

# KALVIN CHANG

kalvinc@cs.cmu.edu • (310) 666-5091 • kalvinchang.github.io • linkedin.com/in/kalvinc

## EDUCATION

### Carnegie Mellon University – School of Computer Science

Master of Science in Language Technologies

GPA: 4.26 / 4.33 (Advisor: David Mortensen)

Selected Coursework: Machine Learning for Structured Data, Multilingual Natural Language Processing, Intro to Machine Learning (Teaching Assistant), Phonetics and Phonemics (Univ. of Pittsburgh)

Pittsburgh, PA

Jan 2022 - May 2023

Bachelor of Science in Computer Science, Concentration in Human-Computer Interaction

Aug 2018 - Dec 2021

GPA: 3.67 / 4.0 (Dean's List: Fall 2020, Spring 2021, Fall 2021; University Honors)

Selected Coursework: Advanced NLP, Speech Processing, Distributed Systems, Design & Analysis of Algorithms, Language Variation & Change, Language Diversity & Cultural Identity, Ethics & Policy Issues in Computing

## SKILLS

**Technical:** Python, Go, Java, C, PyTorch, Amazon Web Services, SML, Ruby, HTML/CSS, Git

**Languages:** English (native), Mandarin Chinese (native), Spanish (CA Bilingual Seal), Taiwanese Hokkien

## EXPERIENCE

### Amazon Web Services

May 2022 - Aug 2022

#### Machine Learning Engineer Intern, AWS AI Themis Team

Seattle, WA

- Designed a phonetic-acoustic feature space to discover granular clusters of high error samples in an **ASR** system
- Built a cascade **phonetic** feature extraction pipeline involving phone recognition and phone embedding lookups
- Extracted acoustic embeddings from a pretrained **wav2vec 2.0** model in **PyTorch**
- Experimented with [CLS]-token pooling and visualized the pooled embeddings using PCA, t-SNE, and UMAP

#### Software Development Engineer Intern, EC2 Instance Quality Team

Jun 2021 - Aug 2021

- Implemented **pagination** and **presigned URL** for large payloads to work around Lambda's 6 MB payload limit
- Sped up the downloading of **S3** objects with **multithreading** for a **Java** Lambda to stop Lambda from timing out
- Onboarded additional features for a Lambda / **React** frontend to help on-call engineers diagnose **EC2** Instances
- Caught elusive bug in the **AWS** Java SDK after collaborating with two SDK teams, leading to a bug fix

### Cardinal Blue Software, Inc.

Feb 2020 - Jul 2020

#### Server Developer Intern, PicCollage

Taipei, Taiwan

- Investigated **PostgreSQL** transition to transactional pooling, which later increased database scalability **18-fold**
- Handled edge cases in a **Ruby on Rails** server's Sign in with Apple auth flow used by thousands of users
- Rerouted requests on the main server's third busiest route to a microservice by using an **Nginx** reverse proxy
- Implemented active cache invalidation to fix stale content bug caused by time-based cache expiry

### Carnegie Mellon University

Jan 2022 - Current

#### Graduate Research Assistant, Language Technologies Institute

Pittsburgh, PA

- Implemented RNN encoder-decoder model in PyTorch for protoform reconstruction and adapted to **Chinese** data
- Expanded an open source G2P package ([epitran](https://github.com/epitran))'s coverage of **lower resource** Chinese varieties

#### Teaching Assistant, 10-601 Intro to Machine Learning

Aug 2022 - Dec 2022

- Refined coursework for a **400+** student course on concepts such as Hidden Markov Models & gradient descent
- Taught recitations, held office hours, answered questions to provide intuition on key concepts and homework

## PUBLICATIONS

- [1] **Kalvin Chang**, Chenxuan Cui, Youngmin Kim, and David Mortensen. 2022. WikiHan: A New Comparative Dataset for Chinese Languages. In *Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022)*, Gyeongju, Korea.

## PROJECTS

### Morphotactics (<https://anaconda.org/conda-forge/morphotactics>)

Feb 2021 - May 2021

- Implemented a **Python** interface for linguists to express a language's **morphology** with **finite state transducers**
- Compiled dependencies among inputted morpheme classes with a depth-first graph search algorithm