多技术路径融合态势，通过不断优化协议算法以及推进软硬件协同发展，从而构建出一个完整的隐私生态闭环[7]。面对金融隐私领域问题，国外隐私保护应用从反欺诈角度入手，协同政企数据进而打击金融犯罪，同时不断探索新的市场需求，逐步推进隐私保护技术在金融领域的落地。

**2. 国内应用发展**

国内隐私保护技术产业化主要从 2018 年左右开始集中发展，行业巨头、大数据公司、金融（科技）企业、隐私计算初创企业为主要市场入局者，电信运营商发挥互联网能力优势，参与隐私计算相关标准制定及技术深入研究，一些大型咨询公司也将其咨询产品/服务向着隐私保护及其应用靠拢，成为隐私保护技术产业链中的重要服务提供商之一[8]。

随着数据要素市场以及数据交易流通的快速发展，隐私计算作为一种保障数据使用过程中的隐私保护技术，而备受国内企业的青睐，其隐私计算技术产品蓬勃发展，已经形成一定的优势，在金融、运营商以及医疗等领域已经有了深度的探索及应用。金融机构、互联网巨头和创业公司都在加大投入进行技术研发和应用落地，政府机关和监管机构也在积极推进相关法律法规、技术标准的制定工作。中国人民银行发布的《金融科技发展规划（2022—2025 年）》明确提到，在技术方面要积极应用多方安全计算、联邦学习、差分隐私、

联盟链等技术，探索建立跨主体数据安全隐私计算平台，在保障原始数据不出域前提下规范开展数据共享应用，确保数据交互安全、使用合规、范围可控，实现数据可用不可见、数据不动价值动。

当前，国内隐私计算主要应用在金融、通信、政务等行业，据信通院数据统计，金融行业招标占总体的 53%，招标方涉及银行、金融机构、保险、证券、银联、交易所、支付机构等。其中，在银行领域应用探索更广泛，占据金融领域的 35%，远远超过保险和证券[9]。隐私计算技术也被多家银行写入公司年报，例如，工商银行表示应用隐私计算技术创新小微企业贷后风险监控，农业银行表示推进利隐私计算技术在营销、风控等领域试点。在具体的应用场景上，联合风控和联合营销是当前国内金融隐私计算应用最为广泛的场景，如图 3 所示。

| | 场景 | 联合统计 | 联合查询 | | 联合建模及预测 | |
|---|---|---|---|---|---|---|
| | 算法 | 联合统计 | 安全求交 | 隐匿查询 | 监督模型 | 无监督模型 |
| 联合风控 | 贷前风控 | | √ | √ | √ | |
| | 贷后风控 | | | | √ | |
| | 信用评级 | | | | √ | |
| | 反欺诈识别 | | √ | √ | | |
| | 黑名单查询 | | √ | √ | | |
| | 合格投资者认证 | √ | | | | |
| 联合营销 | 供应链金融 | | | √ | √ | |
| | 纳新拓客 | | | √ | √ | √ |
| | 存量客户营销 | | | | √ | |
| | 客户画像 | | | | √ | |
| | 信用评级 | | | | √ | |
| | 个性化广告 | | | | √ | |
| | 名单共享 | | √ | √ | | |

图 3 国内隐私计算金融应用场景

针对隐私计算技术发展不成熟的问题，国内给出了可信隐私计算的概念及框架，以满足数据需求方，数据提供方和监管方等各方的需求，一般包含安全可证、隐私保护、流程可控、高效稳定、开放普适等基本特征。可信隐私计算要对全周期隐私信息进行保护，防止未经授权的个人信息泄漏、篡改和丢失。一方面，可以通过区间化、泛化/有效位截断、k-匿名等方法对个人身份标识信息、属性行为数据等进行标识脱敏处理，实现匿名化。另一方面，在计算过程中也可以对信息熵进行识别、度量和控制，满足不同场景、不同数据生命周期内对用户隐私信息保护的要求。基于此，国内隐私计算企业提出了“全匿踪”的隐私计算技术体系，利用匿名化技术，实现数据集的对齐、求交，并基于此完成数据的匿踪统计分析、联合建模、评估预测等操作，从而保护个人隐私信息不泄露。

## 三、隐私保护技术的基本体系

基于隐私保护技术的主要目的：如何保证数据应用过程中不泄漏隐私和如何更有利于数据的应用，隐私保护技术的基本体系通常包括隐私保护脱敏技术，隐私保护计算技术，隐私保护辅助技术三大类别，覆盖数据应用与流通的生命周期。数据发布和应用前，需要对隐私数据或敏感数据进行脱敏处理，避免用户隐私信息的泄露，同时保证用户数据的可用性。常见的隐私保护脱敏技术，包括数据去标识化、K-匿

名、L-多样性等，可以实现对发布数据的匿名保护。而在数据价值联合挖掘阶段，需要采用隐私保护计算技术，在不泄露原始数据的前提下对数据进行分析、计算、建模等，以数据“可用不可见”的方式进行联合应用。隐私保护计算技术主要包括多方安全计算、联邦学习、可信执行环境、差分隐私、可信密态技术等。此外，为保证数据应用过程的合规性和可信性，数据水印、区块链等技术作为隐私保护辅助技术也常常被采用，以增强数据应用的可追溯性、可验证性、可信性等。

### （一）隐私保护脱敏技术

#### 1. 去标识化

去标识化通常是指通过对个人信息的技术处理，使其在不借助额外信息的情况下，无法识别个人信息主体的过程。例如在涉及客户安全数据或者一些商业性敏感数据的情况下，对真实数据进行改造后再提供使用，如身份证号、手机号、卡号、客户号等个人信息都需要进行数据脱敏、去标识化处理。

信标委发布的《信息安全技术-个人信息去标识化指南》，列出了“去标识化”操作的常规思路，其中对“直接标识符”和“准标识符”的处理是关键。“直接标识符”，在 API 等各类技术文档中，经常使用类似“ID”这样的字段来表示，如 微信开放文档 中提及的 uid 和 unionid 就是典型的直接标识符，其典型特征就是（在一定范围内）具备唯一性，

可以就此识别某个具体的个人用户。而“准标识符”，只指可结合其他属性后可唯一识别个人信息主体的字段，如“男/女”“中文/english”“美国/日本”“汉族/回族”、出生日期、总收入等。

常见的去标识化技术有统计技术、密码技术、抑制技术、假名化技术、泛化技术、随机化技术等。

**2. K-匿名**

K-匿名(K-Anonymity)是 Samarati 和 Sweeney 在 1998 年提出的技术，该技术可以保证存储在发布数据集中的每条个体记录对于敏感属性不能与其他的 K-1 个个体相区分，即 K-匿名机制要求同一个准标识符至少要有 K 条记录，因此观察者无法通过准标识符连接记录。

K-匿名通常通过概括（对数据进行更加概括、抽象的描述）和隐匿（不发布某些数据项）技术，发布精度较低的数据，使得同一个准标识符至少有 K 条记录，使观察者无法通过准标识符连接记录。

K-匿名的具体使用如下: 隐私数据脱敏的第一步通常是对所有标识符列进行移除或是脱敏处理，使得攻击者无法直接标识用户。但是攻击者还是有可能通过多个准标识列的属性值识别到企业或个人。攻击者可能通过（例如知道某个企业的所在地区，采购量等）包含企业信息的开放数据库获得特定企业的准标识列属性值，并与其他大数据平台数据进行匹配，从而得到特定企业的敏感信息。为了避免这种情况的

发生，通常也需要对准标识列进行脱敏处理，如数据泛化等。数据泛化是将准标识列的数据替换为语义一致但更通用的数据。

### 3. L-多样化

L-多样性为了解决同质性攻击和背景知识攻击所带来的隐私泄露，Machanavajjhala 等人在 2006 年提出了 L-多样性(l-diversity)模型。

L-多样性(L-Diversity)是在公开的数据中，每一个等价类里的敏感属性必须具有多样性，即保证每一个等价类里，敏感属性至少有 L 个不同的取值，通过这样 L-多样性使得攻击者最多只能以 1/L 的概率确认某个体的敏感信息，从而保证用户的隐私信息不能通过背景知识，同质知识等方法推断出来。

L-多样性的常见统计方法包括:

(1) 不可区分 L-多样性(Distinction L-Diversity): 在同一个等价类中至少出现 L 个不同的敏感属性值;

(2) 基于概率的 L-多样性(Probabilistic L-Diversity): 在一个类型中出现频率最高的值的概率不大于 1/L;

(3) 基于熵的 L-多样性(Entropy L-Diversity): 在一个等价类中敏感数据分布的熵至少是 log(L);

(4) 递归 (C,L)-多样性(Recursive (C,L)-Diversity): 通过递归的方式，保证等价类中最经常出现的值的出现频率不要太高;

(5) 递归(C1,C2,L)-多样性(Recursive(C1,C2,L)-Diversity):通过递归的方式,保证等价类中最经常出现的值的出现频率不要太高, 同时还保证了等价类中频率最低的敏感属性出现的频率不能太低。

## (二)隐私保护计算技术

### 1. 多方安全计算技术

安全多方计算(Secure Multiparty Computation, MPC)问题首先由图领奖获得者姚期智教授于 1982 年提出,安全多方计算协议作为密码学的一个子领域,其允许多个数据所有者在互不信任的情况下进行协同计算,输出计算结果,并保证任何一方均无法得到除应得的计算结果之外的其他任何信息。多方安全计算经过近几十年的发展,形成了多个不同的基础协议,包括秘密分享、混淆电路、同态加密、不经意传输、零知识证明等。

(1)秘密分享

秘密分享(secret sharing,SS)是一种共享秘密的技术,由 Shamir 在 1979 年提出,通过某种方法将秘密数据进行随机拆分,并将这些碎片化数据分发到 N 个参与方,每个参与方都只能拿到一个随机数据,单个参与方或少数几个参与方无法还原出原始秘密数据,但这些由不同参与方持有的随机碎片可以本地进行计算,只有把各个参与方的碎片计算结果汇聚起来才能恢复到正确的完整计算结果。秘密分享能在计

算前后始终保持秘密在参与方之间分享，并且在计算过程中不会泄漏参与方的敏感数据。

在某些应用秘密分享不需要联合所有参与方才能恢复计算结果，可采用（t,n）门限秘密分享（threshold secret sharing）方案， 在有 n 个参与方的场景下，这种方案允许 t 个参与方联合即可以将秘密数据恢复，但任何不多于 t-1 个参与方均无法将秘密数据解开。

秘密分享的实现技术包括: 基于超平面几何的秘密分享，例如 Blakley 方案和 Brickell 方案; 基于插值多项式的秘密分享，例如 Shamir 门限秘密分享方案; 基于中国剩余定理的秘密分享，例如 Mignotte 方案和 Asmuth 方案等。

（2）混淆电路

混淆电路(Garbled Circuit，GC)是姚期智教授在 20 世纪 80 年代提出的安全计算概念。

混淆电路通过布尔电路的观点构造安全函数计算，达到参与者可以针对某个数值来计算答案，而不需要知道他们在计算式中输入的具体数字。在这里关键词是“电路”，实际上所有可计算问题都可以转换为各个不同的电路，例如加法电路，比较电路，乘法电路等。而电路是由一个个门（gate）组成，例如与门，非门，或门，与非门等。

混淆电路里的多方的共同计算是通过电路的方式来实现，例如两个参与方要进行多方计算，他们首先需要构建一个由与门，或门，非门，与非门组成的布尔逻辑电路，每个

门都包括输入线，输出线。混淆电路则通过加密和扰乱这些电路的值来掩盖信息，而这些加密和扰乱是以门为单位，每个门都有一张真值表。计算发起方用密钥加密真值表，并把表打乱后发给计算参与方，通过这种这加密+打乱的过程，达到混淆电路的目的。而计算参与方在接收到加密表后，根据收到的加密真值表，混淆的输入，及自己的 Key，对加密真值表的每一行尝试解密，最终只有一行能解密成功，并提取相关的加密信息。最后计算参与方将计算结果返回给计算发起方。在整个过程大家交互的都是密文或随机数，没有任何有效信息泄露，在达到了计算的目的，同时达到了对隐私数据数据保护的目的。

（3）同态加密

同态加密(Homomorphic Encryption, HE)指的是能够直接使用密文进行特定运算的加密技术。在同态加密计算过程中，无需密钥即可实现操作，而结果仍需密钥解密从而变为明文，在解密后，得到与明文计算相同的结果。

同态加密素有隐私计算的“圣杯”之称。作为面向数据应用侧的密码算法，可实现在数据加密的状态下，密文数据与原始数据仍具备完全一致的计算能力，即密文数据无需解密便可以进行计算使用。

同态加密技术仍旧处于早期成熟阶段。相比于明文计算，同态加密后的计算流程算力消耗巨大同时数据吞吐量较低。此外，由于同态加密后的数据体积增大、将会挤占网络带宽。

因此，诸如全同态加密在运行速度随着数据量增多、计算耗时急剧增多等问题仍有待研究持续推进。

同态加密可直接对密文进行分析、检索。因此在达成保护隐私的前提下，还能实现某些数据操作。同态加密实现了数据使用过程(Data in use)中的加密，适用于部分诚信和恶意环境中，以保护数据安全与隐私。目前适用场景有医疗数据加密、顾客数据分析、多个机构间客户的交叉分析等。

（4）零知识证明

零知识证明(Zero-Knowledge Proof，ZKP)是由S.Goldwasser、S.Micali及C.Rackoff在20世纪80年代初提出的一种基于密码学的验证技术，旨在证明者可以不向验证者泄露任何关于被验证信息的信息的情况下，使得验证者相信某个结论是正确的。

零知识证明通常包括两个参与方，宣称某一结论是真的证明者(Prover)和确认结论是真的验证者(Verifier)。零知识证明的实现机制分为交互式和非交互式两种类型。交互式零知识证明要求证明者和验证者之间通过交互来完成对结论的验证，这种交互式输入通常以一个或多个挑战的形式进行，使得证明者的回答能够使验证者确信结论是真的。非交互式零知识证明，证明者创建一份证明，任何使用这份证明的参与方都可以进行验证，非交互式零知识证明通过大幅减少通信步骤而提高效率。零知识证明的典型算法有ZK-snark、ZK-stark、Plonk、Bulletproof、Aztec等。

零知识证明主要应用于验证场景，借助零知识证明“不泄露信息”的特性，可以在不泄露交易信息的情况下证明区块链上资产转移的有效性，可以用于验证数据的真实性、算法分析结果的可靠性等。

**2. 联邦学习技术**

联邦学习（Federated Learning，FL），又名联邦机器学习，联邦计算。联邦学习是在原始数据不出本地的前提下，通过模型的流通与处理来完成多方联合的机器学习，得到聚合的训练结果。联邦学习的参与方一般包括数据方、算法方、协调方、计算方、结果方、任务发起者等角色。

按照数据特征和样本的分布情况，联邦学习可划分为横向联邦学习、纵向联邦学习和联邦迁移学习。其中，横向联邦学习适用于参与方有重叠的数据特征，而拥有的数据样本不同的情况。经典的例子是谷歌的 Gboard 键盘，当用户在手机上输入一个词， 综合广大其他用户的输入历史利用横向联邦学习模型可以预测出用户想要输出的下一个词。

纵向联邦学习适用于参与方拥有重叠的数据样本，而数据特征不一致的情况。典型的例子是两家公司提供完全不同的服务，但在用户群体上有非常大的交集时，它们可以在各自不同的特征空间上协作，得到一个更好的机器学习模型。

联邦迁移学习是指各参与方的数据集之间只有少量的重叠样本和特征，或者数据集的分布情况差别很大时，利用

源领域和目标域之间的相似性，实现跨领域知识迁移的学习技术。

联邦学习的架构通常支持通用硬件。而隐私计算的其他分支可执行环境是需要特定硬件的。在多方参与的复杂场景下，联邦学习这种硬件无关的特性减少了多方间达成一致的沟通成本。

当前有部分行业的用户执行严格的数据不出域要求，不仅原始数据不能出域，加密后的密态数据也不能出域。在这种情况下，联邦学习“数据不动模型动”的特点能够契合用户的要求，化解数据孤岛难题。

由于联邦学习的在数据安全流通中的显著作用，已被广泛应用于医疗领域、金融领域、智能手机、智能汽车等诸多领域，在保证用户隐私的前提下用于联合数据挖掘和建模。

### 3. 可信执行环境

可信执行环境(Trusted Execution Environment,TEE)是计算平台上由软硬件方法构建的一个安全隔离区域，可保证在安全区域内部加载的代码和数据在机密性和完整性方面得到保护。其目标是确保一个任务按照预期执行，保证初始状态的机密性、完整性，以及运行时状态的机密性、完整性。隔离区域可以是除了正常操作系统的独立系统（如 Trust Zone），进程地址空间（如 Intel SGX）或独立的虚拟机（如通过 AMD SEV 或 Intel TDX 保护的虚拟机）。

针对TEE的相关概念及规范定义，各家软、硬件厂商结合自己的基础架构形态具体实现各不相同。虽然在技术实现上存在差异性，但是仍可抽象出TEE的共同技术特点。

具体而言，TEE存在以下技术特点:

（1）隔离性

隔离是指基于安全芯片实现的内存隔离，由安全芯片从动态内存区域隔离出一片供可信程序执行的物理内存，任何特权应用（如操作系统、Hypervisor）都无法访问这段内存地址空间。X86架构的隔离机制从Intel 80286处理器开始，Intel提出了CPU的两种运行模式，并且逐步衍生出后来的不同的特权界别，再后来提出了安全区域更小的SGX机制实现可信执行环境TEE。同样的，ARM架构通过Trustzone技术实现了相关软硬件的隔离性，实现安全世界与非安全世界的隔离。TEE 通过隔离的执行环境，提供一个执行空间，该空间有更强的安全性，比安全芯片功能更丰富，提供其代码和数据的机密性和完整性保护。

（2）可证实性

可证实往往指的是 TEE 中的远程证实，一方面，由于TEE 的用户往往在远程提交计算任务，TEE 需要向用户证明其所在的软件环境的安全性；另一方面，由于 TEE 中运行的程序是在隔离的内存区域进行执行，外界的应用无法随意进行访问，这就需要 TEE 环境本身向用户证实在其中运行的代码确实是合法的、经过用户认证的代码。TEE 的远程证

实包含软件状态度量以及公钥签名两个步骤，即首先通过收集软件的运行环境信息生成证据，然后通过绑定在硬件中的证实密钥进行签名，远程用户可以通过验证签名来度量证据的合法性。

（3）软硬协同性

虽然标准定义可以通过软件方式或硬件方式实现 TEE，但实际生产场景下，行业内更多通过软硬结合的方式进行安全性的保障与支持。

（4）富表达性

TEE 与单纯的安全芯片或纯软件的密码学隐私保护方案相比支持的上层业务表达性更强，由于只需要定义好业务层面隐私区域和非隐私区域的逻辑划分，而不会对定义隐私区域内的算法逻辑的语言有可计算性方面的限制（图灵完备的）。同时由于 TEE 已经提供了安全黑盒，安全区域内数据无需进行密态运算，从而支持更多的算子及复杂算法。

可信执行环境 TEE 技术因可支持多层次、高复杂读的算法逻辑实现、 运算效率高和可信度量的方式保证 TEE 的运行逻辑的可信及可度量性的特性，受到业界一致认可，越来越多可信执行环境 TEE 的开源框架和产品踊跃而出。

### 4. 差分隐私技术

差分隐私(Differential Privacy，DP)是 Dwork 在 2006 年提出的一种隐私计算技术，目的是提供一种技术保证在数据库查询和统计时，能够最大化数据查询的准确性，同时最

大限度来减少识别单个记录的机会，也就是让查询结果对于数据集中单个记录的变化不敏感，从而攻击者就无法通过加入或减少一个记录，观察查询结果的变化来推测个体的具体信息。

差分隐私中一个关键概念是相邻数据集，假设给定两个数据集 D 和 D’，如果它们有且仅有一条数据不一样，那么这两个数据集可称为相邻数据集。那么如果对于一个随机算法 A 如果其分别作用于两个相邻数据集得到的两个输出分布式难以区分的，那么这个算法就被认为达到差分隐私的效果。

差分隐私主要通过在输入输出中增加适量的随机噪音以确保修改数据集中一条个体记录不会对统计结果造成显著影响，差分隐私的主要噪声机制包括：拉普拉斯噪音(Laplace Noise)，高斯噪音(Gaussian Noise)，指数机制等。基于拉普拉斯机制和高斯机制的差分隐私是通过向查询结果中加入服从拉普拉斯分布和高斯分布的噪声来实现差分隐私保护，适用于数字型结果的差分隐私。基于指数机制的差分隐私，则采用指数机制来引入一个打分函数来实现差分隐私保护，也就是当接收到一个查询后，不是确定性地输出一个结果，而是以一定概率值返回结果，这个概率值是由打分函数确定，指数机制适用于离散型结果的差分隐私。

差分隐私根据数据收集分析中的保护对象的差异可以划分为中心化差分隐私和本地化差分隐私两种类型。中心化差分隐私通常具有可信第三方，第三方利用收集到的数据进

行统计分析，因此差分隐私主要是保护数据收集分析后的结果发放过程，因此差分隐私保护机制是运行在可信第三方上。本地化差分隐私则认为第三方不可信，因此本地差分隐私主要保护数据上传到第三方的过程，因此差分隐私机制是运行在各个数据提供方本地。

差分隐私常用于多方参与的统计分析中，也用于联邦机器学习算法，在中间参数交互过程通过采用差分隐私，避免恶意参与方通过中间结果反推出训练集中的个体特征信息，增强联邦算法的安全性。

**5. 可信密态技术**

现有技术基本能够满足特定场景的隐私计算需求，但是对于目标数据量大、参与方多、场景多样的数据中心，则显现出不足。经发现和研究，将可信执行环境和密码协议结合，能够很好地相互扬长避短，获得更为优秀的综合性质。将这一想法深化和总结，形成了一个新的密态计算路线，即可信密态计算(Trusted-Environment-based Cryptographic Computing，TECC)。

TECC 的核心理念是在可信环境中将数据以密态的形式进行计算，经过进一步研究分析，密态能够消除可信环境的一些安全隐患，例如侧信道、供应链攻击等。而可信环境能够提供防恶意敌手、防合谋攻击等能力，在可信计算技术的加持下，密码协议可以有效地轻量化，并且由公网交互改为

内网（或高速网）交互，大幅提高性能。这样，可信环境和密码协议相互弥补了对方的缺点。

TECC 典型的示例图如图 4 所示:

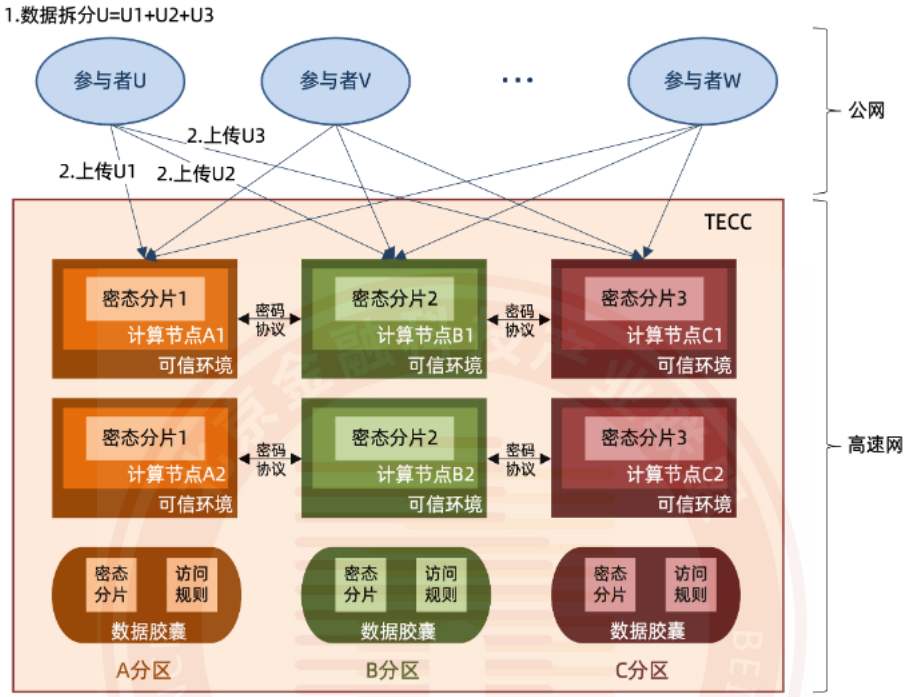


图 4 可信密态计算示例图

数据提供者在本地将数据拆分成多个密态分片数据。单个密态分片数据不泄露原始数据的任何信息。继而并将每个分片数据传递给不同分区的可信计算节点。

每个分区的可信计算节点只接触一份密态分片数据，不接触任何明文，多个分区的可信计算节点通过密码协议（MPC、安全联邦学习等）完成目标计算，单一分区被攻破没有任何安全隐患。

可信计算节点使用可信计算技术（TEE/TPM/全栈可信等），保证运营者无法进行窥探。

密码协议的同一个角色由一个可信计算分区集群承担，可以进行并行化加速。

数据采用密态胶囊形式进行存储，包括数据密态分片，以及与其绑定的访问规则，运营者无法滥用密态数据。

如图 5 展示了 TECC 的系统结构，包含参与方层、调度管理层、分布式计算层、运行环境层、数据层和硬件层。

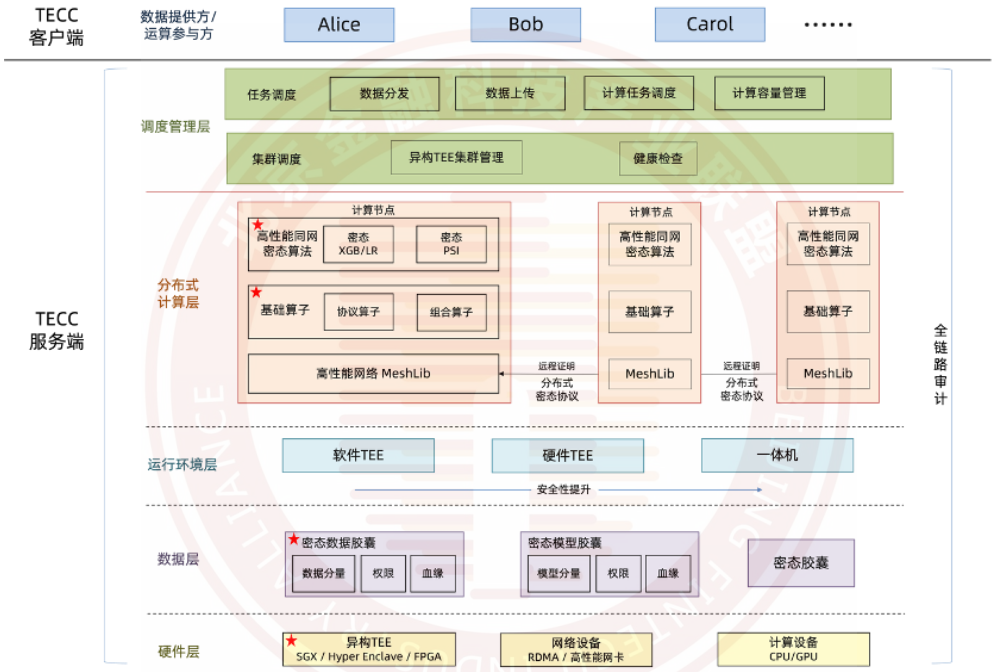


图 5 可信密态计算系统结构

最上层是参与方层，包含数据提供方和数据使用方。TECC 不仅可支持任意个数的参与方，而且也不限制参与方之间的数据分割形式，数据分割形式指的是每一方拥有全部数据的哪一部分。

次上层表示的是调度管理层。数据提供者的密态数据到达 TECC 服务端后，由调度管理层调度至多个分布式计算节点，为了完成该目标，调度管理层还会实时评估各个计算节

点当前的容量，根据任务的需求匹配合适的计算节点。调度管理层还包括集群调度功能，对所有计算节点进行健康检查、扩缩容，保证整个计算集群的高可用。

位于中间层的是分布式计算层，该层使用密码协议完成目标计算。该层的底部是 MeshLib，MeshLib 提供了高性能分布式网络通信能力。中部是基础算子层，内部又包含协议算子和“组合算子”。协议算子指的是底层密码协议中会包含的功能，例如乘、加、截断、与、异或等。“组合算子”指的是在基础算子上使用特定算法搭建上层功能，比如使用 Goldschmidt 算法实现除法、使用 Chebyshev 实现 Sigmoid 等。基于基础算子，可以构建高性能的密态算法，包括隐私求交（PSI）、机器学习（LR、XGBoost、NN 等）、数据密态分析（Select、Join、Sort、Groupby 等）。

运行环境层为分布式计算层提供安全的隔离环境，防止攻击者或者恶意的运营者窥探密态计算的过程。TECC 支持多种可信运行环境，可以是软件 TEE、硬件 TEE、一体机等，也可以是它们的叠加，实际应用可以根据需要的安全程度进行选择。下一层是数据层。TECC 通过密态胶囊将密态数据、权限、血缘绑定。数据的密态性质是基于可信环境保证的，TECC 的运营方也无法窥探数据。权限记录了何种逻辑、哪个参与方才能使用该密态数据，TECC 的运营方也不能滥用密态数据。血缘记录了所有中间结果的原始数据来源，保证数据提供方的控制权能够延伸到他数据产出的所有计算结果。

最底层是硬件层。其中包括 TEE 环境，例如 SGX、HyperEnclave 等，除此之外，硬件层还包括高性能网络设备和计算设备。

### （三）隐私保护辅助技术

#### 1. 数据水印

数据水印是一种将标识信息（如数据归属信息、机构、提供方 id 等）通过一定的规则与算法隐藏在结构化数据中的技术，隐藏后的数据使用价值可以保持不变。数据水印可以在数据泄露后进行追踪溯源，为金融、政务、互联网等数据密集型和信息敏感性行业在数据共享时建立有效的溯源机制，一旦发生数据泄露，可精准溯源到操作数据用户身份、作业及泄露范围和渠道。

数据水印通常具有高隐秘性、高安全性、可监测性、高鲁棒性、高仿真性等特点。其中高隐秘性，是由于数据水印的标记信息通常不易被发现和察觉，最大程度保证原数据可用性；高安全性是被嵌入的数据水印很难被伪造或者修改；可检验性，是用户可通过一定的提取算法高效提取完整数据水印信息，以此宣告数据拥有方对其的版权；高鲁棒性，是数据库中的数据受到攻击方恶意攻击，数据持有方仍然能够精准完整提取出数据水印信息，以此验证其版权；高仿真性，使数据在增加水印后仍可以保持数据原始特征，业务规则关联性，保证开发、测试、培训以及大数据利用类业务不受到数据水印影响。

数据水印包括有鲁邦水印、易损水印、标准水印等类别。鲁棒水印，通常用于数字化图像、视频、音频或电子文档的版权保护，将代表版权人身份的特定信息，如一段文字、标识、序列号等按某种方式嵌入在数字产品中，在发生版权纠纷时，通过相应的算法提取出数字水印，从而验证版权的归属，确保著作权人的合法利益，避免非法盗版的威胁；易损水印，又称脆弱水印，通常用于数据完整性保护，当数据内容发生改变时，易损水印会发生相应的改变，从而可鉴定数据是否完整；标注水印，通常用于标示数据内容。

**2. 区块链**

区块链(blockchain)是一个分布式账本系统，具备数据可溯源、不可篡改、智能合约自动执行和去中心化等技术特点。区块链由于采用了良好的隐私保护策略和数据加密保护技术，能够作为一种辅助技术很好地解决金融场景的信任问题和隐私保护问题。

区块链在金融数据隐私保护上可实现以下支撑能力：区块链可以实现上链前数据真实性交叉验证，上链后数据实现不可篡改性和可追溯性；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；通过数字身份实现数据的确权，并将数据身份与智能合约等关联，实现数据管理、转移和交易；通过协同机制、激励机制的设置与共识，区块链可以连接好数据的提供方和数据的需求方，以及监管单位和其他的周边单位，各方在整个链上进行整体数据传输的合作，

做到所有账务、交易、数据都是联盟当中互信，互相认可，无缝流转；此外，区块链的去中心化和不可篡改性，可以为数据的跨机构协作提供可信存证能力。

## 四、隐私保护技术在个人金融信息保护中的应用

### （一） 隐私保护计算环境

隐私保护计算环境，也就是可信执行环境，是基于CPU硬件构建安全环境，在硬件安全环境中的运算是明文计算，所以计算性能比较高。根据业界的普遍数据来看，基于可信执行环境的计算性能跟明文计算仅有20%的性能损耗，基本可以处理跟明文数据同等数据量级的数据。但可信执行环境需要集中式计算，需要数据集中到某一个可信执行环境节点中计算。所以如果相关场景的数据量较大，且可以集中计算，适合采用多头借贷反欺诈、反电信诈骗等场景中较大规模数据的安全计算。

#### 1. 可信执行环境在多头借贷反欺诈中的应用

近年来，随着互联网及大数据技术的蓬勃发展，互联网金融和数字化金融促进金融信贷飞速发展的同时也催生了一系列风险问题，其中多头借贷等恶意行为发生率随之升高，问题日益突出。多头借贷是指单个借款人向 2 家或 2 家以上的金融机构提出借贷需求的行为。随着借款人的负债率升高，还款压力逐渐增大。有相关数据显示，多头借贷用户的信贷

逾期风险是普通客户的 3-4 倍，贷款申请者每多申请一家机构，违约的概率就上升 20%。这对金融机构的正常运转和金融稳定性带来极大隐患。一般来说，一家机构的金融业务数据很难满足多头借贷的风控需求，但由于数据的零散分布和数据安全保护措施，使得数据孤岛现象严重。

基于可信执行环境的隐私求交算法，可以利用可信执行环境中计算性能高的优势。通过可信执行环境构建可信计算节点，实现跨机构间的数据联合挖掘。通过汇聚多方数据到 TEE 节点，可以实现基于对三方及以上数据复杂计算条件的联合运算。两方之间基于同态算法等的隐私求交，需要进行多次两两求交导致性能较低，采用可信执行环境方案，在进行三方以上的联合求交场景下性能可以显著提升。多个金融机构的消费贷用户信息以及外部数据源对行为特征数据维度的补充，多方进行联合统计、查询求交等相关运算，更好的分析客户综合情况，有效识别客户金融风险。如果同一个用户在短时间内向多个金融机构的消费贷产品都进行消费贷申请操作，则可以将该用户的欺诈风险等级提高。降低金融欺诈风险，提升机构的风控能力。同时面对新型欺诈形式和手段不断衍生的欺诈场景，可信执行环境 TEE 可以实现不同算法对适配 TEE 的自定义编程，根据需求调整查询和计算条件，从而满足不同业务场景下风控策略的差异化细分，增强识别欺诈行为的能力。

可信执行环境 TEE 的计算是基于软硬件的方式在 CPU 中构建一个安全区域，保证其内部加载的程序和数据在机密性和完整性上得到保护。并在计算后对建模过程中产生的数据进行销毁，基于 TEE 节点存在的物理隔离，任何一方均无法获取到原始的数据和计算过程数据，保证数据安全的情况下实现数据“可用不可见”。在各方不泄漏自己的查询条件，也不会被其他人留存本方的数据的前提下利用各机构和平台的数据进行联合构建反欺诈模型，提升反欺诈模型的识别性能，有效降低由于数据缺失和偏差导致的欺诈事件发生概率。

**2. 基于可信执行环境的反电信金融诈骗**

随着国际互联网、通信网络的发展，新型的网络消费方式为广大群众带来了实惠和便利，同时也给不法分子提供了可乘之机。黑色产业链利用新技术新业态也在同步更新升级犯罪工具。2021 年全国破获电信网络诈骗案件数量 39.4 万件，涉案金额达到 3291 亿。电信网络诈骗已经严重危害人民群众的财产安全和合法权益。

由于《数据安全法》《个人信息保护法》的陆续实施，合规监管要求趋严，加上存在技术壁垒，使得海量数据散落在众多机构和信息系统中，形成了“数据烟囱”，价值无法得到充分发挥。隐私计算技术的可用不可见、“可用并可控”“可控可计量”的综合解决方案，降低了金融数据与其他数

据的跨机构、跨行业的共享壁垒，使得数据可以依法合规地共享和融合应用。

基于可信执行环境技术的特点，联合运营商多维度数据（如语音通话时长、流量套餐包、通话对象等）和金融机构内部账户数据（资金往来，一进多出等）进行联合机器学习建模。在运营商和金融机构同时构建可信执行环境节点，双方均在 TEE 节点内完成数据的交互操作与联合运算，参与计算的节点都会经过可信执行环境的安全认证，在安全环境内基于内存做明文运算，训练和预测的效率更高。可信执行环境还可以对参与计算的算法进行一致性度量，使得参与运算的算法符合双方之前的约定。同时不同算法对适配 TEE 的自定义编程，可以实现不同的计算条件，更好地应对不同类型的电信金融欺诈模式。利用模型进行定期筛选具有电信金融诈骗风险的存量账户，用算法辅助风控决策，提高对犯罪特征的识别，实现对电信诈骗的精确打击。

### （二）隐私保护脱敏技术

数据脱敏，指对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护。可以在开发、测试和其他非生产环境以及外包环境中安全地使用脱敏后的真实数据集。数据脱敏后数据的某些字段已经不是真实数据，所以适合于不需要对这些数据进行进一步加工分析的场景，如脱敏数据在不同部门内的共享和汇集。

### 1. 基于数据脱敏技术的企业内部数据汇集

信息化时代，数据是一笔宝贵的资产。所有的公司，每天在进行大量的数据采集、存储、传输、共享，而且涉及的数据规模也越来越大。通过对海量用户数据的梳理、分析，实现价值的传递，但随之而来的数据泄漏的风险也与日俱增。

在行业随着大数据平台，数仓数据分析管理系统的建设，大量的数据需要进行汇集和分析，使得未经管控和脱敏的敏感数据泄露的风险的概率将大大增加。

因此需要通过脱敏技术来保证脱敏后数据的关联性与一致性，脱敏后的数据严格保留原有的数据关系与格式；例如身份证号在多个表中出现，需要保证这些数据经过脱敏后也是一样且可以保证身份证号对应的姓名、地址等一系列的关联；将相关的列作为一个组进行屏蔽，以保证这些相关列中被屏蔽的数据保持同样的关系，例如，城市、省、邮编在屏蔽后保持一致。大数据平台，数仓数据分析核心是对数据进行统计分析，用于相关分析报告或为各个业务系统服务。需要保证数据特征，脱敏后保证敏感数据的一定关联关系成为必然需求。

为保证获取的数据可以满足大数据平台，数仓数据分析需求，保留数据部分要真实有效。脱敏后的数据需要有分析的价值，数据的业务特征不能丧失。大数据分析平台中包括海量的用户隐私和敏感数据，针对大数据环境进行数据脱敏，数据结构复杂，包括结构化、非结构化数据等，脱敏操作复

杂。特定场景下，需对脱敏数据进行还原，例如证券基金通过对脱敏后的转账记录、交易记录等进行分析，评估信用等级等需求。

基于以上数据脱敏系统内置丰富的数据脱敏算法及数据多种脱敏的能力，可适配不同的数据分析场景，从而在保证数据脱敏的前提下，安全地进行各种数据分析。保证主键一致、业务关联一致、有依赖字段的敏感信息脱敏一致、多次脱敏结果保持一致等，整体保证脱敏前后一致性，保持数据间的逻辑关系。支持数据脱敏可恢复操作，具备可逆权限的用户能通过相关算法和密钥进行此操作，充分保证数据安全性。脱敏系统需要支持内置敏感数据发现能力，对源数据中的敏感数据按照脱敏规则，进行变形、遮蔽，仿真等，保证脱敏后的数据保持数据的一致性和业务的关联性。

**2. 基于隐私保护脱敏技术的数据共享**

在数据共享场景中涉及业务数据接口调用与数据分发给第三方，在使用接口开放给外部第三方数据共享或者部分数据分发给外部的第三方或上游组织，若数据中包含未经脱敏处理的敏感信息，数据一旦泄露后果不堪设想。数据脱敏系统保证敏感数据（如，客户号、身份证号码、手机号）在从其所在数据库抽取时，进行相关安全处理；还可以按照第三方或者上游组织的要求根据字段数据的业务特征进行指定方式的脱敏处理，解决对脱敏数据有特殊处理的脱敏需求。

数据脱敏系统内置丰富的敏感数据识别处理模板，包括个人隐私、金融财务、商业机密等，在进行数据脱敏时，将会对特定敏感数据按照指定规则进行脱敏，且可根据实际需求自定义敏感数据，满足多种行业特有数据脱敏需求。

基于场景化的不同，脱敏技术使用需联动数据资产管理平台的分类分级系统，制定基于数据脱敏规则，从而完成基于不同数据使用场景采用不同脱敏算法来实现数据自动化脱敏能力，脱敏结果校验能力，脱敏数据风险分析能力。自动化数据脱敏体系的构建，从敏感数据发现、数据分类分级匹配，数据抽取，数据脱敏处理，数据装载，数据安全分析多个阶段实现了完全的自动化联动能力，优化了原来需高度依赖人工进行数据采集流程和数据脱敏操作割裂的操作流程，将数据提取时间从原来的 1,2 周提升至只需要 1,2 天即可完成，极大地提高了效率，降低的相关成本和数据安全风险。

随着国家对于个人隐私数据保护相关的法律法规不断出台，近年来频发的数据泄露事件，促使企业将数据合规视为首要任务。隐私保护脱敏技术将各类法律法规与企业实际业务相结合，在实现数据安全合规的同时，最大程度上保持数据的可用性与可挖掘性。

## （三）隐私保护计算技术

### 1. 多方安全计算

多方安全计算技术是一种能够基于多方数据协同完成计算目标，并保证不泄露各方隐私数据的密码学技术，其优势在于安全性和准确性高，且通用性强，能够实现隐私集合求交、隐匿查询、联合统计和联合建模等多种功能。但由于涉及密文通信，多方安全计算技术的性能有所下降。因此在实践中，多方安全计算技术适用于对数据保密性和计算结果准确性要求高，但对实时性要求较低的场景。其既可用于多方数据统计计算或比对场景，如资产管理服务、资金流水核验、安全人脸比对等，亦可用于基于多方数据进行联合建模的场景，如信贷风险监测和催收管理模型训练等。

（1）智能化资产管理服务营销

某金融集团旗下的银行应用多方安全计算技术，将客户在集团内部证券、保险、基金和信托等业务线的资产和授信数据进行密文计算，在保护客户隐私的情况下，全面了解该客户的资产和负责情况，为客户打造定制化资管服务方案，同时避免了未经授权的推销。

在本场景中，该银行需要根据该客户在集团内部不同机构沉淀的资产总量和授信总额，为其打造个性化的资产管理服务。此场景需要融合的数据包括该客户的姓名或名称、身份证或统一社会信用代码，以及其在证券、保险、基金以及

信托机构的资产和授信额度信息；此场景的计算主要为客户资产加总和授信额度加总。

应用基于秘密分享的多方安全计算协议，该银行能够在不泄露客户证券、保险、基金及信托资产和授信额度的情况下，对上述数据进行加法联合统计运算。此方案涉及的参与角色包括任务发起方、数据输入方、结果获取方等三个角色。其中，提供资管服务的银行作为任务发起方和结果获取方；数据输入方包括该银行以及集团内的保险机构、证券公司、信托公司和基金公司。

此场景包括四个核心技术环节。一是任务发起和调度环节，该集团的银行在获得客户授权后发起对其的资产额度和授信额度加总任务。二是合约形成环节，该集团的银行、保险公司、证券公司、信托公司和基金公司对计算任务（及资产额度和授信额度加总）和各自所需提供的数据达成一致，形成计算合约。三是数据输入环节，各数据输入方根据计算合约内容将该客户的资金流水明细、保单明细、基金及证券交易数据通过数据节点以密文随机分片(亦称“计算因子”)的形式输入至多方计算平台的不同计算节点，进行联合统计计算。在此过程中，各计算节点互不串通，无法恢复任何原始数据信息。三是结果输出环节，该客户的资产总额及授信总额由密文形式输出给银行并解密，用于银行为该客户制定资管服务方案

2020 年 7 月颁布的《金融控股公司监督管理试行办法》要求金融控股公司及其控股机构在通过数据共享协同开展业务时，应确保合法合规、风险可控及经客户授权的情况下共享数据，以防范数据滥用风险，保护客户隐私。

首先，多方安全计算技术有助于提升关于《个人信息保护法》对于匿名化要求的数据合规[10]。在此方案中，基于秘密分享的多方安全计算协议能够保障输入数据（客户证券、保险、基金以及信托资产和授信额度数据）和输出数据（资产总额和授信总额）的隐私保护，有助于让数据在非受控环境下无法重识别，全方位降低隐私泄露风险。因此，多方安全计算技术可作为匿名化技术方案的一个组成部分，实现一定条件下的匿名化。

其次，在此场景中，多方安全计算技术符合 2021 年 11 月正式实施的《个人信息保护法》的个人信息收集、处理的最小必要原则。《个人信息保护法》第六条规定：处理个人信息应当具有明确、合理的目的，并应当与处理目的直接相关，采取对个人权益影响最小的方式。收集个人信息应当限于实现处理目的的最小范围不得过度收集个人信息。根据此规定，最小必要原则主要包括两部分，一是个人信息应限于实现处理目的的最小范围，二是个人信息处理行为与处理目的直接相关。在该技术方案中，基于秘密分享的多方安全计算协议下，客户在信托公司、保险公司、基金公司和证券公司的资金流水明细、保单明细、基金及证券交易数据被转换

为密码学安全的计算因子，并被送至不同计算节点执行联合统计计算。在此过程中，各计算节点互不串通，无法恢复数据原始信息。另外，计算合约机制保证各方在对数据的使用达成一致后方可执行计算任务，能够严格控制数据的实际使用方式、次数和目的。因此，上述数据无法被任何一方获取并用于目标场景以外的用途，使得数据处理活动和用途与场景目标直接相关，且被限定在必要的范围之中。

最后，多方安全计算技术能够作为符合《数据去标识化共享指南》（DB31/T 1211-2021）中去标识化要求的技术。该指南指出，数据共享域是指以机构间数据共享为目的，由参与机构共同组织和运营的、有安全边界的、受控、互信且制衡的计算机域，具备使去标识化数据在共享域内、对原始控制方以外的其他方匿名的能力。数据共享域适用的去标识化技术满足的条件包括域内匿名、标识隔离。另外，参与数据共享的机构共同构成数据共享域的不同层级服务，互信并制衡。在此场景中，该集团内部的银行、保险、信托、基金等部门构成了符合上述指南所定义的“数据共享域”。在该域中，多方安全计算技术通过将客户证券、保险、基金以及信托资产和授信额度数据以计算因子形式进行计算，实现了上述数据在域内匿名且标识隔离，从而保护客户上述数据的安全。

（2）基于多方安全计算的贷款资金流向追踪

有效监测贷款资金流向及用途是银行防范信用风险重要手段。贷款资金流转往往涉及不同银行，对其流向追踪涉及银行间数据，然而，银行间由于数据互不共享，且资金流水数据属于个人敏感信息，难以直接共享给监管机构。因此，当前银行业跨行贷款资金监测多采用线下方式，其对资金流向的追踪滞后性明显，无法准确掌握贷款资金的真实用途，因而面临较大贷款管理风险与合规风险。对此，某大型国有商业银行在获得用户明确授权的情况下，应用多方安全计算技术，根据贷款资金流入违禁领域的判断准则，对参与行间的账户数据、交易数据和贷款数据进行密文计算，从而判断资金终点是否为违禁领域。

该银行组织三家分支机构分别模拟独立商业银行，在获得客户同意的情况下，应用多方安全计算技术对各参与行的账户数据、交易数据和贷款数据执行基于密文的联合统计计算，并依据资金流入违禁领域的判断准则自动判断资金终点是否违规，实现了对贷款资金流向的线上、实时监测。

具体地，A 银行为资金流向监测方，并与 B 银行、C 银行同为数据提供方。贷后资金追踪流程如下：首先，A 银行发起贷后资金监测任务，指定核查时间区间，准备本行待核查贷款业务数据，B 银行及 C 银行同时准备相关流水数据。其次，三方将相关数据以密码学安全的计算因子形式传输至多方安全计算平台，平台基于核查规则进行密文计算，获取违

规业务资金链路信息。最后，平台将资金链路信息密文发送给 A 银行，供其本地解密后使用。A 银行依据解密结果，实现跨行贷款资金流向监控。

在此场景中，多方安全计算技术能够助力银行提升关于《个人信息保护法》对于匿名化要求的数据合规。在秘密分享的多方安全计算协议下，输入数据（客户的资金流水数据）和输出数据（资金链路信息）以密文分片形式被传输和计算，无任何原始数据被泄露，因此在非受控环境下无法重识别。由此可见，多方安全计算技术在此场景中，可作为匿名化技术方案的一个组成部分，实现一定条件下的匿名化。

其次，在此场景中，多方安全计算技术符合 2021 年 11 月正式实施的《个人信息保护法》的个人信息收集、处理的最小必要原则。《个人信息保护法》第六条规定：处理个人信息应当具有明确、合理的目的，并应当与处理目的直接相关，采取对个人权益影响最小的方式。收集个人信息应当限于实现处理目的的最小范围不得过度收集个人信息。根据此规定，最小必要原则主要包括两部分，一是个人信息应限于实现处理目的的最小范围，二是个人信息处理行为与处理目的直接相关。在基于秘密分享的多方安全计算协议下，客户的资金流水数据被转换为密码学安全的计算因子，并被送至不同计算节点执行联合统计计算。在此过程中，各计算节点互不串通，无法恢复数据原始信息。因此，上述数据无法被某一方获取并用于目标场景以外的用途，使得客户的资金流

水数据的实际用途与贷款资金链路追踪直接相关，且被限定在必要的范围之中。

（3）基于多方安全计算的安全人脸识别

人脸识别在应用过程中所产生的数据在传输、存储、使用等环节中均存在个人信息泄露、滥用、超范围共享等安全风险。对此，某股份制银行应用多方安全计算技术，在人脸识别过程中将人脸特征数据进行基于密文的比对计算，能够保障人脸数据在注册、存储和比对环节的安全。另外，该方案结合数据溯源、区块链等技术，能够对人脸数据的使用进行全流程存证，支持事后问题溯源和审计。

该银行构建的安全人脸识别平台应用秘密分享协议，由多个互不串通的计算节点组成。在注册和存储环节，平台遵从随机密钥分发协议和计算合约，因此即使某个计算节点被恶意攻击，人脸数据仍无法被恢复。具体地，用户在个人终端开通人脸识别支付服务时，终端设备内嵌的算法提取人脸特征值，并将 ID 与人脸特征值发送至本地加密模块。加密模块使用秘密分享协议，对用户 ID 和人脸特征值转换为密文随机切片（亦称“计算因子”），并将不同部分分别传输至银行、支付机构等不同机构的人脸特征库，进行分离存储，后续用于身份验证等环节的人脸比对工作。

在比对环节，平台将人脸特征数据在全密文环境下进行比对，保障了人脸数据的安全。具体地，用户通过应用程序进行支付操作，通过人脸识别核实身份信息，触发人脸数据

比对任务。随后，终端设备利用其采集的人脸图片提取人脸特征值，内嵌加密模块使用秘密分享的方法将人脸特征值转换为密文数据分片（即“计算因子”）形式，并发送到平台的不同计算节点缓存。在管理模块的任务调度下，各机构人脸特征值库中的密文数据分片与缓存的密文数据分片在全密文环境下进行比对，并将比对结果分片返回到终端。随后，解密模块对数据进行拼装解密，应用程序根据解密后得到的身份识别结果完成支付业务。

在此场景中，多方安全计算技术能够助力银行和支付平台提升关于《个人信息保护法》对于匿名化要求的数据合规。在秘密分享的多方安全计算协议下，输入数据（客户的人脸数据）以密文分片形式被存储、传输和计算，且输出数据（比对结果）以密文分片形式传输，业务全流程中无任何原始人脸数据被泄露，因此在非受控环境下无法重识别。由此可见，多方安全计算技术在此场景中，可作为匿名化技术方案的一个组成部分，实现一定条件下的匿名化。

另外，多方安全计算技术符合 2021 年 11 月正式实施的《个人信息保护法》的个人信息收集、处理的最小必要原则。《个人信息保护法》第六条规定：处理个人信息应当具有明确、合理的目的，并应当与处理目的直接相关，采取对个人权益影响最小的方式。收集个人信息，应当限于实现处理目的的最小范围不得过度收集个人信息。根据此规定，最小必要原则主要包括两部分，一是个人信息应限于实现处理目的

的最小范围，二是个人信息处理行为与处理目的直接相关。在基于秘密分享的多方安全计算协议下，客户的人脸特征数据被转换为密码学安全的计算因子，并被送至不同机构的人脸特征库进行分片存储。在比对环节，终端采集的人脸特征数据被转换为计算因子，并被发送至不同计算节点进行计算。综上，在注册存储和比对环节中，人脸特征数据均无法被恢复。因此，上述数据无法被某一方获取并用于目标场景以外的用途，使得用户的人脸数据的实际用途与身份识别直接相关，且被限定在必要的范围之中。

**2. 联邦学习**

联邦学习的核心思想是通过在多个拥有本地数据的数据源之间进行分布式模型训练，在不需要交换本地个体或样本数据的前提下，仅通过交换模型参数或中间结果的方式，构建基于虚拟融合数据下的全局模型。但在交换模型参数时通常采用同态加密等形式，导致计算性能较低。所以在参与运算的每一方数据量在百万级以下时，适合采用联邦学习。

（1）基于联邦学习的金融反欺诈场景

随着人工智能和大数据技术的快速发展，机器学习算法在金融反欺诈的应用场景中广泛，并取得显著成效，金融诈骗行为能够被有效识别。但在巨大的经济利益驱动下，金融诈骗手段层出不穷，并具有愈加复杂与智能化的趋势，传统基于单一企业的数据建模逐渐难以应对不断升级的诈骗手

段，为金融反欺诈风控带来严峻的挑战。与此同时，数据监管日趋严厉，各项隐私保护、数据安全的法令法规相继制定，导致金融行业数据孤岛现象普遍，为跨机构间的数据协作带来各种障碍，迫切需要通过联邦学习等隐私计算技术打通企业间的数据孤岛，构建跨行业数据共享的反欺诈模型，以提升金融反欺诈的效率与精准性。

借助纵向联邦学习技术，可以在保证用户隐私信息、数据安全以及确保企业对数据所有权与控制权的前提下，以安全合规的方式融合银行机构与外部机构，如电商、运营商、政务等多元数据，实现跨行业、跨机构的反欺诈体系建设，构建基于金融行为特征、消费行为特征、通信行为特征、社交行为特征等跨行业特征，构建针对不同细分金融反欺诈业务场景的专有模型，实现跨行业、跨机构的多样性欺诈数据特征互补，从而提升金融行业的反欺诈的精准度。针对某些反欺诈场景中的单一机构样本量不足问题，可以通过横向联邦学习构建金融同业反欺诈模型，解决单一金融机构样本量少难以构建稳定、精准金融反欺诈模型的难题，提升银行业的整体反欺诈能力。针对某些特定场景的小样本建模难题，则可以通过联邦迁移学习来克服样本和特征不足问题，通过知识迁移快速构建目标领域的联邦反欺诈模型。

以银行业的异业金融反欺诈为例，借助联邦学习技术，引入外部数据源，联合银行已有欺诈样本和行内金融特征构建联合反欺诈模型。应用流程通常首先通过隐私集合求交技

术(PSI)进行跨机构间的样本安全对齐，确保参与方直接除了交集样本无法获知或反推其他参与方的非交集部分样本，然后利用纵向联邦学习算法（如联邦 LR、XGBoost 等常用的联邦化有监督学习算法）进行反欺诈模型的构建，并与现有欺诈系统进行对接，具体应用如图 6 所示。

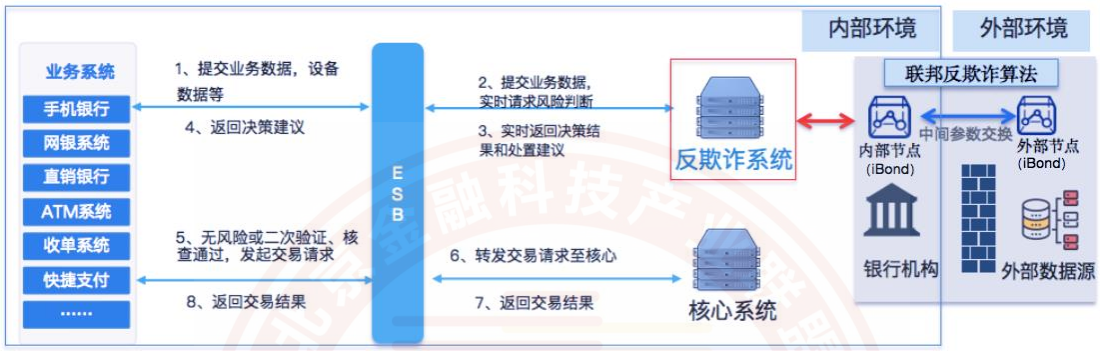


图 6 联邦计算应用场景

基于联邦学习模型能够对用户欺诈行为进行有效识别，能够有效提升商业银行的风险防控能力，并以“数据可用不可见”“数据不动模型动”的方式解决了多方数据协作的合规性和安全性问题，挖掘出更大的数据潜能，并避免了跨机构数据协作过程中原始数据在其他机构的数据留存问题，有效保护数据的所有权和控制权，实现数据价值的安全共享和共创。

（2）基于联邦学习的信用卡分期营销

信用卡业务经过几年的高速发展已经逐步进入业务稳定期，根据中国人民银行公布数据，全国信用卡和借贷合一卡发卡数量增长率连续四年逐年降低。2022 年 7 月 7 日，原中国银保监会、中国人民银行发布《关于进一步促进信用卡

业务规范健康发展的通知》，《通知》要求银行业金融机构不得以发卡量、客户数量等作为单一或主要考核指标，长期睡眠卡率超过 20%的银行业金融机构不得新增发卡。在此背景下，信用卡业务需要从重“量”向重“质”的方向进行转变，信用卡存量客户经营成为信用卡业务的一项工作重点。在信用卡业务中，衍生发展出的消费分期业务在近几年迅速发展，分期手续费收入成为信用卡业务重要的收入来源之一。由于信用卡分期业务具有盈利能力强、业务门槛低的优势特点，同时也可以增强客户活跃度和客户黏性，因此信用卡分期业务的营销成为各大银行的一个重点营销方向。

基于多维数据补充客户画像、构建信用卡分期营销响应模型，开展精准营销，可以提高信用卡分期业务的营销响应率。在构建信用卡分期响应模型时，除了基于行内自有的客户信用卡消费数据，通过引入外部丰富的特征维度如电商消费、通讯行为、兴趣偏好等联合建立响应模型，可以提升模型效果，促进业务转化。

在传统的联合建模方式下，需要将银行的样本数据拿到数据源进行建模或者将数据源的数据拿到银行进行联合建模。在这种传统的联合建模方式下，如果银行的数据拿到数据源进行建模，行方用户的个人身份信息、金融交易等信息就存在着数据泄露的风险；如果数据源的数据拿到银行进行建模，数据源用户的个人身份信息、消费等信息也存在着数据泄露的风险。在数据安全和隐私保护的相关法律法规逐渐

完善、监管日趋严格的大背景下，传统的联合建模方式显然是不能保证数据安全和隐私保护的。

基于多方安全技术和联邦学习技术，可以联合多方开展安全的联合建模。通过基于多方安全计算的隐私求交技术和联邦学习技术，建模双方的原始数据均在各自的服务器私域中，对于模型计算的中间结果采用同态加密等密码学技术进行交互计算。整个联合建模过程中，双方的原始数据均不出各自私域。

某股份制银行想要在行内数据的基础上，利用外部数据联合构建信用卡账单分期业务营销响应模型，期望通过增加丰富的特征维度来提升模型效果，促进分期业务客户转化率。

在本次联合建模项目中，行方提供了行内用户的信用卡消费信息、额度使用信息、信用卡逾期信息、办理分期业务信息等；同时引入了外部的运营商数据进行画像补充，包括通信账单消费、通话偏好、上网偏好等信息。

整个联合建模的全流程是基于隐私计算平台开展的。在数据对齐阶段，基于银行方的建模样本 ID 和数据源机构的全量 ID 进行隐私求交，得到对齐 ID 结果。各数据源机构根据对齐后的共有样本 ID 和回溯时间，匹配各自的特征数据。基于各方提供的特征标签数据，银行方发起模型训练任务。模型训练分别采用纵向联邦评分卡和纵向联邦 Xgboost 进行构建，最终选择效果最优的模型。

在样本对齐阶段，采用基于安全多方计算的隐私求交（PSI）技术，保障求交过程中的 ID 信息安全，同时两方互相不暴露己方的非交集部分样本 ID。在联合模型构建阶段，基于联邦学习技术保障各方原始数据不出己方私域。

（3）基于联邦学习的信贷产品贷前风控

贷前风控是金融机构信贷产品风控管理的第一道防线。在客户申请相关信贷产品前，金融机构需要对申请客户进行资质审查，以尽可能地避免放款后不良贷款或坏账的发生，降低经济损失。申请评分卡是银行等金融机构在贷前风控环节常用的模型，此类风控模型的目的在于预测客户申请时点（申请信用卡、申请贷款）未来一定时间内发生逾期或违约的概率，常用于贷前准入或授信环节。

在客户申请银行的信贷产品时，银行只能获取到客户填写的基本身份信息，显然，这些基本信息是无法全面评估申请客户的风险状况的。因此在贷前风控环节，常常需要结合外部数据来构建贷前评分模型，构建贷前评分模型常用的数据包括中国人民银行征信数据、多头借贷数据、消费交易数据等。

在使用外部数据构建模型时同样面临着传统联合建模所面临的困境：要么行内数据出私域、要么数据源数据出私域，数据出私域就面临着客户身份信息、交易信息等数据泄露的风险。贷前风控业务场景下的多方联合建模也可以基于联邦学习技术来保障建模过程中各方的数据安全。通过联邦

学习，在各建模参与方原始数据不出私域的前提下，完成各方全局模型的构建。

### （四） 隐私保护辅助技术

基于数据水印技术，实现企业内部数据泄露的溯源追责，数据水印技术可以针对数据文件中的敏感数据进行高级别仿真水印标识，将水印信息隐藏在数据库表中，当发生数据泄露和窃取时，通过提取水印信息，可以快速定位风险者，准确查到数据是谁泄漏的、何时泄漏、泄露了什么，从而解决数据扩散后泄漏主体不明确、无法追溯等难题。

随着科学技术的进步，数据的网络化使用越来越广泛，大数据的背后往往隐藏着巨大的社会价值与经济价值。在各个领域都起着关键性作用，数据已逐渐成为当今互联网时代的核心资产。随之数据泄露问题也逐年升高。据有关数据统计,2020 年 1-9 月全球公开披露的数据泄露事件有 2953 起，涉及泄露数据量高达 361.07 亿条。从泄露原因的角度分析，数据泄露既有外部黑客攻击的因素，也有企业内部员工泄露的原因。数据泄露的溯源技术可以从源头上根治黑灰产与数据泄露问题的关键。

数据水印是指在对原始数据进行共享/交换时，通过复杂的水印嵌入算法向原始数据中植入水印标识，在保证数据使用价值不变的情况下，使数据具有可识别分发者、分发对象、分发时间、分发目的等因素。数据水印具有高可用性、高透明无感、高隐蔽性，不易被外部发现破解的特性。当数

据发生泄漏时，可以通过对泄露数据中的水印标识进行提取，通过水印标识编码可以快速查询到数据泄露的单位及责任人，即数据从何处泄露、在什么时间泄露了什么数据。实现数据的准确追溯，提高数据使用过程中的安全性和可溯源能力。

数据作为企业的重要资产，财务报表，人员背景、商业信息等每天都在被大量地使用与交互中。员工的违规操作与非法下载都加大了企业内部数据泄露的风险。数据水印技术在一方面可以帮助企业了解自身安全管理与措施的薄弱环节，另一方面通过有效准确的溯源追责对数据泄露者起到一定的心理威慑作用，从而有效减少数据泄露事件的发生。当员工在下载数据时，系统将水印信息自动嵌入到下载数据中，当发生数据泄露时，企业通过对水印信息的提取，分析溯源到数据泄露者标识、数据下载时间等信息要素。

随着数据化转型的深入推进，企业间大量的数据共享与数据流通在促进数据价值释放的同时，也带来了更多的数据泄露风险。2020 年出台的《数据安全法》以及《个人信息保护法》对数据的使用给出了非常详尽、严格的法律管控范围，对数据保护也提出了明确要求。数据水印技术降低了企业间数据共享时的风险，提高了数据传递间的安全性和可溯源追溯能力，使数据的共享与交换变得安全可控。

# 五、使用隐私保护技术的金融反欺诈案例分析

## （一）风险控制

### 1. 基于联邦学习的银行普惠金融——地方分行中小微企业信贷风控

（1）背景介绍

中小微企业融资难往往受制于企业规模小、缺乏有效的抵押担保资产、由于信息不对称导致信贷风险高等因素，使得信贷需求和信贷供给错配。某银行调研发现，各分行无法结合地方特色对本地客户进行精准风险管理，因此急需利用本地注册企业的各个维度信息打造更全面、更精细化的信贷风控系统，以降低不良率，提升资产质量，提高金融服务的普惠力度。

某银行希望通过引入外部合规数据源进行联合建模，来破解信贷环节中信息不对称问题，而多方数据融合面临原始数据泄露与隐私泄露风险，且面临数据来源多、格式杂，数据处理及多方数据协作技术投入高难题。此前，某银行软件开发中心已探索隐私计算技术，经过长期技术及业务可行性验证，选择部署基于隐私计算技术的多方数据安全协作解决方案，赋能某银行地方分支机构普惠金融业务，并率先在某分行落地应用。

本项目中，技术层面，运用联邦学习等隐私计算技术，在保证原始数据互不可见、合法合规的前提下，实现多方数据安全协作；业务层面，帮助某银行某分行引入运营商某分

公司的数据资源，为分行中小微企业信贷风控补充了具有地方特征的多样化信息。基于联邦学习等技术，帮助某银行软件开发中心、某银行某分行、运营商某分公司基于多样化的企业法人数据特征，进行联合建模，丰富了企业用户画像信息，完善了贷前评估和贷后预警模型。

（2）应用场景合规风险分析

1）交换或融合的数据

银行地方分行侧:

标签信息: 定义为是否存在 M3+逾期，存在则标签为 1; 从未出现逾期标签为 0; 否则剔除此样本。

特征信息: 时点贷款余额、企业注册资本、成立年限、过去 1 年税务缴纳等级等。

运营商地方分公司侧:

特征信息: 个人终端/运营商基本信息，用户使用的手机终端信息，入网信息。

个人通信数据: 用户通过手机终端进行拨打/接听电话，发送短信的相关信息，包括发送，接收的号码，时间及时长等信息。

个人上网数据: 通过对用户通过手机端访问 APP、Web 页面信息，加工的上网标签数据，并根据 APP、网页分类库进行多级分类加工。

2）涉及的相关法规及合规风险分析

《个人信息保护法》《数据安全法》等均对个人隐私信息的保护作出了明确的要求。

银行的个人隐私数据，包括个人的金融信息等是非常敏感的个人数据，需要被高度重视与保护。在与外部的数据联合建模中，银行的数据始终不应离开本地，要求“数据可用不可见”。

运营商的数据具有非常丰富的特征信息，且用户覆盖量较全，是众多领域数据合作的优质数据源。运营商大数据包括个人移动轨迹、上网轨迹、个人通信数据等，这些都属于敏感的个人信息，需要在不泄露个人隐私信息的情况下进行与外部进行联合建模等数据协作。

（3）解决方案（技术架构）内容

在本案例中，某银行某分行借助某总行，软件开发中心历时近 2 年在联邦学习，隐私计算技术的深厚积累，结合运营商某省公司在人群画像挖掘，数据仓库等领域的深厚积累，在满足相关政策法规的要求下，使用某软开中心联邦学习技术和平台领域供应商提供的隐私计算平台和建模服务，创造性地探索了基于联邦学习的普惠金融贷前筛查审批模型和服务方案，并实现了技术的落地，如图 7 所示。

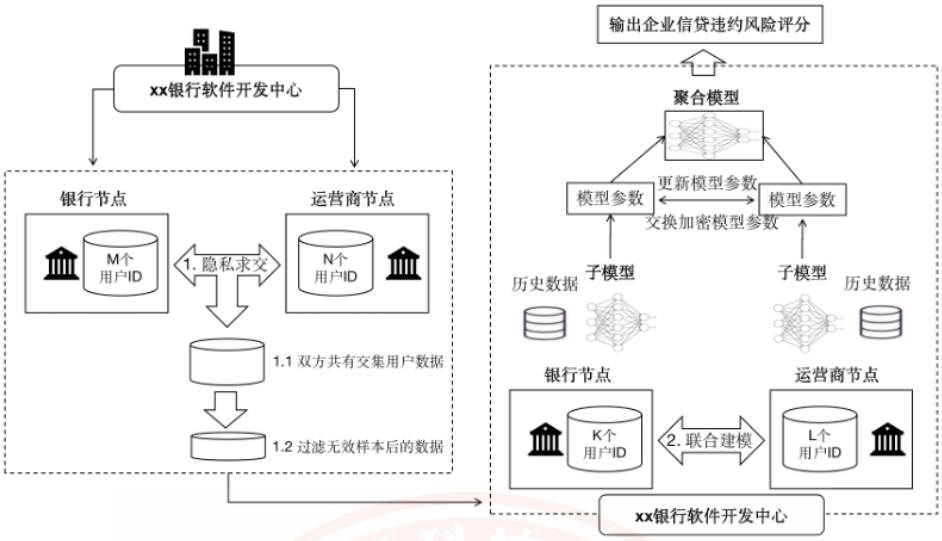


图 7 基于联邦学习的普惠金融贷前筛查审批建模方案

以下技术方案分为业务服务方案、建模方案和算法安全方案。业务服务方案和建模方案具体如下:

业务服务方案：技术在业务落地是隐私计算领域探索的关键目标。本项目中，相关合作方成功探索出，将联邦学习服务的模型，在银行业务中落地的全流程。

服务中，某银行参与系统包括：银行数仓系统，信贷审核系统，联邦学习系统；运营商参与系统包括：数仓系统，联邦学习系统，整体交互流程图如下:

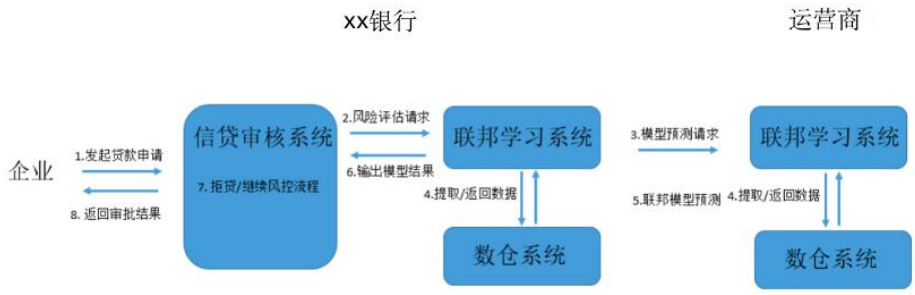


图 8 基于联邦学习的普惠金融贷前筛查审批业务流程图

在交互过程中，首先小微企业向某银行发起贷款审批请求，然后由信贷审批人员联合技术服务人员，以企业法人和主要控制人等手机号信息，应用联邦学习平台（部署在某银行软件开发中心）发起模型服务请求，平台联合移动侧的隐私计算平台，提取双方数仓数据，完成模型运算，并将模型预测结果输出给银行信贷审核系统。

信贷审核系统根据预设置的策略选择继续风控流程，或拒绝贷款。

建模方案：为验证引入运营商数据，对普惠金融场景的业务价值，我们使用如下的建模技术方案，如图 9 所示：

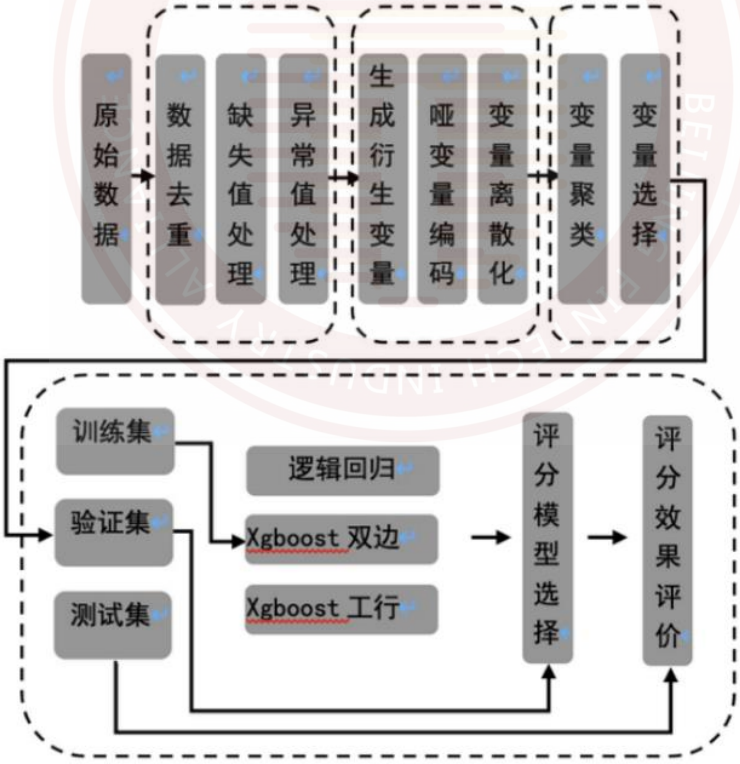


图 9 基于联邦学习的普惠金融贷前筛查审批建模方案

引入数据：本方案中使用了移动通信运营商数据，共包括个人基本信息，用户通信信息，用户上网信息，用户地理

位置信息等，共 1182 个底层特征字段，部分典型字段包括：用户入网时长，读书类 APP 访问天数，通话总时长等。

数据处理：方案中设计了多重的数据处理流程，并应用联邦学习平台在数据不可见的基础上，实现了相应的数据处理工作。包括数据去重，极值去除，缺失值填充等变量异常处理，变量衍生，变量转换等特征工程方法，变量聚类，变量选择等变量探索和筛选方法。

模型训练：方案中按照 7：3 的比例，对样本集进行训练集和验证集的筛选，并应用逻辑回归，XGBoost 等算法完成模型训练和调优。

为验证应用联邦学习技术后对模型效果的提升，我们使用“XGBoost 银行”和“XGBoost 双边”等两种不同的建模方法进行 A/B 测试。

XGBoost 银行：仅使用某银行特征数据入模，使用某银行软件开发中心机器学习建模平台进行模型训练。

XGBoost 双边：使用某银行特征和运营商特征入模，使用某银行软件开发中心联邦学习平台进行模型训练。

XGBoost 银行和 XGBoost 双边使用同样的样本，数据加工方法和模型训练超参数。

模型评估：通过对训练集及验证集下 XGBoost 工行和 XGBoost 双边模型的 KS，AUC 等主流模型效果评估，验证联邦学习技术的应用效果。我们预期引入移动特征后，应提升 XGBoost 模型 5%～8%的效果。

（4）隐私计算技术应用成效

在技术方面，相比于传统的大数据风控系统，一是在系统使用层面降低了对数据分析人员专业能力的要求，普通工程师也可以通过该系统进行风控模型的训练；二是通过融入多方的数据，提高了模型的准确率；三是规范了某银行某分行的数据使用形式，降低了数据泄露的风险；四是带动了某银行某分行以创新技术应用赋能金融数据要素价值发挥的活力。

在业务方面，本次通过隐私计算技术将多方数据进行融合应用，作为某银行贷款初筛模型，已参与数百笔贷款的贷前审批中，目前放款中小微企业尚无出现重大风险企业（出现 M3+逾期）。业务测算效果下，当该模型完整应用于现有业务中(如 2020 年某银行某分行共发放小微企业贷款 400 亿元），结合不同贷款产品的风险情况和贷款年限，预计可降低此行现有小微贷款 0.2%的全周期坏账，年降低此行小微企业不良贷款损失 8000 万元以上。该业务模式推广后，预计会成为运营商某分公司的一大收入突破点，形成规模化收入。

在此基础上，本项目还将打造面向小微商户的开放式融资服务新模式，该服务上线后，客户可在线主动申请业务，由系统开展自动审批，大幅提高业务办理效率和客户体验，并进一步扩大普惠金融服务面。

## 2. 基于 TEE 的联合统计的金融机构信贷风控案例

（1）背景介绍

随着金融机构消费贷快速发展，由于不需要抵押物等，消费贷的欺诈行为也层出不穷，而网络借贷的欺诈行为中以身份冒用最为常见。例如某中介公司以招揽学生做兼职为由，要求学生去银行办理工资卡，利用银行卡和手机号获取了学生的身份证、学籍、学历等信息，而后用绑卡方式向多家金融机构网贷平台申请了多笔信贷业务，造成严重且恶劣的社会影响。

利用隐私计算技术，可以通过对多家金融机构网贷平台的联合隐私计算，获取到恶意诈骗的用户信息，使用借贷数据来判断是否存在多头借贷行为，放贷存在风险。

在该案例中，对多家金融机构的消费贷用户进行联合统计、集合求交等相关运算，如果同一个用户在短时间内向多家金融机构的消费贷产品都进行消费贷操作，则可以将该用户的欺诈风险等级提高，列为高风险用户，一定程度上限制其借贷行为，尽量降低借贷欺诈的出现。

（2）应用场景合规风险分析

1）交换或融合的数据

消费金融机构侧：个人信贷历史记录、个人信贷还款历史记录、个人信贷逾期记录等;

第三方数据机构侧：个人消费行为及特征信息、运营商信息等;

2）涉及的相关法规及合规风险分析

2021 年，我国正式发布《中华人民共和国数据安全法》（以下简称《数据安全法》）、《中华人民共和国个人信息保护法》（以下简称《个保法》）、《信息安全技术 个人信息安全规范》（GB/T 35273-2020）、《个人金融信息保护技术规范》（JR/T 0171-2020），以备受关注的数据安全、个人金融信息安全及权益保护领域在法律层面和技术操作层面提出更高的要求和更清晰的规范。银行要扩大信息数据的维度，增强信贷风险的管控能力的同时，因为联合多渠道、多方面的客户信息，从而产生个人授权不清晰、不合规等个人数据隐私风险问题和在数据共享期间因技术差异导致的数据泄露问题。

（3）解决方案（技术架构）内容

基于 TEE 的联合统计方案如图 10 所示

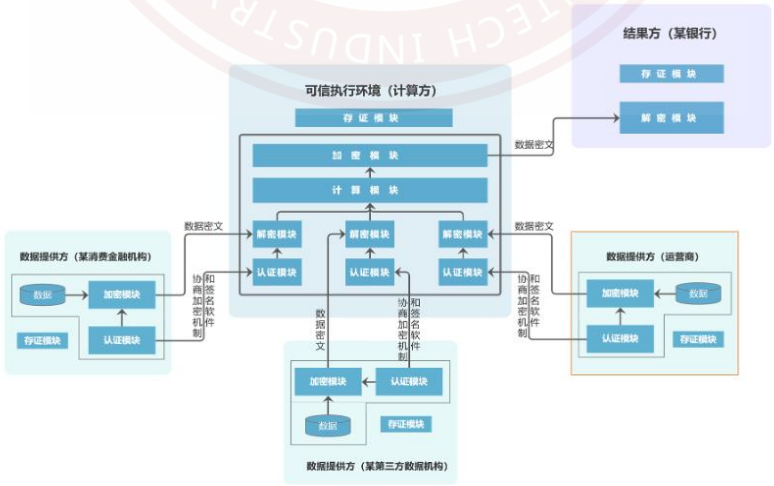


图 10 基于 TEE 的联合统计方案

针对贷前风控模拟业务场景，验证通过可信执行环境可以实现银行对个人多头借贷的行为预测。通过可信执行环境构建的可信计算节点，将银行、消费金融机构、第三方数据机构进行三方联合建模，其中银行提供判定为多头借贷的规则，消费金融机构提供用户个人金融数据，第三方数据机构提供个人借贷的数据，采用有监督-二分类算法（LR/XGBoost/LightGBM），得到多头借贷模型，通过导入银行侧的预测数据，来生成模型预测结果，结果为“是、否”，是代表该用户存在多头借贷行为，否则代表不存在。

基于可信执行环境技术的特点，以软硬件的方法在 CPU 中构建一个安全区域，保证其内部加载的程序和数据在机密性和完整性上得到保护。通过 TEE 构建可信节点，数据提供方和结果获取方均在 TEE 节点内完成数据的交互操作，例如案例中数据提供方消费金融机构和第三方数据机构将数据均具备 TEE 节点，当任务发起方即银行发起一个建模任务后，数据提供方在任务发起方的 TEE 节点中进行数据的汇聚，任务发起方的 TEE 节点来进行数据的特征处理、算法处理、流程编排最终生成模型，并对建模过程中产生的数据进行销毁，由于 TEE 节点存在物理隔离，所以任务发起方无法获取到原始的数据和计算过程数据，TEE 节点仅向任务发起方提供任务的结果，即模型，任务发起方导入预测数据后，训练模型可以产出对应的预测结果，避免数据源提供方包含个人隐私性的原始数据泄露。

（4）隐私计算技术应用成效

效率提升、成本降低

1）运算效率高。基于 TEE 的联合统计，可以实现基于对三方及以上数据复杂计算条件的联合运算，比两两之间逐个进行隐私求交的方案效率更高。

2）系统性能损耗降低。基于可信执行环境（TEE）的方案相比多方安全计算、联邦学习等技术，具有明显的性能优势。基于可信硬件的安全计算技术的性能是原有性能的 2-3 倍，联邦学习和 MPC 的性能是原有性能的百倍以上，所以使用 TEE 的方案对原有系统的性能损耗更低，满足处理实时性要求较高的场景。

安全性及灵活性

1)数据安全保护。TEE 的计算是基于 CPU 和内存的计算，数据完成计算后会销毁，不会被某一方落盘，可以保证金融机构自身的客户数据安全。

2）安全性提升。TEE 方案借助芯片级的隔离机制，为高安全性的代码提供独立的隔离空间，可以在物理环境和整个操作系统环境受控的前提下保护关键数据资产和关键计算过程的机密性、完整性。同时 TEE 也可以增强联邦学习的安全性假设，对梯度数据等在 TEE 中完成，将“半诚实模型”提高到“抵御恶意攻击模型”。

3）算法自定义。多头借贷的欺诈场景中，往往需要根据需求调整查询和计算条件，用户可以实现不同算法对适配TEE的自定义编程，从而实现不同的计算条件。

## （二）反欺诈

### 1. 联邦学习助力银行贷前反欺诈

（1）背景介绍

消费贷近年来日益兴起，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。而互联网机构侧则特征库的用户数据（消费行为、消费特征、网络行为轨迹等）、黑灰产名单库数据等，这些数据有助于银行侧来进行贷前风控的反欺诈。

因此，银行和互联网机构需要融合多方数据进行联合建模，才能构建更加精准的贷前风控反欺诈模型。而为了实现数据安全和隐私保护，双方的数据需要借助联邦学习，来实现数据的“可用而不可见”，同时，充分发挥双方数据的价值。

（2）应用场景合规风险分析

1）交换或融合的数据

银行和互联网机构侧，互相融合的数据主要包括:

在银行侧：主要包括历史还款信息、征信数据和第三方的通用征信分等。

在互联网机构侧：主要包括互联网特征库的用户数据（消费行为、消费特征、网络行为轨迹等）、黑灰产名单库数据等。

2）涉及的相关法规及合规风险分析

在多方数据联合建模的过程中，隐私保护和数据安全是不可忽视的重要环节。依据《数据安全法》《个人信息保护法》等法律法规，需要对个人敏感信息采取严格保护措施，并强化个人信息处理者的义务。依据《JR/T 0171-2020 个人金融信息保护技术规范》等金融领域的行业标准，需要实现对个人金融信息的合规保护。

联邦学习可以有效解决合作中数据隐私保护与特征变量融合之间的矛盾，在双方或多方合作中保障特征变量交换时的数据安全。

（3）解决方案（技术架构）内容

金融机构与互联网机构基于 XX 联邦学习平台，在原始数据不出域的前提下，实现双方数据融合，实现贷前风控反欺诈。其技术架构，如图 11 所示：

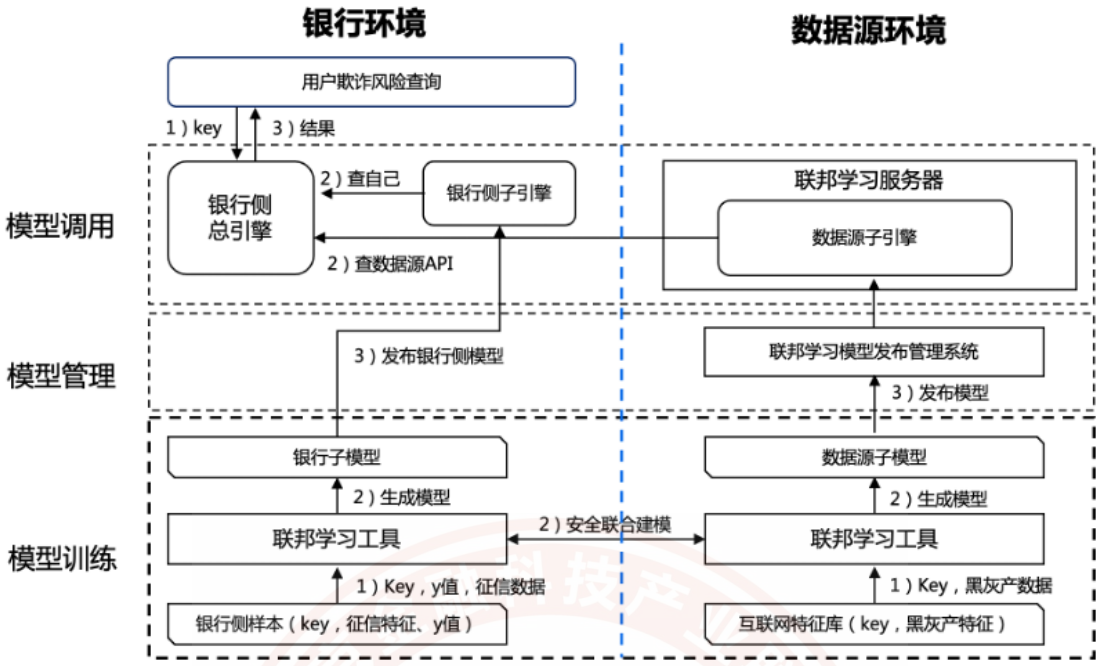


图 11 联邦学习助力银行贷前风控反欺诈的技术架构图

在金融+互联网数据融合的应用场景中，模型训练的业务流程如下:

1）银行侧的联邦学习工具从银行侧样本中提取 key、征信和标签等数据，数据源侧的联邦学习工具从互联网特征库提取 key、黑灰产等数据;

2） 银行侧和数据源侧的联邦学习工具基于提取的数据，进行安全联合建模，并生成欺诈风控的银行子模型和数据源子模型;

3） 银行侧与数据源侧发布各自的子模型。

在金融+互联网数据融合的应用场景中，模型调用的业务流程如下:

1）银行基于用户的 key 从银行侧总引擎查询欺诈风险;

2）银行侧总引擎基于用户的 key，基于银行侧子引擎查询到的自身的子模型分，以及从数据源子引擎的 API 查询到的子模型分，联合计算出用户欺诈风险评分；

3）银行侧总引擎返回用户欺诈风险评分。

在上述流程中，银行机构与互联网机构基于联邦学习，实现安全联合建模与联合预测，在用户的原始数据不出域的前提下，充分发挥双方数据的价值，联合计算出用户的欺诈风险评分，从而实现用户的贷前风控反欺诈。

（4）隐私计算技术应用成效

在贷前风控反欺诈的应用场景中，银行方提供样本标签，互联网机构提供样本特征，双方构建联邦逻辑回归和联邦梯度提升树模型，对客户进行信贷风控评估。在不暴露数据明文的前提下，模型的特征维度被数据服务方丰富优化，风控模型准确性得以提升。

基于 XX 联邦学习平台，银行可以排查多头借贷和超额借贷的情况，对于高风险用户，可以得到来自平台的高风险报警。同时，银行可以在防范风险的同时，避免各方信息泄露。在本案例中，某银行应用联邦学习产品，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。

## 2. 基于联邦学习的银行贷前反欺诈评分模型

（1）背景介绍

在信贷业务中，贷前风控在借贷类产品风险管理中是非常重要的，做好贷前风控管理可以很大程度上减少银行的经济损失。贷前风险主要包括信用风险和欺诈风险，其中信用风险相对有章可循，可以根据借款人的收入、还款、负债等信息来综合评估其还款能力和还款意愿，预测借款人在未来一段时间内的逾期风险。相对于信用风险，欺诈风险是一种更加灵活多变、与时俱进的风险类型。对于欺诈风险，诈骗的种类较多，比如第一方欺诈、第三方欺诈、个人欺诈、团伙欺诈等。诈骗方式也更加专业化、智能化，导致欺诈风险并不能很好地识别，因此反欺诈在金融贷前风控系统中尤为重要。

目前常用的反欺诈体系包括设备指纹反欺诈识别、黑名单核验、反欺诈评分模型等。构建完备的反欺诈体系仅靠银行自有数据是远远不够的，在整个反欺诈体系中，常常需要借助外部数据来完善整个反欺诈体系。

（2）应用场景合规风险分析

1）交换或融合的数据

在引入外部数据构建反欺诈体系时，往往涉及数据的传输、存储和使用的环节。以构建反欺诈评分模型为例

在银行侧，建模需要用到的数据常常包括用户的手机号、身份证号、申请信贷产品填报的收入、家庭状况等基本身份信息；

在外部数据源侧，根据引入的数据类型的不同常常包括用户手机号、设备号、消费行为等信息。

2）涉及的相关法规及合规风险分析

传统方式使用外部数据进行建模往往通过 U 盘、API 等方式进行数据传输，一方面在数据传输的过程可能存在数据泄露的安全隐患，另一方面在银行联合外部数据进行建模时，也会存在银行客户基本信息、建模 Y 标签泄露的风险。

（3）解决方案

在本案例中，通过引入三个不同的外部数据进行特征维度补充，为该银行构建了反欺诈评分模型，如图 11 所示。

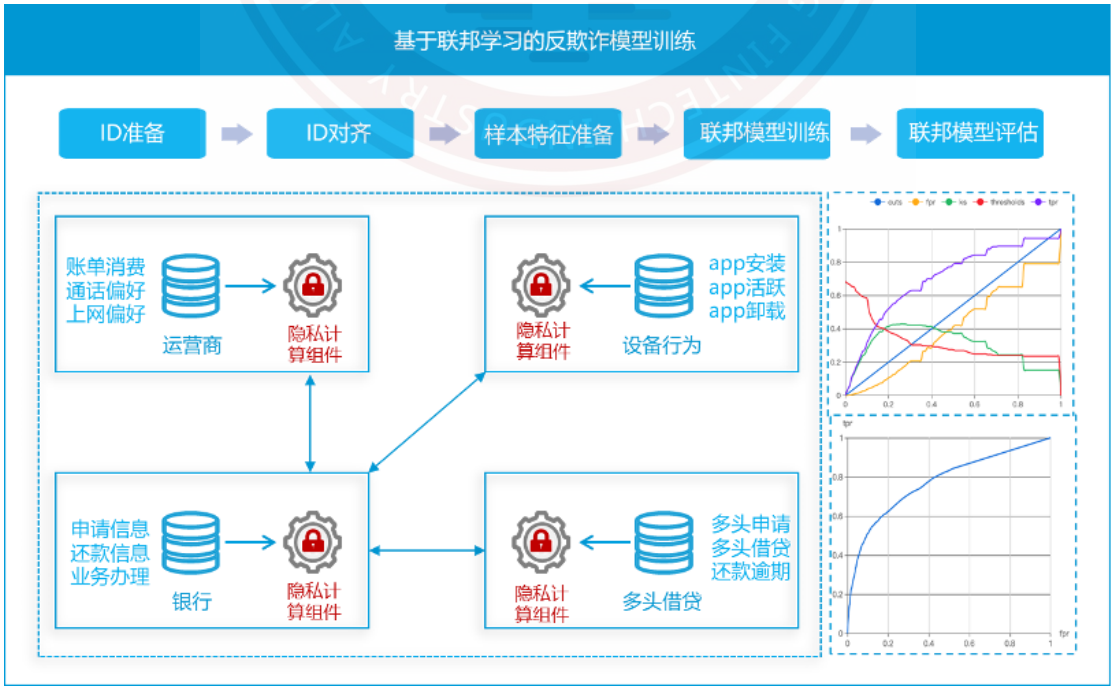


图 11 基于联邦学习的反欺诈模型训练

针对该银行在贷前反欺诈场景下的建模需求，分别引入了运营商数据、多头借贷数据、设备行为数据，提供了上千维度的建模特征。在本次反欺诈联合建模项目中，行方提供了行内用户的基本申请信息、信贷产品欺诈信息等；多头借贷数据源机构提供了多头申请、多头借贷、逾期等信息；设备行为数据源机构提供了 app 安装、活跃、卸载等信息；运营商数据源机构提供了账单消费、通话、上网等信息。

整个联合建模的过程均基于隐私计算平台。在 ID 对齐阶段，运用基于安全多方计算技术的隐私求交（PSI）技术，建模各方的 ID 信息经过多层加密，保证求交过程中各方 ID 信息的安全，同时各方非交集 ID 互不泄露。在联邦建模阶段，底层基于联邦学习技术，各方的原始数据在各自私域服务器上、全程不出各自私域，模型训练过程只传输、交互经过加密的中间计算结果。

（4）隐私计算技术应用成效

通过隐私计算技术，银行机构可以安全地借助外部数据来补充特征维度，构建反欺诈评分模型。经过模型训练，利用外部数据联合构建的反欺诈模型 AUC 达到 0.75+，KS 达到 0.4+，模型效果比只使用行内数据构建的模型提升约 11%。

### 3. 基于多方安全计算的金融反欺诈案例

（1）背景介绍

近年来，在互联网、大数据及人工智能等新兴技术的驱动下，商业银行积极开拓创新产品及服务，业务活动日趋复

杂。与此同时，不法分子依托新兴技术手段，通过银行渠道进行诈骗活动。特别地，信用卡欺诈行为逐渐呈现出组织化、移动化、隐蔽化、场景化等特征，并已形成灰黑产业链，对居民资金安全以及银行业务安全造成严重威胁。在贷前环节，虽然银行能够通过其自身沉淀的黑名单信息在事前筛查并阻挡在本行有欺诈记录的客户，但其并无法了解客户在其他机构的历史行为。然而，金融机构间往往出于数据安全、商业秘密等顾虑而拒绝共享黑名单数据及相关解释逻辑信息，导致各机构对客户金融行为的了解仅局限于自身渠道，无法覆盖客户全面的行为信息。面对信用卡欺诈活动逐渐形成配合严密的产业链条的形势，商业银行普遍面临数据壁垒严重、风控信息维度不全等问题，行业整体欺诈侦测能力难以提升。在此背景下，商业银行亟待建立数据安全共享机制，破除风险信息壁垒，从而强化行业信用卡联合反欺诈能力。

（2）应用场景合规风险分析

1）交换或融合的数据

在信用卡贷前环节，银行间可通过共享信用卡欺诈黑名单，查询申请人是否在其他银行存在欺诈行为，从而在事前有效甄别欺诈可疑主体。其中，被共享的数据主要为申请人的姓名和身份证号。

2）涉及的相关法规及合规风险分析

上述数据的共享受到一系列法规约束。在《个人金融信息保护技术规范》要求金融机构原则上不应共享、转让其收

集的个人金融信息，确需共享、转让的，应充分重视信息安全风险。《个人信息保护法》规定，个人信息处理者应通过制定内部管理制度和操作规程、对个人信息实行分类管理、采取相应加密、去标识化等安全技术措施等手段，防止未经授权的访问以及个人信息泄露、篡改、丢失。因此，若银行间的信用卡欺诈黑名单以明文形式共享，存在较高的个人身份证数据泄露及滥用风险，银行面临的数据合规压力大、成本高。

（3）解决方案（技术架构）内容

为破解数据壁垒，实现各方数据安全合规融合应用，某A银行和B银行采用多方安全计算技术实现了隐匿查询功能，通过安全共享信用卡欺诈黑名单信息，形成了贷前环节的信用卡联合反欺诈试点。具体地，A 银行的黑名单数据通过数据服务节点(DS)以密文随机化碎片的形式(即“计算因子”)输入多方安全计算平台，B 银行将信用卡申请人的身份证号由数据服务节点（DS）以计算因子的形式输入多方安全计算平台。随后，输入数据由平台上的不同计算节点（ES）执行计算，进行隐匿查询功能。最后，计算结果（即 B 银行所查询的身份证号是否在 A 银行的信用卡欺诈黑名单上）通过数据服务节点以密文形式（即“输出因子”）输出至 B 银行，并被解密使用。在此方案下，A 银行和 B 银行能够协同分析黑名单数据，事前拦截有历史欺诈行为的个体，并对可疑线索进行持续监控，谨慎防范欺诈风险，方案如图 12 所示。

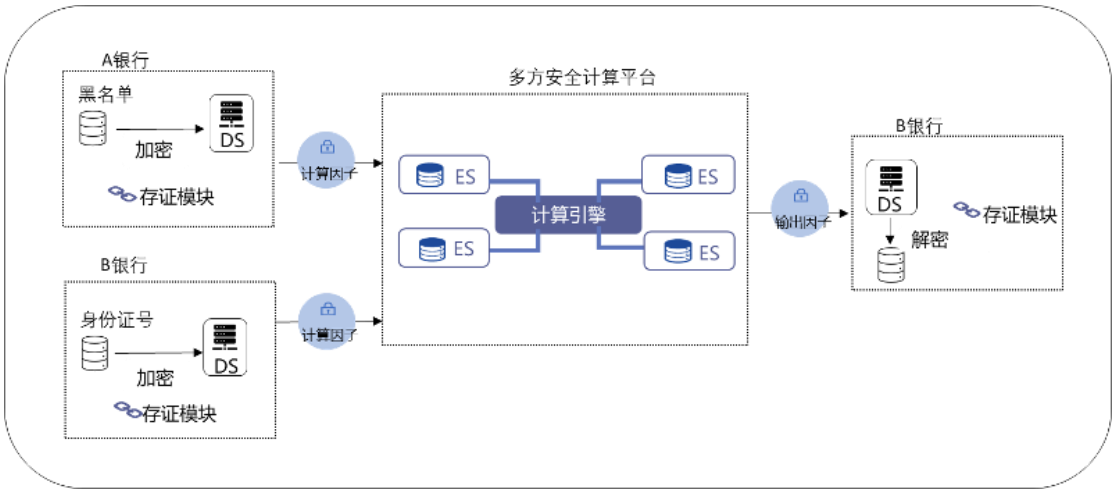


图 12 多方安全计算隐匿查询架构图

在该案例中，多方安全计算技术助力银行提升了关于《个人信息保护法》对于匿名化要求的数据合规，有助于让黑名单数据及身份证号等数据在非受控环境下无法重识别，降低了用户隐私泄露的风险，有效避免了《个人金融信息保护技术规范》中强调的信息安全风险。另外，黑名单数据及个人身份证号被转换为密码学安全的计算因子，并被送至不同计算节点执行隐匿查询。在此过程中，各计算节点互不串通，无法恢复数据原始信息。因此，上述数据无法被任何一方获取并用于目标场景以外的用途，使得数据处理活动和用途与场景目标直接相关，且被限定在必要的范围之中，满足《个人信息保护法》只不过的个人信息收集、处理的最小必要原则。

（4）隐私计算技术应用成效

在该案例中，多方安全计算技术通过满足一定条件下的匿名化，并有效限定黑名单数据在必要的使用目的范围之中，能够助力银行在提升数据合规能力、降低合规成本的前提下，

实现黑名单数据安全共享，为银行业破解数据壁垒、共享风险信息、提升联合反欺诈水平做出标杆性示范。

**4. 基于多方安全图计算的中小微企业融资反欺诈案例**

（1）背景介绍

近年来，中小微企业融资中的金融欺诈问题日益严重，涉及金钱和服务的商业模式（通信、保险、贷款、信用卡申请等多领域）几乎都会受到欺诈攻击。针对金融欺诈问题，金融机构通常利用反欺诈规则模型或机器学习模型来预警，这两种方式都是从历史案例中发现金融欺诈时重复出现的个体行为模式。但随着时间演化、发展和反欺诈技术进步，金融欺诈团伙呈现有组织欺诈趋势，市场急需新技术来对传统反欺诈技术进行补充。

图计算是以图作为数据模型来表达问题并予以解决的过程，需要解决图数据组织及划分、顶点程序调度和计算通信模式等关键技术，目前主要应用在网页排序、社区发现、最短路径等问题。一些电商平台在智能风控场景中也有图计算的应用实践，但并未与多方安全计算技术进行结合，数据隐私上无法得到保障，难以满足合规要求。

基于多方安全图计算的中小微企业融资服务，是国内金融领域首例对外公开运行的多方安全图计算应用。应用基于各自管控的多方安全计算系统平台，确保银行和运营商两方数据不出库，联合建模应用于小微企业普惠金融业务，精准

识别团伙欺诈行为，助力金融机构提高风控能力，通过解决中小企业融资难问题，安全高效服务实体经济。

（2）应用场景合规风险分析

1）交换或融合的数据

本项目通过融合运营商和银行的自有图数据，构建一个虚拟的全局图网络。其中：运营商数据主要包括通话数据、设备数据、短信数据、位置数据等；银行数据主要包括联系人数据、担保数据、交易数据、信贷数据等。

2）涉及的相关法规及合规风险分析

依据中国人民银行发布的《个人金融信息技术保护规范》，个人金融信息全生命周期各阶段都存在对应的风险。在数据采集时，存在用户数据采集目的、方式及范围等是否得到用户授权、是否脱敏等隐私风险；在数据使用时，存在数据是否脱敏、是否去标识、数据是否被使用方故意泄漏、数据计算结果是否保密等隐私风险；在数据存储时，原始信息是否脱敏，是否随意访问，是否存在数据泄露等隐私风险。所以本项目需要考虑数据全生命周期的合规风险，尤其是在数据联合应用的场景下，其应用流程都是非标准化的过程，存在数据流出、泄漏、复制等各种风险。

（3）解决方案（技术架构）内容

通过将图计算和多方安全计算技术融合，在确保各方原始数据不出域的基础上，实现银行和移动运营商关系图谱数据的融合，并基于此建立有监督和无监督反欺诈模型，对用

户和企业欺诈风险进行识别；以融合关系网络的异常监测和离群行为预警等技术手段作为补充，实现全方位欺诈风险预警；图计算关系网络准确识别企业集群背后的复杂关系链条及欺诈风险，助力银行中小微企业精准贷款投放和集群风险管控，提升金融机构风险防控能力及客户贷款体验，并使移动运营商多维度数据价值得到有效发挥，演示如图 13 所示。

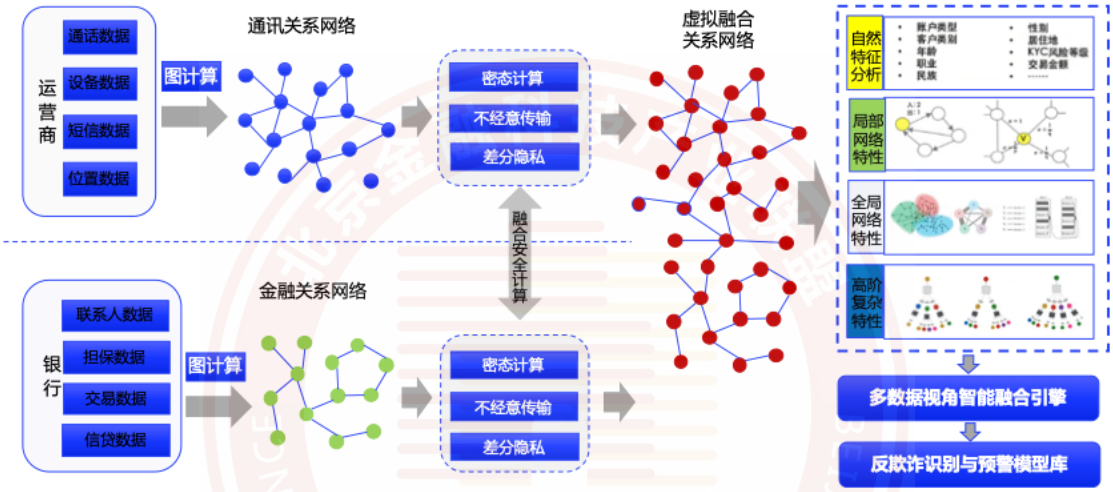


图 13 多方安全图计算技术架构演示图

在模型层面，项目以多元技术融合突破应用边界，采用非对称、独立对称、融合对称三种安全图计算模式，结合差分隐私、密态计算、不经意传输等隐私计算技术，虚拟融合银行及运营商数据关系网络，通过对齐、扩充、传播、聚合等方法刻画关系图谱网络，进行用户自然特征、局部网络特征、全局网络特征和高阶复杂特征分析，在多数据视角智能融合引擎基础上构建反洗钱反欺诈识别预警模型库。其具体流程如下：

1）银行侧对用户黑名单中的手机号进行哈希，哈希方法无法逆推得到用户手机号明文，具有数据去标识化的效果；

2）银行侧对用户黑名单属性值进行半同态加密得到密态值并发送至运营商，密态值也无法逆推得到属性值明文。在这个过程中，同态加密技术保证先计算后解密的效果等同于先解密后计算，因而可以减少计算过程中泄露明文信息的风险，能够保护银行的黑名单信息。

3）在运营商侧，根据哈希后的手机号找到本侧图网络中对应的中心节点，并将中心节点与关联节点以连线的方式刻画出本侧的用户关系。在运营商对来自银行的用户黑名单属性密态值进行密态计算后可得到关联节点的权重值，并在本侧对相应用户进行染色处理。

4）银行侧在进行运营商黑名单查询时，可以通过输入哈希后的手机号，由运营商侧返回节点的黑名单密态值，再由银行侧解密得到用户黑名单属性，从而辅助银行风控人员对用户欺诈行为进行预判。

除此之外，本项目严格遵循央行“用户授权、最小够用、全程防护”的数据治理原则：在数据采集时，通过隐私政策文件等方式明示用户数据采集和使用目的、方式及范围，获取用户授权后再进行采集；在数据使用时，借助标记化、加密等技术，在不归集、不共享原始数据的前提下，仅向外提供脱敏后的计算结果；在数据存储时，通过不可逆加密等技术将原始信息进行脱敏，并与关联性较高的敏感信息进行安全隔离、分散存储，严控访问权限，降低数据泄露风险。同

时，全程都融合高性能加密算法及K-匿名、L-多样性、差分隐私等技术，对敏感数据潜在风险进行检测、预判和处理。

（4）隐私计算技术应用成效

图计算和关系网络的建立，离不开大量数据的使用和分析，而目前对数据的监管愈发严格，加强个人隐私保护和数据安全是国内外的趋势。因此，要提高关系网络对欺诈行为的识别能力，必须处理好数据服务和数据安全及隐私保护之间的关系。多方安全图计算技术提供了全新的反欺诈分析手段。

本项目首次将图计算技术与多方安全计算技术结合，为隐私计算行业首创和重大突破，在技术原理、方案设计、系统搭建、科研专利方面具有极强的前瞻性、领先性、创新性和应用性。基于银行欺诈客户个体信息与移动运营商关系网络，精准识别高危客户群体，主动分析和探查，有效规避欺诈风险；通过对参与方数据进行拓扑图分析，从单点式向集群性银行风险防控技术演进，解决信息不对称、计算结果不准确等问题，实现银行风控模型精准化；通过群体性欺诈开户风险模型，部署相应的监测规则，识别金融欺诈团伙，预警群体性客户的联合欺诈风险。

该项目通过多方安全图计算等技术作为开户真实性意愿审核的辅助手段，达到身份核验手段多元化效果，可视作纸质材料、现场走访审核的有效补充；通过大数据、人工智能等技术手段，依托行内外多维度数据，通过“标准化产品、

数字化风控、集中化运营”为小微企业提供专属线上融资服务，满足客户差异化融资需求；通过将客户移动网络信息与申贷信息进行交叉比对分析和联合建模，在金融反欺诈的识别效率和精确度方面均有大幅提升；精准防范和打击伪冒申贷等扰乱金融秩序的行为，营造健康的普惠金融生态。

### 5. 隐私计算助力医疗保险智能理赔反欺诈应用

（1）背景介绍

传统的保险事后理赔服务中，不仅需求审核纸质的理赔材料，过程更是依靠审核人员基于保险+医学经验积累，人工输出核保理赔结论。不仅流程繁琐，还将耗费较高的人力成本在工作中，且潜在风险较高。如果能够优化这一过程，将有效降低保险运营成本，将更多资源用以服务更广泛的人群，扩大普惠医疗的覆盖范围和服务质量，助力大健康产业发展、大健康服务完善。

商业健康险作为促进多层次医疗保障体系建设的重要组成部分，对于国民医疗健康具有重要意义。2022 年 1 月，原中国银保监会人身险部向全国各人身保险公司下发《关于印发商业健康保险发展问题和建议报告的通知》：“争取与医疗机构信息系统实现充分信息共享，改进结算服务；在确保信息安全和个人隐私权的基础上，强化医疗健康大数据运用，推动医疗支付方式改革，更好地服务医保政策制定和医疗费用管理。”

为了解决千万级在保用户医疗险的理赔体验、成本和效能问题，某财险构建了基于理赔科技平台和隐私计算框架的智能理赔系统，借助数据优势（十万级典型理赔案件）并辅以一定的知识约束，实现医疗凭证深度结构化和“专家级”高置信辅助核赔决策。系统基于百万医疗凭证图像的机器学习，实现了视觉识别+文本分类+文本语义理解的多模态医疗凭证识别模型，对100+种医疗凭证的识别准确率达到95%以上。

随着大数据人工智能等技术的发展推动健康险迈入 3.0 时代，保险行业数字化转型中服务逐步实现向线上迁移，数据流通为保险理赔模式优化带来无限可能，监管趋严的环境下，保险公司在创新服务模式的过程中需要优先重视数据合规，隐私计算成为中立可信技术支撑。在整个系统中，健康险定制多方数据联合分析解决方案是核心模块之一。这一方案多方安全分析（Secure Collaborative Query Language，简称 SCQL）功能，该功能基于多方安全计算技术，使用安全加密算法将多方数据进行联合分析的服务。帮助 XX 财险及其外部医疗数据 ISV 在原始数据不离开本地、数据价值有保护的前提下，进行联合分析。

（2）应用场景合规风险分析

1）交换或融合的数据

保险机构侧：报案数据

第三方数据机构侧：医疗诊断数据

2）涉及的相关法规及合规风险分析

依据个保法、《信息安全技术健康医疗数据安全指南》等相关法律法规要求等，联合分析过程需保证参与方用户数据可用不可见、用户隐私不泄露。

（3）解决方案（技术架构）内容

通过引入创新的 SCQL（MPC SQL）多方联合分析领域专用语言，具体通过如下流程合规引入外部数据源，保障了联合项目各参与方的数据隐私，示意如图 14 所示。

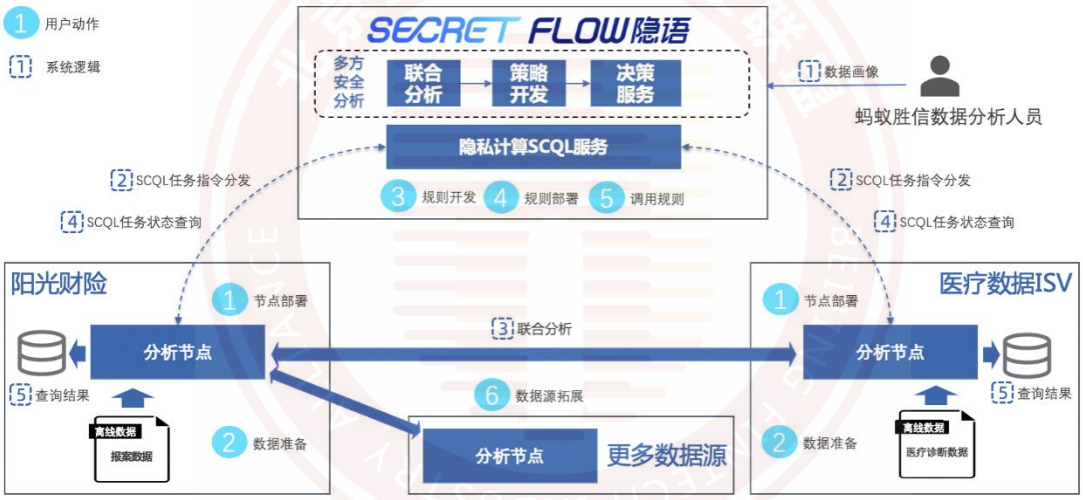


图 14 基于多方安全计算的保险理赔联合分析业务流程

1)节点部署:用户一键 get 部署包,填入节点标识&token 信息后执行脚本，即可轻量化、小时内完成本地节点部署;

2）数据准备（进行前置安全配置）：双方分别将各自的样本数据，加载至各自本地分析节点，并在平台上注册对应样本的数据表结构，同意授权进入多方安全分析项目。在这一环节中,用户可通过隐语的 CCL 前置安全配置功能,在 MPC 相关技术能力支撑下，用户可对数据资产进行分级分类，通

过前置配置来保证安全级别高的数据的安全性，保证多方隐私数据在计算过程中不泄漏。

3）规则开发：基于的丰富 MPC SQL 算子支持，用户可以在脚本中描述基于多个数据源的安全计算，通过“SELECT FROM”“JOIN ON”“GROUP BY”等语句的组合搭配，即可完成联合分析的统计结果生成；通过“SELECT INTO”语句可将交集结果导出至节点本地。

支持：算术计算（+，-，*，/，%）、比较（>，<，>=，<=，=，<>，IN，NOT IN）、逻辑计算（AND，OR，NOT）、窗口聚合（group by ... min，max，avg，sum，count，median ）、控制（IF，CASE WHEN）、排序（RANK，ROW NUMBER，ORDER BY）、日期函数（DATE_DIFF，DATE_ADD）、其他函数（ceil，floor，round...）丰富算子。

4）规则部署：用户可使用 ISV 授权的数据，通过联合分析提供的在线 SCQLIDE，在平台完成在线调试优化规则，且经调试验证后的规则可作为标准规则在更多数据源进行规模化部署。

5）调用规则：规则部署完成后，用户在平台端即可对规则发起调用，平台支持简单的数据分析结果可视，如就诊分布、就诊品次频次等分析结果。

6）数据源拓展：用户还可持续通过外部医疗数据引入进一步提升丰富底层数据能力，进一步提升自身理赔和风控能力，增加核保场景智能决策服务。

该方案可以实现参与方原始数据不出本地节点，通过隐语的 CCL 前置安全配置功能，在 MPC 相关技术能力支撑下，可对数据资产进行分级分类，通过前置配置来保证安全级别高的数据的安全性，保证多方隐私数据在计算过程中不泄漏，从而在保证各参与方数据安全的前提下，完成联合分析。

（4）隐私计算技术应用成效

基于 SCQL 多方联合分析领域专用语言的健康险定制多方数据联合分析解决方案，有利于提升阳性案件识别和调查路径规划能力，覆盖了全国 10+省份的医疗数据，为医疗数据的合规使用提供了一种典型案例。利于有效发现阳性线索、降低错赔风险，通过数字化调查审核控制了理赔运营成本。

智能理赔系统上线后，好医保 30 日线下调查案均成本降低了 40%，赔付率控制在了合理水平，保障了业务持续健康发展；30 日人均结案数量提升了 70%，有效控制了案件积压风险，为该产品规模的进一步扩充留出充足的理赔人力空间。

对于在保用户，无效报案实时拦截率从 0 提升到 50%，高标准满足保险法 1 日内补传材料的要求，不仅显著提升用户报案体验，节省的运营成本更有利于扩大普惠医疗的服务范围、提升普惠医疗的服务效率。

对于用户而言，保险理赔包含用户报案环节和机构审核赔付环节，同时不同险种、不同重疾往往所需的理赔材料各有差异，凭证材料繁多且专业。首先线上报案时，用户理解、

筛选凭证具有难度，很难一次完成提交工作，据相关数据：在重疾险线上报案过程中，用户能够一次性提交完整材料的占比小于 30%；然后进入核赔过程，机构人工核查同样存在耗时长的问题，且风险较高。基于多方数据联合分析方案，智能理赔系统实现了可自动根据病种对用户的凭证把关，给予用户指导降低报案复杂度，并以数字化核赔辅助取代传统的人工核赔，提高核赔效率，报案快加核赔快全面提升了用户的理赔体验。

对于行业而言，现阶段我国医疗行业的数据流通主要发生纵向同场景/同类型机构之间，横向跨场景/多类型机构之间的数据流通发展较缓，此次健康险定制多方数据联合分析解决方案的落地不仅有利于商业健康险的降本增效良好发展，更可拓展应用于医疗行业中的前沿技术合作、创新药研发、高端医疗器械研发应用以及疾病风险评估、疾病预防、分类诊断等众多场景，联通多种类型的医疗健康数据。

### （三）反洗钱

#### 1. 基于联邦学习的银行反洗钱领域应用

（1）背景介绍

近年来，随着严厉打击金融犯罪成为全球金融监管的共识，中国人民银行对于我国金融机构的反洗钱要求不断提升，反洗钱已成为重点监管领域，不少金融机构、支付机构受罚的重灾区。同时，针对信息安全和隐私保护方面，国家陆续出台了《中华人民共和国网络安全法》《中华人民共和国国

家安全法》《中华人民共和国反洗钱法》《个人信息保护法》和《中华人民共和国数据安全法》等多部法律法规，如何使多个参与方在保护数据隐私、满足合法合规要求的前提下进行数据协作，解决数据孤岛问题已然成为金融机构急需解决的问题。

随着互联网应用和技术提升，现在很多洗钱犯罪主体（自然人\法人）利用新技术进行非面对面的交易，并得到受利益诱惑的专业人士的协助，隐蔽性越来越强，执法机构很难直接通过监控犯罪主体来发现洗钱线索。但由于存在现金流通限制，绝大多数的“黑钱”都要在银行或者其他金融机构间流转并留下记录，因此，可疑交易信息进行采集、分析、监测资金交易的异常变动成为发现洗钱线索的最重要的途径。在数字化进程推动下，越来越多的金融机构正在探索引入外部数据，采用隐私计算技术助力反洗钱的开展。通过隐私计算形式，联合多方在保证原始数据不出库的前提下，帮助银行实现反洗钱的联合建模和分析，实现“数据孤岛互联”“数据隐私保护”和“反洗钱业务发展”三者间完美平衡。

（2）应用场景合规风险分析

1）交换或融合的数据

银行侧：银行用户个人的金融信息等;

运营商侧：个人移动轨迹、上网轨迹、个人通信数据等敏感的个人信息;

2）涉及的相关法规及合规风险分析

《个人信息保护法》《数据安全法》等均对个人隐私信息的保护作出了明确的要求。

银行的个人隐私数据，包括个人的金融信息等是非常敏感的个人数据，需要被高度重视与保护。在与外部的数据联合建模中，银行的数据始终不应离开本地，要求“数据可用不可见”。

运营商的数据具有非常丰富的特征信息，且用户覆盖量较全，是众多领域数据合作的优质数据源。运营商大数据包括个人移动轨迹、上网轨迹、个人通信数据等，这些都属于敏感的个人信息，需要在不泄露个人隐私信息的情况下进行与外部进行联合建模等数据协作。

（3）解决方案（技术架构）内容

当前，银行的反洗钱的常用做法是基于内部交易数据、关联账号等建立规则，用来判断哪些动作可归属成洗钱行为，据此搭建模型，应用于实时监控预警和事中控制。然而，由于数据覆盖范围较小且新的洗钱形式不断涌现，传统反洗钱模型召回率和精准率不足，可能会直接影响到判断效果，从而降低信息报送的效率和质量，如图 15 所示。

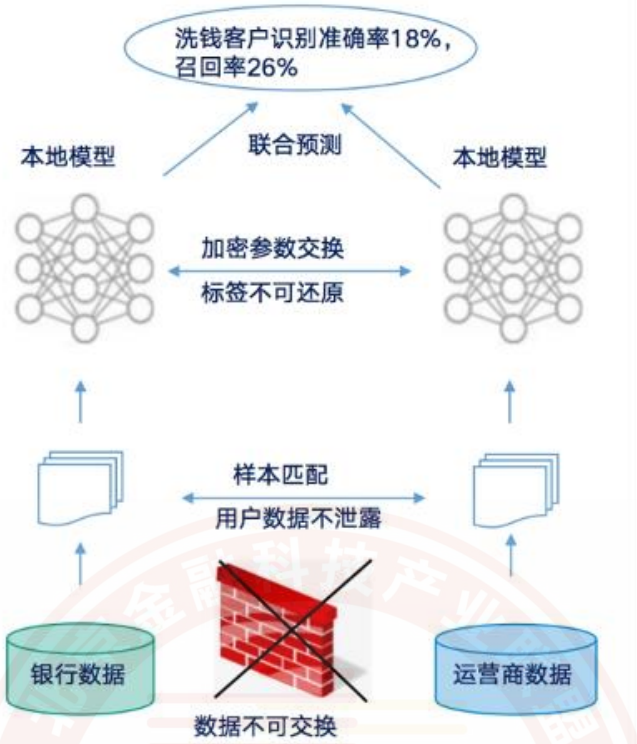


图 15 某银行与通信运营商数据协作示意

银行反洗钱监测中心仅用自有的数据特征对客户洗钱行为的识别能力尚存局限，而通信运营商拥有较完整的缺失数据特征。如果能够引入通信运营商等外部数据，从更多维度识别用户风险，将能显著提高洗钱行为识别的准确率。由于相关法律法规和通信运营商自身数据安全规定的限制，银行无法直接获取。具体而言，银行和通信运营商均存在隐私保护的需求。一方面，银行不希望客户的隐私信息和业务特征在数据协作中被泄露。另一方面，通信运营商也希望在满足业务前提下，保障数据不出门。

因此，在反洗钱场景应用实践过程中，应用联邦学习等隐私计算技术，帮助银行引入通信运营商数据，为银行反洗钱业务中补充了 50+对应特征维度，帮助双方进行隐私计算

联合建模，通过隐私保护集合求交技术对双方用户 ID 进行匹配，帮助双方在隐私保护对前提下进行安全合规对数据协作，提升反洗钱识别能力。

通过联邦学习技术，成功将银行和通信运营商的数据特征结合到一起，建立了基于纵向联邦学习的反洗钱模型，在双方不交互原始数据的前提下，为反洗钱场景模型补充了更多有效数据，模型训练如图 16 所示。

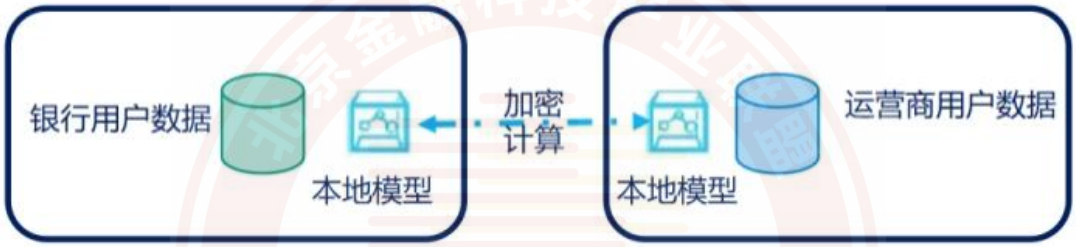


图 16 模型训练示意图

此外，银行通过隐私计算平台，让通信运营商的部分用户个人数据在不出库前提下通过隐私计算平台进行模型训练（联邦学习）和业务规则（联合统计分析），打造反洗钱模型。

（4）隐私计算技术应用成效

反洗钱一直以来是各金融机构监管重点之一，隐私计算技术产品在服务国家战略、辅助金融监管等领域的创新应用，可以助力金融数据的安全开放和应用，更好地开展金融领域数字化和智能化建设，有效防控金融风险。

基于隐私保护的数据协作模式，可以助推同一银行的不同部门间，跨银行间、银行与其他机构间的多方数据安全协

作，提高行业整体业务水平、降本增效，在银行风控及多业务领域开拓创新，如银行联合黑名单、反洗钱、反欺诈等场景。

## 六、隐私保护技术在金融领域应用的展望

随着未来科技的发展，将带来更大的数据爆发、更便捷的连接与交互，及人文与数字世界的深度融合，而这些势必带来更迫切的隐私保护需求。聚焦于金融领域，随着国内个人金融信息保护的法律法规不断完善，将不断促使金融机构严格按照合法、正当、最小必要原则收集、使用和保管客户信息，规范机构内部为商业目的使用个人信息的行为，充分保障消费者隐私和合法权益。而隐私保护技术在金融领域的应用为数据安全共享创造了许多可能性，重塑了用户、企业和政府之间任关系，为数字经济的发展奠定了安全基石。在未来，隐私保护技术在金融领域的发展方向:

### （一）金融隐私保护技术发展方向

#### 1. 提升运算性能

当前的隐私保护技术受限于相关协议以及算法的计算复杂度、通信轮数、通信量等因素，对复杂的联合建模，海量数据融合处理等应用场景的支撑还有待提升。通过设计高效的交互协议、优化基础算法、离线技术辅助等方法提升计算和通信效率将大幅提升隐私保护技术的实用性和拓展性，从而推进隐私保护技术在更多金融场景中快速落地。

### 2. 提升安全可靠性

当前使用的隐私保护技术受运算、通信性能的影响，只能满足半诚实敌手下的安全运算，即参与运算中的每一方都是诚实的，严格按照协议执行算法，若其中一方作恶，则不能保证协议的安全性；同时，在协议执行的过程中也可能面临用户掉线、网络延迟等实际问题，因此，设计安全且可靠的协议也是未来进一步提升隐私保护技术的方向之一，从而切实保证数据在计算过程中的安全性。

### 3. 推进技术融合

单一的隐私保护技术无法支撑整个应用的全部功能，不同的技术路径之间通过相互融合，形成优势互补。同时发挥硬件环境与软件技术各自的优势，通过多模块化综合方案构建通用的隐私保护技术平台，实现基于用户同意的数据服务，进而解决实际的金融隐私问题。

## （二）金融隐私保护应用发展方向

### 1. 推进个人金融信息保护应用发展

随着隐私保护技术的不断进步，更加安全和高效的隐私计算平台将推进个人金融信息保护应用的发展。通过使用隐私保护技术对海量的个人金融数据进行融合处理分析。在保证数据可用不可见、数据不动价值动的前提下，实现联合金融建模与风控，促进数据的有序流动，充分释放数据要素潜能，进而推进金融领域的数字化发展，同时保证用户个人金融数据安全，提升用户与企业之间的信任关系。

### 2. 拓展应用场景

通过探索隐私保护技术在个人金融保护、反欺诈等不同业务场景的具体实践，如面向个人的数据信托和数据权利服务、基于隐私保护的个人边缘代理，以及面向企业的数据访问控制、基于隐私计算的数据共享、基于隐私增强技术的数据分析等，从而在隐私法规和公民数据安全的基础上，拓展出更多有价值的金融创新应用。

### 3. 完善相关法律法规及应用规范

未来，个人将更加强调对数据主权的自主控制，企业也将高度重视用户的隐私保护，相关隐私应用应始终配合法律、政策、标准等相关制度，进一步落实合规要求，同时相关法律法规也应不断完善以适应隐私保护技术的发展，不断提升隐私保护应用在接口使用，产品开发过程中的规范性，从而实现整个隐私保护生态的安全合规。

### 4. 提升产品的可操作性

面对不断增长的个人隐私需求，企业在提供可信、安全、规范的服务时需要考虑产品的敏捷性，灵活性以及用户友好的可操作性，使隐私保护平台能够提供简洁的接口，助力金融隐私应用快速搭建。

# 参考文献

[1] 中国人民银行.《关于银行业金融机构做好个人金融信息保护工作的通知》银发〔2011〕17号。

[2] 中国人民公安大学，江苏通付盾科技有限公司.《金融隐私保护问题分析及对策》(2020)

[3] 郑晓琴，杨燕明，吕伊蒙等.隐私技术与个人信息保护[J]. 中国金融，2021:20.

[4] 中国人民银行.《个人金融信息保护技术规范》JR/T 0171-2020

[5] World Economic Forum. The Next Generation of Data-Sharing in Financial Services: Using Privacy Enhancing Techniques to Unlock New Value (2019)

[6] The Royal Society. Protecting privacy in practice (2019)

[7] European Union Agency for Cybersecurity. Data Pseudonymisation: Advanced Techniques and Use Cases (2021)

[8] 贾轩，白玉真，马智华.隐私计算应用场景综述[J].信息通信技术与政策，2022(5):8.

[9] 中国信息通信研究院.隐私保护计算与合规应用研究报告（2021年）

[10] 隐私计算联盟，中国信息通信研究院云计算与大数据研究所. 隐私计算法律与合规研究白皮书[R]. 2021-12.