Yu-Neng (Allen) CHUANG

(979) 334-0045 ◇ yc146@rice.edu
Website: https://ynchuang.github.io

EDUCATION

Rice University Houston, TX
Ph.D. in Computer Science (Advisor: Dr. Xia "Ben" Hu) *Aug. 2021 - Present*

National ChengChi University Taipei, Taiwan
Master of Science in Computer Science *Feb. 2018 - Jun. 2020*

National ChengChi University Taipei, Taiwan
Bachelor of Science in Mathematical Sciences *Aug. 2013 - Jul. 2017*

RESEARCH INTERESTS

• Uncertainty, Safety, and Explainability of Large Language Models (LLMs) • LLMs Reasoning and Alignment • Multi-Agent Planning • Explainable AI

EXPERIENCE

Rice University Houston, TX
Graduate Research Assistant *Aug. 2021 - Present*

- Developed finetuning and alignment algorithms on uncertainty, explainability, and safety issues of LLMs
- Built efficient framework for LLMs on prompt compression for lower inference throughput and budgets
- Developed efficient explainable frameworks for generating ML models and LLMs

Apple Inc. Cupertino, CA
Machine Learning Research Intern *May 2024 - Aug. 2024*

- Developed uncertainty quantification algorithm for on-device LLMs based on confidence token prediction for instance routing and rejection learning purposes

Samsung Research America Mountain View, CA
Research Intern *May 2023 - Aug. 2023*

- Developed an efficient algorithm of hard prompt compression on large language models with LLM post-training techniques, Proximal Policy Optimization (PPO), deducing 80% of LLM API usage cost and 20% of latency of white box LLMs

Living Analytics Research Centre., Carnegie Mellon University and SMU Singapore
Research Assistant *Jan. 2020 - Apr. 2020*

- Built a ranking method for a personalized job recommendation system with TB-scaled user data in Singapore, which outperformed other state-of-the-art ranking methods by 10.2%

KKBOX Co, Ltd. Taipei, Taiwan
Data Scientist Intern *Sep. 2019 - Jun. 2020*

- Developed a ranking algorithm on a streaming dataset of nearly 1.5 million users to enhance the recommendation system by 15.3% compared with the prior internal recommendation systems

SELECTED PUBLICATIONS

Publications

[C1] **Y.N. Chuang***, G. Wang*, R. Tang, S. Zhong, J. Yuan, H. Jin, Z. Liu, V. Chaudhary, S. Xu, J. Caverlee, and X. Hu. "Taylor Unswift: Secured Weight Release for Large Language Models via Taylor Expansion" Empirical Methods in Natural Language Processing (EMNLP'24)

[C2] J. Yuan*, H. Liu*, S. Zhong*, **Y.N. Chuang**, et. al, and X. Hu. "KV Cache Compression, But What Must We Give in Return? A Comprehensive Benchmark of Long Context Capable Approaches" Empirical Methods in Natural Language Processing (EMNLP'24 Finding)

[C3] G. Wang, **Y.N. Chuang**, F. Yang, M. Du, C.Y. Chang, S. Zhong, Z. Liu, K. Zhou, X. Cai, and X. Hu. "TVE: Learning Meta-attribution for Transferable Vision Explainer" International Conference on Machine Learning (ICML'24)

[C4] **Y.N. Chuang**, T. Xing, C.Y. Chang, Z. Liu, X. Chen, and X. Hu. "Learning to Compress Prompt in Natural Language Formats" Annual Conference of the North American Chapter of the ACL (NAACL'24)

[C5] **Y.N. Chuang***, R. Tang*, and X. Hu. "Secure Your Model: A Simple but Effective Key Prompt Protection Mechanism for Large Language Models" Annual Conference of the North American Chapter of the ACL (NAACL'24 Finding)

[C6] R. Tang, **Y.N. Chuang**, and X. Hu. "The Science of LLM-generated Text Detection" The Communications of the ACM (CACM April Cover), 2024

[C7] **Y.N. Chuang**, R. Tang, X. Jiang, and X. Hu. "SPeC: A Soft Prompt-Based Calibration on Performance Variability of Large Language Model in Clinical Notes Summarization" Journal of Biomedical Informatics. (JBI), 2024

[C8] **Y.N. Chuang***, G. Wang*, F. Yang, Q. Zhou, P. Tripathi, X. Cai and X. Hu. "CoRTX: Contrastive Learning for Real-time Explanations" International Conference on Learning Representations (ICLR'23)

[C9] **Y.N. Chuang***, G. Wang*, M. Du, F. Yang, Q. Zhou, P. Tripathi, X. Cai and X. Hu. "Accelerating Shapley Explanation via Contributive Cooperator Selection" International Conference on Machine Learning (ICML'22 Spotlight)

Preprints

[P1] **Y.N. Chuang**, H. Zhou, P. Sarma, P. Gopalan, J. Boccio, S. Bolouki, and X. Hu. "Learning to Route with Confidence Token" (Arxiv) (Baylearn Symposium Oral (Top 9%)), 2024

[P2] **Y.N. Chuang**, G. Wang, C.Y. Chang, R. Tang, S. Zhong, F. Yang, M. Du, X. Cai, and X. Hu "FaithLM: Towards Faithful Explanations for Large Language Models" (Arxiv), 2024

[P3] G. Wang, **Y.N. Chuang**, H. Chen, Y. Chen, Z. Jiang, M. Bendre, M. Das, Z. Liu, J. Yuan, and X. Hu. "LEMO: Learning Shapley Manifold for Faithful Explanation" (Arxiv), 2024

[P4] **Y.N. Chuang**, G. Wang, F. Yang, Z. Liu, X. Cai, M. Du, and X. Hu. "Efficient XAI Techniques: A Taxonomic Survey" (Arxiv), 2023

OPEN SOURCE PACKAGE

SMORe: Modularize Graph Embedding for Recommendation

- *Developer.* Constructed a large-scale network embedding library for recommendation systems on online streaming services which was developed under C++ with multi-thread processing techniques

LTSM-bundle: Large Time Series Models Training Library

- *Project Leader.* Designed the architecture of the training package, including the prompting design, tokenizer incorporation, and training paradigm.

DiscoverPath: A Knowledge Refinement and Retrieval System for Interdisciplinarity on Biomedical Research (CIKM'23 Best Demo Paper Honorable Mention)

- Designed a KG-based retrieval system designed for biomedical research that aims to assist biomedical researchers in dynamically refining their queries and effectively retrieving articles.
- *Project Leader.* Designed the architecture of the system, encompassing comprehensive full-stack web development, database configuration, and algorithmic design