ASHLEY SHIN

ashley.shin@nih.gov ◇ ashleyhshin@gmail.com ◇ www.ashleyshin.org

Areas of Interest: information retrieval, bioNLP, scientific document processing, representation learning, ML, HCI

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

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

- Worked with Dr. Qiao Jin in Dr. Zhiyong Lu's bioNLP lab, NCBI/NLM/NIH
- Research spanning biomedical natural language processing (bioNLP), information retrieval, machine learning
- Research aimed at improving PubMed, biomedical search engine used by 7 million researchers

PUBLICATIONS

[1] **Ashley Shin**, Qiao Jin, Zhiyong Lu. Highlighting Relevant Tokens in Article Title with Encoders Trained on PubMed Co-Click Logs. *Preprint ready for submission.*

[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 (SIG 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]

HONORS

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

Top 3, 11th 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.

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

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

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]

Sparse Retriever Baseline Implemented BM25 for sparse retrieval baseline as part of LADER ablation study. 35M PubMed articles indexed/searched. *Pyserini/Lucene, Numpy, Pandas.* Led to [3]

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*

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

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