**隐私计算专项科研基金课题目录**

**A类课题：**

1. [大模型推理过程中输入和输出保护的技术研究](#随着人工智能的不断发展与大模型的兴起)
2. [大模型隐私安全算法研究](#以GPT为代表的大语言模型和)
3. [安全高效的联邦大模型训练方法研究](#随着深度学习算法的发展和硬件计算能力的提升)
4. [面向量化神经网络的双方安全推理](#量化神经网络是一种对神经网络模型进行压缩和量化的技术)
5. [数据使用权跨域管控技术研究](#当数据要素作为一种要素)
6. [软硬件结合多方安全计算技术研究](#多方安全计算（MPC）技术是一种经典的密码学算法技术)
7. [GPU加速TFHE/TFHW同态加密方案及方案转换技术研究](#目前实现数据隐私保护主流的同态加密方案包括)
8. [基于CKKS和TFHE切换技术的密文AI推理加速器设计](#隐私保护是目前AI中比较重要的问题)
9. [宽松恶意安全的多方安全计算](#多方安全计算MPC是隐私计算的核心技术路线之一)
10. [隐私增强的跨机构分布式图计算](#PageRank、社区发现、标签扩散等图算法广泛应用于资金图)
11. [通用群模型（GGM）、代数群模型（AGM）下分布式签名及零知识证明安全性研究](#密码技术是构建区块链的重要基础设施)
12. [联营生态中个性化联邦学习相关技术](#在联营生态中，对人群)

**B类课题：**

1. [拆分学习攻防算法研究](#拆分学习是一种解决数据垂直拆分场景下跨域联合建模的方案)
2. [基于 GC 的双方 SHA256 以及 AES 协议实现](#在隐私计算中，有许多C端场景需要MPC对于传统密码套件的支持)
3. [基于 GPU 的加法同态加密算法设计与实现](#等同态加密算法已经在隐私计算领域大量应用)
4. [可信执行环境的状态连续性设计和实现](#可信执行环境（TEE）已经成为机密计算的基石)
5. [隐私计算节点间网络可观测性度量的研究](#隐私计算是一种跨域计算场景)

**A类课题**

**1. 大模型推理过程中输入和输出保护的技术研究**

**背景：**随着人工智能的不断发展与大模型的兴起，大模型推理的使用面越来越广，但其在隐私和安全方面的问题也逐渐暴露。例如，使用者在使用AIGC进行自主创作的过程中，模型的输入和输出可能包含了创作者的思想，这种信息的泄露会损害到创作者利益；而在一些医疗对话场景中，模型的输入和输出可能包含了使用者的隐私信息。基于此，研究大模型推理过程中的的输入和输出的保护技术，使得大模型能够为使用者提供服务的同时保护其创作价值和用户隐私，成为一个亟待解决的问题。

**目标：**

1）原先的生成任务基本不受影响，能完成指定的模型推理功能。

2）不能从保护后的输入内容和输出结果中反推出保护前的内容或语义信息，衡量标准是：保护后推断出的语义信息与保护前的语义信息通过Sentence Bert评测的语义相似度低于0.6。

3）推理性能和硬件成本与明文推理不能相差过大。

4）产出CCF-A类论文1篇。

5）申请发明专利至少1项。

6）源代码：原型技术demo及其技术源代码。

[返回首页](#目录)

**2. 大模型隐私安全算法研究**

**背景：**以GPT为代表的大语言模型和以Diffusion Model为代表的大图像模型均使用互联网上海量的公开数据进行训练，由于互联网数据的复杂多样并且大模型的数据记忆能力较强，大AIGC模型生成隐私安全的图像和文本内容面临着巨大的挑战，包括隐私风险识别、内容脱敏替换、模型隐私数据遗忘等技术方向。我们希望基于历史隐私安全研究的技术积淀，进一步探索大AIGC模型的隐私安全识别、替换、遗忘算法，全面提升隐私测评能力和隐私泄露防控能力，为未来大AIGC模型在更多场景落地进行保驾护航。

**目标：**

1）产出CCF-A顶会论文2-3篇；

2）产出专利1-2篇；

3）产出原型系统并在大模型测评系统中进行验证，提升隐私敏感内容召回并降低对非隐私敏感问题的误打扰。

[返回首页](#目录)

**3. 安全高效的联邦大模型训练方法研究**

**背景：**随着深度学习算法的发展和硬件计算能力的提升，大模型语言模型展现出远超小模型的涌现能力、更高的准确性、更好的泛化能力，可以处理更复杂的任务。然而，大模型的训练需要大量的文本数据，而这些数据往往分布在不同的地方，无法集中到一个中心服务器进行训练。其次，大模型训练还面临着模型安全性问题和数据隐私性问题，因为模型本身对其所有者来说是一种宝贵的财产，而训练数据可能包含敏感信息，因此对模型和数据都需要采取相应的保护措施。此外，大模型的训练过程需要消耗大量的计算资源和时间，训练过程可能会非常耗时，增加了训练的难度和成本。近年来，联邦学习作为一种分布式机器学习技术，因其能够在保护数据隐私的同时进行模型训练而受到广泛关注。但大模型的出现，打破了传统的联邦学习的问题假设，因此也需要探索和研究针对大模型新问题的解决方案，这对大模型未来的实际应用具有重要意义。

**目标：**

1）源代码：原型技术demo及其技术源代码；

2）申请发明专利至少2项；

3）产出CCF-A类论文至少1篇。

4）性能指标：

模型精度方面，所提方法精度与已有联邦模型微调方法（如lora、p-tuning、offset-tuning等）的精度差距小于1%。

安全性方面，能够同时保护参与方的训练数据和模型，能够抵御常见攻击

高效性方面，所提方法训练过程的通信量、通信时长至少与已有方法保持在相同量级。

[返回首页](#目录)

**4. 面向量化神经网络的双方安全推理**

**背景：**量化神经网络是一种对神经网络模型进行压缩和量化的技术，可以在保持相对较高的模型准确性的同时，大大减少模型的大小和计算量，有助于在资源受限的设备上部署。将量化神经网络和加密计算结合起来，可以在保护数据隐私的同时，实现高效的量化神经网络计算。从这一背景出发，神经网络的量化和对应的多方安全计算的算子加速，成为一个亟待解决的技术难题。为此，需要加强同高校的合作，完成相关研究的探索。

**目标：**

1）源代码：技术demo和基于隐语的开源代码；

2）申请发明专利至少1项；

3）CCF-A学术论文至少1篇。

[返回首页](#目录)

**5. 数据使用权跨域管控技术研究**

**背景：**当数据要素作为一种要素，在主体之间流通时，即使它不在持有者的运维域内，持有者仍然能够对其使用进行有效管控，这需要数据使用权跨域管控。

在数据要素流通交易中，只有在保障数据资产的持有权不失控，同时又能实现使用权的跨域管控的情况下，数据要素的经营才能得到保障，相关的提供方、加工处理方、运维方也能各自获得收益。本课题的主要研究内容包括：如何从技术上将数据持有权与使用权分离，实现数据使用权跨域管控，以确保数据流转的全程安全可控。

**目标：**

1）源代码：原型技术demo及其技术源代码；
2）申请发明专利至少3项；
3）产出CCF-A类论文1篇。

[返回首页](#目录)

**6. 软硬件结合多方安全计算技术研究**

**背景：**多方安全计算（MPC）技术是一种经典的密码学算法技术，常见技术方案包括秘密分享、不经意传输、混淆电路等。MPC在隐私计算中有广泛的应用，构成了诸多隐私计算产品中的主要算法协议功能。但是在实际场景中，MPC存在计算量、通信量、安全假设的取舍问题，很难设计一个计算量小、通信量小且安全模型具备普适性的MPC算法。通过软硬件结合的方式，可对MPC算法的纯软件实现进行优化，从而提升方案的普适性，让隐私计算技术能推广到更大的数据集、更高标准的安全场景等业务中。可能的软硬件结合方案包括基于密码硬件、TEE等拓展MPC的安全假设；利用FPGA、GPU等设备进行计算加速；亦或上述方案的综合。因此，需要与高校协作，设计更快、更安全的MPC软硬件结合方案。

**目标：**

1）源代码：原型技术demo及其技术源代码（包括相关文档），源代码需在蚂蚁提供的硬件环境中跑通；

3）协助蚂蚁申请发明专利至少2项；

4）产出CCF-A类论文至少1篇。

[返回首页](#目录)

**7. GPU加速TFHE/TFHW同态加密方案及方案转换技术研究**

**背景：**目前实现数据隐私保护主流的同态加密方案包括CKKS、BGV、BFV、TFHE等。

其中CKKS、BFV、BGV等方案支持整数或实数的密文运算，适合实现线性函数，

而TFHE、FHEW等方案支持位操作，适合实现非线形函数。近几年，学术届不断有论文中提出了两种方案的加速方案及方案间的转换技术，使得加密方案可以同时支持线性函数和非线性函数，从而实现大模型密文推理等更大规模的密文应用。利用GPU等异构运算平台加速FHE算法是学术界、产业界探索的重要方向。通过软硬件结合方案充分利用GPU上的计算与存储资源，将显著提升密文操作及方案转换的运算效率，从而将FHE技术应用到更广泛的业务场景中。据此背景出发，所需的研究方向是利用GPU加速TFHE/FHEW方案中的同态操作以及TFHE/FHEW方案与CKKS方案间的转换。为此，需要加强与高校的技术合作，完成相关领域的探索。

**目标：**

1）源代码：原型技术demo及其技术源代码（包括相关文档），源代码需在蚂蚁提供的硬件环境中跑通；

3）协助蚂蚁申请发明专利至少2项；

4）产出CCF-A类论文至少1篇。

[返回首页](#目录)

**8. 基于CKKS和TFHE切换技术的密文AI推理加速器设计**

**背景：**隐私保护是目前AI中比较重要的问题，很多场景下客户不愿将私人数据发送到服务器执行深度学习任务，全同态加密支持对密文进行任意深度布尔电路或者算术电路处理，并将执行的操作同态地映射到明文上。CKKS方案在实现AI非卷积层，比如Sigmoid、Relu、Maxpooling等，多采用逼近替代的方案，需要频繁执行Bootstrapping，造成计算量大大增加。探讨CKKS和TFHE结合对密文AI推理进行加速，采用CKKS和TFHE切换的方式来提高计算效率。 设计针对密文AI推理的加速器。

**目标：**完成可以支持CKKS和TFHE切换的密文AI推理加速器设计开发和原型验证，加速器AI密文推理性能和PPA指标达到业内领先水平

1）源代码：加速器原型RTL源代码；

2）仿真代码：基于加速器原型，采用CKKS、TFHE实现Resnet-20推理仿真代码，推荐采用C语言，不依赖第三方库；

3）申请发明专利至少1项；

4）产出CCF-A类论文1篇。

[返回首页](#目录)

**9. 宽松恶意安全的多方安全计算**

**背景：**多方安全计算MPC是隐私计算的核心技术路线之一，在工业界有着广泛的应用。不少MPC协议的恶意安全版本的性能远远低于半诚实安全版本，无法满足实际应用的性能要求，而半诚实安全版本面对malicious参与方时存在泄露数据的风险。从这一背景出发，需要研发介于恶意安全和半诚实安全的安全模型下的MPC协议，在对malicious参与方保证数据隐私同时仍保持较高性能。（这样的安全模型包括但不限于：malicious privacy[1]、one-bit leakage security[2]。）为此，需要加强同高校的合作，完成相关研究的探索。

[1] ELSA: Secure Aggregation for Federated Learning with Malicious Actors. S&P 2023

[2] Going Beyond Dual Execution: MPC for Functions with Efficient Verification. PKC 2020

**目标：**

1） 宽松恶意安全MPC协议设计文档；

2） 源代码：原型技术demo及其技术源代码；

3）申请发明专利至少1项；

4）产出CCF-A类论文1篇。

[返回首页](#目录)

**10. 隐私增强的跨机构分布式图计算**

**背景：**PageRank、社区发现、标签扩散等图算法广泛应用于资金图、社交图等图数据的分析，在反欺诈、反洗钱、团伙识别等应用中起到了重要的作用。图数据通常分散到多家机构中，一家机构只有部分的图数据。要获得更好的分析结果，需要多家机构联合起来进行分布式的图计算。由于图数据涉及隐私和商业机密、跨机构带宽低，需要采用技术手段增强分布式图计算的隐私性、提升性能；研究较充分的全密态的安全图计算无法满足大规模图计算的性能要求。从这一背景出发，研发减少中间信息交换和暴露的分布式图算法，将分布式图算法结合差分隐私、多方安全计算、数据脱敏等隐私保护技术，成为一个亟待解决的技术难题。为此，需要加强同高校的合作，完成相关研究的探索。

**目标：**

1）隐私增强的跨机构分布式图计算算法设计文档；

2）源代码：原型技术demo及其技术源代码；

3）申请发明专利至少1项；

4）产出CCF-A类论文1篇。

[返回首页](#目录)

**11. 通用群模型（GGM）、代数群模型（AGM）下分布式签名及零知识证明安全性研究**

**背景：**分布式签名、zk-SNARKs密码技术是构建区块链的重要基础设施，它们分别是MPC钱包、二层网络扩容最核心的技术，其安全性对于交易执行和验证至关重要。为了论证这些密码方案的可证明安全性，业界分别提出了不同的安全论证模型，例如，分布式签名中的门限ECDSA方案在通用群模型（GGM）下论证，主流zk-SNARKs方案，如Plonk、Marlin等在代数群模型（AGM）下进行证明。一般来说，这些模型都有其适用条件和范围，其合理性尚处于研究阶段。近期学术界有工作证明了方案在AGM模型下存在安全性归约，并不意味着其在GGM模型下一定安全。因此，论证这些方案在AGM模型以及GGM模型下的安全性，并研讨是否存在特定攻击影响安全性，是一个有实际意义的课题。为此，需要加强高校合作，完成相关研究。

**目标：**

1） 门限ECDSA方案更高级别安全性证明（UC），落地MPC钱包

2） 现有主流zk-SNARKs方案证明合理性分析报告；

3） AGM、GGM模型使用行业等级规范；

4） 产出CCF-A类论文1篇。

[返回首页](#目录)

**12. 联营生态中个性化联邦学习相关技术**

**背景：**在联营生态中，对人群（如：年轻客群、新市民）的信贷成长是各联营机构共同关心的命题。信贷成长从用户心愿和生命周期需求出发，结合信贷记录、自证他证、行为习惯、金融素养信息，实现用户在场景开放和额度储蓄上的成长。实践过程中，联合任务推荐是获取用户需求和现状，加速用户成长的主要手段。然而由于联营主体和联营机构在运营偏好、用户分布、系统性能上存在差异，并且对信贷记录等核心数据有隐私保护需求，从而需要一套隐私保护下的个性化建模方案。近年来，个性化联邦学习通过中心模型与用户模型之间的协作实现模型个性化，其中联营主体与联营机构在设置多目标、特征空间和样本空间混合交叉使得个性化建模变得困难，同时混合交叉和通讯延迟带来收敛性不足也需要解决。因此，需要加强与高校的合作，完成相关研究探索。

**目标：**

1）源代码：联营业务运营个性化推荐模型；

2）POC报告：商业场景下的分析与POC报告；

3）申请发明专利至少3项；

4）产出CCF-A类论文1篇。

[返回首页](#目录)

**B类课题**

**1. 拆分学习攻防算法研究**

**背景：**拆分学习是一种解决数据垂直拆分场景下跨域联合建模的方案。由于其简单灵活的架构和高效的训练速度，拆分学习在联合营销、联合风控等领域具有广泛的应用潜力。然而，由于拆分学习需要传递隐层和梯度信息，其安全性经常面临挑战。为了分析和加固拆分学习的安全性，需要对拆分学习的攻防算法做系统性的研究。

本项目旨在对业界垂直拆分学习攻击或者防御方法做系统性的整理调研，在隐语框架中实现SOTA的攻击或防御算法并评估效果。

**目标：**

1）拆分学习攻防算法整体报告

2）4类以上符合隐语代码规范的攻防算法实现代码

3）攻防算法的实验报告

[返回首页](#目录)

**2. 基于 GC 的双方 SHA256 以及 AES 协议实现**

**背景：**在隐私计算中，有许多C端场景需要 MPC对于传统密码套件的支持：例如AES、Hash、HMAC 等，而 GC 的相较于其他框架的性能更优。目前隐语计算引擎缺少对传统密码套件算法的支持。

**目标：**

1）根据已有的顶会论文完成半诚实 GC 框架的设计、实现工作，可仅支持针对整数的 Sha256 以及 AES 计算

2.）（可选）实现恶意 GC 框架

**产出：**

1）源代码

2）性能报告

3）发明专利（可选）

[返回首页](#目录)

**3. 基于 GPU 的加法同态加密算法设计与实现**

**背景：**Paillier、Okamoto-Uchiyama（下文简称 OU）等同态加密算法已经在隐私计算领域大量应用，但是相比 FL、MPC等其它隐私计算方案，同态加密的计算性能非常低，是隐私计算大规模应用的瓶颈。虽然隐语也尝试过用 FPGA 等硬件加速 Paillier 并取得一定效果，但是这类方案与硬件型号存在绑定关系，一旦FPGA 型号变化则算法不再适用。GPU 作为一种通用硬件，安装使用门槛低，普及面好，基于 GPU 加速同态加密算法可以使更多用户受益。

**目标：**

1） 完成基于 GPU 加速的 Paillier、OU 算法设计和实现工作，支持主流商用、民用 GPU 卡，并且算法功能正确，无明显安全漏洞。

2） GPU Paillier、OU 的性能要明显优于 CPU 版本，期望 GPU 可以提升5-10倍性能

3） 多 GPU 卡协同：如果机器上安装多张 GPU 卡，可以充分利用 GPU 资源起到协同加速效果（可选）

**课题产出：**

1） 源代码：本课题对应的源代码

2） 性能报告：Paillier、OU 算法在主流 GPU 卡上的性能表现

3） 发明专利：课题创新技术申请专利（可选）

[返回首页](#目录)

**4. 可信执行环境的状态连续性设计和实现**

**背景：**可信执行环境（TEE）已经成为机密计算的基石，是隐私保护计算的重要技术路线之一。TEE提供了程序和敏感数据的机密性和完整性保护，可以保护数据使用中（）data-in-use）安全。但是有状态的TEE应用可能遭受回滚攻击（rollback attach）和分叉攻击（forking attack），比如应用存储在外存的数据虽然是加密的，但是依然可能被回滚至旧版本，因此如何保证TEE状态连续性是TEE在工业落地中面临的实际问题。业界目前存在一些方案，比如Intel SGX monotonic counters、TPM Counter等，但是存在着写入时延高、写入次数有限等缺陷。实现一个高性能、低成本的工业级TEE状态连续性系统成为一个待解决的技术难题，因此希望和高校共同合作，完成相关探索和实现。

**目标：**

1）系统性能要求：写时延<=10ms，状态更新>=5000次每秒。

2）方案应该可落地，部署简洁无复杂依赖（比如不依赖区块链）。

**产出：**

1）源代码和文档：课题对应的源代码和技术文档，需贡献至隐语开源社区。

2）性能测试报告：读写时延、状态更新性能。

3）（可选）课题创新技术申请专利1项。

[返回首页](#目录)

**5. 隐私计算节点间网络可观测性度量的研究**

**背景：**隐私计算是一种跨域计算场景，节点间的网络质量直接影响隐私计算任务的成功率和执行效率。网络可观测性对于网络问题的预警和诊断至关重要。由于不同节点网络结构的差异性和复杂性，传统基于节点网关总体流量的质量度量，已经不能满足生产运维的需求，我们需要一种能够有效度量节点之间端到端的网络可观测性的方案。

本项目旨在分析隐私计算节点之间网络故障诊断的难点，并给出针对性的网络可观测性的度量方案。

**目标与产出：**

1）隐私计算节点间网络可观测性度量的整体报告

2）隐私计算节点间网络可观测性度量的方案和代码

3）隐私计算节点间网络可观测性度量的实验报告

[返回首页](#目录)