

# ANDREW D. SHIN

andrew.shin@nih.gov ◇ andrewshin8@gmail.com ◇ www.andrewshin.org

## EXPERIENCE

---

### Research Fellow

National Library of Medicine (NLM), NIH

Sep 2022 - Present

Bethesda, MD

- Worked with Dr. Qiao Jin, in Dr. Zhiyong Lu's NLP lab, NCBI/NLM/NIH
- Research integrated into PubMed (7 million monthly users / biomedical scientists across globe)
- Research spanning bioNLP, information retrieval, machine learning

## PUBLICATIONS

---

**Andrew Shin**, Qiao Jin, Zhiyong Lu. Multi-stage Biomedical Literature Retrieval System. *work in progress*.

Qiao Jin, **Andrew Shin**, Zhiyong Lu. LADER: Log-Augmented DENSE Retrieval for Biomedical Literature Search. *ACM SIGIR (SIG Information Retrieval) 2023*. [Link]

James Anibal, Adam Landa, Hang Nguyen, Alec Peltekian, **Andrew Shin** ... David Clifton, Bradford Wood. Digital Omicron Detection using Unscripted Voice Samples from Social Media. *Nature Medicine 2023*. [Link]

## PROJECTS

---

**Multi-Stage Document Retrieval System** Implemented system that uses a bi-encoder for retrieval and a cross-encoder model for reranking. Both models initialized with BERT and further trained on query-article search logs of unprecedented scale, with 100M+ query-article pairs, each consisting of a user query and a document clicked by the user. *Pytorch, Hugging Face, FAISS, Numpy, Pandas*

**Sparse Retriever Baseline** Implemented BM25 for sparse retrieval baseline as part of LADER ablation study. 35M PubMed articles indexed/searched. *Pyserini/Lucene, Numpy, Pandas*

**Transformers for PPG/ECG data** Built attention models and novel contrastive learning objectives specifically for extremely long PPG/ECG waveform sequences. *Pytorch, Hugging Face, Scikit-Learn, Numpy, Pandas*

**BearMaps** Wrote the backend for a Google Maps-like web application, with scrolling and zoom in/out for the city of Berkeley, CA. Implemented fastest route with K-D trees and A\* Search Algorithm. *Java, Apache Maven, Junit*

**Gitlet** Working version of Git, with basic functions such as init, commit, push, branch, checkout, merge, etc. *Java*

## HONORS

---

**2nd Place, 2023 BioASQ<sup>1</sup>** Represented NLM at BioASQ 2023, document retrieval subtask. First postbac research fellow to lead a BioASQ NLM team.

## SKILLS

---

**Languages/Tools** Python, Java, Pytorch, Hugging Face, Google Cloud, Git

## EDUCATION

---

**B.A. in Philosophy** University of California, Santa Barbara

2022

*Relevant Coursework:* Real Analysis, Linear Algebra, Topology, Discrete Mathematics, Formal Logic, Algorithms & Data Structures, Philosophy of Language, Philosophy of Science, Metaphysics, Hume.

---

<sup>1</sup>2023 BioASQ Biomedical Semantic Question Answering Challenge, Document Retrieval Subtask (subtask B, phase A). 3rd place in batches 1 and 3, 2nd place in batch 2. Regular BioASQ participants include Google Research.