

2026 | 中南大学计算机学院 第六届青年学者论坛

欢迎有志青年人才来校共谋发展！

聚星城 谋新篇 开新局 创未来

会议手册

主办单位：中南大学计算机学院

2026年7月 湖南·长沙



中南大学计算机学院公众号



中南大学计算机学院官网



中南大学计算机学院第六届青年学者论坛会议日程

时间：2026年7月25日 地点：中南大学（潇湘校区）信息楼535会议室
腾讯会议号：917-194-304 会议密码：202607

中南大学计算机学院第六届青年学者论坛会议日程

时间：2026年7月26日 地点：中南大学（潇湘校区）信息楼535会议室
腾讯会议号：907-161-668 会议密码：202607

日期	时间	内容	报告人	报告形式	主持人
7月25日	09:00-09:15	会议开幕式			冯启龙
	09:15-09:45	跨时空学习：多模态感知、跨域泛化与鲁棒适应 Learning Across Space and Time: Multimodal Perception, Cross-Domain Generalization, and Robust Adaptation	钟嗣儒 香港科技大学（广州）	线下	李志泽
	09:45-10:15	可信与可控人工智能系统：从模型安全到智能体安全 Building Trustworthy and Controllable AI Systems: From Model Security to Agent Safety	李一戈 新加坡管理大学	线下	
	10:15-10:35 茶歇（信息楼 525 会议室）				
	10:35-11:05	面向分布式存储与安全计算的高效编码方案	李杰 香港城市大学	线上	李志泽
	11:05-11:35	实现最小训练数据披露的大语言模型遗忘方案 Less is More: Geometric Unlearning for LLMs with Minimal Data Disclosure	谭晨晨 蒙纳士大学	线上	
	11:35-15:00 午餐及休息				
	15:00-15:30	推进强化学习中的探索：迈向实用化具身控制 Advancing Exploration in Reinforcement Learning: Toward Practical Embodied Control	王一鸣 香港大学	线下	黄俊予
	15:30-16:00	从三维生成模型到具身智能	丁立鹤 香港中文大学	线下	
	16:00-16:30	Likelihood of the Existence of Average Justified Representation	周厚宇 新南威尔士大学	线下	
	16:30-16:50 茶歇（信息楼 525 会议室）				
	16:50-17:20	Learning Consistent and Interactive World Models with Generative Priors	孙文强 香港科技大学	线下	黄俊予
	17:20-17:50	Learning Structured Intelligence from Knowledge and Experience	胡云桐 埃默里大学	线下	
	18:00-19:30 晚餐				
温馨提示：主题报告均含 5 分钟提问交流时间					

日期	时间	内容	报告人	报告形式	主持人
7月26日	09:00-09:30	Expert-level Physical Design in Electronic Designs	韩耀辉 香港中文大学	线上	李少华
	09:30-10:00	基于 AI 驱动的智能芯片设计	来耀 剑桥大学	线上	
	10:00-10:30	From Graph Partitioning Algorithms to Scalable Graph Systems: Algorithm-System Co-Design for Large-Scale Graph Computation	丁泽中 中国科学技术大学	线上	
	10:30-10:50 茶歇（信息楼 525 会议室）				
	10:50-11:20	Learning to Forget: Machine Unlearning, Memorization, and Privacy in Foundation Models	赵凯然 华威大学	线下	李少华
	11:20-11:50	Local Search for Clustering in Almost-linear Time	楼家宁 北京大学	线下	
	11:50-14:30 午餐及休息				
	15:00-15:30	Mean-Shift PCA by Knockoff Mean	李梦达 香港中文大学（深圳）	线下	范成林
	15:30-16:00	生成式模型驱动的通用具身操作策略与基模型	梁志烜 香港大学	线下	
	16:00-16:20 茶歇（信息楼 525 会议室）				
	16:20-16:50	Self-Evolve Agents with In-Distribution Optimization	张钰荻 埃因霍温理工大学	线上	范成林
	16:50-17:20	大语言模型后门的生命周期：后门攻击持久性与可靠移除的研究 The Lifecycle of LLM Backdoors: Backdoor Attack Persistence and Reliable Removal	赵兴一 犹他州立大学	线上	
	17:30-19:30 晚餐				
温馨提示：主题报告均含 5 分钟提问交流时间					

青年学者简介



报告时间：7月25日(周六)09:15-09:45

报 告 人：钟嗣儒

汇报形式：线下

所在单位：香港科技大学（广州）

个人简介：钟嗣儒，香港科技大学（广州）数据科学与分析学域博士研究生，师从梁宇轩教授、乐阳教授；现于付昊桓教授（IEEE Fellow）指导下，在清华大学HPGC研究组与国家超级计算深圳中心开展访问研究。研究方向为时空智能、时间序列基础模型与AI for Science。已在NeurIPS（Spotlight）、ICML、KDD（Oral）、AAAI、WWW（Oral）等国际顶级会议发表论文15余篇，Google Scholar累计引用500余次，连续两年获香港科技大学（广州）DSA Excellent Research Award。受邀担任MM4ST@ACM MM 2025、ICME 2026及MM4TS@AAAI 2026教程讲者，并担任AAAI等会议程序委员会委员，以及NeurIPS、ICLR、KDD、TPAMI等会议和期刊审稿人。曾在贝塔无限记忆大脑团队、华为2012实验室、小鹏汽车自动驾驶中心及腾讯从事机器学习研究与工程实践，兼具学术研究、工业研发与成果落地经验。

报告题目：跨时空学习：多模态感知、跨域泛化与鲁棒适应

英文题目：Learning Across Space and Time: Multimodal Perception, Cross-Domain Generalization, and Robust Adaptation

报告简介：现实世界中的城市、气候、交通与具身环境持续在空间和时间上演化。相关数据不仅具有周期、趋势、空间拓扑和物理传播等复杂结构，还普遍面临多模态观测、跨区域差异和部署环境变化等挑战。围绕“如何让智能模型理解动态世界、泛化到未见场景，并在环境变化中稳健适应”这一核心问题，本报告将介绍三个层层递进的研究方向：首先，通过城市多模态表征与视觉语言模型，增强模型对空间语义与时间动态的感知和建模能力；其次，通过时空因子化基础模型，实现跨领域、跨尺度的知识迁移与泛化；最后，通过样本自适应正则化与外部记忆机制，应对真实环境中的噪声干扰和分布漂移。在此基础上，报告将进一步介绍面向物理场景的具身时空记忆探索，并展望智能模型从被动预测走向持续感知、长期记忆与自主决策。



报告时间：7月25日(周六)09:45-10:15

报 告 人：李一戈

汇报形式：线下

所在单位：新加坡管理大学

个人简介：李一戈，现任新加坡管理大学计算与信息系统学院博后研究员，主要从事人工智能安全、可信机器学习、大语言模型安全与智能体安全研究。以第一作者或通讯作者在ICLR、NeurIPS、ICML、NDSS、ACL、AAAI、CVPR、IEEE TDSC、IEEE TIFS 等国际会议和期刊发表 CCF A 类论文十余篇，谷歌学术引用 2200 余次。当前研究重点包括大模型后门攻击与防御、越狱与安全对齐、智能体风险评测及可控人工智能系统。

报告题目：可信与可控人工智能系统：从模型安全到智能体安全

英文题目：Building Trustworthy and Controllable AI Systems: From Model Security to Agent Safety

报告简介：随着大语言模型逐渐从内容生成工具演化为能够规划、调用工具并与真实环境交互的智能体，人工智能安全问题也从单一模型层面扩展到复杂系统层面。本报告将介绍在人工智能安全领域的系列研究，包括神经网络与大语言模型中的后门攻击与防御、大模型越狱与高风险场景安全评测，以及面向真实智能体运行轨迹的安全风险评估与监控。在此基础上，报告将进一步讨论如何推动人工智能安全从评测与防御迈向运行时监控、主动干预和失效保护，从而构建更加可信、可控的人工智能系统。



报告时间：7月25日(周六)10:35-11:05

报 告 人：李杰

汇报形式：线上

所在单位：香港城市大学

个人简介：李杰，IEEE高级会员，现任香港城市大学Research Fellow。主要研究方向为分布式存储与计算编码、私有信息检索及序列设计。他于2017年获得西南交通大学信息安全博士学位，博士期间曾在美国田纳西大学联合培养一年。2017年至2021年，先后在湖北大学数学系与芬兰阿尔托大学数学系从事博士后研究工作。

2021年3月至2026年3月，任职于华为香港理论部，担任高级研究员。曾主持/参与多项国家自然科学基金项目，在IEEE JSAC/TIT/TIFS/TCOM/INFOCOM/ISIT等高水平期刊/会议上发表学术论文30余篇，他开创了在国际上被命名为Li-Liu-Tang Transformation的编码转换方法及解决了高码率MDS码校验节点最优修复的公开难题，并被收录至《简明编码》百科全书，他的研究成果在华为完成多项成果转化并成功应用于分布式存储产品和鸿蒙系统。他曾荣获了2017年由IEEE Information Theory Society颁发的IEEE Jack Keil Wolf ISIT Student Paper Award。

报告题目：面向分布式存储与安全计算的高效编码方案

报告简介：随着分布式存储系统规模不断扩大，节点故障已成为常态。如何在保证数据可靠性的同时，降低节点修复过程中的带宽和磁盘I/O开销，是系统设计中的关键挑战。最小存储再生（MSR）码能够在保持最优存储冗余的前提下，显著降低故障节点修复所需下载的数据量，是当前存储编码领域的研究热点，然而MSR码巨大的分包数严重制约其实际应用。本报告首先介绍一种通用转换方法，可将任意MDS码转换为MSR码，从理论上解决节点修复难题和MSR码构造难题；之后介绍如何通过牺牲少量修复带宽来大幅降低分包数，使其便于实际部署；最后讨论分布式存储系统下的安全计算问题，包括隐私安全的矩阵乘法方案及其性能优势。



报告时间：7月25日(周六)11:05-11:35

报 告 人：谭晨晨

汇报形式：线上

所在单位：蒙纳士大学

个人简介：谭晨晨，现为澳大利亚蒙纳士大学信息技术学院博士研究生，主要研究方向为大语言模型机器遗忘与隐私保护。其研究重点是如何使大语言模型在保留通用推理与语言生成能力的同时，选择性移除敏感、过时或不再被允许使用的知识，从而提升模型的可控性、安全性与可问责性。在博士阶段，她围绕大语言模型遗忘的有效性、数据效率、模型效用保持及遗忘后幻觉等问题开展了系统研究，相关成果发表于或录用于ICML、NeurIPS、IEEE TNNLS 等国际会议和期刊。她的研究目标是推动大语言模型从“具备生成能力”进一步走向“可控、可遗忘、可验证”，为生成式人工智能的安全部署和责任应用提供技术支持。

报告题目：实现最小训练数据披露的大语言模型遗忘方案

英文题目：Less is More: Geometric Unlearning for LLMs with Minimal Data Disclosure

报告简介：随着大语言模型被广泛应用于真实系统，如何根据隐私保护、数据治理和知识更新需求，移除模型中的特定信息，已成为可信人工智能领域的重要问题。然而，现有大语言模型遗忘方法通常需要重新访问原始训练数据，并依赖大规模遗忘数据和保留数据进行参数更新。这不仅增加了计算与数据成本，还可能在遗忘过程中再次暴露敏感信息。本报告将介绍我们提出的几何遗忘方法 Geometric Unlearning（GU）。该方法不需要访问原始训练语料，而是利用包含目标实体锚点的轻量级合成提示，将模型在提示处理阶段形成的隐状态投影并对齐到“安全规划空间”。该空间为从少量安全参考提示中提取模型期望遗忘行为对应的低秩几何结构。同时，方法通过冻结教师模型的蒸馏约束，减少遗忘过程对非目标知识和模型通用能力的影响。论文在 ToFU 等隐私导向的遗忘基准上进行了评估，结果表明，GU 能够使用极少量合成数据实现较强的目标知识抑制，并较好地保持模型在非目标任务上的性能。报告还将讨论该方法的核心设计以及“行为层面的知识抑制”与“模型内部信息被彻底删除”之间的区别。

青年学者简介



报告时间：7月25日(周六)15:00-15:30

报 告 人：王一鸣

汇报形式：线下

所在单位：香港大学

个人简介：王一鸣，现为香港大学与卡内基梅隆大学联合博士后研究员，同时担任超智技术有限公司首席科学家及XYLab负责人。博士期间接受澳门大学与南洋理工大学的联合博士培养，并获澳门大学计算机科学博士学位，本科毕业于哈尔滨工业大学。其研究主要聚焦强化学习中的高效探索、具身智能与复杂智能决策，重点关注

高维观测、稀疏与延迟奖励、长时序任务及过驱动控制等关键问题。他以第一作者和通讯作者身份在 NeurIPS、ICML、AAAI 和 CVPR 等顶级人工智能会议录用或发表论文数十篇，并获 NeurIPS Spotlight 和 AAAI Oral。除基础研究外，他还致力于推动强化学习在数据中心节能、冷链系统优化、智慧城市与工业智能控制等场景中的应用落地，主持或参与多项产学研合作项目，并担任 ICML、NeurIPS 等国际会议及期刊的审稿人。

报告题目：推进强化学习中的探索：迈向实用化具身控制

英文题目：Advancing Exploration in Reinforcement Learning: Toward Practical Embodied Control

报告简介：强化学习通过智能体与环境的持续交互自主学习决策策略，是实现具身智能与自主控制的重要技术路径。然而，在真实复杂环境中，高维观测、稀疏或延迟奖励、长时序决策以及大规模乃至过驱动动作空间，使传统基于状态计数或预测误差的探索方法面临状态空间爆炸、噪声干扰、表示坍塌和探索效率低等问题。本报告将围绕“如何为强化学习构建结构化、可扩展的探索机制”展开，介绍一系列面向现实具身任务的探索方法。首先，利用与价值函数相关的状态度量构造探索奖励，使智能体优先发现具有决策意义的状态变化；进一步，通过行为度量学习和潜空间方向性探索，在紧凑表示空间中获得更加稳定和多样化的探索行为；针对长时序任务，引入基于状态自预测的潜在前瞻能力，使智能体能够综合未来状态、时间依赖关系与不确定性进行探索；针对仿生机器人等过驱动控制问题，则通过学习可解耦的动作协同模式，将高维动作探索转化为低维、结构化的潜动作搜索。相关方法在高难度游戏、连续控制、室内导航、机器人操作和仿生控制等任务中验证了其在样本效率、探索覆盖、策略性能与泛化能力方面的优势。最后，报告还将结合工业能源优化等实际应用，讨论强化学习探索算法从仿真研究走向现实部署所面临的安全性、可靠性和可扩展性问题。



报告时间：7月25日(周六)15:30-16:00

报 告 人：丁立鹤

汇报形式：线下

所在单位：香港中文大学

个人简介：丁立鹤，香港中文大学MM Lab博士生，导师为薛天帆教授，研究领域围绕三维生成、视频生成与具身智能，致力于高质量、可拆解、可交互的三维物体生成以及世界模型在机器人领域的应用和探索，在 CVPR、ICCV、ICLR、ICML 等国际顶会期刊发表论文16篇，其中第一/共一作8篇，研究成果获ECCV Oral。

报告题目：从三维生成模型到具身智能

报告简介：三维物体作为现实物理世界的基本构成单元，其可生成、可拆解、可交互的特性对具身智能的发展至关重要。本报告前半部分将围绕高质量三维物体生成所面临的数据、先验与算法挑战展开，系统介绍从整体几何纹理建模，到部件语义划分与铰链运动预测，并延伸至服务于机器人仿真训练闭环的全链路技术。后半部分则聚焦于基于视频生成模型的世界动作模型，探讨如何借助视频模型所蕴含的物理先验，增强机器人在真实交互场景中的理解与执行能力。

青年学者简介



报告时间：7月25日(周六)16:00-16:30

报 告 人：周厚宇

汇报形式：线下

所在单位：新南威尔士大学

个人简介：周厚宇，香港城市大学博士，师从李闽滨教授。现任新南威尔士大学算法决策组博士后，师从Toby Walsh教授（澳大利亚两院院士）。曾荣获2023年度香港城市大学杰出学术表现奖与2025年度CCF智能体与多智能体系统优秀博士论文提名奖。主要研究方向包括算法机制设计与社会选择理论，致力于在复杂环境中实现集体决策的效率与公平性，具体涵盖设施选址、公平分配、投票机制等多个前沿议题。其研究成果已在 SODA、NeurIPS、AAAI 等会议发表论文十余篇，并担任等Artificial Intelligence、ACM Transactions on Economics and Computation 等期刊的审稿人。

报告题目：Likelihood of the Existence of Average Justified Representation

报告简介：We study the approval-based multi-winner election problem where n voters jointly decide a committee of k winners from m candidates. We focus on the axiom average justified representation (AJR) proposed by Fernandez, Elkind, Lackner, Garcia, Arias-Fisteus, Basanta-Val, and Skowron (2017). AJR postulates that every group of voters with a common preference should be sufficiently represented in that their average satisfaction should be no less than their Hare quota. It is well-known that a winning committee satisfying AJR is not guaranteed to exist for all multi-winner election instances. In this paper, we study the likelihood of the existence of AJR under the Erdős-Rényi model. We provide a clean and complete characterization of the existence of AJR committees in the case where m is a constant and n tends to infinity.



报告时间：7月25日(周六)16:50-17:20

报 告 人：孙文强

汇报形式：线下

所在单位：香港科技大学

个人简介：2026年7月博士毕业于香港科技大学，入选腾讯混元青云顶尖人才计划，长期致力于世界模型与多模态生成模型等研究领域。学术成果发表于ICML, CVPR, ICCV, TIP, ACM等顶级会议和期刊，作为核心成员参与的项目github仓库star达5k。作为核心成员，参与训练开发腾讯混元交互世界模型，在开源社区和公司内部获得广泛好评。

Learning Consistent and Interactive World Models with Generative Priors

报告简介：The rise of generative models has catalyzed the development of world models that are both geometrically consistent and interactively responsive, offering unprecedented capabilities in simulating real-world environments. However, transitioning from pixel-level video generation to structural world modeling faces key challenges: ambiguous 3D reconstruction from incomplete views, inconsistent 3D/4D generation from a single image, and the tradeoff between real-time interactivity and long-term consistency. We address these challenges through three novel frameworks that enhance the consistency and interactivity of video diffusion models.

First, we tackle the ill-posed problem of 3D reconstruction from sparse views. We introduce ReconX, which reframes reconstruction as a temporal generation task. By constructing a global point cloud as a 3D structural condition, ReconX unleashes pre-trained video diffusion models to synthesize high-fidelity, view-consistent frames, which are then optimized via a confidence-aware 3D Gaussian Splatting (3DGS) scheme, outperforming prior methods across real-world datasets.

Second, we extend the generative prior to consistent 3D and 4D world creation. We present DimensionX, which decouples spatial structures and temporal dynamics in video diffusion models via the ST-Director, a pair of spatial and temporal LoRA directors trained on curated dimension-variant datasets. This decoupling enables photorealistic 3D static scenes and 4D dynamic sequences from a single image. Moreover, we incorporate a trajectory-aware mechanism and an identity-preserving denoising strategy, demonstrating improved spatial and temporal consistency.

Finally, we propose WorldPlay, an autoregressive video diffusion model that achieves both real-time generation and long-term geometric consistency. WorldPlay introduces a dual action representation unifying discrete keyboard inputs with continuous camera poses for robust controllability, and a reconstituted context memory that retrieves and reframes past observations to preserve geometric fidelity over extended horizons. To enable real-time inference without sacrificing memory integrity, we further propose context forcing, a distillation strategy aligning context distributions between teacher and student models.

青年学者简介



报告时间：7月25日(周六)17:20-17:50

报 告 人：胡云桐

汇报形式：线下

所在单位：埃默里大学

个人简介：胡云桐，现为美国 Emory University 计算机科学博士研究生，主要研究方向包括 Graph ML、Generative AI、Agentic Systems。其研究关注如何利用知识、关系、交互和执行轨迹中的结构信息，构建能够进行复杂推理、自主行动并从经验中持续改进的智能系统。相关成果发表于 ICML、NeurIPS、SIGIR、KDD、ACL、NAACL等国际学术会议。

报告题目：Learning Structured Intelligence from Knowledge and Experience

报告简介：当前基础模型主要从大规模数据中学习统计规律，但在处理复杂关系、执行长期任务以及从交互经验中持续改进方面仍存在明显局限。本报告围绕一个核心问题展开：如何使人工智能系统能够发现、利用并不断优化知识与经验中的结构，从而提升其推理、行动和学习能力？报告首先介绍如何将实体关系、科学文献、复杂系统与行为轨迹组织为图结构，并通过图检索、图表示学习和图增强生成，使基础模型能够在大规模结构化知识中进行有效的信息获取与复杂推理。随后，报告将进一步讨论如何将这种结构化建模拓展至智能体系统：将任务依赖、工具调用、智能体通信与执行轨迹表示为可学习的结构，并利用环境反馈自动优化智能体的角色分工、交互拓扑、模型选择和执行策略。报告最后将讨论如何形成知识、行动、反馈与系统优化之间的闭环，以及面向自适应和持续改进智能系统的未来研究方向。



报告时间：7月26日(周日)09:00-09:30

报 告 人：韩耀辉

汇报形式：线上

所在单位：香港中文大学

个人简介：韩耀辉，香港中文大学计算机科学与工程系博士生，导师是Tsung-Yi Ho教授。他在中南大学取得了物理学学士学位。他已经以第一作者在顶级期刊和会议发表了5篇论文，目前的研究方向是大语言模型后训练、电子设计自动化等。

Expert-level Physical Design in Electronic Designs

报告简介：在电子设计自动化领域，尽管现有的物理设计框架已相当丰富，工业界在实际生产中仍高度依赖人类工程师的手动干预，以确保最终布局的可靠性与性能。究其原因，传统基于规则的方法难以像资深专家那样，在布局阶段即对后续布线结果形成有效预判，并据此进行经验驱动的迭代优化；同时，真实场景下的设计约束往往远比常规算法所能处理的更为复杂且难以形式化；此外，工业界频繁出现的特定设计需求，也使传统算法缺乏必要的灵活性以作出针对性调整。针对上述挑战，我们提出了两条互补的解决路径：其一为专家设计迁移（DATE-26），其二是基于大语言模型的智能设计方法（ICML-26、DAC-26）。与既有规则驱动算法相比，我们所提出的方案不仅在标准评测指标上取得了显著提升，而且在生成的实际版图质量上，展现出更强的灵活性与物理合理性，更贴近工业级应用的真实需求。



报告时间：7月26日(周日)09:30-10:00

报 告 人：来耀

汇报形式：线上

所在单位：剑桥大学

个人简介：来耀，剑桥大学计算机系博士后研究员，合作导师为Robert Mullins教授。他于香港大学获得计算机科学博士学位，导师为MMLab罗平教授。博士期间，他曾赴德克萨斯大学奥斯汀分校访问学习，由David Z. Pan教授(IEEE/ACM Fellow)指导。他拥有清华大学硕士学位和复旦大学学士学位。他的研究方向为面向电子设计自动化的人工智能(AI for EDA)，研究成果发表于NeurIPS(3篇Spotlight)、ICML、AAAI、DAC、ICCAD、TCAD等顶级会议和期刊。他曾获香港政府博士奖学金，研究工作入选AAAI 2025最具影响力论文。

报告题目：基于AI驱动的智能芯片设计

报告简介：现代芯片设计是一项劳动密集型工作,需要经验丰富的工程师投入数月的努力。本次报告将探讨人工智能如何从根本上变革这一范式，实现从物理布局到电路综合等关键设计环节的自动化，并从两个视角出发，阐述AI驱动芯片设计的统一愿景。在物理设计层面,报告将介绍强化学习如何攻克数百万电路模块布局这一复杂任务，在线长和可布线性方面取得显著提升。在电路设计层面，将展示大语言模型如何作为智能设计代理,在无需任务特定训练的情况下实现模拟电路综合的自动化。这些方法遵循共同的理念：将传统设计问题重新表述为适合学习的表示形式。使AI能够发现超越传统启发式方法的解决方案。通过弥合机器学习与电子设计自动化之间的鸿沟，这一系列研究旨在实现芯片设计的普惠化，使其更快速易用，减少劳动密集的人工投入。



报告时间：7月26日(周日)10:00-10:30

报 告 人：丁泽中

汇报形式：线上

所在单位：中国科学技术大学

个人简介：Zezhong Ding is a Ph.D. candidate in the Data Darkness Lab (DDL) at the School of Artificial Intelligence and Data Science, University of Science and Technology of China (USTC), supervised by Prof. Xike Xie. He was a visiting student at the Department of Computer Science, Hong Kong Baptist University, from 2024 to 2025, working with Prof. Jianliang Xu. He has published several research papers in prestigious conferences and journals such as SIGMOD, SC, NeurIPS, ICML, KDD, ACL, and TC. His research focuses on data management, machine learning, and computer systems over big data and large models, in particular, big graphs and foundation models. Homepage: zezhongding.github.io

From Graph Partitioning Algorithms to Scalable Graph Systems: Algorithm-System Co-Design for Large-Scale Graph Computation

报告简介：The domain of large-scale graph computation has rapidly emerged as a pivotal force in modern AI, driving transformative advancements across diverse domains, from social network analysis to drug discovery. However, the inherent complexities of real-world graph data—characterized by sheer scale, dynamic evolution, and high velocity—pose formidable challenges for current graph neural network (GNN) training and processing systems. Addressing these bottlenecks necessitates a paradigm shift, moving beyond isolated algorithmic improvements towards a holistic Algorithm-System Co-Design approach.

This talk specifically delves into our works in large-scale graph processing, showcasing an end-to-end co-design approach: from the innovative design of graph partitioning algorithms to the development of robust system frameworks that synergistically optimize both partitioning and GNN training.

青年学者简介



报告时间：7月26日(周日)10:50-11:20

报 告 人：赵凯然

汇报形式：线下

所在单位：华威大学

个人简介：赵凯然，英国华威大学（University of Warwick）机器学习博士研究生，指导老师Prof. Peter Triantafillou与Dr. Eleni Triantafillou。本科及硕士毕业于西安交通大学，师从沈超教授、张萌教授和管晓宏教授；同时获得法国中央理工学院工程师学位（Diplôme d’ Ingénieur）。主要研究方向为AI安全，重点关注生成式模型（包括视觉模型与语言模型）和判别式模型中的机器遗忘（Machine Unlearning）、模型记忆（Memorization）与隐私保护等问题。相关成果发表于 NeurIPS、ICML 等国际人工智能会议，并担任 NeurIPS、ICML、ICLR、TMLR 等国际会议和期刊审稿人，曾任 NeurIPS Machine Unlearning Competition, NeurIPS Workshop等组委会成员，现任Women in Machine Learning (WiML) Board成员。

Learning to Forget: Machine Unlearning, Memorization, and Privacy in Foundation Models

报告简介：随着机器学习模型被广泛应用，如何删除指定训练数据对模型的影响，已经成为隐私保护和可信人工智能中的重要问题。本报告将从“被遗忘权（Right to be Forgotten）”出发，介绍机器遗忘（Machine Unlearning）的基本概念，以及模型记忆（Memorization）在模型能力与隐私风险中的双重作用。围绕“如何让模型在保留已有能力的同时真正忘记指定信息”这一核心问题，结合判别模型和生成模型中的最新研究，探讨数据遗忘难度、模型记忆如何影响遗忘算法的设计，以及如何利用模型内部机制有效降低生成模型复现训练数据的风险。最后，将讨论人工智能时代下的机器遗忘应如何定义与评估，以及未来可信人工智能的发展方向。



报告时间：7月26日(周日)11:20-11:50

报 告 人：楼家宁

汇报形式：线下

所在单位：北京大学

个人简介：楼家宁，北京大学计算机学院前沿计算研究中心博士研究生，师从姜少峰老师，本科毕业于北京大学图灵班。他的研究方向为理论计算机科学，主要关注亚线性算法与大规模数据算法，目前致力于设计流式、动态和大规模并行计算模型下的高效聚类算法。相关研究成果发表于 FOCS、SODA、ICALP、ICML 和 ICLR 等理论计算机科学与机器学习领域的国际学术会议。

Local Search for Clustering in Almost-linear Time

报告简介：We propose the first \emph{local search} algorithm for Euclidean clustering that attains an $O(1)$ -approximation in almost-linear time.

Specifically, for Euclidean k-Means, our algorithm achieves an $O(c)$ -approximation in $\tilde{O}(n^{1 + 1/c})$ time, for any constant $c \geq 1$, maintaining the same running time as the previous (non-local-search-based) approach [Dupr{\e} la Tour and Saulpic, arXiv'2407.11217] while improving the approximation factor from $O(c^6)$ to $O(c)$. The algorithm generalizes to any metric space with sparse spanners, delivering efficient constant approximation in ℓ_p metrics, doubling metrics, Jaccard metrics, etc.

This generality derives from our main technical contribution: a local search algorithm on general graphs that obtains an $O(1)$ -approximation in almost-linear time. We establish this through a new 1 -swap local search framework featuring a novel swap selection rule. At a high level, this rule “scores” every possible swap, based on both its modification to the clustering and its improvement to the clustering objective, and then selects those high-scoring swaps.

To implement this, we design a new data structure for maintaining approximate nearest neighbors with amortized guarantees tailored to our framework.



报告时间：7月26日(周日)15:00-15:30

报 告 人：李梦达

汇报形式：线下

所在单位：香港中文大学（深圳）

个人简介：李梦达，香港中文大学（深圳）数据科学学院博士研究生，师从姚建峰教授。主要从事随机矩阵理论及其在机器学习中的应用研究。本科毕业于法国索邦大学，硕士毕业于巴黎萨克雷大学。在高维统计与机器学习理论方向开展相关研究。

报告题目：Mean-Shift PCA by Knockoff Mean

报告简介：Removing noise is difficult, but adding noise is easy. In this work, we show how to eliminate mean-shift noisy components from PCA by deliberately introducing knockoff mean-shift perturbation. Standard PCA is highly sensitive to shifts in the sample mean: a small fraction of samples from a shifted distribution can cause large deviations in the leading principal components. In high-dimensional regimes, existing Robust PCA approaches cannot handle the mean-shift contamination structure inherent in the mixture model. Using tools from Random Matrix Theory, we prove that the mean-shift spikes are spectrally separable from the stable eigenvalues of the original covariance. Furthermore, the original eigenspace remains asymptotically invariant to the contamination, independent of the mixture weight. Exploiting this spectral stability, we propose a simple, two-stage PCA algorithm by adding knockoff mean that identifies and removes the mean-shift component using only standard PCA operations.



报告时间：7月26日(周日)15:30-16:00

报 告 人：梁志烜

汇报形式：线下

所在单位：香港大学

个人简介：梁志烜，香港大学计算机科学系四年级博士研究生，师从 Ping Luo 教授和 Wenping Wang 教授（IEEE & ACM Fellow）。曾在 阿里巴巴通义千问团队，英伟达研究院及上海人工智能实验室科研实习，并于 2024–2025 年在加州大学伯克利分校 MSC 实验室访问，指导老师为 Masayoshi Tomizuka 教授。他长期从事

具身操作、机器人学习与多模态大模型等方向研究，在扩散模型、机器人操作等领域取得了一系列代表性成果，如 Qwen-RobotManip, Discrete Diffusion VLA, AdaptDiffuser 等工作，在 CVPR、ICML、ICLR、RA-L 等国际顶级会议期刊发表 20 余篇论文，其中第一作者 7 篇。他本科毕业于浙江大学自动化专业，并获竺可桢荣誉学位。本科期间及博士阶段曾获得国家奖学金、香港政府博士奖学金、中国电子学会-腾讯博士生科研激励计划等多项荣誉。毕业后即将前往普林斯顿大学开展博士后研究。

报告题目：生成式模型驱动的通用具身操作策略与基模型

报告简介：本报告将围绕生成式模型驱动的通用具身操作展开，系统介绍以生成式方法实现通用具身策略，及其上的大规模基模型。报告人将从扩散策略与模仿学习出发，介绍利用生成模型实现策略自进化，实现对新任务与新环境的高效适应。在此上，本报告将探讨层次化技能学习、多模态感知融合以及 VLA 模型在复杂操作中的应用，展示其在灵巧操作与三维语义理解中的优势。最后报告人将介绍通用操作具身基模型的最新进展 Discrete Diffusion VLA 和 Qwen-RobotManip。通过报告人的一系列工作，系统揭示生成模型如何推动具身操作迈向更强感知与泛化。

青年学者简介



报告时间：7月26日(周日)16:20-16:50

报 告 人：张钰荻

汇报形式：线上

所在单位：埃因霍温理工大学

个人简介：张钰荻是Eindhoven University of Technology的博士生，由Mykola Pechenizkiy和Meng Fang共同指导。她从事强化学习与大语言模型交叉领域的研究，在ICML, NeurIPS, ICLR, AAAI, EMNLP等顶级会议上以一作身份发表研究。她目前是马克斯·普朗克智能系统研究所的访问学者，由Shiwei Liu博士指导。她曾在微软实习，于山东大学获得硕士和学士学位。

报告题目：Self-Evolve Agents with In-Distribution Optimization

报告简介： Large Language Models have emerged as powerful controllers for interactive agents, yet training them for reliable long-horizon decision making remains hard — a core difficulty being credit assignment under delayed, episodic rewards. We propose Q-Evolve, a self-evolving framework that unifies automatic process-reward labeling and policy learning within a principled in-distribution reinforcement learning paradigm. Each evolving iteration learns an in-distribution critic from a hybrid off-policy dataset that mixes expert demonstrations with agent-generated trajectories, stabilizing Bellman backups in sparse-reward settings via a weighted Implicit Q-Learning objective. The learned value function then derives step-wise process rewards through advantage estimation — dense, reliable supervision without environment backtracking or human annotation. Using these signals, we perform behavior-proximal policy optimization over the same data used for reward labeling, enabling iterative self-improvement without exacerbating distribution shift. On AlfWorld, WebShop, and ScienceWorld, Q-Evolve outperforms strong baselines in sample efficiency, robustness, and overall task performance.



报告时间：7月26日(周日)16:50-17:20

报 告 人：赵兴一

汇报形式：线上

所在单位：犹他州立大学

个人简介：赵兴一，美国犹他州立大学计算机科学博士研究生，师从 Shuhan Yuan 教授。主要研究方向为大语言模型安全（AI Security）与人工智能赋能网络安全（AI for Cybersecurity）。重点研究大模型后门攻击与防御、机器遗忘、对抗鲁棒性以及模型安全对齐；并致力于利用大语言模型的理解与推理能力，实现网络协议与技术规范的自动化安全分析、漏洞发现和安全测试。相关研究成果发表于 ICML、ICLR、EMNLP 和 ACM MobiCom 等国际顶级会议。

报告题目：大语言模型后门的生命周期：后门攻击持久性与可靠移除的研究

英文题目：The Lifecycle of LLM Backdoors: Backdoor Attack Persistence and Reliable Removal

报告简介： 随着大语言模型被广泛应用到各类任务中，后门风险也会在模型训练和微调过程中不断变化。本报告从攻击与防御两个视角，介绍我们在大语言模型后门攻击与移除的研究。在后门攻击任务中，后门容易受到下游微调引起的参数变化影响，导致攻击失败。针对这一问题，我们提出 BAD-BOOM，通过拓宽和平滑后门损失盆地，提高后门在不同微调设置下的鲁棒性。在后门移除任务中，传统的梯度上升虽然可以从表面上移除后门影响，但本质上是将后门现象转移，从而导致新的安全漏洞。针对这一现象，我们提出RGA，通过动态调节遗忘强度并约束参数偏移，在保持正常分类性能的同时，更可靠地消除模型地恶意决策。我们的研究表明，大语言模型后门安全需要贯穿模型训练、下游适配与安全修复的全过程；我们应当系统评估后门能否持续存在、如何发生变化，以及能否被真正移除。

As large language models are increasingly applied to a wide range of tasks, backdoor risks continue to evolve throughout model training and downstream fine-tuning. This talk presents our research on LLM backdoor attacks and removal from both offensive and defensive perspectives. For backdoor attacks, malicious behavior is often sensitive to parameter changes introduced by downstream fine-tuning, which may cause the attack to fail. To address this issue, we propose BAD-BOOM, which broadens and smooths the backdoor loss basin to improve attack robustness across different fine-tuning settings. For backdoor removal, conventional gradient ascent may appear to eliminate the original backdoor effect, while in fact merely shifting the malicious behavior and introducing new security vulnerabilities. To address this phenomenon, we propose RGA, which dynamically adjusts the unlearning strength and constrains parameter deviation. This enables more reliable removal of malicious model behavior while preserving normal classification performance. Our research shows that LLM backdoor security must be considered throughout the entire process of model training, downstream adaptation, and security remediation. Backdoor evaluation should systematically examine whether malicious behavior persists, how it changes, and whether it can be truly removed.

备 忘 录

备忘录

备忘录

中南大学简介:

中南大学坐落在中国历史文化名城——湖南省长沙市，占地面积317万平方米。中南大学是“985工程”重点建设高校、“211工程”首批建设高校、“2011计划”首批牵头高校。学校于2017年9月入选世界一流大学A类建设高校，2022年入选第二轮“双一流”建设高校。2023年，学校在软科世界大学学术排名中位居全球96位，国内第12位。

学校学科门类齐全，拥有完备的有色金属、医学、轨道交通等学科体系，涵盖哲学、经济学、法学、教育学、文学、理学、工学、医学、管理学、艺术学、交叉学科等11大学科门类，辐射军事学，学校是学位授权自主审核单位。现有一级学科国家重点学科6个，二级学科国家重点学科12个，国家重点（培育）学科1个，国家临床重点专科106个；设有31个二级学院104个本科招生专业；博士学位授权一级学科44个，硕士学位授权一级学科48个，博士专业学位授权类别14个，硕士专业学位授权类别31个，博士后科研流动站32个。

拥有一支以两院院士和国家级教学名师为核心、博士生导师为中坚、正副教授为骨干的雄厚师资队伍。现有两院院士16人，教授及相应正高级职称人员2161人，享受政府特殊津贴专家353人。现有国家级科研基地34个，入选中国智库索引来源智库研究机构数19个。2000年以来，学校共获国家科技三大奖119项，其中获国家科技一等奖（含特等奖、创新团队奖）20项，10个项目入选“中国高校十大科技进展”，入选首批国家知识产权示范高校。

计算机学院简介:

中南大学计算机学院拥有计算机科学与技术一级学科博士学位授权点，电子信息、生物与医药专业学位博士授权点，计算机科学与技术一级学科硕士学位授权点，电子信息硕士专业学位授权点，拥有计算机科学与技术博士后流动站，获批国家级首批特色化示范性软件学院。本科专业全面覆盖计算机及相关学科领域，包括：计算机科学与技术、软件工程、数据科学与大数据技术、信息安全。拥有国家级基础学科拔尖学生培养计划2.0基地，计算机科学与技术、软件工程、数据科学与大数据技术、信息安全为国家级一流建设专业。

学院师资力量雄厚，现有教职工177人，教授及高级职称47人，副高职称64人，其中包括国家级教学名师、国家杰青等国家级人才12人，湖南省科技创新领军人才、湖南省杰青等省级人才34人。

学院围绕“双一流”建设目标，坚持对接行业重大需求、促进学科融合，加强创新平台和团队建设，科技成果日渐丰硕，计算机学科进入ESI前1‰。学院现有“医疗大数据应用技术国家工程研究中心”等国家级研究基地2个，省部级研究基地18个；获得省部级科学技术奖20余项，授权发明专利近500项，在国内外重要学术期刊上发表论文近2000篇，承担国家自然科学基金、国家科技计划类项目近300项，省部级及企业委托项目累计300余项，科技成果转让费金额近4000余万。在算法理论与新型计算、具身智能与无人系统、图形图像与可视分析、网络优化与系统安全、大数据计算与智能软件、生物信息计算与数字健康等多个领域形成鲜明的研究特色，为人才培养和社会服务奠定了良好的研究基础。

会议须知

■ 会务组及联系方式

会务组及联络方式：

组长：冯启龙 18711153272

成员：文子豪 16681586006

■ 温馨提示

为保证会场秩序，会议期间，请您全程佩戴大会代表证参会；参会人员须提前15分钟进场，会场内手机调整静音或者震动状态，敬请谅解！

■ 酒店信息

入住酒店：廷泊酒店长沙中南大学地铁站店

酒店电话：0731—81820555

酒店地址：长沙市岳麓街道丰顺路252号惟盛园19栋

酒店退房：退房当天14点前

■ 交通路线



会议须知

■ 天气贴士

会议期间气温较高，请您做好个人防暑降温工作。

7月24日（周五）	27-37℃ 东南风2级	晴	
7月25日（周六）	27-38℃ 南风1级	晴	
7月26日（周日）	27-37℃ 东风1级	晴	