

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)

Relevant Coursework: Machine Learning, Large Language Models, Supply Chain, Algorithms, Database Systems, Statistical Modeling

Atlanta, GA

Aug. 2025 - Dec. 2026

SEOUL NATIONAL UNIVERSITY (SNU)

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

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

Seoul, South Korea

Mar. 2019 - Feb. 2025

SKILLS

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

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

EXPERIENCE

Salesforce

Data Scientist Intern (Industry GTM Team)

Jan. 2025 - Jun. 2025

- Designed **predictive forecasting and market segmentation models** to estimate TAM/SAM across 50+ industries, supporting data-driven demand planning and resource allocation for next fiscal year GTM strategy
- Built automated ETL pipeline (Snowflake-Tableau) and dashboards to analyze sales demand patterns across large-scale datasets, supporting EDA and operational performance monitoring while reducing reporting cycle time by 90%
- 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%) under multiple constraints (toxicity, logical consistency, formality); Led all LLM-related experiments
- Created scalable benchmarking and **evaluation framework to benchmark LLM outputs** (fluency, controllability, content preservation) across multiple models (GPT, LLaMA, Phi), using GitHub for team-wide reproducible research
- Designed and optimized prompts for multiple LLMs, generated synthetic datasets using ChatGPT, and labeled data to support fine-tuning Gen AI models, improving constraint satisfaction by 31% compared to baseline
- 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 building baseline models, evaluation frameworks, and visualizations, organizing data collecting workflows, and designing legal ontologies
- 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

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

- Forecasted call center demand 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

Personalized Recommendation System via Ensemble Modeling | Hyundai Home Shopping Networks

- Boosted purchase rate of target users by 42% via an ensemble of LightGBM and SASRec (sequential recommender)
- Performed feature engineering on user/product-level data (RFM, clusters, Gini coefficient), deployed predictive model on AWS for real-time inference, 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

PUBLICATIONS

- **Locate & Edit: Text Editing-Based Controlled Text Generation for Black-Box LLMs**, under revision, 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