
Member of Technical Staff at [Arena Intelligence](#)

EDUCATION

University of Southern California, Los Angeles, CA

Aug. 2018 – May. 2024

Ph.D. in Computer Science

Advisor: Premkumar Natarajan & Nanyun Peng

Thesis: Towards Generalized Event Understanding in Text via Generative Models

National Taiwan University, Taipei, Taiwan

Sep. 2013 – Dec. 2017

B.S. in Electrical Engineering

WORK EXPERIENCE

Arena Intelligence, San Francisco, CA

Jan 2026 – Present

Member of Technical Staff

- Led end-to-end research for [Code Arena AutoEval](#), spanning data curation, training infrastructure, and reward-model development to provide early model ratings aligned with human preferences on coding tasks.
- Led Code Arena data analysis and developed a [seven-category web development taxonomy](#) from 250,000+ prompts, enabling category-specific leaderboards.
- Develop a checking and validation system for complex RL environments used in model post-training.
- Supported research engagements with frontier AI labs, helping address model evaluation needs through Code Arena.
- Provided technical mentorship and debugging for text-to-image reward modeling and post-training ([C26](#), [P2](#)).

Cloud AI Research, Google, Sunnyvale CA

Aug. 2024 – Jan 2026

Research Scientist

– Post-training for LLM Coding Agent Reasoning:

- * *Gemma Contribution*: Developed an RL post-training framework that accelerates best end-to-end model searching by over 100×, boosting SWE-Bench performance by relatively 30.4% (Gemma) and 54% (OSS prototyping)([P1](#)).
- * *Research*: Pioneered Supervised Reinforcement Learning, a novel method to enhance multi-step reasoning in math and coding. The framework rewards logical actions over simple token mimic (ICLR'26 [[C25](#)]).

– Development and Application of Agentic AI:

- * *Product Impact*: Architected and led the development of the first agent-based framework for Google's Document Understanding product, reducing document-level errors for enterprise customers by over 17%.
- * *Research*: Authored and published foundational research advancing core AI agent capabilities in long-term memory (ICLR'26 [[C24](#)] & ACL'25 [[C19](#)]) and strategic tool utilization (CoLM'26 [[C26](#)] & ACL'25 [[C20](#)]).

– Data Synthesis for LLM Factuality:

- * Transforming low-quality model responses into high-quality training examples for supervised fine-tuning increases model helpfulness by a 7% absolute margin without compromising factual accuracy.
- * Training the *reward model* on a pairwise dataset from our data transformation pipeline improved its factuality benchmark performance by a relative 25%.

Cloud AI Research, Google, Remote / Sunnyvale CA

Sep. 2023 – Mar. 2024

Research Intern; Advised by: Chen-Yu Lee & Chun-Liang Li

- Unsupervised verification algorithm to enhance the grounded generation of LLMs (ACL'24 [[C18](#)])
- Conducted a large-scale empirical study on the behaviors of retrieval-augmented generation (RAG) across various LLMs, including Llama, Vicuna, and Yi.

Microsoft Research, Redmond, WA

May. 2023 – Aug. 2023

Research Intern; Advised by: Silviu Cucerzan

- Developed novel prompting and evaluation methods to assess the ability of LLMs to answer questions using emerging factual knowledge not present in their training data.

Alexa Conversations, Amazon, Sunnyvale, CA

May. 2022 – Aug. 2022

Applied Scientist Intern; Advised by: Avik Ray & Jing Huang

- Developed a data synthesis pipeline to generate *code-switched* text, significantly expanding coverage across multiple language pairs. (ACL'23 [C9])

Alexa Conversations, Amazon, Remote

May. 2020 – Aug. 2020

Applied Scientist Intern; Advised by: Abhay Jha

- Learning from negative labels for improving intent prediction and slot filling in conversational agents.

SELECTED RESEARCH FOCUS

LLM: Post-training ([C25, C27][P1, P2]), Factuality ([C18]), Tool usage ([C20, C26]), Memory ([C19, C24])

Information Extraction: Generative Model ([C6, C7, C8, C12]), Data Synthesis ([C15, C17, C21])

Multilingualism: Data Synthesis ([C9]), Applications ([C6, C15, C21]), Data Efficiency ([C23])

AWARDS

Area Chair Award, ACL 2023 [Link]

2023

- Outstanding paper in *Semantics Track*

Best Paper Award, DLG-AAAI'22 Workshop [Link]

2022

Amazon Alexa Graduate Fellowship [Link]

2018 - 2019

PREPRINTS

- P2.** Yuanhao Ban, I-Hung Hsu, Anastasios Angelopoulos, Wei-Lin Chiang, Ion Stoica, and Cho-Jui Hsieh. Post-Training Frontier Text-to-Image Models by Composing Preference and Rubric Rewards. *arXiv preprint arXiv:2610.02967*, 2026.
- P1.** Gemma team. Gemma 4 Technical Report. *arXiv preprint arXiv:2607.02770*, 2026.

PUBLICATIONS

- C27.** Yuanhao Ban, Tong Xie, Sohyun An, Yunqi Hong, Evan Frick, I-Hung Hsu, Wei-Lin Chiang, Ion Stoica, and Cho-Jui Hsieh. Arena-T2I Hard: Benchmarking and Improving Faithfulness with Dependency-Aware Checklist. In *The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2026. (Spotlight).
- C26.** Tengxiao Liu, Zifeng Wang, Jin Miao, I-Hung Hsu, Jun Yan, Jiefeng Chen, Rujun Han, Fangyuan Xu, Yanfei Chen, Ke Jiang, Samira Daruki, Yi Liang, William Yang Wang, Tomas Pfister, and Chen-Yu Lee. Budget-Aware Tool Use Enables Effective Agent Scaling. In *The Third Annual Conference on Language Modeling (CoLM)*, 2026.
- C25.** Yihe Deng*, I-Hung Hsu*, Jun Yan, Zifeng Wang, Rujun Han, Gufeng Zhang, Yanfei Chen, Wei Wang, Tomas Pfister, and Chen-Yu Lee. Supervised Reinforcement Learning: From Expert Trajectories to Step-wise Reasoning. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026. (*equal contribution).
- C24.** Siru Ouyang, Jun Yan, I-Hung Hsu, Yanfei Chen, Ke Jiang, Zifeng Wang, Rujun Han, Long T. Le, Samira Daruki, Xiangru Tang, Vishy Tirumalashetty, George Lee, Mahsan Rofouei, Hangfei Lin, Jiawei Han, Chen-Yu Lee, and Tomas Pfister. ReasoningBank: Scaling Agent Self-Evolving with Reasoning Memory. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.
- C23.** Shayne Longpre, Sneha Kudugunta, Niklas Muennighoff, I-Hung Hsu, Isaac Rayburn Caswell, Alex Pentland, Serkan Ö. Arık, Chen-Yu Lee, and Sayna Ebrahimi. ATLAS: Adaptive Transfer Scaling Laws for Multilingual Pre-training, Finetuning, and Decoding the Curse of Multilinguality. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.
- C22.** Fangyuan Xu, Rujun Han, Yanfei Chen, Wang Zifeng, I-Hung Hsu, Jun Yan, Vishy Tirumalashetty, Eunsol Choi, and Chen-Yu Pfister, Tomas Lee. SAGE: Steerable Agentic Data Generation for Deep Search with Execution Feedback. In *Findings of the Association for Computational Linguistics: EACL (EACL-Findings)*, 2026.
- C21.** Tanmay Parekh, Yuxuan Dong, Lucas Bandarkar, Artin Kim, I-Hung Hsu, Kai-Wei Chang, and Nanyun Peng. SNaRe: Domain-aware Data Generation for Low-Resource Event Detection. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

- C20. Fan Yin, Zifeng Wang, I-Hung Hsu, Jun Yan, Ke Jiang, Yanfei Chen, Jindong Gu, Long Le, Kai-Wei Chang, Chen-Yu Lee, Hamid Palangi, and Tomas Pfister. Magnet: Multi-turn Tool-use Data Synthesis and Distillation via Graph Translation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.
- C19. Zhen Tan, Jun Yan, I-Hung Hsu, Rujun Han, Zifeng Wang, Long Le, Yiwen Song, Yanfei Chen, Hamid Palangi, George Lee, Anand Rajan Iyer, Tianlong Chen, Huan Liu, Chen-Yu Lee, and Tomas Pfister. In Prospect and Retrospect: Reflective Memory Management for Long-term Personalized Dialogue Agents. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.
- C18. I-Hung Hsu, Zifeng Wang, Long Le, Lesly Miculicich, Nanyun Peng, Chen-Yu Lee, and Tomas Pfister. CaLM: Contrasting Large and Small Language Models to Verify Grounded Generation. In *Findings of the Association for Computational Linguistics: ACL (ACL-Findings)*, 2024.
- C17. I-Hung Hsu^{*}, Zihan Xue^{*}, Nilay Pochhi, Sahil Bansal, Prem Natarajan, Jayanth Srinivasa, and Nanyun Peng. Argument-Aware Approach To Event Linking. In *Findings of the Association for Computational Linguistics: ACL (ACL-Findings)*, 2024. (*equal contribution).
- C16. Kuan-Hao Huang, I-Hung Hsu, Tanmay Parekh, Zhiyu Xie, Zixuan Zhang, Premkumar Natarajan, Kai-Wei Chang, Nanyun Peng, and Heng Ji. TextEE: Benchmark, Reevaluation, Reflections, and Future Challenges in Event Extraction. In *Findings of the Association for Computational Linguistics: ACL (ACL-Findings)*, 2024.
- C15. Tanmay Parekh, I-Hung Hsu, Kuan-Hao Huang, Kai-Wei Chang, and Nanyun Peng. Contextual Label Projection for Cross-Lingual Structure Prediction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2024. (Best Paper Nomination).
- C14. Kuan-Hao Huang, Varun Iyer, I-Hung Hsu, Anoop Kumar, Kai-Wei Chang, and Aram Galstyan. ParaAMR: A Large-Scale Syntactically Diverse Paraphrase Dataset by AMR Back-Translation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023. (Area Chair Award).
- C13. I-Hung Hsu^{*}, Kuan-Hao Huang^{*}, Shuning Zhang, Wenxin Cheng, Premkumar Natarajan, Kai-Wei Chang, and Nanyun Peng. TAGPRIME: A Unified Framework for Relational Structure Extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023. (*equal contribution).
- C12. I-Hung Hsu^{*}, Zhiyu Xie^{*}, Kuan-Hao Huang, Premkumar Natarajan, and Nanyun Peng. AMPERE: AMR-Aware Prefix for Generation-Based Event Argument Extraction Model. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023. (*equal contribution).
- C11. Tanmay Parekh, I-Hung Hsu, Kuan-Hao Huang, Kai-Wei Chang, and Nanyun Peng. GENEVA: Benchmarking Generalizability for Event Argument Extraction with Hundreds of Event Types and Argument Roles. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.
- C10. Keming Lu, I-Hung Hsu, Wenxuan Zhou, Mingyu Derek Ma, and Muhao Chen. Multi-hop Evidence Retrieval for Cross-document Relation Extraction. In *Findings of the Association for Computational Linguistics: ACL (ACL-Findings)*, 2023.
- C9. I-Hung Hsu, Avik Ray, Shubham Garg, Nanyun Peng, and Jing Huang. Code-Switched Text Synthesis in Unseen Language Pairs. In *Findings of the Association for Computational Linguistics: ACL (ACL-Findings)*, 2023.
- C8. Keming Lu, I-Hung Hsu, Wenxuan Zhou, Mingyu Derek Ma, and Muhao Chen. Summarization as Indirect Supervision for Relation Extraction. In *Findings of the Association for Computational Linguistics: EMNLP (EMNLP-Findings)*, 2022.
- C7. I-Hung Hsu^{*}, Kuan-Hao Huang^{*}, Elizabeth Boschee, Scott Miller, Premkumar Natarajan, Kai-Wei Chang, and Nanyun Peng. DEGREE: A Data-Efficient Generation-Based Event Extraction Model. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2022. (*equal contribution).
- C6. Kuan-Hao Huang^{*}, I-Hung Hsu^{*}, Premkumar Natarajan, Kai-Wei Chang, and Nanyun Peng. Multilingual Generative Language Models for Zero-Shot Cross-Lingual Event Argument Extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2022. (*equal contribution).
- C5. I-Hung Hsu, Xiao Guo, Premkumar Natarajan, and Nanyun Peng. Discourse-level Relation Extraction via Graph Pooling. In *The Thirty-Sixth AAAI Conference on Artificial Intelligence Workshop on Deep Learning on Graphs: Methods and Applications (DLG-AAAI'22)*, 2022. (Best Paper Award).

- C4. Rujun Han, I-Hung Hsu, Jiao Sun, Julia Baylon, Qiang Ning, Dan Roth, and Nanyun Peng. ESTER: A Machine Reading Comprehension Dataset for Reasoning about Event Semantic Relations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
- C3. Rujun Han*, I-Hung Hsu*, Mu Yang, Aram Galstyan, Ralph Weischedel, and Nanyun Peng. Deep Structured Neural Network for Event Temporal Relation Extraction. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, 2019. (*equal contribution).
- C2. I-Hung Hsu, Ayush Jaswal, and Premkumar Natarajan. NIESR: Nuisance Invariant End-to-end Speech Recognition. In *Interspeech 2019*, 2019.
- C1. Pin-Jung Chen*, I-Hung Hsu*, Yi-Yao Huang*, and Hung-Yi Lee. Mitigating the impact of speech recognition errors on chatbot using sequence-to-sequence model. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2017. (*equal contribution).

PROFESSIONAL SERVICES

Conference Area Chair

ACL Rolling Review (ARR)	2024 – 2026
--------------------------	-------------

Conference Program Committee / Reviewer

ACL Rolling Review (ARR)	2021 – 2023
Meeting of the Association for Computational Linguistics (ACL)	2021 – 2023
Conference on Empirical Methods in Natural Language Processing (EMNLP)	2021 – 2023
Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)	2022 – 2024
Conference of the European Chapter of the Association for Computational Linguistics (EACL)	2023 – 2024
Conference on Language Modeling (COLM)	2024 – 2025
Conference on International Conference on Machine Learning (ICML)	2026
Conference on Neural Information Processing Systems (NeurIPS)	2024 – 2026
International Conference on Learning Representations (ICLR)	2024 – 2026
AAAI Conference on Artificial Intelligence (AAAI)	2021 – 2025
Artificial Intelligence and Statistics (AISTATS)	2024
CCF International Conference on Natural Language Processing and Chinese Computing (NLPCC)	2022 – 2023

Journal Reviewer

IEEE Transactions on Knowledge and Data Engineering (TKDE)	2024
--	------

SKILLS

Programming: Python, PyTorch, Keras, Tensorflow, JAX, MySQL, Django, Docker, Git, $\LaTeX$, C++, Matlab

Languages: Mandarin (native), English (fluent), Taiwanese (fluent)