Shuzhen Zhang

(+1) 2177215819 muhualiushui@outlook.com github.com/muhualiushui Shuzhen Zhang

Education

University of Washington Seattle Sep.2024 — Mar.2026 (expected)
MS in Statistics GPA 3.7/4.0 *Seattle, WA*

University of Illinois Urbana-Champaign Jan.2022 — May.2024
B.S in Mathematics, Statistics GPA 3.8/4.0 | Minor in Computer Science *Champaign, IL*

Honor: Dean's List | 2024 ATLAS Leadership Scholar [1] | 2024 IML Research Reward [2] | Highest distinction in Statistics | High distinction in Mathematics

Skills

- **Programming languages:** Python, C++, JAVA, SQL, Scala, HTML, R
- **Programming Frameworks:** PySpark, Pytorch, Transformer, Jax, Tensorflow, equinox, Numpy, Pandas, Django
- **Technologies:** Deep Learning, Statistic Inference, Data Analyze, Experiment design, Model Productionize

Work Experiences

Amazon Oct.2024-present
Jr. Applied Scientist *Seattle, WA*

- Worked on a webpage-recommendation embedding experiments; Uncovered and **fixed** a legacy label-loss bug by re-linking click labels to requests via Cradle. Initial results underperformed; traced the cause to **positional bias** in the two-stage ranker from web. Confirmed gain in other web and an Amazon Hackathon with cross-team experiments. **Designed** an architecture to fine-tune ASIN embeddings for a seller-side application, thereby completing the user–item representation logic.
- Co-built a unified Pricing + Manage Your Growth (MYG) channel recommender: productionized ASIN- and seller-level features in the Arya data-generation pipeline with clear lineage; established baselines and simplified high-cardinality features; applied **DensMAP** to surface cold-start clusters (e.g., new ASINs with $\approx$ **20**$\times$ CTR); and introduced `is_asin_existing` to stabilize ranking.
- Owned the end-to-end Vendor Central Home Page **web-recommendation system**—from feature productionization and model onboarding to Weblab experiments—producing thorough documentation; through rigorous analysis, uncovered that user attention follows a **negative-exponential trend** and that labels for recommended items are polarized; earned positive feedback from Senior Scientists during Science Review.
- **AMLC Conference:** Contributed to an deep-learning architecture foundation-model with dded a label-aware/causal signal boosting PR-AUC by **10%** over the baseline model through scaled SageMaker Training/HPO

K2Data Jun.2024-Sep.2024
Machine Learning researcher *Beijing, China*

- Constructed a custom **Fourier Neural Operator** (FNO) with PDE forecasting examples by **Python/JAX**. Validated its industrial value on wind-farm multi-turbine turbulence simulations.
- Applied **deep-learning** methods FNO to a real factory electrolytic-aluminum dataset for detecting fraudulent reactions inside electrolyzers, and helped colleagues and customers easily adopt and understand the Deep learning techniques.

UIUC - ATLAS[1] May.2023-May.2024
Machine learning analyzer *Champaign, IL*

- Replicated a GPT-based **AI town simulation** from a paper using the free model **Mistral + GPT2Tokenizer**, independently rewrote queries and algorithms from the original repository, modified NPC personas, and analyzed the logic of AI actions.
- Created chaos characters via **prompt engineering** and experiment design; observed that the filtered AI exhibited preferences and weaknesses in conditional probability output ability.

Keyword AI May.2023-Aug.2023
Software Development Engineering *San Francisco, CA*

- Identified and resolved errors from a legacy model from company repository, and trained a **NER** model targeting job search keywords using the **BERT** framework, with data collected via Python scripts and prompt engineering with GPT.
- Achieved $\approx$ **99% accuracy** in keyword extraction for correctly spelled user inputs, and completed the frontend and backend deployment settings through **Flask** and Google Cloud with a comprehensive Readme instruction.

Research Experiences

- **Manifold Learning:** Contributed to Manifold Learning (mmp GitHub) with standardized plotting and **unsupervised learning** embedding construction. Developed distortion visualization to uncover the effect of density information on complex biological data (C.elegans) with neighborhood breaks and local density. **Publication:** *Interactive Visualization of Metric Distortion in Nonlinear Data Embeddings using the distortions Package*. (July.2024–Present)
- **Illinois Geometry Lab[2]:** Working with two science project with nominated reward: *Scale-free First Passage Percolation*—refactored legacy **C++** for a **40**$\times$ speed-up; implemented **Dijkstra** path-finding and Bézier-curve visualization for poster presentation. *Quantum Circuit & Quantum ML II*—proposed reducing quantum data from 2^n to 1-D via **Fourier Transform**; proved PAC-learnability with accuracy $1-\delta$ as data $D \to \infty$. (Jan.2023–Dec.2023)