ASHLEY SHIN

ashleyhshin@gmail.com ⋄ www.ashleyshin.org

Interests: information retrieval, natural language processing, representation learning, ML for medicine, ethics

EDUCATION

B.A. in Philosophy University of California, Santa Barbara 2022

Relevant Coursework: Data Structures, Algorithms, Discrete Math, Linear Algebra, Real Analysis, Symbolic Logic, Topology; Ethics, Philosophy of Language, Philosophy of Science, Metaphysics.

EXPERIENCE

National Library of Medicine (NLM), NIH *Bethesda, MD*
Research Fellow Sep 2022 - Present

- Advised by Dr. Qiao Jin and Dr. Zhiyong Lu in bioNLP group, NCBI/NLM/NIH
- Research spanning biomedical natural language processing (bioNLP), information retrieval, and machine learning, aimed at improving PubMed, an academic search engine used by 7 million researchers

HONORS

NSF Graduate Fellowship CSGrad4US, 2023 cohort. Selected based on demonstrated potential in pursuing a doctorate in a CISE field. $159k in total funding upon enrollment in PhD program.

Top 3, BioASQ Challenge[1] 2023 Represented NCBI/NLM at BioASQ, document retrieval subtask. First postbac fellow to lead NLM team at BioASQ. Past NLM participants were postdocs and staff scientists.

NIH Intramural Research Training Award Selected for postbaccalaureate training in biomedical research at the National Institutes of Health

4th place, UCSB ACM-ICPC 2022 Regional algorithmic programming contest.

2nd place, Stanford ProCo 2015 Algorithmic programming contest in the style of ACM-ICPC.

PUBLICATIONS

[1] **Ashley Shin**, Qiao Jin, Zhiyong Lu. Harnessing PubMed User Query Logs for Post Hoc Explanations of Recommended Similar Articles. *Under review 2024.* [link]

[2] **Ashley Shin**, Qiao Jin, Zhiyong Lu. Multi-stage Literature Retrieval System Trained by PubMed Search Logs for Biomedical Question Answering. *CLEF (BioASQ workshop) 2023.* [link]

[3] Qiao Jin, **Ashley Shin**, Zhiyong Lu. LADER: Log-Augmented DEnse Retrieval for Biomedical Literature Search. *ACM SIGIR (Information Retrieval) 2023.* [link]

[4] James Anibal, Adam Landa, Hang Nguyen, Alec Peltekian, **Ashley Shin** ... David Clifton, Bradford Wood. Digital Omicron Detection using Unscripted Voice Samples from Social Media. [link]

PROJECTS

Similar Articles Project Preprocessed PubMed user query-click logs to train a BERT-based model for binary token classification: given a seed article and a "similar article" recommended by PubMed, determine which tokens in the article title to highlight for user convenience. Superior performance over common baselines in internal tests – F_1 of 82.8 (ours) versus word2vec (55.5), SBERT(65.9). *Pytorch, Hugging Face.* Led to [1]

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 255M query-article pairs, each consisting of a user query and a document clicked by the user. *Pytorch, Hugging Face, FAISS, Numpy, Pandas.* Led to [2]

Pubmed Log-Augmented Sparse Retriever Implemented log-augmented sparse retrieval baseline with BM25 as part of LADER ablation study. 35M documents indexed/searched. *Pyserini/Lucene, Numpy, Pandas.* Led to [3]

[1] BioASQ Biomedical Semantic Question Answering Challenge. Past participants include Google Research, UCSD, U. Mass.

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*

SKILLS

Languages	Python, Java, C++, JavaScript
Libraries	Pytorch, Hugging Face Transformers, FAISS, Numpy, Pandas
Tools	LaTeX, Git, Vim, CUDA