# Qijia He

Personal Website | LinkedIn | heqj3@uw.edu

## RESEARCH INTERESTS

- Causal inference: individualized optimal decision-making, variable importance in causal inference, causal transfer learning, and mediation analysis
- LLM applications: multimodal figure-to-code generation, and LLM-guided constrained clustering

## EDUCATION

**University of Washington** — **Seattle, WA**
*Ph.D. in Statistics* — *Sep. 2023 - Jun. 2027 (expected)*
*M.S. in Statistics* — *Sep. 2021 - Mar. 2023*

**Sun Yat-sen University** — **Guangzhou, China**
*B.S. in Statistics* — *Aug. 2017 - Jun. 2021*

## PUBLICATIONS

**Rethinking Human Preference Evaluation of LLM Rationales**
*Ziang Li, Manasi Ganti, Zixian Ma, Helena Vasconcelos,* ***Qijia He****, Ranjay Krishna*
XLLM-Reason-Plan @ Conference on Language Modeling (COLM), 2025. [Link]

**The Role of Placebo Samples in Observational Studies**
Ting Ye, ***Qijia He****, Shuxiao Chen, Bo Zhang*
Journal of Causal Inference, 2025. [Link]

**Generalizing the Intention-to-Treat Effect of an Active Control from Historical Placebo-Controlled Trials: A Case Study of the Efficacy of Daily Oral TDF/FTC in the HPTN 084 Study**
***Qijia He****, Fei Gao, Oliver Dukes, Sinead Delany-Moretlwe, Bo Zhang*
Journal of the American Statistical Association, 2024. [Link]

**Estimating Individualized Treatment Rules by Optimizing the Adjusted Probability of a Longer Survival**
***Qijia He****, Shixiao Zhang, Michael L LeBlanc, Yingqi Zhao*
Statistical Methods in Medical Research, 2024. [Link]

**Statistical Learning Methods for Estimating Optimal Individualized Treatment Rules from Observational Data**
***Qijia He****, Yingqi Zhao*
Handbook of Statistical Methods for Precision Medicine, 2024. [Link]

**Research on the Development Trend and Social Effect of Digital Economy** (In Chinese).
*Yan Zeng,* ***Qijia He****, et al.*
China Social Sciences Press, 2021. [Link]

## WORK EXPERIENCE

**Apple** — **Cupertino, CA**
*Machine Learning Data Scientist Intern, Apple Ads* — *Jun. 2025 - Sep. 2025*

- **Causal Effect Estimation**: Developed an instrumental variable framework with DoubleML, integrating MLP models to estimate long-term causal effects of budget pacing more robustly
- **Constrained Clustering**: Developed a constrained K-means pipeline with Apple Foundation Model embeddings, LLM-guided constraints, and performance correlations, using graph-based handling to cluster campaigns by semantic attributes and pacing behavior.

## RESEARCH EXPERIENCE

**University of Washington** — **Seattle, WA**
*Research Assistant in Department of Computer Science and Engineering* — *Jun. 2025 - Present*
Advisor: Prof. Ranjay Krishna

**Multimodal Figure-to-Code Generation**

- Built the first scientific figure dataset for SVG generation through web crawling and synthetic augmentation, using Vtracer to capture shapes and OCR to extract text, extending beyond prior icon-focused works

- Designed a semantic SVG tokenizer, bridging natural language reasoning with structured graphics code generation
- Fine-tuned Qwen model on curated datasets with LLaMA-Factory, enhancing accuracy in complex figure rendering from text and image inputs

**University of Washington** **Seattle, WA**
*Research Assistant in Department of Statistics* *Aug. 2021 - Present*
Advisor: Prof. Alex Luedtke and Prof. Bo Zhang

**Variable Importance for Heterogeneous Treatment Effects under Missing Data**
- Developed an explainable AI approach for estimating variable importance of treatment effect heterogeneity, and adapted for two-stage sampling designs
- Leveraged RKHS-based semiparametric inference to ensure statistical validity and interpretability when integrating with modern machine learning models

**Generalizability and Transportability in Causal Inference**
- Developed a novel causal inference framework to estimate treatment effects of the active control using historical placebo-controlled trial data
- Derived historical-data-driven estimates under point/partial identification, with strategies for sensitivity analysis

**Causal Mediation Analysis for Surrogate Endpoint Evaluation**
- Developed weighted estimators of controlled risk to evaluate surrogate endpoints in the presence of a continuous mediator and positivity violations
- Derived influence-function-based estimators enabling doubly robust, model-agnostic mediation analysis for real-world biomedical applications

**Optimal Adjusted Probability Learning for Individualized Treatment Rules (ITRs) with Censored Data**
- Developed an individualized treatment recommendation system to optimize a proposed a new criterion, enhancing clinical benefit interpretation for clinicians and patients
- Developed a real-time drug recommendation score to support physicians' decisions based on patient risk factors

## PROJECTS

**Graph-Based Error Propagation with Expectation–Maximization (EM) Algorithm for Robust Model Evaluation**
- Proposed a graph-based EM algorithm for robust LLM evaluation under noisy labels, combining synthetic benchmarks with PyTorch-Geometric to model systematic bias and input-dependent noise

**Prediction-Powered Inference (PPI) With Deep Learning Models**
- Implemented the PPI framework for assumption-free inference with black-box ML models
- Developed a PyTorch LeNet CNN pipeline with SciPy calibration and permutation conformal methods

## PRESENTATIONS

**Generalizing the Intention-to-Treat Effect of an Active Control from Historical Placebo-Controlled Trials**
Joint Statistical Meetings. Portland, OR, 2024.
American Causal Inference Conference. Seattle, WA, 2024
The Translational Data Science Integrated Research Center Retreat. Kirkland, WA, 2023.
20th Annual STI & HIV Research Symposium. Seattle, WA, 2023.

**Approximate Bayesian Computation (ABC)-Calibrated Microsimulation Model for Predicting HIV-1 Prevention Efficacy of Broadly Neutralizing Antibodies**
HVTN Africa Regional Meeting. Cape Town, South Africa, 2024.

## TEACHING AND SERVICE

**Reviewer:** Journal of Applied Statistics, Conference on Language Modeling (COLM)

**Teaching Assistant:** Elements of Statistical Methods (STAT 311)

## SKILLS

- Languages: Python, R, SQL, Latex, Matlab
- Framework/Library: PyTorch, PySpark, scikit-learn, Hugging Face Transformers, OpenAI API