

Chia-Hsuan (Michael) Lee

+1 669-204-2924 • chiahsuan.li@gmail.com
chiahsuanlee chia-hsuan-lee.github.io

Work Experiences

- **Capital One** **New York, U.S.**
2024 – Present
Applied Researcher,
Working on language model alignment and conversational agents at AI foundation team.

Education

- **University of Washington** **Seattle, U.S.**
2019 – 2024
PhD, Electrical and Computer Engineering Department,
(Natural Language Processing Group) Advisor: Mari Ostendorf
- **National Taiwan University** **Taipei, Taiwan**
2017 – 2019
Master, Computer Science,
(Language Processing and Machine Learning Group) Advisors: Lin-shan Lee and Hung-Yi Lee
- **National Taiwan University** **Taipei, Taiwan**
2012 – 2017
B.S., Electrical Engineering,

Research Experiences

- **Research Assistant, UW NLP Research Group** **Seattle, U.S.**
Sep 2019 - Aug 2024
Advisor: Mari Ostendorf,
 - Proposed an **SLM/LLM dynamic routing framework driven by a retrieval-based router** that is finetuned to effectively leverage LM exemplars. The routing results outperform LLM systems and achieve 50% computational efficiency.
 - Created a new model for **LMs to perform structured predictions on task-oriented dialogues with schema-driven prompting** and achieved state-of-the-art results on a competitive dialogue state tracking task.
 - Proposed an **in-context learning framework for dialogue state tracking with an intent-aligned retriever** and achieved state-of-the-art few-shot results on a competitive dialogue state tracking task.
 - Built a **human-in-the-loop framework for generating long and natural conversations by leveraging instruction-following language model**. Showed effectiveness in improving state tracking on information-gathering dialogues.
- **Student Researcher, Google Brain Research** **New York, U.S.**
Jun-Sep 2022
Advisors: Ankur Bapna, Yu Zhang, Tara Sainath,
Built a family of **multimodal speech-text long-context language models with quantized representation learning** and demonstrated improvements on long-form ASR and spoken question answering tasks.
- **Research Intern, Google Machine Translation Research** **Remote Position**
Jun-Sep 2021
Advisor: Melvin Johnson,
Proposed a new **pretraining objective for multilingual language models to comprehend and generate long documents**. Achieved state-of-the-art results in document translation tasks and analyzed effects of data quality and pretraining objectives for multilingual language models.
- **Research Intern, Microsoft Research NLP Group** **Remote Position**
Jun-Sep 2020
Advisor: Matthew Richardson, Alex Polozov,
Constructed a new cross-domain text-to-SQL dataset of real Web databases, along with realistic evaluation schemes. Demonstrated **incorporating database documentation benefits understanding structural in-domain knowledge** and improves parser performance.
- **Research Intern, Apple Siri Team** **Cupertino, CA.**
Jul-Sep 2019
Advisor: Jerome Bellegarda,
Worked on pretraining of multilingual language models.

Publications

- [1] **OrchestraLLM: Efficient Orchestration of Language Models for Dialogue State Tracking**
Chia-Hsuan Lee, Hao Cheng, Mari Ostendorf. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2024*
- [2] **Does Collaborative Human-LM Dialogue Generation Help Information Extraction from Human Dialogues?**

Bo-Ru Lu*, Nikita Haduong*, **Chia-Hsuan Lee**, Zeqiu Wu, Hao Cheng, Paul Koester, Jean Utke, Tao Yu, Noah A. Smith, Mari Ostendorf. *Conference on Language Modeling (COLM)*, 2024.

- [3] **In-Context Learning for Few-Shot Dialogue State Tracking**
Yushi Hu, **Chia-Hsuan Lee**, Tianbao Xie, Tao Yu, Noah A. Smith, Mari Ostendorf. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2022
- [4] **DOCmT5: Document-Level Pretraining of Multilingual Language Models**
Chia-Hsuan Lee, Aditya Siddhant, Viresh Ratnakar, Melvin Johnson. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2022
- [5] **MIA 2022 Shared Task: Evaluating Cross-lingual Open-Retrieval Question Answering for 16 Diverse Languages**
Akari Asai, Shayne Longpre, Jungo Kasai, **Chia-Hsuan Lee**, Rui Zhang, Junjie Hu, Ikuya Yamada, Jonathan H. Clark, and Eunsol Choi. *In Proceedings of the NAACL Workshop on Multilingual Information Access (MIA)*, 2022
- [6] **Dialogue State Tracking with a Language Model using Schema-Driven Prompting**
Chia-Hsuan Lee, Hao Cheng, Mari Ostendorf. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021
- [7] **KaggleDBQA: Realistic Evaluation of Text-to-SQL Parsers**
Chia-Hsuan Lee, Oleksandr Polozov, Matthew Richardson. *Annual Conference of the Association for Computational Linguistics (ACL)*, 2021.
- [8] **Machine Comprehension of Spoken Content: TOEFL Listening Test and Spoken SQuAD**
Chia-Hsuan Lee, Hung-yi Lee, Szu-Lin Wu, Chi-Liang Liu, Wei Fang, Juei-Yang Hsu, Bo-Hsiang Tseng. *IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP)* 2019
- [9] **Mitigating the Impact of Speech Recognition Errors on Spoken Question Answering by Adversarial Domain Adaptation**
Chia-Hsuan Lee, Yun-Nung Chen, Hung-Yi Lee. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2019
- [10] **ODSQA: Open-Domain Spoken Question Answering Dataset**
Chia-Hsuan Lee, Shang-Ming Wang, Huan-Cheng Chang, Hung-Yi Lee. *IEEE Spoken Language Technology Workshop (SLT)* 2018
- [11] **Spoken SQuAD: A Study of Mitigating the Impact of Speech Recognition Errors on Listening Comprehension**
Chia-Hsuan Lee, Szu-Lin Wu, Chi-Liang Liu, Hung-yi Lee. *Annual Conference of the International Speech Communication Association (Interspeech)* 2018

Preprints

- [1] **CorrectionLM: Self-Corrections with SLM for Dialogue State Tracking**
Chia-Hsuan Lee, Hao Cheng, Mari Ostendorf. *arxiv*, 2024.
- [2] **Cross-Lingual Transfer Learning for Question Answering**
Chia-Hsuan Lee, Hung-Yi Lee. *arxiv*, 2019.

Professional Services

- Co-organizer, **Multilingual Information Access Workshop at NAACL 2022**
- Program Committee (Reviewer) EMNLP 2022-2023, ARR 2022-present, NAACL 2022, COLING 2022

Honors and Awards

- Research scholarship, **Ministry of Education of Taiwan** 2021-2023
- Rushmer Fellowship, **University of Washington** 2019
- The Phi Tau Phi Scholastic Honor Society of Taiwan (top 3% of EECS class) 2021

Teaching Experiences

- **Teaching Assistant** **Seattle, WA**
[CSE 473] *University of Washington, Introduction to Artificial Intelligence* Sep, 2023 – Dec, 2023

- **Teaching Assistant**
[EE 596] University of Washington, LLMs: From Transformers to GPT **Seattle, WA**
Jan, 2024 – March, 2024
- **Head Teaching Assistant**
[EE 5184] National Taiwan University, Machine Learning **Taiwan**
Sep, 2017 – Jan, 2018

Skills

- Software: Python, C/C++, Matlab, R, Tensorflow, PyTorch, Theano, Jax