

**Homepage:** <https://chunhuizng.github.io>

**Email:** [chunhui.cheung@gmail.com](mailto:chunhui.cheung@gmail.com)

**Address:** 15 Thayer Dr, Dartmouth College, Hanover, NH 03755

**Phone:** 1-7816521380

## Research Focus

- **Reasoning & Planning:** MCTS-LLM, PRM-RL, working memory
- **Multimodality:** vision, audio (speech/music), and language
- **Post-Training:** instruction tuning, DPO, reinforcement learning
- **Mechanistic interpretability:** Memory for long-context modelling

## Education

**Dartmouth College** Hanover, NH, USA  
Ph.D. in Computer Science Aug. 2023 – Apr. 2026 (expected)  
Advisor: [Professor Soroush Vosoughi](#)

**Brandeis University** Waltham, MA, USA  
M.S. in Computer Science Sep. 2021 – Jun. 2023  
GSAS Research Fellowship Recipient

**Northeastern University** China  
B.S. in Computer Science Sep. 2017 – Jun. 2021  
Outstanding Honor Thesis Award

## Experience

**Honda Research Institute USA** San Jose, CA, USA  
Research Intern Jun. 2024 – Sep. 2024  
*Project: Multimodal LLM Post-Training (8B - 70B) and Synthetic Data Generation*  
Developed a **global-local multi-agent planning system** for understanding human behaviors in **multimodal environments**, achieving **4.6%** improvement over state-of-the-art solutions. Implemented post-trained small agents to facilitate inference-time alignment and scaling to 405B parameters. Research paper under review at ICML 2025.  
Host: [Dr. Shao-Yuan Lo](#)

## Under Review

**Overcoming Multi-step Complexity in Theory-of-Mind Reasoning: A Scalable Bayesian Planner**  
[ICML 2025 submission](#) | [Code](#) | *A global-local structure that uses Bayesian inverse planning for global planning, then allows LLMs to fully focus on local reasoning.*

**Chunhui Zhang**, Sean Dae Houlihan, Kwonjoon Lee, Nakul Agarwal, Zhongyu Ouyang, Soroush Vosoughi, Shao-Yuan Lo

**Growing Through Experience: Scaling Episodic Grounding in Language Models**

[ACL 2025 submission](#) | [Code](#) | *Post-trained agentic LLMs (DPO on 3B, 8B, and 70B models) on MCTS-sampled data from physical simulators to enhance planning capabilities.*

**Chunhui Zhang**, Elsie Wang, Zhongyu Ouyang, Xiangchi Yuan, Soroush Vosoughi

**Knowing More, Acting Better: Hierarchical Representation for Embodied Decision-Making**

[ACL 2025 submission](#) | *Post-trained Multimodal LLMs as embodied agents in physical indoor environments.*

**Chunhui Zhang**, Zhongyu Ouyang, Zheyuan Liu, Soroush Vosoughi

**Publications**

**Pretrained Image-Text Models are Secretly Video Captioners**

[Preprint](#) | [Code](#)

*Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025) Main Conference.*

*An RL (reinforcement learning) post-training recipe that trains a **Top-2** multimodal LLM captioner on PaperswithCode Leaderboard, outperforming industry MLLM captioners.*

**Chunhui Zhang\***, Yiren Jian\*, Zhongyu Ouyang, Soroush Vosoughi

**Temporal Working Memory: Query-Guided Segment Refinement for Enhanced Multimodal Understanding**

[Preprint](#) | [Code](#)

*Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025) Findings.*

*Extending my previous papers from NAACL 2025 and EMNLP 2025, this research utilizes working memory to enable multimodal LLMs to handle long multimodal contexts.*

{Xingjian Diao\*, **Chunhui Zhang\***}, Weiyi Wu, Zhongyu Ouyang, Pei-jun Qing, Ming Cheng, Soroush Vosoughi, Jiang Gui

Is It Navajo? Accurate Language Detection in Endangered Athabaskan Languages

[Preprint](#) | [Code](#)

*Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025) Main Conference.*

Ivory Yang, Weicheng Ma, **Chunhui Zhang**, Soroush Vosoughi

Working Memory Identifies Reasoning Limits in Language Models

**Chunhui Zhang**, Yiren Jian, Zhongyu Ouyang, Soroush Vosoughi

*Accepted to the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).*

Learning Musical Representations for Music Performance Question Answering

Xingjian Diao, **Chunhui Zhang**, Tingxuan Wu, Ming Cheng, Zhongyu Ouyang, Weiye Wu, Soroush Vosoughi, Jiang Gui

*Findings of the Association for Computational Linguistics: Empirical Methods in Natural Language Processing (Findings of EMNLP), 2024.*

Expedited Training of Visual Conditioned Language Generation via Redundancy Reduction

Yiren Jian, Tingkai Liu, Yunzhe Tao, **Chunhui Zhang**, Soroush Vosoughi, Hongxia Yang

*Annual Meeting of the Association for Computational Linguistics (ACL, Oral Presentation), 2024.*

Aligning Relational Learning with Lipschitz Fairness

{Yaning Jia, **Chunhui Zhang**}, Soroush Vosoughi.

*International Conference on Learning Representations (ICLR), 2024.*

*Note: Co-first author Jia was a master student who was mentored by me. Thanks Jia.*

Mitigating Emergent Robustness Degradation on Graphs while Scaling-up

{Xiangchi Yuan, **Chunhui Zhang**}, Yijun Tian, Yanfang Ye, et al.

*International Conference on Learning Representations (ICLR), 2024.*

*Note: Co-first author Yuan was a master student who was mentored by me. Thanks Yuan.*

Graph Mixed Supervised Learning via Generalized Knowledge

Xiangchi Yuan, Yijun Tian, **Chunhui Zhang**, Yanfang Ye, Nitesh V Chawla, et al.

*ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2024.*

GCVR: Reconstruction from Cross-View Enable Sufficient and Robust Graph Contrastive Learning

Qianlong Wen, Zhongyu Ouyang, **Chunhui Zhang**, Yiyue Qian, Chuxu Zhang, Yanfang Ye

*The Conference on Uncertainty in Artificial Intelligence (UAI), 2024.*

Symbolic Prompt Tuning Completes the App Promotion Graph

Zhongyu Ouyang, **Chunhui Zhang**, Shifu Hou, Shang Ma, Chaoran Chen, Toby Li, Xusheng Xiao, et al.

*European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), 2024*

How to Improve Representation Alignment and Uniformity in Graph-based Collaborative Filtering?

Zhongyu Ouyang, **Chunhui Zhang**, Shifu Hou, Chuxu Zhang, Yanfang Ye

*International AAAI Conference on Web and Social Media (ICWSM), 2024.*

Breaking the Trilemma of Privacy, Utility, and Efficiency via Controllable Machine Unlearning

{Zheyuan Liu, Guangyao Dou}, Yijun Tian, **Chunhui Zhang**, Eli Chien, Ziwei Zhu

*ACM International World Wide Web Conference (WWW/TheWebConf), 2024.*

When Sparsity Meets Contrastive Models: Less Data Can Bring Better Class-Balanced Representations

**Chunhui Zhang**, Chao Huang, Yijun Tian, Qianlong Wen, et al.

*International Conference on Machine Learning (ICML), 2023.* AAAI-DCAA 2023 Best Paper Runner-up Award

Chasing All-Round Graph Representation Robustness: Model, Training, and Optimization

**Chunhui Zhang**, Yijun Tian, Mingxuan Ju, Zheyuan Liu, et al.

*International Conference on Learning Representations (ICLR), 2023.*

Mind the Gap: Mitigating the Distribution Gap in Graph Few-shot Learning

**Chunhui Zhang**, Hongfu Liu, Jundong Li, Yanfang Ye, et al.

*Transactions on Machine Learning Research (TMLR), 2023.*

Fair Graph Representation Learning via Diverse Mixture-of-Experts  
{Zheyuan Liu, **Chunhui Zhang**}, Yijun Tian, Erchi Zhang, et al.  
*ACM International World Wide Web Conference (WWW/TheWebConf), 2023.*

*Note: Co-first author Liu (in alphabetical order) was an undergraduate who was mentored by me. Thanks Liu.*

Boosting Graph Neural Networks via Adaptive Knowledge Distillation  
Zhichun Guo, **Chunhui Zhang**, Yujie Fan, Yijun Tian, et al.  
*AAAI Conference on Artificial Intelligence (AAAI), 2023.*

Heterogeneous Graph Masked Autoencoders  
Yijun Tian, Kaiwen Dong, **Chunhui Zhang**, et al.  
*AAAI Conference on Artificial Intelligence (AAAI), 2023.*

Heterogeneous Temporal Graph Neural Network Explainer  
Jiazheng Li, **Chunhui Zhang**, Chuxu Zhang.  
*ACM International Conference on Information and Knowledge Management (CIKM), 2023.*

Label-invariant Augmentation for Semi-Supervised Graph Classification  
Han Yue, **Chunhui Zhang**, Chuxu Zhang, and Hongfu Liu.  
*Conference on Neural Information Processing Systems (NeurIPS), 2022.*

Co-Modality Imbalanced Graph Contrastive Learning  
Yiyue Qian, **Chunhui Zhang**, Yiming Zhang, Qianlong Wen, Yanfang Ye, et al.  
*Conference on Neural Information Processing Systems (NeurIPS), 2022.*

Look Twice as Much as You Say: Scene Graph Contrastive Learning for Self-Supervised Image Caption Generation  
**Chunhui Zhang**, Chao Huang, Youhuan Li, Xiangliang Zhang, Yanfang Ye, et al.  
*ACM International Conference on Information and Knowledge Management (CIKM), 2022.*

GraphBERT: Bridging Graph and Text for Malicious Behavior Detection on Social Media

Jiele Wu, **Chunhui Zhang**, Zheyuan Liu, Erchi Zhang, Steven Wilson, et al.

*IEEE International Conference on Data Mining (ICDM)*, 2022.

Towards Tailored Models on Private AIoT Devices: Federated Direct Neural Architecture Search

**Chunhui Zhang**, Xiaoming Yuan, Qianyun Zhang, Guangxu Zhu, Lei Cheng, and Ning Zhang.

*IEEE Internet of Things Journal (IEEE-IoTJ)*, Feb. 2022.

## Honors and scholarships

ACL Oral Presentation Award	2024
Graduate School of Arts and Sciences Fellowship	2021 – 2023
GSAS Ph.D. Student Conference Award	2023
Travel and Research Grant	2022
CIKM Travel Grant Award	2022
AAAI-DCAA Best Paper Runner-up Award	2023

## Teaching experience

**Teaching Assistant, Computer Science, Brandeis** Fall 2021 & Spring 2023

CS 133A: Graph Mining

Graphs are capable of modeling complex social, technological, and biological systems. This course covers the core concepts, models, and algorithms of graph mining.

**Teaching Assistant, Computer Science, Brandeis** Spring & Fall 2022

CS 165B: Deep Learning

This course covers the core methods and algorithms of deep learning techniques.

## Service and outreach

**Program Committee/Conference Reviewer**

NeurIPS 2023, NeurIPS Datasets and Benchmarks track 2023, AAAI 2023, Learning on Graphs 2023, NeurIPS 2022, CIKM 2022, ICDM 2022, IEEE HPCC 2020

**Journal Reviewer**

IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Network Science and Engineering, ACM Transactions on Intelligent Systems and Technology, Neurocomputing, Big Data