# WONHA SHIN

Phone: +1 551-245-2356   |  Email: leahnote01@gmail.com
Linkedin: https://www.linkedin.com/in/wshin7  |  Project Portfolio: https://leahnote01.github.io/

## EDUCATION

| **University of Rochester** | **Rochester, New York** |
|---|---|
| Master of Data Science  GPA 3.43 / 4.0 | *Aug 2023 – Dec 2024* |

| **Pusan National University** | **Busan, South Korea** |
|---|---|
| Bachelor of Arts in International Business and Economics  GPA 3.73 / 4.0 **(Cum Laude)** | *Mar 2011 – Aug 2018* |

## PROFESSIONAL EXPERIENCE

**VOESH New York**                                                                                 **South Plainfield, New Jersey**
*Administrative Accounting Specialist, Admin/Accounting*                                   *Oct 2018 – Mar 2023*
- Developed a predictive time-series framework integrated with Oracle DB, improving sales forecasting accuracy by 15% for 10,000+ clients. Revamped ERP systems, reducing data errors by 70% and enhanced cross-team efficiency.

**EASTERN AMERICAN CDC (Non-Profit Federal Community Bank)**                   **Englewood, New Jersey**
*Credit Analyst Intern, Credit Assessment*                                                        *May 2018 – Oct 2018*
- Conducted credit analysis for 100+ SBA 504 loan applications, mitigating financial risks and supporting funding decisions for small businesses. Assisted in securing new clients through active financial analysis and risk assessment.

## PROJECTS & SKILLS

- **LLM-Based Real-Time Translation System with Kafka and MLOps Automation:** Spearheaded a fully end-to-end real-time translation system as a solo engineer, designing and building every component of the pipeline. Architected a scalable solution using **Kafka** for distributed streaming, **FastAPI** for asynchronous API orchestration, and **Hugging Face Transformer** models for neural machine translation. Built a robust **MLOps** framework, integrating **Prometheus** and **Grafana** for system monitoring and **Weights & Biases (W&B)** for drift detection, retraining and model versioning.
- Achieved ≤ 0.001s translation latency by optimizing FastAPI's async architecture and tuning Kafka streaming throughput. Independently owned the full ML system lifecycle—from data ingestion and model integration to API performance, monitoring, and experiment tracking—demonstrating deep expertise in both ML engineering and MLOps.

- **200 Days Learning Posting Challenge ('24–'25):** Completed a 200-day deep learning and MLOps-focused study challenge, mastering foundational and advanced topics in ML, AI, and Deep Learning—from regression, PCA, decision trees, and SVMs to CNNs, transformers, and LLMs. Explored advanced themes including efficient transformers, bias mitigation, AI ethics, conformal prediction, and MLOps practices such as reproducibility, monitoring, and deployment.
- Documented daily progress with technical write-ups on GitHub and LinkedIn, engaging the broader AI/ML community.

- **E-Cigarette Perception Capstone Project (Chief Operations Manager):** Directed a full-cycle NLP project analyzing 100,000+ multilingual social media posts to uncover public perceptions of e-cigarette use and support public health initiatives. Led team coordination, modeling, MLOps planning, data governance, and research design.
- Supervised human annotation workflows, developed labeling guidelines, and ensured labeling quality and consistency.
- Applied sentiment analysis and LDA topic modeling to identify health risks, misinformation, and demographic patterns.
- Currently preparing findings for publication, with focus on health discourse, language bias, and cross-linguistic trends.

- **Real-Time Tweet Sentiment Analysis Pipeline :** Built a real-time sentiment analysis pipeline using **PySpark**, **Databricks**, and **AWS**, streaming tweets from S3, classifying them with a **pretrained transformer model,** and storing results in **Delta Lake**. Ensured structured streaming and windowed triggers for reliable, low-latency real-time processing.

- **Data Mining: Cross-Cultural NLP Analysis of Luxury Hotel Reviews in Europe- LDA Topic Modeling :** Led a transformative NLP project, dissecting over 515,000 hotel reviews to reveal nuanced cultural patterns across Europe's diverse traveler demographics. Leveraged LDA Topic Modeling and Scikit-Learn to unlock empowering insights.

- **Skills:** Python (NumPy, Pandas, Scikit-learn, FastAPI, Hugging Face Transformers), Machine Learning (Deep Learning, Natural Language Processing, Time-Series Analysis, Predictive Modeling, Sentiment Analysis, Topic Modeling), MLOps (Kubernetes, Docker, Prometheus, Grafana, Weights & Biases, Model Versioning, Drift Detection, CI/CD, Deployment), Big Data & Cloud (Kafka, PySpark, Databricks, AWS, Delta Lake), Databases (SQL, Oracle DB), Version Control (Git).