ASHLEY SHIN

ashleyhshin@gmail.com ⋄ www.ashleyshin.org ⋄ ashleyshin@ucsd.edu

interests: information retrieval, recommender systems, NLP, representation learning, AI for medicine, HCI, ethics

EDUCATION

PhD in Computer Science University of California, San Diego 2024 – present
Advisor: Julian McAuley

EXPERIENCE

Research Fellow National Library of Medicine, National Institutes of Health (NIH) 2022 – 2024
Advisors: Qiao Jin, Zhiyong Lu
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.

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

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]

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. Models trained with 255M query-article pairs, constructed from PubMed user search logs. *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]

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*

SKILLS

Languages Python, Java, C++
Libraries Pytorch, Hugging Face Transformers
Tools Git, CUDA

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