Yifei Yuan

+45 50340456 | yfyuan01@gmail.com
ETH Zurich
https://yfyuan01.github.io/
https://scholar.google.com/citations?user=2LQnqqAA

EMPLOYMENT

ETH Zurich

Postdoc

I am currently a postdoc researcher at Swiss Data Science Center (SDSC), ETH Zurich. I am co-founded by UZH, working on Swiss political NLP.

University of Copenhagen

Postdoc

I previously served as a postdoc fellow (now as a guest researcher) at Coastal NLP group at University of Copenhagen, working with Prof. Anders Søgaard, founded by Novo Nordisk and Rigshospitalet, Denmark.

RESEARCH INTERESTS

Natural language processing (NLP), Multimodality, Information retrieval (IR), Machine learning (ML)

Multimodal representation learning; Health AI; Conversational information seeking; Generative AI; Human-centered AI

EDUCATION

The Chinese University of Hong Kong	Aug 2019 - Aug 2023
Doctor System Engineering and Engineering Management	Hong Kong SAR
I obtained my PhD degree at CUHK Text Mining Group, supervised by Prof. Wai Lam.	
Harbin Institute of Technology	Jul 2015 - Jul 2019
Bachelor Department of Computer Science and Technology	China
I received my Bachelor's degree on Bioinformatics at HIT in 2019 (GPA: 3.82).	

RESEARCH EXPERIENCE

Visiting Scholar	Jul 2022 - Apr 2023
IRLab at UvA	Amsterdam, Netherlands
I was a visitor to the IRLab of University of Amsterdam, supervised and collaborated with Dr. Mohammad Aliannejadi and Prof. Maarten de Rijke.	
Research Intern	Nov 2021 - Jun 2022
Alibaba Group	Hangzhou, China
I was a research intern at the AliCloud Team of Alibaba, working as a data scientist on conversational systems under large language models (LLMs).	
Research Student	Apr 2017 - Jun 2019
SCIR Lab of HIT	Harbin, China
I was a research student at the SCIR lab of HIT, supervised by Prof.Bin Qin.	
Research Intern	Nov 2018 - Jun 2019
Business Intelligence Lab, Baidu	Beijing, China
I was an intern at the Business Intelligence Lab, Baidu, lead by Prof.Hui Xiong, working as a data scientist on point of interest (POI) analysis.	

TEACHING EXPERIENCE

Lecturer at KU	Nov 2023 - Present
2023, 2024 ATNLP (Co-teaching with Desmond Elliot)	

Invited Lecturer at UvA	Oct 2022 - Nov 2022

2022 Intelligent Interactive Systems (Co-teaching with Frank Nack)

Teaching Assistant at CUHK	Feb 2021 - Jun 2021

2020R2-SEEM 2420 Operations Research I

Teaching Assistant at CUHK	Feb 2021 - Jun 2021

2020R2-SEEM 3450 Engineering Innovation and Entrepreneurship

Teaching Assistant at CUHK	Feb 2020 - Jun 2020

2019R2-SEEM 3570 Stochastic Models

PROFESSIONAL ACTIVITY

Proceeding Chair, ECIR 2026

Area Chair, ACL 2025

Program Committee, SIGKDD 2025

Reviewer, ICLR 2025

Reviewer, COLING 2025

Program Committee, CIKM 2024

Program Committee, SIGIR 2024

Reviewer EMNLP 2024

Reviewer, TOIS 2024

Reviewer, ACL 2024

Reviewer, NAACL 2024

Reviewer, EACL 2024

Reviewer, WWW 2022

Reviewer, ACL 2022

Reviewer, ACM Computing Surverys 2022

Secondary Reviewer, ECIR 2022, CVPR 2021, WWW 2021, COLING 2020 ...

Honors & Awards

SIGIR Student Travel Grant	Dec 2021
SIGIR Student Travel Grant	Jun 2021
Second Prize of People's Scholarship, Harbin Institute of Technology	Aug 2017
National Scholarship	May 2016
Outstanding Technical Project Award of People's Network Scholarship	Apr 2016

INVITED TALK

Talk at European Summer School on Information Retrieval 2024 (Conversational IR)	Jul 2024
Speaker	University of Amsterdam, Amsterdam
Talk at NLP-SIG Workshop (LLM-based Clarification in Conversational Search)	May 2024
Speaker	National University of Singapore, Singpoare
Talk at Swiss Data Science Center, ETH	Feb 2025
Speaker	ETH, Switzerland
Tutorial at the WebConf 2025	
Speaker	

Tutorial Session 'Query Understanding in LLM-based Conversational Information Seeking' at the Web Conference, Sydney.

PUBLICATIONS

Revisiting the Othello World Model Hypothesis

Revisiting the Othello World Model Hypothesis

Yifei Yuan, Anders Søgaard

ICLR World Model Workshop 2025

Unlocking Markets: A Multilingual Benchmark to Cross-Market Question Answering

Yifei Yuan, Yang Deng, Anders Søgaard, Mohammad Aliannejadi

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).

FoodieQA: A Multimodal Dataset for Fine-Grained Understanding of Chinese Food Culture

Wenyan Li, Xinyu Zhang, Jiaang Li, Qiwei Peng, Raphael Tang, Li Zhou, Weijia Zhang, Guimin Hu, Yifei Yuan, Anders Søgaard, Daniel Hershcovich, Desmond Elliott

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).

Asking Multimodal Clarifying Questions in Mixed-Initiative Conversational Search

Yifei Yuan, Clemencia Siro, Mohammad Aliannejadi, Maarten de Rijke, Wai Lam

Proceedings of the 2024 ACM Web Conference (WWW 2024).

A Comparative Analysis of Faithfulness Metrics and Humans in Citation Evaluation

Weijia Zhang, Mohammad Aliannejadi, Jiahuan Pei, Yifei Yuan, Jia-Hong Huang, Evangelos Kanoulas

Proceedings of the 17th International Natural Language Generation Conference (INLG 2024).

CO3: Low-resource Contrastive Co-training for Generative Conversational Query Rewrite

Yifei Yuan, Chen Shi, Runze Wang, Liyi Chen, Renjun Hu, Zengming Zhang, Feijun Jiang, Wai Lam

Proceedings of the LREC-COLING 2024.

On the Multi-turn Instruction Following for Conversational Web Agents

Yang Deng, Xuan Zhang, Wenxuan Zhang, Yifei Yuan, See-Kiong Ng, Tat-Seng Chua

Proceedings of the 2024 Annual Meeting of the Association for Computational Linguistics (ACL 2024).

Let the LLMs Talk: Simulating Human-to-Human Conversational QA via Zero-Shot LLM-to-LLM Interactions

Zahra Abbasiantaeb, Yifei Yuan, Evangelos Kanoulas, Mohammad Aliannejadi

Proceedings of the 15th ACM International Conference on Web Search and Data Mining, 2024 (WSDM 2024).

Exploring Visual Culture Awareness in GPT-4V: A Comprehensive Probing

Yong Cao, Wenyan Li, Jiaang Li, Yifei Yuan, Antonia Karamolegkou, Daniel Hershcovich

Arxiv Preprint 2024

Social Media Fashion Knowledge Extraction as Captioning

Yifei Yuan, Wenxuan Zhang, Yang Deng, Wai Lam

Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region (SIGIR-AP 2023).

Knowledge-enhanced Mixed-initiative Dialogue System for Emotional Support Conversations

Yang Deng, Wenxuan Zhang, Yifei Yuan, Wai Lam

Proceedings of the 2023 Annual Meeting of the Association for Computational Linguistics (ACL 2023).

Mcqueen: a benchmark for multimodal conversational query rewrite

Yifei Yuan, Chen Shi, Runze Wang, Liyi Chen, Feijun Jiang, Yuan You, Wai Lam

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).

Sentiment Analysis of Fashion Related Posts in Social Media

Yifei Yuan, Wai Lam

Proceedings of the 15th ACM International Conference on Web Search and Data Mining, 2022 (WSDM 2022).

Aspect Sentiment Quad Prediction as Paraphrase Generation

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, Wai Lam

Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021).

Conversational Fashion Image Retrieval via Multiturn Natural Language Feedback

Yifei Yuan, Wai Lam

Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021 (SIGIR 2021).

Point-of-Interest Oriented Question Answering with Joint Inference of Semantic Matching and Distance Correlation

Yifei Yuan, Jingbo Zhou, and Wai Lam.

Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (AACL 2020).

AGENT-CQ: Automatic Generation and Evaluation of Clarifying Questions for Conversational Search with LLMs

Clemencia Siro, Yifei Yuan, Mohammad Aliannejadi, Maarten de Rijke

Arxiv Preprint

Student Supervison

Albert Arvad Ottosen, Yingming Wang, Jacob Schwartz Olsen, Shengwei Ding, Chenye Zou

Master student at University of Copenhagen

Kimia Ramezan

Bachelor student at Sharif University of Technology

Jiatong Li

Master student at University of Science and Technology of China

SKILLS LIST

Progamming Languages: Python (skilled), C++, Matlab, SQL, Latex

Frameworks & Tools: Pytorch, Tensorflow, Keras, Jupyter, Git, Google Could Platform

Languages

Native: Mandarin

Full professional proficiency: English