CHENG-I JEFF LAI

Mountain View, CA jefflai108@gmail.com linkedin.com/in/jefflai108 Google Scholar

> Research scientist building real-time speech-to-speech and audio-text LLM systems. Led core user-memory work for Meta's voice AI companion — from post-training data curation and eval to real-time infra integration. Previously architected sparse MoE and audio-visual speech models scaled to 512 TPUs / 256 GPUs. PhD MIT; experience at Meta MSL, WaveForms AI, Google DeepMind, and others.

TECHNICAL SKILLS

Speech / ML	Speech-to-speech, ASR, TTS, real-time voice agents, audio-visual LLMs
Systems	Python, PyTorch, JAX, distributed GPU training, data curation pipelines

PROFESSIONAL EXPERIENCE

Recent Applied Voice & Foundation Model Work

Meta Superintelligence Labs

Research Scientist Aug 2025 – Present

- Scaling real-time audio-text LLMs from research to production, focusing on long-term memory and user-specific personalization for voice AI companions.
- Leading user-memory post-training data curation and evaluation as part of the core real-time voice post-training recipe.
- Co-leading infra integration for real-time memory extraction and personalized response generation in streaming voice interactions.

WaveForms AI — Acquired by Meta

Member of Technical Staff May 2025 – Jul 2025

- Architected the initial audio-text MoE architecture for real-time voice modeling, transitioning the stack from dense to sparse experts.
- Orchestrated large-scale mid-training across 256 GPUs, optimizing expert load balancing, throughput, and training stability.
- Owned model-training tradeoffs across routing, checkpointing, and systems throughput for sparse audio-text models.

Google DeepMind, Speech

Student Researcher Jun 2024 – Dec 2024

- Built Audio-Visual Gemma (3B/10B/28B), a foundation model family for natively visually-coherent speech generation.
- Led large-scale audio-visual data curation (285k hours of spoken captions) and orchestrated training on 512 TPUs.
- Achieved SOTA results in Spoken Visual QA, Video-to-Speech generation, and Audio-Visual Sound Understanding.

Selected Prior Research Experience

Apple Foundation Models

Research Scientist Intern May 2023 – Aug 2023

- Developed instruction-following ASR models that follow natural language prompts for controllable transcription.

Meta FAIR

Research Scientist Intern May 2022 – Sep 2022

- Investigated grounding sources for word discovery, applied to direct speech-to-speech translation systems.

MIT-IBM Watson AI Lab

Research Scientist Intern Jun 2021 – Sep 2021

- Developed pruning-based self-supervised ASR methods recognized as a NeurIPS 2021 Spotlight.
- Designed disentanglement techniques for self-supervised speech representations.

Amazon AI

Applied Scientist Intern May 2020 – Aug 2020

- Built self-supervised spoken language understanding models under limited and noisy data regimes.

EDUCATION

Massachusetts Institute of Technology (MIT)

Sep 2019 – Aug 2025

Ph.D. and S.M. in Electrical Engineering and Computer Science

Thesis: Language Modeling from Visually Grounded Speech · Merrill Lynch Fellowship

Johns Hopkins University

Sep 2015 – Dec 2018

B.S. in Electrical Engineering

Vredenburg Scholarship · Dean's List (All semesters)

SELECTED PUBLICATIONS — Google Scholar: 3,400+ citations

1. **Cheng-I Jeff Lai***, Zhiyun Lu, Liangliang Cao, Ruoming Pang. "Instruction-Following Speech Recognition," *NeurIPS 2024 workshop*.
2. **Cheng-I Jeff Lai***, Freda Shi*, Puyuan Peng*, Yoon Kim, Kevin Gimpel, Shiyu Chang, Yung-Sung Chuang, Saurabhchand Bhati, David Cox, David Harwath, Yang Zhang, Karen Livescu, James Glass. "Textless Phrase Structure Induction From Visually-Grounded Speech," *ASRU 2023*.
3. Alexander H. Liu*, **Cheng-I Jeff Lai***, Wei-Ning Hsu, Michael Auli, Alexei Baevski, James Glass. "Simple and Effective Unsupervised Speech Synthesis," *Interspeech 2022*.
4. **Cheng-I Jeff Lai**, Yang Zhang*, Alexander H. Liu*, Shiyu Chang*, Kaizhi Qian, Yi-Lun Liao, Yung-Sung Chuang, Sameer Khurana, David Cox, James Glass. "PARP: Prune, Adjust and Re-Prune for Self-Supervised Speech Recognition," *NeurIPS 2021* **(Spotlight)**.
5. Yuan Gong, **Cheng-I Jeff Lai**, Yu-An Chung, James Glass. "SSAST: Self-Supervised Audio Spectrogram Transformer," *AAAI 2022*.
6. Shu-wen Yang, Po-Han Chi*, Yung-Sung Chuang*, **Cheng-I Jeff Lai***, Kushal Lakhotia*, Yist Y. Lin*, Andy T. Liu*, Jiatong Shi*, Xuankai Chang, et al. "SUPERB: Speech Processing Universal PERformance Benchmark," *Interspeech 2021*.