



# CCIR 2023

## 第二十九届全国信息检索学术会议

The 29th China Conference on Information Retrieval, CCIR 2023

2023年11月24日~26日·北京



# Contents

<b>关于 CCIR</b>	<b>4</b>
CCIR2023 . . . . .	4
<b>会议组织</b>	<b>5</b>
<b>日程一览</b>	<b>6</b>
<b>会议程序</b>	<b>7</b>
11月23日日程安排 . . . . .	7
11月24日日程安排 . . . . .	7
11月25日日程安排 . . . . .	9
11月26日日程安排 . . . . .	11
<b>大会主旨报告</b>	<b>12</b>
报告一：AI大模型时代--新的机遇与挑战 . . . . .	12
报告二：Large Language Models for Information Retrieval: Opportunities and Challenges . . . . .	13
<b>大会主题报告</b>	<b>15</b>
报告一：工业界推荐系统十大挑战问题 . . . . .	15
报告二：大模型的发展与技术思考 . . . . .	16
<b>工业论坛报告</b>	<b>17</b>
报告一：大模型中文语料的采集 使用和风险 . . . . .	17
报告二：大语言模型在现代搜索引擎中的应用 . . . . .	17
报告三：原子大模型，从通才迈向专才 . . . . .	18
报告四：大规模类案检索系统-信息检索技术在司法场景的应用 . . . . .	18
报告五：中国移动“弈衡”大模型评测体系研究及技术发展趋势浅析 . . . . .	19
报告六：AI在短视频领域的机遇和挑战 . . . . .	19
报告七：基于可视化技术的数字孪生工厂实践及探索 . . . . .	20

<b>青年论坛报告</b>	<b>21</b>
报告一：认知驱动的文档检索技术 . . . . .	21
报告二：可信推荐系统的研究实践与思考 . . . . .	22
报告三：基于图神经网络的推荐系统定向注毒攻击方法 . . . . .	23
报告四：科技大模型“墨子”及其学术应用 . . . . .	24
<b>录用论文列表</b>	<b>25</b>
<b>赞助单位</b>	<b>30</b>

# 关于 CCIR

信息检索旨在满足人类在互联网上快速准确地获取信息与知识的需求，研究成果将支撑国家战略决策，推动互联网和人工智能领域的发展，提升整个社会的生产效率，并对社会生活各个领域产生重大影响。全国信息检索学术会议（CCIR）由中国中文信息学会（CIPS）举办，一路伴随着中国互联网产业的成长，是信息检索领域的旗舰会议。

## CCIR2023

第二十九届全国信息检索学术会议（The 29th China Conference on Information Retrieval, CCIR 2023）将于 2023 年 11 月 24-26 日在北京举行，此次会议由中国中文信息学会主办，由中国中文信息学会信息检索专委会、清华大学承办。本次会议与首届 ACM SIGIR-AP (Information Retrieval in the Asia Pacific) 会议联合举办。

会议包含一系列学术活动，除传统的海内外知名学者的大会报告、会议论文报告、Poster 交流、评测活动外，还将组织青年学者论坛以及面向热点研究问题的前沿讲习班等。大会也邀了信息检索国际顶级学术会议 SIGIR 2023 的部分论文作者进行报告交流。

# 会议组织

大会主席： 张敏（清华大学）  
窦志成（中国人民大学）  
王仲远（快手）

程序委员会主席： 殷大伟（百度）  
徐君（中国人民大学）

组委委员会主席： 马为之（清华大学）  
范意兴（中国科学院计算技术研究所）

宣传主席： 白婷（北京邮电大学）

出版主席： 陈旭（中国人民大学）

网络主席： 谢晓晖（清华大学）

青年论坛主席： 张鹏（天津大学）  
毛佳昕（中国人民大学）

评测主席： 刘伟东（中国移动通信研究院）  
朱小飞（重庆理工大学）  
何向南（中国科学技术大学）

赞助主席： 聂礼强（哈尔滨工业大学（深圳））  
李晨亮（武汉大学）

产业论坛主席： 李晨亮（武汉大学）

财务主席： 孙培杰（清华大学）

评奖主席： 黄萱菁（复旦大学）

# 日程一览

时间	时长	11月23日 星期四	
14:00-17:00	180分钟	Workshop - 快手	
19:00-21:00	120分钟	专委会工作会议	
时间	时长	11月24日 星期五	
8:30-8:50	20分钟	大会开幕式	
8:50-9:00	10分钟	合影	
9:00-10:00	60分钟	大会主旨报告 1	
10:00-10:40	40分钟	大会主题报告 1	
10:40-11:00	20分钟	茶歇	
11:00-12:00	60分钟	工业论坛	
12:00-13:30	90分钟	午休	
13:30-15:00	90分钟	大会最佳候选论文报告	
15:00-15:20	20分钟	茶歇	
15:20-17:00	100分钟	新委员论坛	
18:00-20:00	120分钟	晚宴	
时间	时长	11月25日 星期六	
8:30-9:30	60分钟	大会主旨报告 2	
9:30-10:10	40分钟	大会主题报告 2	
10:10-10:30	20分钟	茶歇	
10:30-12:00	90分钟	青年论坛	
12:00-13:30	90分钟	午休	
13:30-15:30	120分钟	国际顶会顶刊论文报告	
15:30-15:50	20分钟	茶歇	
15:50-16:50	60分钟	评测报告会	
16:50-17:20	30分钟	大会闭幕式	
时间	时长	11月26日 星期日 专题讲习班（与 SIGIR-AP 合办）	
8:30-11:45	195分钟	专题报告 1	专题报告 2
11:45-13:00	75分钟	午休	
13:00-16:30	210分钟	专题报告 1	专题报告 2

# 会议程序

## 11月23日 星期四

14:00-17:00	<b>Workshop - 快手</b> 地点：北京西郊宾馆银杏大厅（一号楼三层）
19:00-21:00	<b>专委会工作会议</b> 地点：北京西郊宾馆金缘厅（5号楼二层北侧）

## 11月24日 星期五

地点：北京西郊宾馆银杏大厅（一号楼三层）

<b>大会开幕式（主持人：张敏）</b>	
8:30-8:50	1. 中文信息学会领导致辞 2. 信息检索专委会领导致辞 3. 程序委员会致辞
8:50-9:00	合影
<b>大会主旨报告 1（主持人：张敏）</b>	
9:00-10:00	AI 大模型时代--新的机遇与挑战，李航，字节跳动
<b>大会主题报告 1（主持人：徐君）</b>	
10:00-10:40	推荐系统的十个挑战性问题，董振华，华为诺亚实验室
10:40-11:00	茶歇
<b>工业论坛（主持人：李晨亮，聂礼强）</b>	
11:00-11:10	1. 大模型中文语料的采集使用和风险，梁斌，八友科技
11:10-11:20	2. 大语言模型在现代搜索引擎中的应用，王帅强，百度
11:20-11:30	3. 原子大模型，从通才迈向专才，张俊祺，原子回声
11:30-11:40	4. 大规模类案检索系统-信息检索技术在司法场景的应用，杜斟，航天国政
11:40-11:50	5. “弈衡” 大模型评测体系研究及技术发展趋势浅析，刘伟东，中国移动

11:50-12:00	6. AI 在短视频领域的机遇和挑战, 宋洋, 快手
12:00-12:10	7. 基于可视化技术的数字孪生工厂实践及探索, 彭海军, 首亨软件
12:10-13:30	午休
<b>大会最佳候选论文报告 (主持人: 殷大伟, 徐君)</b>	
13:30-13:45	1. 李昆泽, 张宇. 自适应的流水线式无监督问题生成方法
13:45-14:00	2. 吕由钢, 郝继泰, 王梓涵, 高莘, 任鹏杰, 陈竹敏, 马军, 任昭春. 层次化推理链驱动的多被告法律判决预测
14:00-14:15	3. 魏屹芃, 林原, 鄂福顺, 相天雨, 王英霖, 许侃, 林鸿飞. 基于比较学习的多模态情感分析方法
14:15-14:30	4. 叶扬帆, 冯骁骋, 冯夏冲, 于鑫淼, 秦兵. 基于大型语言模型的无监督细粒度长文本会议摘要框架
14:30-14:45	5. 张兴, 郭宇璨, 靳小龙, 官赛萍. 论元角色多指称的主题事件论元抽取
14:45-15:00	6. 阎世宏, 马为之, 张敏, 刘奕群, 马少平. AID-GCN: 基于图卷积网络推荐系统自适应交互去噪
15:00-15:20	茶歇
<b>新委员论坛 (主持人: 兰艳艳)</b>	
15:20-15:45	刘慧婷 (安徽大学), 毛先领 (北京理工大学), 吴书 (中科院自动化所), 陈晋鹏 (北京邮电大学), 高小燕 (北京工业大学)
15:45-16:10	郭磊 (山东师范大学), 洪鑫 (清华大学), 侯旻 (合肥工业大学), 李辉 (厦门大学), 廉涛 (太原理工大学)
16:10-16:35	刘海峰 (南京师范大学), 刘正皓 (东北大学), 刘政 (智源人工智能研究院), 王翔 (中国科学技术大学), 吴之璟 (北京理工大学)
16:35-17:00	谢若冰 (腾讯), 俞蔚捷 (对外经贸大学), 张孟奇 (山东大学), 朱余韬 (中国人民大学), 邹立新 (武汉大学)
18:00-20:00	晚宴, 地点: 全聚德 (清华园店)

# 11月25日 星期六

地点：北京西郊宾馆银杏大厅（一号楼三层）

大会主旨报告 2（主持人：窦志成）	
8:30-9:30	Large Language Models for Information Retrieval: Opportunities and Challenges, 翟成祥，美国伊利诺尔大学香槟分校
大会主题报告 2（主持人：殷大伟）	
9:30-10:10	大模型的发展与技术思考，王小川，百川智能
10:10-10:30	茶歇
青年论坛（主持人：张鹏，毛佳昕）	
10:30-10:45	1. 认知驱动的文档检索技术，吴之璟，北京理工大学
10:45-11:00	2. 可信推荐系统的研究实践与思考，马为之，清华大学
11:00-11:15	3. 基于图神经网络的推荐系统定向中毒攻击方法，白婷，北京邮电大学
11:15-11:30	4. 科技大模型“墨子”及其学术应用，毛先领，北京理工大学
11:30-12:00	5. Panel
12:00-13:30	午休
国际顶会顶刊论文报告（主持人：何向南，王宏宁）	
13:30-13:45	1. Alleviating Matthew Effect of Offline Reinforcement Learning in Interactive Recommendation. <b>Chongming Gao</b> , Kexin Huang, Jiawei Chen, Yuan Zhang, Biao Li, Peng Jiang, Shiqi Wang, Zhong Zhang, and Xiangnan He
13:45-14:00	2. Triple Structural Information Modelling for Accurate, Explainable and Interactive Recommendation. <b>Jiahao Liu</b> , Dongsheng Li, Hansu Gu, Tun Lu, Peng Zhang, Li Shang, and Ning Gu
14:00-14:15	3. When Search Meets Recommendation: Learning Disentangled Search Representation for Recommendation. <b>Zihua Si</b> , Zhongxiang Sun, Xiao Zhang, Jun Xu, Xiaoxue Zang, Yang Song, Kun Gai, and Ji-Rong Wen
14:15-14:30	4. Large Language Models are Versatile Decomposers: Decomposing Evidence and Questions for Table-based Reasoning. Yunhu Ye, <b>Binyuan Hui</b> , Min Yang, Binhu Li, Fei Huang, and Yongbin Li
14:30-14:45	5. ML-LJP: Multi-Law Aware Legal Judgment Prediction. <b>Yifei Liu</b> , Yiquan Wu, Yating Zhang, Changlong Sun, Weiming Lu, Fei Wu, and Kun Kuang

14:45-15:00	6. Adapting Generative Pretrained Language Model for Open-domain Multimodal Sentence Summarization. <b>Dengtian Lin</b> , Liqiang Jing, Xuemeng Song, Meng Liu, Teng Sun, and Liqiang Nie
15:00-15:15	7. T2Ranking: A Large-scale Chinese Benchmark for Passage Ranking. Xiaohui Xie, <b>Qian Dong</b> , Bingning Wang, Feiyang Lv, Ting Yao, Weinan Gan, Zhijing Wu, Xiangsheng Li, Haitao Li, Yiqun Liu, and Jin Ma
15:15-15:30	8. MR2: A Benchmark for Multimodal Retrieval-Augmented Rumor Detection in Social Media. <b>Xuming Hu</b> , Zhijiang Guo, Junzhe Chen, Lijie Wen, and Philip S. Yu
15:30-15:50	茶歇
<b>评测报告会（主持人：刘伟东，朱小飞，何向南）</b>	
15:50-15:55	1. 评测总体情况介绍
15:55-16:05	2. 赛题一介绍
16:05-16:20	3. 赛题一冠军团队报告
16:20-16:30	4. 赛题二介绍
16:30-16:45	5. 赛题二冠军团队报告
16:45-16:50	6. 结束语
<b>大会闭幕式（主持人：刘奕群）</b>	
16:50-17:20	1. 最佳论文颁奖
	2. 评测颁奖
	3. 宣布下一届举办地
	4. 大会总结

# 11月26日 星期日（与 SIGIR-AP 合办）

地点：北辰洲际酒店二层 Room A 和 Room B

Session1 (Room A)	
8:30-11:45	<b>User Simulation for Evaluating Interactive Information Access Systems</b> Chengxiang Zhai, University of Illinois at Urbana-Champaign
Session2 (Room B)	
8:30-11:45	<b>Rethinking Conversational Agents in the Era of LLMs: Proactivity, Non-collaborativity, and Beyond</b> Yang Deng, National University of Singapore Wenqiang Lei, Sichuan University Minlie Huang, Tsinghua University Tat-Seng Chua, National University of Singapore
11:45-13:00	午休
Session3 (Room A)	
13:00-16:30	<b>Recent Advances in Generative Information Retrieval</b> Yubao Tang, CAS Key Lab of Network Data Science and Technology, ICT, CAS; University of Chinese Academy of Sciences Ruqing Zhang, CAS Key Lab of Network Data Science and Technology, ICT, CAS; University of Chinese Academy of Sciences Jiafeng Guo, CAS Key Lab of Network Data Science and Technology, ICT, CAS; University of Chinese Academy of Sciences Maarten de Rijke, University of Amsterdam
Session4 (Room B)	
13:00-16:30	<b>Large Language Models for Recommendation: Progresses and Future Directions</b> Kejin Bao, University of Science and Technology in China Jizhi Zhang, University of Science and Technology in China Yang Zhang, University of Science and Technology in China Wenjie Wang, National University of Singapore Fuli Feng, University of Science and Technology in China Xiangnan He, University of Science and Technology in China

# 大会主旨报告

## 报告一：AI 大模型时代--新的机遇与挑战

讲者：李航，字节跳动

时间：11 月 24 日 9:00-10:00



**个人介绍** 李航，字节跳动科技有限公司研究部门负责人。ACM Fellow、ACL Fellow, IEEE Fellow。京都大学毕业，东京大学博士。曾就职于 NEC 公司中央研究所，微软亚洲研究院，华为技术有限公司诺亚方舟实验室。主要研究方向自然语言处理、信息检索、机器学习、数据挖掘。

**报告摘要** ChatGPT 的问世揭开了人工智能历史的新篇章，也孕育着人们未来生活和工作的巨大变化。在这个报告中，我将介绍自己对 AI 大模型，特别是语言大模型的观察和思考。总结语言大模型技术的特点，介绍自己对面向未来的一些研究课题的看法。重点讨论语言大模型与心智语言，语言大模型与多模态处理、语言大模型与逻辑推理等问题。AI 大模型的新时代开始了，我们需要在已有工作的基础上，定义新问题，研究新方法，开发新技术，开启 AI 研究的新征程。

## **报告二：Large Language Models for Information Retrieval: Opportunities and Challenges**

**讲者：翟成祥，美国伊利诺尔大学香槟分校**

**时间：**11 月 25 日 8:30-9:30



**个人介绍** ChengXiang Zhai (<http://czhai.cs.illinois.edu>) is a Donald Biggar Willett Professor in Engineering of the Department of Computer Science at the University of Illinois at Urbana-Champaign (UIUC), where he is also affiliated with School of Information Sciences, Department of Statistics, and the Carl R. Woese Institute for Genomic Biology. He received a Ph.D. in Computer Science from Nanjing University in 1990, and a Ph.D. in Language and Information Technologies from Carnegie Mellon University in 2002. He worked at Clairvoyance Corp. as a Research Scientist and a Senior Research Scientist from 1997 to 2000. The general goal of his research is to build intelligent systems to leverage big data, especially big text data, to augment human intelligence and support complex user tasks in many domains such as healthcare, education, and scientific discovery. His specific research interests include intelligent task agents, computational models of users, human-like natural language processing, and intelligent infrastructure for human-AI collaboration. He has over 400 publications in these areas with over 40,000 citations and an h-index of 92 in Google Scholar. He holds 5 US patents. He is America Editor of the Springer Information Retrieval Book Series and a Senior Associate Editor of ACM Transactions on Intelligent Systems and Technology. He served as a program co-chair of ACM SIGIR'09, WWW'15, and as a conference co-chair of CIKM'16, WSDM'18, and IEEE BigData'20. He is an ACM Fellow and a member of the ACM SIGIR Academy, and received numerous awards, such as ACM SIGIR Gerard Salton Award, ACM SIGIR Test of Time Award (three times), the 2004 Presidential Early Career Award for Scientists and Engineers (PECASE), Alfred P. Sloan Research

Fellowship, IBM Faculty Award, HP Innovation Research Award , UIUC Rose Award for Teaching Excellence, and UIUC Campus Award for Excellence in Graduate Student Mentoring. He has graduated 41 PhD students and over 50 MS students.

**报告摘要** Recent years have seen great success of large language models (LLMs) in performing many natural language processing tasks with impressive performance, especially tasks that directly serve users such as question answering and text summarization. While statistical language models have been applied to information retrieval since many decades ago, these new large language models go far beyond traditional language models in their representation learning capacity, which enabled them to both understand natural language semantically and generate fluent meaningful natural language text. As such, they open up unprecedented opportunities for transforming information retrieval research and applications. However, concerns such as hallucination undermine their trustworthiness, limiting their actual utility when deployed in real-world applications, especially high-stake applications where trust is vital. In this talk, I will systematically examine the opportunities and challenges that LLMs have created for information retrieval research and applications. Specifically, I will address the following questions: 1) How can LLMs be leveraged to improve the current search engines? 2) What is the potential for LLMs to transform the future information retrieval applications? 3) What are the major challenges in applying LLMs to information retrieval? 4) Given the anticipated growth of LLMs, what will future information retrieval systems look like?

# 大会主题报告

## 报告一：工业界推荐系统十大挑战问题

讲者：董振华，华为诺亚方舟实验室，dongzhenhua@huawei.com

时间：11月24日 10:00-11:00

**个人介绍** 董振华博士是华为诺亚方舟实验室技术专家、项目经理，其目前的研究兴趣是推荐系统、因果推断、信息检索及机器学习应用。他所领导的研究团队将推荐系统、机器学习、因果推断等技术落地于应用市场、信息流、广告、即时服务、音乐、视频等多个产品，显著提升业务效果。基于以上实际的场景和数据，董振华博士申请了超过 40 项专利，并在 TKDE, TOIS, TORS, SIGIR, RecSys, KDD, WEB, AAAI, CIKM 等期刊、会议发表推荐系统、信息检索、因果推断相关论文 60 余篇，并有译著中文版《奇点临近》。他还做为 RecSys24 industry chair, RecSys22/23 sponsor chair, PC, SPC 服务于 RecSys, SIGKDD, SIGIR, WSDM, CIKM 等研究社区。董振华博士本科毕业于天津大学，博士毕业于南开大学，期间在明尼苏达大学 GroupLens 实验室作为访问学者交流学习 1 年。



**报告摘要** 从 2013 年开始，华为诺亚方舟实验室帮助很多产品构建推荐系统和搜索引擎，帮助用户在合适的时间、地点、场景获取满足其个性化需求的信息、服务。我们的推荐系统每天都为数以亿计的用户提供新闻、歌曲、视频、图书、应用、主题壁纸等个性化内容服务。海量的数据和多元的场景也为我们提供了研究前沿推荐系统技术的机会。与此同时，我们也见证了推荐系统过去 10 年的技术趋势：从简单、浅层的协同过滤、线性模型、低秩模型到复杂、深层的神经网络、预训练模型。基于以上使命、机会和技术趋势，我们在推荐系统的研究和落地过程中发现了很多非常难的问题。在本次演讲中，我们将分享 10 个重要、有趣的推荐系统挑战难题，并期望有机会与推荐系统研究社区的同仁一起合作，解决以上难题，创建更好的推荐系统。

## 报告二：大模型的发展与技术思考

讲者：王小川，百川智能

时间：11月25日 9:30-10:30



**个人介绍** 王小川，百川智能创始人兼CEO，清华大学计算机科学与技术专业工程博士，清华大学天工智能计算研究院联席院长。高中时期用吴文俊消元法，首次在微型机下完成初等几何命题的全部证明。曾获国际信息学奥林匹克竞赛（IOI）金牌、北京市科学技术奖一等奖、中国青年科技奖、中国电子学会科学技术奖科技进步奖一等奖、钱伟长中文信息处理科学技术一等奖。

**报告摘要** 大模型正引领我们走进智能时代，其引发的变革将超越工业革命、电力革命和信息革命所引起的变革总和。大模型就像互联网时代提供信息化基础的芯片一样，会在社会各个方面“点亮智能”。语言模型是AGI的起点，因为大模型将语言变成数学模型，掌握了语言背后的知识、思考、沟通和文化，大模型不仅将重塑现有行业，更将创造全新价值。虽然大模型展现出了强大的通用性和泛化能力，但现在的大模型还有很多不足，比如幻觉问题、时效性问题、安全问题等。其中很多问题可以通过搜索技术来解决。搜索技术不仅可以从超大规模数据获取和清洗、算法与工程优化等方面助力大模型研发，还能够减少幻觉的生成，有效提升模型的时效性、安全性。

# 工业论坛报告

## 报告一：大模型中文语料的采集 使用和风险

讲者：梁斌，八友科技

时间：11月24日 11:00-11:10

**报告摘要** 大模型在2023年2月底在中国飞速发展，短短半年时间内涌现出了超过100家大模型团队，我们通过8年数据采集工作的积累服务了其中的50多家，并拜访了其中大部分团队，在服务过程中，加深理解了大模型团队在数据方面的需求和使用方法。通过本报告主要介绍关于大模型数据的采集，使用及其整个过程涉及的红线和风险。



## 报告二：大语言模型在现代搜索引擎中的应用

讲者：王帅强，百度

时间：11月24日 11:10-11:20

**报告摘要** 近年来，基于预训练的大语言模型目前已经成为自然语言处理和人工智能等领域最令人瞩目的研究方向，并且在学术界和工业界都引发了广泛的关注。现代搜索引擎是自然语言处理和人工智能技术最重要的应用场景，大语言模型已经在当前的搜索系统迭代和产品优化中大规模落地，并大幅提升了搜索效果。本报告将具体介绍大语言模型在现代商用搜索引擎不同阶段中的一系列相关应用和探索，最后探讨未来大语言模型在信息检索领域值得关注的研究和应用方向。



## 报告三：原子大模型，从通才迈向专才

讲者：张俊祺，原子回声

时间：11月24日 11:20-11:30

**报告摘要** 原子大模型由原子回声自主研发，包括 Atom-1B、Atom-7B 和 Atom-13B 三个通用基础模型，在语言理解、文本生成、任务执行等方面达到国内先进水平。同时构建了智能体标准化定制平台 AtomHub，基于个人和企业的专有知识库，通过检索增强与知识注入等技术实现模型快速定制，让每位用户都可以拥有自己的专才大模型。



## 报告四：大规模类案检索系统-信息检索技术在司法场景的应用

讲者：杜斟，航天国政

时间：11月24日 11:30-11:40

**报告摘要** 类案检索对于全面准确落实司法责任制和统一法律适用具有十分重要的意义，本次报告主要介绍在一亿审判文书规模上的类案检索系统研发及效果，该系统主要特点是同时支持基于关键词和向量的检索，提出了 CaseFormer 模型，使得检索效果更加精准高效。



## 报告五：中国移动“弈衡”大模型评测体系研究及技术发展趋势浅析

讲者：刘伟东，中国移动

时间：11月24日 11:40-11:50

**报告摘要** 中国移动作为国内领先的信息服务企业，为了积极致力于推进大模型技术的健康发展，我们深入开展了“弈衡”大模型评测体系研究，旨在建立全面、客观和中立的评测标准，针对业界主流的20余款大模型进行了深入的评测验证，从多维度揭示了目前主流大模型的主要挑战和发展趋势，助推大模型技术发展与产业成熟。

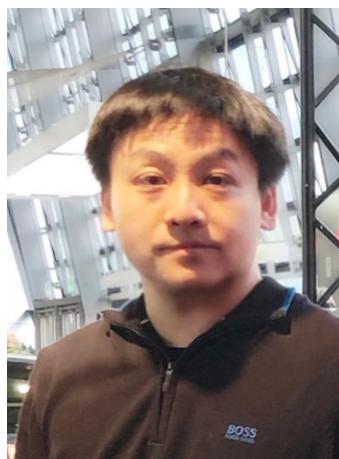


## 报告六：AI在短视频领域的机遇和挑战

讲者：宋洋，快手

时间：11月24日 11:50-12:00

**报告摘要** 快手AI前沿技术探索及合作布局，包括快手在内容生产，内容理解，内容分发三大方面的前沿技术成果和未来规划，重点展开快手作为主流的短视频平台在推荐系统方面的挑战，实战和探索，涉及多个业界首次落地的技术方案和首次公开的工业界推荐数据集，在学术界和工业界都获得了广泛的影响力。



## 报告七：基于可视化技术的数字孪生工厂实践及探索

讲者：彭海军，首亨软件

时间：11月24日 12:00-12:10

**报告摘要** 利用自主研发的数字孪生可视化应用平台，帮助烟厂、军工及智能制造场景在虚机空间创建物理世界的数字映射，并借助数、算法等手段，实现对物理世界的感知、诊断、预测释放数据价值，辅助业务决策。



# 青年论坛报告

## 报告一：认知驱动的文档检索技术

讲者：吴之璟，北京理工大学

时间：11月25日 10:30-10:45



**个人介绍** 吴之璟，北京理工大学计算机学院助理教授、特别副研究员，硕士生导师，主要研究方向为信息检索、自然语言处理。2022年于清华大学获得博士学位，曾前往新加坡国立大学、皇家墨尔本理工大学进行长期访问交流。在 TOIS、SIGIR、TheWebConf、WSDM、CIKM 等国际期刊与会议上发表论文十余篇。

**报告摘要** 文档检索是信息检索领域的关键技术。结合用户认知行为设计检索模型有助于构建更智能、更符合用户需求的检索系统，提高用户的搜索体验和信息获取效率。本报告将介绍用户认知行为驱动的文档检索模型，并分享我们在借助大语言模型识别关键认知内容上的一些探索。

## 报告二：可信推荐系统的研究实践与思考

讲者：马为之，清华大学

时间：11月25日 10:45-11:00



**个人介绍** 马为之，清华大学智能产业研究院助理研究员，曾于清华大学计算机系获得学士、博士学位及从事博士后工作，主要研究方向为智能信息获取、智慧医疗等。已发表 50 余篇高水平国际会议/期刊论文，担任 ACM TOIS 助理编辑（CCF A 类国际期刊）。曾获钱伟长中文信息处理科学技术一等奖，以及 SIGIR 2020 最佳论文提名等多项国内外论文奖励，先后入选清华大学“水木学者”计划和中国科协“青年人才托举工程”。

**报告摘要** 推荐系统作为智能信息获取的重要手段，被广泛应用在信息流、电子商务等场景，极大提升了海量用户的信息获取效率。现有推荐系统通常侧重结果的精准性优化，对于平台上内容消费者和内容生产者的可控性、公平性等复杂需求关注不足。团队针对相关领域开展了一系列研究，本报告将主要介绍在公平性和高质量推荐上的相关工作，首先将简要回顾发表在 TOIS 上的公平性综述论文；随后将从如何从内容侧寻求展示机会与效能（utility）的公平推荐，以及如何对于推荐交互实现高质量优化驱动的引导式可信推荐系统两个方面介绍近期工作。

# 报告三：基于图神经网络的推荐系统定向注毒攻击方法

讲者：白婷，北京邮电大学

时间：11月25日 11:00-11:15



**个人介绍** 白婷，北京邮电大学计算机学院讲师，2019年于中国人民大学获得博士学位，专注于个性化推荐系统研究。在国际重要会议与期刊发表论文30余篇，在国际会议CIKM, KDD DLP-Workshop中获得两项最佳论文/最佳论文候选奖；担任国际会议IEEE ICDM, IEEE ICKG, SIGIR-AP领域主席；2020年获评中国中文信息学会“优秀博士学位论文奖”，2022年获得中国电子学会“科技进步一等奖”，2023年入选百度首届“全球AI华人女性青年学者”榜单。

**报告摘要** 基于图神经网络（Graph Neural Networks，简称 GNNs）的推荐系统在面对注毒攻击时表现出脆弱性。注毒攻击是指攻击者通过注入虚假信息干扰推荐系统，对用户产生同质化的影响，使得注入的虚假信息被尽可能多的用户进行点击。由于用户不会对其不感兴趣的推送物品进行点击，因此这种无差别的注毒攻击方式会导致资源浪费。因此，我们根据目标用户的兴趣偏好，进行有针对性的注毒攻击研究，实现针对特定目标用户群体的精确攻击，并最小化对非目标用户的影响。论文提出了一种基于图神经网络的自动定向注毒攻击框架 AutoAttack，该框架不仅能够自动构建注入虚假用户节点的特征，还能够自动生成虚假用户节点与现有物品的连边，达到为特定目标用户推荐目标物品的攻击目的。为了评估 AutoAttack 框架在推荐系统上的攻击性能，我们选取了不同的 GNN 推荐算法，并在四个真实数据集上进行了大量实验验证，实验证明了 AutoAttack 框架在基于图神经网络推荐系统中应用的有效性和灵活性。

## 报告四：科技大模型“墨子”及其学术应用

讲者：毛先领，北京理工大学

时间：11月25日 11:15-11:30



**个人介绍** 毛先领，北京理工大学副教授，博导。主要研究信息检索与数据挖掘。目前担任计算机学会中文信息技术专委会委员、中文信息学会青工委执委以及语言与知识专委会委员；已在 SIGIR、WWW、SIGMOD、ACL、TOIS、TKDE 等国际期刊会议上发表 50 余篇论文；部分成果获 2 项最佳论文奖和 3 项省部级科技进步奖。

**报告摘要** 科学技术是生产力发展的核心动力，承载人类最前沿科技知识的论文数据也正在指数级快速增长。围绕海量科技数据与研究人员有限时间之间的矛盾，课题组研发了科技大模型“墨子”，并将其应用于前期研发的跨模态科技搜索引擎、科技知识图谱、科技论文阅读器和科技对话系统等学术产品中，希冀提升科研工作者的科研效率。

# 录用论文列表

1. 一种注意力引导知识增强的事件因果关系识别方法; Bo Xu, Jinchen Sun, Hongfei Lin and Linlin Zong.
2. App 搜索场景下的意图推荐方法研究; Yuqi Zhou, Sunhao Dai, Chaoliang Zhang, Gang Wang, Baotian Hu and Jun Xu.
3. 自适应的流水线式无监督问题生成方法; 李昆泽 and 张宇.
4. Opinion Evolution Model based on Long-Term and Short-Term Opinions; Yitong Wang, Xianyong Li, Yuhang Cheng, Yajun Du, Xiaoliang Chen and Yongquan Fan.
5. 层次化推理链驱动的多被告法律判决预测; 吕由钢, 郝继泰, 王梓涵, 高莘, 任鹏杰, 陈竹敏, 马军 and 任昭春.
6. 基于因果常识的事件共指消解技术研究; 王禹琦 and 张宇.
7. 结合局部与全局信息关联的相关工作生成; Kaixin Niu, Zhongqing Wang, Hongling Wang and Mengling Han.
8. 基于多模态特征融合的对话回复生成方法; Linqin Li, Dong Zhang, Suyang Zhu and Shoushan Li.
9. 一种融合位置信息和交互注意力的方面级情感分析方法; Jiajing Li, Sheng Li, Yuanyuan Dai, Tao Meng, Xiaoqing Luo and Hongfei Yan.
10. 基于稳定学习的多兴趣序列推荐网络; 刘昭呈, 朱振熙 and 刘强.
11. 面向排序学习的概率分布优化模型研究; Fengxu Zhao, Jian Wang, Yuan Lin and Hongfei Lin.

12. 基于对抗型排序学习的混合推荐算法; Kan Xu, Xinzhuo Wu, Yuan Lin, Qian Gu, Hongfei Lin and Zhang Xie.
13. Multimodal sentiment analysis method based on comparative learning ; Yipeng Wei, Yuan Lin, Fushun E, Tianyu Xiang, Yinglin Wang, Kan Xu and Hongfei Lin.
14. 上下文感知增强的多轮个性化对话检索方法研究; Yanbing Chen and Lin Li
15. 基于大型语言模型的无监督细粒度长文本会议摘要框架; Yangfan Ye, Xiaocheng Feng, Xiachong Feng, Xinmiao Yu and Bing Qin
16. Gender Biased Legal Case Retrieval System on Users' Decision Process; Ruizhe Zhang, Qingyao Ai, Yiqun Liu, Yueyue Wu and Beining Wang
17. 基于提示增强原型网络的小样本多标签方面类别检测; Bai Yu, Chaofeng Guan and Xianlei Zhou
18. 基于提示学习的电信网络诈骗案件分类方法; Jie Ji, Chengjie Sun, Lili Shan, Boyue Shang and Lei Lin
19. Matrix Product Operator based Sequential Recommendation Model; Peiyu Liu, Bowen Yao, Ze-Feng Gao and Wayne Xin Zhao
20. 基于自监督的预训练在推荐系统中的研究综述; 杨纪元, 马沐阳, 任鹏杰, 陈竹敏, 任昭春, 辛鑫, 蔡飞 and 马军
21. Multi-View Based Method for Tourism Similar Question Matching; Hao Wang, Yang Li, Suge Wang, Jian Liao and Jianxing Zheng
22. 基于多提示学习的方面类别情感分析方法; 刘锦行, 李琳 and 吴任伟
23. 基于表示对齐增强和外部知识融合的低资源跨语言文本检索; Yanbin Wang, Mao Cunli, Yongbing Zhang, Yuxin Huang, Shengxiang Gao and Ran Song

24. 基于情感分类的可解释证据提取方法; Shan He, Dongfang Li, Baotian Hu and Jindi Yu
25. Multimodal Conversation Emotion Recognition Based on Clustering and Group Normalization; Luo Qi and Gou Gang
26. 面向域外说话人适应场景的多层次解耦个性化语音合成; Shengxiang Gao, Shangbin Mo, Yuanzhang Yang, Zhengtao Yu, Linqin Wang and Ling Dong
27. Fine-grained Intent Recognition from Pediatric Medical Dialogues with Contrastive Learning; Wenbo Li, Qing Dong, Chao Liu and Qi Zhang
28. 基于多特征融合的中文医疗关系抽取; Dandan Zhao, Zhihao Zhang, Jiana Meng, Wen Su, Yingchun Long and Junpeng Zhang
29. 基于弱监督学习的稠密向量检索模型; Yi Jia and Jiaxin Mao
30. 基于位置感知时空图卷积网络的交通流量预测; 王子彤 and 李晨亮
31. A Survey of Dimensionality Reduction and Retrieval Algorithms for High Dimensional Data; Wei Shao, Gaoyu Zhu, Lei Yu and Jiafeng Guo
32. 面向图神经网络表征学习的类别知识探针; 黄兴宇, 赵明宇 and 吕子钰
33. 基于细粒度问题类型与答案一致性校验的医学视觉问答; 舒婷戴, 利军刘 and 小兵杨
34. 融合量子干涉信息的双重特征文本表示模型; Hui Gao, Peng Zhang and Jing Zhang
35. 论元角色多指称的主题事件论元抽取; Xing Zhang, Yucan Guo, Xiaolong Jin and Saiping Guan
36. 基于 EEG 的短视频用户情感倾向研究; Bangde Du, Qingyao Ai, Yiqun Liu and Ziyi Ye
37. 基于生物实体关系感知的图神经排序; Yichen He, Xiaofeng Liu and Shoubin Dong

38. 基于领域对比自适应模型的大学生焦虑心理分析; Weiwei Zhu, Yijia Zhang, Guantong Liu, Mingyu Lu and Hongfei Lin
39. 基于内在奖励的强化学习推荐探索策略; Yuanqing Yu, Weizhi Ma and Min Zhang
40. 融合依存句法分析的言论抽取方法; da Li, Xiangyu Zhang, Xichou Zhu, Lei Yu and Jiafeng Guo
41. AID-GCN: 基于图卷积网络的推荐系统自适应交互去噪; Shihong Yan, Weizhi Ma, Min Zhang, Yiqun Liu and Shaoping Ma
42. 一种基于核心论元的篇章级事件抽取方法; Chengjie Sun, Zongwei Li, Lili Shan and Lei Lin
43. 多尺度差分算子收敛语义的命名实体识别方法; Caiwei Yang, Yanping Chen, Yongbin Qin and Ruizhang Huang
44. Contrastive Dynamic Graph for Sequential Recommendation; Yu Cui, Jiawei Chen and Can Wang
45. 结合一阶逻辑的弱监督中文命名实体识别方法; Xi Tang and Dongchen Jiang
46. 一种融合词性编码信息的中文拼写校正模型; 孙胜 and 王鹏鸣
47. 基于融合特征多重异构网络的新浪微博社交机器人检测方法; Zhang Huaibo, Liu Xiaona, Feng Haoyuan, Yin Zhiyi, Shen Huawei and Liu Xin
48. SSHGCN: 基于音形异构图卷积的中文纠错方法; 任俊 and 黄瑞章
49. 中文去毒任务的研究; Jiangsheng Liu, Jiali Zuo and Yuting Hu
50. 基于动态邻居选择的知识图谱事实错误检测方法; Liang Gui and Shizhu He
51. 基于二维增强融合机制的事件论元抽取; Luxiang Wang, Yanping Chen, Hui Huang, Ruizhang

52. 基于盒嵌入的无监督任务型对话系统中的域外意图检测 Unsupervised OOD Detection in Task-Based Dialogue Systems using Box Embedding; Heyang Xu, Pei Wang and Yutao Mou
53. 构建解释增强的对话式推荐系统; 郭书宇, 孙维纬, 任鹏杰, 陈竹敏, 辛鑫, 马军 and 任昭春
54. 大语言模型作为高效的零样本排序器; Zheng Chen, Weiwei Sun, Xinyu Ma, Lingyong Yan, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin and Zhaochun Ren
55. 融合多重特征的噪声网络对齐方法; 咸宁, 范意兴, 郭嘉丰 and 廉涛
56. MVAE: 用于反讽检测的多维注意力变分自编码器; 刁宇峰, 杨亮, 李世琪, 樊小超 and 林鸿飞

# 赞助单位



