Sunny Yoon

✉ syoon333@gatech.edu   📞 +1 470 838 2401   in seohee-yoon   🔗 Portfolio   ⭘ GitHub

Education

Master of Science (M.S.) in Computer Science   *Expected Dec 2025*
Georgia Institute of Technology - Atlanta, GA   *GPA: 3.87/4.00*
Bachelor of Science (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
  ○ Investigate bias in large language models (LLMs), detecting harmful stereotypes against mental health entities and advancing Responsible AI that improve safety in real-world applications
  ○ Leverage natural language processing (NLP) techniques with statistical analysis on 10K+ LLM responses to quantify consistency and divergence patterns for bias mitigation
  ○ Validate toxicity classifier reliability through a hybrid evaluation pipeline (LLM evaluators + human review, 500 samples/model)

Data Scientist Intern, LG Uplus - Seoul, South Korea   *Jun 2025 – Aug 2025*
  ○ Developed a novel context-aware recommendation system on a Docker-based MLOps platform, integrating DyCARS and HGT models on 50M+ interactions from 137K+ users to capture short- and long-term preferences
  ○ Improved recommendation metrics, achieving 15× higher diversity and 3× broader coverage over history-based methods, mitigating repetitive suggestions and enhancing retention
  ○ Optimized inference pipeline, achieving 30% faster latency and 20% lower memory usage through feature engineering and batch-size tuning, reducing resource costs and supporting nationwide scalability

Data Scientist Intern, PTKOREA - Seoul, South Korea   *Jun 2024 – Aug 2024*
  ○ Managed 20+ daily user-reported issues on business performance dashboards via Jira/Tableau, improving stakeholder confidence
  ○ Resolved dashboard data discrepancies with SQL/Excel/Adobe Analytics to deliver accurate KPI metrics for business decision-making
  ○ Collaborated within a cross-functional team of engineers and managers to conduct QA testing, detect anomalies, and ensure reliable system deployment

Undergraduate Research Assistant, C21U Lab - Atlanta, GA   *Aug 2022 – Dec 2023*
Supervisors: Dr. Jeonghyun Lee, Dr. Meryem Yilmaz Soylu
  ○ Led 4 data scientists in classifying 1K+ assessments by difficulty to support objective models for effective student evaluation, leveraging the Depth of Knowledge framework
  ○ Designed reproducible data pipelines with text based feature extraction, improving regression model precision by 25% to strengthen the reliability of difficulty predictions for assessment design

Projects

Rabbit Hole Project
  ○ Analyzed LLM jailbreak responses to identify 12 target groups using k-means clustering, and applied PCA for intuitive visualization
  ○ Examined harmful stereotype patterns through lexical, structural, and semantic analyses by group using NLP (frequency analysis, sequence matching, and cosine similarity) for LLM safety evaluation

Implicit Emotion Classification
  ○ Boosted classification accuracy from $64\% \rightarrow 88\%$ by integrating RoBERTa with a sentiment knowledge graph and applying soft-position embedding for smoother knowledge integration
  ○ Improved classifier efficiency by reducing training time 20% through EDA, feature selection and noise reduction with a visible matrix

Respiratory Diagnosis Assistant
  ○ Developed a GRU-based respiratory diagnosis web application, integrating S3 and MongoDB for large-scale audio management, showcasing end-to-end software development from model to deployment
  ○ Enhanced audio processing (noise injection, pitch shifting, MFCC features) to improve accuracy from $70\% \rightarrow 83\%$, demonstrating a real-time, low-cost prototype for telemedicine and mobile health

Skills

Programming Languages: Python, SQL, PySpark, Scala, R, Java, JavaScript
Machine Learning: Scikit-learn, Pandas, NumPy, Microsoft Azure Machine Learning, TensorFlow/PyTorch
Data Visualization & Tools: Tableau, Power BI, Matplotlib, Seaborn, Excel, D3.js, Adobe Analytics, Git, Atlassian tools (Jira, Confluence)
Cloud & Big Data: AWS (S3, RDS), Hadoop, MongoDB, Docker, Spark, GCP, FastAPI, Relational Databases (MySQL, PostgreSQL)