

Brynn Yein Lee

brynnylee.github.io/portfolio | linkedin.com/in/yeinl/ | brynnlee.ds@gmail.com | Los Angeles, CA

Work Experience

Graduate Research Assistant | AIR Lab (Audio Information Research)

New York, U.S.

Department of Electrical and Computer Engineering, University of Rochester

Sep 2024 – Present

- **Elevated** a voice deepfake detection pipeline by upgrading nural network architectures (AASIST) and integrating 20K+ diverse datasets (synthesized, cloned, recorded), enhancing model generalization and cross-domain scalability to date.
- **Designed and deployed** a scalable dataset splitting logic with seamless integration of SAMO-specific loss functions, preserving data integrity across the entire modeling pipeline.
- **Engineered** a robust augmentation pipeline using RawBoost to address extreme class imbalance and **incorporated** error analysis workflows to refine model accuracy, achieving a current EER reduction to 0.05 during validation.

Data Scientist Intern | IngenID

New York, U.S.

Provider of voice biometrics solutions (24M+ transactions annually)

Jul 2024 – Aug 2024

- **Synthesized** a 14K audio dataset by integrating and fine-tuning 18 generative AI and neural network models (Text-to-Speech, voice cloning), enabling the development of a robust deepfake detection model with enhanced accuracy and robustness.
- **Developed and executed** a data collection strategy with 28 key parameters, optimizing parameter distributions to collect 400K metadata values, reducing dataset bias, and enhancing generalization.

Data Scientist & Co-founder | Artclub

Songdo, South Korea

Online course platform for Korean students targeting U.S. art schools (Achieved \$3M seed funding)

Jul 2020 – Jan 2022

- **Conducted and automated** NLP analysis of 50K+ messages from Korea's #1 user-ranked chat community in the category.
- **Increased** purchases by 25% and eliminated manual analysis of 300K+ messages annually, saving \$6,500 per year.
- **Built** ETL pipelines and **engineered** features for a recommendation system, decreasing dropout rates by 10%.

Data Analyst | Living Jin

Seoul, South Korea

E-commerce retailer of Korean food for U.S. market (#1 in category on Amazon US)

Feb 2019 – Jul 2020

- **Analyzed** 7,800+ product reviews using NLP to uncover customer behavior patterns, boosting repeat purchases by 13% and driving the launch of two new products targeting loyal customers.
- **Optimized** Amazon ad campaigns for a \$4M annual sales product line by refining keywords and budgets, increasing RoAS by 17%, resulting in an additional \$680K in annual revenue.
- **Improved** project outcomes through ongoing 1:1 mentorship by Nvidia Korea's ex-CEO.

Technical Skills

- **Tools:** Python (TensorFlow, PyTorch, Scikit-learn, Hugging Face, Pandas, NumPy, LightGBM, XGBoost, GeoPandas, Librosa), SQL, R | **Databricks** (Spark, Mlflow, Delta Lake) | **AWS** (S3, SageMaker) | Docker, Git | Tableau, Power BI, WandB
- **Techniques:** Machine Learning (Predictive Modeling, Time Series Forecasting, A/B Testing, Bayesian Inference), NLP (LLMs, Sentiment Analysis, Topic Modeling), Deep Learning (RNN, LSTM, Transformer Architectures), Generative AI | **Data Engineering** (Data Pipelines, Data Augmentation) | **Geospatial Data Analysis** (ArcGIS) | Model Deployment

Projects

Emergency Response & Resource Optimization for City of Rochester Fire Department

Dec 2024

- **Led** the student team to analyze 1.6M+ incident records using time-series models (Prophet, SARIMAX) and geospatial tools (Folium, GeoPandas), forecasting incidents and optimizing resources for 500 personnel managing 50K annual dispatches.
- **Delivered** predictive models with (85% R^2 accuracy), **evaluated** fire stations placemebts, and **recommended** reallocation strategies for 19 vehicles, through comprehensive reports and html interactive maps.
- **Secured** recognition with media features on the [University of Rochester News Center](#).

Real-Time Streaming X(Tweet) Sentiment Analysis on Influencers

Jun 2024

- **Designed and implemented** a scalable sentiment analysis pipeline with Apache Spark, Delta Lake, AWS S3, and Databricks, processing 200K+ tweets through a medallion architecture to analyze influencer sentiment trends and drivers.
- **Deployed** a Hugging Face transformer model for real-time sentiment classification, seamlessly integrated with MLflow for inference tracking and performance monitoring, achieving 75% precision.

Academic Papers

Zhang, Y., Lee, B., & Duan, Z. (Expected Jan 2025). "SAMO+: Enhanced Speaker Attractor Multi-Center Learning for Voice Anti-Spoofing."

Education

University of Rochester

Rochester, NY

M.S. in Data Science (Merit-based scholarship recipient)

Aug 2023 – Dec 2024

Yonsei University

Wonju, South Korea

B.S. in Physics (Merit-based scholarship recipient)

Mar 2014 – Feb 2019