Haopeng Zhang

haopeng.zhang@hawaii.edu
309C, 1680 East-West Rd
Honolulu, HI 96822
404-376-1886
Website
Google Scholar

RESEARCH INTERESTS

Natural Language Processing
Summarization, Generation, Low-Resource Languages

Generative AI
Large Language Models, Multimodal Foundation Models

AI4Science
Healthcare, Linguistics, Sensing, Nutrition

Graph Mining
Graph4NLP

ACADEMIC EMPLOYMENT

University of Hawaii at Manoa (R1)   Aug. 2024 - Present
Tenure-track Assistant Professor in Information and Computer Sciences Department
Director of Advanced Language Processing and Human Analytics (ALOHA) Lab

EDUCATION

University of California, Davis   Jun. 2024
Ph.D. in Computer Science   Advisor: Dr. Jiawei Zhang
Thesis: Building Intelligent and Reliable Summarization Systems

Georgia Institute of Technology   Dec. 2018
M.S. in Computational Science and Engineering
M.S. in Electrical and Computer Engineering

University of Illinois Urbana-Champaign   May 2016
B.E. in Electrical Engineering

INDUSTRY EXPERIENCE

Amazon Web Services (AWS) AI Lab   Santa Clara, CA
Applied Scientist Intern   Jun. 2023 - Sep. 2023

Megagon Labs   Mountain View, CA
Research Intern   Mar. 2023 - Jun. 2023

Tencent AI Lab   Seattle, WA
Research Intern   Jun. 2022 - Sep. 2022

Salesforce Research   Remote
Research Intern   May 2021 - Aug. 2021

Scholarships and Awards

- **Co-PI, NSF Division of Information & Intelligent Systems (IIS)** 2024–2028
  - Project: *SCH: Personalized Machine Learning for Repeat Adverse Health Events using Novel Multimodal Self-Supervised Pretraining Methods*
  - Award Amount: $1.2M (Personal Share: $450K)

- **UH Mānoa Faculty Research Travel Fund** 2025
  - Award Amount: $2,000

- **OpenAI Researcher Access Program** 2025
  - Computing Credits: $7,500

- **Google Cloud Research Credits Program** 2024–2025
  - Computing Credits: $5,000

- **NSF ACCESS Allocations Program** 2024

- **UC Davis Graduate Group in Computer Science (GGCS) Research Fellowship** 2023
  - Award Amount: $15,000

- **FSU Travel Award** 2020

- **Adelaide D. Wilson Graduate Fellowship** 2019

- **UIUC Russell E. Berthold Scholarship** 2015

Publications

Citations: 900 | h-index: 9 (as of Jun. 10, 2025)
* indicates equal contributions | underscore indicates my Ph.D. students.

Conference Proceeding

[1] Yili Ren*, **Haopeng Zhang***, <u>Haohan Yuan</u>, Jingzhe Zhang, and Yitong Shen, "Wi-chat: Large language model powered wi-fi sensing", in *The 3rd International Workshop on Advances in Environmental Sensing Systems for Smart Cities (**EnvSys'25**)*, 2025.

[2] Yili Ren*, **Haopeng Zhang***, <u>Haohan Yuan</u>, Jingzhe Zhang, and Yitong Shen, "Wi-Chat: Large Language Model Powered Wi-Fi Sensing", in *Proceedings of the 23rd ACM International Conference on Mobile Systems, Applications, and Services (**MobiSys'25**) - Posters*, ACM, 2025.

[3] Bardia Konh, Mitchell Kamrava, Rex Imanaka, and **Haopeng Zhang**, "Haptic simulation workstation for prostate brachytherapy training", in *Proceedings of the American Brachytherapy Society (**ABS 2025**) Annual Conference*, American Brachytherapy Society, Nashville, TN, 2025.

[4] Yiqiao Huang, Zhaorun Chen, He Qi, **Haopeng Zhang**, Hanchao Yu, and Zhuokai Zhao, "Boosting multimodal reasoning with implicit chain-of-vision finetuning", in *Multimodal Algorithmic Reasoning workshop at CVPR 2025 (**MAR 2025**)*, 2025.

[5] Zizhong Li, **Haopeng Zhang**, and Jiawei Zhang, "Learning by ranking: Data-efficient knowledge distillation from black-box llms for information retrieval", in *2025 International Joint Conference on Neural Networks (**IJCNN 2025**)*, 2025.

[6] <u>Haohan Yuan</u>, Siu Cheung Hui, and **Haopeng Zhang**, "A structure-aware generative model for biomedical event extraction", in *30th International Conference on Database Systems for Advanced Applications (**DASFAA 2025**)*, 2025.

[7] <u>Haohan Yuan</u> and **Haopeng Zhang**, "DomainSum: A hierarchical benchmark for fine-grained domain shift in abstractive text summarization", in *Findings of the Association for Computational Linguistics: **NAACL 2025***, Albuquerque, New Mexico, Apr. 2025, pp. 2219–2231.

[8] Zizhong Li, **Haopeng Zhang**, and Jiawei Zhang, "Unveiling the magic: Investigating attention distillation in retrieval-augmented generation", in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (**NAACL 2024**)*, Mexico City, Mexico, Jun. 2024, pp. 745–754.

[9] **Haopeng Zhang**, Hayate Iso, Sairam Gurajada, and Nikita Bhutani, "XATU: A fine-grained instruction-based benchmark for explainable text updates", in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (**LREC-COLING 2024**)*, Torino, Italia, May 2024, pp. 17 739–17 752.

[10] **Haopeng Zhang**, Xiao Liu, and Jiawei Zhang, "SummIt: Iterative text summarization via ChatGPT", in *Findings of the Association for Computational Linguistics: (**EMNLP 2023**)*, Singapore, Dec. 2023, pp. 10 644–10 657.

[11] **Haopeng Zhang**, Xiao Liu, and Jiawei Zhang, "Extractive summarization via ChatGPT for faithful summary generation", in *Findings of the Association for Computational Linguistics: (**EMNLP 2023**)*, Singapore, Dec. 2023, pp. 3270–3278.

[12] **Haopeng Zhang**, Sangwoo Cho, Kaiqiang Song, Xiaoyang Wang, Hongwei Wang, Jiawei Zhang, and Dong Yu, "Unsupervised multi-document summarization with holistic inference", in *Findings of the Association for Computational Linguistics: (**IJCNLP-AACL 2023**)*, Nusa Dua, Bali, Nov. 2023, pp. 123–133.

[13] **H. Zhang***, X. Liu*, and Jiawei Zhang, "DiffuSum: Generation enhanced extractive summarization with diffusion", in *Findings of the Association for Computational Linguistics: (**ACL 2023**)*, Toronto, Canada, Jul. 2023, pp. 13 089–13 100.

[14] **Haopeng Zhang**, Xiao Liu, and Jiawei 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, Jul. 2023, pp. 37–47.

[15] **Haopeng Zhang**, Xiao Liu, and Jiawei 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, Dec. 2022, pp. 10 167–10 176.

[16] **Haopeng Zhang**, Semih Yavuz, Wojciech Kryscinski, Kazuma Hashimoto, and Yingbo Zhou, "Improving the faithfulness of abstractive summarization via entity coverage control", in *Findings of the Association for Computational Linguistics: **NAACL 2022***, Seattle, United States, Jul. 2022, pp. 528–535.

[17] **Haopeng Zhang** and Jiawei Zhang, "Text graph transformer for document classification", in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2020**)*, Online, Nov. 2020, pp. 8322–8327.

Journal

[18] **Haopeng Zhang**, Philip S. Yu, and Jiawei Zhang, "A systematic survey of text summarization: From statistical methods to large language models", *ACM Computing Surveys*, Apr. 2025, ISSN: 0360-0300.

[19] **Haopeng Zhang**, "Building intelligent and reliable summarization systems", Ph.D. dissertation, University of California, Davis, 2024.

Preprint

[20] William O'Grady, **Haopeng Zhang**, and Miseon Lee, "The negative concord mystery", *arXiv preprint*, 2025.

[21] Chengqi Zheng, Jianda Chen, Yueming Lyu, Wen Zheng Terence Ng, **Haopeng Zhang**, Yew-Soon Ong, Ivor Tsang, and Haiyan Yin, "Mermaidflow: Redefining agentic workflow generation via safety-constrained evolutionary programming", *arXiv preprint arXiv:2505.22967*, 2025.

[22] Haohan Yuan and **Haopeng Zhang**, "Strucsum: Graph-structured reasoning for long document extractive summarization with llms", *arXiv preprint*, 2025.

[23] Kaiying Lin, Hsiyu Chen, and **Haopeng Zhang**, "Formosanbench: Benchmarking low-resource austronesian languages in the era of large language models", *arXiv preprint*, 2025.

[24] **Haopeng Zhang**, "Multimodal summarization of text, images, and video: A unified survey", *arXiv preprint*, 2025.

[25] Zizhong Li, **Haopeng Zhang**, and Jiawei Zhang, "Token-level precise attack on rag: Searching for the best alternatives to mislead generation", *arXiv preprint*, 2025.

[26] Zizhong Li, **Haopeng Zhang**, and Jiawei Zhang, "A revisit of fake news dataset with augmented fact-checking by chatgpt", *arXiv preprint:2312.11870*, 2023.

[27] Jiawei Zhang, **Haopeng Zhang**, Li Sun, and Congying Xia, "Graph-bert: Only attention is needed for learning graph representations", *arXiv preprint:2001.05140*, 2020.

Patent

[28] **Haopeng Zhang**, Semih Yavuz, Wojciech Kryscinski, Kazuma Hashimoto, and Yingbo Zhou, *Systems and methods for abstractive document summarization with entity coverage control*, US Patent 11741142, Aug. 2023.

PROFESSIONAL SERVICE

- **University Service**
  - Affiliate Member, University of Hawai'i Cancer Center (UHCC) — 2025–Present
  - Member, Hawai'i Data Science Institute — 2024–Present
  - Member, Hawai'i EPSCoR Change Hawai'i (HI) Initiative — 2024–Present
  - Graduate Studies Committee Member, ICS department — 2024–Present
  - Data Science Committee Member, Communication and Information Sciences (CIS) Program — 2024–Present

- **Panel and Proposal Review Service**
  - Panelist, NSF CISE Program — 2025, 2024

- **Conference and Workshop Service**
  - Organizer, EMNLP 2025 Workshop on New Frontiers in Summarization (NewSumm) — 2025
  - Area Chair, Conference on Empirical Methods in Natural Language Processing (EMNLP) — 2025
  - Area Chair, Annual Meeting of the Association for Computational Linguistics (ACL) — 2025
  - Area Chair, ACL Rolling Review (ARR) — 2025
  - Minitrack Chair, Americas Conference on Information Systems (AMCIS) — 2025

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

- IEEE Transactions on Big Data
- ACM Transactions on Embedded Computing Systems
- IEEE Transactions on Computational Social Systems
- IEEE Computer Magazine
- Computer Speech and Language (Elsevier)
- Artificial Intelligence Review (Springer)
- Expert Systems with Applications (Elsevier)
- Engineering Applications of Artificial Intelligence (Elsevier)
- PeerJ Computer Science
- Frontiers in Artificial Intelligence

- **Conference Reviewer & Program Committee Member**
  - Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): 2024, 2025
  - Annual Meeting of the Association for Computational Linguistics (ACL): 2023, 2024
  - Conference on Empirical Methods in Natural Language Processing (EMNLP): 2023, 2024
  - ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD): 2023, 2024
  - International Joint Conference on Neural Networks (IJCNN): 2023, 2024
  - SIAM International Conference on Data Mining (SDM): 2024
  - International Conference on Pattern Recognition (ICPR): 2024
  - ACL Rolling Review (ARR): 2023, 2024
  - Conference of the European Chapter of the Association for Computational Linguistics (EACL): 2023
  - EMNLP 2023 Workshop on New Frontiers in Summarization (NewSumm): 2023
  - ACL 2023 Workshop on Computational Approaches to Discourse (CODI): 2024
  - AAAI Conference on Artificial Intelligence (AAAI): 2021, 2022

Teaching Experience

$\dagger$ denotes graduate-level courses.

- **Instructor, University of Hawai'i at Mānoa**
  - ICS 691† : Natural Language Processing and Large Language Models   Fall 2025
  - ICS 635† : Machine Learning   Spring 2025
  - ICS 435: Machine Learning Fundamentals   Spring 2025
  - ICS 661† : Advanced Artificial Intelligence   Fall 2024

- **Teaching Assistant, University of California, Davis**
  - ECS 189G: Deep Learning   Winter 2022

- **Teaching Assistant, Florida State University**
  - COP 5725† : Advanced Database Systems   Fall 2021, Spring 2021
  - COP 4531: Complexity and Analysis of Data Structures and Algorithms   Spring 2020
  - COP 4710: Theory and Structure of Databases   Fall 2020, Fall 2019

- **Teaching Assistant, Georgia Institute of Technology**
  - ECE 3450: Semiconductor Devices   Fall 2017, Spring 2017

Mentorship

Underscore indicates my Ph.D. students.

<u>Cade Kane</u> *Ongoing*
Ph.D. Student, University of Hawai'i at Mānoa
Research Area: Hawaiian Natural Language Processing

<u>Haohan Yuan</u> *Ongoing*
Ph.D. Student, University of Hawai'i at Mānoa
Research Area: Information Extraction
Collaborated Publications: [6], [7], [22]

Zizhong Li *Ongoing*
Ph.D. Student, University of California, Davis
Research Area: Retrieval-augmented Generation
Collaborated Publications: [5], [8], [25], [26]

Xiao Liu *Ongoing*
Ph.D. Candidate, University of California, Davis
Research Area: Text Summarization
Collaborated Publications: [10], [11], [13], [14], [15]