2021

# 隐私机密计算蓝皮书

## 数据向善

# 报告编写组

**参编单位（排名不分先后）：**

中国移动通信联合会

中国科学院信息工程研究所

中国信息通信研究院

上海数据交易中心

四川省生物信息学学会数据共享与安全分会

浙江大学人工智能研究所

杭州锘崴信息科技有限公司

光之树（北京）科技有限公司

杭州金智塔科技有限公司

软通动力信息技术（集团）股份有限公司

每日互动股份有限公司（个推）

厦门建发信息技术有限公司

深圳市力合微电子股份有限公司

北京久其软件股份有限公司

连连数字科技股份有限公司

荣科科技股份有限公司

北京灵伴即时智能科技有限公司

科技谷（厦门）信息技术有限公司

浙江省数据安全服务有限公司

联通数字科技有限公司

**参编学者专家（按姓名笔画排序）：**

| | | | | |
|---|---|---|---|---|
| 王　伊 | 王　舒 | 王　帅 | 王文浩 | 王华雄 |
| 王润垠 | 王　榕 | 王运祥 | 叶新江 | 戎艳中 |
| 刘元成 | 孙　林 | 李　帜 | 张　珣 | 张佳辰 |
| 陈　斌 | 陈　博 | 陈思恩 | 林　明 | 周诗文 |
| 庞在虎 | 郑　灏 | 郑小林 | 赵金清 | 聂拥军 |
| 倪健中 | 黄　颖 | 龚佶敏 | 温双有 | 颜亦军 |

**特别鸣谢以下专家对于本书的指导和建议：**

李正茂　中国电信集团总裁

李慧镝　中国移动集团副总裁

倪光南　中国工程院院士、中国科学院计算机研究所教授

张　平 中国工程院院士、北京邮电大学教授

闻　库　工信部通信发展司原司长，中国通信标准协会秘书长

陆书春　中国互联网金融协会秘书长

谢麟振　工业和信息化部电子司原副司长、北京大学教授、博导

蒋林涛　中国信息通信研究院科技委主任

周鸿祎 360集团董事长

徐文伟 华为公司董事、战略研究院院长

吕述望 中国科学院大学教授、信息安全国家重点实验室原主任

黄铁军 北京大学信息科学技术学院教授、计算机科学技术系系主任

蔡维德 北京航空航天大学教授、互链网创始人

贺知明 电子科技大学广东电子信息工程研究院 副院长、教授、博导

张同须 中国移动研究院院长

张云勇 中国联通产品中心总经理

沈红群 中国移动金融科技公司董事长

耿学峰 中国移动通信集团公司技术部副总经理

宋雨伦 联通数字科技有限公司副总裁、数据智能事业部总经理

黄 颖 软通动力信息技术(集团)观股份有限公司 集团副董事长

杜正平 京东集团前副总裁、华为云前副总裁

朱 波 原华为互联网业务总裁

# 版权声明

# 目录

# 序言

隐私机密计算自从上世纪80年初中国学者姚期智教授发表第一篇安全两方计算相关论文起，至今已经将近半个世纪。2013年，又一位中国学者王爽教授提出并发表了第一篇联邦学习的文献，使得相关技术从学术领域进入到工业领域。随着数据应用行业的发展，隐私机密计算近年来得到迅猛发展。

**江山如此多娇，引无数英雄竞折腰**

数据要素和隐私机密计算是百年未有之大机遇，中国移动通信联合会密切跟踪关注这一人类史上罕见的大技术趋势，认识到隐私机密计算将会为互联网、IT新基建、数据智能等各个领域带来“**数百年未有之大变局**”，它将和区块链技术一起，重构整个IT产业，进而推动社会的又一次变革。

中国移动通信联合会，作为国家重要行业协会，顺应产业发展潮流，相应成立了数据融合委员会；并携手产学研各界领先的专家学者，一线的市场技术管理者，费时近半年组织编写这一蓝皮书。从发展历程，系统构架到技术体系和业务场景，多角度多维度对隐私机密计算进行了广泛而深入研讨。越深入越发现，隐私机密计算对行业、社会乃至国家发展的重要性和及时性。

首先《中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见》提出了数据作为生产要素的高度，隐私机密计算是这一政策基础性、核心性的技术支撑，有望推动整个IT产业的革新和发展。其次，在数据隐私、数据安全得到空前重视的今天，在面临以欧盟GDPR等相关法案的国际竞争的今天，隐私机密计算技术具有特别的意义，能够配合相关的法律法规政策，在隐私保护和共享共用之间取得良好平衡，为促进我国数字经济的发展、”一带一路“战略、数据安全跨境流动等方面起着基础性支撑作用。

中国移动通信联合会认为隐私机密计算将会带来一个巨大的全新的数据价值交易市场，随着IT系统的更换的生命周期，不久将会迎来一次大规模的隐私计算新基建建

设。隐私机密计算将成为下一次工业革命的引擎，推动工业互联网发展壮大。我国已迅速发展成为隐私机密计算主要的产业发展地区。当然，也应该看到隐私机密计算技术发展还处于初级阶段，有大量工作需要做，相关行业对其还不甚了解。本蓝皮书就是为了各行业对其有一个初步的、系统的、客观的理解。

**俱往矣，数风流人物，还看今朝**

中国移动通信联合会愿意与各界携手，促进相关产品技术的落地应用。推进隐私机密计算的发展，做大做强我国数据产业。推动数据要素化，共同为实现中华民族伟大复兴贡献自己的力量。

**倪健中**

中国移动通信联合会执行会长

2021辛丑年端午

# 1. 隐私机密计算趋势

## 1.1 隐私机密计算的技术发展简史

隐私机密计算是一个系统工程技术，来源于当代密码学、数学、硬件等多个领域。

当代密码学起源于 1977 年，Ron Rivest、Adi Shamir 和 Leonard Adleman 发明了非对称式加密（又称公开密钥加密）算法 RSA，突破了长期以来的瓶颈，达到了新的阶段。密码学通过数学理论将数据转化为密文状态，无私钥不能读取其内容，解决了不安全环境下隐私存储与通信的问题，但在使用环节存在空白。当信息拥有者不得不提交数据使用第三方服务时，他就面临着信息泄露的风险，其他环节的加密状态也就失去了意义。针对这种情况，学术界开展了加密状态下进行数据计算的研究。

1978 年 Ron Rivest、Leonard Adleman 和 Michael L. Dertouzos 提出了同态加密问题，并在同年提出了满足乘法同态的算法 RSA。在此之前，密码学研究关注的都是数据在存储和传输过程中的静态安全，而同态加密问题的提出将加密技术的研究从静态引向动态，是理论上的巨大革新，也开创了隐私机密计算的先河。

1982 年，百万富翁问题引入了多方安全计算概念。 姚期智教授在他的论文《Protocols for Secure Computations》中提出了百万富翁问题，即两个百万富翁在没有可信第三方、不透露自己的财产状况的情况下，如何比较谁更富有。

**20 世纪 80 年代**，MIT 研究员 Shafi Goldwasser、Silvio Micali 和 Charles Rackoff 提出了零知识证明的概念。零知识证明涉及两个参与方：证明者和验证者。它的目的是解决如下问题：证明者如何向验证者证明自己拥有某一特定的数据，但证明过程不能透露任何有关该数据的信息。

经过学界的不断研究和发展，以同态计算、多方安全计算和零知识证明为代表的理论进步，为隐私机密计算奠定了坚实的基础，但是这些算法实践中，所需资源巨大，需要条件较为严格，在应用中遇到很多难以克服的困难。

2006年，OMTP工作组率先提出了一种双系统解决方案：即在同一个智能终端下，除了多媒体操作系统外再提供一个隔离的安全操作系统，这一运行在隔离的硬件之上的隔离安全操作系统用来专门处理敏感信息以保证信息的安全。该方案即TEE的前身，TEE(Trusted Execution Environment)，也叫可信执行环境，TEE所能访问的软硬件资源是与外部OS分离的。TEE提供了授权安全软件的安全执行环境，同时也保护资源和数据的保密性、完整性和访问权限。在服务器端利用TEE技术来进行安全计算，也被称为机密计算（Confidential Computing）。TEE是一种较为成熟的技术解决方案，目前已经在商业应用中被广泛使用。

2013年，联邦学习系统构架层面真正的突破来自于由王爽教授团队在SCI学术期刊Journal of Biomedical Informatics发表的《Expectation Propagation Logistic Regression (EXPLORER): Distributed privacy-preserving online model learning》[1]，这是全球第一篇在线安全联邦学习的文献，该论文提出了数据“可用不可见”问题，在不需要分享原始个体数据的情况下，利用多个数据源进行带有隐私保护的联合建模的概念。同年该团队发表了开源联邦学习框架“WebGLORE: a web service for Grid Logistic Regression”[2]，该底层技术服务于多个医疗网络数据的联邦建模需求。通过合理的构架，融合相关的技术，实现了隐私机密计算的真正落地。近年来陆续有很多行业部署实施了相关项目，比较大的项目有 pSCANNER[3]。这些成果引起了业内广泛的重视和参考借鉴。

2017年谷歌在官方博客中发文，提出了联邦学习在移动端的应用。此外，香港科技大学新明工程学讲席教授、计算机科学和工程学系主任杨强教授团队2017年提出了“Distant Domain迁移学习”的理论体系[4]。此外，为了解决数据割裂、数据孤岛问题，杨强教授团队于2018年提出并于2020发表了安全联邦迁移学习论文，结合了联邦学习和迁移学习[5]。

**2019年**，微软云发布基于TEE的机密计算云服务，即Azure Confidential Computing，2020年Google Cloud也跟进发布了谷歌机密计算云服务。

**2021年**，Intel SGX发布最新版本，突破算力及内存限制问题，有效释放商业应用场景。同时，ARM V9架构发布，也将机密计算列为V9架构中的重要内容。此外，多家国内芯片厂商也开始纷纷布局服务器端的TEE及其生态建设工作。

到今天，隐私机密计算已经被广泛接受，开始被应用到商业系统中。同时，硬件、软件、算法领域都在不断进行研究改进，以满足日益增长的业务需求。

## 1.2 隐私机密计算的关键时间窗口

隐私机密计算虽然源自于美国，但是目前，我们国家已经走到了世界前列。如果能够适时地配套有效的政策，发挥我们的体制优势，发挥政府、行业协会、头部企业、行业成员的各自优势，给予必要灵活的法律政策的支持，必能促进隐私机密计算的发展，更能够创造性的配合中央的数据要素化的产业政策。从而抓住当前的发展窗口期，努力形成绝对领先的隐私机密计算的产业优势，为整体的数据产业发展和应对社会经济发展的挑战做出基础性、核心性的贡献。

### 1.2.1 历史层面

我们正处于又一次社会变革的风口，纵观20世纪末，信息产业带动了全世界的经济发展，促使人类的技术生活上升了一个台阶。2019年11月，发布的《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》这一文件中，数据，被定义为新的生产要素，并与传统生产要素，如土地、劳动力等一同参与市场分配。2020年4月，中共中央、国务院印发《关于构建更加完善的市场化配置体制机制的意见》中进一步强调了数据这一新兴生产要素作为信息化建设有力抓手的重要性，并对未来的市场化配置改革方向作出指示。这标志着，主要生产力从工业化向信息化迈进的过程已经开始，同时也意味着，数据的重要性和价值，将被重新定义和认识。

十四五规划中第五篇：加快数字化发展，建设数字中国，其中重点强调了数字经济发展，这里面隐私机密计算可以起到核心基础关键作用：

- 云计算：采用隐私机密计算的云计算将会逐步普及。
- 大数据：采用隐私机密计算的大数据将会替换原有的模式。
- 物联网：采用隐私机密计算的具有更高安全等级的物联网将会被更大范围的采用。
- 工业互联网：工业互联网采用隐私机密计算将会具有更高的安全等级，更强的处理能力。
- 区块链：隐私机密计算和区块链的结合将会形成互补效用，提升应用效果。

在具体的公共服务、智慧城市、数字乡村、数字生活、数字政府方面，隐私机密计算也会起到非常重要的作用。

隐私机密计算不仅仅是局限于隐私保护，而是打开新世界的一个窗户，是一种颠覆性的技术手段。就像是铁器发明那样，新技术产业的掌控者将会具有降维打击能力。随着社会进步，以隐私机密计算为核心的新一代技术革命将会是对整个信息产业的一次洗牌。隐私机密计算将会全面颠覆社会的产业结构，使人类文明进入一个新的阶段。我国经历了40年的改革开放，积累了雄厚基础，拥有广大的市场，有可能在这方面和美国并驾齐驱，甚至有所超越，让中国再次领先。

### 1.2.2 技术层面

2020年《麻省理工科技评论》发布的“全球十大突破性技术”榜单中，入选的“差分隐私”技术引起了各界高度关注。美国政府希望运用这一技术在2020年人口普查中更好地保护公民隐私。毫无疑问，数字文明中的隐私数据保护已经成为大数据市场的核心问题。无论是出于数据获取合规的考虑，还是出于数据应用的考虑，企业都正在加大对数据隐私保护的力度。根据国际调研机构Gartner最新的一份战略科技趋势报告预测，隐私增强计算成为2021年重点深挖的9项技术之一。并且Gartner认为到2025年，将有一半的大型企业机构使用该技术在不受信任的环境和多方数据分析中用来处理数据。

具体来说，在技术方面有以下几个基础核心作用：

- 使得符合严格的隐私保护法律法规成为可能；
- 使得大数据行业进一步发展成为可能，数据赋能企业发展到赋能行业社会；
- 促进数据要素化，数据资产化；
- 使得AI能够进一步发展；
- 使得区块链落地成为可能；
- 促进硬件和IT设施建设。

每个方面都会带动一个规模化市场的发展。[1]

### 1.2.3 市场层面

随着隐私保护等相关法律法规的颁布实施，以及个人保护意识的增强，相关企业将会部署增强的系统，并全面引入部署隐私保护技术，来达到合规目的。我们估算作为企业IT系统的核心功能，这部分功能业务改善投资将会占有不同类型IT系统投资的10%-40%左右，将会在下一个IT系统更新周期启动。据赛迪顾问的分析报告，2018年IT服务市场规模大约为6685.7亿元，2021年将会突破1万亿元。随着隐私机密计算的逐步应用，21世纪20年代中期，这部分对应的隐私机密计算市场每年在1000亿到4000亿之间。[2]

隐私机密计算技术使得大数据行业可以克服原有瓶颈，为行业带来新的发展，再加上数据资产化的趋势，整个大数据产业将会焕然一新。通过隐私机密计算技术的加持，充分实现数据价值挖掘，极大的增强社会生产力。以隐私机密计算为核心的大数据解决方案，将会逐步淘汰原有大数据应用方式，而对隐私机密计算的投资也将会持续加强。根据赛迪报告[3]大数据市场在2021年达到将近5000亿市场规模，隐私机密计算将会占据

[1] 从目前看市场最大的构成部分仍然来自于传统硬件部分：服务器和存储，网络设备，通常占比超过40%。其次为IT服务和商业服务，两者共占35%的比例，由25%的大数据相关软件所构成。未来服务器和储存设备都将会强化隐私机密计算的支持，例如具有可信计算环境功能的服务器未来将会占据将近100%的市场，没有此功能的服务器将会退出市场。此外将会出现和类似于于人工智能加速器一样的辅助隐私机密计算的专用硬件，相关软件将会增加或兼容隐私机密计算功能，相关服务或商业服务，将会以隐私机密计算为基础。

[2] 赛迪顾问《2018年IT服务市场数据》，对全球IT服务市场和我国的IT服务市场做了数据解析

[3] 赛迪顾问《2019-2021年 中国大数据市场预测与展望数据》

大数据行业的主要份额，而Gartner认为到2025年，将有一半的大型企业机构使用隐私机密计算在不受信任的环境和多方数据分析用例中处理数据。隐私机密计算技术实现数据资产化，保护数据所有者权益。其市场规模将会比原有估计大很多，预计市场总额每年约5000亿-8000亿。

目前AI技术发展迅猛，但是在落地时遇到很多困难和阻力，AI技术落地也存在合规和利益问题，同时也需要监管，而隐私机密计算可以在其中发挥核心基础作用，包括存量市场和增量市场。存量市场中的视频、音频、文本数据为基础的识别要求集成隐私保护功能，具体应用场景包括自动驾驶、医药研发、人脸识别、金融反欺诈等等。增量市场中，如IoT、交通、网络安全等，隐私机密计算也是新基建和核心底层。2023年，中国人工智能市场规模将达到979亿美元[4]，估计届时隐私机密计算的应用在其中约占20%-60%。21世纪20年代中期后，这部分中对应的隐私机密计算市场约为每年1300亿到3900亿之间[5]。

此外，隐私机密计算技术，使得区块链落地成为可能。单独的区块链虽然受到全世界的重视，但是在具体落地时候举步维艰，这其中一根本原因就是区块链的功能从技术角度局限在存证或溯源，但是完成具体业务不仅需要链上存储，还需要“计算”，链上计算，或链下计算。这个“计算”如果不是可信的，则链上存储很难体现其应有的价值。因此，区块链只有结合隐私机密计算才能落地，或者说隐私机密计算可以采用区块链作为辅助技术手段，更好完成具体业务。据IDC预测，2024年中国区块链市场整体支出规模将达到22.8亿美元，年复合增长率高达51%[6]。我们预计真正落地的区块链项目必须和隐私机密计算结合。这其中隐私机密计算约占30%-60%的投资总额，约40亿到80亿元的市场总额。

隐私机密计算是革命性的技术更新，将会带动新一代IT技术生命周期替换。并且会和公有云、私有云等模式紧密捆绑，整个体系会为隐私机密计算而优化。

[4] IDC FutureScape 全球人工智能市场2020预测 中国启示

[5] 赛迪顾问《2018年IT服务市场数据》，对全球IT服务市场和我国的IT服务市场做了数据解析

[6] 2020年8月更新的本年度第二期《IDC全球区块链支出指南》

对于行业来说有以下促进：

- 医疗健康领域：新药研发、临床辅助诊断、医保风控、科研、医疗AI等；
- 金融/营销：联合征信、精准营销、联合风控、客户画像等；
- 政务行业：数据能力开放、一网通管、联合安防、政企互联、智慧医疗、智慧城市、应急管理和应急响应等；
- 2C行业：精准获客、品牌营销、跨域营销、联名品牌策划等。

从行业上来看，2020年中国大数据技术与服务市场中的主要行业包括金融（银行、保险、证券与投资等）、政府（包括中央政府与地方政府）、营销和医疗，其总和占总体的50%以上。在金融行业中，大数据分析技术赋能于金融反欺诈、风控、信贷等业务。在政府行业中，智慧城市、公共安全、交通、气象等各部门对大数据应用较多。在营销行业中，通过个人位置数据，精准营销、信用评估等是大数据技术主要的应用方向。在医疗健康及制药领域，因医疗行业数据监管和合规要求严格，对隐私机密计算的需求是明确和广泛的，同时，在制药领域、真实世界研究、临床辅助诊断等场景都具有非常广泛的应用市场。

大数据行业正在面临越来越严格的法律法规限制，原有非隐私机密计算技术在应用过程中面临各种挑战，无法解决；这也正是隐私机密计算的用武之地。

### 1.2.4 法律层面

首先，随着信息技术的发展，人们逐渐意识到数据的重要性和数据挖掘所带来的价值，使得数据拥有方不再有意愿将数据进行无偿赠与或转让。而当数据作为一种市场资源进行交换时，它就具有了与传统财产所拥有的相同的特性，即财产权专属性、不可随意移转性、不可无利益的交易性等[6]。这也是数据的所有方倾向于将数据把控在手里，当作财产保护的重要原因之一。其次，数据作为一种载体，比其本身更重要的，应当是其中含有的信息和信息所转化成的价值，尤其是有关隐私安全的部分，例如上文所提到的基因数据所关联的个人及其血亲的健康状况等。因此，由于数据的性质已发生变化—

—已切实转变为一种可交易、可产生价值的市场资源及生产要素[6]，数据及其所含有的信息、信息安全等应当被纳入法律保护。

尽管近几年，我国有关信息安全保护方面的立法得到了进一步的加强和细化，但总体来说仍处于探索阶段。2009年，隐私权法律地位得到确立，被写入《侵权责任法》[7]中，这也是隐私权首次被纳入相关法律。同年，《刑法修正案（七）》[8]新增加了非法获取公民个人信息罪以及出售、非法提供公民个人信息罪，对相关的违法行为提供量刑参考。2012年，《关于加强网络信息保护的决定》[9]通过并颁布，其中对个人信息的范围和内涵提供了更清晰明确的界定。这也是我国第一次较为系统地在法律层面对个人隐私信息安全保护方面进行规定。2013年，国家工业和信息化部发布《电信和互联网用户个人信息保护规定》[10]，随后又对《消费者权益保护法》[11]进行了修订。这一阶段，相关法律着重强调个人信息权的基本法律地位，即个人信息权应当作为消费者保护的一项基本权利，同时进一步细化了个人信息权利保护方面的内容。2016年，全国人大及其常委会制定了《中华人民共和国网络安全法》[12]，它明确规定网络运营商有义务对用户的隐私安全提供保护，并应当建立健全的用户信息保护制度，这也是迄今为止关于公民个人信息保护最为全面的立法。《民法典》的颁布，进一步完善了个人信息保护的相关法律体系，特别是“隐私+一般信息”双重保护模式的确立，对于实现个人信息的有效保护，具有极为重要的意义。2020年10月《个人信息保护法（草案）》保护个人信息权益，规范个人信息处理活动，保障个人信息依法有序自由流动，促进个人信息合理使用。草案规定：**侵害个人信息权益的违法行为，情节严重的，没收违法所得，并处5000万元以下或者上一年度营业额5%以下罚款，5%的额度甚至超过了在个人信息保护方面规定“最严”的欧盟（GDPR）的4%。**另一方面本次立法，非常鼓励企业自身对系统进行评估，鼓励企业采用新技术应对法律监管所带来的自身业务发展的挑战。2021年6月《中国人民共和国数据安全法》通过并发布，将于2021年9月1日起施行；其中提出国家将对数据实行分级分类保护、开展数据活动须履行数据安全保护义务，承担社会责任等。

2017年，全国信息安全标准化技术委员会发布了《信息安全技术个人信息安全规范》[13]，这一份标准主要针对个人隐私信息所面临的安全问题，包括非法收集、滥用、泄露等，旨在遏制这些乱象。该标准规范了个人信息收集者和控制者在收集、保

存、使用、共享、转让、公开披露等信息处理环节中的相关行为。2019年发布的《信息安全技术个人信息去标识化指南》[14]则提出了个人信息去标识化的过程和管理措施，从技术角度提供了个人信息去标识化的指导，为相关组织、机构的个人信息去标识化工作提供指导，同时也为网络安全相关主管部门、第三方评估机构等组织开展个人信息安全监督管理、评估等工作提供依据。

相关法律的颁布和行业标准的设立和细化都表明，我国在隐私安全领域的发展正逐渐走向成熟。为了进一步引导和规范行业的健康发展，我们希望出台的法律规定中，增加技术性和可操作性的条款，可以给出详细明确的行为指导，这样司法机关可以清晰的界定责任边界，进一步落实法律法规的执行。

### 1.2.5 政策层面

隐私机密计算变得重要还是因为其底层构成——数据，已经成为一种生产要素开始促进社会生产的发展。2020年4月，《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》（简称《意见》）正式公布，将数据定性为土地、劳动力、资本、技术之外的第五大生产要素。这也标志着社会发展正式进入数据要素时代。2021年3月“中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要”（简称“十四五”规划）经第十三届全国人民代表大会第四次会议审查批准，正式发布。“十四五”规划对于大数据的发展作出了重要部署，其中多次提及数据安全，如何保障数据安全流通，隐私机密计算技术正是最优解决方案。2021年5月，国家发展改革委、中央网信办、工业和信息化部、国家能源局联合印发了《全国一体化大数据中心协同创新体系算力枢纽实施方案》，明确提出布局全国算力网络国家枢纽节点，启动实施“东数西算”工程，构建国家算力网络体系建设数据共享开放、政企数据融合应用等数据流通共性设施平台，试验多方安全计算、区块链、隐私计算、数据沙箱等技术模式，构建数据可信流通环境。

### 1.2.6 数据经济层面

全球经济数字化发展趋势愈加明显，传统产业加速向数字化、网络化、智能化转型升级，数字经济规模持续扩大。数字经济增加值规模由2018年的30.2万亿美元扩张至2019年的31.8 万亿美元，年规模增长了1.6万亿美元，数字经济已成为全球经济发展的新动能。2019年，德国、英国、美国数字经济占GDP均超60%，三国的占比分别为63.4%、62.3%和 61.0%，数字经济已占主导地位。2019年中国数字经济在国民经济中地位进一步提升，占GDP比重达到36.2%，对经济增长的贡献达到67.7%。Gartner还预测，到2025年，将有**一半的大型企业机构使用隐私增强计算在不受信任的环境和多方数据分析用例中处理数据**。

综上所述：数据时代，政策导向明确，企业间数据流转为大势所趋，个人隐私保护迫在眉睫，隐私机密计算将成数字经济时代的核心新基建。

## 1.3 隐私机密计算的技术需求

随着大数据产业的发展和相关法律法规的颁布落实，大数据产业对隐私机密计算的主要技术需求着重在以下几个方面：

### 1.3.1 隐私查询

匿踪私密查询(Private Information Retrieval，PIR)，简单来说就是查询方仅知道匹配的查询结果并且不留查询痕迹。通过采用隐私机密计算技术包括但不限于非对称加密、不经意传输等密码学技术，构建出多方查询时的数据交互加密通信通道，在整个查询交互过程中进行数据混淆、数据加密、数据传输、数据解密及匹配，从而让数据服务方无从知晓查询方的查询信息，查询方无从知晓数据服务方除查询信息外的其余信息，达到数据隐私保护、防止信息泄露、制止数据缓存的目的。

在某些商业应用过程中，当某查询方需要对一个或多个数据源通过某平台方进行查询或匹配并返回结果时，可在不泄露查询方查询条件给数据源方和平台方的情况下得到加密的查询或匹配结果。最大限度保护数据查询方的查询条件、数据源方的原始数据、平台方的匹配模型等各参与方的信息安全和商业机密。同时，可借助区块链技术，支持查询存证及授权，满足查询方信息不泄露的要求。

### 1.3.2 隐私建模/分析

随着大数据的进一步发展，重视数据隐私和安全已经成为了世界性的趋势，同时，大数据行业数据分散，呈现数据孤岛现象。如何在满足用户隐私保护、数据安全、商业机密保护和法律法规的要求前提下，进行跨组织、跨地域的数据合作，进而实现联合建模和联合分析，是困扰大数据和人工智能的一大难题。此外，在进行建模时，数据源的样本量和样本维度对模型的精度起到关键作用。尤其在医疗、金融领域，解决数据的多源性问题显得尤为迫切。隐私机密计算的优势在于多中心数据虚拟融合，使得数据“可用不可见”的情况下实现联合建模/分析，并且有效保护多源输入数据和计算过程的全流程隐私保护。隐私机密计算能够帮助合规安全利用数据，实现数据价值可信流动。

### 1.3.3 隐私推理

传统的模型推理是将模型和数据汇聚到一起，或者一个可信第三方，模型或者知识库在传输和执行过程中都存在安全隐患，可能会导致模型信息泄露或模型输入的个体信息泄露。其在使用中的安全保护、以及模型输入信息的隐私保护是当前大数据应用需要解决的重要问题。利用隐私机密计算实现隐私推理，实现商业模型应用的全流程隐私保护。

## 1.4 隐私机密计算的安全需求

隐私机密计算的一个重要特征是参与计算的是多方（这里需要特别指出数据提供方在有些隐私计算技术中并不是一定是隐私计算方），相对于其他类型的计算，计算方式

为单方，隐私机密计算的安全需求存在很大的不同。传统的计算，其数据计算能力均为逻辑上单点计算，其计算的正确性很大程度上依赖于下面几个方面：

1）计算的数据是否可信，是否是准确的数据。

2）算法实现的工程正确性和防止外部攻击。

而隐私机密计算则复杂的多，由于其计算是相对独立的多个方参与，其可靠性不仅仅包括参与计算的数据准确性和算法工程的正确性，还包括如何防止外部攻击，以及对于计算参与方的安全假设，比如：半诚实模型和恶意模型假设。半诚实模型（semi-honest model）：参与方在接触和处理其他参与方私有数据的时候会严格遵守协议规范，但会尽其所能地从经手的数据中挖掘出有效信息。这样的参与方也称为半诚实参与方；恶意攻击模型（malicious attack model）：参与方可能做出任何行为，譬如背离协议或与他人串通，尽其所能地获得关于隐私数据的有效信息。这样的参与方也称为不诚实参与方。由此可见，隐私机密计算于传统计算相比在安全层面有着更高的要求。

## 1.5 隐私机密计算的普及

隐私机密计算从开始提出到现在，发展了10多年，为了普及相关的概念技术，从2014年起首届全球iDASH隐私机密计算大赛举办，得到全世界主要科技公司关注。至今，比赛已经举办了8年，备受赞誉和认可，世界权威学术机构和相关媒体，都支持和报道该项赛事。

iDASH（全称是：integrating data for analysis, anonymization, and sharing），是美国八大生物医学计算中心之一，也是其中唯一负责生物医学数据分享过程中隐私安全与保护的中心。该中心创建于2010年底，位于University of California San Diego，该中心主要负责开发新的算法、工具、计算基础设施和服务来帮助生物医学数据进行安全高效的分析和共享。王爽教授从2011年开始参与iDASH中心生物数据隐私安全相关技术的研发，于2014年，作为创始人和PI之一创办了“iDASH全球隐私安全保护竞赛”。2014年至今，全球对于隐私安全的保护意识都在提升，比赛也得到各国关注。2020年，包括

我国的阿里巴巴，腾讯、百度、浙江大学等全世界有20多个国家，将近200家队伍参加了比赛。

2020年11月，中国人民银行正式发布《多方安全计算金融应用技术规范》（JR/T 0196—2020，以下简称《规范》）金融行业标准。《规范》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。《规范》的发布促进实现“数据可用不可见”、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。

## 1.6当前市场状况

近年来，国内隐私机密计算行业发展十分迅速，2018年国内对于隐私机密计算的认知还非常有限，只有若干学术机构和厂商进行研究，加起来不过3~4家。由于隐私机密计算的巨大发展潜力，同时相关领域比赛的推动，近两年来，无论是头部互联网企业，或是初创型科技企业，都在纷纷入局隐私机密计算。

隐私机密计算领域各参与企业的资源生态、技术路线和行业布局均有不同。互联网大厂主要优势是丰富的数据生态和应用组件；产业背景公司的主要优势是垂直行业的专注积淀和应用能力。但这些公司都是这一两年才介入隐私机密计算领域，很多是利用公开的科研文献或开源代码进行修改，其实用性、可靠性有待验证，而且对隐私机密计算理解有待加强，对业务的理解也需要纠偏。此外，还有存在对客户宣传偏颇的问题。因此有必要对相关的技术进行梳理，以方便用户进行选择。

# 2. 隐私机密计算的基础技术

隐私机密计算，是一个较新的概念，类似的概念有隐秘计算，隐私安全计算、隐私保护计算、安全计算、可信计算、隐匿计算、机密计算等等，这些相关技术概念着重于隐私保护算法或数据保护算法。而隐私机密计算更强调是系统构架，通过创造性的采用

相关技术，不仅能保护隐私还能够保护商业秘密，以及在保护基础上进行海量数据处理、执行复杂算法的系统能力。

现有的用于保护数据隐私的解决方案主要基于以下基础技术：同态加密（HE），多方安全计算（MPC），可信执行环境（TEE），联邦学习、差分隐私（DP）、区块链、零知识证明（ZKP），以及脱敏、假名等相关的传统技术。

## 2.1 同态加密

**同态加密** (Homomorphic Encryption) 的思想首先由Rivest、Adleman 和 Dertouzos于1978年提出[15]，其理论上的可行性在2009年被证明[16]。其特点在于计算过程基于加密数据，并且解密后的计算结果和传统计算结果一致。这种特性使得其特别适用于云计算或者外包计算应用场景中的隐私保护（secure data outsourcing），使用者可以将自己的敏感数据加密后传输至不受信任的云计算服务中心进行加密计算生成加密的计算结果，然后将计算结果下载到本地通过私钥解密后使用。如下图2.1所示，根据其技术特点的灵活性，同态加密可以分为：

(1) 部分同态加密 (partial homomorphic encryption)：仅支持加法或乘法运算

(2) 半同态加密(somewhat homomorphic encryption)：支持有限次累计乘法计算［17］

(3) 全同态加密 (fully homomorphic encryption)：支持任意次数的加法和乘法运算操作；同时同态加密的复杂度（对于计算、存储等开销）也随着显著增加

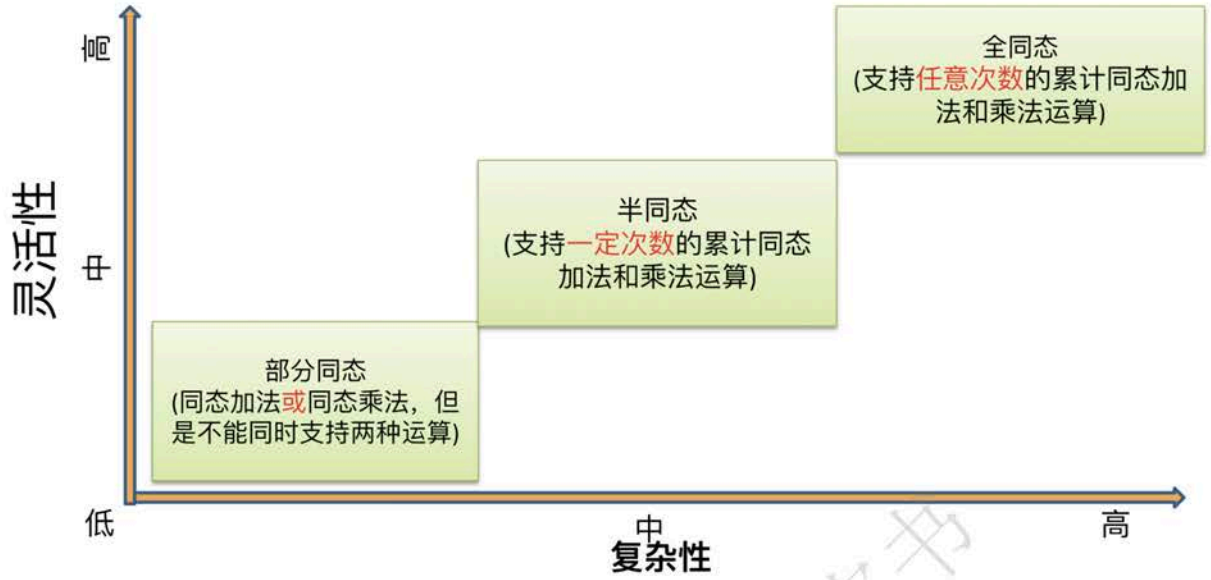


**图2.1:不同同态加密技术的对比**

从技术理论上讲，同态加密可以支持任何可以用低阶多项式表示的函数，包括加、减和乘。例如，对加密矢量或矩阵的许多算术运算［18］可以支持。借助同态加密，用户数据受到强大的加密保护，可以将其外包并安全地存储在不受信任的存储节点上并支持对密文执行计算（例如，公共云）。对于一个给定的同态加密公钥pk和私钥*sk*，以及需要计算的数据 $m_1$ 和 $m_2$ ，同态加密的计算特性可以表示如下：

$$\mathrm{Dec}_{sk}\left(\mathrm{Enc}_{pk}\left(m_1\right)\oplus\mathrm{Enc}_{pk}\left(m_2\right)\right)=m_1+m_2$$

$$\mathrm{Dec}_{sk}\left(\mathrm{Enc}_{pk}\left(m_1\right)\otimes\mathrm{Enc}_{pk}\left(m_2\right)\right)=m_1\times m_2$$

其中⊕和⊗是对密文进行加法和乘法的算术运算符。使用这两个基本运算符，我们可以构造更丰富的可以用低阶多项式表示的函数。

同态加密改变了如何在不受信任的云环境中进行私有数据的计算。 图2.2 说明了目前的云计算架构，我们主要使用云作为存储。在云中进行的计算需要解密，如果存在安全漏洞或内部攻击，则解密可能会泄漏敏感信息［19］-［21］。

相反， 图2.3 说明了同态处理直接作用于加密数据而没有解密的情况。在受信任的用户端之外，敏感数据在黑盒中计算（由于加密）。入侵云中的对手只能看到密文，而密文不会显示有关原始明文的信息。

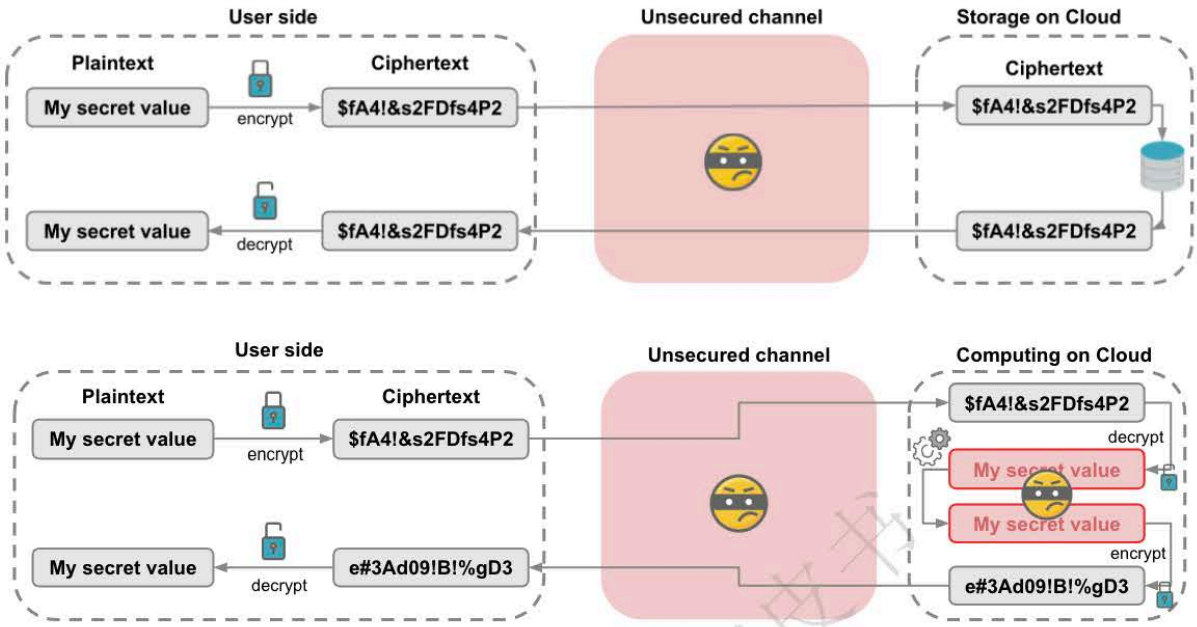


**图2.2： 云计算用于数据存储（上面）和计算（下面）。**

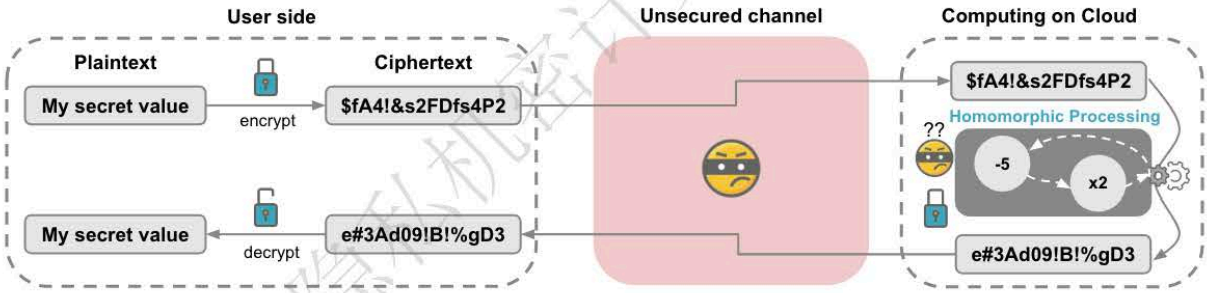


**图2.3： 具有同态加密的云计算。**

从技术上讲，基于*Learn with error*（LWE）问题构建的同态加密的解决方案[22]-[24]可以提供强大的安全保证，例如可以抵抗*已知的*量子计算算法。强大的安全保证对于高度敏感的数据非常重要。因为诸如基因组数据之类的敏感数据需要在个体的整个生命周期甚至更长的时间内得到保护。此外，基因组数据在同一个家族的血亲之间是相关的[25]，[26]。例如在医学领域，同态加密被用于安全外包计算和存储加密数据查询[27]-[29]，遗传风险计算[30]，[31]，GWAS中的统计分析[28]，[32]-[35]和编辑距离计算[36]，[37]。同态加密技术目前主要的局限性在于计算的效率相对比较低、加密后的数据很难适配通用计算，只能服务制定的计算场景。同态加密技术通常适用的

安全模型是半诚实安全模型，适用的计算场景为安全外包计算，即私钥拥有方通常为数据拥有方的情况，其中不受信任的计算提供方通过公钥和同态加密的密文完成计算。

## 2.2 多方安全计算

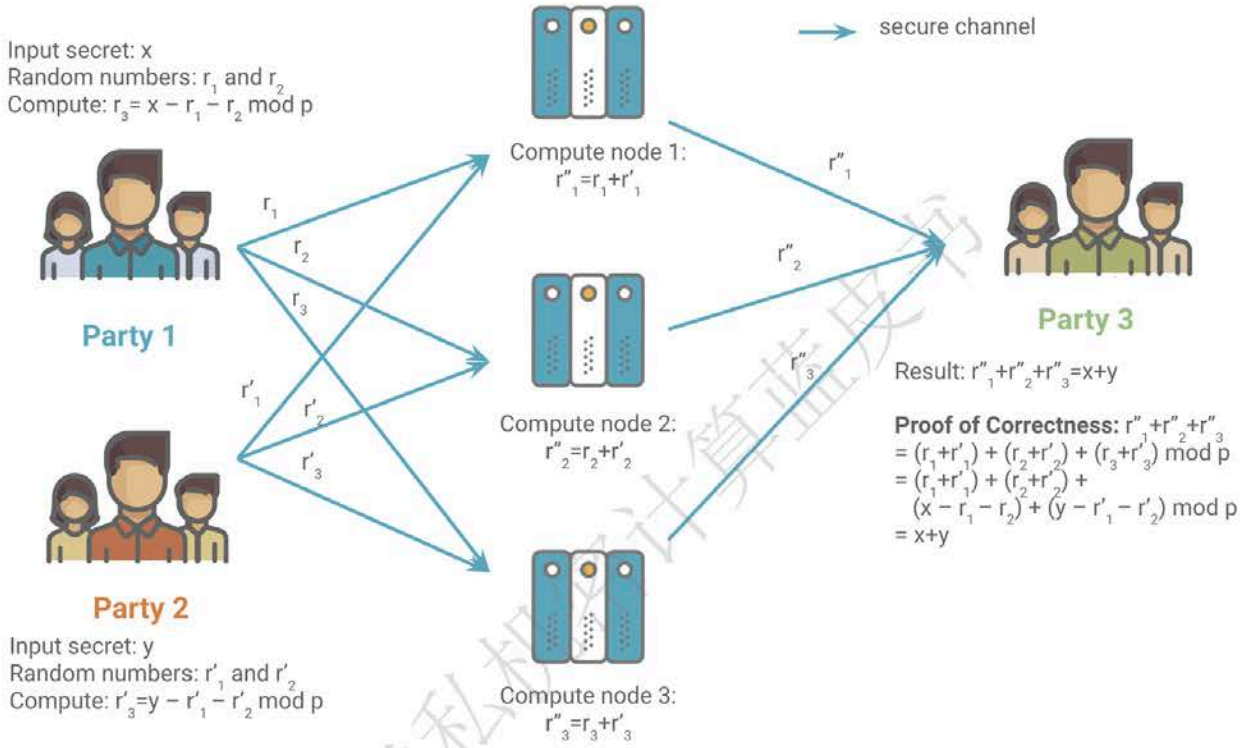


**图2.4：根据SS协议的份额计算MPC协议的示例**

***多方安全计算***（MPC）是一种交互式协议，用于计算多方之间的某些功能。根据参与方个数不同，可分为两方计算和多方计算。两方计算主要是基于乱码电路（Garbled Circuit，简称GC），三方的是基于机密分享（Secret Sharing，简称SS），其底层是通过不经意传输（oblivious transfer，简称OT）机制来支持。

基于混淆电路的协议更适用于两方逻辑运算，通讯轮数固定。另一类基于秘密分享的多方安全计算中，数据输入和计算中间值都会以「密文分片」的方式存在。秘密分享技术可以把隐私数据切割为 2 份或更多份后，将随机分片分发给计算参与方；之后，就可以利用分片间存在的相关性来实现在分片上计算并重建得到隐私数据计算结果，这个过程保护了数据隐私又允许多方联合对数据进行计算。

如图2.4说明了MPC如何通过SS协议计算两个秘密的份额之和。假设双方1和2有秘密值$x$，$y$和第3方希望得到的结果 $x + y$ ，但不知道各方的值。

只要没有窜谋，MPC协议的优点从理论上讲是安全可信的，而不是依赖于计算假设。数据是秘密共享的，而不是加密的。每个服务器上的共享都不会在没有所有服务器相互勾结的情况下显示有关机密的信息。通过结合使用SS和MPC协议，我们可以在非窜谋服务器之间执行任何算术功能，同时隐藏用户的私有数据。如以上示例中所示，计算加法非常有效，但是乘法的计算会更为复杂，需要计算服务器之间的一轮额外的通信来交换计算中使用的元素。在生物医学中，已经对MPC进行了研究，以支持安全序列相似性比较［38］-［43］，多中心逻辑回归［44］-［47］等。例如，最近的《科学》论文［48］和《自然生物技术》论文［49］证明了使用MPC分别进行小规模基于基因组的诊断和支持大规模百万个人关联研究而不暴露患者的可行性。MPC在一些应用场景已经应用，但是也存在一些局限性，例如，MPC没有办法有效支持超过三个以上的安全计算节点；计算节点间不能窜谋，否则分片的密文有隐私泄露风险；在MPC通常的安全假设下，节点计算是必须诚实的按计算流程操作，任何违背计算流程的活动不能被MPC所保护；MPC通讯量比较大，计算前需要动态的把数据进行加密分割并在计算参与方之间分享，并且分享的数据出于安全性考虑很难复用，存在巨大的通信带宽需求。

## 2.3 可信计算环境

**可信计算环境**（Trusted Execution Environment）是指计算机处理器的一块隔离的安全区域，它可以确保在其内部加载的代码和数据在机密性和完整性方面得到保护[50]，有效的防止了底层操作系统或虚拟平台被挟持后对数据和代码的攻击。同时，可信计算环境还提供了对于不授信第三方安全计算环境的远程验证功能。其实现主要包括ARM TrustZone、Intel SGX软件防护扩展、AMD SEV安全加密虚拟机、基于RISC-V的开源框架Keystone、Intel TDX可信域扩展以及ARM CCA机密计算架构等。

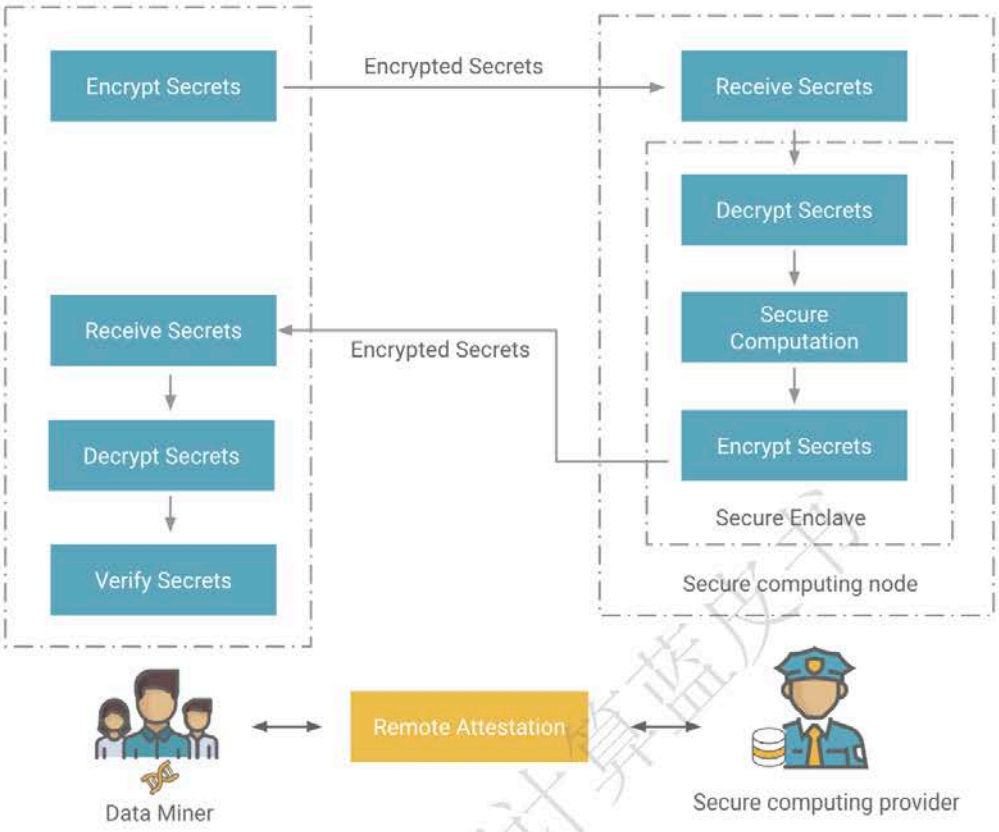


**图2.5：典型的SGX框架的概述，该框架由数据所有者，不受信任的云服务提供商（CSP）和安全区域组成。**

可信执行环境（TEE）在硬件中提供了隔离的内存和计算空间（enclave），可以在其中高效而安全地分析敏感数据。如图2.5，SGX是Intel主流CPU（例如Skylake或更高型号）的功能，可在现成的台式机、工作站和服务器上商业购买。SGX使开发人员能够创建安全的隔离区域以便在隐私保护下计算敏感的数据。安全区内存使用强大的加密算法进行加密，并且仅在将数据加载到CPU高速缓存和寄存器中时才解密。初始化安全区时，安全区的初始代码和数据将保存在加密的内存中，称为 *Enclave Page Cache*或EPC。然后，远程方可以通过称为远程验证的过程比较enclave的测量来验证enclave的身份。使用远程验证，敏感数据的所有者将获得对在安全区域内运行的代码的信任，因此可以使远程服务器（例如，云）的安全区域处理私密数据，从而确保目标服务器执行预期的计算。此CPU功能已用于保护敏感数据操作：例如，Microsoft Azure最近在其公共云中引入了一个具有SGX功能的服务器的机密计算平台［51］以保护从区块链财务操作到SQL Server中存储的数据的所有内容和微软的基础架构。其他主要的云提供商，例如

Google云平台、阿里云，也提供了具有SGX功能的系统。该技术已广泛应用于数字版权管理、付款保护、密码和秘密密钥管理、安全文档共享等，甚至被考虑用于保护电子病历。目前使用TEE的局限在于需要特定硬件的支持，但是随着支持TEE功能的计算节点的普及，这个局限性正在逐渐弱化。另外，一些相关的研究表明，TEE在某些计算任务中可能会受到侧信道攻击的问题[52]，但是通过一些技术手段可以减轻或者避免相关问题[53]，[54]。

## 2.4 联邦学习

**联邦学习**（Federated Learning）是一种在计算过程中分享中间统计结果而不泄露原始数据的分布式算法和框架，实现了数据在多中心协同计算研究中的隐私保护[55]，[56]。其特点是在保证了计算结果准确性和精度的前提下，实现了各个数据中心原始数据的隐私保护。因此，联邦学习被广泛应用于多中心大数据的协同研究分析和隐私保护中，同时适用于各种不同数据格式（结构化电子病历，非结构化数据，基因数据以及医学图像数据等）。

联邦学习一般认为有两种架构：（1）客户端/服务器模式（图2.6.a）；和（2）去中心化模式(图2.6.b)。客户端/服务器模式一般适用于预测全局模型参数和开展各种统计学检验。基于不同的联邦算法设计，各个客户端在本地进行基于原始数据的隔离计算，然后将本地计算的模型参数（中间结果，不涉及原始隐私数据）发送到服务器进行汇总计算。服务器根据各个客户端的本地统计结果更新全局模型参数，如有需要，服务器发送更新的全局模型参数到各个客户端进行多次迭代计算进而得到具有可靠精度的最终计算结果。去中心化模式使用于各种分布式算法，比如稀疏线性回归，主成分分析以及向量支持机等等[57]-[59]。其特点在于不需要中心服务器，各个相邻的客户端不断交换本地计算的中间结果，进而得到精度可靠的全局计算结果。无论哪种架构，联邦学习实体之间只传输中间结果，中间结果不涉及任何原始数据信息，从而实现了敏感数据的隐私保护。

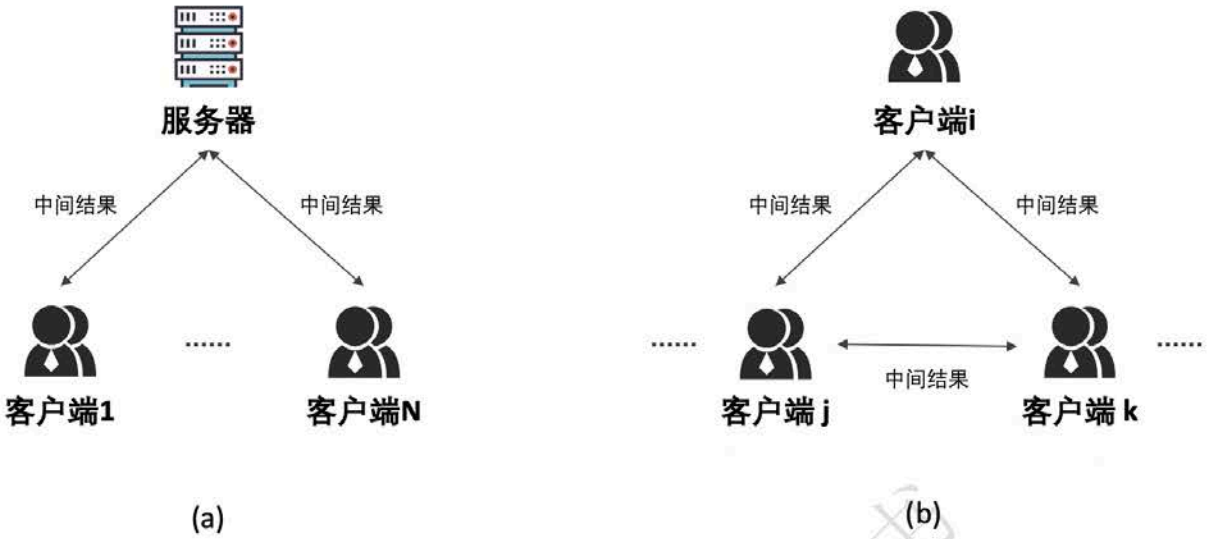


图2.6. 联邦学习的两种架构模式：(a) 客户端/服务器模式；(b) 去中心化模式

联邦学习模式下，各个数据中心的数据可以具有两种格式：（1）水平分割数据格式（各个数据中心持有相同的特征，不同的样本）；和（2）垂直分割数据格式（各个数据中心持有相同的样本，不同的特征）。前者适用于各个数据中心持有不同的群体，每个群体具有相同数据特征的应用场景。例如针对罕见病的跨中心数据合作研究[60]，其优点在于增大了数据量，使得研究计算结果更具普遍性。后者适用于各个数据中心具有相同的群体，各数据中心的群体具有不同数据特征的应用场景，其优势在于丰富了样本的数据特征，使得研究的群体画像结果更加全面准确。

需要注意的是，近期的研究表明，联邦学习在某些应用场景下可能造成部分隐私泄露，比如在某些图片处理的联邦学习框架中，中间结果包含了图片数据的梯度信息，通过不断监控和分析计算过程中的梯度信息，研究者发现可以部分恢复一些原始图片[61]。因此，联邦学习的应用(例如肿瘤图片的联合分析研究)应该根据实际情况结合不同的安全技术手段实现更严格的隐私保护，比如中间结果的交换使用传输层加密协议（SSL/TLS），对中间结果进行加噪声处理，以及在中心服务器部署可信计算环境等等。

## 2.5 安全联邦学习

联邦学习是一个开创性的发明，但是仍然存在不完美的方面，针对普通联邦学习的不足，业界对其提出了不同改进的方案，2012年王爽教授提出的联邦学习的文献中已经

针对联邦学习的缺陷提出了增强方案，2014年王爽教授针在普通联邦学习技术基础上结合多种相关密码学技术，做出大量的系统性的改进，形成了安全联邦学习，这是目前为止最优秀的隐私机密计算技术之一，其克服原有的技术方案的弊端和风险，既能够消除信息泄露的问题，还同时具有较高的执行效率和处理能力。并且已经在一些系统中得到应用和验证。

**安全联邦学习**技术主要解决隐私机密数据多数据源的联合分析需求，是综合相关技术的最优解决方案。其中，TEE有相对较高的执行效率、较高的处理能力；多方安全计算能够比较好的处理两方或三方数据问题；同态加密能够有较高的抗量子攻击能力和简单的数据处理能力；密码学技术适用于协助进行数据管理、身份认证等内容；区块链用于协助实现业务流程管理、审计管理、计费等功能。通过综合利用上述技术组合，经过算法优化，能够处理海量数据，满足特定业务场景的需求。通过使用安全联邦学习技术，使得隐私机密计算能够作为大数据平台的底层核心基础设施（原生内核），打破数据孤岛，可以实现多行业、多部门、多中心的数据联合计算。可实现在符合我国的网络安全法、GDPR、HIPAA等严格隐私保护法律法规情况下的多中心多维度实时大数据分析计算。传统联邦学习和安全联邦学习的比较如下表：

| 技术 | 成熟度 | 性能 | 算法能力 | 安全性 | 依赖可信方 | 概要 |
|---|---|---|---|---|---|---|
| 普通联邦学习(FL) | 高 | 高 | 中 | 中 | 部分需要 | 普通联邦学习，存在风险需要结合其他技术改进，才能合规。 |
| 安全联邦学习(SFL) | 高 | 高 | 高 | 高 | 可不需要 | 综合运用TEE、密码学等隐私机密计算方法，适用于各种业务场景。容易合规。 |

## 2.6 区块链技术

**区块链**技术以其去中心化、防篡改、可追溯、高度可扩展等特点，正与大数据、云计算、AI、5G等技术快速融合，并应用到医疗、政务、金融等重要领域。区块链的分布式存储、不可篡改、加密性、可追溯性等特性，可有效保持信息的透明性，能够保证在

不同机构共享信息。但是，严格意义上来说，区块链并不属于隐私机密计算范围 ，其主要用于数据记录，和传统数据库所起到的功能等价。它可以和隐私机密计算技术结合应用，用于实现审计、存证、追溯等功能，和隐私机密计算互补，更好实现业务场景落地。

## 2.7 隐私保护相关传统技术

### 2.7.1 脱敏

**数据脱敏**是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。在涉及客户安全数据或者一些商业性敏感数据的情况下，在不违反系统规则条件下，对真实数据进行改造并提供测试使用，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。

数据脱敏从技术上可以分为静态数据脱敏和动态数据脱敏两种。静态数据脱敏一般应用于数据外发场景，例如需要将生产数据导出发送给开发人员、测试人员、分析人员等；动态脱敏一般应用于直接连接生产数据的场景，例如运维人员在运维的工作中直接连接生产数据库进行运维，客服人员通过应用直接调取生产中的个人信息等。换句话说，静态脱敏可以理解为“原始数据的仿真替换”，而动态脱敏则更强调“在使用的同时脱敏”。静态脱敏可将脱敏设备部署于生产环境与测试、开发、共享环境之间，通过脱敏服务器实现静态数据抽取、脱敏、装载。动态脱敏采用代理部署方式：物理旁路，逻辑串联。应用或者运维人员对数据库的访问必须都经过动态脱敏设备才能根据系统的规则对数据访问结果进行脱敏。

无论是静态脱敏还是动态脱敏其最终都是为了防止组织内部对隐私数据的滥用，防止隐私数据在未经脱敏的情况下从组织流出。满足组织既要保护隐私数据，同时又保持监管合规，满足合规性。

### 2.7.2 假名

**假名化**是指通过生成新的字符来替代原标识符（通常为直接标识符）的数据处理方式。假名化的概念在GDPR、CCPA（California Consumer Privacy Act）和各种标准文件中出现过。假名化技术是指用生成的新字符，即假名（pseudonym），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。WP29工作小组《Opinion 05/2014》中列举了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以应对上文所述的彩虹表破解的常用技术手段，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。

通过加密标识符生成的假名的方式，用于还原标识符的信息为密钥；通过哈希函数和令牌化技术生成假名的情况下，通常会额外生成一张假名与原始标识的映射表单用来还原标识符。根据GDPR和CCPA等法律的要求，这些密钥或者映射表单等可用来还原标识符的“额外信息”需要与假名化后的个人信息分开存储以保证个人信息的安全。假名化可以在一定程度上可以减轻数据主体的风险并帮助数据控制者满足数据保护的义务，使数据交换或共享合规化。

### 2.7.3 传统技术的限制

传统技术脱敏和假名由于其容易理解，技术成熟，在很多项目中得到广泛的应用。和通常的认知不同，科学研究已经发现[62]，[63]，脱敏作为一种保护隐私的技术，其

保护能力并不完善，仍然存在系统性的漏洞使得其保护隐私的能力大打折扣。并且，由于其对于原始数据的价值的覆盖，在很多情况下脱敏的数据不能满足很多应用场景的需求。例如，基因组数据即使脱敏，因为它们包含独特的、可遗传的遗传标记，这些信息可用于远程家庭搜索，通过将脱敏受试者与身份已知的远亲联系起来，可以识别脱敏受试者。因此，基因组数据脱敏本身不足以保护隐私。

另一个角度上看，也应该认识到，脱敏和假名只是对数据静态的敏感信息处理，并不是计算过程中的保护处理。因此，严格意义上说并不属于隐私机密计算技术的范畴，但是由于在行业内的广泛使用和认知，在这里一并介绍。

## 2.8 总结

隐私机密计算是一个系统工程技术，需要以合理的技术构架驱动，才能满足业务需求。

# 3.隐私机密计算的整体框架

上一章我们介绍隐私机密计算中的一些核心技术，这些技术有些已经熟知，有些则还是处于学术探讨阶段，有些则已经在不知不觉中被广泛使用。但是单独使用其中任何一项技术都不能完整实现隐私保护的要求。这里需要一个科学合理的整体框架来实现新的计算范式，满足隐私机密计算的标准。

在具体介绍整体框架前，我们首先要考虑一个可靠的、完整的、可信的、可用的大数据隐私机密计算技术解决方案，需要满足以下维度的能力要求：合规能力、隐私保护能力、计算正确性能力、安全性证明能力、计算效率、大数据处理能力等。其次，我们还要对隐私机密计算涉及的这些技术有比较准确清晰的认识，不同技术的比较如下表所示：

| | | 秘密分享(SS) | 混淆电路(GC) | 同态加密(HE) | 联邦学习(FL) | 可信执行环境(TEE) | 安全联邦学习（SFL） |
|---|---|---|---|---|---|---|---|
| 安全模型 | 计算参与方数量 | 3 | 2 | ≥2 | 理论上无限制 | 理论上无限制 | 理论上无限制 |
| | 常用安全模型 | 半诚实模型 | 半诚实模型 | 半诚实模型 | 可信参与方 | 恶意模型 | 恶意模型 |
| | 安全假设 | 参与方不能串谋，否者泄露加密原始数据 | 参与方不能串谋，否者泄露加密原始数据 | 如果超过两方，需要可信第三方管理私钥 | 计算算法不会输出原始数据 | 每个参与方可以拥有单独的动态密钥 | 计算算法不会输出原始数据，每个参与方可以拥有单独的动态密钥 |
| | 数据保护 | 原始数据不出边界，密文碎片出边界 | 原始数据不出边界，密文碎片出边界 | 原始数据不出边界，原数据加密后出边界 | 原始数据不出边界，交换统计信息 | 数据加密后出边界 | 原始数据不出边界，交换加密统计信息 |
| 计算能力 | 算力要求 | 高 | 较高 | 非常高 | 低 | 低 | 低 |
| | 通信要求 | 非常高 | 非常高 | 较低 | 低 | 低 | 低 |
| | 基础计算能力 | 电路可以表示的算法 | 电路可以表示的算法 | 加法乘法可以表示的算法 | 大多数算法 | 几乎任何算法 | 大多数算法 |
| | 实现方式 | 软件 | 软件 | 软件 | 软件 | 硬件 | 软硬件结合 |
| | 系统扩展性 | 低 | 低 | 低 | 高 | 高 | 高 |
| 当前主要缺陷 | | 计算和通讯量大，通常支持三个计算参与方 | 计算和通讯量大，通常支持两个计算参与方 | 不能复杂计算，加密后数据不支持通用计算 | 中间过程交换的参数可能泄露数据隐私信息 | 处理能力有待提升 | 综合技术解决方案，技术复杂度高 |
| 发展前景 | | 数学原理限制，依赖于通用计算能力和通讯能力 | 数学原理限制，依赖于通用计算能力和通讯能力 | 数学原理限制，依赖于通用硬件计算能力 | 改进后的联邦学习，是非常有前景的发展方向 | 发展迅速 | 解决交换参数泄露隐私弊端，安全性高，可适用于各种业务场景。 |

从上表可知，这些基础技术都不是完美的技术，单独看都存在着这样那样的不足，不能很好地实际应用。隐私机密计算，需要一个综合各种技术的解决方案才能具体落地，满足业务需求。具体来说，选择一种技术作为核心框架，选择其他技术作为支撑互补，以满足完整的业务要求。

# 4.隐私机密计算应用场景

## 4.1 医疗

### 4.1.1 基因分析

全基因组关联研究（GWAS）是从人类全基因组范围内找出存在的序列变异，即单核苷酸多态性（SNP）并筛选出与疾病相关的SNPs，以帮助进行疾病诊断或是预防的研究方法。它常用于复杂疾病研究，包括脊柱炎、肿瘤、糖尿病和高血压等。利用全基因组关联分析对遗传机制的研究有助于开发新药物、发展新疗法和开展预防工作，提高整体国民健康水平。然而，全基因组关联分析的研究过程中存在着诸多挑战。

1）“数据孤岛”问题：GWAS依赖大量基因数据积累，样本量不足是各项GWAS研究中最常见的问题和难点; 基因数据大多独立存在于各家医院、科研机构或是基因库内，缺乏关联和交互方式，形成了“数据孤岛”。

2）隐私安全问题：基因数据具有个人识别性，一旦泄露将造成难以预计的损失，且伤害会蔓延至信息遭泄露个体的血亲，因为他们拥有相似的基因片段。

### 4.1.2 医疗数据匿踪查询系统

匿踪私密查询(Private Information Retrieval，PIR)简单来说就是查询方仅知道匹配的查询结果并且不留查询痕迹。通过隐私机密计算可以打造医疗数据匿踪查询系统，主要采用RSA非对称加密、不经意传输等密码学技术，构建出多方查询时的数据交互加密通信通道，在整个查询交互过程中进行数据混淆、数据加密、数据传输、数据解密及匹配，从而让数据

服务方无从知晓查询方的查询信息，查询方无从知晓数据服务方除查询信外的其余信息，达到数据隐私保护、防止信息泄露、制止数据缓存的目的。

### 4.1.3 临床数据分析及新药辅助开发

药物研发是AI技术应用的重要场景之一。药物研发要经历靶点的发现与验证、先导化合物的发现与优化、候选化合物的挑选及开发和临床研究等多个阶段。传统的药物研发耗时耗力，且成功率低。AI助力药物研发，可大大缩短药物研发时间、提高研发效率并控制研发成本。借助AI并利用大数据分析辅助新药研发，进行药物疗效研究、药物市场分析、药物副作用等的分析存在尚待解决的问题：1）缺乏社区医疗数据。2）缺乏病人的随访数据。3）缺乏病人的消费习惯和使用偏好数据。4）数据源覆盖面有限。5）数据维度不够。

利用隐私机密计算技术，打破数据孤岛，联合多中心多维度数据源，高效利用数据的同时保证患者个人隐私数据安全。

### 4.1.4 医学影像分析

健康医疗大数据时代，大量医疗数据被源源不断采集，并被使用到生物医学研究中。其中医学影像学数据是一个非常重要的组成部分。在医学影像实际问题中，人工智能模型精度和效果往往是由训练样本的数据量及其质量决定的。但是由于数据孤岛问题、传统数据脱敏的局限性带来的隐私问题、数据监管问题等，无法实现数据安全有效被利用。大数据分享和分析带来了信息隐私和模型保护两方面的挑战。

隐私机密计算可在医疗影像学中的应用覆盖智能辅助诊断疾病、智能勾画靶区、智能判断病理切片、影像设备的图像重建，以及其他智能辅助诊断方案。具体包括：

1）病灶识别：基于AI技术帮助医务工作者找出病灶区域，避免人工操作带来的失误，并且可以节约人力成本，为医生的诊断提供快速、可靠和精准的辅助诊断参考。

2）病灶分析：基于AI技术根据医学图像对病人病情作出初步判断和分析，为医生最后决策提供参考。

例如：1）诊断糖尿病性视网膜疾病——根据视网膜底的图片来识别糖尿病性视网膜病

变程度，提高筛查效率。 2）辅助诊断肺炎、结核病——根据病人胸部影像来快速准确的筛查出肺炎、结核病，医生可以在此基础上做进一步诊断。

## 4.2 金融

### 4.2.1 金融征信

传统征信服务的数据来源主要分为两部分：一是对接银行、政府部门的自建数据库，主要提供商业、信贷等金融数据；二是依靠申请者自行提交材料，但由于审核渠道和能力有限，这部分数据往往水份很大，真假难辨，很难成为有效、可靠的证据或依据。

国内目前已形成从央行到各部委以及区域政府的各级信用数据库。近几年来，不少民营大数据征信公司也逐渐形成了自己的数据库。然而，由于采集和维护数据的成本极高，这些信息源和数据库往往被当作竞争优势。各系统和公司对于自己的数据库“严防死守”，各数据库之间往往相互隔离，数据信息因此难以流转。这样一来，不仅造成了数据资源的浪费，也导致跨领域或跨地域的失信行为很难被检测出来。

隐私机密计算则能很好地解决上述问题，作为大数据平台，隐私机密计算平台可以实现多行业、多部门、多中心的数据联合计算。结合区块链溯源技术，在接入不同的数据源之后能自动采集和读取相关的信息。例如，当需要查询某家企业的征信情况时，可以自动根据需求找到相应的原始数据并进行计算。一是免去了该家企业申请中需要准备各种材料的麻烦，二是由于信息获取方式从被动到主动，大幅降低了瞒报漏报的可能性。因此能改善传统征信部分数据依赖申请者自行申报导致数据真假难辨或是错报漏报的问题，保证计算结果的完整性和真实性，从而构建多维度更全面的征信体系。

### 4.2.2 金融风控

金融机构在保证其资金正运转时，会利用大数据构建风控模型，通过技术手段消灭或减少风险事件发生的各种可能性。银行信贷风控与营销过程中往往需要政府、企业、个人数据提供支撑。特别是针对缺乏抵押资产的中小微企业，基于数据的风控营销机制变得更为重要。然而传统模式下，由于数据安全问题，银行（特别是区域银行）难以高效且合规地获得企业与个人数据。隐私机密计算技术可以在不泄露各方原始数据的前提下进行分布式模型推断或

者训练，政府部门可以将政务数据共享给金融机构，通过隐私机密计算进行合规数据分析，实现精准风控。

### 4.2.2 交易策略隐私保护

金融行业的券商、基金公司等客户，在期权投资管理过程中积累了相关的投资策略经验，并希望可以通过期权交易策略可以帮助用户获得更好的投资收益。但是期权交易策略作为策略提供者的资产在对外赋能的时候也希望其策略能得到保护，并能获得利益。同时，策略提供者也希望其所提供的策略能够不断从市场使用中获得反馈，对策略进行学习优化以提高其性能。传统模式下，策略在被使用过程中缺乏充分保护，存在被泄露的风险。同时基于策略的交易数据也包含敏感的交易信息。

通过隐私机密计算技术可以打造的交易策略隐私保护及学习系统，可以为策略应用和反馈过程中的相关隐私信息提供保护，使得交易策略能够被应用的同时其机密信息也得到保护，并为策略提供者提供反馈机制，从而可辅助动态交易策略学习，优化策略。

## 4.3 政务

### 4.3.1 医疗核保

2020年7月份国办印发《关于推进医疗保障基金监管制度体系改革的指导意见》，提出加快推进医保基金监管制度体系改革，构建全领域、全流程的基金安全防控机制，严厉打击欺诈骗保行为。随着大数据、AI技术的发展，利用大数据甄别合理、必要的医疗行为，“识别”出欺骗行为，减少骗保现象，看牢“救命钱”，实现医保智能监控。

而在实施过程中，数据的日趋碎片化，形成了数据孤岛，传统的大数据AI分析难以适应行业的需要。隐私机密计算技术可构建医保智能风控网络，很好地解决了数据隐私保护的问题，有效地聚合了多渠道、多平台的数据源；包括政府自有居民数据、医院就诊数据、居民购药、资金交易等数据，满足医保核保要求。

### 4.3.2 医保控费

2019年全国基本医保总支出为19945.73亿元，比上年增长11.9%。2020年10月19

日，国家医疗保障局官方发布了《国家医疗保障局办公室关于印发区域点数法总额预算和按病种分值付费试点工作方案的通知》，提出将统筹地区医保总额预算与点数法相结合（DRG），实现有限医保基金的优化利用，并走向医院、患者、医保的三方共赢。

具体而言，DRG旨在通过对疾病进行科学分组，再利用大数据分析，在医保基金的可用财力内，测算并制定每个病种组别相对应的医保付费标准，再由医保基金向提供服务的医疗机构进行支付，以最大化发挥医保的保障效应。

DRG模型的精准度依赖于多元大数据的支持（医保、医院、政务、金融等），传统的大数据AI分析难以有效合规的利用这些多源信息。隐私机密计算技术可合规高效的利用多方大数据构建精准的DRG模型，有效地聚合了多渠道、多平台的数据源，为医保控费提供有效解决方案。

### 4.3.3 政务数据开放

2015年9月国务院发布了《关于促进大数据发展的行动纲要》，标志着大数据在我国的发展与应用上升到国家战略层面。2020年4月《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》，意见首次将“数据”与土地、劳动力、资本、技术等传统要素并列为要素之一，提出要加快培育数据要素市场，推进政府数据开放共享。随着中央将政务大数据开放纳入顶层设计，各级政府在逐步加快大数据的共享开放，以安全、融合、集约和服务化为特点的政务大数据平台将是未来智慧型政府的建设重点。

政务数据是国家的重要战略资源，涉及城市管理的人、事、物等方方面面，其中包含了大量的敏感信息。隐私机密计算的特性可为政务大数据平台提供，自上而下、由内到外，从政府到企业到群众，同时联合社会各级组织企业等能力，最终形成全面的智慧化数据化的政务体系，为社会发展提供更好的服务。利用隐私机密计算技术，实现政务数据的安全合规共享。盘活政务数据资产，通过政务数据的挖掘对外赋能，服务社会，服务人民群众，借助大数据创新技术辅助决策，提高行政管理能力。

# 5. 隐私机密计算的评价方法

一个可持续发展的完整可靠、可信、可用的大数据隐私机密计算技术解决方案，需要考虑以下几个方面：

## 5.1 合规角度

业务模式是否能够满足现有和未来的法律法规要求？特别是符合国家相关法律法规要求的能力，满足特定客户特定场景的业务合规需求的能力。数据处理的全生命周期是否合规？如何落实监管部门的要求？如何平衡数据处理的成本与收益？如何设计出更加弹性化的解决方案？

特别是根据最新的通过的《中华人民共和国数据安全法》，明确企业合规义务，企事业单位需要对数据进行评估，列举了一系列需要遵守的合规义务，例如开展安全培训、完善制度建设、风险评估监测、报告安全事件、落实数据分级分类等制度，这些都需要方案能够有效支撑。

## 5.2 技术角度

为满足业务要求，符合相关法律法规，提供全方位的技术支持，同时满足相关处理能力以及性能要求。具体有以下几个方面：

- 数据保护：指在不向任何节点透露原始数据的前提下的计算的能力，以及保护计算算法、计算结果、查询条件等能力，这是隐私机密计算的核心。
- 计算正确性：证明计算工作实际上是使用规定的函数。在无信任的网络中，证明以正确的方式执行某个函数是非常重要的。
- 安全性证明：证明计算实际上是在安全环境中进行，能够在恶意环境下抵御攻击破坏。
- 计算效率：计算效率是能够有效处理问题的能力，特别是对计算资源要求是否足够低。
- 计算能力：是否能够满足政务行业、金融保险行业、医疗等行业处理海量数据的能力，满足特定业务场景的需求；不仅能够保护涉及个人隐私数据，也能保

护公司商业机密。

# 6. 隐私机密计算未来发展

在新基中包括5G基站建设、特高压、城际高速铁路和城市轨道交通、新能源汽车充电桩、大数据中心、人工智能、工业互联网七大领域，是以技术创新为驱动，以信息网络为基础，面向高质量发展需要，提供数字转型、智能升级、融合创新等服务的基础设施体系。这其中，数据作为核心生产要素受到格外重视，但是由于数据载体形式的特殊性，如何保证数据所有者权益，保护数据安全，发挥数据的价值，是大数据产业进一步发展的一大挑战。

隐私机密计算作为数据应用的新型技术，可满足数据资产化的需求，促进大数据产业发展，为新基建提供强有力支撑。可以预见的是，随着国家对隐私数据监管的加强，特别是《中华人民共和国数据安全法》已在6月通过并发布，以及《个人信息保护法》的推进和后续的落地，企业与个人对自身数据价值重视程度的提高，数据资产化的发展。作为大数据行业，人工智能行业的技术基础设施，隐私机密计算需求将在未来几年实现爆炸式增长。

隐私机密计算未来的发展在于赋能业务场景落地，目前看来，隐私机密计算企业落地案例往往集中的金融、医疗、政务、营销等领域。有相关业务积累的公司会具有一定的优势，不同业务的业务逻辑千差万别，对于业务逻辑熟悉也是能否扎实落地的关键。竞争格局方面，互联网大厂体系玩家的主要优势在于其已有的生态体系、流量入口、以及其相对稳定性; 创业企业的主要优势为中立性、亲和的服务力和快速的决策能力。互联网大厂吸引人才优势在于其品牌效应和工作稳定性，而创业企业优势在于员工个人价值体现和极大的股权增值空间。总结隐私机密计算创业公司代表企业如下表所示：

| 隐私计算代表企业 | 企业类型 | 主要创始人员 | 成立时间 | 技术方向 | 落地行业 |
|---|---|---|---|---|---|
| 锘崴科技 | 创业公司 | CEO: 郑灏，原硅谷资深信息学科学家，上交本科、Georgia Tech博士；<br>CTO: 王爽教授，前UCSD隐私机密计算教授，安全联邦学习的开拓者，iDASH创始人，全球隐私计算顶尖专家，同济大学、华西医院特聘教授。 | 2019 | 隐私机密计算，TEE、安全联邦学习、MPC等技术 | 医疗/政务等 |
| 光之树科技 | 创业公司 | CEO：张佳辰，哈佛商学院MBA学位和肯尼迪政府学院公共政策硕士学位，曾任亚马逊高级产品经理；<br>密码学科学家：袁晨，上海交通大学网络空间安全学院院长聘副教授，多篇论文发表于美密、欧密、亚密等国际顶级学术期刊。 | 2017 | 区块链、联邦学习、MPC、TEE | 金融等 |
| 华控清交 | 创业公司 | CEO：前高盛全球合伙人张旭东<br>CTO:徐葳教授，现任清华大学交叉信息研究院助理院长兼博士生导师，智能金融科技研究院副院长 | 2018 | 以MPC为基础，结合数据脱敏、区块链等技术 | 金融、能源、政务等 |
| 富数科技 | 创业公司 | CEO：张伟奇，上交硕士，拥有十五年人工智能、推荐引擎技术经验；<br>CTO: 朱学政, 中山大学硕士，历任文讯信息技术合伙人，阿里巴巴B2B技术架构师，挖财技术总监。 | 2016 | 联邦学习、MPC | 金融等 |
| 星云 | 创业公司 | CEO: 陈沫 CEO, 中科院计算所博士, 曾任真格基金 Venture Partner、ofo VP；<br>CTO：张骏雪，香港科技大学博士，数据中心网络专家。 | 2018 | 联邦学习 | 金融/医疗等 |
| 金智塔 | 创业公司 | CEO：李岩，阿里中供铁军早期员工，前联想渠道经理，成章创客前CEO；<br>CTO: 新文，浙江大学博士，UIUC访问研究员。 | 2012 | MPC为主，联邦学习为辅 | 金融/政务 |
| 蓝象智联 | 创业公司 | CEO：徐敏，原阿里金融云总经理；<br>CTO：童玲，原蚂蚁集团首席架构师。 | 2019 | MPC | 金融等 |

# 结束语

随着大数据及人工智能领域不断发展，隐私保护和数据应用的安全问题不断显现，人们对这一问题的重要性也有了新的认知。相关行业的需求仍然在不断快速增长，人们迫切地需要一套完善的方案以解决这一问题，从而实现其他技术和行业的进一步发展。由此发展出了包括隐私机密计算在内的新兴技术以解决相关的问题，同时我国政府也从法律层面对相关行业进行了规范。尽管如此，我国在隐私保护领域上仍处于探索阶段，还有很多问题尚待解决。希望各方共同努力，共同促进产业发展，为提升我国数据产业的发展水平贡献力量。

2021隐私计算蓝皮书交稿付梓之际，《中华人民共和国数据安全法》正式发布，这无疑对隐私机密计算是一个有力推动。中国移动通信联合会数据融合委员会诚意邀请各位产学研专家学者，进一步研究探讨，期待下一版的蓝皮书，将会提供更丰富和更深入的内容，为促进产业发展做出更大的贡献。

# 免责申明

在任何情况下，本蓝皮书中的信息或表述的意见并不构成对任何人的投资建议，本蓝皮书所载的资料、工具、意见及推测只作为参考之用，并非作为或被视为出售或购买证券或其他投资标的的邀请或向人做出邀请。在任何情况下，蓝皮书的编著机构不对任何人因使用本蓝皮书中的任何内容所引致的任何损失付任何责任。

本蓝皮书主要以电子版形式分发，间或也会辅以印刷品形式分发，所有蓝皮书版权均归属编著机构所有。未经编著机构事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播蓝皮书的全部或部分内容，不得将蓝皮书内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其他用途。如需引用、刊发或转载本蓝皮书，需注明出处，且不得对本蓝皮书进行任何有悖原意的引用、删节和修改。

所载资料仅供一般参考用，并非针对任何个人或者团体的个别情况而提供。虽然我们已致力提供准确和及时的资料，但我们不能保证资料在阁下收取时或日后仍然准确。任何人士不应在没有详细考虑相关的情况及获取适当专业意见下依据所载资料行事。

# 参考文献


[1] S. Wang, X. Jiang, Y. Wu, L. Cui, S. Cheng, and L. Ohno-Machado, “EXpectation Propagation LOgistic REgRession (EXPLORER): distributed privacy-preserving online model learning,” *J. Biomed. Inform.*, vol. 46, no. 3, pp. 480–496, Jun. 2013.

[2] W. Jiang *et al.*, “WebGLORE: a web service for Grid LOgistic REgression,” *Bioinformatics*, vol. 29, no. 24, pp. 3238–3240, Dec. 2013.

[3] “pSCANNER.” http://pscanner.ucsd.edu/ (accessed May 03, 2021).

[4] B. Tan, Y. Zhang, S. Pan, and Q. Yang, “Distant Domain Transfer Learning,” *AAAI*, vol. 31, no. 1, Feb. 2017, Accessed: May 03, 2021. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/10826.

[5] Y. Liu, Y. Kang, C. Xing, T. Chen, and Q. Yang, “A Secure Federated Transfer Learning Framework,” *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 70–82, Jul. 2020.

[6] “徐汉明专栏(第20期).” http://www.hubeitoday.com.cn/post/6/128454 (accessed Sep. 25, 2020).

[7] “中华人民共和国侵权责任法（主席令第二十一号）.” http://www.gov.cn/flfg/2009-12/26/content 1497435.htm (accessed Sep. 25, 2020).

[8] “中华人民共和国刑法修正案（七）（主席令第十号）.” http://www.gov.cn/flfg/2009-02/28/content 1246438.htm (accessed Sep. 25, 2020).

[9] “全国人大常委会关于加强网络信息保护的决定.” http://www.gov.cn/jrzg/2012-12/28/content 2301231.htm (accessed Sep. 25, 2020).

[10] “《电信和互联网用户个人信息保护规定》发布.” http://www.gov.cn/gzdt/2013-07/19/content 2451360.htm (accessed Sep. 25, 2020).

[11] “中华人民共和国消费者权益保护法.” http://www.gov.cn/jrzg/2013-10/25/content 2515601.htm (accessed Sep. 25, 2020).

[12] 朱英, “中华人民共和国网络安全法.” http://www.gov.cn/xinwen/2016-11/07/content 5129723.htm (accessed Sep. 25, 2020).

[13] “全国信息安全标准化技术委员会.” https://www.tc260.org.cn/front/postDetail.html?id=20180124211617 (accessed Sep. 25, 2020).

[14] “国家标准 - 全国标准信息公共服务平台.” http://std.samr.gov.cn/gb/search/gbDetailed?id=91890A0DA4AB80C6E05397BE0A0A065D (accessed Sep. 25, 2020).

[15] R. L. Rivest, L. Adleman, M. L. Dertouzos, and Others, “On data banks and privacy homomorphisms,” *Foundations of secure computation*, vol. 4, no. 11, pp. 169–180, 1978.

[16] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in *Proceedings of the forty-first annual ACM symposium on Theory of computing*, Bethesda, MD, USA, May 2009, pp. 169–178, Accessed: May 10, 2020. [Online].

[17] A. Acar, H. Aksu, A. Selcuk Uluagac, and M. Conti, “A Survey on Homomorphic Encryption Schemes: Theory and Implementation,” *arXiv [cs.CR]*, Apr. 12, 2017.

[18] Z. Shan, K. Ren, M. Blanton, and C. Wang, “Practical Secure Computation Outsourcing: A Survey,” *ACM Comput. Surv.*, vol. 51, no. 2, pp. 31:1–31:40, Feb. 2018.

[19] T. A. Hemphill and P. Longstreet, “Financial data breaches in the U.S. retail economy: Restoring confidence in information technology security standards,” *Technol. Soc.*, vol. 44, pp. 30–38, Feb. 2016.

[20] T. Moore, “On the harms arising from the Equifax data breach of 2017,” *Int. J. Crit. Infrastruct. Prot.*, vol. 19, pp. 47–48, Dec. 2017.

[21] N. Manworren, J. Letwat, and O. Daily, “Why you should care about the Target dat

a breach," *Bus. Horiz.*, vol. 59, no. 3, pp. 257–266, May 2016.
[22] V. Lyubashevsky, C. Peikert, and O. Regev, "A Toolkit for Ring-LWE Cryptography," in *Advances in Cryptology – EUROCRYPT 2013*, 2013, pp. 35–54.
[23] O. Regev, "The learning with errors problem," *Invited survey in CCC*, vol. 7, 2010, [Online]. Available: https://pdfs.semanticscholar.org/4cb8/ebc5bc8adc450c40albed072a607a287f2bb.pdf.
[24] M. Kim, Y. Song, S. Wang, Y. Xia, and X. Jiang, "Secure Logistic Regression Based on Homomorphic Encryption: Design and Evaluation," *JMIR Med Inform*, vol. 6, no. 2, p. e19, Apr. 2018.
[25] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti, "Addressing the concerns of the lacks family: Quantification of kin genomic privacy," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, Nov. 2013, pp. 1141–1152.
[26] Z. Huang, E. Ayday, J. Fellay, J. P. Hubaux, and A. Juels, "GenoGuard: Protecting Genomic Data against Brute-Force Attacks," in *2015 IEEE Symposium on Security and Privacy*, May 2015, pp. 447–462.
[27] K. Shimizu, K. Nuida, and G. Rätsch, "Efficient privacy-preserving string search and an application in genomics," *Bioinformatics*, vol. 32, no. 11, pp. 1652–1661, Jun. 2016.
[28] G. S. Çetin, H. Chen, K. Laine, K. Lauter, P. Rindal, and Y. Xia, "Private queries on encrypted genomic data," *BMC Med. Genomics*, vol. 10, no. Suppl 2, p. 45, Jul. 2017.
[29] M. Kim, Y. Song, and J. H. Cheon, "Secure searching of biomarkers through hybrid homomorphic encryption scheme," *BMC Med. Genomics*, vol. 10, no. Suppl 2, p. 42, Jul. 2017.
[30] E. Ayday, J. L. Raisaro, P. J. Mclaren, J. Fellay, and J.-P. Hubaux, "Privacy-Preserving Computation of Disease Risk by Using Genomic , Clinical , and Environmental Data," 2013.
[31] P. J. McLaren *et al.*, "Privacy-preserving genomic testing in the clinic: a model using HIV treatment," *Genet. Med.*, 2016.
[32] Y. Zhang, W. Dai, X. Jiang, H. Xiong, and S. Wang, "FORESEE: Fully Outsourced secuRe gEnome Study basEd on homomorphic Encryption," *BMC Med. Inform. Decis. Mak.*, vol. 15 Suppl 5, p. S5, Dec. 2015.
[33] S. Wang *et al.*, "HEALER: homomorphic computation of ExAct Logistic rEgRession for secure rare disease variants analysis in GWAS," *Bioinformatics*, vol. 32, no. 2, pp. 211–218, Jan. 2016.
[34] W. Lu, Y. Yamada, and J. Sakuma, "Efficient Secure Outsourcing of Genome-wide Association Studies," 2015.
[35] M. Kim and K. Lauter, "Private Genome Analysis through Homomorphic Encryption," *BMC Med. Inform. Decis. Mak.*, vol. 15 Suppl 5, no. Suppl 5, p. S3, Dec. 2015.
[36] Y. Zhang *et al.*, "SECRET: Secure Edit-distance Computation over homomoRphic Encrypted daTa," 2015.
[37] J. H. Cheon, M. Kim, and K. Lauter, "Homomorphic Computation of Edit Distance," in *Financial Cryptography and Data Security*, Jan. 2015, pp. 194–212, Accessed: Oct. 27, 2016. [Online].
[38] S. Jha, L. Kruger, and V. Shmatikov, "Towards Practical Privacy for Genomic Computation," in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, May 2008, pp. 216–230.
[39] X. S. Wang, Y. Huang, Y. Zhao, H. Tang, X. Wang, and D. Bu, "Efficient Genome-Wi

de, Privacy-Preserving Similar Patient Query based on Private Edit Distance," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15*, Oct. 2015, pp. 492–503.

[40] R. Wang, X. Wang, Z. Li, H. Tang, M. K. Reiter, and Z. Dong, "Privacy-preserving genomic computation through program specialization," in *Proceedings of the 16th ACM conference on Computer and communications security - CCS '09*, Nov. 2009, p. 338.

[41] G. Asharov, S. Halevi, Y. Lindell, and T. Rabin, "Privacy-Preserving Search of Similar Patients in Genomic Data," *IACR Cryptology ePrint Archive*, vol. 2017, p. 144, 2017.

[42] R. Zhu and Y. Huang, "Efficient privacy-preserving general edit distance and beyond," Cryptology ePrint Archive, Report 2017/683, 2017. http://eprint. iacr. org/2017/683 10 April 2017, date last accessed, 2017. [Online]. Available: https://pdfs.semanticscholar.org/4462/4f5e9213843443719b3a52c6e5b40281c89f.pdf.

[43] M. M. A. Aziz, D. Alhadidi, and N. Mohammed, "Secure approximation of edit distance on genomic data," *BMC Med. Genomics*, vol. 10, no. 2, p. 41, Jul. 2017.

[44] H. Shi *et al.*, "Secure Multi-pArty Computation Grid LOgistic REgression (SMAC-GLORE)," *BMC Med. Inform. Decis. Mak.*, vol. 16 Suppl 3, p. 89, Jul. 2016.

[45] Y. Wu, X. Jiang, J. Kim, and L. Ohno-Machado, "Grid Binary LOgistic REgression (GLORE): building shared models without sharing data," *J. Am. Med. Inform. Assoc.*, vol. 19, no. 5, pp. 758–764, Sep. 2012.

[46] Y. Wu, X. Jiang, S. Wang, W. Jiang, P. Li, and L. Ohno-Machado, "Grid multi-category response logistic models," *BMC Med. Inform. Decis. Mak.*, vol. 15, p. 10, Feb. 2015.

[47] K. El Emam, S. Samet, L. Arbuckle, R. Tamblyn, C. Earle, and M. Kantarcioglu, "A secure distributed logistic regression protocol for the detection of rare adverse drug events," *J. Am. Med. Inform. Assoc.*, vol. 20, no. 3, pp. 453–461, 2013.

[48] K. A. Jagadeesh, D. J. Wu, J. A. Birgmeier, D. Boneh, and G. Bejerano, "Deriving genomic diagnoses without revealing patient genomes," *Science*, vol. 357, no. 6352, pp. 692–695, Aug. 2017.

[49] H. Cho, D. J. Wu, and B. Berger, "Secure Genome Crowdsourcing for Million-Individual Association Studies," *Nat. Biotechnol.*, 2018.

[50] M. Sabt, M. Achemlal, and A. Bouabdallah, "Trusted execution environment: what it is, and what it is not," in *2015 IEEE Trustcom/BigDataSE/ISPA*, 2015, vol. 1, pp. 57–64.

[51] M. Russinovich, "Introducing Azure confidential computing," *Financial Times*.

[52] Y. Xu, W. Cui, and M. Peinado, "Controlled-channel attacks: Deterministic side channels for untrusted operating systems," in *Security and Privacy (SP), 2015 IEEE Symposium on*, 2015, pp. 640–656.

[53] G. Chen *et al.*, "Racing in Hyperspace: Closing Hyper-Threading Side Channels on SGX with Contrived Data Races," in *2018 IEEE Symposium on Security and Privacy (SP)*, May 2018, pp. 178–194.

[54] W. Wang *et al.*, "Leaky Cauldron on the Dark Land: Understanding Memory Side-Channel Hazards in SGX," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, Dallas, Texas, USA, 2017, pp. 2421–2434.

[55] W. Dai, S. Wang, H. Xiong, and X. Jiang, "Privacy preserving federated big data analysis," *Guide to Big Data Applications*, 2018, [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-53817-4_3.

[56] Xin Wu, Hao Zheng, Zuochao Dou, Feng Chen, Jieren Deng, Xiang Chen, Shengqian Xu, Shengqian Xu, Guanmin Gao, Mengmeng Li, Zhen Wang, Yuhui Xiao, Kang Xie, Shuang Wan

g, Huji Xu, “A Novel Privacy-Preserving Federated Genome-wide Association Study Framework and its Application in Identifying Potential Risk Variants in Ankylosing Spondylitis.” Accepted on May 2020.

[57] G. Mateos, J. A. Bazerque, and G. B. Giannakis, “Distributed Sparse Linear Regression,” *IEEE Transactions on Signal Processing*, vol. 58, no. 10. pp. 5262–5276, 2010, doi: 10.1109/tsp.2010.2055862.

[58] I. D. Schizas and A. Aduroja, “A Distributed Framework for Dimensionality Reduction and Denoising,” *IEEE Transactions on Signal Processing*, vol. 63, no. 23. pp. 6379–6394, 2015, doi: 10.1109/tsp.2015.2465300.

[59] P. A. Forero, A. Cano, and G. B. Giannakis, “Consensus-based distributed linear support vector machines,” *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks - IPSN '10*. 2010, doi: 10.1145/1791212.1791218.

[60] F. Chen *et al.*, “PRINCESS: Privacy-protecting Rare disease International Network Collaboration via Encryption through Software guard extensionS,” *Bioinformatics*, vol. 33, no. 6, pp. 871–878, Mar. 2017.

[61] W. Wei *et al.*, “A Framework for Evaluating Gradient Leakage Attacks in Federated Learning,” *arXiv [cs.LG]*, Apr. 22, 2020.

[62] K. Benitez and B. Malin, “Evaluating re-identification risks with respect to the HIPAA privacy rule,” *J. Am. Med. Inform. Assoc.*, vol. 17, no. 2, pp. 169–177, 2010.

[63] M. Gong *et al.*, “Evaluation of Privacy Risks of Patients’ Data in China: Case Study,” *JMIR Medical Informatics*, vol. 8, no. 2. p. e13046, 2020, doi: 10.2196/13046.