Jiaying Ye

+1 (206) 928-3766 | yjy0509@uw.edu | Google Scholar

Education

University of Washington Seattle, WA

B.S. in Applied Mathematics – Data Science & Bachelor of Informatics *Expected Autumn 2026*

- GPA: 3.96/4.00 (Dean's List for 10 consecutive quarters)

Core Skills

Courses: Machine Learning, Numerical Analysis, Differential Equations, Mathematical Statistics, Scientific Computing, Databases and Data Modeling, Quantitative Finance

Programming Languages: Python (Pandas, NumPy, Scikit-learn), R, Java, SQL
Data Engineering: ETL pipelines, data warehousing
Machine Learning: PyTorch, TensorFlow
Data Visualization: Excel, Tableau, QuickSight
Languages: English (Fluent), Mandarin (Native)

Research Experience

Ubicomp Lab, University of Washington Seattle, WA

Research Assistant *Jan 2026 – Present*

- Validated a 350+ hour in-the-wild PPG dataset across 4 wearable form factors; implemented ECG-based quality filtering pipeline in Python to ensure benchmark integrity.
- Implemented and evaluated 7 heart rate estimation models (supervised: ResNet, DCL, CNN-LSTM; self-supervised: SimCLR, BYOL) under 20-fold LOSO protocol, enabling rigorous cross-site performance comparison.
- Analyzed multi-site and PPG–accelerometer fusion across 8 device configurations, informing optimal sensor selection strategies for wearable health monitoring research.

Math AI Lab, University of Washington Seattle, WA

Research Assistant *Sep 2025 – Present*

- Fine-tuned Qwen3-Embedding-8B and Octen-Embedding-8B using SimCLR-style stochastic multi-view contrastive learning with dual DataLoader objectives in PyTorch on 118K math statement triples; generated hard negatives via LLM prompting (Qwen3.5-9B) to improve contrastive training signal.
- Diagnosed and fixed a learning rate schedule bug in sentence-transformers causing premature LR decay with unequal multi-loader setups; implemented per-pair-type evaluation across 6 formal/informal retrieval directions to verify training gains.
- Achieved state-of-the-art on MELD and mathlib retrieval benchmarks, outperforming 6 competitive baselines including Qwen3-Embedding and harrier-oss.

Stanford Rural Education Action Program Remote

Research Assistant *Aug 2024 – Present*

- Initiated a research project analyzing the effects of maternal socioeconomic status on early childhood development, employing linear and logistic regression models to analyze 600+ datasets.
- Engineered an end-to-end data analysis pipeline using R, performing data cleaning, feature engineering, and regression modeling, which improved data readiness for analysis by 10%.
- Developed and interpreted multiple regression models, identifying maternal education level as the most significant predictor ($p < 0.01$); currently drafting a manuscript for peer-reviewed publication.

National Clinical Research Center for Child Health Hangzhou, China

Research Assistant *Jun 2022 – Jul 2024*

- Engineered a standardized data preprocessing pipeline in Python that automated spectral normalization and baseline correction, eliminating manual intervention and ensuring consistent data quality across 500+ samples.
- Analyzed ML-driven diagnostic frameworks, evaluating the performance of models such as SVM, Random Forest, and CNN across cancer, pathogen, and metabolic disease diagnostics.
- Evaluated advanced neural networks such as 1D-ResNet for bacterial identification (99.7% accuracy), identifying spectral biomarkers and model decision bases via saliency mapping and feature importance analysis.

Work Experience

VMware LLC Hangzhou, China

Virtualization Infrastructure Intern *Jun 2025 – Sep 2025*

- Developed PowerCLI automation scripts for VM provisioning and deployment, reducing single-VM setup time from 2 hours to 15 minutes; scripts were adopted by the team to deploy 100+ demo VMs.
- Validated cluster stability via stress testing on a 2-host vSphere environment, providing performance data that directly informed resource allocation decisions for production deployment.

Publications

[1] Does My Embedding Reflect That $A = B$? Evaluating Mathematical Equivalence in Embedding Models
Jiaying Ye, Samarth Rao, Leo Carlin, Kedar Chintalapati, Saharsh Bhargava, Rachit Jaiswal, Michael Zhou, Jared Darlington, Vasily Ilin, Jarod Alper, Henry Kvinge
AI4Math Workshop@ICML, 2026

[2] Multi-site PPG: An In-the-Wild Physiological Dataset from Emerging Multi-site Wearables
Jiayi Shao, **Jiaying Ye**, Shengyao Liu, Zachary Englhardt, Girish Narayanswamy, Vikram Iyer, Qiuyue (Shirley) Xue
NeurIPS Datasets and Benchmarks Track, 2026

[3] Recent Research Progress in the Integration of Raman Spectroscopy with Machine Learning Algorithms for Disease Diagnosis
Zefeng Zheng*, Ying Cui*, Boyou Zhang*, Yang Li, Yongjie Liu, **Jiaying Ye**, Feng Yuan, Khek-Yu Ho, Jing Liu, Lijia Wang
Nano Research, 2025

[4] Raman Spectroscopy of Deuterium-Labeled Macromolecules: Emerging Strategies in Biomedical Research
Zefeng Zheng*, Ying Cui*, Boyou Zhang*, Yang Li, Yongjie Liu, Shanji Fan, **Jiaying Ye**, Feng Yuan, Jing Liu, Lijia Wang
International Journal of Biological Macromolecules, 2025

[5] Online Symbolic Distillation: Learning from Teacher Continuations at Student States
Haojin Wang*, Dylan Zhang*, Huaibo Chen, Suhao Yu, Yihang Sun, Zhanyang Jin, **Jiaying Ye**, Dianqi Li, Prasanna Sattigeri, Kamal Youcef-Toumi, Hao Peng
Preprint, 2026

[6] ParetoRSI-Gym: Can AI Agents Improve Language Models Across Multiple Domains?
Haojin Wang, Yangzhen Wu, Zeyi Chen, Jian Xie, **Jiaying Ye**, Lili Meng, Shuyun Tang, Shangding Gu, Dawn Song
Preprint, 2026
https://arxiv.org/abs/2610.34386