ZHIQI HUANG

zhiqi.huang@capitalone.com ⊙ 339-999-7777
https://zhiqihuang.github.io

RESEARCH INTEREST

Information retrieval, natural language processing, data mining, and machine learning.

EDUCATION

University of Massachusetts, Amherst *June 2024*
Ph.D. in Computer Science
Advisor: Prof. James Allan

University of Maryland, College Park *May 2015*
M.A. in Statistics

Sun Yat-sen University *June 2013*
B.S. in Applied Mathematics

WORK EXPERIENCE

Capital One Jul. 2023 - Present
Applied Researcher *Boston, MA, USA.*

Work at the AI Foundation at Capital One, developing Generative AI applications focused on large language models (LLMs) customization and retrieval-augmented generation (RAG).

Google Oct. 2023 - Jan. 2024
Research Intern *New York, NY, USA.*

Work on a project related to retrieval-augmented automatic speech recognition (ASR).

National Engineering Lab, Shenzhen University Jun. 2017 - Jul. 2018
Research Fellow *Shenzhen, China*

Work on uncertainty modeling and method for inductive learning.

Quad2Media Jul. 2015 - Mar. 2017
Data Scientist *Boston, MA, USA*

Design and manage a large-scale user profile system for the marketing campaign.

SELECTED PUBLICATIONS

[1] Zhichao Xu, **Zhiqi Huang**, Shengyao Zhuang, and Vivek Srikumar. *Distillation versus contrastive learning: How to train your rerankers.* The Proceedings of 14th International Joint Conference on Natural Language Processing (IJCNLP) & 4th Asia-Pacific Chapter of the Association for Computational Linguistics (AACL), 2025

[2] **Zhiqi Huang***, Vivek Datla*, Chenyang Zhu, Alfy Samuel, Daben Liu, Anoop Kumar, and Ritesh Soni. *Confidence-Based Response Abstinence: Improving LLM Trustworthiness via Activation-Based Uncertainty Estimation.* 2nd Uncertainty-Aware NLP Workshop at EMNLP 2025.

[3] Chi Zhang, Vivek V. Datla, Aditya Shrivastava, Alfy Samuel, **Zhiqi Huang**, Anoop Kumar, and Daben Liu. *An Automatic Method to Estimate Correctness of RAG.* The Proceedings of the 31st International Conference on Computational Linguistics (COLING), 2025.

[4] **Zhiqi Huang**, Puxuan Yu, Shauli Ravfogel, James Allan. *Language Concept Erasure for Language-invariant Dense Retrieval*. The Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024.

[5] **Zhiqi Huang**, Diamantino Caseiro, Kandarp Joshi, Christopher Li, Pat Rondon, Zelin Wu, Petr Zadrazil, Lillian Zhou. *Optimizing Large-Scale Context Retrieval for End-to-End ASR*. INTERSPEECH, 2024.

[6] **Zhiqi Huang**, Puxuan Yu, James Allan. *UMass at TREC 2023 NeuCLIR Track*. The 32nd Text REtrieval Conference (TREC), 2023.

[7] Puxuan Yu, Razieh Negin Rahimi, **Zhiqi Huang**, James Allan. *Search Result Diversification Using Query Aspects as Bottlenecks*. The 32th ACM International Conference on Information and Knowledge Management (CIKM), 2023.

[8] **Zhiqi Huang**, Hansi Zeng, Hamed Zamani, James Allan. *Soft Prompt Decoding for Multilingual Dense Retrieval*. The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2023.

[9] **Zhiqi Huang***, Shahrzad Naseri*, Hamed Bonab, Sheikh Muhammad Sarwar, James Allan. *Hierarchical Transformer-based Query by Multiple Documents*. The 13th International Conference on the Theory of Information Retrieval (ICTIR), 2023. **Best Paper Honorable Mention**.

[10] **Zhiqi Huang**, Puxuan Yu, James Allan. *Improving Cross-lingual Information Retrieval on Low-Resource Languages via Optimal Transport Distillation*. The 16th ACM International Conference on Web Search and Data Mining (WSDM), 2023.

[11] **Zhiqi Huang**, Hamed Bonab, Sheikh Muhammad Sarwar, Razieh Negin Rahimi, James Allan. *Mixed Attention Transformer for Leveraging Word-Level Knowledge to Neural Cross-Lingual Information Retrieval*. The 30th ACM International Conference on Information and Knowledge Management (CIKM), 2021.

[12] **Zhiqi Huang**, Razieh Negin Rahimi, Puxuan Yu, Jingbo Shang, James Allan. *AutoName: A Corpus-Based Set Naming Framework*. The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2021.

[13] Puxuan Yu, Razieh Negin Rahimi, **Zhiqi Huang**, James Allan. *Learning to Rank Entities for Set Expansion from Unstructured Data*. The 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2020.

[14] Puxuan Yu, **Zhiqi Huang**, Razieh Negin Rahimi, James Allan. *Corpus-based Set Expansion with Lexico-syntactic Features and Distributed Representations*. The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2019.

[15] **Zhiqi Huang**, Ryan McKenna, Gerome Miklau, Michael Hay, Ashwin Machanavajjhala. *PSynDB: Accurate and Accessible Private Data Generation*. The 45th International Conference on Very Large Data Bases (VLDB), 2019.

[16] Ran Wang, Sam Kwong, Yuheng Jia, **Zhiqi Huang** and Lang Wu. *Mutual Information Based K-Labelsets Ensemble for Multi-Label Classification*. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2018.

[17] **Zhiqi Huang**, Xizhao Wang, *Sensitivity of Data Matrix Rank in Non-iterative Training*, Neurocomputing Volume 313 (2018): p386-391.

(∗equal contribution)

AWARDS

Best Paper Honorable Mention 2023
ACM ICTIR

Dissertation Writing Fellowship 2023 and 2024
CICS, UMass Amherst

Winner of WSDM Cup 2023 2023
Second prize on multilingual information retrieval (MIRACL) challenge

PROFESSIONAL SERVICES

Program Committee Member (Reviewer) for ACM SIGIR (22, 23, 24, 25), ACM CIKM (21, 22, 23, 24, 25), ACM WWW (23, 24, 25, 26), IEEE BigData (24).

Reviewer of IEEE Transactions on Knowledge and Data Engineering (TKDE).

TEACHING

Teaching Assistant: COMPSCI 240 Reasoning Under Uncertainty, Spring 2020, UMass Amherst.

Teaching Assistant: COMPSCI 689 Machine Learning, Fall 2019, UMass Amherst.

Teaching Assistant: COMPSCI 356 Digital Forensics, Spring 2019, UMass Amherst.