

# HONGRU WANG

Emails: [hrwang@se.cuhk.edu.hk](mailto:hrwang@se.cuhk.edu.hk) / [hongru.carrywang@gmail.com](mailto:hongru.carrywang@gmail.com)

[Homepage](#) | [Thesis Slides \(v2\)](#) | [Google Scholar](#)

## RESEARCH STATEMENT

My research focus revolves around internal cognitive tools and external physical tools of personalized language agents, designed to seamlessly unifying reasoning and acting from tool perspective instead of treat them in isolation. My long-term objective is to achieve the *impossible triangle* between safety, personalization and autonomy of language agent to learn from interactions internally or externally.

## EDUCATION

<b>The Chinese University of Hong Kong</b> Ph.D. candidate supervised by Prof. Kam-Fai Wong Dialogue System, Large Language Models, Tool Learning	<b>2021/09 - Now</b> GPA: 3.82/4.00
<b>The Chinese University of Hong Kong</b> M.Eng. Computer Science and Engineering Data Mining, Big Data Technology, Machine Learning	<b>2019/09 - 2020/11</b> GPA: 3.60/4.00
<b>Communication University of China</b> B.Eng. Computer Science and Technology Computer Operation System, Computer Networks, Data Structures and Algorithms	<b>2015/09 - 2019/09</b> GPA: 3.27/4.00

## (CO-)FIRST AUTHOR CONFERENCES

10. AppBench: Planning of Multiple APIs from Various APPs for Complex User Instruction  
[Hongru Wang](#), [Rui Wang](#), [Boyang XUE](#), et. al, [Jeff Z. Pan](#), [Kam-Fai Wong](#)  
**EMNLP 2024** LLM , Tool Learning (Apple Intelligence)
9. TPE: Towards Compositional Reasoning over Conceptual Tools with Multi-persona Collaboration  
[Hongru Wang](#), [Huimin Wang](#), [Lingzhi Wang](#), et. al, [Kam-Fai Wong](#)  
**NLPCC 2024** LLM , Tool Learning (Meta-reasoning theory / Cognitive tools)
8. Enhancing Large Language Models Against Inductive Instructions with Dual-critique Prompting  
[Rui Wang\\*](#), [Hongru Wang\\*](#), [Fei Mi](#), [Boyang XUE](#), [Yi Chen](#), [Kam-Fai Wong](#), [Ruifeng Xu](#)  
**NAACL 2024** LLM , Safety \* denotes equal contribution
7. UniRetriever: Multi-task Candidates Selection for Various Context-Adaptive Conversational Retrieval  
[Hongru Wang](#), [Boyang Xue](#), [Baohang Zhou](#), et. al, [Kam-Fai Wong](#)  
**LREC-COLING 2024** Dialogue System , Tool Learning
6. M3Sum: A Novel Unsupervised Language-guided Video Summarization  
[Hongru Wang](#), [Baohang Zhou](#), [Zhengkun Zhang](#), [Yiming Du](#), [David Ho](#), [Kam-Fai Wong](#)  
**ICASSP 2024** LLM , MM
5. Large Language Models as Source Planner for Personalized Knowledge-grounded Dialogue  
[Hongru Wang](#), [Minda Hu](#), [Yang Deng](#), et. al, [Irwin King](#), [Kam-Fai Wong](#)  
Findings of **EMNLP 2023** Dialogue System , LLM **Best Paper Award 🏆 @Doctoral Forum**
4. Cue-CoT: Chain-of-thought Prompting for Responding to In-depth Dialogue Questions with LLMs  
[Hongru Wang\\*](#), [Rui Wang\\*](#), [Fei Mi](#), et. al, [Ruifeng Xu](#) and [Kam-Fai Wong](#)  
Findings of **EMNLP 2023** Dialogue System , LLM [Blog](#)

3. MCML: A Novel Memory-based Contrastive Meta-Learning Method for Few Shot Slot Tagging  
Hongru Wang, Zezhong Wang, Wai-Chung Kwan, Kam-Fai Wong  
**IJCNLP-AAACL 2023** Dialogue System
2. Integrating Pretrained Language Model for Dialogue Policy Learning  
Hongru Wang, Huimin Wang, Zezhong Wang and Kam-Fai Wong  
**ICASSP 2022** Dialogue System , RLHF (probably first work of RLHF)
1. CUHK at SemEval-2020 Task 4: CommonSense Explanation, Reasoning and Prediction ..  
Hongru Wang, Xiangru Tang, Sunny Lai, et. al, Gabriel Pui Cheong Fung and Kam-Fai Wong  
SemEval of **COLING 2020**. Commonsense

## (CO-)FIRST AUTHOR JOURNAL

---

2. KddRES: A Multi-level Knowledge-driven Dialogue .. Towards Customized Dialogue System  
Hongru Wang, Wai-Chung Kwan, Min Li, Zimo Zhou and Kam-Fai Wong  
**Computer Speech and Language 2024 (Tsinghua B, JCR Q2, IF: 4.3)** Dialogue System
1. A Survey on Recent Advances and Challenges in Reinforcement Learning Methods ..  
Wai-Chung Kwan\*, Hongru Wang\*, Huimin Wang and Kam-Fai Wong  
**Machine Intelligence Research 2023 (JCR Q1, IF: 6.4)** Dialogue System , RLAIIF

## NON-FIRST AUTHOR CONFERENCES

---

17. AutoPSV: Automated Process-Supervised Verifier  
Jianqiao Lu, Zhiyang Dou, Hongru Wang, et. al, Zhijiang Guo  
**NeurIPS 2024** LLM (Meta-reasoning theory)
16. Knowledge Conflicts for LLMs: A Survey  
Rongwu Xu, Zehan Qi, Zhijiang Guo, Cunxiang Wang, Hongru Wang, Yue Zhang, Wei Xu  
**EMNLP 2024** LLM [Blog](#)
15. VLEU: a Method for Automatic Evaluation for Generalizability of Text-to-Image Models  
Jingtao Cao, Zhang Zheng, Hongru Wang, Kam-Fai Wong  
**EMNLP 2024** MM
14. Less is More: Making Smaller Language Models .. Subgraph Retrievers for Multi-hop KGQA  
Wenyu Huang, Guancheng Zhou, Hongru Wang, et. al, Mirella Lapata, Jeff Z. Pan  
Findings of **EMNLP 2024** LLM (Differentiable Search Index (DSI))
13. Enhancing Biomedical Knowledge RAG with Self-Rewarding Tree Search and PPO  
Minda Hu, Licheng Zong, Hongru Wang, et. al, Kam-Fai Wong, Yu Li, Irwin King  
Findings of **EMNLP 2024** LLM
12. DPDLLM: A Black-box Framework for Detecting Pre-training Data from Large Language Models  
Baohang Zhou, Zezhong WANG, Lingzhi Wang, Hongru Wang, et. al, Kam-Fai Wong  
Findings of **ACL 2024** LLM
11. Medical Dialogue: A Survey of Categories, Methods, Evaluation and Challenges  
Xiaoming Shi, Zeming Liu, Li Du, Yuxuan Wang, Hongru Wang, et. al, Shaoting Zhang  
Findings of **ACL 2024** LLM
10. REGA: Role Prompting Guided Multi-Domain Adaptation for Large Language Models  
Rui Wang, Fei Mi, Yi Chen, Boyang XUE, Hongru Wang, Qi Zhu, Kam-Fai Wong, Ruifeng Xu  
Findings of **NAACL 2024** LLM , Role Playing

9. SELF-GUARD: Empower the LLM to Safeguard Itself  
Zezhong Wang, Fangkai Yang, Lu Wang, Pu Zhao, Hongru Wang, et. al, Kam-Fai Wong  
**NAACL 2024** LLM , Safety
8. JoTR: A Joint Transformer and Reinforcement Learning Framework for Dialog Policy Learning  
Wai-Chung Kwan, Huimin Wang, Hongru Wang, et. al, Kam-Fai Wong  
**LREC-COLING 2024** Dialogue System
7. MCIL: Multimodal Counterfactual Instance Learning for .. Multimodal Information Extraction  
Baohang Zhou, Ying Zhang, Kehui Song, Hongru Wang, Yu Zhao, Xuhui Sui, Xiaojie Yuan  
**LREC-COLING 2024** LLM , MM
6. ReadPrompt: A Readable Prompting Method for Reliable Knowledge Probing  
Zezhong Wang\*, Luyao YE\*, Hongru Wang, Wai-Chung Kwan, David Ho, Kam-Fai Wong  
Findings of **EMNLP 2023** LLM
5. Improving Factual Consistency for Knowledge-Grounded Dialogue Systems via Knowledge ..  
Boyang XUE\*, Weichao Wang\*, Hongru Wang, et. al, Xin Jiang, Qun Liu, Kam-Fai Wong  
Findings of **EMNLP 2023** Dialogue System
4. Prompting and Evaluating Large Language Models for Proactive Dialogues ..  
Yang Deng, Lizi Liao, Liang CHEN, Hongru Wang, Wenqiang Lei, Tat-Seng Chua  
Findings of **EMNLP 2023** LLM
3. Towards Robust Personalized Dialogue Generation via Order-Insensitive Representation ..  
Liang Chen, Hongru Wang, Yang Deng, Wai Chung Kwan, Zezhong Wang, Kam-Fai Wong  
Findings of **ACL 2023** (short) Dialogue System
2. Retrieval-free Knowledge Injection through Multi-Document Traversal for Dialogue Models  
Rui Wang, Jianzhu Bao, Fei Mi, Yi Chen, Hongru Wang, et. al, Kam-Fai Wong, Ruifeng Xu  
**ACL 2023** Dialogue System
1. DIGAT: Modeling News Recommendation with Dual-Graph Interaction  
Zhiming Mao, Jian Li, Hongru Wang, Xingshan Zeng, Kam-Fai Wong  
Findings of **EMNLP 2022** Others

## WORKSHOP, TUTORIAL AND OTHERS

---

4. Analysing the Residual Stream of Language Models Under Knowledge Conflicts  
Yu Zhao, Xiaotang Du, et. al, Hongru Wang, Xuanli He, Kam-Fai Wong, Pasquale Minervini  
MINT Workshop of **NeurIPS 2024**
3. PerLTQA: A Personal Long-Term Memory Dataset for Memory Classification, Retrieval, and ..  
Yiming Du, Hongru Wang, Zhengyi Zhao, et. al, Kam-Fai Wong  
SIGHAN Workshop of **ACL 2024**. **Best Paper Award** 😊.
2. OSPC: Detecting Harmful Memes with Large Language Model as a Catalyst  
Jingtao Cao, Zheng Zhang, Hongru Wang, Bin Liang, Hao Wang, Kam-fai Wong  
Online Safety Prize Challenge, **WWW 2024**. **Champion Solution** 😊.
1. Empowering Large Language Models: Tool Learning for Real-World Interaction  
Hongru Wang, Yujia Qin, Yankai Lin, Jeff Z. Pan, Kam-fai Wong  
Tutorial, **SIGIR 2024** LLM , Tool Learning

## PREPRINTS

---

8. MlingConf: A Comprehensive Study of Multilingual Confidence Estimation on LLMs  
Boyang XUE\*, Hongru WANG\*, Rui Wang, et al., Bin Liang, Kam-Fai Wong  
Under Review, ARR Oct (3 3.5 3.5) [LLM](#)
7. Self-DC: When to retrieve and When to generate? Self Divide-and-Conquer for Compositional ..  
Hongru Wang\*, Boyang Xue\*, Baohang Zhou, et. al, Kam-Fai Wong  
Under Review, ARR Oct (3.5 3.5 3.5) [LLM](#) , [Tool Learning](#)
6. Steering Knowledge Selection Behaviours in LLMs via SAE-Based Representation Engineering  
Yu Zhao, et al., Hongru WANG, Xuanli He, Kam-Fai Wong, Pasquale Minervini  
Under Review, ARR Oct (3.5 4 4) [LLM](#)
5. SeqAR: Jailbreak LLMs with Sequential Auto-Generated Characters  
Yan Yang, Zeguan Xiao, Xin Lu, Hongru WANG, et al., Guanhua Chen, Yun Chen  
Under Review, ARR Oct (3 4 3) [LLM](#)
4. Compound-QA: A Benchmark for Evaluating LLMs on Compound Questions  
Yutao Hou, Yajing Luo, Zhiwen Ruan, Hongru WANG, et al., Guanhua Chen  
Under Review, ARR Oct (3 3 3.5) [LLM](#)
3. Can LLMs Evaluate Complex Attribution in QA? Automatic Benchmarking Using Knowledge Graphs  
Nan Hu, Jiaoyan Chen, Yike Wu, Guilin Qi, Hongru Wang, Sheng Bi, Tongtong Wu, Jeff Z. Pan  
Under Review, ICLR 2025 [LLM](#)
2. A Survey of the Evolution of Language Model-Based Dialogue Systems  
Hongru Wang, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, Kam-Fai Wong  
Under Review [LLM](#) [Dialogue System](#) [Blog](#)
1. UniMS-RAG: A Unified Multi-source Retrieval-Augmented Generation for Personalized ..  
Hongru Wang, Wenyu Huang, Yang Deng, et. al, Jeff Z Pan, Kam-Fai Wong  
Under Review [LLM](#) , [Dialogue System](#)

## EXPERIMENTENCE

---

Visiting Student <b>BlenderLab</b> (Mentor: Prof. Heng Ji)	University of Illinois Urbana-Champaign	2024/12 - Now
Research Intern <b>ByteDance (SZ)</b> (Mentor: Dr. Deng Cai, Dr. Wanjun Zhong)	Seed-LLM-Harizon (Doubao)	2024/08 - 2024/12
Visiting Student <b>EdinburghNLP</b> (Mentor: Prof. Jeff Z. Pan)	University of Edinburgh	2024/04 - 2024/11
Research Assistant <b>CUHK</b> (Mentor: Prof. Kam-Fai Wong)	MoE Key Lab of High Confidence Software Technologies	2019/12 - 2021/07

## GRANTS / FUNDING

---

3. Overseas Research Attachment Programme (50,000 HKD, ORAP 2023-2024)
2. A Knowledge Graph Based Dynamic Video Extractive Summarization System (PRP/054/21FX)
1. "SEVES: Semantic-driven Effective Video Extractive Summarization System" – Technology and Business Development Fund (200,000 HKD, TBF22ENG004)

## TEACHING ASSISTANT

---

SEEM 3450 Engineering Innovation and Entrepreneurship  
SEEM 3490 Information Systems Management  
SEEM 5730 / ECLT 5910 Information Technology Management

## COMPETITIONS & AWARDS

---

Top 1 at Online Safety Prize Challenge, WWW 2024	international
Reaching Out Awards (2022-2023)	school
Top 10 at ICLR 2021 Workshop MLPCP Track 1 (rank 7th)	international
Third Price at SMP2020-ECDT Few-shot Spoken Language Understanding	international
Top 10 at SemEval2020-Task4: CommonSense Detection and Explanations	international
Distinguished Academic Performance Scholarship (2019-2020), CUHK CSE	school
Meritorious Winner, Mathematical Contest In Modeling (2018)	international
A software copyright of "WeCampus WeChat mini-program" (2018SR562540)	national
China National Radio Scholarship	school

## COMMUNITY CONTRIBUTION

---

Reviewer: ICLR, ACL/EMNLP/NAACL/COLING, AACL, IJCAI, ARR, ...  
Organizer: [ToolsMeetLLM@SIGIR2024](mailto:ToolsMeetLLM@SIGIR2024)  
Co-founder and Committee: [NLP Academic Exchange Platform \(NICE\)](#) (Fans: 10w+; Views: 15w+)