Alan (Haoxin) Li

haoxin.li@yale.edu ☎ 206-257-9667 Webpage LinkedIn: Alan Li

Education

Yale University, Computer Science — **New Haven, CT**
PhD Student at Computer Science, YaleNLP — Sept 2024–Present
Advised by Prof. Arman Cohan.

University of Washington, Paul G. Allen School of CSE — **Seattle, WA**
B.S./M.S. Combined at Computer Engineering | GPA B.S. 3.89/M.S. 3.96 — Sept 2020–Jun 2024
Mentored by Prof. Noah A. Smith, Prof. Jungo Kasai, and Mr. Phillip Keung.

Work Experiences

Research Scientist Intern | *Kotoba Technologies, Tokyo Japan* — Summer 2024
- Developed **end-to-end speech-to-speech models** for Japanese ↔English translation. (Demo)
- Developed **offline real-time speech-to-text** Japanese ↔English translation models. (Demo)

Machine Learning Engineer Intern | *DGene Digital Technology, Shanghai* — Winter 2021–Summer 2021
- Developed deep learning **UNet** based **class-independent image matting and semantic segmentation** algorithms on 4K images that features iterative refinements with user inputs.
- Developed in-house **multi-camera calibration** libraries for 3D reconstruction with **C++ (Pybind11)** backend.

Research Experiences

Research Assistant, PhD | *YaleNLP (New Haven, CT)* — Autumn 2024–Present
Supervised by Prof. Arman Cohan, collaborating with AI2.
- **Efficient domain adaptation**.
  - ◇ Building strong scientific LLMs with fully open-source data.
  - ◇ Exploring efficient recipe for domain adaptation and continual pretraining.

Research Assistant | *Noah's ARK Lab (Seattle, WA)* — Winter 2022–Spring 2024
Supervised by Prof. Noah A. Smith, Prof. Jungo Kasai, and Mr. Phillip Keung.
- **Model architecture** and **model efficiency**.
  - ◇ Proposed efficient transformer architectures by alleviating intrinsic sequence representation redundancy.
  - ◇ Explored efficient finetuning with Direct Preference Optimization (DPO).
- **Information retrieval** and **generative retrieval**.
  - ◇ Proposed summarization-based document IDs for generative retrieval.

Publications

2024 ***ACID: Abstractive, Content-based IDs for Document Retrieval with Language Models*** (Link)
Haoxin Li, Phillip Keung, Daniel Cheng, Jungo Kasai, Noah A. Smith.
Keywords: generative retrieval, document identifiers, large language models (LLMs).
Preprint.

2023 ***NarrowBERT: Accelerating Masked Language Model Pretraining and Inference*** (Link, Code, Video)
Haoxin Li, Phillip Keung, Daniel Cheng, Jungo Kasai, Noah A. Smith.
Keywords: masked language model (MLM), efficiency, sparsity.
To appear in ACL 2023 (main conference).