# Steven Liu

+1 (519)721-5934 | bingfanliu36@outlook.com | LinkedIn link

## WORK EXPERIENCE

**Applied Scientist Intern** | *Amazon.com, United States* Jun. 2024–Sep. 2024

- Analyzed 2 billion log entries to uncover **position bias** and limitations in single-item recommendation in the legacy model, redefining the problem as a **list-wise** ranking task.
- Implemented a **Transformer-based** personalized **recommender** that ranks full product lists for **100+ million users**.
- Deployed the model end-to-end on **AWS** (**S3**, **IAM**, **SageMaker**) with automated CI/CD pipelines.
- Delivered a **5 – 10 %** lift in **top-k Precision, Recall, and NDCG**, yielding more than **$400 million** annual revenue.
- Presented the approach via org-wide talks and an internal **research conference (oral)**, upskilling 4000+ engineers and scientists.

**Machine Learning and MLOps Engineer Intern** | *Celayix Software, Canada* Sep. 2023–May 2024

- Migrated codebase from TensorFlow1 to **TensorFlow2,** leveraging **AWS** (**S3**, **Lambda**, **Step Functions**, **SageMaker**) services.
- Integrated **Neural Collaborative Filtering** for shift **recommendation**, saving clients up to **$2600** per employee annually.
- Wired automated rollouts through the **TeamCity CI/CD** pipeline, enabling zero-downtime weekly updates of the model.

**Statistical Consultant Intern** | *SFU Big Data Hub, Canada* May 2023– Dec. 2023

- Consulted on rock analysis using **DBSCAN clustering** model, improving **the Silhouette coefficient** metric by 7%.
- Processed **audio data** from language exams for rating estimation using **functional regression method**, achieving 98% R-square.

**Quantitative Model Developer** | *Scotiabank, Canada* Jun. 2021–Aug. 2022

- Led the development of a **machine learning** pipeline for **stress-testing** 90% of bank products valued at **$32 billion**.
- **Selected** the top 10 risk **features** to quantify impacts on bank solvency under **tail risk scenarios**.
- Engineered backend algorithms to reconcile bank balance sheets for assets exceeding **$1.3 trillion**.

**Founding Data Analyst** | *Yumi Organics, Canada* May 2021–Aug. 2022

- Built interactive **data dashboards** for sales analysis across 100+ cities, driving a 20% revenue increase to $5 million.
- Analyzed KPIs, supply chain and campaign, aiding company success and CEO recognition as **Forbes 30 Under 30** in 2023.

**Statistician** | *McGill University Health Centre, Canada* Nov. 2020–May 2021

- Designed a **recommendation algorithm** for kidney transplantation reducing the odds of graft loss by 10%.
- Discovered 500+ critical genetic variables linked to transplant failure using Cox and **network clustering** (WGCNA) models.
- Performed **statistical hypothesis testing** on genetic variable distribution across 10 provinces and diverse ethnic groups.

**Machine Learning Researcher - Signal Data Processing** | *University of Waterloo, Canada* Sep. 2019–Aug. 2020

- Developed an **SVM-based** classifier for binary **classification** with signal input data, achieving 87% classification accuracy.
- Identified 64 neural features for drug addiction classification using **EEG signals**.

**Data Scientist Intern** | *United Nations, Remote* Jun. 2019–Sep. 2019

- Forecasted poverty levels for 5 impoverished countries using a **CNN-backbone** model with satellite **images** in **PyTorch**.
- Achieved 83% accuracy in predicting severe rainfall events using **PCA** and **Random Forest**.

## EDUCATION

**Doctor of Philosophy, Statistics |** *Simon Fraser University* Sep. 2022–Present

- Research: Machine Learning, Computer Vision, Signal Processing, Multimodal Learning, Causal Inference.
- Invited Review Services: CVPR, STAT COMPUT, STAT MED, STAT BIOSCI, JABES, CJS, ECOSTA.

**Master of Mathematics, Statistics |** *University of Waterloo* Sep. 2019–Aug. 2020

**Bachelor of Arts, Honors Economics Joint Honors Mathematics |** *University of Waterloo* Sep. 2016–Apr. 2019

## SKILLS

**Programming Languages/Software**: Python, R, Java, C, SQL, PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, Numpy, Matplotlib, AWS, Git, Bitbucket, TeamCity, PySpark, PowerBI and Tableau.

**Technical Skills**: Machine Learning, Computer Vision, Recommender System, Generative AI, Data Science, Deep Learning, Time Series Analysis, Experimental Design, Causal Inference, Data Visualization, Product Analysis, Data Structure, LVM, LLM, CI/CD, Quantitative Finance, Risk Management, Linux and OOP.

# Steven Liu

+1 (519)721-5934 | bingfanliu36@outlook.com | LinkedIn link

## PROJECTS

**3D Computer Vision - TensoRF Replication**
- Replicated TensoRF for modeling 3D object based on the influential ECCV 2022 paper achieving PSNR score of 30.

**Fraud Detection – A Semi-Supervised Learning Approach**
- Implemented an auto-encoder model in PyTorch that detected 99% of fraudulent records while addressing label imbalance.

**Experimental Design / AB Testing**
- Applied response surface method to reduce simulated Netflix users' average browsing time by 20%.

## PUBLICATIONS

**Robust NeRF++,** *In Progress*
- Developed a robust NeRF model resilient to image artifacts across varying poses.

**Causal Inference in Estimating Time-varying Treatment Effect under Endogeneity,** *In Progress*
- Addressed endogeneity in time-varying treatment effect estimation by proposing a two-stage copula model.

**FunKAN: A Kolmogorov-Arnold Network for Signal-to-Signal Learning using Functional Data Analysis,** *In Progress*
- Proposed a KAN-based network for function-to-function learning with signal-based inputs and outputs.

**LongSurvMamba: A State-space Model for Survival Prediction using Longitudinal Images,** *Submitted to Neurocomputing*
- Introduced a state-space model for interpretable survival prediction through sequential image analysis.

**LongSurvFormer: Transformer Joint Modeling for Survival Prediction using Longitudinal Images,** *Submitted to Biostatistics*
- Engineered a Transformer model to analyze sequential medical images (MRI, CT scans) for Alzheimer's Disease prediction.

**LoFPCA_Cox: Dynamic Risk Prediction with Time-varying Images,** *Submitted to Computational Statistics and Data Analysis*
- Proposed a semi-parametric method for survival risk prediction using time-varying images and incomplete labels.

**Optimal Subsampling for Generalized Additive Models for Large Scale Datasets,** *Statistics and Computing (2025)*
- Introduced an efficient sampling method for parameter estimation in GAMs applied to large-scale datasets.

**Transformer-based Position Aware Item Recommender,** *Amazon Machine Learning Conference (2024)*
- Developed a Transformer-based, position-aware recommender that mitigates position bias for list-wise item recommendations.

**L1-regularized Functional Support Vector Machine,** *Statistics and Its Interface (2024)*
- Engineered an L1-regularized SVM tailored for signal inputs, including ultrasound, audio, EEG, ECG and fMRI data.

**The Calculated Panel of Incompatible Epitopes in Equitable Access to Transplantation,** *American Transplant Congress (2021)*
- Proposed the cPIE statistic to enhance identification of blood and epitope-compatible donors.

**Towards National Organ Sharing: Fair Distribution of Eplets in Canada,** *American Transplant Congress (2021)*
- Analyzed epitope frequency strategies to ensure regional and national HLA compatibility.

**Is Equitable Access to Transplantation Possible in the Era of HLA Epitope Compatibility,** *American Transplant Congress (2021)*
- Investigated the impact of ancestry on transplantation outcomes using frequency-based hypothesis testing.