附件一：申报主题

申报主题介绍

（各主题均不限于给定的建议研究方向，申请人可自行拓展决定。）

1. AI for Science

1.1 生物大模型的研究和应用

生物大模型的研究和应用是当前人工智能与生命科学交叉领域的热点。通过整合海量生物数据，如基因组和蛋白质序列，生物大模型能够揭示复杂的生物规律，推动精准医学、药物研发和合成生物学等领域的突破。随着数据量和计算能力的提升，生物大模型有望进一步推动生命科学的范式变革，为人类健康和生物技术发展带来深远影响。本课题旨在探索在生物信息学中的大模型算法和工具的研究。

建议研究方向：

1) 用于分析和挖掘多组学数据的大模型算法和工具；
2) 用于蛋白质结构预测和设计的大模型算法和工具；
3) 用于制药的多模态药物分子设计大模型算法和工具。

返回目录

1.2 AI 技术在二氧化碳地质封存领域的应用研究

二氧化碳地质封存是指通过前沿技术手段将二氧化碳注入深部地质储层，实现二氧化碳与大气长期隔绝的过程。在二氧化碳地质封存领域，地质构造仅在特定条件下可被选为储存场地，以确保不存在重大泄漏风险和重大环境或健康风险。通过深度学习等 AI 技术求解多相流等地质模型，可对封存选址决策、封存安全评估等关键问题进行预测预判。本课题旨在探索 AI 技术在二氧化碳地质封存领域的应用，基于腾讯与国家科研团队共同研发的二氧化碳封存可视化智能化平台孵化地质封存领域 AI 垂类应用模型。

建议研究方向：

1) 基于 AI 的二氧化碳地质封存反演与多元 CO_2流演化模型：混合数值模拟与深度神经网络，研究完全可微分的多相流偏微分方程神经求解器，精准预测二氧化碳地质封存的储层响应，并在小样本条件下实现外插与泛化。在低保真度参数场和有限监测数据的情况下，构建 AI 驱动的预测框架，实现多元 CO_2流体在地层中的精准运移模拟。结合多物理场耦合方法，为多元 CO_2流的长期封存安全性评估提供技术支撑；
2) CO_2泄露异常点空间推断及发展风险预测模型：根据脉冲测试压力扰动，构建 AI 预测模型，区分泄漏与非泄漏的压力响应，建立异常探测器解释 CO_2泄漏风险的指示指标；
3) 基于 AI 的二氧化碳地质封存盖层完整性预测：结合微地震监测数据与 AI 算法，对封存盖层的完整性进行评估与预测，构建封存盖层破裂预测模型，实现对封存过程盖层完整性的实时预警；

4) 基于 AI 的二氧化碳地质封存混合密度预测：用大量实验和监测的开放数据，结合多种流体影响因素，建立二氧化碳与咸水层流体的混合密度 AI 预测模型，实现混合密度预测模型动态更新；
5) 多源数据融合驱动的 CO_2地质封存动态优化：基于 AI 融合井下传感器、地震成像等多源异构实时数据，突破井下-地面-空天多源异构数据的融合难题；考虑地质不确定性条件下在线更新储层参数场，突破传统静态建模局限并准确预测羽流迁移；同步优化井网布局与注入方案，实现地质封存效率最优。

返回目录

2. 大模型及其应用研究

2.1 大语言模型在视频搜索个性化排序的应用研究

视频搜索场景具有三大典型特性：1）需求个性化显著，表现为视频搜索中一部分请求与用户前序消费视频相关，一部分与历史搜索有关，一部分与用户历史消费视频相关；2）视频内容核心语义位置分散，其核心语义表达除视频本体外，还分散在标题、封面 OCR、内容 OCR、ASR、评论等多个文本域，其中内容 OCR、ASR 和评论往往内容多、噪音多，需要精准提取有效信息；3）数据稀疏性问题突出，存在大量长尾搜索需求。本课题旨在通过大语言模型（LLM）优化搜索个性化排序算法，实现技术突破，提升搜索效果并推动应用落地。

建议研究方向：

1) 大模型表征：基于 LLM 的 Query、Doc、User 向量表征，更精准的建模用户、Query、视频内容，应用于排序特征；
2) 大模型排序：基于 LLM 的排序研究（可以应用在精排或重排阶段；可以在已有特征上模型结构改造，或端到端改造）。

返回目录

2.2 分布式 LLM 推理系统设计与关键技术研究

随着 DeepSeek R1 等千亿级大模型的突破性进展，模型性能提升与产业落地需求之间的矛盾日益凸显。超大参数规模导致推理时延（TTFT/TPOT）显著增加与机器资源成本指数级增长，已成为制约 AI 应用商业化的核心瓶颈。特别是在高并发搜索场景下，首 Token 响应延迟直接影响用户体验，持续 Token 生成速率关乎服务流畅度，而系统吞吐量则直接决定服务部署的经济可行性。本课题旨在面向搜索业务场景，构建高效分布式推理系统架构，提升大模型的吞吐以及性能。

建议研究方向：

1) PD 分离：为了降低 Prefill 和 Decode 相互影响，并且提升 Decode 的吞吐，探索合适的 PD 分离系统方案；
2) 全局负载均衡：设计高效的集群调度算法，使最大程度地提升推理并行性以提升整体的吞吐；

3) 通信并行优化：随着 MoE 架构逐步使用，除了 Prefill 的计算瓶颈以及 Decode 的内存带宽瓶颈，也带来了通信瓶颈。探索可能的通信优化方式，如何计算并行处理等，提升整体的吞吐能力；
4) 全局 KV Cache 管理系统：结合业务场景探索全局 Cache 系统的方案；
5) 投机采样：结合搜索场景探索有效的投机采样技术，提升推理速度和准确性。

返回目录

2.3 基于大语言模型长文本生成的幻觉抑制技术研究

近年来，基于大规模预训练语言模型，尤其是广泛采用的基于 Transformer 的预训练语言模型的方法，已成为自然语言生成的主流范式。依托强大的大型语言模型，我们能够生成更多样化、更流畅的文本。然而，“模型幻觉”问题仍然是大模型面临的一大挑战，特别是在处理长序列文本时，这一问题可能会更加严重，导致生成内容偏离真实情况。从学术和工业界的角度来看，模型的潜在幻觉问题将极大地限制其在实际应用场景中的使用，可能引发法律和伦理风险。本课题旨在探索视频剧本生成场景下幻觉控制问题，希望能借助大模型生成更高质量、更符合标准的剧本内容。

建议研究方向：

1) 训练/推理阶段的可控生成解幻研究：聚焦通过不同阶段技术实现可控文本生成(CTG)，解决深度模型可解释性不足的挑战，以适配在实际应用中的特定约束和生成更高质量的文本数据；
2) 基于信息检索增强的解幻方法研究：针对信息检索增强技术在长文本生成中的局限性（如上下文理解缺失、知识幻觉加剧），探索新型检索增强方案，重点解决长文本剧本生成任务中语义连贯性与知识可靠性问题；
3) 基于知识编辑的模型解幻方法研究：针对模型参数中存储的错误知识或过时知识导致的幻觉放大问题，研究知识编辑（KE）技术，通过精准修改 LLM 参数以注入特定知识，同时最小化对无关知识的干扰，抑制长文本生成中的偏差累积效应。

返回目录

2.4 个性化全模态大模型研究

随着大模型技术的迅猛发展，支持文本、图片、视频和语音的全模态大模型成为一个新的趋势，并在视频理解、语音交互和具身智能等领域展现出巨大的应用潜力。然而，现有的全模态大模型仍面临诸多挑战，例如难以实现用户个性化适配和人机交互实时性差等。本课题旨在通过融合文本生成、语音识别、视频理解等多模态能力，构建一个“千人千面”的个性化全模态大模型，使其能够根据用户的偏好、情感和需求进行动态调整和持续学习，并通过实时视频-语音交互等自然的人机交互方式提供个性化的智能服务。

建议研究方向：

1) 视频理解：探索高效的视频理解方案，增强模型的细节感知及长视频理解能力，确保模型能够快速准确地提取关键信息并理解复杂场景；
2) 视频交互：探索视频、语音等实时交互大模型方案，使其能够支持多模态输入和输出，保证系统在动态交互中的响应速度和自然性；

3) 个性化大模型：探索具备根据用户画像和历史交互进行个性化适配的大模型方案，并通过持续学习技术不断动态优化模型；
4) 具身交互：探索全模态技术与具身智能结合的方案，提升环境感知、决策和交互能力，赋能机器人、智能眼镜等设备，增强其智能化水平。

返回目录

2.5 基于多模态大模型的图像理解研究

当前多模态大模型虽已广泛应用于图片标注、描述生成等场景，但在垂类、内容领域（如冷兵器、人物套装等专业内容）的密集描述（Dense Caption）任务中仍存在准确性不足、幻觉现象严重等问题。提升图像的密集描述能力具有多重价值：1）在生成领域，精准的密集描述可显著增强图像生成模型的可控性与质量，是突破 AIGC 技术瓶颈的关键；2）在内容管理方面，可为人物道具、皮肤等资产构建统一的细粒度画像，支撑个性化推荐系统，可用密集描述取代 Clip 特征，服务于信息流和视频的推荐；3）在交互编辑场景，实现基于对话的图像编辑（如 Gemini 2.0 Flash 所示范）需以精准的图像理解为前提；4）在图像检索领域，可通过密集描述实现文本到图像的跨模态细粒度检索；5）在内容审核环节，密集描述能进一步增强低质、敏感内容的识别能力。基于此，本课题旨在构建垂类多模态大模型系统，重点突破专业内容的密集描述技术，并集成问答、分割、生成与编辑等能力，形成端到端的解决方案。

建议研究方向：

1) 领域自适应微调：针对人物皮肤、武器等专业内容，优化多模态大模型对 3D 模型部件级特征的密集描述能力；
2) 人类偏好知识的利用：运用强化学习、偏好学习等方法提升模型在游戏原画、3D 模型渲染图理解上的表现；
3) 多专家系统集成：集成 OCR、分割、姿态检测、人脸、光照估计等能力；联合优化图像 RAG，增强模型在垂类、专业领域能力表现。

返回目录

2.6 医学大模型中的若干技术研究

医疗领域是大模型技术应用的理想场景。这主要基于两方面优势：1）二十年医疗信息化建设沉淀了海量数据资源；2）丰富的医疗场景为大模型落地提供了多元机会。正因如此，医疗大模型研究已成为学界关注热点。本课题旨在深度探索大模型技术在医学领域的创新应用与关键技术问题突破。

建议研究方向：

1) 深度思考能力：构建具备深度思考能力的大模型，能够在诊断、用药、检查推荐等方面输出思考过程。模型应根据实际问题复杂性，动态调整思考深度，确保思考内容符合循证逻辑；
2) 医学幻觉问题：与通用领域相比，医学领域对幻觉的容忍度较低。降低在各种医学任务中产生幻觉的可能性是一个重要的研究课题。可能的研究方向包括结合已有的医学知识，强化学习和优化解码器等方法；

3) Agent 研究：包括单一用途的医学 Agent（例如家庭医生 Agent、辅助诊疗 Agent、院外管理 Agent）以及多 Agent 协同（例如通过不同科室的 Agent 合作完成会诊任务）；
4) 医学大模型的训练过程优化：探索如何更好地优化 Post-Pretrain、SFT、DPO、GRPO 和 PPO 中的数据和算法，以提升模型的能力；
5) 多模态研究：将医学图像、视频、音频和文本等多模态信息集成到大语言模型（LLM）中，并探索其在医疗人工智能生成内容（AIGC）方面的应用。

返回目录

2.7 微信刷掌支付前沿 AI 研究

微信刷掌支付是微信支付推出的一种新型支付手段，于 23 年正式发布。通过微信刷掌支付自研的 AI 算法和软硬件，用户可以像使用扫码一样通过自己的手掌满足购物、出行、就餐、开锁等日常需求。目前微信刷掌支付已在国内北京大兴地铁、深圳大学、国家奥体中心、7-11 便利店等场景落地，凭借其在便捷、隐私、可靠等方面的体验优势获得了用户的广泛好评。微信刷掌支付未来的目标是服务全国乃至全世界的用户，需要解决包括部分人群使用体验差、高隐私安全要求下数据获取/问题分析困难、假体攻击、硬件成本高等问题。本课题旨在通过 AI 算法研究，提升微信刷掌支付应对上述问题的能力。

建议研究方向：

1) 多模态模型在刷掌支付中的应用，包括数据生成、判别、AIGC 判断等；
2) 识别系统问题自动分析 Agent 建设研究；
3) 动作交互研究（连续姿态定位、动作意图识别、生成等方向）；
4) 多线索活体判别研究（例如材质、静脉、心跳等线索）；
5) 更低计算算力研究：低算力模型结构设计(低 bit、NAS 等)。

返回目录

2.8 大模型训练框架设计与关键技术研究

随着 Deepseek 等千亿级大模型的效果取得突破性进展，模型训练的复杂性和资源需求也在不断增加。包括预训练、微调和强化学习等不同阶段，同时更长序列长度的训练需求也日益凸显，为提升训练效率和稳定性，亟需探索新的训练系统架构和关键技术。本课题旨在面向大模型训练场景，构建高效训练系统架构，提升大模型的训练效率和性能。

建议研究方向：

1) 长序列长度支持：探索适用于更长序列长度的训练方法和系统架构，解决长序列训练中的内存和计算瓶颈，提升训练效率和模型性能；
2) FP8 训练：研究 FP8 训练技术，降低训练过程中的计算和内存开销，提升训练效率并保证效果；
3) MoE 训练性能优化：对 MoE 部分的专家通信和分配等策略等进行调研优化，提升训练速度效率；
4) 强化学习-资源调度：设计高效的资源调度策略，优化强化学习训练过程中的资源分配和使用，提升训练效率；

5) 强化学习-通信性能优化：优化强化学习训练过程中各模型之间参数同步和信息传输，重点优化通信性能。

返回目录

2.9 高性能大语言模型训练优化

大语言模型的突破深刻改变了我们的工作模式与生活方式，然而它的预训练仍面临着高成本、低效率等问题的挑战。本课题旨在确保模型效果不降级的前提下，探索通过模型结构创新和模型蒸馏等技术提升大语言模型预训练的效率、降低训练成本。

建议研究方向：

1) 稀疏注意力机制探索：探索稀疏注意力机制的新方案，加快训练速度，增强其对长文本的处理能力，为大规模语言模型的实际应用提供更高效的支持；
2) 大语言模型的融合架构研究：在模型参数不降级的情况下，通过多种先进架构的融合，提升训练效率，同时提升大模型的理解和推理能力；
3) 大语言模型蒸馏技术的研究：探索同词表和跨词表的大语言模型预训练知识蒸馏技术，提升大语言模型预训练的收敛速度和效果。

返回目录

2.10大模型架构探索及高效推理研究

大模型的超强能力在各项任务中被广为验证。随着大模型的普及与其在业务中的广泛应用，高效经济的大模型推理模型逐渐吸引了更多的研究兴趣。本课题旨在探索大模型新架构及高效推理方案，希望能够在模型性能损失尽可能小的情况下，最大化提升大模型推理效率。

建议研究方向：

1) 大模型新架构探索及优化，如（混合）Mamba 架构优化、MoE 结构和策略优化等；
2) 注意力机制改进，包括线性注意力算法研究等；
3) KV Cache 压缩技术及量化技术研究；
4) 大模型基础架构优化，如激活层、归一化层等的优化。

返回目录

2.11更高效、表达能力更强的大语言模型

Transformer 架构重塑了语言、语音、视觉等领域的研究范式。然而，一些研究表明，Transformer 架构在定理证明、搜索等任务上仍面临挑战；在处理超长上下文时 Transformer 架构会遇到计算效率问题。本课题旨在构建下一代模型架构，在表达能力及计算效率两个方面进行更有效的平衡。

建议研究方向：

1) 高效模型架构：研发高效架构，在保持与 Transformer 架构性能相当的前提下，计算成本有显著降低；

2) 表达能力更强的模型架构：设计新架构在效果和效率方面均超越 Transformer 架构，如，如何更好结合 Memory、多模态信息等。

返回目录

2.12原生多模态大模型探索

现阶段主流的多模态大模型架构更多是以语言大模型（LLM）为重心，桥接挂载图片&视频、语音&音频等其它模态。这种方式虽具备一定的多模态能力，但仍存在以下关键局限：1）模型的多模态能力是在 LLM 后叠加的，难以实现对真实物理世界的细粒度建模、推理与预测；2）模型是静态架构，一旦训练完成便不能从与环境的交互之中学习新的知识和能力；3）推理和记忆等能力都封装在同一套参数里，导致记忆难以动态更新，会造成严重的遗忘问题。本课题旨在探索原生多模态大模型方向，从训练初期即同步发展语言、音频与视觉能力，重点打造模型在以下方面的原生能力：真实物理世界理解、多模态复杂推理与自适应持续学习。

建议研究方向：

1) 原生多模态模型架构：研究新的模型架构和训练方法，从初始阶段即实现多模态融合（Fusing），在多模态任务上超越当前以 LLM 为核心的建模方式；
2) 时空建模能力：与当前以 LLM 为重心、更偏 Chatbot 的模型不同，原生多模态模型具备更强的对时间-空间进行建模的能力，从而能够处理、重构、理解和预测多模态数据（包括 2D+时间和 3D+时间的音视频），探索具有真实世界理解、多模态推理和自适应持续学习的大模型；
3) 多模态复杂推理能力：复杂推理能力是大模型最基本的能力，须具备多模态理解、多模态形式的中间复杂推理以及答案生成能力；
4) 多模态编码与表征学习：设计适用于音频、图像、视频等多模态数据的连续或离散表示方式；
5) 长期记忆能力：探索多模态模型如何建立、更新与调用长期记忆，以支持持续学习与历史知识积累。

返回目录

3. 强化学习及 AI Agent

3.1 高效的强化学习模型迁移问题研究

强化学习模型在智能决策、机器人控制、Game AI 等领域的应用展现出巨大的技术潜力。然而，现有的强化学习算法仍然有计算成本高、样本利用率低等瓶颈，制约了强化学习在真实场景下低成本应用的可能性。在真实业务场景中，模型被应用的环境会快速更迭，可能是环境版本的迭代，也可能是相似任务的不同环境，如何能够高效地训练模型是一个非常严峻的挑战。本课题旨在聚焦于强化学习模型在实际业务领域中的跨场景迁移能力的优化。以实际业务场景作为研究的模拟环境，重点研究多智能体在协同场景下知识迁移的机制以及轻量化的迁移框架构建，提升模型在跨任务环境中的训练效率与决策可靠性。

建议研究方向：

1) 相似任务不同环境的强化学习模型迁移（例如：同一任务环境的不同数值版本或者相似任务的不同环境）；
2) 强化学习技能在不同复杂任务中的复用（例如：同一能力用在不同的任务中）；
3) 宏观决策知识的蒸馏与共享（例如：复杂宏观决策的时机）。

返回目录

3.2 智能体场景下的决策模型预训练和后训练

近年来，基于海量数据和算力训练的大语言模型展现出令人印象深刻的通用问题解决能力。大语言模型的训练通常经历两个训练流程：基于海量数据的预训练和基于少量样本的后训练。相比之下，智能体场景下的决策模型，由于其传统训练方法（基于特定场景的模仿学习或强化学习）和模型表达能力的限制，缺乏不同任务和场景的迁移能力。在这个背景下，本课题旨在探索更高效的决策模型架构和训练方法，使其展现出类似大语言模型的通用问题解决能力，适用于不同场景和任务、上下文学习、推理等。

建议研究方向：

1) 智能体场景下的决策模型预训练；
2) 智能体场景下的决策模型后训练；
3) 智能体场景下的决策模型人类偏好对齐；
4) 智能体场景下的决策模型策略多样性生成。

返回目录

3.3 面向测试及探索的 Game AI 及智能体

Game AI 及智能体一直是业界关注的研究方向，近年来大语言模型的发展为相关方向的研究带来了更多可能。在这个课题中，我们并不优先关注 Game AI 的对抗及竞技能力，而是希望结合多模态大模型的能力，面向测试和体验，探索其在通用操作、探索能力、小成本训练以及与人类交互等方面的技术边界。本课题旨在引入视觉信息作为学习及交互的模态，实现对用户行为的理解。并通过基座模型、定制化小模型的结合，以及智能体技术、结合感知、规划、行动、验证等机制，使 Game AI 具备更泛化的规则理解、自主决策和实时交互能力。让 AI 不仅能进行对局，还能测试漏洞，并能像人类用户一样推进任务、体验产品，甚至参与到复杂任务策略的制定中。最终实现面向测试、产品使用而设计的多模态智能体。

建议研究方向：

1) 基于大模型的智能体：探索通过蒸馏或与小模型、行为树等传统方案的结合，实现具备实时操作能力的 Game AI；
2) 探索性 Agent：研究智能体在任务执行，世界探索方面的能力；
3) 高效的强化学习或模仿学习研究：可利用包括视频、操作等信息，实现游戏规则理解、玩家行为模仿，从而实现基础的操作和任务执行能力；
4) GUI Agent：针对 GUI 界面的互动的研究。研究内容可包括但不限于任务规划、意图识别、记忆、工具使用等方向的探索，推进智能体的通用操作能力。

返回目录

3.4 面向金融场景的多 Agent 自学习自演进前沿技术研究

受益于 LLM 的强大能力，基于 LLM 构建的 Agent 在多领域应用中表现出色。但术业有专攻，金融不同场景专业性强且要求不同、部分任务复杂度高，单一 Agent 难以满足要求，通过将复杂的金融任务分解为多个子任务（如行情分析、财报解读），并利用不同 Agent 的专业能力分工协作，可有效改善问题解决效果。与此同时，金融场景较高的动态性和不确定性（如行情、重大事件）给 Agent 效果稳定性带来了新的挑战，基于强化学习的自进化能够帮助 Agent 在未知环境中优化行为策略，实现长期自治学习。但减少 Agent 之间的冲突，实现多 Agent 的协同进化仍具有挑战。本课题旨在探索金融场景下多 Agent 协作的最优范式，并通过强化学习实现 Agent 的持续自我进化，提升其在复杂金融任务中的表现。这不仅具有理论创新性，还对投研、投顾、投资等实际金融应用具有重要指导意义。

建议研究方向：

1) 探索金融场景多 Agent 协作与强化学习结合机制，达到自我进化目的，提升 Agent 在金融领域的适应性和泛化能力；
2) 金融场景 Agent 协作范式研究（多 Agent 如何协作）：探索金融场景下多 Agent 协作的最优架构与机制（角色分工、协作策略、通信协议等）；
3) 金融场景多 Agent 强化学习研究（多 Agent 如何进化）：探索在金融动态环境中多 Agent 的自我进化机制和基于强化学习的协同训练（奖励设计、优化策略、持续学习等）。

返回目录

3.5 基于多智能体高效协同的跨域业务体验动态推理研究

随着视频点播/直播服务的普及，业务体验（QoE，如卡顿率、首帧时延）直接影响用户的在线时长。传统 QoE 评估依赖客户端上报，但存在时间滞后导致观测不及时、采样粒度偏粗导致数据波动大等问题。本课题旨在研究多智能体协同推理方法，通过服务端可观测的 QoS 指标，实时动态推理客户端 QoE，实现跨域（服务端-网络-客户端）的协同动态监测与优化，支撑低卡顿、高速率的视频点播/直播服务。

建议研究方向：

1) 多智能体协同推理：服务端智能体动态监测、预测、评估网络系统 QoS（如 CDN 节点负载、带宽波动、拥塞信息等），客户端轻量化代理反馈关键 QoE 事件，通过联邦学习、知识蒸馏等技术手段，实现跨域数据融合，提高动态监测的准确性与可靠性；
2) QoS→QoE 因果推理模型：基于历史数据构建因果图（如“带宽骤降→缓冲区耗尽→卡顿”），梳理区分显性指标（如延迟）与隐性指标（如报文重传对卡顿的影响），通过在线学习和反事实分析等技术手段，建立动态推理引擎，指导传输控制。

返回目录

3.6 基于大语言模型推理能力的自主决策智能体研究

在浏览器使用场景中，用户在跨应用操作（如旅行规划需联动地图、内容搜索、服务供应商）、复杂任务处理（如科研文献综述自动检索、摘要、对比）和个性化服务编排等需求中，存在三大核心痛点：1）信息过载导致决策效率低下，用户需在数十个页面跳转中手动筛选信息；2）跨站点操作割裂，无法形成自动化工作流；3）个性化需求难以被静态功能模块满足。传统基于规则的系统在动态环境适应性和任务泛化能力上存在根本性局限。本课题旨在探索智能体在数字环境中的自主决策与持续学习能力突破，基于目标规划→执行→反馈→优化的闭环逻辑，结合大语言模型（LLM）及多模态感知技术，融合强化学习（RL）方法，覆盖个人设备与企业场景需求。推动智能体从“被动工具”向“生产力要素”发展，突破多模态感知与跨任务学习瓶颈。

建议研究方向：

1) 自主决策与多智能体协作：基于 LLM 的全局规划，将复杂目标分解为子任务序列，通过环境感知与交互动态反馈调整策略；多智能体系统采用角色分工，主智能体与领域专家智能体相互协作，完成交互链编排，共同完成同一任务；
2) 多模态数字环境感知与交互：打通浏览器、APP、操作系统等多源异构的多模态数据获取，将用户操作行为（如鼠标点击、手势滑动）转化为结构化指令，构建安全高性能的运行环境，使 LLM 模型具备从环境感知到行动执行的能力；
3) 持续学习与跨任务知识迁移：将闭环中积累的任务执行轨迹转化为结构化知识库，通过强化学习探索生成新监督数据集，从而不断提升任务成功率，同时完成更高维度的业务抽象，实现跨领域知识迁移与复用。

返回目录

4. 生成式人工智能技术研究

4.1 面向空间智能的高质量三维场景生成研究

随着 3D AIGC 技术发展，研究焦点已从单一物体的生成拓展到更为复杂的 3D 场景生成。高质量、可交互的 3D 场景在游戏设计、具身智能仿真、AR/VR 等领域中需求广泛，但复杂场景的构建费时费力。本课题旨在研发新一代面向空间智能的 3D 场景生成技术，可感知图像、视频中的三维空间信息，并在未知空间区进行推理、补全，实现高精度、可编辑、符合物理世界规律的 3D 场景生成。

建议研究方向：

1) 相机轨迹可控的视频生成：充分利用视频生成模型中蕴含的 3D 知识，实现输入相机视角生成对应场景视频；
2) 新视角合成的新架构研究：探索可泛化的新视角合成方法及相关 3D 表达；
3) 稀疏视角三维场景重建：输入单视角或稀疏视角，训练重建模型，生成 3D Gaussian Splatting 等表达；
4) 可物理交互的三维场景生成研究：将物理属性引入到三维场景生成中，实现生成的场景可进行光影、运动等物理仿真；

5) 多视图一致的深度估计与相机估计：针对多视图或视频输入训练模型，实现可靠的深度估计。

返回目录

4.2 多模态检索增强生成研究

在大模型时代，检索增强生成（RAG）在大模型落地到实际产品的环节中发挥着至关重要的作用。例如基于大模型+RAG 的 AI 搜索新形态大大提升了用户获取信息的效率和服务体验（如腾讯元宝-联网搜索）。除了全网信息外，RAG 也支持用户针对自己的知识库/数据库进行高效地提问和信息整理（如腾讯 Ima）。文本 RAG 算法趋于成熟，而多模态 RAG 仍是一个亟待研究的方向。具体而言，不同模态的数据（如视频、音频、图像，文本等）难以纳入统一的知识库体系中，同时多模态融合理解与推理的质量与效率仍有待提升。当前，OmniSearch、TaiChu-mRAG 等原型系统已进行了一些初步探索。本课题旨在探索更加高效高质且可应用的多模态 RAG 方案，在多模态检索及多模态理解生成上作出核心贡献。

建议研究方向：

1) 多模态对齐/检索研究：探索创新性的多模态对齐/检索方案，实现不同模态间的语义融合，支持多模态-多模态的检索，支持含推理能力的检索；利用表征学习、特征索引等技术改进检索效率，以适应大规模实时检索需求；
2) 面向通用/特定领域的多模态 RAG：面向通用或特定领域开发定制化多模态检索增强生成系统，减少大模型幻觉；针对多模态 RAG 返回的不同模态知识，设计训练有效的多模态大模型对上述信息进行理解和推理，最终预测出准确的答案。

返回目录

4.3 多模态统一模型推理增强的图像布局生成

在数字内容创作、广告设计、虚拟场景构建以及用户界面(UI)设计等众多实际业务场景中，根据文本描述或草图意图自动生成高质量，符合要求的图像布局具有巨大的应用价值和效率提升潜力。本课题旨在探索如何利用多模态理解生成统一模型的推理能力，以提升图像布局生成的效果，构建一个能够深度理解文本描述、视觉特征等多模态信息，并在此基础上进行有效推理的统一模型。通过增强模型的推理能力，实现对图像布局更精确、更可控的生成，从而提高图像编辑的可控性和可解释性。

建议研究方向：

1) 多模态推理机制研究：探索如何在统一模型中融合文本和图像等多模态信息，构建高效的推理机制。研究如何利用注意力机制、图神经网络等方法，实现多模态信息之间的深度交互和关联推理；
2) 图像布局生成的可控性与可解释性：研究如何利用多模态推理结果，实现对图像布局生成过程的精确控制。探索如何提取和可视化模型的推理过程，提高生成结果的可解释性；
3) 多模态统一模型训练与优化：研究如何构建大规模、高质量的多模态数据集，用于统一模型的训练。探索设计有效的损失函数和优化算法，提高模型的生成效果和泛化能力；
4) 多模态上下文感知图像生成:探索如何让模型理解图像中不同元素之间的关系，并且根据这些关系来生成更符合上下文的图像。

4.4 营销创意素材的可控生成研究和应用

近年来，人工智能生成内容（AIGC）技术通过多模态大模型与生成式算法的突破，正在快速渗透至数字营销领域以重构创意素材生产全链路。在增长营销场景中，素材创意质量与迭代效率对营销效果至关重要，而传统人工制作模式面临两大核心瓶颈：1）创意对个体经验高度依赖，易受知识边界限制，导致同质化问题；2）创意素材制作流程冗长，从策划到视觉设计实现跨多角色协作，时间与人力成本高。本课题旨在基于 AIGC 技术的智能化营销，实现营销创意过程的提效和增强，让人摆脱繁杂的执行实现，聚焦更高质量和个性化的创意灵感思考。

建议研究方向：

1) 基于 DeepResearch 技术的创意生成模型：结合 DeepResearch 技术架构的意图补足、任务拆解及 DeepSearch 优势和多样异构的知识库，研究创意生成这类复杂任务的知识引用和推理问题，产生更准确合理、创意多样的生成结果；
2) 主体细节保真可控的生成式图像编辑：在保证主体角色、商品细节高度一致条件下，研究通过 Prompt 控制局部精准编辑，多图交叉融合等图像编辑任务；
3) 主体细节保真可控的生成式视频编辑：在保真主体角色、商品细节高度一致条件下，研究通过 Prompt、Image 条件控制视频语义区域替换、风格改变等编辑任务。

返回目录

4.5 基于多模态大模型的 AIGC 生成内容识别技术研究

近年来，人工智能生成内容（AIGC）技术通过多模态大模型的赋能，实现了文本、图像、音频、视频等跨模态内容的自动化与智能化生产，深刻改变了内容创作、传播与消费的范式。而 2025 年是 AIGC 内容大爆发的一年，我们在互联网平台上获取的内容 AIGC 占比正在以肉眼可见的加速度倍增。本课题旨在探索如何更好的应用 AIGC 能力服务我们的社会、产品，用户。

建议研究方向：

1) AIGC 内容判别：可信生成式人工智能的多模态内容安全检测研究，高效准确地区分和识别 AIGC 生成内容和真实世界内容；
2) AIGC 安全风险管控：探索生成式人工智能安全风险管控方法，识别用户生成的风险内容，让 AIGC 的输出更加规范合规；
3) AIGC 生成意图判断：生成式人工智能的用户意图识别研究，结合模型上下文用户需求，以及前文生成的图片去精准识别用户的生成需求，以生成准确图片数据。

返回目录

4.6 基于大模型的音乐理解和生成研究

音乐能够跨越文化和地域的界限，触动人们的情感。然而，理解和生成音乐是一项极具挑战性的任务，因为音乐包含了丰富的结构、风格和情感元素。大模型如 Transformer 已经在语言、语音和视觉等领域取得了显著的成果，但在音乐理解和生成方面的研究仍然有待深

入。本课题旨在探索如何利用大模型来理解音乐的复杂结构，以及如何生成具有特定风格和情感的音乐。

建议研究方向：

1) 音乐理解的大模型：研究如何利用大模型来理解音乐的复杂结构，包括旋律、和声、节奏等元素，以及音乐的风格和情感；
2) 音乐生成的大模型：研究如何利用大模型来生成具有特定风格和情感的音乐，包括创作新的旋律、和声、节奏等，以及模仿或创新特定的音乐风格；
3) 多模态音乐理解和生成：研究如何结合音乐与其他模态的信息，如歌词、视觉元素等，以提升音乐理解和生成的效果和效率。

返回目录

4.7 高效可控实时视频生成技术研究

视频生成技术发展迅速，已从实验室研究快速迈向产业应用阶段，但在实际落地中仍存在诸多难点与挑战：1）稳定性与一致性：长视频易出现角色变化、背景偏移及运动失真等问题；2）可控性：视频生成模型多模态指令对齐与泛化有待提升；3）实时性与效率：视频生成模型推理时间久，成本高。本课题旨在开发高效、可控、实用的 AI 视频生成技术，这些问题对于推动视频生成技术的广泛应用具有重要意义。

建议研究方向：

1) 多模态与强化学习：研究不同模态间的新型融合方式，探索强化学习对生成策略的优化；
2) 高效计算：探索高效计算网络结构；
3) 设计视频检测工具，构造自查模块，优化视频生成物理逻辑。

返回目录

4.8 影视高品质的可控视频生成

视频生成技术近年来发展迅猛，将视频生成技术应用到影视、动漫等内容制作环节有巨大的商业价值，能够激发更多创作灵感与形式。然而，影视行业对于视频质量有着特殊要求。内容侧，生成的视频内容需要满足情节需求以及导演、编剧和创作者的镜头表达；视频质量侧，需要满足清晰度、画质、色彩、美学，物理规律等一系列标准。目前行业内的开源或者闭源商业化视频生成模型在这两方面仍然存在诸多问题。本课题旨在研究针对生成视频的质量评价方法，研究精细化的人机交互手段，指导生成高可控及高质量的视频，从内容和画质两方面满足影视制作等工业化需求。

建议研究方向：

1) 基于 2D/3D 镜头运镜的可控文/图生视频：结合前沿的基础视频生成基座模型进行方法探索，引入 2D/3D 镜头运镜控制，生成符合输入运镜运动模式的视频；
2) 基于文本、镜头运镜指令的视频编辑模型：结合文本、镜头运镜指令（2D/3D 空间下）的视频编辑探索，实现对既有视频的运镜、拍摄手法的修改；

3) AIGC 视频的质量评价与质量控制方法研究：针对文生视频/图生视频等主流 AIGC 方案形成质量评价方法，提供对于画质、美学、物理规律等方面的评价手段，指导生成模型的优化方向，提升生成内容的质量以及可用性；
4) 视频增强模型：基于当前视频生成模型，构建生成式视频增强模型，将定量化或可微视频评价体系等引入视频生成框架，提升视频生成结果（画质提升，分辨率 2K/4K 级别），满足时序一致性、稳定性等细节要求；
5) 高品质大运动下的视频插帧模型：探索高品质生成式视频插帧模型，满足主体、全局运动模式下的细粒度插帧，在满足高品质视频生成的同时提升视频生成效率。

返回目录

4.9 面向虚拟现实的实时神经双目帧生成算法研究

随着显示技术的发展,VR 设备以及裸眼 3D 显示器都为消费者提供了独特的沉浸式体验。然而，这种显示技术对渲染系统提出了严苛要求：为实现双视角连续成像，传统方案需并行渲染高分辨率视图，导致计算负载增长。尤其在动态场景中，渲染延迟与功耗问题严重制约了其在移动终端和实时交互场景中的应用。虽然已有一些实时超分辨率和帧外推的研究，但针对立体视觉设计的超分辨率和帧生成方法仍未被充分研究，直接应用过往方法仍会带来一定的采样和着色冗余。本课题旨在探索如何利用帧生成技术提高渲染效率，提高生成渲染画面的质量和稳定性，以及其在 VR 场景的应用落地。

建议研究方向：

1) 基于实时渲染的帧生成:通过深度挖掘已有的渲染管线中的信息，如 Depth，G-buffer 来生成双目渲染图像。可以结合不同的渲染管线构造，如前向渲染、延迟渲染管线等，利用渲染的时序冗余，设计不同的内插或者外插的帧生成算法；
2) 基于连续视频帧的帧生成:虽然渲染管线能够提供很多额外诸如 Depth，Motion Vector 信息，但是对于透明、特效、折射这些效果的帮助有限。可以利用视频生成的方案，探索图像空间的变换和生成算法，在保证时序稳定和双目一致性的情况下，利用单目渲染图像生成双目渲染结果。

返回目录

5. 安全与量子计算

5.1 基于溯源图分析的威胁狩猎技术研究

随着网络攻击的复杂性和隐蔽性不断提升,传统的威胁检测方法已难以应对高级持续性威胁（APT）等新型攻击。溯源图分析作为一种新兴技术，能够通过构建系统端上采集行为的因果关系图，有效还原攻击链并检测隐蔽威胁。本课题旨在结合入侵检测规则、终端安全分析、溯源分析和用户行为分析（UEBA），探索基于溯源图分析的威胁狩猎技术，提升对复杂攻击行为的识别能力，并实现实时、高效的威胁检测、调查与响应。

建议研究方向：

1) 终端安全分析中的溯源图构建与优化，探索低开销、高精度的系统行为追踪技术；

2) 基于溯源图 APT 攻击检测与场景还原，结合 MITRE ATT&CK 知识库框架，提升攻击链识别精度；
3) 基于图表征学习的威胁狩猎技术，利用图嵌入和图匹配算法，实现复杂攻击行为的自动化检测与分类；
4) 用户行为分析（UEBA）与溯源图融合，通过建模系统实体间的交互关系，识别潜在的异常威胁行为。

返回目录

5.2 基于 LLM 与全图风控融合的微信黑产团伙挖掘

微信作为亿级别用户量级的社交平台每天面临着巨大的安全挑战。如何准确的识别和打击黑产保障用户和平台安全是微信的重要命题。黑产通常以团伙的方式进行活动，其活动会涉及多个实体和复杂的关系网络，传统的分析方法难以处理这种复杂性。黑产团伙的活动具有高度的动态性和隐蔽性，同时也在不断变换手段和网络结构以逃避检测，还会涉及多种类型的数据。本课题旨在探索如何整合这些多源异构数据，为更全面的风险评估提供技术保障。

建议研究方向：

1) 多源异构数据图的构建：如何有效的整合数据源，更全面地描述黑产团伙的网络结构和感知黑产团伙的构建和变化；
2) 图算法与大模型的结合：如何利用大模型强化对于图结构和图数据的理解，优化团伙挖掘算法，更全面和实时的发现黑产团伙；
3) 模型的可解释性：如何利用大模型和图来提升模型的可解释性。

返回目录

5.3 基于密码学的 AI 智能体身份认证研究

得益于 AI 技术的不断提升，AI 智能体可以响应自然语言，通过协同多种形态的智能体执行复杂任务。智能体协同在提供便利的同时，也带来了安全挑战。智能体交互缺乏严格的身份认证机制，可能导致身份欺骗、未授权访问、人身伤亡及财产损失等问题。传统的身份认证机制基于知识因素（如口令）、拥有因素（如数字证书）和生物特征（如指纹、人脸）等，大多围绕人类用户和中心化平台设计，不完全适用于分布式协作场景下的智能体之间高效、自动化的身份验证需求。本课题旨在基于密码学技术，研究适用于智能体网络的身份认证系统，推动更加开放、协作、可信的智能体应用发展。

建议研究方向：

1) 探索基于密码学的智能体身份标识方法；
2) 设计智能体身份管理系统，实现智能体身份的分布式存储、验证和撤销；
3) 研究智能体之间的身份认证协议，支持多智能体的安全协作。

返回目录

5.4 基于深度思考的内容安全大模型研究

深度思考在代码、数学、科学、逻辑等领域上带来了显著的效果提升。然而，在文本内容安全领域，深度思考还没有得到合理应用。判断一段文本是否安全，需要经过一系列的思考过程。比如审核一篇新闻，需要首先粗读，理解新闻的主题；然后精读，挖掘其中潜在的有害信息，并按照标准进行比对和分析；最后，对上面的分析进行汇总，判断新闻的有害类型。本课题旨在将深度思考能力合理应用在文本内容安全场景上，一方面提升模型效果，一方面通过将深度思考内容提供给用户，提升模型的可解释性。

建议研究方向：

1) 探索深度思考大模型在内容安全审核场景中的应用效果和方式；
2) 在不影响模型效果的前提下，对深度思考内容进行压缩，并提升深度思考内容的可读性；
3) 探索使用深度思考的边界，比如模型能根据问题类型，动态决定是否启用深度思考能力。

返回目录

5.5 大模型时代全球人机识别与验证码对抗行为研究

在生成式大模型的推动下，行为模式模拟和操作指令生成展现出强大潜力，突破了传统自动化工具的局限，对业务安全中的人机识别机制提出了全新挑战。攻击者利用大模型生成高度拟人化的行为，使传统基于规则或单一特征的防护体系面临失效风险。防守方亟需构建涵盖理论模型、实时检测和动态对抗的全面防御体系。同时，验证码作为对抗 Bot 的核心手段，也面临多模态大模型带来的新威胁。大模型在图像生成与识别上的突破，使验证码攻防对抗的复杂度升级，需从理论、工程和策略层面建立动态对抗机制。本课题旨在聚焦于大模型驱动的 Bot 行为特征与验证码防御，研究其行为识别方法与对抗策略，并在验证码系统、API 防护等场景中进行实践验证。

建议研究方向：

1) 探索大模型在行为模式识别中的认知边界，特别是在时序特征和异常上下文关联等复杂行为模式中的薄弱环节；
2) 研究融合大模型语义理解的多级行为流量检测框架，结合预训练特征提取与对抗样本防御，实现对新型 Bot 流量的精准识别；
3) 研究多模态大模型在生成的时序行为解析技术，构建持续进化的对抗闭环；
4) 探讨多模态大模型在识别机器自动操作中的应用及检测方法；
5) 研究验证码防御中的图像混淆方式与识别资源开销的关系，探索高成本攻击的抑制策略。

返回目录

5.6 量子信息和量子计算理论研究

量子计算机在处理某些特定问题时，理论上速度要优于经典计算机，使得量子计算在密码学、化学、金融、机器学习等领域具有广泛的应用潜力。但目前量子计算机的硬件能力仍存在较大瓶颈，在比特门保真度和比特规模等方面都面临巨大挑战。为支撑远期应用落地，针对量子纠错方案和量子算法方面的开发工作，亦需要领域专家团队继续深耕和突破。本课

题旨在探索量子算法及复杂性理论、量子错误缓解与纠正和量子机器学习等量子计算相关的基础理论问题。

建议研究方向：

1) 量子算法与复杂性理论：研究量子信息中的基本问题，例如量子算法的效率和可行性，量子算法的优势和局限性，量子电路的优化等；
2) 量子错误缓解和量子编码理论：研究可用于实际的量子错误缓解，及更有效的量子纠错机制；
3) 量子机器学习与人工智能：结合量子计算和人工智能技术，研究新型量子机器学习算法，变分量子算法的分析与设计，以及可能的应用场景。

返回目录

6. 数据库及数据挖掘技术研究

6.1 基于 LLM 的数据库缺陷检测技术探索

数据库是存储和管理关键业务数据的核心系统，其稳定性直接关系到企业的业务连续性、数据的完整性和隐私保护。腾讯数据库产品技术发展迅速，市场增长迅猛，但是数据不一致、数据丢失和偶发性系统崩溃的质量问题日益凸显。例如，金融、电信等重点行业对数据库系统的使用广度和强度极高，对系统可靠性提出了极为严苛的要求，必须杜绝因数据丢失、不一致或质量风险导致的业务损失。近年来，随着大语言模型（LLM）技术的突破性进展，为数据库缺陷检测体系带来了范式革新。本课题旨在融合大语言模型（LLM）与程序分析技术，探索数据库内核缺陷的智能检测方法，发现系统潜在的稳定性问题、数据一致性和数据丢失等风险，保障业务稳定运行。

建议研究方向：

1) 智能驱动的模糊测试生成：通过强化学习与符号执行等技术生成符合 SQL 规范以及自研语法特性测试样本，并通过语义变异策略检测数据库逻辑缺陷；
2) 基于缺陷根因分析的测试生成：构建缺陷交互知识图谱定位缺陷模式，结合大模型生成测试样本并通过变异覆盖缺陷变种；
3) 基于代码分析的测试覆盖增强：基于代码流图分析识别潜在缺陷路径，利用大模型推理生成定向测试用例提升覆盖率。

返回目录

6.2 LLM 赋能的智能数据库系统研究

LLM for DB 技术旨在提升数据库系统的智能化水平、查询处理效率及自动化管理能力，通过自然语言交互、语义理解和智能决策增强数据库的易用性与性能。一方面，LLM 通过深度语义解析与生成技术，将自然语言查询转换为高效的数据库执行计划，结合提示工程优化复杂查询的执行路径（如查询重写、多表关联优化、索引推荐等），降低用户使用和优化数据库的门槛；另一方面，LLM 驱动的智能数据管理框架能够动态分析数据特征与负载模式，实现自适应配置管理（参数、索引、物化视图）、异常诊断与自愈，减少人工干预成本。本

课题旨在探索大规模数据库场景下的 LLM 赋能技术，聚焦智能查询优化、自治数据管理等关键问题。

建议研究方向：

1) LLM 智能查询优化：探索自然语言查询向 SQL 及高效执行计划的转换技术，将 LLM 应用于 NL2SQL、执行计划生成等环节，打造端到端查询调优工具链；
2) LLM 增强自治数据库管理：融合联邦学习等技术，搭建基于 LLM 的数据库自治管理系统，提升跨集群自适应管理能力；
3) 知识驱动 LLM 智能诊断：将 DBA 诊断经验转化为知识图谱，借助图谱搜索诊断路径，利用 LLM 生成诊断报告。

返回目录

6.3 基于机器学习的查询优化器前沿技术探索

查询优化器是数据库的核心组件，其产生的查询计划的质量直接决定了资源开销和查询时延。为提高计划准确性，优化器亟待解决基数估计不准确、计划错误不感知、搜索空间不完备等问题。近年来，基于机器学习以及 LLM 的技术取得了长远进步，应用到查询优化中有颇多优势：一方面，AI 模型具备强大的模式识别能力，能够充分利用统计信息和查询反馈，摆脱传统优化器中可能错误的规则与假设依赖，实现更精准的查询结果估计；另一方面，AI 优化器允许优化器借助模型的精准预测能力，能够在更大规模的计划集中进行预测与筛选工作，进行更完备的探索以找到最优执行计划。本课题旨在探索基于机器学习的查询优化器产生最优计划，研究模型的持续演进、混合负载下的查询过滤机制和模型泛化性增强等问题。

建议研究方向：

1) 模型持续演进问题：探究在动态负载下维持优化效果的模型演进方法，应对长时间跨度场景中可能出现的数据漂移等问题；
2) 混合负载下的查询过滤机制：在降低推理开销的同时保证模型的充分训练和对负载变化的充分感知；
3) 模型泛化性增强技术：探究模型的泛化性问题，探索不同负载类型之间实现模型迁移与预训练的有效途径，降低模型训练成本并提高模型在多样化负载下的适应性；
4) LLM 赋能小模型：利用 LLM 积累的查询优化经验补齐小模型短板，结合联邦学习、持续学习、RAG 等技术，实现更精准的计划选择。

返回目录

6.4 多元存储介质下向量检索的研究与应用

向量检索在实际应用中面临的内存消耗大、索引构建时间长、召回率受数据分布影响较大等问题。本课题旨在通过优化算法，显著提升向量检索在不同存储介质（Memory、SSD、GPU）下的效率、准确性及稳定性，降低资源消耗，为各类依赖向量检索的应用场景，如信息检索、推荐系统、图像识别等，提供更可靠、高效的技术支撑，推动相关领域的进一步发展与应用拓展。

建议研究方向：

1) 高效向量压缩方法：研发能有效压缩向量数据的算法，在保证检索精度的前提下，大幅降低内存占用。通过设计合适的压缩编码方式，减少存储向量所需的内存空间；
2) 基于数据分布自适应的检索策略：深入分析数据分布特征，开发能根据数据分布动态调整检索策略的算法。针对不同分布的数据，自动选择最优的检索方法，提高召回率，确保向量检索算法在各种数据分布下都能稳定高效运行；
3) 向量检索（SSD）索引结构优化：充分利用 SSD 的 IO 能力，降低索引构建时的内存、时间消耗，提高检索效率，降低综合成本；
4) 针对复杂数据分布的 GPU 检索技术：研究适合 GPU 处理的向量检索技术，提升在复杂数据分布下的召回率；
5) 基于图结构的向量数据库设计与传统高维向量数据库对比，更好的平衡索引构建、更新与查询效率、成本。

返回目录