YEHU CHEN

(734) 846-0174 chenyehu@wustl.edu yahoochen97.github.io

SUMMARY

Ph.D. Data Scientist from WashU STL specializing in machine learning, quantitative analytics, and causal inference. Proven track record in predictive models, recommendation system, MLOps and large language model (LLM) that drive actionable business insights. Skilled in experimental design, data visualization, and cross-functional collaboration, with research published in top-tier conferences/ journals. **Actively looking for full-time opportunities in data science in Summer/Fall 2025.**

EDUCATION

- **Ph.D. in Data Science**, Washington University in St Louis. GPA: 3.9/4.0. 2019 - Present
  Research topics: Bayesian Machine Learning, Gaussian Process, Quantitative methods, Causal Inference
- **Bachelor of Science in Computer Science**, University of Michigan. Summa Cum Laude. 2017 - 2019

WORK EXPERIENCE

- **Data Science Research Assistant, Washington University, St Louis, MO** Fall 2019 - now
- **Software Engineer Intern, Foxit Software Inc, Fremont, CA** Summer 2018
  Engineered key features for solutions in C++/C# including automation/UI for advanced PDF integration.
- **Research Intern, Shanghai Fudan Microelectronics Group, Shanghai, China** Winter 2017
  Conducted research and implement algorithm in Java to support microelectronics engineering solutions.

RESEARCH & PUBLICATION

- **Dynamic Forecasting for 2020 U.S. Senate Elections** *Publish at* **Political Analysis, 2023**
  Collaborate with *CNN* and design forecasting models for senatorial elections by collecting polling data from major polling companies using web-scraping techniques in python. Successfully forecast outcomes of 33/35 races in 2020 with lower MSE of predicted vote shares than other forecasters, including *The Economist* and *FiveThirtyEight*.
- **Treatment Effect Estimation in Panel Text Data** *Publish at* **AISTATS, 2023**
  Propose and implement a novel difference-in-difference model based on Gaussian Process with more precise causal effect estimation and calibrated event predictions in Matlab. Apply the model and large language model to analyze broadcast transcripts that deepens the understanding of supply-side roles by mainstream news media.
- **Personalised Assessment of Big-Five Personality and Behaviors** *Publish at* **Neurips, 2024**
  Design and execute pilot studies on personalized assessment by conducting experience sampling surveys and building novel deep learning measurement model in pytorch. Substantive findings manage to reconcile a long-lasting psychological debate and contribute to grant award from National Science Foundation of $500,000.
- **Quantitative Latent Measurement using Dynamic Item Response Theory** *Publish at* **APSA, 2023**
  Develop new quantitative models in R and publish packages of Bayesian item response theory that provides better justification and prediction of economic and legal behaviors. Analyze large data archives such as American Panel Survey and Supreme Court dispositions that yield meaningful insights to researchers and clients.
- **Public Opinion Survey using Small-Area Estimation with Post-Stratification** *Work in progress*, **2024**
  Cooperate with National Geospatial-Intelligence Agency and build machine learning models for efficiently estimating small-area public opinion from large nationally representative surveys. Significantly reduce required sample sizes by over 25% and hence alleviating the operational and time-related demands associated with data collection.

TECHNICAL SKILLS

- **Programming** C/C++/C#, Python, R, Matlab, SQL, Java, JavaScript, HTML, Latex, Linux, GIT
- **Machine Learning** Tensorflow, Pytorch, GPyTorch, AWS, Anaconda, Jupyter, Google Colab, Pyro, Stan
- **Data Science** Statistics, Analysis, Database, Visualization, Critical thinking, Communication, Tableau