Seohee Sunny Yoon

syoon333@gatech.edu +1 470 838 2401 seohee-yoon-268482263 Portfolio GitHub

Education

M.S. in Computer Science *Expected Dec 2025*
Georgia Institute of Technology - Atlanta, GA *GPA: 3.87/4.00*

B.S. in Computer Science *May 2024*
Georgia Institute of Technology - Atlanta, GA *GPA: 3.72/4.00*

- Awarded merit-based scholarships for both B.S. and M.S. ($20,000 annually, 2023–2025)

Work Experience

Graduate Research Assistant, SocWebLab - Atlanta, GA *Feb 2025 – Present*
Supervisor: Professor Munmun De Choudhury

- **Investigating bias in large language models (LLMs)**, detecting bias against mental health entities and advancing **Responsible AI research** with implications for safe AI deployment
- **Analyzing 10K+ LLM responses** using statistical measures to uncover consistency and divergence of responses and generate bias mitigation strategies
- Validating classifier reliability via a **hybrid evaluation pipeline** (LLM evaluators + human review, 300 samples/model across 3 classes), aligning with **industry evaluation practices**

Data Science Intern, LG Uplus - Seoul, South Korea *Jun 2025 – Aug 2025*

- Built a **context-aware web/app recommendation system** for 137K users and 50M+ interactions, leveraging **location and time context** and ensembling two models to capture users' short- and long-term preferences
- Benchmarked against history-based recommendation, the ensemble achieved **15× higher diversity** and **3× greater coverage**, reducing user disengagement from repetitive results
- Optimized inference pipeline for **30% faster latency** and **20% lower GPU memory**, enabling scalability for **nationwide telecom services**

Data Analysis Intern, PTKOREA - Seoul, South Korea *Jun 2024 – Aug 2024*

- Managed and triaged **20+ daily customer-reported issues** on dashboards via Jira/Tableau, improving reliability and **stakeholder confidence**
- Resolved **client data discrepancies** with SQL, Excel, Adobe Analytics, ensuring compliance with diverse client requirements under tight deadlines
- **Led pre-release QA testing for dashboards** with engineers/managers, delivering more accurate performance and improvements in user satisfaction

Undergraduate Research Assistant, C21U Lab - Atlanta, GA *Aug 2022 – Dec 2023*
Supervisors: Dr. Jeonghyun Lee, Dr. Meryem Yilmaz Soylu

- **Spearheaded a data science team of 4 to integrate Depth of Knowledge (DOK) levels into 1,200+ student assessments**, enabling instructors to design more effective, data-driven assessments
- Implemented **TF-IDF-based feature extraction** to identify key terms critical for difficulty estimation, complemented by data augmentation to improve dataset quality
- Evaluated multiple regressors to identify the optimal model, achieving a **25% improvement in prediction precision** and strengthening assessment reliability

Projects

Rabbit Hole Project

- Analyzed 10K+ jailbreak responses with **k-means clustering**, uncovering 12 adversarial strategy groups and applied **PCA** for dimensionality reduction to visualize response distributions
- Identified representative sentences via **embedding similarity**, characterizing narratives and providing insights for LLM safety evaluation

Implicit Emotion Classification

- Optimized preprocessing pipeline for an implicit emotion dataset of over 30K text samples across 6 classes, reducing model training time by **20%** through regular expressions and feature engineering
- Boosted classification accuracy/F1 from **64% to 88%** by integrating **BERT with a sentiment knowledge graph**, providing insights for implicit emotion detection in user-generated content

Respiratory Diagnosis Assistant

- Developed a GRU-based respiratory diagnosis assistant with a React prototype, improving **accuracy from 70% to 83%** through advanced audio preprocessing (noise injection, pitch shifting, MFCC feature extraction)
- Demonstrated potential for **real-time, low-cost respiratory screening** in telemedicine and mobile health applications, highlighting value beyond model benchmarks

Skills

- **Programming Languages:** Python, SQL, PySpark, Scala, R, Java, JavaScript
- **Machine Learning:** Scikit-learn, Pandas, NumPy, Microsoft Azure Machine Learning, PyTorch, TensorFlow
- **Data Visualization:** Tableau, Matplotlib, Seaborn, Excel, D3.js, Adobe Analytics
- **Cloud & Big Data:** AWS (S3, RDS), Spark