

2026 CAAI-腾讯犀牛鸟研究计划 AI Lab 专项 研究课题

(申报材料不需包含各课题建议的全部研究问题, 选择其中部分或一项开展深入研究即可)

■ 多模态大模型.....	1
多模态理解前沿技术探索.....	1
1.1 多模态理解模型预训练研究.....	1
1.2 多模态理解模型后训练研究.....	2
多模态生成前沿技术探索.....	3
1.3 原生多模态架构下的视频生成技术研究.....	3
1.4 高效的视频生成方法研究.....	3
1.5 知识驱动的长视频生成方法研究.....	4
多模态理解生成统一的建模方法研究.....	4
1.6 多模态生成理解统一的模型架构研究.....	4
1.7 生成与理解的统一表示方法研究.....	5
1.8 面向生成理解统一的视觉编码器研究.....	5
高效多模态建模方法探索.....	6
1.9 面向大语言模型的离散-连续状态对齐与扩散生成研究.....	6
1.10 高效大模型蒸馏方法研究.....	7
世界模型与 3D 构建技术探索.....	7
1.11 视频世界模型的多模态交互算法研究.....	7
1.12 3D 世界模型的可仿真场景建模算法研究.....	8
1.13 高精 3D 物体生成和编辑方法研究.....	8
1.14 基于统一表征的 3D 生成理解大模型技术研究.....	9
可信数据构建技术探索.....	10
1.15 面向大模型的高价值数据知识网络构建方法研究.....	10
1.16 跨模态数据语义对齐与分层聚类算法研究.....	10
1.17 面向大模型训练数据的智能质检与价值衡量方法研究.....	11
■ 大模型智能体.....	11
智能体构建与持续学习研究.....	11
1.18 智能体模型基础架构探索.....	12
1.19 智能体模型的高效训练研究.....	12
1.20 多模态智能体持续学习研究.....	13
1.21 多模态智能体 Reasoning RL 的高效探索与训练优化研究.....	13
GUI Agent 的构建与学习研究.....	14
1.22 GUI Agent 的构建与交互学习研究.....	14
1.23 GUI Agent 多轮强化方法与训练框架研究.....	16
1.24 面向 GUI 场景的 VLM Agent 长程决策与自进化机制研究.....	16
具身智能体研究.....	17
1.25 面向具身智能场景的多模态理解通用能力探索.....	17
1.26 面向具身智能场景的通用鲁棒交互能力探索.....	18
■ 开放课题.....	19
1.27 下一代智能范式探索研究开放课题.....	19

多模态大模型

◇ 多模态理解前沿技术探索

近年来，多模态理解大模型（MLLM/VLM）在图文理解与跨模态推理任务中取得快速进展，正逐步成为通用智能系统的重要基础能力。随着模型规模与能力提升，其性能瓶颈逐渐由模型结构转向训练流程与感知建模机制。基于此，本课题围绕多模态模型训练全流程，系统探索预训练与后训练阶段的关键技术，以提升模型的理解能力、感知精度与泛化表现。

1.1 多模态理解模型预训练研究

研究背景：

当前多模态理解模型的核心能力主要在预训练阶段形成，但现有预训练范式在大规模训练条件下逐渐暴露出系统性瓶颈。一方面，主流跨模态投影与融合结构难以在保证训练稳定性的同时充分保留视觉细粒度信息，限制了跨模态协同感知能力的进一步提升；另一方面，随着模型规模扩大，通用图文数据驱动的预训练方式在能力扩展效率与泛化表现上已显现不足。

从训练流程看，多模态模型能力呈现明显的阶段性特征：预训练阶段决定跨模态表征质量与能力上限，中期训练（Mid-training）影响能力扩展效率，而后续对齐训练高度依赖预训练阶段所形成的能力基础。与此同时，现有 VLM 在精细视觉感知方面仍存在系统性短板，其根源在于预训练阶段监督信号偏重语义层级，缺乏对像素级或区域级感知能力的系统建模。

因此，有必要围绕多模态模型的预训练阶段，系统探索跨模态对齐结构、Mid-training 能力扩展策略以及视觉感知建模方法，为构建更稳定、高效且具备更强感知泛化能力的多模态模型奠定基础。

研究问题：

本课题聚焦多模态理解模型的预训练阶段，围绕当前主流预训练范式中存在的关键瓶颈。

- 1) **跨模态对齐结构的稳定性与信息保真问题：**针对现有 VLM 预训练中跨模态投影结构层次浅、对视觉细粒度信息保留不足、在大规模训练条件下对齐不稳定等问题，研究更适配预训练阶段的跨模态映射与融合机制，并分析其在对齐质量、训练稳定性与计算成本之间的权衡；
- 2) **Mid-training 阶段能力扩展的有效性问题：**针对预训练后期能力提升主要依赖后训练阶段、领域能力扩展效率低的问题，探索在 Mid-training 阶段引入结构化能力注入的方法，研究多源数据在不同能力维度中的组织与调度策略，以提升能力扩展效率并增强通用泛化能力；
- 3) **视觉感知建模的监督与表征不足问题：**针对现有预训练阶段视觉监督偏重语义对齐、缺乏像素级或区域级感知约束的问题，研究更适配 VLM 的视觉表征与 token 化方式，构建细粒度感知训练目标，系统提升模型在复杂视觉理解场景下的感知能力。

研究目标：

本课题旨在探索面向多模态理解模型的预训练方法体系。在模型架构层面，研究更优的跨模态映射与融合方式，减轻模态间信息损失并提升跨模态协同能力；在训练流程层面，形成有效的 Mid-training 能力注入与数据配方设计方法，为模型能力拓展奠定基础；在感知建模层面，通过优化视觉表征、tokenizer 与训练目标设计，提升模型在精细视觉理解任务中的整体表现。通过上述研究，推动多模态模型在端到端能力与泛化性能上相较现有预训练范式实现稳步提升。

1.2 多模态理解模型后训练研究

研究背景：

在多模态模型完成大规模预训练后，其是否能够在真实应用场景中稳定发挥理解、推理与决策能力，关键取决于后训练阶段对模型行为与输出方式的有效塑形。当前主流多模态模型通常依赖监督微调（SFT）、奖励建模（RM）与强化学习（RL）等方法进行对齐与能力强化，但在复杂任务与长链路推理场景中，现有后训练范式仍难以充分释放模型潜在能力。

具体而言，SFT 过程对数据分布与训练深度高度敏感，若缺乏合理约束，容易导致模型能力激活不均或过早收敛；奖励模型在过程合理性、输出风格与跨任务泛化方面仍存在不足，难以为后续强化学习提供高质量、可扩展的反馈信号。同时，在多轮交互、Agent 与 VLA 等多步决策场景下，传统 RL 方法在稳定性与探索效率方面面临显著挑战，限制了模型在复杂决策任务中的持续优化能力。

因此，有必要围绕多模态模型的后训练阶段，系统研究 SFT、RM 与 RL 的协同设计与优化方法，构建更加稳健、可扩展的对齐与决策训练范式，为多模态模型在复杂应用场景中的可靠部署提供支撑。

研究问题：

- 1) **SFT 与 RL 衔接优化：**研究 SFT 训练数据萃取方法，在控制数量的同时确保多样性与训练效果；设计系列指标对 SFT 过程进行监控，在保证模型能力激活的同时提供更大的探索空间；研究 SFT 训练深度/数据量对端到端效果影响，以更好结合 SFT 与 RL，取得更好的性能提升；
- 2) **RM 优化：**研究 PRM，对回答过程进行有效监督，减少 false positive 对 RL 的破坏，同时提升过程合理性及推理效率；研究 rubric RM，对不同问题生成个性化 checklist 并结合 checklist 进行打分，进一步提升 RM 泛化性、分数合理性，为 RL 提供更准确、全面的奖励信号；
- 3) **RL 优化：**研究强化学习训练策略，如课程学习、RLHF/RLVR 编排、模型融合（On-policy Distillation）等，在兼顾不同能力维度提升的同时，重点提升 RL 训练的稳定性与可扩展性；研究减少噪声数据、训推不一致等因素对训练稳定性影响的方法，使 RL 能够持续探索并扩展模型智能上限；
- 4) **面向多轮与多步决策场景的 RL 优化：**针对 Agent、VLA 等多轮与多步决策场景中 RL 稳定性与探索效率进一步下降的问题，研究更适配长序列决策的 RL 建模与训练方法，高效利用长序数据与过程奖励，同时缓解 off-policy 训练带来的熵坍塌与探索不足问题，提升模型在复杂决策任务中的执行效率与精度。

研究目标：

本课题旨在研究 VLM 的后训练。SFT 方面，需要研究 SFT 数据分布与训练深度，更好衔接 RL，取得整体端到端的性能提升。RM 方面，需要研究过程监督、风格监督等，不仅要提升模型准确性，还要提升输出内容的风格与合理性，以更契合用户需求与偏好。同时，对一些难校验与开放性问题，需要研究 rubric 奖励机制，以取得更全面、更合理、泛化性更好的奖励结果。RL 方面，需要研究训练策略，包括难度逐步提升的课程学习，专项训练与模型融合（On-policy Distillation），训练任务编排等，以确保模型在各个专项都获得充分的提升。设计方法和策略提升 RL 训练稳定性，使 RL 实现长训与 scaling 效应。另外，针对 VLA、agent 等场景下的长多轮、长序列决策问题，需设计更适配的 RL 建模及训练方法，结合过程监督提升模型决策/执行效率，同时缓解长序列下过度 off-policy 导致的熵探索问题。通过系列优化，提升整体后训练效果，相比当前范式有显著进步及 scaling 效应。

◇ 多模态生成前沿技术探索

近年来，视频生成模型在短视频和片段级内容生成取得了突破性进展，特别是在扩散模型等生成方法推动下，模型已经可以生成视觉质量和语义一致性都较高的数秒级视频内容。然而，视频生成依然面临诸多挑战，系统探索原生多模态融合、高效建模、知识驱动生成等技术，将为视频生成基础模型的发展提供坚实支撑。

1.3 原生多模态架构下的视频生成技术研究

研究背景：

视频生成正逐步向多模态和多任务融合方向发展。融合文本、图像、视频、音频等多种模态的生成，以及支持基于提示词和参考图深圳视频内容的生成，实现更强的语义理解与生成能力。当前各模态往往独立优化，跨模态适配与整合难以获得最优效果。因此，探索原生多模态架构下的多模态融合与协同，包括“思考模式”、基于图文视频生成、视频续写与编辑、以及音画一致性等新型任务，成为视频生成领域的重要趋势，这要求基础模型具备更高的表达力和灵活性。

研究问题：

- 1) **原生多模态视频架构：**研究原生多模态多任务融合的视频新架构，支持包括文、图、视频、音频的多模态融合输入和生成；
- 2) **新架构下的训练推理范式：**探索在原生多模态新型架构下，如何更简洁和高效的实现如思考模式生成、原生视频编辑、视频续写、参考图音视频生成等任务的统一训练推理范式；
- 3) **新架构下的强化学习：**探索原生多模态架构下的强化学习后训练算法，验证多任务多模态之间的强化学习的泛化性。

研究目标：

本课题旨在设计和实现原生多模态融合的基础架构，探索多任务协同机制，使文本、图像、视频、音频等输入信息能够高效融合，共同驱动视频生成过程。重点突破在原生多模态新架构上，如何能简洁和高效的统一思考模式生成、原生视频编辑、续写与参考图音视频生成等任务的训推范式，同时探索支持原生多模态架构的新型强化学习算法，验证多任务多模态之间的强化学习的泛化性。通过多模态生成与编辑公开基准上进行系统验证，达到或超过当前主流单模态方法的性能表现。

1.4 高效的视频生成方法研究

研究背景：

在实际应用中，随着视频序列长度的不断提升，视频的计算复杂度呈指数的增长，这对于实时交互反馈和更长的视频生成而言，效率都是当前最大的困难。如何设计高效的训练范式，显著降低计算和推理成本，成为亟需突破的技术难题。

研究问题：

- 1) **计算与存储效率：**探索设计计算复杂度随视频长度亚线性增长或近似常数增长的生成框架，保证生成质量的前提下，减少显存和算力的依赖；
- 2) **生成策略与推理方式：**保证生成质量效果的快速生成推理策略。

研究目标：

本课题研究重点设计随视频长度亚线性增长或近似常数增长的生成范式，显著提升生成效率并降低计算/显存成本。研究生成策略如分段生成、规划-执行融合等，提升模型的可控性和扩展性，并在视频公开基准上进行效果验证。

1.5 知识驱动的长视频生成方法研究

研究背景：

现有主流方法通常通过大规模数据驱动统计学习，在视频片段生成中取得成功，但对显式知识（如常识、物理规则、叙事逻辑、领域知识等）的建模能力有限，尤其是在长视频生成场景，如何有效引入结构化或半结构化知识，实现合理、可控、可推理的内容创作，是视频生成基础模型迈向更高智能水平的关键环节。

研究问题：

- 1) **知识的表示与形式：**探索在长视频生成中，“知识”的表示形式，以及对不同类型知识（常识、领域知识、叙事规则）的建模方式；
- 2) **知识与视频生成的耦合机制：**探索知识在生成过程中的注入方式，以及如何将知识约束转化为对生成模型可微、可学习的约束；
- 3) **知识驱动的长时规划与推理：**探索如何将长视频生成建模为一个基于知识的规划-执行过程，并在生成的长视频中保证知识的一致性。

研究目标：

本课题致力于探索结构化或半结构化知识驱动的视频生成方法，包括常识、物理规则、领域知识和叙事逻辑的表示、融合与模型化。重点突破知识约束在生成全流程中的耦合与反馈机制，实现基于知识的规划-执行式生成和长时推理，保证叙述和内容的合理性、多样性及可控性。拟在知识注入的视频生成公开基准上验证知识建模有效性，达到超越现有无知识视频生成模型的可控/一致性表现。

✧ 多模态理解生成统一的建模方法研究

本方向致力于解决多模态生成与理解在当前研究范式中仍相对割裂的问题，探索生成与理解统一的建模方法，打破生成与理解之间的壁垒，为构建可解释、可推理、可泛化多模态智能系统的打下关键基础。

1.6 多模态生成理解统一的模型架构研究

研究背景：

当前多模态生成与理解采用不同的基础模型架构：生成常用基于扩散模型的架构；而理解常用基于自回归（AR）的模型架构。相对割裂的模型架构设计成为生成、理解任务统一建模的一个重要阻碍。最近一段时间，多模态生成领域内已有多个研究工作探索基于 AR 的多模态生成模型架构，并在生成质量与生成效率上展示了不错的潜力。延续当前研究趋势，本课题致力于进一步探索 AR 模型架构在生成理解统一上的潜力与可行性，为多模态生成理解统一模型架构打下基础。

研究问题：

基于自回归的视觉联合理解生成方法：探索基于视觉理解模型的视觉生成方法，以基于自回归架构的视觉理解模型为基座，充分利用理解模型的能力解析用户的多样化、精细化需求，生成语义词元（Token）引导解码器渲染所需的图像或者视频内容。

研究目标：

设计并实现基于自回归架构的统一视觉理解生成框架，打破理解与生成的壁垒，并验证模型在 Scaling 至更大参数时的性能增益。通过充分统一视觉理解和生成模型，在复杂结构生成（如海报排版、四格漫画）任务上，语义一致性指标和美学指标要优于当前主流 Diffusion 方法，并支持分钟级长视频的生成。

1.7 生成与理解的统一表示方法研究

研究背景：

多模态生成与理解任务，对图像/视频等多模态数据的表示方法有不同的要求。生成模型更关注如何从潜在空间或条件信号中合成高保真视觉内容，要求数据表征带有更多细节信息；而理解模型则强调对视觉内容进行语义抽象与判别，依赖具有判别性的特征表示，但这些表示通常不具备可逆性，难以支持高质量生成。本课题致力于探索解决这一多模态表征方法里的基本矛盾，为生成与理解统一打下基础。

研究问题：

- 1) **统一表示的形式与语义结构：**探索能够支持生成与理解统一的表示方法应具备的基本属性；
- 2) **生成与理解的双向约束机制：**生成任务可以约束和提升理解表示的表征学习方法，以及可以利用理解任务（如分割、追踪、动作识别）提升生成过程的结构合理性与可控性等学习范式；
- 3) **统一表示的学习范式：**统一表示的学习范式以及大规模可扩展的训练方法。

研究目标：

探索生成理解统一表示方法的理论基础，实现统一表示的新型学习范式，并在多模态生成与理解基准测试集上，验证方法的有效性，达到与独立的生成、理解表示方法持平的表现。

1.8 面向生成理解统一的视觉编码器研究

研究背景：

多模态生成通常对多模态数据通过自编码器（VAE 等）进行数据的压缩与表征学习。而多模态理解模型通常使用 ViT 等自监督或者语义监督的编码器进行压缩。生成与理解任务的编码器设计目前还存在着显著差异。本课题致力于探索统一的视觉编码器设计，使得同一个编码器模型可以得到适合理解与生成两类任务的数据压缩表征。

研究问题：

- 1) **统一视觉编码：**对像素信息和语义信息的统一压缩、解码；
- 2) **视觉动态感知编码：**对视觉信息的非均匀压缩，即动态编码；
- 3) **基于统一视觉编码的自监督训练范式：**结合新型编解码方式的视觉自监督学习模型，实现视觉模型的 Next Token Prediction 范式。

研究目标：

设计并实现对图像/视频像素与语义统一的编码与解码模型，为生成理解统一模型打下基础；实现视频模型的动态编码；实现基于该编码与解码模型的视觉自监督学习范式；在生成与理解的标准测试集上，达到或超过独立生成、理解模型的表现。

✧ 高效多模态建模方法探索

本课题旨在打破大模型在模态扩展与推理效率上的双重壁垒。通过构建离散-连续协同的扩散生成框架，解决文本与多模态数据（图像/视频）的统一建模与高效并行生成难题；同时，融合 On-policy 蒸馏与跨模态知识迁移技术，在保证模型生成质量与泛化能力的前提下，显著降低训练与部署成本。

1.9 面向大语言模型的离散-连续状态对齐与扩散生成研究

研究背景：

当前，以自回归（Autoregressive, AR）机制为主导的大语言模型在自然语言处理领域取得了统治地位。然而，AR 模型严格的“从左到右”逐词生成方式带来了推理延迟高、难以进行并行加速以及缺乏全局规划能力（如无法灵活控制文本中间或结尾内容）等固有局限。与此同时，扩散模型（Diffusion Models）在图像生成领域证明了其强大的非自回归生成能力与可控性。将扩散过程引入语言建模（Diffusion LLMs），有望通过并行解码大幅提升生成速度，并赋予模型更强的文本编辑与受控生成能力。然而，由于文本数据本质上是离散的，而扩散模型通常定义在连续空间中，二者之间存在天然的模态鸿沟。现有的尝试往往面临着离散-连续空间映射困难、生成文本逻辑性弱于 AR 模型以及训练收敛难等挑战。

研究问题：

本课题旨在探索构建高性能的扩散大语言模型，重点解决离散文本与连续扩散机制的对齐难题。

- 1) **离散-连续对齐与质量提升：**设计高效的嵌入空间映射策略或离散状态扩散算法，缩小与自回归模型在生成质量上的差距；
- 2) **并行可控生成与编辑机制：**利用扩散模型的迭代去噪特性，实现非自回归的并行文本生成与精细化的可控文本编辑（如指定上下文填空、风格控制）；
- 3) **高效推理与采样优化：**优化采样算法，在保证生成多样性的同时显著降低推理成本，推动下一代高效、可控语言模型的技术演进。

研究目标：

设计并实现一套端到端的 Diffusion LLM 全栈训练与推理框架，覆盖预训练、指令微调（SFT）及强化学习（RL）全流程，交付完整源代码及实验报告。

研发新型离散/连续文本扩散算法，在同等参数量下，生成质量（PPL/BLEU）对齐主流自回归模型；利用非自回归并行解码优势，实现推理速度（Latency）较自回归模型提升 5-10 倍；构建基于扩散轨迹的 RL 对齐算法（如 DPO-Diffusion），在指令遵循任务（如 AlpacaEval）中胜率显著优于基线模型。

1.10 高效大模型蒸馏方法研究

研究背景：

大语言模型虽能实现超高性能，但参数量、计算成本、部署门槛等问题严重限制其落地，而蒸馏技术通过“知识迁移”让小模型具备接近大模型的性能，同时兼顾效率与成本，成为大模型从“实验室”走向“产业界”的关键桥梁。如果说大模型的价值在于“探索 AI 能力的上限”，那么蒸馏技术的价值在于“让这种上限触手可及”。此外，通过蒸馏可提取某一任务（如法律文本分析）的核心知识，让小模型在该任务上的性能超越通用大模型，还可以通过蒸馏使得模型可以在低延迟、低内存、低功耗的约束下仍然表现出强大的性能。事实上，蒸馏技术在海内外各个大模型公司广泛使用。最近，on-policy distillation 也进一步证明了蒸馏的强大潜力。本课题希望沿着 on-policy distillation 的方向继续探索，找出新的蒸馏范式和蒸馏算法，进一步提升中小尺寸模型的综合能力。

研究问题：

本课题旨在探索更加高效的大语言模型蒸馏算法与蒸馏范式。

- 1) **高效的教师-学生 (Teacher-Student) 蒸馏算法：**通过教师-学生或者大尺寸模型引导等方式将大尺寸模型的知识融入到小模型中，进而让小模型快速提升各项综合能力，例如代码生成、数学推理、诗词创作等；
- 2) **基于数据增广的蒸馏方法：**探索利用数据增广的方法（例如拼接强弱模型对于同一个 prompt 的输出等、让小模型完形填空大模型的输出等）提升弱模型生成更好回答的能力；
- 3) **跨词表、跨分词器的蒸馏范式与方法：**不同语言模型使用的 tokenizer、词表等有所差异，因此需要设计良好的方法实现跨词表、跨 tokenizer 的蒸馏方法，使得任意尺度、任意种类的强大模型均可用于小模型的微调。

研究目标：

研究更好的 on-policy distillation 方法或者通过重要性采样进行 off-policy distillation，此外还可以探索跨 tokenizer 的高效蒸馏方法与范式，乃至大模型自反馈、自进化的范式。完成大模型蒸馏算法创新，在中小尺寸模型上蒸馏效果优于现有最好的方法。

✧ 世界模型与 3D 构建技术探索

世界模型是 AI 迈向空间智能的关键路径，其核心在于让机器能够动态预测、推理并与物理环境交互，为构建高拟真数字世界奠定基础。然而，当前技术在长视频生成的逻辑一致性、3D 资产生成的细节对齐与跨模态控制方面仍面临挑战。本章节围绕世界模型与 3D 构建，旨在通过一系列关键技术突破，打通从多维感知到三维创造的全链路，为元宇宙、具身智能、数字孪生等前沿应用提供核心驱动。

1.11 视频世界模型的多模态交互算法研究

研究背景：

现有视频世界模型的交互方式以键盘和鼠标等控制信号为主，交互模态较为单一。本课题旨在探索多模态交互的视频世界模型，支持文本、语音、图像等多种模态的输入控制，及复杂任务中对长时序逻辑、三维空间一致性及多模态指令的精细控制需求。突破模型在长上下文 (Long Context) 记忆与多模态交互方面的瓶颈，探索建立具备记忆机制和上下文学习能力的交互式视频世界模型。

研究问题:

- 1) **提高世界模型的长上下文能力(Long context learning)** 长时序视频生成的计算效率与一致性: 研究基于线性复杂度架构或稀疏注意力机制的视频生成算法, 解决长视频生成中的“灾难性遗忘”与显存爆炸问题;
- 2) **探索世界模型 memory 的实现机制:** 探索显式和隐式不同的实现方式, 研究静态和动态场景的 memory 的不同要求和评估方法;
- 3) **扩展世界模型的多模态交互方式:** 实现更多样更多模态的交互方式, 探索不同交互方式融合的方法。

研究目标:

探索基于 memory 的长时序视频生成算法, 实现实时的长视频生成。实现支持多模态交互的视频世界模型, 扩展交互方式的多样性和各种交互方式的融合。

1.12 3D 世界模型的可仿真场景建模算法研究

研究背景:

近期, HunyuanWorld 模型以及 Worldlabs Marble 模型为代表的 3D 世界模型关注度迅速增加。本课题旨在探索超高清、带有物理属性的 3DGS 场景生成与重建技术。研究内容为利用多视角全景图等作为核心中间表达, 结合先进的生成模型, 最终输出超高清的、带有物理属性的 3D 高斯溅射(3DGS) 场景以及对应高质量碰撞 Mesh, 并能导入 Isaac Sim 等仿真平台。

研究问题:

本研究聚焦面向生成图片/视频的高精度 3DGS 重建及在具身仿真中的应用。

- 1) **超高质量的 3DGS 重建算法:** 可以从全景图及多视角视频中重建出 Marble 级别的超高清 3DGS, 清晰度大幅超越传统的 3DGS 优化方法;
- 2) **高效、高质量的 3DGS 转 Mesh 算法:** 可以从 3DGS 场景中导出高质量 Mesh 表达, 可编辑、可仿真;
- 3) **具身智能仿真的 3DGS 建模全链路:** 可以生成/重建出带有物理属性的可重打光的 3DGS 场景资产, 并能用于 Isaac Sim 仿真。

研究目标:

- 1) 技术: a) 基于全景图、多视图、视频等输入, 研究超高质量的带有材质属性的 3DGS 重建算法, 重建 PSNR 超过 3DGRUT、Worldlabs Marble 等业界方法 20%, 逆渲染及重打光效果超过 3DGRUT 等 20%, 重建时间 3 分钟以内; b) 研究高质量的 3dgs 转 Mesh 算法, 并验证可将 PLY 3DGS 与 GLB Mesh 资产联合导入仿真平台, 并实现机械臂仿真 Demo;
- 2) 系统: 交付以上源码及完整系统, 并验收可兼容主流仿真平台(Isaac Sim, MuJoCo 等)。

1.13 高精 3D 物体生成和编辑方法研究

研究背景:

当前, 基于图像的 3D 模型生成技术在精度方面取得了显著进展, 但仍面临细节对齐、生成可控性和编辑灵活性等方面的挑战。现有技术难以将图像中的精细纹理、几何细节和艺术风格精确匹配到 3D 模型, 导致生成的模型在细节表现上不够真实或与源图像存在偏差。同时, 用户难以精确控制生成过程中的特定属性如模型结构、局部形状或材质, 限制了根据特定需求定制 3D 模型的能力。此外, 生成后的 3D 模型在后期编辑时效率低下, 当前的编辑工具难以高效地进行非破坏性或

高层次语义编辑。这些局限性严重影响了 3D 资产的生产效率和质量，尤其在需要快速迭代和个性化定制的场景下，这些瓶颈尤为突出。

研究问题：

本课题将聚焦这些核心痛点，通过搭建高质量数据引擎和强化学习技术，精准提升图片与 3D 模型的细节匹配度。同时构建 基于“图片 + 组件”的可控生成框架，实现 3D 资产的精准定制与灵活编辑。

- 1) **高精度细节对齐机制的构建：**针对 2D 图像与 3D 模型间的几何与纹理细节对齐挑战，设计并实现创新的学习框架，精确捕捉图像从宏观到微观的细节信息，并高质量地映射到 3D 模型上，从而显著提升真实感和精细度；
- 2) **基于“图像 + 组件”的语义可控生成：**构建 3D 数据引擎和生成框架，使用户能通过图像、3D 条件和高层次语义指令（如添加、移除、替换组件或调整局部形状）精细控制和引导 3D 模型生成，实现模型的灵活修改和定制；
- 3) **3D 模型生成与编辑质量的量化评估及强化学习奖励函数构建：**建立全面且有说服力的评估体系，量化生成 3D 模型在几何精度、纹理真实感、语义一致性、编辑遵循度、美学方面的表现，并将其转化为有效的强化学习奖励信号用于模型优化和自动化评测。

研究目标：

在 3D 物体几何和纹理生成任务上达到行业领先水平，并从 0 到 1 构建和上线高可控、易交互的 3D 生成编辑框架。

1.14 基于统一表征的 3D 生成理解大模型技术研究

研究背景：

3D 多模态生成与理解是 AI 与 3D 世界进行交互的重要能力，其目标是构建能“看懂、生成、编辑”3D 世界的智能系统，支撑元宇宙、机器人、AR/VR 等核心应用。然而，现有方法仍面临三大核心瓶颈：一是文本/图片与 3D 几何、语义的跨模态统一表征难，模态差异导致生成内容与输入指令“语义错位”；二是高质量 3D 生成的细节与结构缺失，尤其在复杂场景（如室内布局）和精细几何（如物体纹理、部件结构）上，常出现“形状模糊、物理不合理”等问题；三是 3D 语义理解与编辑的精确性不足，难以实现“按指令修形状、按场景摆物体”的灵活交互。

研究问题：

本课题聚焦“3D 多模态的统一建模”，围绕表征对齐、生成增强、语义交互展开研究，推动 3D 智能从“能生成”迈向“高质量、可理解、易编辑”。

- 1) **面向 3D 的统一表征形式：**设计新的表征空间，可以将文本、图片、3D 等模态统一编码到该特征空间，实现跨模态对齐，促进 3D 多模态生成与理解任务；
- 2) **3D 理解与生成一体化模型：**研发 3D 多模态架构，包括数据构造、架构设计、模型选型、loss 设计等，最终输出兼容多种模态输入输出的 3D 多模态模型，可同时支持理解与生成任务；
- 3) **细粒度 3D 语义理解：**研究 part 或更细粒度的 3D 理解模型，支持在 3D 模态进行细分、定位、描述等任务，并逐渐从物体级理解升级至场景级理解。

研究目标：

设计并实现基于统一表征理解的生成、理解、编辑一体化框架，打破 3D 理解与生成的壁垒，并验证模型在 Scaling 至更大参数时的性能增益。

通过充分统一视觉理解和生成模型，在 grounding、caption、reasoning、文生 3D、图生 3D、3D 编辑等任务上的精度与语义指标不低于当前主流单任务方法。

◇ 可信数据构建技术探索

可信数据会涉及数据管理、数据质检、数据价值几个方面的研究探索。本研究从专家知识体系、数据语义表征两个方面探索数据组织管理的关键算法，将数据管理问题建模为细粒度分类和层级聚类问题。面对更极致的数据质量追求，人工质检无法满足需求，基于 Agent 探索先进的智能质检算法，同时设计数据实验对数据价值论证，旨在优化训练效率、降低算力成本及提升模型效果。

1.15 面向大模型的高价值数据知识网络构建方法研究

研究背景：

互联网中海量数据以自然孤立的形式存在，缺乏结构化的知识管理，高价值知识及其逻辑关联被掩埋。该研究旨在解决如何高效从海量数据中识别领域高价值数据、挖掘数据间语义/知识关联关系。高价值数据知识网络的构建，可以透视现有数据在知识体系上的分布情况、建立跨模态数据关联打破数据孤岛、主动发现数据缺口并牵引数据获取，为大模型提供灵活、精准、高质量训练数据与推理知识库。

研究问题：

- 1) **数据知识体系构建：**面向重点关注领域，设计多层次知识体系，体系完备合理可支持动态迭代更新；
- 2) **知识点关系网络构建：**基于知识体系，进一步构建完备知识点及知识点之间的关系、属性；知识点建设可结合人工专家，从数据自身出发更新迭代知识点集合；
- 3) **知识挂载：**面向多模态数据（文本、图像、视频、音频、3D、传感器等），基于已建设的数据知识网络，识别高价值多模态数据、理解挖掘多模态数据内容，实现对数据知识网络的数据挂载，完成跨模态高价值数据知识网络的建设。

研究目标：

提出知识体系构建与数据知识挂载的领域通用算法，落地重点领域（准召达到 95%以上），交付源代码及相关说明文档；在多模态大模型业务中落地应用，有明确量化指标提升，包括但不限于数据检索/数据生产提效、赋能大模型 RAG 等。

1.16 跨模态数据语义对齐与分层聚类算法研究

研究背景：

海量互联网数据以多模态、混合模态的形式存在，如一个简单的网页就可能同时包括文、图、视频、音频等模态。当前业界数据挖掘更多地专注于单模态、与文本结合的模态对，缺乏对跨模态、混合模态数据挖掘的算法研究，尤其是面向全模态（文本、图像、音频、视频、3D、传感数据等）数据的跨模态表征学习，并进一步基于表征的全模态数据分层聚类。针对庞大的跨模态数据，探索跨模态数据层次化聚类方法，支持大模型数据治理、数据分布调优、检索增强生成 RAG 等。基于混合模态大模型基座构建跨模态统一向量化模型，支持跨模态间语义互检索、混合模态数据检索，为下一代大模型数据处理范式进行技术探索和储备。

研究问题：

- 1) **跨模态数据表征：**面向文本、图像、视频、音频、3D、传感器等数据，研究跨模态表征学习算法，实现不同模态之间的通用语义对齐、全局/局部语义对齐，以及跨模态实体联动的语义对齐；
- 2) **混合模态数据表征：**在跨模态表征基础上，支持对混合模态数据（如图文交错的文档、音视频-字幕数据等）的表征，以及指令遵循的多模态语义表征（如指令中包括需要表征关注的目标）；
- 3) **多模态数据分层聚类：**基于跨模态/混合模态表征，探索高效的数据分层聚类算法，实现对多模态数据的分层聚类推理，并支持指令遵循的定制化知识体系聚类。

研究目标：

完成跨模态/混合模态数据表征算法、多模态分层聚类算法，在多模态 Emb 的 benchmark 中取得第一梯队效果，交付源代码及相关说明文档；在多模态大模型业务中落地应用，包括但不限于数据生产/数据检索优化、赋能大模型 RAG。

1.17 面向大模型训练数据的智能质检与价值衡量方法研究

研究背景：

随着大模型的快速发展，数据规模与质量是决定模型性能的核心要素。然而海量训练数据在领域专业性、知识准确性、细粒度 caption 完整性、实体 IP 准确性等方面普遍存在问题，严重制约了多模态大模型在垂类领域的高可用性。需要对多模态训练数据进行自动化质检，攻克大模型领域知识不足、细粒度特征捕捉难、模态对齐幻觉等难题。此外，随着数据规模的增长，除了数据本身的质量问题，数据的价值也同样重要，即数据对当前模型是否存在价值增益。研究高效的数据智能质检与价值衡量算法，对于优化训练效率、降低算力成本及提升模型效果具有重要的意义。

研究问题：

- 1) **智能质检算法：**面向文本与多模态数据，基于 Agentic 质检与 ReAct 框架，研究大模型训练数据的智能质检算法，实现对数据的领域专业性、知识准确性、细粒度 caption 完整性、实体 IP 准确性等的质检；
- 2) **数据价值实验：**面向数据对于模型的训练价值，设计数据价值验证实验框架，通过实验对比，可以找出让模型学得快、效果好的核心数据，剔除无效冗余数据。

研究目标：

完成基于 Agent 的大模型数据智能质检算法、通用的数据价值验证实验框架，质检准确率达到 95%以上，数据价值实验帮助模型减少 20%训练成本，交付源代码及相关说明文档。

大模型智能体

◇ 智能体构建与持续学习研究

本课题致力于数字智能体的模型架构探索、新型训练范式，以及在线持续学习方法，从而探索与构建可处理超长上下文、无限长记忆的智能体模型架构，并通过探索高效探索、经验能力复用和在线学习等新范式，提高智能体的学习效率，实现自演进的智能体模型。

1.18 智能体模型基础架构探索

研究背景：

随着智能体模型（Agent）成为 LLM 模型的重要应用场景，当前范式下的语言模型架构暴露出诸多缺点，包括缺乏有效的长期记忆，参数无法在推理过程中更新等，导致 Agent 在解决真实世界任务时，严重依赖上下文工程（Context Engineering）将最准确的信息裁剪到模型的上下文窗口中。因此，LLM 及多模态模型在整个 Agent 系统中，仍然只是一个被动的组件，而非能主动沉淀经验的智能体。如何设计更高效的智能体模型架构，可实现对过往经验的持续高效记忆，成为亟需解决的研究课题。

研究问题：

本课题主要研究支持长期记忆的智能体模型架构，重点解决记忆“存在哪里”和“如何更新”两大核心问题，目标实现在无需人工干预的上下文工程，智能体能够自主处理轨迹长度超过一百万 tokens 的复杂长任务，并具备随任务经验积累而自我进化的能力。

- 1) **记忆与推理分离的模型架构：**探索类似人脑结构的“小推理模型 + 长期记忆”模型架构，实现模型的参数化记忆；
- 2) **高效的长上下文处理能力：**研究能够高效处理长记忆与上下文信息的新型模型架构，实现模型的长记忆；
- 3) **模型原生的上下文工程能力：**将 Context Engineering 承担的的信息过滤、压缩、索引等能力内化为模型的原生表征能力，使模型能自主决定信息的留存与遗忘。

研究目标：

设计并实现高效处理 1-10M tokens 的上下文窗口的新型模型架构，并在长上下文理解与推理任务上，达到 SOTA 的表现。

1.19 智能体模型的高效训练研究

研究背景：

随着大模型与智能体（Agent）技术的快速发展，智能体正在从单一任务执行器演进为具备感知、推理、规划、行动和学习能力的通用系统。然而，当前智能体训练范式在很大程度上仍依赖于大规模数据、大量交互和高昂算力成本，其训练效率与可扩展性正逐渐成为制约智能体进一步发展的核心瓶颈。与此同时，人类和动物在复杂环境中的学习却表现出显著的高效性：能够通过少量示例、有限试错和抽象知识快速掌握新技能。这一对比凸显出当前智能体训练范式在训练目标、数据组织、学习机制和反馈利用方式上的根本不足。因此，亟需探索高效智能体训练范式，从根本上提升智能体的样本效率、计算效率和泛化能力，使其能够在有限资源条件下实现持续学习与能力成长。

研究问题：

本课题主要研究面向开放环境的智能体高效学习与持续自演进方法，重点解决经验数据的高效在线利用与弱监督信号下的过程指导两大核心问题。目标是建立一套在交互中持续学习的自演进框架，使智能体能够在无人工干预的情况下，通过积累历史经验和实时的价值评估自主更新模型参数。

- 1) **无可验证奖励情况下的自演进学习：**摆脱对可验证奖励信号的依赖，探索开放式验证或模型内生的奖励机制，实现智能体在只有模糊反馈甚至无反馈环境下的自主进化；
- 2) **知识、模型与环境的协同：**探索如何将先验知识、世界模型或规则引入训练过程以减少盲目试错，提高训练效率；

- 3) **跨任务与环境学习**：探索跨任务、跨环境的智能体知识迁移与能力复用，避免从零开始训练的模式，提高学习效率。

研究目标：

设计并实现可在线高效学习的智能体学习范式：可在线持续学习的智能体模型；可高效利用经验数据的学习范式；无可验证奖励信号下的智能体训练范式；并在开放任务环境中验证学习方法的有效性，在公开的复杂任务测试集上，达到 SOTA 的表现。

1.20 多模态智能体持续学习研究

研究背景：

近年来，大语言模型与多模态基础模型在互联网产业中快速落地，推动智能系统从“被动工具”向“自主智能体（Agent）”演进。在搜索推荐、智能客服、办公自动化和图形用户界面自动操作（GUI Automation）等场景中，智能体需要长期在线运行，持续感知环境变化，并根据用户反馈和任务结果不断调整自身行为策略。然而，现有多模态智能体大多依赖静态预训练或一次性离线微调，难以适应真实环境中任务分布、界面结构、用户偏好和交互模式的持续变化，表现为智能体性能退化快、维护成本高、可用性和泛化能力不足。以 GUI 智能体为代表的互联网应用场景中，智能体需要同时理解屏幕图像、界面文本和用户指令等丰富的多模态信息，并执行点击、输入等操作。当软件版本更新或业务流程调整时，现有方法通常需要重新采集数据并整体训练模型，难以满足在有限成本条件下互联网产品快速迭代的现实需求。因此，研究具备持续学习能力的多模态智能体，是提升智能体长期可用性与业务场景下决策稳定性的关键科学问题，具有显著的紧迫性和应用价值。

研究问题：

本课题围绕“多模态智能体在长期交互中的持续学习与稳定决策”展开研究。

- 1) **统一多模态状态表示与建模**：面向图像、语音、文本与动作的统一多模态状态表示与建模方法；
- 2) **自适应在线策略优化**：结合强化学习与在线反馈机制，使智能体能够通过环境奖励和用户反馈持续优化行为策略；
- 3) **持续学习与知识保存**：设计防止灾难性遗忘的持续学习机制，在学习新界面、新任务的同时保持既有能力；
- 4) **安全高效的在线系统维护**：研究适用于工业场景的高效在线更新与安全约束方法，保障系统长期稳定运行。

研究目标：

围绕多模态智能体在真实长期交互环境中的持续学习与稳定决策问题，本课题旨在构建具备在线适应能力、抗遗忘能力与工程可部署友好的多模态智能体持续学习训练方法和框架，使智能体能够在任务分布、界面结构和用户行为不断演化的条件下保持稳定性能并持续学习提升决策质量。

1.21 多模态智能体 Reasoning RL 的高效探索与训练优化研究

研究背景：

以 DeepSeek-R1 和 OpenAI o1 为代表的大型推理模型（LRM）通过推理时计算扩展（Inference-time Scaling），验证了 System 2 思维模式在工程范式上的可行性，极大地强化了模型在复杂逻辑任务上的推理能力。然而，当前的 Reasoning RL 仍面临严峻的效率瓶颈：一是 Credit Assignment 困境。主流算法通常将稀疏的 Outcome Reward 等权分配给响应序列中的所有 Token。这种粗粒度的信号反馈极易导致反馈信号的语义漂移，不仅在长达数万 Token 的 CoT 中

难以精准定位关键决策点，还引入了显著的采样方差，导致梯度更新不稳定。二是高维动作空间下的探索失效。近期研究表明，基于 RLVR 的优化往往表现为对基座既有分布的“过度锐化”，即侧重于提升 Pass@1 成功率，却往往以牺牲 Pass@k 的长尾覆盖度为代价。这导致模型难以在极大的动作空间中探索出基座分布外的新解法。在处理如 ARC-AGI-2 等需要 In-context Logic Synthesis 的任务，或 FrontierMath、AIME 等高难度非模版化推理任务时，模型难以通过有限采样探索到有效解。对于 4B/8B 等轻量级模型，其 RL 探索上限往往被其基座分布的先验知识所锚定。此外，模型频繁出现“正确答案、错误路径”的逻辑一致性崩塌，暴露出其对路径模式匹配的依赖，亟需探索精准的过程监督方法。

研究问题：

针对大模型 Reasoning RL 中面临的挑战，本课题聚焦于探索高效的大模型强化学习机制，旨在从算法层面解决大模型在强化学习中的探索困难与训练低效问题，提升模型在复杂推理任务中的表现。

- 1) **优化探索与利用 (Exploration-Exploitation) 的平衡：**从专家数据引导与探索算法优化层面入手，研究如何在高维动作空间中进行高效策略搜索，以缓解稀疏奖励信号下的冷启动难题，提升采样效率；
- 2) **研究精准的过程监督与信度分配机制：**针对 Credit Assignment 困境导致的梯度不稳定问题，研究精准的过程监督方法或自我批判机制，通过建立细粒度的反馈机制，将整体结果奖励有效回溯至关键推理步骤，以实现 Token-level 的奖励归因，缓解奖励估计量的高方差问题，从而显著提升模型在复杂长链任务中的样本利用率；
- 3) **提升模型逻辑一致性与泛化能力：**通过显式建模推理路径的合理性，纠正逻辑一致性崩塌，引导模型从“结果模式匹配”转向“逻辑路径理解”，从而深度强化模型处理复杂未知任务的推理鲁棒性。

研究目标：

设计并实现高样本效率的 Reasoning RL 探索算法，结合专家数据引导与过程监督技术，解决大模型在高维动作空间下探索难、稀疏奖励下收敛难的问题，提升模型在复杂推理任务中的表现。

在 4B/8B 等轻量级模型上对所提算法进行验证，使其在 AIME、FrontierMath 或 ARC-AGI-2 等高阶基准上的逻辑推理能力得到有效提升；在有限资源下充分挖掘模型性能边界，并发表 Tech Report 及开源相关优化模型。

✧ GUI Agent 的构建与学习研究

图形用户界面 (GUI) 是人与数字世界交互的普适通道。本领域的核心价值在于，让人工智能超越内容生成与对话，真正获得在通用数字环境中“看得懂、会操作、能学习”的具身行动能力，从而将大模型的认知智能转化为可直接驱动各类软件、完成复杂流程的数字化生产力。

1.22 GUI Agent 的构建与交互学习研究

研究背景：

大模型与多模态技术正以前所未有的速度推动人工智能领域发展，尤其在提升人机交互效率、实现复杂任务自动化方面展现出巨大潜力。GUI Agent (图形用户界面智能体) 作为连接大模型智能与复杂软件操作环境的关键桥梁，是当前具身智能和应用自动化领域的研究焦点。然而，将 GUI

Agent 从实验室环境推向真机/真实应用环境，依然面临一系列核心挑战：1）效率低：重复性任务仍需每次从头进行多步视觉推理，延迟高；2）可靠性差：长序列决策的错误累积导致成功率随步数显著下降；3）可控性弱：黑盒推理使用户难以理解和修正 Agent 行为，且缺乏对不同环境的适应性。与此同时，现实办公场景中存在大量结构化、可复用的重复性任务（如数据录入、报表生成、跨应用信息同步等）。对于这类任务，相比让 Agent 每次重新推理，更理想的方式是：用户通过一次演示（Demonstration）教会 Agent 完成任务，Agent 从中提取出可复用的工作流（Workflow），后续执行时直接调用该工作流并适配到新的输入参数。这种“Programming by Demonstration”的交互范式更符合用户心智模型，也能显著提升执行效率和可靠性。

基于此，本课题致力于建立一套从高保真环境搭建与高质量数据构建、Agent 强化学习优化到演示驱动工作流生成的完整技术链路，加速推动 GUI Agent 在复杂、动态的真实应用场景中实现泛化性强、鲁棒性高、操作流畅的自动化能力，并力求在学术界发表具有高水平影响力的研究成果。

研究问题：

- 1) **真机-仿真并行环境搭建与高质量数据体系构建**：探索并实现一套高鲁棒性、低延迟的真机（真实手机/桌面操作系统）执行环境，研究高效的跨平台 GUI 状态感知与动作注入机制，建立真机环境与高保真仿真环境的数据和动作空间对齐机制，实现从仿真到真机的高效经验迁移（Sim-to-Real）；针对真实场景中的长尾、多步任务，设计新型的人机协作式或自监督式数据采集范式，研究跨模态（视觉、文本、动作序列）数据标注与对齐的自动化方法，特别是针对非结构化 GUI 元素和复杂任务指令的处理，以低成本构建大规模、高质量的训练数据集；
- 2) **基于强化学习的通用决策能力优化**：探索利用离线或在线强化学习算法，对基于模仿学习训练的基础 GUI Agent 模型进行精调和策略优化，以解决 OOD（Out-of-Distribution）泛化性和长期规划问题。研究基于人类专家反馈的高效集成方法，使 Agent 能够更好地理解和遵循复杂指令，并快速适应真实环境中的动态变化和潜在错误；
- 3) **基于演示学习的工作流生成与泛化**：研究如何从用户的单次或少量 GUI 操作演示中，利用多模态大模型提取任务本质意图，区分核心操作与偶发行为，并将具体操作抽象为可复用、参数化的工作流步骤；设计工作流参数化泛化算法，使其能适配不同的输入参数和界面布局变化；构建人机协作机制，允许用户理解、验证和编辑生成的工作流，实现从“指令驱动”向“演示驱动”的范式跃迁。

研究目标：

本课题旨在研究 GUI Agent 环境和数据体系搭建、训练方法和演示驱动 GUI 自动化框架设计。建议研究目标包括，选择其中部分或一项开展深入研究即可。

- 1) 环境搭建方面，需要搭建真机与仿真交互环境，确保 Agent 在真实、动态环境中的准确交互，同时确保仿真环境到真机的高效经验迁移；
- 2) 数据构建方面，需设计高效、低成本的数据采集及标注链路，同时保证数据数量、质量及多样性；
- 3) 训练方法方面，需研究适配 Agent 的强化学习算法，对过程及结果进行准确、合理的奖励，使 Agent 在与环境的交互中高效探索与学习，不断提升 Agent 执行效率与准确性，更好完成任务且具备极好的任务泛化能力；
- 4) GUI 自动化框架方面，需要开发“演示驱动”GUI 自动化框架，支持从用户演示中自动生成可编辑、可泛化的 Agent 工作流，在相关基准上验证成功率超过 80%，且相比端到端 Agent 推理效率提升 2 倍以上。

1.23 GUI Agent 多轮强化方法与训练框架研究

研究背景：

随着大模型（LLMs/VLMs）从文本对话向“Agentic System”演进，基于视觉语言模型的 GUI Agent 成为实现通用人工智能的关键路径。不同于传统的 API 调用，GUI Agent 通过模拟人类视觉感知和键鼠/触控操作，能够跨应用处理复杂任务。然而，当前的 GUI Agent 主要依赖静态数据的监督微调（SFT），在面对长程（Long-horizon）、动态变化的真实环境时存在显著局限：1）泛化性差：SFT 模型难以应对未见过的 UI 布局和复杂交互逻辑；2）误差累积：在多轮交互中，单步决策错误往往导致整个任务失败；3）缺乏自我进化：无法像人类一样通过环境反馈（Trial-and-Error）持续优化策略。

研究问题：

本课题旨在研究面向 GUI Agent 的多轮强化学习（Multi-turn RL）方法。

- 1) **探索与利用的平衡（Exploration-Exploitation）**：针对 GUI 场景动作空间巨大且奖励稀疏（Sparse Reward）的问题，研究基于轨迹级（Trajectory-level）的价值评估与奖励设计，引入过程监督与反思机制（Self-Reflection）；
- 2) **训练稳定性与效率**：探索解耦的训练架构与异步采样机制，解决长序列决策中的梯度高方差问题，提升训练吞吐量；
- 3) **推理时计算扩展（Test-time Scaling）**：探索如何通过增加推理侧的“思考时间”（如 Chain-of-Thought, Search），结合 RL 训练出的价值模型，提升 Agent 在复杂任务中的规划与纠错能力。

研究目标：

本课题致力于构建一套高效、可扩展的 GUI Agent 多轮强化学习训练框架，并产出高性能的 Agent 模型。

- 1) 技术指标：a) 在主流 GUI Benchmark（如 OSWorld, AndroidWorld, WebArena）上，经过 RL 训练后的 Agent 任务成功率相较于 SFT 基座模型获得显著提升；b) 实现训练效率优化，相比传统 PPO 基线，训练吞吐量获得提升；c) 在长程复杂任务上的完成率达到业界领先水平；
- 2) 产出物：a) 源代码与框架：交付包含环境交互、数据采样、RL 训练（如 GRPO/PPO 变体）及评估的全流程代码库。

1.24 面向 GUI 场景的 VLM Agent 长程决策与自进化机制研究

研究背景：

随着多模态大模型（VLM）的爆发，构建能直接操作图形用户界面（GUI）的智能体（Agent）已成为通向通用人工智能的重要路径。然而，在真实的工业级 GUI 环境中，Agent 仍面临严峻挑战。首先，GUI 任务通常跨越数十步交互，具有极长的决策链路（Long Horizon），且仅在任务结束时才有稀疏的奖励信号。由于缺乏有效的信用分配（Credit Assignment）机制，模型难以定位关键动作，导致样本利用率极低。其次，受限于上下文窗口及环境的动态性，模型难以有效建模长程依赖，在多步执行中易产生信息遗忘与决策漂移。此外，传统的强化学习方案多依赖于静态策略，缺乏 In-context Learning 带来的灵活性，无法像人类一样从失败中学习，导致模型在面对未知场景时难以泛化。近期，Test-time Compute（测试时计算）的兴起为这一难题带来了曙光，结合世界模型（World Models）对动作后果的预演推断以及 Reflexion（反思）机制，展示了模型通过“慢思考”实现实时纠错的巨大潜力。但如何使 VLM Agent 在复杂 GUI 场景下，通过试错实现策略的闭环自进化，仍是当前亟待攻关的科研高地。

研究问题：

本课题旨在探索面向 GUI 场景的 VLM Agent 长程决策与自进化机制，系统性地解决真实环境下任务链路长、反馈稀疏及泛化性差等核心挑战。

- 1) **探索过程评估体系以优化 Credit Assignment:** 针对 GUI 任务中稀疏奖励导致的训练效率瓶颈，研究如何构建高效合理的 Critic 评估体系，实现对中间决策步骤的精准价值评判，从而解决信用分配难题，提升模型在复杂多步交互场景下的样本学习效率；
- 2) **研究世界模型与层次化规划以增强决策连贯性:** 探索基于世界模型 (World Model) 的环境演化推演或层次化规划方法，强化 Agent 对操作后果的预判能力，从而抑制长程任务执行中的决策漂移，确保指令执行的连贯性与准确性；
- 3) **探索基于记忆机制与实时反馈的策略自进化范式:** 研究如何增强 VLM Agent 的 In-context Learning 潜力，构建结合分层记忆模块 (Memory) 与实时反馈的自我反思机制。通过对历史经验的结构化存储与检索，突破上下文窗口限制，实现在未知场景下的自主纠错与策略迭代。

研究目标：

设计并实现具备“规划-推理-记忆-反思-进化”闭环能力的 GUI VLM Agent 系统。通过构建长程任务决策框架与动态环境建模，显著增强 Agent 在复杂多变场景下的泛化水平与自我纠偏能力。打通 GUI Agentic RL 训练的 Infra，构建价值评估、记忆、多智能体协同等核心能力模块；实现模型性能的 Test-time Scaling，支持通过多轮交互试错自主优化策略并显著提升任务成功率；交付一套完整的 GUI Agent 系统源代码及技术文档。

✧ 具身智能体研究

具身智能场景对多模态模型提出了新的要求：智能体需要在动态、可交互的物理环境中持续融合视觉、语言、空间状态与动作反馈信息，支撑感知—理解—行动的闭环过程。现有通用 VLM 多采用单次交互的静态理解范式，难以应对具身场景中强时序依赖、感知—行动耦合以及跨时刻一致性需求，在深度估计、空间定位和动态场景理解等精细感知能力上存在明显不足。具身场景的 VLA 模型尝试在 VLM 基础上引入动作空间建模，实现端到端的感知与决策统一，但在复杂真实环境中的鲁棒性和通用性面临挑战。为解决这些问题，需要探索多模态模型设计的新范式，系统研究面向具身智能的通用感知与鲁棒交互能力建模方法，从而为具身智能从“可用”走向“可靠”提供坚实基础。

1.25 面向具身智能场景的多模态理解通用能力探索

研究背景：

随着多模态大模型的发展，视觉-语言模型 (VLM) 在图文理解和视觉问答等静态感知任务中取得了显著进展，逐渐成为多模态智能系统的重要基础。然而，具身智能场景对多模态理解提出了更高要求：智能体需要在动态、可交互的物理环境中持续融合视觉、语言、空间状态与动作反馈信息，支撑感知—理解—决策的闭环过程。现有通用 VLM 多采用“单次输入—单次输出”的静态理解范式，难以应对具身场景中强时序依赖、感知—行动耦合以及跨时刻一致性需求，尤其在深度估计、空间定位和动态场景理解等精细感知能力上仍存在明显不足。上述局限表明，有必要突破以静态图文对齐为核心的 VLM 设计范式，探索面向动态交互环境的多模态理解通用能力，为具身智能提供更加稳健、可靠的模型基础。

研究问题：

本课题以多模态 VLM 基座模型为研究对象，聚焦其在具身智能场景下的通用理解能力建模，而非针对具体具身任务进行策略或控制层面的设计。

- 1) **具身与通用场景的细粒度感知增强：**面向具身智能典型场景和通用视觉理解场景，提升多模态模型的细粒度感知能力。提升模型在时序感知、空间定位与动态环境理解中的稳定性与精度；
- 2) **多模态理解-环境交互深度融合：**针对具身智能场景的高复杂性与多样性，研究多模态理解能力与环境交互过程的深度融合机制。增强 VLM 在跨场景、跨任务条件下的泛化能力，为后续具身智能系统的推理、规划与决策提供通用、可靠的多模态理解基础。

研究目标：

探索面向具身智能场景的多模态理解通用能力建设，开发面向精细感知和决策的多模态模型架构、训练范式和数据管线，在顶尖多模态模型基础上实现进一步感知、行动规划性能提升，在相关方向产出原创研究成果。

1.26 面向具身智能场景的通用鲁棒交互能力探索

研究背景：

在具身智能系统中，仅具备多模态感知与理解能力仍不足以支撑真实环境中的长期自主运行，模型还需能够将感知结果稳定、可靠地转化为行动决策。近年来，视觉-语言-行动模型（VLA）尝试在 VLM 基础上引入动作空间建模，实现感知—理解—行动的端到端闭环推理，为具身智能提供统一的感知与决策框架。然而，现有 VLA 模型仍处于探索阶段，在复杂真实环境中的鲁棒性和通用性面临显著挑战。具身场景具有高度动态性和不确定性，环境变化、感知噪声与执行误差会对交互过程产生累积影响，要求模型在长时序、多轮交互中保持行为稳定性与策略一致性，但现有方法在环境或任务变化时往往性能退化明显。因此，有必要在多模态理解基座能力之上，系统探索面向具身智能的通用、鲁棒交互能力建模方法，构建稳定适应复杂环境变化的 VLA 基座模型及统一评测框架，推动具身智能由“可用”迈向“可靠”。

研究问题：

本课题在多模态理解基座模型基础上，进一步面向具身智能中的交互与行动决策阶段，聚焦 VLA 模型的通用性与鲁棒性问题。

- 1) **多模态统一基座模型架构：**探索能够统一建模视觉、语言、空间状态与动作反馈等多模态信息的基座模型架构与表示方法。探索统一感知任务和动作空间的统一建模方案和模型架构，更好将通用感知理解能力迁移到 VLA 模型；
- 2) **具身交互鲁棒性与评测体系构建：**针对具身场景，提升 VLA 模型在交互任务上的鲁棒性和通用性。探索提升 VLA 模型的鲁棒性和可迁移性的方法，构建全面高效的 VLA 评测框架。

研究目标：

面向具身智能复杂交互场景，探索通用鲁棒的 VLA 基座模型设计与训练方法，形成统一的多模态-行动建模框架与评测体系，在复杂交互任务的稳定性、泛化性和鲁棒性方面实现系统性提升，并产出具有原创性的研究成果。

开放课题

1.27 下一代智能范式探索研究开放课题

本开放课题旨在突破现有技术范式，面向更加本质和长远的智能演进问题。