YEHU CHEN

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

SUMMARY

Ph.D. machine learning engineer and researcher specializing in ML system development, Bayesian and deep learning models, causal inference, and MLOps. Experienced in designing, implementing, and deploying end-to-end machine learning pipelines and large language models (LLMs) that generate actionable insights and optimize business processes. Strong software engineering background with expertise in production-ready code, experimental design, and cross-functional collaboration. Published in leading ML and social science venues.

EDUCATION

- **Ph.D. in Data Science**, Washington University in St Louis. GPA: 3.9/4.0.   2019 - 2025
  Focus: Machine Learning, Quantitative Methods, Causal Inference (A/B testing)

- **Bachelor of Science in Computer Science**, University of Michigan. Summa Cum Laude.   2017 - 2019

WORK EXPERIENCE

- **AI Research Assistant, WashU DI2 Accelerator, St Louis, MO**   May 2025 - Aug 2025
  Engage in research and software engineering on how emerging artificial intelligence technologies intersect with legal frameworks, ethical accountability and public understanding of digital privacy.

- **Data Science Research Assistant, Washington University, St Louis, MO**   Fall 2019 - May 2025
  Design predictive and computational models in Python/R/SQL and conducted causal inference analyses with diff-in-diff method. Built end-to-end data workflows for preprocessing, feature engineering, and large-scale survey analytics. Communicated insights through dashboards, peer-reviewed publications, and stakeholder reports.

- **Software Engineer Intern, Foxit Software Inc, Fremont, CA**   Summer 2018
  Engineered and optimized C++/C# components for advanced PDF automation and user interface plugins.

SELECTED PROJECTS & PUBLICATIONS

- **Agentic AI for Secure Web Crawling**   *In progress*, **2025**
  Designed and implemented an AI agent framework for automated web crawling using Playwright and LLM. Integrated security modules to analyze vulnerabilities and support regulatory tooling for policymakers.

- **Personalized Personality Assessment via Deep Learning**   *Publish at* **NeurIPS, 2024**
  Led survey design and developed PyTorch-based models for experience-sampled personality traits. Findings resolved long-standing debates and contributed to a $500k NSF grant.

- **Election Forecasting for 2020 U.S. Senate Races**   *Publish at* **Political Analysis, 2023**
  Built a polling-based forecasting pipeline (Python, web scraping) that correctly predicted 33 out of 35 races. Outperformed FiveThirtyEight and The Economist in vote share MSE.

- **Causal Analysis of Media Coverage via Gaussian Process Models**   *Publish at* **AISTATS, 2023**
  Estimated treatment effects from panel text data using advanced diff-in-diff models (Python, MATLAB). Incorporated LLMs to preprocess and encode broadcast transcripts for causal inference.

- **Dynamic Item Response Modeling for Behavioral Prediction**   *Publish at* **APSA, 2023**
  Developed Bayesian IRT models in R for modeling latent traits and predicting legal/economic outcomes. Released open-source packages used in political and social science research.

TECHNICAL SKILLS

- **Programming**   C/C++/C#, Python, R, Matlab, SQL, Java, JavaScript, HTML, Latex
- **Machine Learning**   MLOps, Deep Learning, LLM, AWS, Google Colab, Jupyter, TensorFlow, PyTorch, Pyro
- **Software Engineer** Linux, Git, CI/CO, Automated Testing, Data Pipelines, Docker/Kubernetes
- **Soft Skills**   Communication, Collaboration, Data Storytelling, Experimental Design