Zhaoyi Sun

850 Republican St, Seattle, WA 98109 │ 332-209-6062 │ zhaoyis@uw.edu

EDUCATION

Univeristy of Washington   **Seattle, WA**
Doctor of Philosophy in Biomedical and Health Informatics | GPA: **3.85/4.0**   *June 2028 (Expected)*
Weill Cornell Medical College, Cornell University   **New York, NY**
Master of Science in Health Informatics | GPA: **4.0/4.0**   *July 2022*
Nanjing University   **Nanjing, China**
Bachelor of Science in Chemistry | GPA: **3.62/4.0**   *June 2019*

RESEARCH EXPERIENCE

University of Washington   **Seattle, WA**
Research Assistant   *September 2023 – present*

- Under the guidance of Prof. Meliha Yetisgen, focused on applying NLP and LLMs to solve clinical problems
- Collaborated with Microsoft on creating the benchmark datasets for MediQA shared tasks
- Working on creating a database of radiology incidental findings including related concepts, image follow-up status, and associated cancer related outcomes
- Preparing a review on medical misinformation, and a multimodal multilingual medical misinformation dataset

Weill Cornell Medical College, Cornell University   **New York, NY**
Research Assistant   *July 2022 – August 2023*

- Under the guidance of Prof. Yifan Peng, focused on developing and applying computational approaches to biomedical text data and medical images, and evaluating large language models on clinical tasks
- Evaluated the performance of large language models on medical evidence summarization (GPT-3.5 and GPT-4) and impressions generation of radiology reports (GPT-4)
- Developed NLP pipelines to identify online health education materials to prevent postpartum depression using feature-based methods (TF-IDF and Random Forest), deep-learning-based methods (BERT), and ChatGPT
- Adapted a suicide-specific SDoH ontology and used NLP to identify individual-level SDoH-related social risks from death investigation narratives
- Published a scoping review on multi-modality joint learning in biomedical images and texts

Graduate Research Assistant   *December 2021 – July 2022*

- Under the guidance of Prof. Thomas Campion, focused on extracting structured and unstructured information from EHRs to support cancer clinical trials
- Conducted manual annotation of patients' ECOG/KPS scores and BRAF/KRAS mutation status in clinical notes in WCM's EHR system using REDcap
- Developed NLP pipelines using Python to automate data extraction from over 300 physician notes and genomic test reports for candidate identification for cancer clinical trials

Capstone Project Manager   *October 2021 – July 2022*

- Designed algorithms using Python for multiple-label classification of biomedical texts concerning hallmarks of cancer with deep learning
- Performed statistical analyses using R for predictors of skeletal-related events among bladder cancer patients with bone metastases
- Investigated the impact of built environments and social determinants on people's health status in NYC using SQL, Access, and Python
- Explored the contribution of imaging standards (DICOM and PACS) to healthcare interoperability and computer-aided diagnosis

Institute of Drug R&D, Medical School, Nanjing University   **Nanjing, China**
Research Assistant   *July 2017 – January 2020*

- Under the guidance of Prof. Yiqiao Hu and Prof. Jinhui Wu, focused on the combination of radiotherapy and immunotherapy for treating tumors and smart drug delivery systems
- Investigated the roles of DNA damage and STING pathway in tumor immunotherapy
- Conducted cell culture, animal experiments, organic synthesis, and immune characterization (ELISA and Flow Cytometer)
- Processed animal data for visualization using GraphPad as well as performed medical statistics analyses
- Published several papers in SCI core journals as the first author and co-author, one of them has been cited nearly 200 times in the last two years

Institute of Computing Technology, Chinese Academy of Sciences   **Beijing, China**
Data Analyst Internship   *October 2018*
- Constructed a peptide section search engine based on MATLAB and mass spectra
- Led a team to review literature about bioinformatics, genomics, and proteomics and prepared a systematic literature review

School of Chemistry and Chemical Engineering, Nanjing University   **Nanjing, China**
Research Assistant   *March 2017 – July 2017*
- Explored the methodology of gold-catalyzed organic synthesis
- Worked with Prof. Shaozhong Wang, prepared a series of alkynol compounds, and characterized them by ^{1}H NMR and infrared spectroscopy (IR)

TEACHING AND PROFESSIONAL EXPERIENCE

Weill Cornell Medical College, Cornell University   **New York, NY**
Teaching Assistant   *December 2022 – April 2023*
- Worked as a teaching assistant for the Natural Language Processing course, performed duties including mentoring, lecturing, researching, and grading for 30+ students

Nanjing Guozhun Data Co., Ltd   **Nanjing, China**
Data Analyst   *September 2020 – August 2021*
- Processed over 750+ data analysis orders from individuals, government, and research institutions using R, SQL, and Tableau
- Provided R programming tutoring for 22 college students
- Prepared 16 reports concerning COVID-19 data visualization in R and prediction of confirmed cases in the US and India

HONORS AND AWARDS

- **Certificate of Academic Excellence**, Weill Cornell Graduate School of Medical Sciences   *July 2022*
- **The People's Scholarship in China,** Nanjing University   *November 2018*
- **The People's Scholarship in China**, Nanjing University   *November 2017*

PUBLICATIONS

Journal Articles
- Lin M, Li T, **Sun Z**, *et al*. Improving Fairness of Automated Chest X-ray Diagnosis by Contrastive Learning. *Radiology: Artificial Intelligence*, *6*(5), e230342
- Patra BG, **Sun Z**, *et al*. Automated classification of lay health articles using natural language processing: a case study on pregnancy health and postpartum depression. *Frontiers in Psychiatry* 2023, 14, 1258887
- **Sun Z**, *et al*. A survey on multimodal deep learning in biomedical images and texts. *Journal of Biomedical Informatics* 2023, 104482
- Tang L, **Sun Z**, *et al*. Evaluating large language models on medical evidence summarization. *NPJ Digital Medicine* 2023, 6 (1), 158
- Wang S, Dang Y, **Sun Z** *et al*. An NLP Approach to Identify SDoH-related Circumstance and Suicide Crisis from Death Investigation Narratives. *Journal of the American Medical Informatics Association* 2023, 30 (8), 1408-1417
- **Sun Z**, *et al*. Evaluating GPT-4 on Impressions Generation in Radiology Reports. *Radiology* 2023, 307 (5), e231259

- Wang C, **Sun Z**, *et al.* Maintaining manganese in tumor to activate cGAS-STING pathway evokes a robust abscopal anti-tumor effect. *Journal of Controlled Release* 2021, 331, 480-490
- Wang C, Dong H, Liu H, **Sun Z**, *et al.* Using HPLC to analyze (S)-oxiracetam and four related substances in the bulk drug of (S)-oxiracetam. *Journal of Pharmaceutical and Biomedical Analysis* 2020, 180, 113072
- **Sun Z**, *et al.* Hydrogel-based controlled drug delivery for cancer treatment: a review. *Molecular Pharmaceutics* 2020, 17 (2), 373-391

Conference Papers

- Yim WW, Fu Y, **Sun Z**, *et al.* Dermavqa: A multilingual visual question answering dataset for dermatology. *In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 2024 (pp. 209-219)
- Yim WW, Abacha AB, Fu Y, **Sun Z**, *et al.* Overview of the MEDIQA-M3G 2024 Shared Task on Multilingual Multimodal Medical Answer Generation. *In Proceedings of the 6th Clinical Natural Language Processing Workshop* 2024 (pp. 581-589)
- Abacha AB, Yim WW, Fu Y, **Sun Z**, *et al.* Overview of the mediqa-corr 2024 shared task on medical error detection and correction. *In Proceedings of the 6th Clinical Natural Language Processing Workshop* 2024 (pp. 596-603)
- Yim WW, Abacha AB, Fu Y, **Sun Z**, *et al.* Overview of the MEDIQA-magic task at ImageCLEF 2024: multimodal and generative telemedicine in dermatology. *CLEF* 2024

Posters and Presentations

- **Sun Z**, *et al.* An Ontology-Guided Analysis to Unveil Disparities of Social History Documentation in Discharge Notes. *AMIA 2024 Annual Symposium,* San Francisco, 2024
- **Sun Z**, *et al.* An Enhanced Multimodal Multilingual Dataset for Medical Misinformation Detection. *The 12th IEEE International Conference on Healthcare Informatics,* Orlando, 2024
- **Sun Z**, *et al.* Innovative Analysis of Social Determinants to Enhance Suicide Prevention for Black Youth. *AMIA 2023 Annual Symposium,* New Orleans, 2023

SERVICES

- **Journal Reviewers:** Journal of Biomedical Informatics (JBI), Journal of Medical Internet Research (JMIR)
- **Conference Reviewers:** American Medical Informatics Association (AMIA), IEEE International Conference on Healthcare Informatics 2025 (IEEE ICHI 2025)
- **Membership:** Member of AMIA, Member of IEEE

SPECIALIZED SKILLS AND CERTIFICATIONS

- **Technical Skills:** Python, R, SQL, Tableau, MATLAB, GraphPad, Origin, & MS Office (Access, Excel, Outlook, PowerPoint, Word)
- **Research Skills:** Natural Language Processing, Large Language Models, Machine Learning, Data Analysis, Data Visualization, Literature Reviews, Statistical Analysis
- **Licenses & Certifications:** Google Data Analytics Specialization, IBM Data Science Specialization