
EDUCATION**Harvard University, T.H. Chan School of Public Health**

Boston, MA

*Master of Science in Health Data Science; GPA: 4.00**Aug. 2022 – May 2024 (Expected)*

Relevant courses: Data science programming in R, Data science programming in Python, Machine learning, Statistical inference, Artificial intelligence (MIT), Computer system and network security (MIT), Efficient Deep Learning (MIT), Large scale symbolic systems (MIT)

University of North Carolina at Chapel Hill

Chapel Hill, NC

Bachelor of Science in Computer Science (Highest Honor) & Bachelor of Science in Physics; GPA: 3.99 Aug. 2018 – May 2022

Relevant courses: Algorithms, Computer organization, Databases, Data structures, Distributed systems, Object-oriented programming

PROGRAMMING SKILLS

- **Languages:** Python, Java, MySQL, R, Scheme, C++, JavaScript, HTML, CSS, Bash, Prolog, SML, Lisp, TypeScript, C
- **Technologies:** Jupyter Notebook, Keras, TensorFlow, Pytorch, Hugging Face, LLMs, Scikit-Learn, RStudio, Git, Stata, Latex, Docker

RELEVANT EXPERIENCE**Mayo Clinic, Department of Otolaryngology - ENT- Head and Neck Surgery**

Rochester, MN

Data Science Intern

Project: Automating Data Abstraction for the Generation of Clinical Registries: *Jun. 2023- Present*

- Engineered a natural language processing (NLP) pipeline using large language models (LLMs) to extract information from unstructured clinical notes to automate the clinical registries generation process.
- Formulated the LLM prompt, built an autograder to evaluate the pipeline performance, and gave a presentation on the phased results on the Harvard DSI South Africa Training Program monthly seminar in collaboration with two data scientists at Mayo Clinic.

Project: Automatic Decision System for Rhinology and Otology Appointments: *Jun. 2023- Present*

- Built an automatic decision system to automate the approval of the Mayo ENT Rhinology and Otology appointments based on historical patient symptom descriptions on Qualtrics and historical physician approval data on Triage.
- Designed a system that was projected to automate 63% and 25% of the Rhinology and Otology appointments, collaborated with two data scientists at Mayo Clinic, and presented the outcomes to the chief physicians of the Rhinology and Otology division.

Dana-Farber Cancer Institute, Department of Data Science

Boston, MA

Data Science Research Assistant

Project: Annotation of ECOG PS from Unstructured Oncology Notes and Survivability Analysis: *Jul. 2023- Present*

- Identified performance status (PS) labels in unstructured clinical notes using a text-based search, trained a CNN model and a transformer-based model to predict the ECOG PS, and evaluated the correlation between ECOG PS and survival outcomes.
- Developed a model with 95.5% accuracy, found strong correlation between ECOG PS and survival outcomes, worked with two medical oncologists and one data scientist at Dana-Farber, and drafted a research paper.

Brigham and Women's Hospital, Department of Medicine

Boston, MA

Data Science Research Trainee

Project: Extract Patient Entities from Free-Text EHR Data Using NLP Models: *Aug. 2023- Present*

- Constructed a natural language processing (NLP) pipeline by implementing transformer-based and large language models to extract social determinants of health (SDOH) variables from unstructured text.
- Worked under the guidance of one epidemiologist and one data scientist at Brigham, and presented during the weekly group meetings.

Harvard University, T.H. Chan School of Public Health

Boston, MA

Graduate Student

Project: Chest X-ray Classification for Pneumonia: *Feb. 2023-May. 2023*

- Created KNN, XGBoost, and CNN models that predict whether patients have pneumonia based on X-ray images of their lungs.
- Achieved accuracies of 72%, 74%, and 90% on the test data set for the KNN, XGBoost, and CNN models respectively, collaborated with four Harvard health data science students, composed a project report, and delivered a final project presentation.

PUBLICATIONS

- **Bowen Gu**, Hao Wang, and Kaizhuo Chen, “Towards a Comprehensive AI Teaching Assistant Based on Course Forum”, Published on Carolina Digital Repository.
- Prasun Dewan, Samuel George, **Bowen Gu**, Zhizhou Liu, Hao Wang, and Andrew Wortas, “Broad Awareness of Unseen Work on a Concurrency-based Assignment”, Published on Workshop on Education for High Performance Computing.
- Prasun Dewan, Andrew Wortas, Ken Liu, Sam George, **Bowen Gu**, and Hao Wang, “Automating Testing of Visual Observed Concurrency”, Published on Workshop on Education for High Performance Computing.