Hejie Cui

Postdoctoral Scholar
Stanford University

E-mail: hejie.cui@stanford.edu
Homepage: hejiecui.com

Research Interests

Health AI: large language models and large vision-language models post-training for healthcare applications.
Multimodal Knowledge Extraction: multimodal knowledge extraction and integration with images, texts, tabular data, etc.
Data Science and Data Mining: data mining, health data science.

Academic Experience

Stanford University, Palo Alto, CA, USA   May 2024 – Now
Postdoc Scholar, Department of Biomedical Informatics & Department of Computer Science
Advisor: Prof. Nigam Shah (Medicine), Prof. Sanmi Koyejo (CS)
Research Topic: *Large Language Model (LLM) Evaluation and Post-training for Healthcare*

Emory University, Atlanta, GA, USA   Sept 2019 – May 2024
Ph.D., Department of Computer Science
Advisor: Prof. Carl J. Yang
2023 EECS Rising Star, Laney-EDGE Graduate School Diverse Scholars in the Sciences
Thesis Title: *AI-Assisted Healthcare with Multimodal Structured Knowledge Extraction and Augmented Reasoning*
Thesis Committee: Dr. Carl J. Yang (Emory), Dr. Joyce C Ho (Emory), Dr. Yana Bromberg (Emory), Dr. Fei Wang (Cornell)

Emory University, Atlanta, GA, USA   Sept 2019 – Dec 2022
Master of Science in Computer Science, in the Middle of PhD Program

Tongji University, Shanghai, CHINA   Sept 2015 – Jul 2019
B.E., School of Software Engineering
Advisor: Prof. Lin Zhang
Summa cum laude, Valedictorian
GPA: 4.96/5.0, Rank: 1/164

Professional Experience

Microsoft Research, Redmond, WA, USA   May 2023 – Aug 2023
Research Intern @ Augmented Learning and Reasoning Group
Worked on serendipity chat message recommendations for the Teams platform.

Amazon, Seattle, WA, USA   May 2022 – Aug 2022
Applied Scientist Intern @ Product Knowledge Graph Team
Worked on attribute information extraction from texts and images on the Amazon shopping platform.

Google, Remote   Jan 2022 – Apr 2022
Google's CS Research Mentorship Program
Research on graph neural networks for retrieval and recommendation on social networks.

Sensetime Research, Beijing, CHINA   Dec 2018 – Jul 2019
Research Intern @ Intelligent Health Team
Worked on pulmonary vessel segmentation for 3D chest CT images.

Queens University, Kingston, ON, CANADA   Jul 2018 – Oct 2018
Research Intern @ Perk Lab (The Laboratory for Percutaneous Surgery)
Worked on real-time web-cam video images classification.
Advisor: Prof. Gabor Fichtinger

Microsoft, Beijing, CHINA   Aug 2017 – Aug 2017
Visiting Student @ Microsoft Research Asia
Worked on HoloLens mixed reality applications.
Best work award for university summer camp Hackathon.

Honors and Awards

Publications [Google Scholar]

Peer-Reviewed Publications

[NeurIPS 2024]   **Biomedical Visual Instruction Tuning with Clinician Preference Alignment**
Hejie Cui, Lingjun Mao, Xin Liang, Jieyu Zhang, Hui Ren, Quanzheng Li, Xiang Li, Carl Yang
*Neural Information Processing Systems (**NeurIPS**), 2024*

[NeurIPS 2024]   **Microstructures and Accuracy of Graph Recall by Large Language Models**
Yanbang Wang, **Hejie Cui**, Jon Kleinberg
*Neural Information Processing Systems (**NeurIPS**), 2024*

[AMIA 2024]   **LLMs-based Few-Shot Disease Predictions using EHR: A Novel Approach Combining Predictive Agent Reasoning and Critical Agent Instruction**
Hejie Cui, Zhuocheng Shen, Jieyu Zhang, Hui Shao, Lianhui Qin, Joyce C Ho, Carl Yang
*American Medical Informatics Association Annual Symposium, (**AMIA-Annual Symposium**), Oral, 2024*

[ACL 2024]   **Knowledge-Infused Prompting: Assessing and Advancing Clinical Text Data Generation with Large Language Models**
Ran Xu, **Hejie Cui**, Yue Yu, Xuan Kan, Wenqi Shi, Yuchen Zhuang, Wei Jin, Joyce Ho, Carl Yang
*Annual Meeting of the Association for Computational Linguistics **(ACL-Findings)**, 2023*
NeurIPS Workshop on SyntheticData4ML, 2023

[KDD 2024]   **TACCO: Task-guided Co-clustering of Clinical Concepts and Patient Visits for Disease Subtyping based on EHR Data**
Ziyang Zhang, **Hejie Cui**, Ran Xu, Yuzhang Xie, Joyce C Ho, Carl Yang
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2024

[ISBI 2024]   **Multi-task Learning for Brain Connectome Analysis in the ABCD study**
Xuan Kan, **Hejie Cui**, Keqi Han, Ying Guo, Carl Yang
IEEE International Symposium on Biomedical Imaging (ISBI), 2024

[IC2S2 2024]   **Network Recall by Large Language Models**
Yanbang Wang, **Hejie Cui**, Jon Kleinberg
*International Conference on Computational Social Science, (**IC2S2**), Oral, 2024*

[SPIE 2024]   **Federated Learning for Cross-Institution Brain Network Analysis**
Han Xie, Yi Yang, **Hejie Cui**, Carl Yang

*Proceedings of the SPIE Medical Imaging Conference (**SPIE**), **Oral**, 2024*

[NeurIPS 2023]   **Open Visual Knowledge Extraction via Relation-Oriented Multimodality Model Prompting**
Hejie Cui, Xinyu Fang, Zihan Zhang, Ran Xu, Xuan Kan, Xin Liu, Manling Li, Yangqiu Song, Carl Yang
*Neural Information Processing Systems (**NeurIPS**), 2023*
ICML Workshop of Knowledge and Logical Reasoning in the Era of Data-driven Learning, 2023

[PSB 2024]   **BrainSTEAM: A Practical Pipeline for Connectome-based fMRI Analysis towards Subject Classification**
Alexis Li (Mentored High-Schooler), Owen Yang, **Hejie Cui**, Carl Yang
*Pacific Symposium on Biocomputing (**PSB**), **Oral**, 2024*

[PSB 2024]   **FedBrain: Federated Training of GNNs for Connectome-based Brain Imaging Analysis**
Yi Yang, Han Xie, **Hejie Cui**, Carl Yang
*Pacific Symposium on Biocomputing (**PSB**), 2024*

[IEEE-BHI 2023]   **Dynamic Brain Transformer with Multi-Level Attention for Functional Brain Network Analysis**
Xuan Kan, Tony Gu, **Hejie Cui**, Ying Guo, Carl Yang
*International Conference on Biomedical and Health Informatics (**IEEE-BHI**), 2023*

[KDD 2023]   **R-Mixup: Riemannian Mixup for Biological Networks**
Xuan Kan, Zimu Li, **Hejie Cui**, Yue Yu, Ran Xu, Shaojun Yu, Zilong Zhang, Ying Guo, Carl Yang
*ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**), 2023*

[ACL 2023]   **PV2TEA: Patching Visual Modality to Textual-Established Information Extraction**
Hejie Cui, Rongmei Lin, Nasser Zalmout, Chenwei Zhang, Jingbo Shang, Carl Yang, Xian Li
*Annual Meeting of the Association for Computational Linguistics (**ACL-Findings**), 2023*
*KDD Workshop on Knowledge Augmented Methods for NLP (**KnowledgeNLP**), **Oral**, 2023*

[CHIL 2023]   **PTGB: Pre-Train Graph Neural Networks for Brain Network Analysis**
Yi Yang, **Hejie Cui**, Carl Yang
*The Conference on Health, Inference, and Learning (**CHIL**), **Oral**, 2023*

[ISBI 2023]   **Deep DAG Learning of Effective Brain Connectivity for fMRI Analysis**
Yue Yu, Xuan Kan, **Hejie Cui**, Ran Xu, Yujia Zheng, Xiangchen Song, Yanqiao Zhu, Kun Zhang, Razieh Nabi, Ying Guo, Chao Zhang, Carl Yang
*IEEE International Symposium on Biomedical Imaging (**ISBI**), 2023*
Extended Abstract accepted to the IEEE International Conference on Big Data

[ISBI 2023]   **Transformer-based Hierarchical Clustering for Brain Network Analysis**
Wei Dai, **Hejie Cui**, Xuan Kan, Ying Guo, Sanne van Rooij, Carl Yang
*IEEE International Symposium on Biomedical Imaging (**ISBI**), 2023*
Extended Abstract accepted to the IEEE International Conference on Big Data

[AAAI 2023]   **Neighborhood-regularized Self-Training for Learning with Few Labels**
Ran Xu, Yue Yu, **Hejie Cui**, Xuan Kan, Yanqiao Zhu, Joyce C. Ho, Chao Zhang, Carl Yang
*International Conference on Artificial Intelligence (**AAAI**), **Oral**, 2023*

[IEEE TMI]   **BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks**
Hejie Cui, Wei Dai, Yanqiao Zhu, Xuan Kan, Antonio Aodong Chen Gu, Joshua Lukemire, Liang Zhan, Lifang He, Ying Guo, Carl Yang
*IEEE Transactions on Medical Imaging (**TMI**), **IF: 11.037**, **167 GitHub Stars**, 2022*

[BIBM 2022]   **Multi-View Brain Network Analysis with Cross-View Missing Network Generation**
Gongxu Luo, Chenyang Li, **Hejie Cui**, Lichao Sun, Lifang He, Carl Yang
*IEEE International Conference on Bioinformatics and Biomedicine (**BIBM**), 2022*

[NeurIPS 2022]   **Brain Network Transformer**
Xuan Kan, Wei Dai, **Hejie Cui**, Zilong Zhang, Ying Guo, Carl Yang
*Proceedings of the Conference on Neural Information Processing Systems (**NeurIPS**), **Spotlight**, 2022*

[CIKM 2022]	**On Positional and Structural Node Features for Graph Neural Networks on Non-attributed Graphs** **Hejie Cui**, Zijie Lu, Pan Li, Carl Yang *International Conference on Information and Knowledge Management (**CIKM** Short)*, **Spotlight**, *2022* **Most Influential CIKM Paper of 2022**
[KDD 2022]	**Data-Efficient Brain Connectome Analysis via Multi-Task Meta-Learning** Yi Yang*, Yanqiao Zhu*, **Hejie Cui**, Xuan Kan, Lifang He, Ying Guo, Carl Yang *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD**)*, *2022*
[MICCAI 2022]	**Interpretable GNNs for Connectome-Based Brain Disorder Analysis** **Hejie Cui**, Wei Dai, Yanqiao Zhu, Xiaoxiao Li, Lifang He, Carl Yang *International Conference on Medical Image Computing & Computer-Assisted Intervention (**MICCAI**)*, **Oral**, *2022*
[EMBC 2022]	**Joint Embedding of Structural and Functional Brain Networks with Graph Neural Networks for Mental Illness Diagnosis** Yanqiao Zhu, **Hejie Cui**, Lifang He, Lichao Sun, Carl Yang *International Conference of the IEEE Engineering in Medicine and Biology Society (**EMBC**)*, *2022*
[MIDL 2022]	**FBNetGen: Task-aware GNN-based fMRI Analysis via Functional Brain Network Generation** Xuan Kan, **Hejie Cui**, Joshua Lukemire, Ying Guo, Carl Yang *International Conference on Medical Imaging with Deep Learning (**MIDL**)*, **Oral**, *2022*
[SDM 2022]	**Structure-Aware Hard Negative Mining for Heterogeneous Graph Contrastive Learning** Yanqiao Zhu, Yichen Xu, **Hejie Cui**, Carl Yang, Qiang Liu, Shu Wu *SIAM International Conference on Data Mining (**SDM**)*, *2022*
[ECIR 2022]	**How Can Graph Neural Networks Help Document Retrieval: A Case Study on CORD19 with Concept Map Generation** **Hejie Cui**, Jiaying Lu, Yao Ge, Carl Yang *The European Conference on Information Retrieval (**ECIR** Short)*, *2022*
[ECML 2021]	**Zero-Shot Scene Graph Relation Prediction through Commonsense Knowledge Integration** Xuan Kan, **Hejie Cui**, Carl Yang *The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (**ECML-PKDD**)*, *2021*
[MICCAI 2019]	**Pulmonary Vessel Segmentation based on Orthogonal Fused U-Net++ of Chest CT Images** **Hejie Cui**, Xinglong Liu, Ning Huang *International Conference on Medical Image Computing and Computer-Assisted Intervention (**MICCAI**)*, *2019*
[ICBC 2018]	**A Method to Predict the Performance and Storage of Executing Contract for Ethereum Consortium-Blockchain** Huijuan Zhang, Chengxin Jin, **Hejie Cui** *IEEE International Conference on Blockchain (**ICBC**)*, *2018*

Preprints and Workshop

[Preprint]	**A Survey on Knowledge Graphs for Healthcare: Resources, Applications, and Promises** **Hejie Cui**, Jiaying Lu, Shiyu Wang, Ran Xu, Wenjing Ma, Shaojun Yu, Yue Yu, Xuan Kan, Chen Ling, Joyce Ho, Fei Wang, Carl Yang *ICML Workshop on Interpretable Machine Learning in Healthcare, 2023* *Under Review*
[Preprint]	**Domain Specialization as Key to Make Large Language Models Disruptive: A Comprehensive Survey** Chen Ling, Xujiang Zhao, Jiaying Lu, Chengyuan Deng, Can Zheng, Junxiang Wang, Tanmoy Chowdhury, Yun Li, **Hejie Cui**, Xuchao Zhang, Tianjiao Zhao, Amit Panalkar, Wei Cheng, Haoyu Wang, Yanchi Liu, Zhengzhang Chen, Haifeng Chen, Chris White, Quanquan Gu, Jian Pei, Liang Zhao **Honorably mentioned by the 2024 Economic Report of the President from the White House** *Under Review*

[Preprint]	**Multimodal Fusion of EHR in Structures and Semantics: Integrating Clinical Records and Notes with Hypergraph and LLM**
	Hejie Cui, Xinyu Fang, Ran Xu, Xuan Kan, Joyce C Ho, Carl Yang
	On arXiv

Tutorials

[ISBI 2024]	**Brain Network Analysis with Graph Neural Network**
	Hejie Cui, Xuan Kan, Ying Guo, Xiaoxiao Li, Lifang He, Liang Zhan, Carl Yang
	*IEEE International Symposium on Biomedical Imaging (**ISBI**), Tutorials, 2024*

[ICIBM 2023]	**Brain Network Analysis with Graph Neural Network**
	Hejie Cui, Xuan Kan, Xiaoxiao Li, Lifang He, Liang Zhan, Ying Guo, Carl Yang
	*International Conference on Intelligent Biology and Medicine (**ICIBM**), Tutorials, 2023*

Talks & Presentations

Invited Talks

Towards Trustworthy Foundation Model Adaptation and Evaluation for Health	University of Michigan AI Symposium 2024
Advancing Healthcare with Multimodal Structural Knowledge	Standford/UCLA/Harvard 2023
Serendipitous Recommendation over Communication Graph via Group Fairness	Microsoft Research 2023
Introduction to Multimodal Learning and Knowledge Acquisition	Amazon Inc. 2022
GNNs for Brain Network	University of Rochester 2022
Effective GNNs for Multimodality Graph Mining	Microsoft AI 2022

Conference & Workshop Oral Presentations

PV2TEA: Patching Visual Modality to Textual-Established Information Extraction	KnowledgeNLP-KDD'23
PTGB: Pre-Train Graph Neural Networks for Brain Network Analysis	CHIL'23
Interpretable GNNs for Connectome-Based Brain Disorder Analysis	MICCAI'22
BrainNNExplainer	Interpretable Machine Learning in Healthcare ICML'21
Node Feature Choice for GNNs on Non-attributed Graphs	Deep Learning on Graphs KDD'21

Open-Source Projects

BrainGB [Link]
The first benchmark for brain network analysis with graph neural networks.
With a Python pip package, tutorials, and benchmark website.
Received 179 GitHub stars.

Teaching Experience

CS 253 Data Structures and Algorithms, Exercise Lesson Instructor and Head TA	Spring 2021
CS 584 Natural Language Processing for Biomedical Applications, Head TA	Fall 2020
CS 584 Deep Learning, Head TA	Spring 2020

Mentoring

Lingjun Mao, Undergraduate RA at UC Berkeley	12/2023 – 05/2024
Co-authored paper: NeurIPS'24	
Alexis Li (Female), Senior High School Student in Hamilton High School at Chandler	08/2022 – 03/2023
Co-authored paper: PSB'24	
David Dai, Undergraduate RA at Emory University, Master Student in Stanford	05/2021 – 11/2022
Co-authored paper: MICCAI'22, ISBI'23	

Owen Yang, Undergraduate RA at Emory University, PhD in Duke   02/2022 – Present
Co-authored paper: KDD'22, CHIL'23, PSB'24

Tony Gu, Undergraduate RA at Georgia Institute of Technology   01/2022 – 08/2023
Co-authored paper: TMI, IEEE-BHI'23

Xinyu Fang, Undergraduate RA at Emory University   05/2022 – Present
Co-authored paper: NeurIPS'23, ICML-KLR'23

Services

Program Committee Member | Reviewer

International Conference on Machine Learning (**ICML**)	Since 2024
Annual Meeting of the Association for Computational Linguistics (**ACL**)	Since 2024
The International Conference on Learning Representations (**ICLR**)	Since 2024
Conference on Health, Inference, and Learning (**CHIL**)	Since 2023
International Conference on Learning Representations (**ICLR**)	Since 2023
Machine Learning for Health (**ML4H**)	Since 2023
Conference on Neural Information Processing Systems (**NeurIPS**)	Since 2023
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**SIGKDD**)	Since 2022
International Conference on Data Mining (**ICDM**)	Since 2022
The Web Conference (**WWW**)	2022
Conference on Artificial Intelligence, Main Track and Social Impact Track (**AAAI**)	2022
International Conference on Medical Image Computing and Computer Assisted Intervention (**MICCAI**)	Since 2020
SIAM International Conference on Data Mining (**SDM**)	2022
Learning on Graphs Conference (**LOG**)	2022
Medical Image Analysis (**MIA**)	Since 2023
IEEE Transactions on Medical Imaging (**TMI**)	Since 2022
IEEE Transactions on Knowledge and Data Engineering (**TKDE**)	Since 2022
IEEE Transactions on Big Data (**Big Data**)	Since 2021
NeurIPS Workshop on Generative AI & Biology	2023
ICML Workshop on Knowledge and Logical Reasoning in the Era of Data-driven Learning	2023
NeurIPS Workshop on AI for Science: Progress and Promises	2022
NeurIPS Workshop on Temporal Graph Learning	2022
CIKM Workshop on Federated Learning with Graph Data	2022
ICML Workshop on Interpretable Machine Learning in Healthcare	Since 2021
ICCV Workshop on Computer Vision for Automated Medical Diagnosis	2021

Workshop Organizer

IEEE BigData Workshop on Neural Networks for Brain Connectome Analysis: Theories, Methods, and Applications	2022

Volunteer

Conference on Health, Inference, and Learning (**CHIL**)	2023
Conference on Information and Knowledge Management (**CIKM**)	2022
International Conference on Machine Learning (**ICML**)	2021, 2020

Diversity, Equity, and Inclusion (DEI)

Emory Laney-EDGE Graduate School Diverse Scholars in the Sciences
Member of Emory OpenBMI
Member of Emory ACM-W Women in Computing Chapter