

# Haopeng Zhang

Email: haopengz@hawaii.edu

Tel: 404-376-1886

Website: hpzhang94.github.io

Google Scholar

## RESEARCH INTERESTS

---

- **Natural Language Processing**

*Graph4NLP, Text Summarization, Large Language Models*

- **Graph Mining**

*Graph Neural Network, Social Network Analysis*

## EDUCATION

---

**University of California, Davis**

Ph.D. Candidate in Computer Science

Transferred from Florida State University (FSU)

Thesis: Intelligent and reliable summarization systems.

Expected: Jun. 2024

Advisor: Dr. Jiawei Zhang

**Georgia Institute of Technology**

M.S. in Computational Science and Engineering

M.S. in Electrical and Computer Engineering

Dec. 2018

**University of Illinois, Urbana and Champaign (UIUC)**

B.E. in Electrical Engineering

May 2016

## PROFESSIONAL EXPERIENCE

---

**Amazon Web Services (AWS) AI Lab**

**Applied Scientist Intern** (Mentors: Dr. Lan Liu, Dr. Peng Qi, Dr. Zhiheng Huang)

Worked on exploring different optimization strategies of retrieval-augmented generation systems for large-scale industry search engines.

Santa Clara, CA

Summer 2023

**Megagon Lab**

**Research Intern** (Mentors: Dr. Hayate Iso, Dr. Sairam Gurajada)

Proposed and created the first large-scale benchmark for explainable fine-grained text editing task. The research paper is published in LREC-COLING 2024.

Mountain View, CA

Spring 2023

**Tencent AI Lab**

**Research Intern** (Mentors: Dr. Sangwoo Cho, Dr. Kaiqiang Song)

Worked on graph-based unsupervised multi-document summarization; proposed a holistic framework that balances summary salience and diversity. The research paper is published in IJCNLP-AAACL 2023.

Seattle, WA

Summer 2022

**Salesforce Research**

**Research Intern** (Mentors: Dr. Semih Yavuz, Dr. Yingbo Zhou)

Proposed a new approach to address the entity-level summary hallucination problem by controlling entity coverage and domain transfer for abstractive summarization with intermediate data. The research paper is published in NAACL 2022.

Remote

Summer 2021

**Kidswant Company**

**Data Mining Research Intern** (Mentor: Dr. Hang Liu)

Worked on extracting lifecycle rules for different classes of products and improved the recommendation system; Developed tool with Django and Pyecharts for dynamic data visualization.

Nanjing

Summer 2018

## PUBLICATIONS

---

- [1] Z. Li, **H. Zhang**, and J. Zhang, “Unveiling the magic: Investigating attention distillation in retrieval-augmented generation”, in *2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2024)*, 2024.
- [2] **H. Zhang**, H. Iso, S. Gurajada, and N. Bhutani, “Xatu: A fine-grained instruction-based benchmark for explainable text updates”, in *The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 2024.
- [3] **H. Zhang**, X. Liu, and J. Zhang, “Summit: Iterative text summarization via chatgpt”, in *Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, Singapore: Association for Computational Linguistics, 2023.
- [4] **H. Zhang**, X. Liu, and J. Zhang, “Extractive summarization via chatgpt for faithful summary generation”, in *Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, Singapore: Association for Computational Linguistics, 2023.
- [5] **H. Zhang**, S. Cho, K. Song, X. Wang, J. Zhang, and Y. Dong, “Unsupervised multi-document summarization with holistic inference”, *The 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AAACL 2023)*, 2023.
- [6] **H. Zhang**, X. Liu, and J. Zhang, “DiffuSum: Generation enhanced extractive summarization with diffusion”, in *Findings of the Association for Computational Linguistics: ACL 2023*, Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 13 089–13 100.
- [7] **H. Zhang**, X. Liu, and J. Zhang, “Contrastive hierarchical discourse graph for scientific document summarization”, in *Proceedings of the 4th Workshop on Computational Approaches to Discourse (ACL CODI 2023)*, Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 37–47.
- [8] **H. Zhang**, X. Liu, and J. Zhang, “HEGEL: Hypergraph transformer for long document summarization”, in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 10 167–10 176.
- [9] **H. Zhang**, S. Yavuz, W. Kryscinski, K. Hashimoto, and Y. Zhou, “Improving the faithfulness of abstractive summarization via entity coverage control”, in *Findings of the Association for Computational Linguistics: NAACL 2022*, Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 528–535.
- [10] **H. Zhang** and J. Zhang, “Text graph transformer for document classification”, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, Online: Association for Computational Linguistics, Nov. 2020, pp. 8322–8327.
- [11] Z. Li, **H. Zhang**, and J. Zhang, “A revisit of fake news dataset with augmented fact-checking by chatgpt”, under review, 2023.
- [12] **H. Zhang** and J. Zhang, “Centrality meets centroid: A graph-based approach for unsupervised document summarization”, *arXiv preprint arXiv:2103.15327*, 2021.
- [13] J. Zhang, **H. Zhang**, L. Sun, and C. Xia, “Graph-bert: Only attention is needed for learning graph representations”, *arXiv preprint arXiv:2001.05140*, 2020.

## SCHOLARSHIPS AND AWARDS

---

- UC Davis Graduate Group in Computer Science (GGCS) Research Fellowship 2023
- FSU Travel Award 2020
- FSU Adelaide D. Wilson Graduate Fellowship Endowment Fund 2019
- UIUC Russell E. Berthold Scholarship 2015

## SERVICE

---

- **Journal Reviewer**

*IEEE Transactions on Big Data*

*IEEE/ACM Transactions on Audio, Speech, and Language Processing*

*Artificial Intelligence Review*

*Computer Speech & Language*

*PeerJ Computer Science*

- **Conference Reviewer**

*SDM 24, EMNLP 23, IJCNN 23/24, ACL 23, EACL 23, KDD 23/24, AAAI 21/22, NAACL 24, CODI 24, ACL Rolling Review*

- **Program Committee**

*The 4th New Frontiers in Summarization Workshop @EMNLP 2023*

## TEACHING EXPERIENCE

---

- **Teaching Assistant@UC Davis**

*Deep Learning (ECS189G)*

Winter 2022

- **Teaching Assistant@FSU**

*Advanced Database Systems (COP5725)*

Fall 2021, Spring 2021

*Complexity and Analysis of Data Structures and Algorithms (COP4531)*

Spring 2020

*Theory and Structure of Databases (COP4710)*

Fall 2020, Fall 2019

- **Teaching Assistant@Georgia Tech**

*Semiconductor Devices (ECE3450)*

Fall 2017, Spring 2017

## MENTORSHIP

---

- **Xiao Liu** (M.S. student → Ph.D. student, UC Davis)

*I have been mentoring Xiao for two years since he was a M.S. student and working with him through the Ph.D. program. I guided and helped him in several research projects on extractive text summarization, structural document representation, and large language model applications. We collaborated and published six papers in top-tier NLP conferences.*

- **Zizhong Li** (Ph.D. student, UC Davis)

*I have been mentoring Zizhong for a year on her research projects involving fake news detection with large language models and retrieval augmented generation. Our collaborative research paper is published in NAACL 2024.*

## REFERENCE

---

- **Jiawei Zhang** (Assistant professor, Department of Computer Science, UC Davis)  
*email: jiwzhang@ucdavis.edu*
- **Xin Liu** (Professor, Department of Computer Science, UC Davis, IEEE Fellow)  
*email: xinliu@ucdavis.edu*
- **Xiaoyang Wang** (Senior Researcher, Tencent AI lab)  
*email: shawnxywang@global.tencent.com*