

Chia-Yuan (Scott) Chang

☎ 979-721-2046 ✉ cychang@tamu.edu

🌐 LinkedIn 🏠 Home Page

Research Interests

- **Large Language Models:** Extending context window of LLMs without fine-tuning [3]; RAG applications [1][6]; Linearizing and hybrid LLMs with model distillation.
- **Generative Models:** Developing a data generation solution for concept drift issue [2].
- **Domain Generalization:** Model-agnostic frameworks for domain generalization [7].
- **Fairness in Healthcare:** Fair machine learning algorithms for healthcare tasks [14].

Education

Texas A&M University

Ph.D. in Computer Science (Advisors: Dr. Na Zou and Dr. Xia Hu)

College Station, TX

Aug. 2021 – Expected Aug 2025

National Cheng Kung University

Master of Science in Structures and Materials

Tainan, Taiwan

Sep. 2013 – Jun. 2015

Publications

- [1] **C.Y. Chang**, Z. Jiang, et al., and N. Zou, “MAIN-RAG: Multi-Agent Filtering Retrieval-Augmented Generation,” The 63rd Annual Meeting of the Association for Computational Linguistics ([ACL’25 Main](#))
- [2] **C.Y. Chang**, Y.N. Chuang, Z. Jiang, K.H. Lai, A. Jiang, and N. Zou, “CODA: Temporal Domain Generalization via Concept Drift Simulator,” The 31st ACM SIGKDD International Conference on Knowledge Discovery & Data Mining ([KDD’25](#))
- [3] H. Jin, X. Han, J. Yang, Z. Jiang, Z. Liu, **C.Y. Chang**, H. Chen, and X. Hu, “LLM Maybe LongLM: Self-Extend LLM Context Window Without Tuning,” The Forty-first International Conference on Machine Learning ([ICML’24 Spotlight](#))
- [4] G. Wang, Y.N. Chuang, F. Yang, M. Du, **C.Y. Chang**, et al., and X. Hu, “TVE: Learning Meta-attribution for Transferable Vision Explainer,” The Forty-first International Conference on Machine Learning ([ICML’24](#))
- [5] Y.N. Chuang, S. Li, J. Yuan, G. Wang, K.H. Lai, L. Yu, S. Ding, **C.Y. Chang**, et al., and X. Hu, “Learning to Compress Prompt in Natural Language Formats,” 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics ([NAACL’24 Main](#))
- [6] G. Wang, J. Ran, R. Tang, **C.Y. Chang**, Y.N. Chuang, Z. Liu, V. Braverman, Z. Liu, and X. Hu, “Assessing and Enhancing Large Language Models in Rare Disease Question-answering,” arXiv ([JBI \(submitted\)](#))
- [7] **C.Y. Chang**, Y.N. Chuang, G. Wang, M. Du, and N. Zou, “DISPEL: Domain Generalization via Domain-Specific Liberating,” arXiv’23 ([ICDM’25 \(submitted\)](#))
- [8] Y.N. Chuang, G. Wang, **C.Y. Chang**, et al., and X. Hu, “Large Language Models As Faithful Explainers,” arXiv ([NeurIPS’25 \(submitted\)](#))
- [9] Y.N. Chuang, T. Xing, **C.Y. Chang**, et al., and X. Hu, “Understanding Different Design Choices in Training Large Time Series Models,” arXiv ([NeurIPS’24 \(submitted\)](#))
- [10] H. Jin, X. Han, J. Yang, Z. Jiang, **C.Y. Chang**, and X. Hu, “GrowLength: Accelerating LLMs Pretraining by Progressively Growing Training Length,” arXiv ([NeurIPS’24 \(submitted\)](#))
- [11] Y. Wang, X. Han, **C.Y. Chang**, D. Zha, U. Braga-Neto, and X. Hu, “Auto-PINN: Understanding and Optimizing Physics-Informed Neural Architecture,” Thirty-seventh Conference on Neural Information Processing Systems ([NeurIPS’23 AI4Science](#))

- [12] Y.N. Chuang, G. Wang, **C.Y. Chang**, et al., and X. Hu “DiscoverPath: A Knowledge Refinement and Retrieval System for Interdisciplinarity on Biomedical Research,” ACM International Conference on Information and Knowledge Management ([CIKM’23 Demo](#))
- [13] **C.Y. Chang**, Y.N. Chuang, K.H. Lai, X. Han, X. Hu, N. Zou, “Towards Assumption-free Bias Mitigation,” arXiv ([IJCAI’24 \(submitted\)](#))
- [14] **C.Y. Chang**, J. Yuan, S. Ding, Q. Tan, K. Zhang, X. Jiang, X. Hu, and N. Zou, “Towards Fair Patient-Trial Matching via Patient-Criterion Level Fairness Constraint,” AMIA 2023 Annual Symposium ([AMIA’23](#))
- [15] S. Ding, Q. Tan, **C.Y. Chang**, et al., and X. Hu, “Multi-Task Learning for Post-transplant Cause of Death Analysis: A Case Study on Liver Transplant,” AMIA 2023 Annual Symposium ([AMIA’23](#))
- [16] Y.N. Chuang, G. Wang, F. Yang, Z. Liu, **C.Y. Chang**, et al., and X. Hu, “Efficient XAI Techniques: A Taxonomic Survey,” arXiv ([Preprint](#))
- [17] Y.N. Chuang, K.H. Lai, R. Tang, M. Du, **C.Y. Chang**, N. Zou, X. Hu, “Mitigating Relational Bias on Knowledge Graphs,” ACM Transactions on Knowledge Discovery from Data ([TKDD](#))
- [18] **C.Y. Chang***, C.W. Lu*, and C.J. Wang, “A Multi-step-ahead Markov Conditional Forward Model with Cube Perturbations for Extreme Weather Forecasting,” The 36th Annual AAAI Conference ([AAAI’21](#))
- [19] **C.Y. Chang**, and et al., “Query Expansion with Semantic-based Ellipsis Reduction for Conversational IR,” The Twenty-Ninth Text REtrieval Conference ([TREC’20](#))

Experience

Amazon, Rufus

Palo Alto, CA

Applied Scientist

Jun. 2025 – Present

- RoPE-based long-context extension research to further extend the effective context window after pre-training
- Sparse attention mechanisms research, implementation, and evaluation
- Large-scale LLMs pre-training

Amazon, Rufus

Palo Alto, CA

Applied Scientist Intern

Nov. 2024 – Mar. 2025

- Research SOTA linearizing and hybrid LLMs
- Propose and implement a new hybrid LLM architecture

Visa Research

Foster City, CA

Research Intern

May. 2024 – Aug. 2024

- Develop a retrieval-augmented generation (RAG) framework to identify and rule out unrelated chunks without tuning
- Explore potential application scenarios of RAG for business models

Texas A&M University

College Station, TX

Graduate Research Assistant

Aug. 2021 – Present

- Developed model-agnostic algorithms that focus on generalization issues in machine learning models
- Developed fairness machine learning frameworks for healthcare tasks by task-specific regularizations
- Proposed the efficient pre-training paradigm for large language models during pre-processing

Academia Sinica

Taipei, Taiwan

Research Assistant

Oct. 2019 – Dec. 2020

- Developed a ranking algorithm for financial news recommendation systems to reduce 41% of traders’ daily reading time
- Researched and published papers in time-series forecasting and large language model applications
- Won the 2nd place award in an information retrieval competition via a T5-based coreference and reranking framework

EZTABLE

Taipei, Taiwan

Backend Engineer

Jan. 2019 – Sep. 2019

- Designed/developed APIs for the website and mobile app deployed on cloud services, including AWS and GCP
- Utilized CI/CD tools (Jenkins and Drone) for automatic testing and exporting reports
- Improved cloud infrastructures that reduced 19.2% server loading and 23.8% cost

Projects

Recommender System Algorithms | *E.SUN Bank*

Nov. 2019 – Dec. 2020

- Developed recommender system algorithms with the machine learning method
- Fine-tuned BERT pre-trained models for leveraging given limited labeled data
- Built and trained an RNN-based model to tackle a time-series problem
- Designed data pre-processing and post-processing pipelines to enhance the efficiency of experiments
- Led and organized the project members and communicated with E.SUN Bank partners

TripChat [\[Link\]](#) | *Side Project*

Sep. 2018 – Nov. 2018

- Developed front-end part with JavaScript library React
- Implemented RESTful APIs with Node.js and a real-time co-editing map with Socket.IO
- Implemented Cache for low updated frequency data with Redis
- Built the environment as a Docker Image and deployed on AWS EC2

Awards and Honors

CIKM 2023 Best Demo Paper Honorable Mention [12]

Oct. 2023

AMIA 2023 Best Student Paper Finalist [15]

Oct. 2023

NSF Travel Award for Quality and Productivity Research Conference (QPRC) 2023

Jun. 2023

2nd and 4th Place Award, TREC CAsT, Text REtrieval Conference [19] [\[Link\]](#)

Oct. 2020

National Chen Kung University Scholarship Award (Top 10%)

2014, 2015

Invited Talks

KDD 2023 Machine Learning in Finance Tutorial – *Algorithmic Fairness in Finance* [\[Link\]](#)

Aug. 2023

QPRC 2023 Short Course – *Fair Machine Learning in Healthcare* [\[Link\]](#)

Jun. 2023