

Qianyi Wang

617-359-8333 | joannawqy@gmail.com | [LinkedIn](#) | [Portfolio](#)

EDUCATION

Harvard University

May 2026

M.S. Data Science – *Institute for Applied Computational Science*

Notable Coursework: Machine Learning, Advanced Practical Data Science MLOps, Quantitative Methods for Natural Language Processing, Multilevel and Longitudinal Model, Causal Inferences, Time Series

University of British Columbia

May 2024

B.A. Economics and Statistics

Technical Skills: Python, SQL, R, Scala, PyTorch, Tensorflow, NumPy, Scikit-Learn, Pandas, MapR, Google Cloud, Git, Spark, BASH Scripting, Google Analytics, BigQuery, Excel, Tableau, Power BI, Google Looker Studio

PROFESSIONAL EXPERIENCE

Digital Analyst

May 2023 - Sept 2023

British Columbia Lottery Corporation

Vancouver, BC

- Created **Looker Studio dashboard** through customized **SQL Queries** in **BigQuery**, monitored over 300,000 user sessions, identifying key trends and **conversion rate** bottlenecks that led to a 15% improvement in **user engagement** and a 10% reduction in **bounce rate**.
- Analyzed navigation patterns on over 20 betting features utilizing **Google Analytics**, mapped user engagement through **funnel analysis** with **Looker Studio**, resulting in UX redesign that increased bet placement rates by 12%.
- Performed **hypothesis testing** on banner CTR in **Python**, analyzing 100 banner placements over 6 months; found correlations between banner position and a 10% higher CTR, informing content and design optimizations.

Data Engineer

Sept 2022 - May 2023

Insurance Corporation of British Columbia

Vancouver, BC

- Developed 4 **data pipelines** handling over 2,000 daily transactions using **Scala** on the **Apache Spark** platform.
- Performed maintenance on Scala scripts monthly on the Apache Spark platform, utilized **SQL** within **Hadoop** to ensure daily ingested data integrity.
- Coordinated a 4-month **data migration** project, transferring eight datasets (spanning millions of records) from EDW to ADS with zero downtime.

RESEARCH EXPERIENCE

Network Meta-analyses on Energy Consumption for Climate Mitigation | *R*

Apr 2024 - Present

- Led **meta-data** and **network meta-data** analysis on a dataset of 360 metadata, using **multilevel clustered data** with **robust error estimates** and **network analysis technique** to synthesize insights on energy conservation.
- Authored 3 manuscripts, contributing to the development of **energy-saving policies** with an estimated reach of 100,000 households worldwide.

Climate-Hydrology Analysis in North American Watersheds | *R, Python*

Sept 2023 - Present

- Applied **Multilevel Mixed Effect model** on analyzing climate-hydrology data across 6,000 watersheds, generating insights used in policy recommendations to mitigate climate change.
- Employed **hierarchical clustering** to categorize catchment into five clusters based on hydrology characteristics.

Wildfires and the Distribution of Risk for Commercial Properties | *R, Python*

Mar 2025 - Present

- Conducting econometric modeling, spatial analysis to examine wildfire risk distribution among commercial properties
- Managing micro-level datasets on property values and wildfire events performing data cleaning and integration.

PROJECTS

American Statistical Association (ASA) DataFest 2023 Winning Team

- Leveraged **PyTorch** and **BERT** embeddings to analyze conversations on ABA's pro bono legal services platform, extracting semantic patterns that differentiate answered from unanswered legal queries.
- Identified key linguistic patterns using **TF-IDF** and **SHAP analysis**, uncovering predictive phrases and structural cues that influence response likelihood.
- Combining BERT embeddings, TF-IDF keyword features, and metadata to train a **logistic regression classifier**, achieving 87% accuracy in predicting whether a question would be answered to optimize platform response rates.

Debiasing Pretrained Language Models with Auto-Debias

- Applied **Auto-Debias**, an **orthogonal projection-based method** to mitigate age and disability biases in **BERT** embeddings, reducing **SEAT** effect sizes from 0.51 to 0.04 (disability) and 0.51 to 0.13 (age).
- Implemented layer-wise debiasing across token- and sentence-level embeddings, preserving semantic integrity while maintaining NLP performance on **GLEU** benchmark tasks.
- Evaluated debiasing effectiveness using **cosine similarity metrics** on **News-Commentary v15 corpus** and achieved bias reduction without sacrificing model accuracy in downstream tasks.