

ZIHAN XU

+1-551-244-0066 | zix4004@med.cornell.edu | stephaniexuuu.github.io

New York, New York State, United States

EDUCATION

• Weill Cornell Medicine

Master of Science in Health Informatics

09/2023 - 08/2024

New York, United States

- GPA: 4.11 / 4.00
- Core Course: Research Methods and Visualisation, Healthcare Organization and System, Natural Language Processing in Medicine, Data Management in Medicine, Clinical Informatics, Biostatistics, Health Info Standards & Interoperability
- Thesis: A qualitative study of participants' experiences with Engage & Connect Psychotherapy for Postpartum Depression through interviews

• University of Melbourne

Bachelor of Science in Computational Biology

07/2020 - 07/2023

Melbourne, Australia

- Core Course: Calculus, Linear Algebra, Probability, Statistics, Data Management, Machine Learning, Biostatistics, Mathematical Biology, Computer Science, Genetics and Genomics, Microbiology and Immunology, General Biology

PUBLICATIONS

- [1] B. Idnay, Z. Xu, et al. (2024). **Environment Scan of Generative AI Infrastructure for Clinical and Translational Science**. Manuscript submitted for publication in *NEJM Catalyst Innovations in Care Delivery*.
- [2] Z. Wang*, Z. Xu*, et al. (2024). **ReDose: Benchmark for Drug related entities extraction from Reddit**. Manuscript submitted for publication in *Scientific Data*.
- [3] Z. Xu*, M. Lin*, Y. Zhou, Z. Xu et al. (2024). **Establishing dermatopathology encyclopedia DermPathNet with Artificial Intelligence Based Workflow**. Manuscript submitted for publication in *JAMA Dermatology*.
- [4] G. Zhang, Z. Xu, et al. (2024). **A MapReduce Approach to Effectively Utilize Long Context Information in Retrieval Augmented Language Models**. Manuscript submitted for publication in *npj*.
- [5] S. Cheng, Y. Wei, Y. Zhou, Z. Xu et al. (2024). **Deciphering genomic codes using LLMs: a scoping review**. Manuscript submitted for publication in *JAMIA*.
- [6] Z. Bai, Z. Xu et al. (2024). **A hybrid Natural Language Processing pipeline for extracting long COVID symptoms from RECOVER initiative clinical notes**. Manuscript submitted for publication in *npj*.

RESEARCH EXPERIENCE

• Weill Cornell Medicine

Senior Research Assistant

Nov 2023 - Present

New York, United States

- Employed advanced natural language processing techniques to analyze complex medical issues, such as post-acute sequelae of COVID-19 and the impact of drug usage using data extracted from social media, improving understanding of treatment outcomes and patient experiences
- Proficiently utilized machine learning and NLP libraries including TensorFlow, PyTorch, NLTK, and spaCy for text preprocessing and model development, enhancing the predictive capabilities of healthcare models.
- Engineered and launched a clinical information website using HTML, JavaScript, and CSS, significantly improving patient engagement and access to interactive health tools and resources
- Collaborated extensively with a team of researchers to draft and submit papers to prestigious journals and conferences such as AMIA, contributing to the advancement of knowledge in the application of NLP in healthcare

PROJECTS

• Natural Language Processing for Evidence-based Medicine: A Systematic Review

Aug 2024 - Present

Tools: Covidence, IEEE, ACM, ACL, PubMed, drawio

- Conducted a systematic review of existing literature, analyzing over 600 articles to identify how NLP techniques are applied in evidence-based medicine and assessing their impact on improving the efficiency and accuracy of medical research and clinical practice
- Utilized PubMed, ACM, IEEE and ACL for retrieving technical papers, and Covidence for review processes, data extraction and comprehensive medical article retrieval
- Collaborated with a team of researchers to summarize findings and discussed potential NLP strategies that could address current gaps in the use of private electronic health records and patient data
- Prepared a manuscript detailing the review outcomes and proposed future directions for integrating NLP more effectively into clinical decision support systems

- **RECOVER: Hybrid NLP Framework for PASC-related Symptoms Extraction**

Mar 2024 - Present

Tools: Python, BERT Models, AWS Sagemaker, GraphPad Prism

- Developed a hybrid NLP framework that combines rule-based and machine learning modules to extract and analyze symptoms related to Post-Acute Sequelae of COVID-19 (PASC) from electronic health records (EHRs)
- Employed Medtext for comprehensive clinical text analysis, including section segmentation and named entity recognition, which significantly improved the system's capability to process diverse clinical narratives and extract relevant symptom data
- Fine-tuned multiple BERT models, including BioBERT and ClinicalBERT, on a unique dataset of clinical notes, resulting in enhanced model performance for assertion detection and increased accuracy in identifying PASC-related symptoms
- Conducted large-scale data analysis across multiple sites using optimized models, which included extensive benchmarking and validation of the models' effectiveness

- **CTSA Survey: Assessing LLM Adoption in Healthcare**

Jul 2024 - Present

Tools: Python, NVivo, Qualtrics



- Led a comprehensive survey across 36 CTSA institutions to evaluate the integration and effectiveness of LLMs in healthcare, focusing on governance, ethics, and technology enhancement strategies.
- Analyzed data using quantitative methods for multiple-choice and Likert-scale questions, and performed thematic analysis on qualitative responses to open-ended questions.
- Synthesized survey findings into strategic recommendations for the deployment and governance of GenAI in healthcare institutions, highlighting key areas such as stakeholder involvement and ethical considerations
- Created detailed visualizations of survey results, facilitating clear communication of complex data and insights to stakeholders and decision-makers

- **ReDose: A Transformer-Based NLP System for Extracting Drug Dosages and Effects**

May 2024 - Jul 2024

Tools: Python, Transformer, PyTorch, HuggingFace



- Developed a transformer-based Named Entity Recognition (NER) model to extract drug, dosage, and effect entities from unstructured medical text, achieving an F1-score of 0.82 for DRUG entities
- Implemented a custom BERT-based architecture with CRF and BiLSTM layers to improve entity extraction performance, processing over 6,000 clinical notes
- Experimented with large language models such as Llama and GPT to assess their potential in enhancing the extraction process and improving overall model performance
- Conducted additional processing to comply with Reddit's policy, which allowed only the publication of annotated data, requiring careful handling of the dataset

- **DermpathNet: Open-access Dermatopathology Database**

Apr 2024 - June 2024

Tools: HTML, JavaScript, CSS



- Engineered the frontend architecture for DermpathNet, an educational and research platform providing access to an extensive dermatopathology image database
- Implemented interactive features using HTML, CSS, and JavaScript to facilitate user-friendly navigation and efficient access to over 7,700 dermatopathology images
- Developed a dynamic search functionality that categorizes images based on diagnostic ontologies, enhancing the utility for educational and diagnostic purposes
- Collaborated closely with AI specialists and dermatopathologists to integrate AI-powered tools for image analysis, ensuring the website's alignment with clinical standards

TEACHING EXPERIENCE

- **Teaching Associate**

Sep 2024 - Present

Weill Cornell Graduate School of Medical Sciences

- Assisted in teaching *Introduction to Biostatistics* with over 60 students, focusing on core statistical concepts and their practical applications.
- Led the Stata labs, guiding students through hands-on exercises, helping them apply statistical methods using Stata for data analysis.
- Provided ongoing support to students by answering questions during lectures and labs, as well as holding weekly office hours to address individual concerns, clarify course materials, and provide further assistance with assignments and software usage.
- Graded assignments and provided detailed feedback on homework, helping students improve their understanding of topics such as hypothesis testing, regression, and data visualization.

PROFESSIONAL EXPERIENCE

• Frost & Sullivan

Dec 2021 - Apr 2022

Consulting Analytics Intern

Shanghai, China

- Deployed cutting-edge data analytics to conduct marketing research and analysis on competitive products
- Researched mainly on cancer study and the effective drugs of monoclonal and polyclonal antibodies
- Supported the team in writing the data analysis part of the prospectus and the clinical trials studies
- Collected and processed big data from databases and publications with computational tools, extracted content from sources and journals, and organized them into reports
- Performed epidemiology analysis on different indications, including research on their prevalence and incidence
- Successfully listed KELUN Biotech on the Hong Kong stock exchange

HONORS AND AWARDS

• Academic Excellence

July 2024

Weill Cornell Graduate School of Medical Sciences

- Recognized for outstanding academic achievement in the field of Health Informatics, demonstrating exceptional proficiency in coursework and research.
- Awarded based on rigorous criteria including academic performance, innovative research contributions, and the potential for significant impact in medical science.

• Melbourne Global Scholar Award

Aug 2022

University of Melbourne

- Recognized the undertaken overseas study as part of the undergraduate coursework.
- Awarded based on academic merits.

• Melbourne Plus

Oct 2023

Awarding Institution/Organization

- Recognized the capabilities as a people leader, community engager, sustainability advocator, and innovator
- Connected with a wider network of professionals and organizations, leading to collaborative opportunities

• Leaders in Community Awards

Dec 2021

Awarding Institution/Organization

- Recognized the undertaken community volunteering, increasing the university co-curricular and extra-curricular participation, and developed their employability and leadership skills
- Satisfied 40 hours volunteer ventures, 20 hours of university action and professional development

SKILLS

- **Technical Skills:** Programming (Python, R, Java, C, Stata, SAS, SQL), Web Development (HTML, CSS, JavaScript), Databases (MySQL, MSSQL, Neo4j)
- **Data Science & Analytics:** Machine Learning (TensorFlow, scikit-learn, Pandas, NumPy, SciPy), Data Visualization (Matlab, Power BI, Tableau, GraphPad Prism), Statistical Tools (SPSS, R packages, Excel)
- **Tools & Technologies:** DevOps (Docker, Git, AWS, Azure), FHIR, Epic, LaTeX, NVivo, Qualtrics
- **Specialized Areas:** Predictive Modeling, Natural Language Processing (NLP), Artificial Intelligence, Statistical Analysis, Quantitative & Qualitative Research

VOLUNTEER EXPERIENCE

• AMIA Student Volunteer

July 2024 - present

American Medical Informatics Association

- Served as a Student Co-Lead for the NLP workgroup in the Year-in-Review, directing the literature review process, engaging with authors, and synthesizing findings for annual reports.
- Presented on the Closing Keynotes with Dr. James Cimino
- Acted as a Panelist in the First Look Panel, discussing emerging trends and providing insights into the future of health informatics.
- Coordinated session logistics at the AMIA 2024 symposium, ensuring seamless operations and enhancing the experience for attendees.
- Mentored a high school student in health informatics, significantly contributing to their understanding and enthusiasm for the field.

• Volunteer Fundraiser

June 2021 - July 2021

United Nations

- Led an online fundraising campaign for Malala Sports Foundation, supported by the United Nations, to address the Covid-19 emergency impacts on African children
- Coordinated efforts to improve hygiene, medical support, and nutrition for children severely affected by the pandemic
- Promoted campaign initiatives across digital platforms, enhancing awareness and contributions to meet critical needs

PROFESSIONAL MEMBERSHIPS

- **American Medical Informatics Association** Student Membership *August 2024 - Present*
- **Melbourne International Relations Society** Marketing Director *Jan 2022 - Dec 2022*
- **Women in Technology Unimelb** Events Team Member *Jan 2022 - Dec 2022*

REFERENCES

1. **Yifan Peng, PhD, FACMI**
Associate Professor, Population Health Sciences
Weill Cornell Medicine
Email: yip4002@med.cornell.edu
Relationship: Course instructor, Supervisor
2. **Arindam RoyChoudhury, PhD**
Associate Professor, Population Health Sciences
Weill Cornell Medicine
Email: arr2014@med.cornell.edu
Relationship: Course instructor, TA supervisor
3. **Jose F Florez-Arango, MD, PhD**
Associate Professor, Clinical Population Health Sciences (Pending Appointment at Rank)
Weill Cornell Medicine
Email: jff4001@med.cornell.edu
Relationship: Professor