ANDREW D. SHIN

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

EXPERIENCE

Research Fellow Sep 2022 - Present
National Library of Medicine (NLM), NIH *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 Document Retrieval System for Biomedical Literature. *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 T Anibal, Adam Landa, Hang Nguyen, Alec Peltekian, **Andrew D Shin** ... David Clifton, Bradford Wood. Digital Omicron Detection using Unscripted Voice Samples from Social Media. *Nature Medicine 2023.* [Link]

PROJECTS

Multi-Stage Biomedical Document Retrieval System Implemented system which uses a bi-encoder for retrieval and a cross-encoder model for reranking. Both models initialized with PubMedBERT and further trained on PubMed query-article search logs of unprecedented scale, with 255 million query-article pairs, each consisting of a user query and a document clicked by the user in their PubMed search. *Pytorch, Hugging Face, FAISS, 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*

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, California. Used K-D trees and A* Search Algorithm for implementing the fastest route, given starting and ending points. *Java, Apache Maven, Junit*

Gitlet In a team of 4, wrote a working version of Git, with all the basic functionalities such as init, add, commit, push, pull, rm, branch, checkout, merge, etc. *Java*

HONORS

2nd Place, 2023 BioASQ[1] Represented NLM at BioASQ 2023, document retrieval subtask. First postbac research fellow to lead a BioASQ NLM team. Regular participants include Google Research.

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.

[1] 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.