



系列交叉论坛

第

66

期

ChatGLM：从千亿模型到ChatGPT的一点思考

2023年10月12日（周四）

19:30-21:00

报告人

唐杰教授

主持人

戴琼海院士

腾讯会议

ID: 937-635-937

Password: 101273



扫码观看直播

唐杰

清华大学计算机系教授，ACM/AAAI/IEEE Fellow。研究兴趣包括人工智能、知识图谱、数据挖掘、社交网络，最近专注大语言模型：ChatGLM（开源下载近1000万）等。曾获ACM SIGKDD Test-of-Time Award（十年最佳论文）、国家科技进步二等奖。

报告摘要

近年超大规模预训练模型取得突飞猛进的发展，OpenAI发布大规模自然语言预训练模型GPT-3/4，同时推出了AI聊天机器人ChatGPT，推理能力大大加强，能够回答非常复杂的推理问题，实现多模态融合。清华自研了GLM-130B中英双语模型，参数1300亿，算法采用自研的GLM，支持低成本、低资源使用，仅4张3090即可驱动，同时支持在NVIDIA、海光DCU、华为Ascend 910和神威Sunway处理器上进行训练与推理。本报告将介绍团队在研发GLM-130B过程中碰到的困难和收获，同时介绍对ChatGPT技术点的理解，分析下一步要实现ChatGPT需要走的路和面临的技术挑战。