

# Alan (Haoxin) Li

 [haoxin.li@yale.edu](mailto:haoxin.li@yale.edu)  206-257-9667  [Webpage](#)  [LinkedIn: Alan Li](#)

## Education

---

### Yale University, Computer Science

*PhD Student at Computer Science, YaleNLP*

Advised by Prof. Arman Cohan.

New Haven, CT

Sept 2024–Present

### University of Washington, Paul G. Allen School of CSE

*B.S./M.S. Combined at Computer Engineering | Fully Funded M.S.*

Mentored by Prof. Noah A. Smith, Prof. Jungo Kasai, and Mr. Phillip Keung.

Seattle, WA

Sept 2020–Jun 2024

## Work Experiences

---

### Research Scientist Intern | Kotoba Technologies, Tokyo Japan

Summer 2024

- Developed **efficient real-time speech-to-text** Japanese ↔English translation models. ([Demo](#))
- Developed **multimodal end-to-end audio language models** for Japanese ↔English translation. ([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.*

- ◇ Leveraging test-time reasoning to train scientific LLMs.

### 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.*

- ◇ Proposed efficient transformer architectures by alleviating intrinsic sequence representation redundancy.

## Publications

---

### 2025 ***Survey on Evaluation of LLM-based Agents*** ([Link](#))

Asaf Yehudai, Lilach Edelstein, **Alan Li**, Guy Uziel, Yilun Zhao, Roy Bar-Haim, Arman Cohan, Michal Shmueli-Scheuer

*Preprint. Under Review.*

### ***Multi-Perspective Integration via Mixture-of-Agents for Enhanced Healthcare QA Summarization***

Dongsuk Jang, **Alan Li**, Arman Cohan

**#2** at [CL4Health@NAACL2025 Leaderboard](#). *Under Review.*

### 2024 ***SciRIFF: A Resource to Enhance Language Model Instruction-Following over Scientific Literature***

David Wadden\*, Kejian Shi\*, Jacob Morrison, **Alan Li**, Aakanksha Naik, Shruti Singh, Nitzan Barzilay, Kyle Lo, Tom Hope, Luca Soldaini, Zejiang Shen, Doug Downey, Hannaneh Hajishirzi, Arman Cohan

*Under Review.*

**Summarization-Based Document IDs for Generative Retrieval with Language Models** ([Link](#))

**Alan Li**, Phillip Keung, Daniel Cheng, Jungo Kasai, Noah A. Smith.

*EMNLP 2024 (Workshop: NLP for Wikipedia).*

2023 **NarrowBERT: Accelerating Masked Language Model Pretraining and Inference** ([Link](#), [Code](#), [Video](#))

**Alan Li**, Phillip Keung, Daniel Cheng, Jungo Kasai, Noah A. Smith.

*ACL 2023 (main conference).*

---

\*These authors contributed equally to this work.