

# 李建 简历

☎ (+86) 13541051104

✉ lijian-jack@qq.com

🌐 jack57lee.github.io

## 教育背景

香港中文大学 (The Chinese University of Hong Kong)

中国香港, 2015.8–2020.6

博士学位, 计算机科学与工程, 导师: 吕荣聪 (Michael R. Lyu) 教授

研究方向: 自然语言处理与信息检索

斯坦福大学 (Stanford University)

美国旧金山, 2019.7–2020.3

访学博士, 计算机科学, 导师: Monica S. Lam 教授

电子科技大学 (UESTC)

中国成都, 2011.9–2015.6

学士学位, 电子信息工程, 排名 2/360, 两获国家奖学金

## 工作经历

腾讯公司–AI Lab/数据平台部–NLP 高级研究员

2022.3–至今

工作主题: 预训练大模型, 数字人对话, 广告算法

近三次绩效: 5 星 / 5 星 / 4 星

工作内容:

1. 基于 LLM 的数字人对话: (a) 数据建设, 基于 LLM 搭建训练数据构造管线, 可配置可拓展, 产出符合人设、世界观、多场景的多轮对话数据。(b) 检索增强 RAG, 将 NPC 人设背景、游戏世界知识等构建为向量数据库, 基于自研多粒度检索算法, 按需为玩家 query 检索相应知识以支撑模型回复, 同时训练回复模型对检索到知识的灵活运用能力。(c) 记忆与认知, 构建对话 AI 的记忆系统, 抽象对话内容并持续更新, 按需检索同时训练回复模型高情商运用记忆的能力。(d) 任务式对话, 构建类智能客服的对话能力, 通过意图识别、参数抽取、检索增强、工具调用、text-to-sql 等技术, 打造可控的对话剧情与任务推进。

目前已在腾讯 3 个大型游戏上线智能对话 Agent 服务, 个人为算法负责人。

2. NLP 大模型训练: (a) 参与腾讯混元大模型的偏好对齐 (RLHF) 模块, 包括: 跟踪 RM 数据的标注与清洗, 提出多种标注过滤策略, 并采用 GPT4 增补数据, 以达到更高质量的 RM 训练数据; 进行 reject sampling 实验并提出以标注人员的不一致性来构建 disagreement margin loss, 提高了混元模型在困难问题上的表现。(b) 垂域大模型训练: 基于开源通用领域模型进行下游 SFT 增训 (例如游戏), 采用多任务、MoE 等训练方式, 提升模型在下游领域的表现。

3. 学术研究: (a) 研究检索增强的 LLM (RAG), 通过 bandit 方法提升 RAG 中检索路由模块的效果与效率 (论文 [1][2]), 提升 RAG 中 LLM 对检索到噪声文档的鲁棒性 (论文 [3])。 (b) 研究模型偏好对齐, 提出对抗式偏好对齐方法 APO, 将 RM 与 LLM 交替迭代更新, 达到优于 reject sampling 等方法的效果 (论文 [4])。

华为公司–诺亚方舟 AI 实验室–NLP 研究员

2020.7–2022.2

工作主题: 基于预训练语言模型的搜索与推荐

工作内容:

1. 研究基于预训练 BERT 的信息流推荐技术, 即根据用户的新闻浏览历史, 匹配内容最相关的新闻。创新性提出多兴趣用户建模框架 MINER, 从多角度建模用户的历史行为和阅读兴趣。该方法在微软新闻推荐榜单 MIND 上排名第一 (Link, 2021 年 9 月), 论文发表在顶会 ACL 2022 (论文 [6], 第一作者)。

该算法应用至华为浏览器新闻信息流服务, 通过与精排模型 DCN 融合, 线上用户点击率 (CTR) 相对提升 2.5%

2. 研究基于图学习的新闻推荐方法, 分别把用户阅读历史和候选新闻建模成两张图, 提出算法 DIGAT 交互地更新用户图和新闻图, 论文已发表在顶会 EMNLP 2022 (论文 [9], 指导实习生完成)。

3. 研究文本生成与开放域问答技术, 提出文本生成算法 DPrior 发表在顶会 ACL 2022 (论文 [8]); 参加 NeurIPS 2020 开放域问答比赛, 改进基于 BERT 的文档检索和基于 T5 的文本生成模块, 在榜单上机构排名第五。

## 论文发表

谷歌学术 链接: Link 引用: 2100+

\* denotes Co-First Author.

- [1] Xiaqiang Tang, **Jian Li** (Corresponding Author), Nan Du, Sihong Xie. [Adapting to Non-Stationary Environments: Multi-Armed Bandit Enhanced Retrieval-Augmented Generation on Knowledge Graphs](#). In AAAI 2025.
- [2] Xiaqiang Tang\*, Qiang Gao\*, **Jian Li**\*, Nan Du, Sihong Xie. [MBA-RAG: a Bandit Approach for Adaptive Retrieval-Augmented Generation through Question Complexity](#). In COLING 2025.

[3] Wenhao Yu, Hongming Zhang, Xiaoman Pan, Kaixin Ma, **Jian Li**, Hongwei Wang, Dong Yu. [Chain-of-Note: Enhancing Robustness in Retrieval-Augmented Language Models](#). In *EMNLP 2024*.

[4] Pengyu Cheng\*, **Jian Li**\*, Yifan Yang\*, Yong Dai, Tianhao Hu, Peixin Cao, Nan Du. [Adversarial Preference Optimization: Enhancing Your Alignment via RM-LLM Game](#). In *findings of ACL 2024*.

[5] Xianghong Fang, **Jian Li**, Qiang Sun, Benyou Wang. [Rethinking the Uniformity Metric in Self-Supervised Learning](#). Accepted in *ICLR 2024*.

[6] **Jian Li**, Jieming Zhu, Guohao Cai, Lifeng Shang, Zhenhua Dong, Xin Jiang, Qun Liu. [MINER: Multi-Interest Matching over BERT for News Recommendation](#). In *findings of ACL 2022*.

[7] Qiwei Bi\*, **Jian Li**\*, Lifeng Shang, Xin Jiang, Qun Liu, Hanfang Yang. [MTRec: Multi-Task Learning over BERT for News Recommendation](#). In *findings of ACL 2022*.

[8] Xianghong Fang, **Jian Li** (Corresponding Author), Lifeng Shang, Xin Jiang, Qun Liu, Dit-Yan Yeung. [Controlled Text Generation Using Dictionary Prior in Variational Autoencoders](#). In *findings of ACL 2022*.

[9] Zhiming Mao, **Jian Li** (Corresponding Author), Xingshan Zeng, Lifeng Shang, Xin Jiang, Kam-Fai Wong, Qun Liu. [DIGAT: Modeling News Recommendation with Dual Graph Interaction](#). In *findings of EMNLP 2022*.

[10] Xianghong Fang, Haoli Bai, **Jian Li**, Zenglin Xu, Michael Lyu, Irwin King. [Discrete Auto-regressive Variational Attention Models for Text Modeling](#). In *IJCNN 2021*.

[11] **Jian Li**, Xing Wang, Zhaopeng Tu, Michael R. Lyu. [On the Diversity of Multi-Head Attention](#). In *Journal of Neurocomputing 2021 (SCI Journal)*.

[12] Silei Xu, Giovanni Campagna, **Jian Li**, Monica S. Lam. [Schema2QA: Answering Complex Queries on the Structured Web with a Neural Model](#). In *2020 International Conference on Information and Knowledge Management (CIKM 2020)*.

[13] **Jian Li**, Xing Wang, Baosong Yang, Michael R. Lyu, Zhaopeng Tu. [Neuron Interaction Based Representation Composition for Neural Machine Translation](#). In *2020 AAAI Conference on Artificial Intelligence (AAAI 2020)*.

[14] **Jian Li**, Baosong Yang, Zi-Yi Dou, Xing Wang, Michael R. Lyu, Zhaopeng Tu. [Information Aggregation for Multi-Head Attention with Routing-by-Agreement](#). In *2019 Conference of North American Chapter of the Association for Computational Linguistics (NAACL 2019)*.

[15] Baosong Yang, **Jian Li**, Derek Wong, Lidia S. Chao, Xing Wang, Zhaopeng Tu. [Context-Aware Self-Attention Networks](#). In *2019 AAAI Conference on Artificial Intelligence (AAAI 2019)*.

[16] **Jian Li**, Zhaopeng Tu, Baosong Yang, Michael R. Lyu, Tong Zhang. [Multi-Head Attention with Disagreement Regularization](#). In *2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*.

[17] **Jian Li**, Yue Wang, Irwin King, Michael R. Lyu. [Code Completion with Neural Attention and Pointer Networks](#). In *2018 International Joint Conference on Artificial Intelligence (IJCAI 2018)*.

[18] **Jian Li**, Pinjia He, Jieming Zhu, Michael R. Lyu. [Software Defect Prediction via Convolutional Neural Network](#). In *2017 IEEE International Conference on Software Quality, Reliability and Security (QRS 2017)*.

[19] Pinjia He, Jieming Zhu, Shilin He, **Jian Li**, Michael R. Lyu. [Towards Automated Log Parsing for Large-Scale Log Data Analysis](#). In *IEEE Transactions on Dependable and Secure Computing (TDSC 2017)*.

[20] Pinjia He, Jieming Zhu, Shilin He, **Jian Li**, Michael R. Lyu. [An Evaluation Study on Log Parsing and Its Use in Log Mining](#). In *2016 IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2016)*.

获奖与荣誉

CUHK Global Scholarship Programme for Research Excellence	2019–2020
CUHK Overseas Research Attachment Programme Scholarship	2019–2020
CUHK Full Postgraduate Studentship	2015–2019
国家奖学金 (Top 2%)	中国教育部, 2012 & 2013
唐立新奖学金 (Top 1%)	电子科大, 2014
成电杰出学生	电子科大, 2015
四川省优秀毕业生	四川省, 2015
美国大学生数学建模竞赛 (ICM) 一等奖	COMAP, 2014

技能

- 编程语言: Python, Scala, C/C++.
- 工具: PyTorch, Tensorflow, Spark, Linux, Git, Vim.