Yu-Neng (Allen) CHUANG

(979) 334-0045 ◇ ynchuang@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 (NCCU) Taipei, Taiwan
Master of Science in Computer Science *Feb. 2018 - Jun. 2020*

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

RESEARCH INTERESTS AND SKILLS

- Machine learning: Efficiency and Trustworthy issues (Explainability, Uncertainty, and Safety) on Large Language Models (LLMs), Explainable Artificial Intelligence, AI in Healthcare
- Data Mining: Recommender Systems, Graph Neural Network, Network Embedding

RESEARCH EXPERIENCE

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

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

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

- Developed an efficient algorithm of hard prompt compression on large language models, deducing the 80% of LLM API usage cost and up to 4 times lower 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 1 million target users in Singapore, which outperformed other state-of-the-art ranking methods by 10.2%
- Developed an open-source package for textual-based recommendation systems to better exploit TB-scaled textual information with TB-scaled users' interaction information

Institute of Information Science., Academia Sinica Taipei, Taiwan
Research Assistant *Dec. 2020 - Jun. 2021*

- Developed privacy-aware recommendation systems to protect users' private information from being attacked through the API callback
- Developed a two-tower sequential-based model on time series data for high-contributed customer prediction and loan prediction, which achieves 530% of improvement in online testing

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

- Developed a recommendation algorithm that utilized the TB-scale user feedback dataset, enhancing the 15.3% of the performance compared with the prior internal ticket and video recommendation systems
- Investigated the data distribution of the graph embedding space to improve the recommendation performance on the TB-scaled music streaming dataset, yielding up to 17.7% improvement in offline testing

PUBLICATIONS

Conference and Journal Papers

[C1] **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)*

[C2] R. Tang, **Y.N. Chuang**, and X. Hu. "The Science of LLM-generated Text Detection" *The Communications of the ACM (CACM, 2024)*

[C3] **Y.N. Chuang**, G. Wang 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 Best Demo Paper Honorable Mention)*

[C4] **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)*

[C5] **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)*

[C6] **Y.N. Chuang***, C.M. Chen*, C.J. Wang, M.F. Tsai, Y. Fang, and E.P. Lim. "TPR: Text-aware Preference Ranking for Recommender Systems" *ACM International Conference on Information and Knowledge Management (CIKM'20 Oral)*

[C7] **Y.N. Chuang***, C.J. Wang*, C.M. Chen, and M.F. Tsai. "Skewness Ranking Optimization for Personalized Recommendation" *Conference on Uncertainty in Artificial Intelligence (UAI'20 Oral)*

[C8] S.C. Lin, **Y.N. Chuang**, S.F. Yang, M.F. Tsai, and C.J. Wang*. "Negative-aware Collaborative filtering" *ACM Conference on Recommender Systems (RecSys'19)*

[C9] **Y.N. Chuang**, Z.Y. Huang, and Y.L. Tsai. "Variational Grid Setting Network" *International Conference on Asian Language Processing (IALP'17)*

Preprints and Under Review

[P1] **Y.N. Chuang**, G. Wang, C.Y. Chang, Mengnan Du, R. Tang, X. Cai, F. Yang, and X. Hu. "Large Language Models As Faithful Explainer" (*Arxiv 2024*)

[P2] **Y.N. Chuang**, T. Xing, C.Y. Chang, Z. Liu, X. Chen, and X. Hu. "Learning to Compress Prompt in Natural Language Formats" (*Arxiv 2023*)

[P3] **Y.N. Chuang***, R. Tang*, and X. Hu. "Secure Your Model: A Simple but Effective Key Prompt Protection Mechanism for Large Language Models" (*Arxiv 2023*)

[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*)

[P5] **Y.N. Chuang**, K.H. Lai, R. Tang, M. Du, C.Y. Chang, N. Zou, and X. Hu. "Mitigating Relational Bias on Knowledge Graphs" (*Arxiv 2022*)

[P6] **Y.N. Chuang**, Cheng-Te Li. "Privacy-Preserving Representation Learning with Gradient Obfuscation against Attribute Inference for Recommendation" (*Arxiv 2021*)

[P7] G. Wang, **Y.N. Chuang**, F. Yang, et al., and X. Hu. "LETA: Learning Transferable Attribution for Generic Vision Explainer" (*Arxiv 2023*)

[P8] C.Y. Chang, **Y.N. Chuang**, Z. Jiang, K.H. Lai, A. Jiang, N. Zou. "CODA: Temporal Domain Generalization via Concept Drift Simulator" (*Arxiv 2023*)

[P9] C.Y. Chang, **Y.N. Chuang**, G. Wang, M. Du, and N. Zou. "DISPEL: Domain Generalization via Domain-Specific Liberating" (*Arxiv 2023*)

[P10] C.Y. Chang, **Y.N. Chuang**, K.H. Lai, X. Han, X. Hu, N. Zou. "Towards Assumption-free Bias Mitigation" (*Arxiv 2023*)

OPEN SOURCE PACKAGE

DiscoverPath: A Knowledge Refinement and Retrieval System for Interdisciplinarity on Biomedical Research

- Designed a KG-based retrieval system designed for biomedical research 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.

SMORe: Modularize Graph Embedding for Recommendation

- Developed real-time online streaming algorithm for online music streaming services in Spotify Inc. and KKBOX Inc.
- *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

TextRec: General Recommendation System with Textual Information

- A package of general ranking algorithms on learning textual information with user-item interaction utilized in the Healthcare AI systems of ASUS Inc.
- *Project Leader.* Designed the architecture of the package and implemented API based on the package for better accessing recommendation results

HONORS AND AWARDS

- Best Student Paper Award Finalist, Texas Medical Center AI Summit *Feb. 2024*
- Best Demo Paper Award Honorable Mention, CIKM *Oct. 2023*
- 4th Place, ACM RecSys Challenge *Sep. 2020*
- Dean's List Award, NCCU *Aug. 2020*
- 1st Place, TREC CAsT Competition *Aug. 2019*

PROFESSIONAL SERVICES

Reviewer (Since 2020): KDD, NeurIPS, ICLR, ICML, IJCAI, CIKM, IEEE TAI, IEEE TIST, IEEE ICHI

Yu-Neng (Allen) Chuang Rice University

Research Statement

Overview My research is centered on the fields of explainability, safety, and uncertainty issues in artificial intelligence and machine learning, especially on large language models (LLMs). Despite the impressive accomplishments in existing machine learning (ML) techniques, current ML research lacks substantial resources to sustain its trustworthiness. Consequently, the direct application of these ML advancements in high-stack settings becomes challenging, primarily because of the concerns related to uncertain performance with black-box decision-making and rigid regulatory requirements, which unleash the importance of enhancing the trustworthiness of ML. In this manner, my research aims to meet industrial and high-stack requirements through academic solutions, unlocking their potential for wider applications and fostering substantial societal impact.

Research Achievements In the domain of trustworthy machine learning, I have contributed through peer-reviewed papers at the top machine learning conferences and journals, including ICML, ICLR, CACM, UAI, CIKM, JBI, etc. I have been honored with the CIKM Best Demo Paper Award Honorable Mention and recognized with a Spotlight Paper at ICML 2022. From an industrial perspective, my contributions have significantly applied to the development of the Meta Ads Transparency Tool and have been integrated into Spotify for its online music recommendation services.

Previous Work Over the past few years, I envision improving ML trustworthiness under the settings of high-stack and societal applications, making them more accessible and efficient. To alleviate the uncertainty, I have developed methods to enhance machine learning applications, including (1) enhancing the reliability of manually designed hard prompts in pre-trained LLMs for clinical note summarization: I developed the SPeC [1] framework to mitigate performance uncertainty brought by different prompts, ensuring more robustness outputs and reducing the presence of incorrect information from LLMs; and (2) introducing a user-friendly system designed to facilitate the verification of concepts within biomedical prompts for LLMs: I proposed DiscoverPath [2] to construct the biomedical knowledge graphs, which can effectively assist experts in refining biomedical prompts that directly used in LLMs with less misinformation. To improve the transparency, I have paved the way for increasing accessibility of explainable methods on real-time ML systems, such as (1) accelerating the estimation of Shapley Values: I proposed CoRTX [3] to leverage semi-supervised learning with limited labels for deriving real-time explanations alongside the predictions of real-time ML systems, such as autopilot system, significantly reducing computational costs and time; (2) uncovering the decision-making process of black-box LLMs: I developed xLLM [4] framework to provide the faithful natural language explanations that truthfully reflect the decision-making behavior of black-box LLMs; (3) reducing computational complexity of Shapley Values: I proposed SHEAR [5] to deduce the complexity in Shapley Values by observing feature interaction regarding the input of ML systems. Besides, I also work on fairness in machine learning [6, 7].

Future Plan My goal is to expand, improve, and promote the adoption of trustworthy machine learning methodologies, particularly in healthcare and large language models. In the healthcare domain, I will focus on enhancing the transparency of generative AI, which greatly assists healthcare experts in trusting and judging AI predictions before they take any medical actions. As for LLMs, I will explore efficient LLM inference and LLM safety, making LLMs more accessible and safe while applying them to real-world and high-stack applications, such as precision medicine and drug discovery.

References

[1] 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*, p. 104606.

[2] Y.-N. Chuang, G. Wang, C.-Y. Chang, K.-H. Lai, D. Zha, R. Tang, F. Yang, A. C. Reyes, K. Zhou, X. Jiang, *et al.*, "Discoverpath: A knowledge refinement and retrieval system for interdisciplinarity on biomedical research," *Proceedings of the 32th ACM International Conference on Information & Knowledge Management*, 2023.

[3] Y.-N. Chuang*, G. Wang*, F. Yang, Q. Zhou, P. Tripathi, X. Cai, and X. Hu, "Cortx: Contrastive framework for real-time explanation," in *The Eleventh International Conference on Learning Representations*, 2023.

[4] Y.-N. Chuang, G. Wang, C.-Y. Chang, R. Tang, F. Yang, M. Du, X. Cai, and X. Hu, "Large language models as faithful explainers," *arXiv preprint arXiv:2402.04678*, 2024.

[5] 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," in *International Conference on Machine Learning*, pp. 22576–22590, PMLR, 2022.

[6] C.-Y. Chang, Y.-N. Chuang, K.-H. Lai, X. Han, X. Hu, and N. Zou, "Towards assumption-free bias mitigation," *arXiv preprint arXiv:2307.04105*, 2023.

[7] Y.-N. Chuang, K.-H. Lai, R. Tang, M. Du, C.-Y. Chang, N. Zou, and X. Hu, "Mitigating relational bias on knowledge graphs," *arXiv preprint arXiv:2211.14489*, 2022.

Unofficial Academic Transcript

S01401301 Yu-Neng Chuang
Feb 28, 2024 02:53 pm

This is not an official transcript.

Note: Esther displays GPA to the second decimal place. The Office of the Registrar will continue to use the GPA to the third decimal place when determining Academic Standing such as the President's Honor Roll, probation, and suspension.

Unofficial Academic Transcript - Rice University

Institution Credit Transcript Totals Courses in Progress

Transcript Data

STUDENT INFORMATION

Birth Date: Nov 30, 1994

Curriculum Information

Current Program

Major and Department: Computer Science, Computer Science

***Transcript type:Official is NOT Official ***

AWARDED:

Sought: Doctor of Philosophy **Degree Date:**

Curriculum Information

Major: Computer Science

INSTITUTION CREDIT -Top-

Term: Fall Semester 2021

Academic Standing:

Subject	Course	Level	Title	Grade	Credit Hours	Quality Points	R
COMP	536	GR	SECURE & CLOUD COMPUTING	B	3.000	9.00	
COMP	600	GR	GRADUATE SEMINAR	S	1.000	0.00	I
COMP	800	GR	GRADUATE RESEARCH	S	5.000	0.00	I

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA

Current Term:	9.000	9.000	9.000	3.000	9.00	3.00
Cumulative:	9.000	9.000	9.000	3.000	9.00	3.00

Term: Spring Semester 2022

Academic Standing:

Subject	Course	Level	Title	Grade	Credit Hours	Quality Points	R
COMP	518	GR	IOT PROGRAMMING DATA ANALYSIS	A	4.000	16.00	
COMP	590	GR	COMPUTER SCIENCE PROJECTS	A+	4.000	16.00	
COMP	600	GR	GRADUATE SEMINAR	S	1.000	0.00	I
COMP	800	GR	GRADUATE RESEARCH	S	4.000	0.00	I

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	13.000	13.000	13.000	8.000	32.00	4.00
Cumulative:	22.000	22.000	22.000	11.000	41.00	3.72

Term: Summer Semester 2022

Academic Standing:

Subject	Course	Level	Title	Grade	Credit Hours	Quality Points	R
COMP	800	GR	GRADUATE RESEARCH	S	10.000	0.00	I

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	10.000	10.000	10.000	0.000	0.00	0.00
Cumulative:	32.000	32.000	32.000	11.000	41.00	3.72

Term: Fall Semester 2022

Academic Standing:

Subject	Course	Level	Title	Grade	Credit Hours	Quality Points	R
COMM	601	GR	ORAL COMMUNICATION SKILLS	S	2.000	0.00	
COMP	600	GR	GRADUATE SEMINAR	S	1.000	0.00	I
COMP	800	GR	GRADUATE RESEARCH	S	9.000	0.00	I

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	12.000	12.000	12.000	0.000	0.00	0.00
Cumulative:	44.000	44.000	44.000	11.000	41.00	3.72

Term: Spring Semester 2023

Academic Standing:

Subject	Course	Level	Title	Grade	Credit Hours	Quality Points	R
COMP	600	GR	GRADUATE SEMINAR	S	1.000	0.00	I
COMP	800	GR	GRADUATE RESEARCH	S	8.000	0.00	I

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	9.000	9.000	9.000	0.000	0.00	0.00
Cumulative:	53.000	53.000	53.000	11.000	41.00	3.72

Unofficial Transcript

Term: Summer Semester 2023

Academic Standing:

Subject	Course	Level	Title	Grade	Credit Hours	Quality Points	R
COMP	800	GR	GRADUATE RESEARCH	S	9.000	0.00	I

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	9.000	9.000	9.000	0.000	0.00	0.00
Cumulative:	62.000	62.000	62.000	11.000	41.00	3.72

Unofficial Transcript

Term: Fall Semester 2023

Academic Standing:

Subject	Course	Level	Title	Grade	Credit Hours	Quality Points	R
COMP	600	GR	GRADUATE SEMINAR	S	1.000	0.00	I
COMP	800	GR	GRADUATE RESEARCH	S	8.000	0.00	I

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Current Term:	9.000	9.000	9.000	0.000	0.00	0.00
Cumulative:	71.000	71.000	71.000	11.000	41.00	3.72

Unofficial Transcript

TRANSCRIPT TOTALS (GRADUATE) -Top-

	Attempt Hours	Passed Hours	Earned Hours	GPA Hours	Quality Points	GPA
Total Institution:	71.000	71.000	71.000	11.000	41.00	3.72
Total Transfer:	0.000	0.000	0.000	0.000	0.00	0.00
Overall:	71.000	71.000	71.000	11.000	41.00	3.72

Unofficial Transcript

COURSES IN PROGRESS -Top-

Term: Spring Semester 2024

Subject	Course	Level	Title	Credit Hours
COMP	600	GR	GRADUATE SEMINAR	1.000
COMP	800	GR	GRADUATE RESEARCH	8.000

Unofficial Transcript

Esther Privacy

RELEASE: 8.7.1