Wenyu Zeng

Shoreline, WA, 98155 | wz34@uw.edu | 412.692.0641 | Github | LinkedIn

EDUCATION

University of Washington, Seattle, WA 09/2023 – 05/2028
Ph.D. student, Biomedical and Health Informatics

George Washington University, Washington, DC 08/2019 - 05/2021
Master of Science, Data Science

University of Pittsburgh, Pittsburgh, PA 08/2015 - 04/2019
Bachelor of Science, Statistics | Minor in Religious Studies and Chemistry

RESEARCH EXPERIENCE

UW BioRepository and Integrated Neuropathology (BRAIN) Lab 03/2024 - Present
Department of Lab Medicine and Pathology, University of Washington, Seattle, WA
Research Assistant
Supervisors: Jimmy Phuong, MPH, Ph.D., C. Dirk Keene, M.D., Ph.D.

- Develop robust, reproducible, and scalable computational pipelines to analyze spatial transcriptomic data.
- Integrate spatial transcriptomics and single-cell RNA sequencing data to study glioblastoma's heterogeneity.
- Analyze 90GB of large-scale data using high-performance computing clusters and cloud services.

National Center for Advancing Translational Sciences (NCATS)
National Institutes of Health (NIH), Rockville, MD
Post-Baccalaureate Intramural Research Training Award Research Fellow 12/2021 – 09/2023
Supervisor: Ewy Mathé, Ph.D.

- Built numerous machine learning models to predict COVID-19 severity levels and chemical toxicity.
- Developed optimal machine learning models in an automation pipeline for ADME (Absorption, Distribution, Metabolism and Excretion) projects.
- Assisted lab mates in validating and troubleshooting developed R packages.
- Presented findings to teammates and institute-wide meetings to discuss research efforts.

Graduate Data Science Summer Program Intern at Informatics Core 06/2021 - 08/2021
Mentor: Dac-Trung Nguyen, M.S.

- Implemented Naïve Bayesian Classifier using features' kernel density functions.
- Modified model results for continuous data based on previous discrete data.
- Developed charts, tables and figures in Python to measure the distance between data distributions.
- Discussed research effort with mentor and presented results to institute-wide meetings.

US-China Health Summit, Boston, MA
Volunteer & Summer intern 02/2020 - 09/2020
Supervisor: Jing Ma, Ph.D.

- Built compartmental models to analyze number of COVID-19 cases and predict COVID-19 trends.
- Scraped online news articles using chrome network and applied sentiment analysis for prediction.
- Collected and organized countries online health policy data and presented the results to 40+ person audience.
- Interpreted and presented data and results to lab mates and team leaders to discuss research effort.

WORK EXPERIENCE

George Washington University, Washington, DC
Graduate Teaching Assistant 09/2020 - 05/2021

- Provided technical support to data warehousing graduate students for SQL and NoSQL databases' labs.
- Explained time series modeling algorithms to graduate students, held office hours, and graded assignments.

Western Cloud Data Technology Co., Ltd, Beijing, China
AWS Strategic Partner Intern 05/2019 - 08/2019

- Prepared recruiting reports for the HR Team by analyzing hiring needs across the organization using R & Excel.
- Coordinated interview process and collaboratively assisted contract preparations for statistical analysis.

University of Pittsburgh, Pittsburgh, PA
Computing Lab Monitor, Computing Services and Systems Development (CSSD) 01/2018 - 04/2019
- Provided solutions for customers by troubleshooting Microsoft-related issues.

PRESENTATIONS

3MT Competition, ITHS, Reveal Glioblastoma Heterogeneity through Spatial Transcriptomics, October 2025

AMIA Annual Symposium, An Ontology-Guided Analysis to Unveil Disparities of Social History Documentation in Discharge Notes, November 2024

AMIA Informatics Summit, Predicting Chemical Toxicity using a Hierarchical Bayesian Approach with Tox21 Assay Data, March 2023 (Also presented at NCATS Informatics Retreat, Oct 2022)

Postbac Poster Day, NIH, Predicting COVID-19 Severity Levels Using Pre-COVID Metabolomic and Clinical Data, April 2022 (Also presented at NCATS Poster Day)

Three Minute Talk (TMT) Competition, NCATS, How to Reduce Data Dimension in Predictive Models, April 2022

NIH Summer Research Presentation, NIH, Predicting Bioassay Activity Outcome: Bayesian Modeling and Probabilities Programming, August 2021 (Also presented at NCATS Summer Research Symposium, July 2021)

PUBLICATION

*Equal contribution
Li, L. X., Chung, R., Chen, F., **Zeng, W.,** Jeon, Y., & Zaslavsky, O. (2025, April 28). *Learning from elders: Making an LLM-powered chatbot for retirement communities more accessible through user-centered design*. arXiv.org. https://doi.org/10.48550/arXiv.2504.08985

Wenyu Zeng, Arjun Yadaw, Jaleal Sanjak, Dac-Trung Nguyen, Ruili Huang, Ewy Mathé, *Predicting Chemical Toxicity by Applying a Hierarchical Bayesian Approach with Priors to the Tox21 Assay Data* (Manuscript under preparation)

Kato, R.*, **Zeng, W**.*, Siramshetty, V. B., Williams, J., Kabir, M., Hagen, N., Padilha, E. C., Wang, A. Q., Mathé, E. A., Xu, X., & Shah, P. (2023). *Development and validation of PAMPA-BBB QSAR model to predict brain penetration potential of novel drug candidates.* Frontiers in pharmacology, 14, 1291246. https://doi.org/10.3389/fphar.2023.1291246

FELLOWSHIPS & AWARDS

- ITHS TL1 Translational Research Fellowship Training Program (NIH-funded) 2025 – 2026
  - Year-long research training fellowship supporting translational biomedical and health informatics research.
- NCATS Post-Baccalaureate Poster Day Outstanding Poster Award 04/2021
- Post-Baccalaureate Intramural Research Training Award 2021 – 2022

SKILLS

Programming & Data Analysis: Python (pandas, scikit-learn, PyTorch, scaleSC), R (Seurat, ggplot2, tidyverse), SQL
Bioinformatics: Spatial and single-cell transcriptomics (Seurat, Scanpy, Harmony, WGCNA, GSEA)
Tools & Visualization: Jupyter, Git, Linux, Vega-Lite, matplotlib, Tableau, High Performance Computing