**文献综述**

1. **安全多方计算与联邦学习的背景及相关理论**
2. **安全多方计算**

安全多方计算起源于1982年姚期智的百万富翁问题。在没有可信第三方的情况下，两个百万富翁如何不向对方揭露自己的真实财富而比较出谁更有钱。工作原理可简单概括为在各方不可信的情况下，不泄露输入数据来安全地进行多方协同的计算，实质上是一种加密技术，使参与数据共享各方隐私得到保护的同时，共同合作完成某个数据加工函数的计算任务，使各方得到各自需要的正确结果，即通过安全多方计算技术使用数据者看不到明细数据但可以加工数据，在没有可信第三方时，保证数据所有权和使用权隔离，平衡了数据共享融合和数据隐私保护方面。

这与区块链去中心化的场景十分吻合。区块链和安全多方计算能够相互促进，形成更加安全可靠的用户交易环境。一方面，安全多方计算增加了区块链的安全性和功能性，不信任的用户在不暴露各方输入金额的情况下，利用所有用户的输入联合计算出结果，且结果的正确性是可验证的；另一方面，区块链提供了安全可靠的信息通讯与数据存储平台，互不信任的用户在区块链上形成一致性的交易历史，提供防篡改、智能合约、去中心化和公开溯源等特性。魏立斐等人（2020）指出，安全多方计算技术具有可用性高、通信开销大、效率低的特点，它适合分布式学习、联邦学习等场景。

**2.联邦学习**

传统算法在工业界、金融界的应用正在受到挑战。杨强（2020）指出：在当前大数据驱动的社会环境下，数据隐私安全成为了全民探讨的重要议题。信息技术的发展，离不开由我们在智能终端（手机及其他设备等）上产生或推断出的个人数据，如浏览习惯、点击频次等，来推动个性化应用和服务的发展。尤其在 AI 领域，这一情况更为明显，依赖于持续的数据感知、收集，并上传至服务端进行深度分析与训练，AI 才能迎来蓬勃发展。但底层未经审查、不透明的数据收集和聚合协议，很可能造成严重的数据安全威胁和隐私风险。

联邦学习是一种支持大规模分散数据集进行模型训练的分布式机器学习方法该方法避免了在机器学习过程迁移数据集,可解决数据隐私、所有权和位置等基本问题。 不久前，由中国人民大学未来法治研究院主办、中央网信办《数据安全立法研究》重大项目课题组协办的《数据安全法（草案）》暨数据法治研讨会在线上举行，来自中央网信办政策法规局、多所高校、企业智库、知名科研机构的近 20 名代表及资深专家学者与会。会议以《数据安全法（草案）》修改为主题，就数据安全法私法视野、数据法视野、公法视野等角度进行了讨论。专家们都明确了一个态度就是数据的保护一定会越来越严谨。而这也对需要数据的各行各业提出了挑战和风险，联邦学习经研究可以很好的被投入到保险公司和电信公司的运营中，而我们需要解决联邦学习在金融业中的技术壁垒和如何商业化。

1. **多方安全计算与联邦学习的研究现状**

秘密分享是构造安全多方计算的重要密码技术。秘密分享是将秘密以适当的方式拆分，拆分后的每一个份额由不同的参与者管理，单个参与者无法恢复秘密信息，只有若干个参与者一同协作才能恢复秘密消息。当其中任何相应范围内参与者出问题时，秘密仍可以完整恢复，也就是说，即使存在恶意参与方，安全多方计算仍然能够保证最终输出的正确性，且在此过程中不暴露任何参与方的输入。然而，目前绝大多数安全多方计算不适用于区块链场景。首先，因为乘法门的执行需要消耗一定量的通信，所以实现起来通信轮数多，通信量大，导致算法本身和区块链平台运行效率降低。在2016年美密会议上，DANMGARD 等人提出，安全多方计算的轮数至少是电路的乘法深度。此外，由于区块链中的节点是不可信的，大量的通信数据也给节点带来繁重的验证负担。

为应对区块链受通信轮数与通信量严重限制的问题，ZHOU 等人基于区块链可验证计算研究低轮通信安全多方计算，在Shamir秘密分享算法递进式基础上提出一轮通信安全多方计算方案——AnyNest。用户能够以一轮通信的方式在不暴露自身信息的情况下联合所有用户执行计算，并且最终结果的正确性可以验证。结合安全多方计算和区块链，MIT Engima将区块链用于医疗数据的分享和隐私计算，Engima在底层区块链之上提供了秘密分享和安全计算的技术，用于保障上层医疗数据的安全共享和智能合约的自动化隐私支付。安全多方计算部分在Engima的链下执行，公链将主要通过验证安全多方计算的证明来负责可执行的代码正确执行审计功能。BENHAMOUDA等人在超级账本上实现了一个竞标系统，该系统向超级账本添加私有数据，节点将其私有数据的密文存储在链上，并在交易需要这一私有数据时使用安全多方计算。

2015年，以太坊创始人BUTERINV（2016）在以太坊白皮书中描述，智能合约是一个运行在安全环境下的计算机程序，当特定的条件满足时，程序就可以直接控制数字资产的转移，实现自我执行和自我验证，不需要人为的干预。其中区块链技术是支撑智能合约得以实现的基础架构。另一方面，在央行法定数字货币与区块链技术的研究中，区块链技术可以被应用于数字货币钱包地址的管理、交易信息的监管以及数字票据的交易监管。

但是随着实际应用范围的进一步扩大，在诸如大宗或频繁交易中，区块链的公开可验证性为用户隐私带来了安全挑战，同时其性能问题，特别是在交易吞吐量和可扩展性方面，也限制了区块链技术的进一步发展。对此，王童（2019）等人提出基于区块链的信息共享及安全多方计算模型，通过链上存储与链下存储相结合的方式来存储数据，即将索引信息存储于链上，大规模数据信息则在链下进行存储，以此解决区块链大规模数据的分析计算问题，同时利用安全多方计算模型进行隐私保护。随着云计算的不断发展，越来越多不同领域内的应用也逐渐转移到云平台上。蒋瀚（2016）等人提出利用云作为外部服务器的同时，还可以作为安全多方计算的一个辅助设施，同时通过与安全多方计算结合，更好地完成保护隐私的数据处理、加密数据的处理。

另一方面，安全多方计算被证明，当超过一半的参与者不诚实时，公平性就无法得到保证。对此，黄建华（2020）等人提出了基于区块链的公平安全多方计算，利用区块链具有去中心化、去信任及不可窜改等特点，以及其上的智能合约不仅可以不为人控制地自动执行代码，存储数据，还能和普通账户一样存储价值的机制来解决公平性问题，这个新方法就是在安全多方计算协议中加入经济惩罚机制，具体来说就是参与方在执行计算之前先缴纳押金，计算结束后对参与方行为进行验证，诚实者的押金将被退回，恶意者的押金将平分给诚实者。董祥千（2018）等人认为当前数据开放共享模型存在一定缺陷:(1)以关键字为基础的数据检索无法高效发现可连接数据集;(2)数据交易缺乏透明性，无法有效检测及防患交易参与方串谋等舞弊行为; (3)数据所有者失去数据的控制权、所有权、数据安全无法保障。对此，他们提出去中心化数据共享模型结构解决以上问题。

1. **多方安全计算与联邦学习的实际应用发展前景**

联邦学习作为一项全新的基于分布式的机器学习算法，其可能的应用前景是广阔的。杨强（2020）指出：联邦学习希望做到各个企业的自有数据不出本地，而联邦系统可以通过加密机制下的参数交换方式，即在不违反数据隐私法规情况下，建立一个虚拟的共有模型。这个虚拟模型就好像大家把数据聚合在一起建立的最优模型一样。但是在建立虚拟模型的时候，数据本身不移动，也不泄露隐私和影响数据合规，也就是“数据不动，模型动”。这样，建好的模型在各自的区域仅为本地的目标服务。在这样一个联邦机制下，各个参与者的身份和地位相同，而联邦系统帮助大家建立了“共同富裕”的策略，也就是“风险不增，效益增”。

基于以上特性，学者提出了其在多方面的应用可能。尽管因为技术原因，目前金融行业尚未进行尝试；然而其在其他行业中的应用可以给金融打下基础，确定方向，提供一个合适的应用模式参考。

在疫情防控方面，徐萍，何家俊，岳小尧（2020）认为：作为新冠肺炎疫情所有信息的重要入口，互联网成为这次疫情主要的“信息源”平台。在这场疫情阻击战中，以计算机技术为依托，融合多种算法深度挖掘和实时跟踪全网数据，从高校管理侧打造防疫战线，发挥计算机技术战“疫”价值将十分重要。联邦学习技术，应用在针对高校新冠肺炎疫情防控的管理模型上，可以在保证每位师生的信息安全前提下，快速分析高校所收集的师生数据，从而及时安排相应的防疫措施，避免新冠肺炎疫情进一步扩散。

在教育领域，李默妍（2020）认为：随着隐私泄露问题的凸显，在保护学习者隐私的基础上使用来自多方的数据以提升人工智能应用的性能成为智能时代亟待解决的问题。联邦学习与教育数据挖掘的各类算法相结合，能够解决教育数据挖掘中可能存在的隐私保护问题。联邦学习方法能够从原理上保障数据隐私且容易整合到现有的教育应用中，在面对不可公开数据如学生成绩时更是有优良的表现。

在通信领域，陈国润等（2020）认为：利用联邦学习能有效地甄别和预防通信诈骗。出于用户隐私和数据安全的原因，各方原始数据信息不能直接进行交换，传统建模面临困难。联邦学习概念的提出能有效解决基于隐私保护的分布式安全联合建模问题，通过对比传统机器学习和联邦学习在预测精度、训练时间等维度的差别，他们认为联邦学习在电信运营商和公安机关的数据集上联合建模具有可行性。

相较于联邦学习，安全多方计算作为一项密码学技术，提出比前者早了将近10年，研究也更为成熟，也有学者在金融领域提出了一些设想。

在机械工业上，李禾，王述洋（2008）认为：安全多方计算新的应用方向可以是以传统模型为主的机械设计工程领域，如产品的异地设计和制造 、敏捷和虚拟制造技术等；而在保护隐私基础上合作的安全多方计算将成为热点问题。

在供应链金融上，董绍辉 ，西宝 ，田丽娜（2008）认为：传统供应链产能分配过程中信息泄露将对供应链效率产生负面影响。而基于安全多方计算的按比例和线性安全分配机制可以确保制造商产能分配活动顺利进行 、零售商获得合理的分配量的同时，不泄露参与者的私有信息。

同时，ARPA 联合创始人章磊提出了多方安全计算的五种应用场景，包括基金联合收益计算、联合个人征信、供应链金融、机器学习和联合风险价值计算。在供应链金融的应用中，其解决了传统供应链实现过程中，传统的商票不可拆分、金融机构存在较大的授信风险、以及信息有效传递较难导致的融资等问题。在他的模型中，通过 MPC 和零知识证明等加密技术，可帮助区块链实现智能合约的公开审计确认能力与实际数据保密性的分离，让企业不再担心核心商业信息的泄露。

**四．工业应用的尝试与分析**

在具体应用上，目前有着不同的尝试。杨强（2020）认为：目前人工智能

大部分应用场景存在小数据和数据孤岛，需要数据孤岛和合规的数据整合。微众银行构建了首个工业级框架，其白皮书（2020）认为：联邦学习作为一种保障数据安全的建模方法，在销售、金融等行业中拥有巨大的应用前 景。在这些行业中，受到知识产权、隐私保护、数据安全等诸多因素影响，数据无法被直接聚合来进行机器学习模型训练。此时，就需要借助联邦学习来训练一个联合模型。微众给出了智慧金融—FL实现检测多头借贷问题与智慧医疗—IBM“沃森”的关于联邦学习应用的具体案例。

在金融行业，刘洋（2020）认为：目前，在金融领域中，微众银行主要将联邦学习技术用于小微企业信贷以及个人贷款的风险管理，旨在解决银行在此类场景下存在的「数据维度缺失」、「数据低频」等问题。在银行中，我们可能收集到的数据包括征信报告以及用户在银行的一些表现数据。然而，由于用户使用银行的频次非常低，我们还需要结合第三方的数据（例如，来自互联网及其它数据源的数据）进行风控建模。实际上，在互联网金融领域中，利用上游数据的概念已经较为常见，但是随着监管越来越严，数据出库的难度日益提升，我们希望通过联邦学习技术搭建一个桥梁，连接银行数据以及其它的互联网数据。我们可以通过联邦学习技术打通企业征信和发票数据的交互渠道。此时，在合作企业段，数据可能包括这些小微企业的流水，这些流水是高频且相关的，流水的状况可以很好地反映一个企业的资金流和经营状况。通过使用联邦学习技术，我们可以对企业的征信进行更好的预测和评分。

关乎如何对待数据提供者，翁翕（2020）等人认为：对于数据参与者来说，联邦学习应用框架有必要使其有意愿参与联盟，愿意贡献全部与最好的数据，贡献数据的成本是私人信息，而得到的报酬应当与数据的数量与质量成正比，并具有稳定性。长期性。

**五．技术监管与未来治理**

在对这些新兴技术的监管上，应当回归数据价值。Hyesung Kim（2020）等人认为：区块链技术与联邦学习技术有着相似性与兼容性，基于区块链设计联邦学习模型是一种好的建模方式。基于此，一些区块链的未来监管思路将对新技术的监管有所启发。杨强（2020）认为：网络结构上，均是一种去中心化网络结构，稍有区别是联邦第三方会承担汇聚模型，区块链完全是点对点的；安全算法上，联邦学习使用同态加密，区块链使用哈希算法、非对称加密；数据角度上，区块链的每个节点上记录了完整的密文数据，联邦学习的数据均保留在用户本地；奖励机制上，区块链的节点间通过竞争记账获得奖励，联邦学习中依据各方的共贡献分配奖励。

在未来监管上，杨东（2019）认为：区块链的制度设计应当实现从“监管”转向“治理”的思维转变。当下区块链治理的主要问题在于市场被错误的理念支配，规制者缺乏技术规制工 具。“共票”是区块链上集投资者、消费者与管理者三位一体的共享分配机制，同时也能对数据赋权、确权、赋能，能为以数据为核心的数字经济激发新动能。

参考文献

1. 魏立斐，陈聪聪，张蕾，李梦思，陈玉娇，王勤：《机器学习的安全问题及隐私保护》，载《计算机研究与发展》2020年第10期
2. DANMGARD I, NIGLSEN J, POLYCHRONIADOU A, et al. On the Communication Required for Unconditionally Secure Multiplication [C]//Lecture Notes in Computer Science: 9815. Heidelberg: Springe Verlag, 2016: 459-488
3. ZHOU L, WANG L, SUN Y, et al. Ant Nest: Fully Non-Interactive Secure Multi-Party Computation[J].IEEE Access, 2018, 6: 75639-75649.
4. BENHAMOUDA F, HALEVI S, HALEVI T. Supporting Private Data on Hyperledger Fabric with Secure Multiparty Computation[C]//Proceedings of the 2018 IEEE International Conference on Cloud Engineering. Piscataway: IEEE, 2018: 357-363.
5. 杨强：《联邦学习：人工智能的最后一公里》，载《智能系统学报》，第 15 卷第 1 期
6. 徐萍，何家俊，岳小尧：《基于边缘学习和联邦学习的新冠肺炎（COVID-19）高校防控管理模型研究》，载《当代教育论坛》，2020年第2期
7. 李默妍：《基于联邦学习的教育数据挖掘隐私保护技术探索》，载《电化教育研究》，2020.11.013
8. 陈国润，母美荣，张蕊，孙丹，钱栋军：《基于联邦学习的通信诈骗识别模型的实现》，载《运营与创新》，.2020年111
9. 李 禾，王述洋，《安全多方计算的应用研究》，载《中国安全科学学报》，第 18卷 第 3期
10. 董绍辉 ，西 宝 ，田丽娜，《基于安全多方计算的供应链产能分配机制》，载《物流管理》，第 23卷 第 5期
11. Mcmahan H B , Moore E , Ramage D , et al. Communication-Efficient Learning of Deep Networks from Decentralized Data[J]. 2016.
12. Bonawitz K, Eichner H,Grieskamp W,et al．Towards Federated Learning at Scale: System Design[J]．2019,CoRRabs /1092.01046
13. Yang Q,Liu Y,Chen T,etal．Federated Machine Learning: Conceptand Applications [J]．ACM TransＧ actions on Intelligent Systems and Technology,2019, 10(2):1２:1-12:19;
14. 杨强．《GDPR对 AI的挑战和基于联邦迁移学习的对策》，载《中国人工智能学会通讯》, 2018年第8期
15. BUTERINV.ANext-generationSmartContractandDecentralizedApplicationPlatform [EB/OL]. [2020-04-26]. http://www.fintech.academy/wp-content/uploads/2016/06/EthereumWhitePaper.pdf.
16. 王慧，王励成，柏雪等：《区块链隐私保护和扩容关键技术研究》，载《西安电子科技大学学报》,2020年第47卷第5期
17. 王童，马文平，罗维：《基于区块链的信息共享及安全多方计算模型》，载《计算机科学》，2019年第46卷第9卷
18. 蒋瀚，徐秋亮：《基于云计算服务的安全多方计算》，载《计算机研究与发展》2016年第53卷第10期
19. 黄建华，江亚慧，李忠诚：《利用区块链构建公平的安全多方计算》，载《计算机应用研究》2020年第37期第1卷
20. 董祥千，郭兵，沈艳，段旭良，申云成，张洪：《一种高效安全的去中心化数据共享模型》，载《计算机学报》，2018,年第41期第5卷
21. 刘洋，《联邦学习技术在金融领域的研究与应用》,《2020 中国科技峰会系列活动青年科学家沙龙——人工智能学术生态与产业创新》
22. 杨强等，《联邦学习白皮书V2.0》
23. 杨强，《AI向善，数据孤岛和联邦学习》，https://dl.ccf.org.cn/lecture/lectureDetail?id=4340567999891456
24. 翁翕，丛明舒，于涵：《联邦学习之激励机制设计》，https://dl.ccf.org.cn/lecture/lectureDetail?id=4841322416883712
25. 杨东：《“共票”：区块链治理新维度》，《东方法学》，2019年第三期
26. Hyesung Kim, Jihong Park, Mehdi Bennis, Seong-Lyun Kim，《On-Device Federated Learning via Blockchain and its Latency Analysis》，https://arxiv.org/abs/1808.03949v1

评述：

我们发现，由于这两项技术的不成熟性，特别是联邦学习，2016年才被正式提出，2019年才完成一个基本的框架，目前学术界对其技术讨论比较热烈，计算机界对其前景普遍看好，但其中技术细节还待讨论，绝大部分文献均是在研究其技术优化手段。在应用层面的讨论较少。我们认为一方面该技术开始应用于工业实践案例较少，时间较短，目前鲜有十分具有代表性的案例出现，因此分析文章鲜有素材；另一方面，一些技术性文章也提及，目前的框架有其进步性，但会导致成本上升、性能下降、算法整体复杂度大幅上升，工业界普遍呈观望态势。

我们深入了解两项技术后，认为其确有革命性的价值。大部分文章中都提到，多方安全计算与联邦学习能够打破数据壁垒，在扩大企业掌握数据量与加强数据隐私保护这对矛盾中找到了一个近乎完美的解决方案，如若成熟应用将解决数据匮乏、画像模糊、冷启动等等许多现行模型难以克服的缺点，对几乎所有有数据应用需求的行业都将造成冲击。

我们重点研究了讨论微众银行FATE框架的文章，分析了其应用案例。在其金融应用中，

不可否认的是，多方安全计算与联邦学习极大提高了数据合规性。我们发现，这从另一个角度来理解的话，就是现行以隐私保护为主导的数据监管模式已经不适用于该技术的监管。该技术将会在现行隐私保护法许可的情况下挖掘用户特征信息，导致大型企业能够掌握更多个人信息为企业自己服务，事实上将形成更大规模的数据垄断与对个人的更为严重的数据剥削。因此我们认为，在监管方面必须要强调数据价值，保障数据提供者的合法权益；同时应该限制企业行为，监管机关有必要加入企业训练数据模型的联盟，监督企业是否正当使用了用户的数据。

我们认真学习了杨东老师的共票理论。共票为数据赋能，作为大众参与创造数据的对价的想法给了我们启发。在联邦学习的数据联盟中，数据提供者天然处于弱势地位，受到掌握大量数据的企业的控制。而在联盟中引入共票作为激励机制，能够有效保障数据价值；同时能够规范企业数据收集方法，便于企业收集稳定、正确的数据，实现双赢。