中国科学技术大学
University of Science and Technology of China

高水平学术前沿讲座

Differentially Private Training in the LLM Era

报 告 人： 王帝 阿卜杜拉国王科技大学

报告时间： 2026 年 05 月 13 日 15:00-16:00

报告地点： 高新校区 1 号学科楼 A544 会议室

主办单位： 信息科学技术学院　　几何计算与智能生物制药实验室

报告摘要：

Differentially private stochastic gradient descent (DP-SGD) has been extensively studied in recent years from a variety of perspectives. However, applying DP-SGD to large language model (LLM) training introduces several additional challenges. First, unlike the end-to-end training pipeline commonly used in classical deep learning, LLM development typically involves multiple training stages, each of which may impose different privacy requirements. Second, DP-SGD is difficult to scale to large models due to its inherent computational and memory overhead.

In this talk, Dr.Wang will present our recent work on addressing these challenges from both theoretical and systems perspectives. He will first discuss how to make DP-SGD scalable for LLM pretraining with little to no sacrifice in throughput or memory efficiency. He will then show how to enable scalable differentially private fine-tuning for extremely large LLMs under limited computational resources. Next, he will present theoretical limitations of differentially private alignment, and show how these barriers can be addressed through new structure-aware methods, together with corresponding lower-bound analyses. Finally, he will discuss future research directions along these lines.

报告人简介：

王帝博士现任阿卜杜拉国王科技大学（KAUST）计算机科学助理教授，并兼任统计学项目、生成式人工智能卓越中心及计算生物科学研究中心成员。他的研究方向为可信机器学习与生成式人工智能，重点关注机器学习模型中的隐私、安全、鲁棒性与可控记忆问题，具体包括差分隐私、机器遗忘、知识编辑与遗忘、对抗鲁棒性、可解释人工智能和概念擦除等。王博士于纽约州立大学布法罗分校获得计算机科学与工程博士学位，此前分别获得西安大略大学数学硕士学位和山东大学数学与应用数学学士学位。他的研究成果曾获得 USENIX Security 荣誉提名论文奖、和 ACML 最佳论文奖等认可。 他已在人工智能，机器学习，网络安全，数据库和理论计算机顶级期刊和会议例如 ICML，NeurIPS，ICLR，JMLR，IEEE S&P，USENIX，IEEE Transactions on Information Theory，SIGMOD，VLDB 等发表超过 170 篇论文。他多次担任国际顶级学术会议和期刊的领域主席和副主编，并且多次受邀于国际顶级学术会议和学术机构做报告。近些年王帝教授主持并且参与了超过 5000 万元的科研项目。

欢迎感兴趣的师生参加