Haotian(Matthew) Ma

+1-206-696-6842 | matthewma2022@gmail.com
LinkedIn | Google Scholar | Personal Website
Long Island City, New York - 11101, United States

Education

- **Columbia University** *September 2022 - May 2024*
  Master of Science in Biostatistics New York, NY
  - GPA: 3.72/4.0
- **University of Washington** *September 2018 - June 2022*
  BA in Biochemistry, BS in Psychology Seattle, WA
  - GPA: 3.89/4.0

Research Experience

- **Deepcell: Cell Segmentation Model-building Project** *June 2025 - Present*
  Research Intern, Virginia Commonwealth University, Supervised by Dr. Jinze Liu Richmond, VA
  - Reproduced Cellpose 4.0 on internal multi-channel microscopy data; built a reproducible pipeline for tiling, channel selection/normalization, and post-processing
  - Defined an evaluation set and metrics; established a tuned cellpose-SAM baseline
  - Curated training data and implemented augmentation and class-imbalance handling
  - Fine-tuned a SAM-assisted segmentation model and compared against the off-the-shelf cellpose-SAM
- **Decoding Suicide Risk: A Computational Analysis of Youth Decedent Profiles** *March 2025 - Present*
  Senior Research Assistant, Weill Cornell Medicine, Supervised by Dr. Yifan Peng New York, NY
  - Applied unsupervised topic modeling (Latent Dirichlet Allocation) to identify and compare the latent factors of suicide "circumstances" and "crisis" between youth and the general decedent population
  - Demonstrated that while core risk factors are stable across age groups, youth suicide profiles are uniquely characterized by a higher prevalence of childhood trauma, school-related stress, and acute intimate partner conflicts
- **Meta-Research on Evidence of Variant (Re)Classification** *December 2024 - March 2025*
  Research Assistant, Weill Cornell Medicine, Supervised by Dr. Yifan Peng and Dr. Chunhua Weng New York, NY
  - Employed NLP techniques to analyze variant re-classifications such as breast cancer variants BRCA
  - 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
  - Processed and analyzed large-scale datasets on sources such as ClinVar and PubMed
  - Discovered research gap between evolving evidence for variant (re)classification and lacking standards
- **Natural Language Processing in Support of Evidence-based Medicine** *September 2024 - January 2025*
  Research Assistant, Weill Cornell Medicine, Supervised by Dr. Yifan Peng and Dr. Chunhua Weng New York, NY
  - Systematically reviewed 129 studies to analyze the application of Natural Language Processing (NLP) across the core framework of Evidence-Based Medicine (EBM)
  - Analyzed the technical evolution of NLP for EBM, from early machine learning to modern LLMs, in applications like automated evidence extraction, synthesis, and summarization
  - Outlined critical challenges and future research directions, highlighting the need for specialized datasets and the potential of few-shot learning to advance clinical decision support
- **Master's Practicum Project: Comorbidity in Psychiatric Disorders in Children** *September 2023 - March 2024*
  Project Manager, Columbia University, Supervised by Dr. Seonjoo Lee New York, NY
  - Performed data-cleaning and exploratory data analysis on ABCD data, encompassing 11,000+ participants and 90 diagnoses
  - Categorized participants using cluster analysis techniques, including K-means and Latent Class Analysis; investigated cluster differences using various statistial methods
  - Implemented the Ising model and Lasso-regularized logistic regression to estimate diagnoses
  - Performed community detection algorithms and clique percolation methods for graph structure learning

Teaching Experience

- **Graduate Teaching Assistant (HINF 5016)** *September 2024 - March 2025*
  Weill Cornell Graduate School of Medical Sciences Supervised by Dr. Yifan Peng New York, NY
  - Assisted in teaching Introduction to LLMs, NLP, and Health Informatics for over 60 students
  - Led weekly Python labs; held weekly office; Assessed assignments, quizzes, and projects

- **Graduate Teaching Assistant (SYSEN 5630)** *September 2024 - March 2025*
  Cornell University Graduate School Supervised by Dr. Yifan Peng Ithaca, NY
  - Coordinated course operations (Zoom scheduling, announcements, Canvas recordings/materials)
  - Evaluated homework, and projects by providing structured feedbacks; held virtual office hours

- **Graduate Teaching Assistant (CTIV 5057)** *September 2024 - March 2025*
  Weill Cornell Graduate School of Medical Sciences Supervised by Dr. Yifan Peng New York, NY
  - Managed class attendance in a weekly bases; graded homework and group projects
  - Held in-person final presentation session; provided on-going online support for question answering

- **Teaching Assistant (PSYCH 318)** *January 2022 - March 2022*
  University of Washington Supervised by Dr. Laura Little Seattle, WA
  - Led weekly in-person active-learning sections; prepared and graded worksheets
  - Held office hours and monitored forum QA; administered assignments and grading

Publications

C=Conference, J=Journal

[C.3] Zihan Xu, **Haotian Ma** (co-first), Gongbo Zhang, Yihao Ding, Chunhua Weng, Yifan Peng (2025). **Natural Language Processing in Support of Evidence-based Medicine: A Scoping Review**. In *Findings of the Association for Computational Linguistics: ACL 2025*. Association for Computational Linguistics. Jul 27–Aug 1, 2025, Vienna, Austria. (Findings, accepted/in press). Preprint DOI: 10.48550/arXiv.2505.22280

[C.2] **Haotian Ma**, Zihan Xu, Wendy Chung, Chunhua Weng, Yifan Peng (2025). **A Pilot Meta-Research on Evolving Evidence Behind Genetic Variant (Re)Classification**. In *Proceedings of MEDINFO 2025: 20th World Congress on Medical and Health Informatics* (*Studies in Health Technology and Informatics*, vol. 329), pp. 108–112. IOS Press. Aug 9–13, 2025, Taipei International Convention Center (TICC), Taipei, Taiwan. DOI: 10.3233/SHTI250811

[C.1] Max Lovitt, **Haotian Ma**, Song Wang, Yifan Peng (2024). **Suicide Risk Assessment on Social Media with Semi-Supervised Learning**. In *Proceedings of the 2024 IEEE International Conference on Big Data (BigData 2024)*, pp. 8541–8549. IEEE. Dec 15–18, 2024, Washington, DC, USA. DOI: 10.1109/bigdata62323.2024.10825422

[J.1] Song Wang, Yishu Wei, **Haotian Ma**, Max Lovitt, Kelly Deng, Yuan Meng, Zihan Xu, Jingze Zhang, Yunyu Xiao, Ying Ding, Xuhai Xu, Joydeep Ghosh, Yifan Peng (2025). **A Multi-Stage Large Language Model Framework for Extracting Suicide-Related Social Determinants of Health**. *Communications Medicine* (accepted, in press). Preprint DOI: 10.48550/arXiv.2508.05003.

Volunteer Experience

- **uClimb2.0 Python Learning Project** *June 2021 - March 2022*
  Research Volunteer, University of Washington, Supervised by Dr. Chantel Pratt Seattle, WA
  - Collected data from 100+ student subjects; organized and cleaned 3200+ raw data using Python
  - Conducted a series of statistical modelings including residual analysis, A/B Testing, and regression analysis
  - Collaborated with other teams on moderating the experimental design of the study

Skills

- **Programming Languages:** Python, R, Numpy, Pandas, Sklearn, Pytorch, Ten- sorflow, CSS
- **Data Analysis:** RStudio, SAS, MySQL, PostgreSQL, Tableau, Microsoft Office Suite

Professional Memberships

- **American Medical Informatics Association (AMIA)**, Member *July 2025 - Present*