

SAEHEE EOM

Atlanta, GA • (404) 203-5710 • saehee.data@gmail.com • www.linkedin.com/in/saehee-eom/



EDUCATION

GEORGIA INSTITUTE OF TECHNOLOGY

Master of Science in Analytics (Data Science)

Atlanta, GA

Aug. 2025 - Dec. 2026

Relevant Coursework: Natural Language Processing, Machine Learning, Data Visualization, Database Systems, Statistical Modeling

SEOUL NATIONAL UNIVERSITY (SNU)

Bachelor of Science in Artificial Intelligence, Bachelor of Arts in Linguistics

Seoul, South Korea

Mar. 2019 - Feb. 2025

Relevant Coursework: Natural Language Processing, AI, Computer Vision, Data Structures, Algorithms, Databases, Probability Theory

SKILLS

Programming & ML: Python (Hugging Face, OpenAI APIs, PyTorch, TensorFlow, Scikit-Learn), R, Java, C++

Data Engineering & Cloud: SQL (BigQuery, MySQL, SQLite), Spark, Hadoop, AWS, GCP, Databricks, Docker

EXPERIENCE

Salesforce

Data Scientist Intern (Industry GTM Team)

Jan. 2025 - Jun. 2025

- Designed scoring algorithm and recommender system to predict company-level enterprise SaaS opportunity close rates, directly informing next FY GTM sales strategy and prioritization of high-potential industries among 50+ sectors
- Built scalable ETL pipeline (Snowflake-Tableau) using Tableau Prep and developed role-based access control (RBAC) dashboards, reducing manual reporting effort by 90% and providing tailored visibility into sales performance
- Collaborated with global stakeholders in Korea, USA, and Japan to unify data definitions and access protocols

[SNU SKI-ML Lab](#) | Natural Language Processing (NLP) Research Laboratory

Research Intern (Advisor: Prof. Jay-yoon Lee)

Jun. 2024 - Jan. 2025

- Co-authored paper on Controlled Text Generation, improving fluency (+23% BERTScore) and reducing perplexity (-22%) with an energy model-based approach to handle multiple constraints (toxicity, logical consistency, formality)
- Created scalable evaluation framework to benchmark LLM outputs (fluency, controllability, content preservation) across multiple LLMs (LLaMA, Phi, Qwen) and shared on GitHub for team-wide reproducible research
- Designed and optimized prompts for multiple LLMs, generated synthetic datasets using ChatGPT, and labeled data to support GenAI training and downstream tasks, improving constraint satisfaction by 31%
- Led journal club sessions on NLP, RAG, autonomous agents, LLM applications, embeddings, and prompt engineering

[LBOX](#) | Korea's leading and fastest-scaling LegalTech startup (Series C)

Data Scientist Intern

Jul. 2022 - Jan. 2023

- Co-authored an EMNLP 2022 workshop paper (NLLP) on Information Extraction (IE) by organizing data collecting workflows, designing legal ontologies, and building baseline models, evaluation frameworks, and visualizations
- Created 200+ KPI dashboards (MongoDB-SQL) and weekly market reports (Python Web Scraping) to monitor user data, campaign results & competitor trends, enabling data-driven initiatives and driving a 161% YoY sales increase

PROJECTS

1) CRM Targeting via Ensemble Modeling | Hyundai Home Shopping Networks

- Boosted purchase rate of target users by 42% via an ensemble of LightGBM and SASRec (sequential recommender)
- Engineered user/product-level features (RFM, clusters, Gini coefficient), deployed the model on AWS for real-time prediction and pitched insights to non-technical stakeholders, contributing to a 14% lift in campaign engagement
- Analyzed 10K+ user reviews via sentiment and text analysis to identify UX pain points, proposed structured review input templates, and quantified expected engagement gains through data-driven simulations

2) Call Volume Forecasting for Call Center Planning | Hyundai Home Shopping Networks

- Forecasted call center volume using CatBoost and LSTM (MAE < 100, 29% lower monthly error) and deployed via Streamlit, enabling managers to optimize workforce allocation, reduce HR costs, and minimize service delays

PUBLICATIONS

- Locate & Edit: Text Editing-Based Controlled Text Generation for Black-Box LLMs*, under review, Jul. 2025
- Syntactic Complexity and LLM Performance: Exploring the Role of Prompt Design*, bachelor's thesis, Dec. 2024
- Data-efficient End-to-end IE for Statistical Legal Analysis*, [EMNLP Workshop \(NLLP 2022\)](#), Dec. 2022