Jixin "Jay" Li

617-230-7140 | li.jix@northeastern.edu | jixinli.info | linkedin.com/in/jixin-li-432b7a123 | google scholar

Career Objective

Ph.D. candidate in *Personal Health Informatics* with expertise in statistical analysis, predictive modeling, human-computer interaction, and innovative applications of generative AI. Seeking a **Data Scientist** or **User Experience Researcher** role to apply research-driven problem solving and data-informed insights in healthcare or mental health domains. Authorized to work in the U.S. under **OPT**.

Skills

Python, R, Java, Android Mobile App Development, Git, AWS, SQL, Shell (Bash), LaTeX, Microsoft Office, LLM Prompt Engineering, statistical machine learning, deep learning, time series analysis, statistical inference, survey methodology, quantitative and qualitative usability research

Work and Research Experience

Northeastern University, mHealth Research Group   Boston, MA

LLM-Powered AI Assistant for Statistical Analysis in Health Research   *Nov 2024 - present*
- Designed and implemented an LLM-based chatbot with a RAG pipeline to empower health researchers in understanding and applying complex statistical models through question answering, research question development, model configuration guidance, and results interpretation
- Conducted user testing and interviews with health researchers to assess the prototype tool's efficacy in supporting statistical analysis workflows

Adaptive-Timing Ecological Momentary Assessment (EMA) Using Mobile Sensing and Active Learning   *Aug 2024 - Present*
- Engineered 400 features from mobile sensing data such as app usage, location, and motion to detect momentary emotion at 5-minute intervals
- Implemented active learning with uncertainty quantification to identify the most informative moments for prompting self-reported surveys

Adaptive-Length EMA Surveys for Sustainable Longitudinal Assessment on Subjective Experiences   *Feb 2023 - Feb 2024*
- Applied Bayesian Networks to model momentary informativeness of survey items to enable adaptive-length EMA surveys in longitudinal studies
- Evaluated the proposed method on multiple EMA datasets, reducing average survey length by 34%-56% with minimal information loss

Contextual Biases in Microinteraction Ecological Momentary Assessment Non-response   *May 2021 - Mar 2022*
- Applied multi-level statistical models (in R with lme4) to analyze wearable sensor and EMA survey data, identifying key contextual factors influencing participant compliance and informing the design of context-sensitive EMA methods for improved long-term response rates
- Developed an end-to-end data pipeline utilizing parallel computing on a cluster and Pandas for efficient processing of large-scale system logs and sensor data (70,000 days from 300 participants), facilitating downstream survey response and health-related behavior analyses

Automated Semantic Enrichment of Personal Mobility Data using POI labels from OpenStreetMap   *Sep 2021 - Oct 2022*
- Created an open-sourced tool that automatically generate hierarchical taxonomies of OpenStreetMap POI data for visited place annotation
- Benchmarked performance of using popular geo-databases (e.g., Google Maps) to label frequently visited smartphone-detected locations

MixWILD: GUI-based Desktop Application for Intensive Longitudinal Data Analysis   *Sep 2019 - Present*
- Built a Java desktop GUI application to help health researchers configure statistics models for longitudinal data analysis without coding
- Performed in-depth usability testing with health researchers using combined qualitative and quantitative methods, including user log analysis, observation, questionnaires and interviews to understand user needs and inform iterative system improvements for user experience

Northeastern University, Khoury College of Computer Sciences   Boston, MA

Teaching Assistant   *Sep 2014 - Dec 2024*
- Designed and taught lectures for a 60-student senior-level machine learning course on deep learning for sequential data modeling
- Developed and presented tutorials on PyTorch and cluster computing to support student learning of deep learning and scalable model training

Learnable, Inc.   Boston, MA

Data Scientist   *Oct 2017 - Feb 2019*
- Enabled real-time pricing for transportation delay insurance by scraping weather data and predicting delay risk using machine learning models
- Built multi-level machine learning models to automate hierarchical categorization of exercises digitized from math and physics textbooks

Education

Northeastern University   Boston, MA

Ph.D., Personal Health Informatics, Khoury College of Computer Sciences   *Expected May 2026*

Columbia University   New York City, NY

M.A., Statistics   *Dec 2017*

University of Michigan   Ann Arbor, MI

B.A., Major in Psychology, Minor in Statistics | University Honor   *Apr 2014*

Publications and Presentation

Authored and co-authored 10 peer-reviewed publications (1 **Distinguished Paper Award**, 36 citations) and presented at 3 top-tier conferences