Andrew Koh

dhecloud | andrewkjj@gmail.com

Education

2020 – 2024 **Doctor of Philosophy (Ph.D.) Computer Science**
Singapore, Nanyang Technological University
Doctorate Thesis: **Audio Captioning and Retrieval with improved Cross-Modal Objectives**

2015 – 2019 **Bachelor of Engineering (Computer Engineering)**
Singapore, Nanyang Technological University
CGPA: 4.06 / 5.00 (Second Upper Class Honours)

Projects

SakuraSensei: A Japanese Conversational AI Tutor 2025

- Using **LangChain**, built a fully context-aware Japanese conversational Telegram bot with a custom persona Sakura Sensei and memory persistence.
- Implemented a multi-dataset **RAG** pipeline using scraped JLPT grammar/vocab, JMDICT, Tatoeba, and JaSquad; embedded using multilingual transformer models.
- Developed automated Japanese news explanation using multi-agent workflows, BeautifulSoup4 scraping, safety filtering, and memory injection strategies.
- Engineered an automated cloze-question generator using YouTube transcripts, difflib distractors, **Whisper + VAD segmentation**, and kanji-level extraction.
- Optimized retrieval by restructuring metadata filtering, increasing search depth, and upgrading embedding models for Japanese semantic similarity.

FaceChangerGIFBot: Swap your face into funny GIFs and clips 2025

- Developed a real-time face swap Telegram bot that processes user-provided photos, GIFs, and videos using ONNX-based inference.
- Built a usage tracking system with daily free tries, lifetime quotas, and premium tiers paid via Stripe.
- Added premium-only features, watermarking, size/duration validation, and a content moderation pipeline.
- Implemented persistent webhook hosting by replacing Ngrok with a Cloudflare Tunnel–based solution.
- Designed a "Featured Videos" system where users can spend tries to swap their face into pre-selected clips (e.g., music videos).

Acoustic Event Detection for Emergency Telephony Data 2023-2024

- Built a streaming audio event-detection system with a custom PyTorch model capable of classifying 3-second audio windows into nine acoustic event classes (e.g., speech, sirens, explosions, gunshots.
- Developed the full inference pipeline—including log-mel spectrogram feature extraction, silence detection, and top-k caption generation in SRT/XML/JSON—along with Dockerized deployment and a configurable streaming API.
- Integrated offline VAD, optimized model parameters, and provided tooling for folder-based and real-time audio prediction.

Hand Pose Estimation for Rheumatoid Arthritis 2018

– Implemented Region Ensemble Network from scratch in PyTorch and dlib for hand pose estimation.
– Developed real-time tracking application using Microsoft Kinect and augmented dataset with geometric transformations.

Work Experience

Research Assistant, Temasek Laboratories Jan 2020 – Dec 2023
– Managed a research team of graduate and undergraduate students.
– Conducted research and authored papers in Audio Processing including Automated Audio Captioning, Language-Based Audio Retrieval, and Acoustic Scene Classification.
– Collaborated with external organizations to bring AI models to production.

Research Assistant, Temasek Laboratories Jan 2019 – Dec 2020
– Implemented and trained a state-of-the-art Question-Answer System (QASA) using PyTorch for production.
– Improved QASA with transformer-based architectures.
– Documented experiments and utilized High-Performance Computing platforms for training.
– Created a demo webserver for (testing).

Research Assistant, Nanyang Technological University Aug 2017 – Feb 2018
– Created a hybrid recommendation engine using Singular Value Decomposition (SVD) in python
– Used Natural Language Processing to extract keywords from tweets and determine a measure of similarity
– Collaborated with front end web developers to develop a listening web app for Twitter

Research Intern, SAP Asia Pte Ltd May 2017 – July 2017
– Trained deep learning models (tf-faster-rcnn) for computer vision for object localization and classification
– Documented experiements, analyzed test results and oversaw data collection.
– Set up a micro-service using Flask for use in conjunction with a machine learning platform

Other Experience

Freelance Web Design Projects 2025
Created and maintained custom websites for small businesses and personal clients using modern front-end frameworks like Wordpress.

National Data Science Challenge Singapore 2019 Feb 2019 – Mar 2019
Achieved Top 10 / 360 teams. Trained BERT in NLP and CV to improve state-of-the-art architectures and parallelized experiments to enhance leaderboard scores.

Publications

[1] A. Koh and C. E. Siong, "Language-based audio retrieval with converging tied layers and contrastive loss," in *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2022, pp. 1644–1648. DOI: 10.23919/APSIPAASC55919.2022.9979840.

[2] A. Koh, S. Tiwari, and C. E. Siong, "Automated audio captioning with epochal difficult captions for curriculum learning," in *2022 Asia-Pacific Signal and Information Processing Association Annual*

Summit and Conference (APSIPA ASC), 2022, pp. 1058–1063. DOI: 10.23919/APSIPAASC55919.2022.9980242.

[3] A. Koh, X. Fuzhao, and C. E. Siong, "Automated audio captioning using transfer learning and reconstruction latent space similarity regularization," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 7722–7726. DOI: 10.1109/ICASSP43922.2022.9747676.

[4] T. Khandelwal, R. K. Das, A. Koh, and E. S. Chng, *Leveraging audio-tagging assisted sound event detection using weakified strong labels and frequency dynamic convolutions*, 2023. arXiv: 2304.12688 [eess.AS].

[5] B. P. Yap, A. Koh, and E. S. Chng, "Adapting BERT for word sense disambiguation with gloss selection objective and example sentences," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, Online: Association for Computational Linguistics, Nov. 2020, pp. 41–46. DOI: 10.18653/v1/2020.findings-emnlp.4. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.4.

Skills

Keywords	Python, PyTorch, TensorFlow, dlib, NLP, Computer Vision, GPTs, transformers
Languages	English (Native), Chinese (Conversational), Japanese (Conversational)