

# Jason Chun Lok Li

jasonli0707.github.io · Email · Google Scholar · GitHub · LinkedIn

*AI research engineer working on generative modeling, representation learning, and efficient inference. Recent work spans full-duplex spoken dialogue models and neural audio codecs for music generation, including synthetic data generation, post-training, and flow-matching decoders. Ph.D. from the University of Hong Kong, with first- or co-first-authored publications at ICLR, AAAI, IJCAI, and EMNLP.*

## EDUCATION

### University of Hong Kong

- **Doctor of Philosophy** (Electrical and Electronic Engineering) (Department GPA: 3.9) Dec 2020 – Sep 2024
  - Supervised by Prof. Ngai Wong (Primary) and Prof. Edmund Y. Lam
- **Bachelor of Engineering** (Electronic Engineering) (CGPA: 3.5) Class of 2020

## INDUSTRY EXPERIENCE

### Senior Research Engineer at Audio Speech Lab, Huawei Hong Kong Research Center (HKRC) Oct 2024 – Present

- Developing an full-duplex spoken dialogue model with text output, owning synthetic dialogue data generation, model architecture, and post-training for turn-taking, backchannels, interruption handling, and semantic rejection
- Led a team of 4 engineers and researchers developing a neural audio codec for music generation, spanning self-supervised encoders (BestRQ, MAE, MLM), vector quantization (VQ, FSQ, RVQ), and generative decoders (flow matching, Wave-VAE); evaluated internal models against MuCodec and X-Codec
- Fine-tuned a pretrained audio language model for ASR, running ablations on training strategies (LoRA vs. full SFT), CTC decoding and dynamic frame-rate adapters

### Research Intern at Fermat Lab, Huawei Hong Kong Research Center (HKRC) Apr 2024 – Oct 2024

- Conducted research on MLLM agents for mobile application navigation
- Internship concluded with a co-first-authored EMNLP 2025 paper: KG-RAG, a knowledge graph-driven retrieval-augmented generation framework for GUI agent decision-making

### Research Intern at TCL Corporate Research (HK) Co., Ltd Sep 2023 – Feb 2024

- Benchmarked LLM inference acceleration using vLLM and low-bit quantization, evaluating throughput, memory usage, and model quality
- Investigated parallel and speculative decoding methods, including Lookahead Decoding (LADE), Prompt Lookup Decoding (PLD), and Retrieval-Based Speculative Decoding (REST)
- Conducted research on efficient diffusion models for edge devices

### Research Assistant at HKU AI-WiSe | University of Hong Kong Jun – Nov 2020

- Designed and developed a cross-platform mobile application “UMeAir” (available for [iOS](#) & [Android](#)) from scratch for air pollution monitoring and forecasting using Flutter
- Implemented a route advisor which provides suggestions based on traveling distance and estimated air pollution exposure
- Developed a Flask application to host APIs for fine-grained air quality estimation
- Leveraged various third party technologies including Google Map API (Directions, AutoComplete, Geocoding), Firebase Authentication and OpenWeatherAPI
- Supervised by Prof. Victor O.K. Li and Dr Jacqueline C.K. Lam

### Research Engineer (Part-time) | MatrixSense Jul – Nov 2020

- Reimplemented an end-to-end goal-oriented dialog system named Meta-Dialog System (MDS) (Dai et al., 2020)
- Conducted research on few-shot cross-domain & cross-lingual task-oriented dialog system

### Research Intern (Part-time) | Fano Labs Sep 2019 – May 2020

- Implemented an integrated model of the Pointer-Generator Network (See et al., 2017) and BERTSUM (Liu et al., 2019)
- Supervised by Dr Albert Y.S. Lam

### Summer Intern | ASTRI Jun – Aug 2019

- Researched and developed a proof-of-concept blockchain application for sharing storage via ASTRI distributed file system
- Implemented smart contracts using Go & Go-kit
- Presented findings to an audience of size 70+ including C-level executives at the Hong Kong's leading research lab ASTRI

### Student Research Assistant (Part-time) | University of Hong Kong Nov 2018 – Mar 2019

- Developed a cross-platform mobile application “SmartEye” for electricity consumption monitoring using React Native
- Implemented a notification system with Firebase Cloud Messaging(FCM)
- Supervised by Prof. Victor O.K. Li

### R&D Engineering Intern | Chun Shing Development (HK) Ltd July – Oct 2018

- Designed and developed a proof-of-concept smart recycling bin using Arduino and ESP8266 WiFi module
- Developed the backend waste management system with PHP, SQL and CodeIgniter framework
- Demonstrated on the Eco Expo Asia 2018

## PUBLICATIONS

### PEER-REVIEWED PUBLICATIONS

- M. Xu, J. Chen, Y. Zhao, **Jason Chun Lok Li**, Y. Qiu, Z. Du, M. Wu, et al. VP-Bench: A Comprehensive Benchmark for Visual Prompting in Multimodal Large Language Models. **AAAI 2026**.
- D. Xu, H. Yang, Y. Zhao, P. Zhang, J. Chen, W. Ma, Z. Hou, M. Wu, X. Li, S. Hu, Z. Guan, **Jason Chun Lok Li**, L.-M. Po. From Exploration to Exploitation: A Two-Stage Entropy RLVR Approach for Noise-Tolerant MLLM Training. **CVPR 2026**.
- H. Yang, **Jason Chun Lok Li**, K. Li, W. Ma, M. Xu, Y. Zhao, L.-M. Po. RSTFA: Efficient Training-Free Human-Preference Alignment via Rejection Sampling for Text-to-Image Diffusion Models. **IEEE TIP 35, 4571–4586 (2026)**. [[paper](#)]
- Z. Guan\*, **Jason Chun Lok Li**\*, Z. Hou\*, P. Zhang, D. Xu, Y. Zhao, M. Wu, J. Chen, T.-T. Nguyen, P. Xian, W. Ma, S. Qin, G. Chesi, N. Wong. KG-RAG: Enhancing GUI Agent Decision-Making via Knowledge Graph-Driven Retrieval-Augmented Generation. **EMNLP 2025**. [[paper](#)]
- R. Lin\*, **Jason Chun Lok Li**\*, J. Zhou, B. Huang, J. Ran, N. Wong. Lite it fly: An All-Deformable-Butterfly Network. **IEEE TNNLS 36(1), 1919–1924 (2025)**. [[paper](#), [code](#)]
- M. Xu, M. Wu, Y. Zhao, **Jason Chun Lok Li**, W. Ou. LLaVA-SpaceSGG: Visual Instruct Tuning for Open-vocabulary Scene Graph Generation with Enhanced Spatial Relations. **WACV 2025**.
- B. Li\*, Z. Ai\*, B. Jiang, B. Huang, **Jason Chun Lok Li**, J. Liu, Z. Tu, G. Wang, D. Yu, N. Wong. BDLUT: Blind Image Denoising with Hardware-Optimized Look-Up Tables. **SID Symposium Digest 56(1), 282–285 (2025)**. **Distinguished Student Paper**. [[paper](#)]
- **Jason Chun Lok Li**\*, S. T. S. Luo\*, L. Xu, N. Wong. ASMR: Activation-Sharing Multi-Resolution Coordinate Networks for Efficient Inference. **ICLR 2024**. [[paper](#), [poster](#), [slides](#), [code](#)]
- **Jason Chun Lok Li**\*, C. Liu\*, B. Huang, N. Wong. Learning Spatially Collaged Fourier Bases for Implicit Neural Representation. **AAAI 2024**. [[paper](#), [poster](#), [code](#)]
- B. Huang\*, **Jason Chun Lok Li**\*, J. Ran, B. Li, J. Zhou, D. Yu, N. Wong. Hundred-Kilobyte Lookup Tables for Efficient Single-Image Super-Resolution. **IJCAI 2024**. [[paper](#), [code](#)]
- C. Zhang\*, S. T. S. Luo\*, **Jason Chun Lok Li**, Y.-C. Wu, N. Wong. Nonparametric Teaching of Implicit Neural Representations. **ICML 2024**. [[paper](#), [poster](#), [slides](#), [code](#), [page](#)]
- **Jason Chun Lok Li**\*, R. Lin\*, J. Zhou, E. Y. M. Lam, N. Wong. A Unifying Tensor View for Lightweight CNNs. **ASICON 2023**. [[paper](#)]
- J. Ran, R. Lin, **Jason Chun Lok Li**, J. Zhou, N. Wong. PECAN: A Product-Quantized Content Addressable Memory Network. **DATE 2023**. [[paper](#)]

### PREPRINTS

- B. Huang, **Jason Chun Lok Li**, C. Liu, N. Wong. Poisoning-based Backdoor Attacks for Arbitrary Target Label with Positive Triggers. **arXiv:2405.05573 (2024)**. [[paper](#)]
- **Jason Chun Lok Li**, Y. Han, J. Lam, V. Li, S. Matthews, L. Cheung, ..., I. Gozes. Correlation Between the Bilingual Status and the Onset Age of AD and MCI Subjects: Evidence from the ADNI Dataset. **Preprints (2021)**. [[paper](#)]

\* Equal contribution.

### HIGHLIGHT PROJECTS/ACTIVITIES

- Teaching Assistant, HKU – ELEC3442 Embedded Systems (Spring 2022, 2023); ELEC2840 Engineering Training, IoT Module (Summer 2021)
- 1st Runners-up, Innotech Law Hackathon: Belt & Road Justice Challenge (2018) – built an NLP web app (transcription, translation, QA, NER) improving multilingual access to justice
- Cathay Pacific 24-hour Hackathon (2018); iOS Application Development HK Regional Contest (2018)
- Core member, A.I. Paper Reading Group – hosted weekly research paper discussions (2018 – 2020)

### AWARDS:

- Postgraduate Scholarship (PGS) (2020-2024)
- Talent Development Scholarship (2018)
- Hong Kong Electric Electrical Energy Scholarship (2017)
- Silver Award for Volunteer Service (2016)
- MTR Corporation Scholarship Certificate (2015)

### SKILLS

**Programming & Frameworks:** Python, PyTorch, JAX, TensorFlow, Dart, Go, SQL

**Generative Modeling & Representation Learning:** Flow matching, self-supervised learning, vector quantization, neural audio codecs

**Model Training & Post-training:** SFT, LoRA, synthetic dialogue data generation

**Efficient Inference:** Low-bit quantization, parallel and speculative decoding, vLLM

**Applications:** Full-duplex spoken dialogue, ASR, music generation

**Application Development:** Linux, Flutter, Flask, Firebase, Google Cloud

**Languages:** Cantonese (Native), Mandarin (Fluent), English (Fluent)