**目录**

**一、应用方向**

1. [基于大模型的跨域场景下的联合表征建模研究](#在当前的人工智能领域)

2. [民生场景下的多轮交互设置下的大模型插件能力研究](#ChatGPT等大规模预训练模型)

3. [端侧大模型图像实时生成技术研究](#AIGC大模型和多模态通用感知大模型加速)

4. [大模型在金融领域的推理能力探索及推理过程的白盒化呈现](#通过机器辅助金融分析，从过去的计量金融到量化分析)

5. [基于医疗大模型与临床数据的专病诊疗与靶点发现](#支付宝健康频道在疫情后)

**二、通用模型**

1. [大模型知识增强和事实正确性研究](#大模型在问答对话和逻辑推理等方面都展现出了强大的能力)

2. [基于小样本的高效LLM蒸馏学习](#LLM在大量任务上都取得了非常好的效果)

3. [高性能类Transformer模型架构研究](#随着大模型技术的快速发展)

4. [基于多模态大模型的零样本视觉任务研究](#优异的效果在很大程度上源于通过图像)

5. [结合图学习和智能体的仓库级代码生成大模型研究](#随着大模型在软件开发的广泛应用)

6. [语音合成大模型基础研究](#随着人工智能技术的不断发展)

7. [3D生成基础模型研究](#近年来，深度学习技术的突破性进展)

8. [大语言模型与知识图谱联合建模与应用](#大语言模型已经被证明其基于自然语言强大的世界理解和推理能力)

9. [大模型预训练可控推理算法研究](#伴随海量数据信息和大模型独特的涌现能力)

10. [大模型超级对齐关键技术研究](#大模型已经展现出强大的能力，但是如何确保模型输出可靠)

**三、大模型Infra**

1. [多模态大模型向量检索索引技术研究](#在大模型技术广泛应用和数据规模持续增长的背景下)

2. [大模型推理极致优化：性能、成本与敏捷迭代](#在大规模部署时面临性能)

3. [面向万卡级训练作业提效的异构高性能AIDC软硬件协同优化研究](#大模型大规模分布式训练，对系统的性能，稳定性)

**四、数据方向**

1. [大模型训练数据品质度量与效果评估方法](#对于大模型的训练，需要海量的数据输入)

2. [基于大语言模型的数据分析 Copilot 应用研究](#随着蚂蚁业务发展，用户对数据分析的需求呈快速增长的趋势)

**五、安全方向**

1. [基于预训练大模型和污点分析的业务漏洞智能化检测技术研究](#业务逻辑漏洞的自动化检测是业内长期以来的一个难题)

2. [LLM基于检索增强的安全可控生成技术专项研究](#ChatGPT等大语言模型具有较强的文本生成能力)

3. [大模型安全可信对抗攻击](#当前语言生成大模型取得了一系列突破性的进展)

4. [安全高效的大模型隐私保护技术研究](#大模型应用涉及到用户数据的处理)

5. [基于可控大模型数据生成技术研究](#大模型在训练过程中对数据的数量和质量有着极高的要求)

**一、应用方向**

**1. 基于大模型的跨域场景下的联合表征建模研究**

**背景：**在当前的人工智能领域，大型预训练模型已经成为研究和应用的热点，它们因其卓越的学习能力和广泛的适用性受到了极大的关注。在支付宝平台上，搜索广告推荐系统扮演着连接商户与消费者之间的重要桥梁。随着支付宝服务场景的不断丰富和用户行为数据的积累，不同领域间的知识迁移和综合利用成为一个挑战。由于各领域数据的分布差异性和特征表征的多样性，如何通过联合表征学习来消除领域间的鸿沟，结合和提炼跨域间的通用语义与知识，是一项颇具挑战的研究工作。本项目旨在探索基于大模型的跨域场景下的联合表征建模。通过研究和开发新的算法与框架，使大模型能够更好地理解并结合来自不同领域的信息，建立起一个统一且高效的知识表征系统，在多域知识融合、跨领域迁移学习等应用中表现出较强的能力。实现对支付宝用户行为和偏好的深层次理解，同时充分考虑不同场景下的用户需求，为用户提供更为精准、相关和个性化的内容。

**目标：**

1）提出大模型跨域联合表征建模算法，提升搜广推相关的业务模型综合效果5%以上；

2）发表2篇CCF-A类论文；

3）申请专利1项以上。

[返回主页](#目录)

**2. 民生场景下的多轮交互设置下的大模型插件能力研究**

**背景：**ChatGPT等大规模预训练模型（LLM）在完成各种NLP通用任务上展现出了强大的文本理解、推理能力。然而，受限于大模型本身的静态知识，在一些特定领域的任务上，仍然存在一定的局限性（例如大模型无法进行天气查询）。大模型新范式工具学习将一些信息查询类、操作类、计算类等能力封装成API（例如ChatGPT plugins），让LLM来编排API调用并消费API的结果，充分结合LLM的理解推理能力和垂直领域的原子服务来完成复杂的任务。 在实际业务场景中，我们发现用户很少将一个复杂的任务用一句话描述清楚，很多时候用户需要通过多轮交互的方式对需求进行补充或更正，在交互过程中可能需要机器人对用户进行反问或引导。而目前大模型API编排和应用在多轮对话和主动引导方面的能力还有所欠缺。

因此我们需要对多轮交互设置下，大模型如何更好的结合API完成任务进行深入的研究和探索。

**目标：**

1）NL2API数据集，多轮交互增强的大模型代码和参数；

2）发表CCFA类论文1篇，申请1项技术专利；

3）多轮交互大模型的线上服务，并带来业务效果提升。

[返回主页](#目录)

**3.端侧大模型图像实时生成技术研究**

**背景：**AIGC大模型和多模态通用感知大模型加速，及部署到端侧是一个不可忽视的趋势。这种趋势的背后有多个原因：实时性和低延迟的需求，隐私和数据安全的考量，离线能力的重要性以及提供更好的用户体验和个性化。通过将大模型和多模态通用感知大模型推向端侧，可以实现实时的推理和响应，增强数据的隐私和安全性，提供离线运行能力，并为用户提供个性化的服务。以扩散模型为例，基于StableDiffusion（SD）的图像合成技术需要多步推理来实现，对端侧计算资源和耗时提出了挑战，尽管业界已经提出了许加速方法，但它们为减少计算量也通常牺牲了生成质量，并且对不同的模型的适配增加了额外的训练成本。

同时，随着最近AIGC视频生成技术的发展，目前的视频生成算法耗时过长，严重影响算法的商业化应用。

为了解决这些限制，我们将从以下几个方面进行进一步研究：

1） 知识蒸馏。提出了一种新穎且通用的SD加速模块，使其可以作为插件引入各种SD微调模型中，无需额外训练的情况下提升10倍～20倍的推理速度。

2） 网络结构优化。针对端侧算力设计更加轻0级的网络结构，以减少模型的参数量和计算复杂度。采用深度可分离卷积、轻量级注意力机制等技术，来设计可以满足端侧实时推理的AIGC网络。

3） 模型压缩和量化。通过模型压缩和0化技术，减少A℃C模型的参数0和计筒复杂度。可以使用剪枝、稀疏化、低比特量化等方法来进一步减少模型的大小和计算需求，并在一定程度上保持模型的精度。

4） 插件式的加速方案在LLM、多模态通用感知大模型上的拓展验证探索。

5） 通用AIGC视频加速方案。

**目标：**

1）发表CCFA类论文2篇；

2）申请2项技术专利；

3）提供实时加速模块在蚂蚁内部用户中使用，在至少1个业务场景落地上线。

[返回主页](#目录)

**4.大模型在金融领域的推理能力探索及推理过程的白盒化呈现**

**背景：**通过机器辅助金融分析，从过去的计量金融到量化分析，都是业内一直在实践和探索的方向。但金融市场可被数值化，可被数学建模的内容是相对有限的，大量的主观决策仍依赖人类专家完成。过去曾经有基于知识图谱进行机器主观决策的尝试，但收效甚微。在大模型技术爆发后，其在语言理解、推理上的能力，暗示其在主观分析领域有极大的潜力：金融主观分析，本质上是分析师通过语言描述的逻辑，在大量的文字洪流中提取关键信息的过程。而这正是大语言模型擅长之处。在实践过程中，大模型确实表现出一定的效果，但还有一些必须要解决的问题，如控制模型遵从人类专家的推理过程、在推理过程中保持强金融逻辑、对于推理过程进行人类可理解可验证的白盒化呈现等。

**目标：**

1）金融类解读分析盲评，胜率70%超kimi、文心等竞品；

2）财富业务专业金融内容gc占比50%以上，解读内容采纳率超70%；

3）CCFA类会议论文1～2篇。

[返回主页](#目录)

**5. 基于医疗大模型与临床数据的专病诊疗与靶点发现**

**背景：**支付宝健康频道在疫情后，越来越多用户持续使用支付宝。目前来看，用户(患者)有很多长周期疾病，包括慢性病、精神类疾病，是需要很长时间的观察随访，并需要规律性筛查和必要治疗。传统的医疗专病诊疗方法通常基于患者主动就医后医生的经验和少量的临床数据，缺乏长周期、全面的、个体化的评估。

近年来，基于大模型的深度学习在医疗领域取得了显著的进展，该课题将研究如何充分利用有限的临床数据来构建医疗专病诊疗大模型，利用大规模的临床数据来分析和挖掘医疗专病的相关特征和靶点，以实现精准的及早洞察、筛查、以及辅助诊断和治疗。提高医疗专病的防治水平，改善数亿患者的生活质量。

**目标：**

1）研发医疗专科诊疗大模型，诊断正确率对比行业大模型高50%；

2）在蚂蚁至少2个业务场景落地上线；

3）发表CCFA类论文1～2篇或影响因子10以上的跨学科医疗期刊；

4）申请2项专利。

[返回主页](#目录)

**二、通用模型**

**1.大模型知识增强和事实正确性研究**

**背景：**大模型在问答对话和逻辑推理等方面都展现出了强大的能力，但是目前大模型仍然存在比较严重的幻觉问题，难以保证生成内容的事实正确，特别是对于一些时效性较强或专业性较强的领域，大模型的输出更加无法获得用户的信任，而信任作为支付宝的核心用户心智，这些问题会严重阻碍大模型在实际场景中的落地和效果达成。

大模型的真实性问题可以从事实真实性（factualness）和事实一致性（faithfulness）两个方面进行增强，一方面通过不同的方式给大模型注入事实知识，加强大模型对正确知识的认知，从而提高模型产出的事实准确性，另一方面，如何通过包括强化学习在内的各种训练方式提升大模型底座真实性，需要通过反馈技术调整大模型输出分布，基于内部知识做到知之为知之，不知为不知，降低模型的幻觉。课题涉及到如何构建跟真实性相关的奖励数据以及收集人工的反馈数据，但目前业界在这方面的标准以及数据很有限，尤其中文的更少。目标是形成一套解决模型真实性方案，降低目前大模型基座幻觉，可通过但不局限大模型真实度校准、真实性奖励模型、基于检索的多奖励模型建设等技术。

**目标：**

1）大模型知识评估相关数据集；

2）发表CCFA类论文2篇，申请2项技术专利；

3）知识增强的通用方案、文档、代码、框架实现等，沉淀一套大模型知识增强的工具链，方便开展实验进行迭代，将相应精调，蒸馏等能力解耦并平台化，增强可复用性。

[返回主页](#目录)

**2.基于小样本的高效LLM蒸馏学习**

**背景：**LLM在大量任务上都取得了非常好的效果，但是存在推理性能瓶颈。模型压缩，比如量化、稀疏化、剪枝、张量分解等, 可以降低一部分参数空间，进而提高推理速度，但是依然存在较大的计算负载。考虑到业界的大模型应用场景，往往只需要特定领域的知识或者能力，因此可以通过蒸馏大模型，得到一个在具体领域表现良好的小模型，大幅提升计算效率。另外在实际业务场景中，领域特定的样本获取成本较高，因此探索基于小样本的高效蒸馏算法具有重要意义。

**目标：**

1）结合蚂蚁大模型，提升蒸馏模型的压缩比；

2）在蚂蚁至少2个业务场景落地上线；

3）发表CCFA类论文2篇；

4）申请1项技术专利。

[返回主页](#目录)

**3.高性能类Transformer模型架构研究**

**背景：**随着大模型技术的快速发展，大量的应用形式开始涌现。多模信息输入/输出（如图像、语音等）、实时交互，是这些新型应用中关键特性以及持续强化的趋势。但是大模型庞大的参数规模，以及Transformer结构在高分辨率图像、长文本序列上遇到的巨大计算挑战，成为多模态大模型应用广泛落地的技术卡点。从以上的应用卡点中可以推导其中的关键技术需求：

1）突破Transformer计算瓶颈，探索类Transformer架构，同时能够与当前大模型的技术及演进无缝兼容；

2）轻量级、兼容异构计算环境，支撑在web端、手机端高效率运行，以及高分辨率图像、长序列文本输入。

**目标：**

1）发表CCFA类论文1～2篇，申请2项技术专利；

2）在手机端运行的原型系统。

[返回主页](#目录)

**4.基于多模态大模型的零样本视觉任务研究**

**背景：**多模态大模型(MLLM)优异的效果在很大程度上源于通过图像到文本表达空间的映射，实现图像模态和文本语义空间的对齐，进而复用LLM对于文本的强大零样本理解能力实现对图文的理解。这种方法为图像解译提供了新的视角和思路，尤其对开放集合下的视觉任务。本项目旨在研究通过图文模态语义关联，实现多模态大模型对开放集合下的零样本视觉任务支持， 构建形式统一的常见的检测、分割等视觉任务表示， 并提高MLLM模型在不同视觉任务上的泛化性。

**目标：**

1）发表CCFA类论文1～2篇，申请2项技术专利；

2）以多模态基础模型支撑的多视觉任务的原型系统。

[返回主页](#目录)

**5.结合图学习和智能体的仓库级代码生成大模型研究**

**背景：**随着大模型在软件开发的广泛应用，代码生成的自动化进程不断加速，尤其在处理具有庞大代码库的仓库级别应用方面扮演了关键角色。在这种规模的应用中，代码生成模型需要处理并整合复杂的应用逻辑、库依赖和系统架构等信息，这增加了对生成代码准确性的要求。与现有自然语言处理模型不同的是，仓库级别的代码生成对模型的精度和可靠性要求极高，小的错误可能引发链式反应，导致整个编译或运行环境失败。另一方面，程序分析技术能够从代码中生成丰富的图结构信息，包括但不限于调用图、AST（抽象语法树）、CFG（控制流图）、DFG（数据流图），这些图表示可以更好地捕捉代码间的依赖关系和逻辑结构。然而如何将这些图结构信息与代码大模型融合，从而在复杂的软件开发项目中提高代码生成的准确率和执行效率，仍是一个尚未充分解决的问题。图表征学习、图搜索技术，甚至结合检索式的增强学习（RAG)、以及测试和执行工具反馈的智能体为解决这一问题提供了新的方法论。集成这些技术的研究可以提升代码生成模型对语义的理解深度，并增强它们在大规模项目中的适用性。本项目拟从利用先进的图学习技术、智能体交互技术，通过更准确的模型预测输出，减少人为干预，提升端到端需求生成体验，有助于加速大规模软件项目的研发进程。

**目标：**

1）交付代码原型/模型等，集成当前的代码大模型，产出有效性的测试报告；

2）发表CCF-A/B类或者蚂蚁认可的论文1+篇；

3）申请1+项技术专利。

[返回主页](#目录)

**6.语音合成大模型基础研究**

**背景：**随着人工智能技术的不断发展，语音合成技术已经被广泛应用于各个领域，如智能客服、语音助手等。高质量、高自然度的语音合成技术可以实现更加智能、更加自然的人机交互，为用户提供更好的使用体验。传统的语音合成技术在语音音质、自然程度等方面存在局限性，并且依赖大量数据训练，定制音色的成本较高；而近期流行的语音克隆技术，虽然降低了对数据的依赖，但会降低合成语音的效果。语音合成大模型一方面能大幅减少定制新音色依赖的数据量，降低业务定制音色的成本，甚至提供UGC的语音合成能力；另一方面可以提升语音合成的效果，包括音质、自然程度等，是语音领域最新的研究方向，对语音合成任务有很大的提升。

主要解决的问题：构建大规模语音数据集，结合最新的论文和开源项目研发适合的模型结构和训练方法，通过研究声纹、情感等方面的技术提高语音合成的自然度和拟真度，探索大模型高效部署和线上服务的系统方案。目前我们已完成单人高质量的语音合成系统研发，下一步的目标是：构建高质量、多样化的语音合成系统，支持多说话人音色合成的同时增加one-shot音色克隆的功能。具体而言，希望只需要提供少量的目标说话人语音样本数据，就能复刻出目标说话人的音色和说话风格。生成高质量的合成语音，实现自然流畅的语音交互，一直都是我们追求的目标。目前所在团队既研究声音防伪检测，又研究语音合成，攻防一体。多样化、高质量的合成语音是AIGC时代的必备基础设施，本研究课题无论是在技术上还是应用上都具有重要意义。

**目标：**

1）整理好的TTS训练、测试数据集，语音合成大模型的代码和模型参数；

2）可使用的TTS服务引擎；

3）发表CCFA类论文至少1篇，申请至少1项技术专利；

4）语音合成大模型的线上服务。

[返回主页](#目录)

**7. 3D生成基础模型研究**

**背景：**近年来，深度学习技术的突破性进展，特别是GAN、VAE、Diffusion等模型在图像、视频合成领域的成功应用，为3D内容生成提供了全新的思路与方法。与此同时，图形学、计算机视觉等多学科交叉融合，催生出如NeRF、3D-GS等一系列前沿工作。尽管如此，现有技术在3D内容生成方面尚不成熟，在以下几个方面仍存在较大提升空间：

- 生成效果：现有技术方案生成的三维模型在几何、贴图、结构、可视化效果等方面还达不到美术人员的产出效果，离工业应用还有差距。

- 跨模态学习与融合：如何有效利用多模态数据（如2D图像、文本描述、语音指令等）进行3D生成，实现跨模态信息的有效整合与利用，尚未形成系统化的理论框架与实践路径。

- 模型解释性与可控性：现有的3D生成模型往往被视为“黑箱”，缺乏对生成过程的清晰理解与有效控制，这在需要严格遵守设计规范或伦理准则的应用场景中尤为重要。

- 数据的采集与高效处理：相较图像、文本数据，可获取的3D数据量不大，但数据格式、标准不统一，处理难度较大。如何设计高效的数据采集、规范化、压缩、索引、检索策略，以及适应分布式计算环境的并行化训练方法，是提升3D生成模型实用性的关键问题。

综上所述，本课题“3D生成基础模型研究”旨在顺应技术发展趋势，响应产业变革需求，填补学术研究空白，通过与高校开展深度合作，共同攻克上述难题，构建一套先进、普适、智能的3D生成基础模型体系，为相关行业的创新发展提供强大支撑。

**目标：**

1）整理3D数据集，研发高质量的3D生成模型，重点解决某一个方向的问题，达到SOTA效果；

2）CCF-A类论文1篇、专利1篇及对应的代码和模型。

[返回主页](#目录)

**8. 大语言模型与知识图谱联合建模与应用**

**背景：**大语言模型已经被证明其基于自然语言强大的世界理解和推理能力，然而其广泛应用还存在3个主要问题：1）inference慢、消耗资源大；2）可靠力差：其生成模式不可控、距离部署到真实工业生产环境仍不现实；3）知识力差：缺乏精细化知识供给的条件下，单纯大模型推理导致结果缺乏知识性、逻辑性且冗余。知识图谱作为当前主流的可靠知识库可以有效缓解以上问题，然而知识图谱的问题在于其高人力成本和推理泛化性缺失。

目前，蚂蚁支付场景如实体图谱扮演了对齐跨域知识的角色，同时金融场景如信贷、理财、保险均包含大量专业金融知识。如何结合这些动态、专业知识对业务具有极大价值。

如何利用知识图谱增强大语言模型的可靠力以及知识力。例如：如何为大语言模型提供强大的context信息，如何利用知识图谱引导大语言模型做上层概念逻辑推理，如何让大语言模型利用图谱结构，以及结合图谱中的事实知识帮助推理生成，从而有效支持知识生成可控、避免幻觉、提升推理逻辑性和知识性、避免高成本预训练，具有重要的研究意义。

同时，大语言模型通过已有知识图谱、对应schema甚至web知识，进而提升知识图谱的质量，两者如何结合彼此增强，也具备重要研究价值。

**目标：**

1）论文与专利：发表CCFA类论文2篇，申请2项技术专利；

2）通过知识图谱增强大模型基座的知识和推理能力，核心榜单提升5%。

[返回主页](#目录)

**9.大模型预训练可控推理算法研究**

**背景：**伴随海量数据信息和大模型独特的涌现能力，预训练大模型在语言、视觉、跨模态数据合成上取得了突破性进展，引发了广泛的关注。尽管有着海量的数据知识积累，预训练大模型依旧显示出对于分布外样本的脆弱性，其推理过程的可控性仍有待提高。现有模型主要通过自然语言和用户交互，很难精细控制数据的生成过程，不免造成匹配度低、效果差的问题，甚至产生违反道德、乃至法律的结果，是制约相关模型进一步落地的关键问题。

本项目聚焦于可控概率推理基础理论与算法，从能量指引、分布外样本检测、对抗鲁棒等技术手段出发，解决大模型预训练及微调过程中的可控推理问题，并在蚂蚁相关场景验证，包括：1）提升在微贷业务中大模型的预测精度，提升微贷业务上大模型风控能力，提升高风险用户识别、高风险交易识别，减少资损；2）提高在数字生活业务上广告、搜索等场景转化效率，提升广告业务CPM、搜索业务PVCTR。

**目标：**

1）发表CCFA类论文至少2篇；

2）在蚂蚁相关场景验证所提可控生成理论与算法的有效性。部署可控生成算法，提升有关大模型可控生成、与人类意图对齐的能力，提升大模型应用与人类感知的一致性；

3）聚焦于可控概率推理基础理论与算法，提出大模型可控生成的新颖解决思路(目标提升可控生成相关性能～4%)，实现并大模型可控生成技术工具包1～2项。

[返回主页](#目录)

**10. 大模型超级对齐关键技术研究**

**背景：**大模型已经展现出强大的能力，但是如何确保模型输出可靠、可控、可信的内容，避免造成对人类的危害，这需要对模型进行与人类的对齐操作，而随着模型能力的增强，模型在部分场景中已经表现出了远超人类的能力，如何能够保证这些超过人类能力的模型能够遵从人类的指令，并且表现出与人类一致的行为表现，是目前在大模型研究领域一直在探索的方向。大模型对齐中常见的工作主要还是通过人类来构造大规模的高质量监督数据来进行，比如使用基于人类反馈的强化学习（RLHF）进行训练，目前这些方法存在一定的瓶颈，包括依赖大量人类的标注从而可扩展性较差，还有人类对于超级智能的监督是否一直可靠等。

本课题是期待解决蚂蚁自研大模型在安全可信领域的技术瓶颈。大模型的自主研发符合蚂蚁的技术战略，而大模型对齐是大模型自主研发上必不可少的一环。本次立项着重于解决大模型对齐过程中的问题，解决项目背景中提到的目前的技术瓶颈，为大模型安全性中涉及到的伦理域的风险问题、指令攻击类型问题、如何平衡harmless、helpful、honest，带来技术突破和技术创新。

**目标：**

1）顶级学术会议（CCF-A类/阿里A类）论文2篇；

2）安全性的公开数据集一个；

3）超级对齐系统。

[返回主页](#目录)

**三、大模型Infra**

**1.多模态大模型向量检索索引技术研究**

**背景：**在大模型技术广泛应用和数据规模持续增长的背景下，向量数据库已成为AI技术链路的核心环节。在多模态的交互过程中，文本、图片数据都需要向量化（embedding）保存在向量数据库中。如果大模型是AI时代的计算核心、向量数据库就是大模型的存储核心。

在蚂蚁的多模态应用场景中，如大模型对话机器人、多媒体内容检索、安全信息查询和生物特征认证等，对向量数据库的需求日益增长。

与此同时向量检索领域还有很多问题尚待解决，比如

1）向量索引的高内存消耗问题；

2）如何有效利用硬件加速提升索引的构建和查询时间；

3）如何有效的将标量与向量进行混合索引；

4）如何根据不同场景的特征自适应选择最优的近邻图算法；

5）如何通过索引结构实现索引合并，以解决类似LSM-Tree存储引擎合并带来的构建损耗。

本研究项目旨在通过优化向量索引技术，结合蚂蚁现有的存储基础设施，研发新一代的向量数据库产品，提升蚂蚁在向量数据库以及大模型基础设施上的技术实力。

**目标：**

1）发表CCFA类论文至少1篇，申请至少1项技术专利；

2）对比当前的开源算法和产品能够实现20%以上的整体性能提升或者相同程度的成本优化。

[返回主页](#目录)

**2.大模型推理极致优化：性能、成本与敏捷迭代**

**背景：**AIGC大模型（LLM，stable diffusion模型等）在大规模部署时面临性能、成本与敏捷迭代等多方面的挑战。例如：

1）性能：在总吞吐和RT方面能否足接近经典DNN，达到10X性能提升？

2）成本：能否在典型推理卡（包括国产卡）例如A10级或更廉价的卡上运行百千亿参数模型，以节省千万级别成本；

3）敏捷：考虑到多元化的异构算力（包括国产卡），提升模型优化开发效率。

围绕以上挑战，本课题将探索以下领域方法，包括但不局限于：

1）轻量化或减少迭代的方法：包括量化、剪枝、投机采样等，或模型创新（例如Mamba等）。

2）高效算子：针对核心算子，优化性能或资源占用等。

3）高效编译：动态适配多种加速卡（包括国产GPU）配置和运行时环境。

4）高效分布式调度与运行时：包括batch、抢占策略、高效的缓存设计等。

5）超长序列优化：高效支持序列长度>100K。

6）其他可显著提升推理性能（降低成本）的机制和方法，包括CPU/GPU协同优化等

**目标：**

1）典型大模型（语言或多模）实测有显著的性能提升；

2）CCF-A类论文1～2篇；

3）2项技术专利。

[返回主页](#目录)

**3.面向万卡级训练作业提效的异构高性能AIDC软硬件协同优化研究**

**背景：**大模型大规模分布式训练，对系统的性能，稳定性，迭代效率提出了多方面的挑战：如何有针对万亿级参数在万卡以上规模上还能发挥高性能的算力吞吐，是业界一个非常有挑战的课题，尤其是在中美竞争日趋激烈的今天，我们还需要综合考虑众多国产加速硬件的协同适配和使用。为了达到这个目标，我们需要在软件层解决，通信，显存，计算等多方面的瓶颈以及大规模组网效率和稳定性；AIDC建设方面也需要设计更高效的网络拓扑链接，存储性能以及计算效率来最大化在软件层释放给用户更强大的算力。研究方向包含但不限于通讯感知的全图AI编译优化，分布式策略自动优化，大规模先进AIDC建设等方向。

**目标：**

1）研究成果有效支撑蚂蚁内部万亿级参数，万卡规模大模型训练微调加速；

2）结合DLRover基于上述研究成果打造蚂蚁和高校双方在业界的技术影响力；

3）发表CCFA类论文至少1～2篇，沉淀专利1～2项。

[返回主页](#目录)

**四、数据方向**

**1.大模型训练数据品质度量与效果评估方法**

**背景：**对于大模型的训练，需要海量的数据输入，以LLM大模型为例，当前蚂蚁以采集互联网公开数据超过PB级别，但是网络数据的品质往往参差不齐，实际可用于模型训练的不足10%，并且数据处理方式多样，处理过程成本较高，最终数据对于模型训练的效果影响不能很精准的度量。整个过程存在巨大的采集成本和训练成本的浪费，如果能够建立对于数据集质量的评估标准，并且通过模型评测效果，来动态调整数据集处理配置和数据集数据采集策略，来持续的提升数据集质量，那么对模型训练将带来极大的效果提升和成本节省。

**目标：**

1）数据质量提升方案（评估方法，处理手段，包含但不限于训练数据处理、数据集refine、数据评估方法）；

2）方案带来大模型效果提升20%；

3）发表CCFA类论文1～2篇；

4）申请2项技术专利。

[返回主页](#目录)

**2.基于大语言模型的数据分析 Copilot 应用研究**

**背景：**随着蚂蚁业务发展，用户对数据分析的需求呈快速增长的趋势，但目前用户在数据分析过程中遇到的产品门槛高的问题，阻碍了数据分析产品的普及以及用户获取数据进行决策的效率。

近期以 ChatGPT 和 Microsoft Copilot 为典型的基于大语言模型（LLM）的产品，极大地拉高了用户对数据分析产品的期待，而目前数据分析产品的产品使用门槛以及对大语言模型（LLM）应用的现状，跟用户的期待之间存在巨大的 GAP。因此，为了降低数据分析产品的门槛，提高用户进行数据分析的效率，我们需要在继续升级完善产品功能覆盖率、提升智能分析能力、优化产品体验的同时，结合大语言模型（LLM），研发更加智能、高效、易用的工具，通过自然语言进一步降低用户的取数分析的门槛，从而进一步提升用户数据驱动决策的效率。

此前，DeepInsight 已经基于自然语言查询（NLQ）等场景进行了大量积累，一旦有大模型的支持，有望形成数据分析产品的领先地位。

**目标：**

1）新的数据分析产品形态，在蚂蚁内部用户中使用；

2）产出数据分析领域的大模型，提升自然语言分析的准确率，期望超越业界的SOTA；

3）发表论文1～2篇，专利1～2个。以及对应的源代码/原型系统；

4）有条件的情况下，带来1～2个专业人才的转化引进。

[返回主页](#目录)

**五、安全方向**

**1.基于预训练大模型和污点分析的业务漏洞智能化检测技术研究**

**背景：**业务逻辑漏洞的自动化检测是业内长期以来的一个难题。过去主要利用程序分析技术去分析代码并通过漏洞特征规则进行漏洞的检测。如通过污点传播技术，我们可以分析变量之间的数据流向关系，从而实现对SQL注入、系统命令执行等安全技术漏洞问题的检测。而业务逻辑漏洞需要检测的漏洞特征包含了大量业务信息及语义，如注释语义、文档含义和函数变量命名约定的理解等，如水平权限漏洞需要识别某个接口方法是否需要进行鉴权及鉴权是否正确；敏感数据泄露漏洞需要识别某个变量是否包含敏感业务数据。

此外，由于污点分析技术本身存在的理论瓶颈以及工程实现上的限制，其分析结果会不可避免地出现较大的误报和漏报，需要大量人工投入对结果进行二次确认。虽然传统的安全工作流程、漏洞扫描技术在过去几十年内长足发展，是软件安全保证的基石，但它本质仍然是劳动密集型的。

因此，我们期望融合大模型到污点分析漏洞检测技术中，提升检测结果的准确性和有效性，在极大程度上解决业务语义强相关的业务逻辑漏洞的自动化检测问题。

**目标：**

1）发表CCFA类论文1～2篇；

2）申请1～2个技术专利；

3）业务逻辑相关漏洞召回率从35%提升至90%、准确率从20%提升至80%。

[返回主页](#目录)

**2. LLM基于检索增强的安全可控生成技术专项研究**

**背景：**ChatGPT等大语言模型具有较强的文本生成能力，其生成的文本流畅度高、结构完整、上下文连贯，其出色的表现有时甚至超越了人类水平。然而，大语言模型在事实准确性方面有时表现较差，特别是在具体安全场景中，当场景问题同预训练数据差异较大时，LLM输出可能会忽视事实并编造答案，出现所谓“幻觉”问题。在实际的工作应用中，我们希望可以借力LLM在文本生成方面的优秀表达能力，同时控制其生成内容的表达范围，使其可以更安全的应用于合规、安全咨询、智能交互等对于生成文本有相对严格规范要求的场景。

本项目旨在研究生成式大语言模型的可控性，通过知识增强等方式提升语言模型的事实性、逻辑性和文本质量。结合知识图谱、RAG（Retrieval-Augmented Generation）检索增强技术，进一步提高模型的知识推理能力和生成内容的可控性。通过高效融合企业内部积累的规则、政策、知识库以及外部的百科信息、舆情等开放资源，打造基于知识图谱的知识推理、可控文本生成和多样化文本输出方法，并在安全风险认知助理、反洗钱优质报文生成等真实场景做有效性验证。**目标：**

1） 发表CCF-A类论文1～2篇；

2） 专利发明2～4篇；

2） 在反洗钱报文生成、智能交互EDD以及安全风险智能咨询服务中落地应用。

[返回主页](#目录)

**3.大模型安全可信测评**

**背景：**当前语言生成大模型取得了一系列突破性的进展，在越来越多的复杂任务上取得了接近甚至于超越人类的性能表现，这使得开发自有的语言生成大模型、将语言生成大模型部署到各种各样的现实应用中的需求日渐增长，蚂蚁各业务部门也针对其业务场景开发了各类的产品大模型。然而，由于开发、部署、使用过程中的不当，以及恶意攻击者的存在，给语言生成大模型的安全可信运行带来了很大的挑战，亟需对语言生成大模型在实际运行条件下的安全可信进行全方面的评测，以提前发现模型的安全漏洞，为相应防御手段的开发提供重要基础。研究意义：以ChatGPT为代表的语言生成大模型近期取得了突破性的进展和广泛的应用，这也使其暴露在了充满风险的运行环境下，对其安全可信造成了极大的挑战。然而，目前针对先进的语言生成大模型的安全可信评测的研究仍处于起步阶段，也缺乏全面、有效的语言生成大模型安全可信评测平台。本项目拟从大模型开发-部署-应用的完整流程出发，分析各阶段中大模型所面临的安全可靠可控风险，并研究相应的安全评测方法，为构建全面、有效的语言生成大模型安全可信评测平台提供基础。

**目标：**

1）交付代码原型/模型等，并对目前运行的产品大模型，提供安全评测报告；2）发表CCFA类论文2篇；

3）申请2项技术专利。

[返回主页](#目录)

**4.安全高效的大模型隐私保护技术研究**

**背景：**大模型应用涉及到用户数据的处理，对这些涉及隐私的数据进行保护十分必要。考虑到大模型应用即时交互所要求的高性能，基于密码的方法在性能上力有不逮，所以一个直接的想法是将大模型系统跑在CPU/GPU上的可信执行环境中。但是，这么做是否足够？是否需要详细分析典型大模型系统在TEE中的隐私泄露风险，并基于此设计高效的安全机制？我们希望本课题能比较全面和深入地探索大模型的隐私泄露风险，并设计安全高效的大模型隐私保护技术。

**目标：**

1）原型系统代码；

2）发表CCFA类论文1篇。

[返回主页](#目录)

**5.基于可控大模型数据生成技术研究**

**背景：**大模型在训练过程中对数据的数量和质量有着极高的要求。据市场调研，后续大模型训练数据中，合成数据比例将发生指数级增长，最大可占比75%。

蚂蚁集团在大模型数据标注方面，一直在持续发力，在优秀的标注人员情况下，大模型训练数据仍然面临着一定的短缺情况。

在文章《Physics of Language Models - Knowledge Capacity Scaling Laws》充分展示了数据质量对大模型的影响。

这是因为深度学习模型通过从大量数据中学习特征和模式来实现其功能，而合成数据在这一过程中扮演着至关重要的角色。

合成数据具有非常好的一些性质：

数据稀缺性：对于一些特定任务或领域，获取真实的、标注好的训练数据可能非常困难和昂贵。合成数据提供了一种可行的解决方案，可以在没有足够真实数据的情况下进行模型训练。

数据多样性：即使在数据量足够的情况下，真实数据的样本多样性可能仍然有限。合成数据能够帮助提供更广泛的场景和变化，从而增强模型的泛化能力和健壮性。

隐私和安全：特别是在处理敏感信息（如个人身份信息、医疗记录等）时，使用合成数据可以避免直接暴露真实个人数据的风险，符合隐私保护的要求。

模拟罕见事件：对于一些罕见但重要的事件，通过合成数据生成相应的训练样本，可以帮助模型学习到这些情况下的处理策略，提高其在特定任务上的表现。

基于以上性质，合成数据如果在可控大模型的作用下进行生成，必须要保证以下条件：

数据准确性：高质量的数据意味着数据准确无误，能够真实反映预测或分类任务的实际情况。模型的性能直接受到训练数据准确性的影响。

数据一致性：数据需要具有一致性，以保证训练过程的稳定性。不一致的数据会导致模型无法学习到有效的特征和规律，影响训练效果。

数据代表性：训练数据需要能够代表实际应用中将要遇到的各种情况，以确保模型在现实世界中的有效性和泛化能力。质量差的数据可能会有偏见或无法涵盖重要的场景。

数据无噪声：高质量的数据应尽可能减少噪声干扰，因为数据中的噪声会误导模型学习错误的信息，降低模型的预测准确率。

所以针对可控大模型的数据合成，需要进行前沿技术研究。

**目标：**

1）发表1篇A类论文；

2）形成一个可以复用并稳定的可控生成技术框架；

3）大模型带来的合成数据产品竞争力提升，助力签约1+客户。

[返回主页](#目录)