

## Summary

---

ML engineer and applied researcher with industry experience in **LLM evaluation** (Google/Gemini) and **ads bidding** (Moloco). Strong background in statistical modeling and data-driven analysis; research focus on **semi-supervised learning** problems.

## Education

---

**Ph.D. in Statistics and Operations Research**, University of North Carolina at Chapel Hill (UNC)

Chapel Hill, NC

Aug 2021 – May 2026

- Advisor: Dr. Vladas Pipiras

**M.S. in Statistics**, Seoul National University (SNU)

Seoul, South Korea

Mar 2019 – Feb 2021

- Advisor: Dr. Hee-Seok Oh

**B.S. in Statistics, Minored in Computer Science and Engineering**, Seoul National University (SNU)

Mar 2015 – Feb 2019

## Work Experience

---

**Google** | Software Engineer Intern | Google Research

San Francisco, CA

May 2025 – Aug 2025

- Developed an LLM factuality evaluation metric of **Gemini** responses to search queries, using **Python**.
- Validated the metric against human raters and auto-raters; collaborated with Research scientists on iterative metric design.

**Moloco** | Software Engineer Research Intern | Ads Bidding

Seattle, WA

- Built experimentation workflows to evaluate ads bidding quality and understand revenue impact.
- Identified data distribution shift in ML pipeline; developed reweighting strategy ([arXiv](#)) using **SQL** and **Python**.
- Partnered with ML engineers to translate experimental findings into production system improvements.

**Lawrence Livermore National Laboratory (LLNL)** | NSF Mathematical Sciences Graduate Intern

Livermore, CA

- Built data-driven methods in [libROM](#), a scalable C++ library for high-performance physics simulations.

May 2022 – Jul 2022

## Research Experience

---

*Semi-supervised learning for data-efficient ML; rare/extreme event prediction; neural network architectures.*

**Multi-Fidelity and Semi-Supervised Learning Methods for Data-Efficient Estimation (UNC)** [1,2,4]

- Developed **semi-supervised and multi-fidelity methods** leveraging abundant low-cost (unlabeled) data to improve estimation with limited high-cost (labeled) data, with focus on predicting **rare/extreme events**. Published in top statistics journals.

**Neural Network Architectures for Spatio-Temporal Modeling (UNC)** [3]

- Designed a latent-space architecture for learning dynamics of spatio-temporal data in latent space using **PyTorch**. Published at [NeurIPS 2024 Workshop](#).

**Multi-scale Time-Series Clustering (SNU)** [5]

- Built clustering methods for large-scale behavioral data (wearable devices), detecting multi-scale time-series patterns.

**Course Projects (UNC)** RL: [RLHF for Human Preference Alignment](#) / ML: [Generative diffusion models](#) / OT: [Optimal Transport](#)

## Teaching Experience

---

**Instructor** [UNC] STOR 155, Introduction to Data Models and Inference (24F, sole instructor for 50 students; full course ownership).

**Teaching Assistant** [UNC] Statistical Theory (25F), Introduction to Deep Learning (22Sp), Methods of Data Analysis (22Sp, 21F), etc.

## Selected Honors and Awards

---

**Cambanis-Hoeffding-Nicholson Award (2022)** UNC, for outstanding academic performance in first-year doctoral program.

**NSF Mathematical Sciences Graduate Internship (2022)** Oak Ridge Institute for Science and Education.

**Korean Government Scholarship for Overseas Study (2021, \$80K)** Korean Government, 5 students in the Intelligent Infrastructure field selected nationwide for a doctoral study abroad program.

**The Presidential Science Scholarship (2015, Tuitions and incentives)** Top national scholarship for outstanding STEM students.

## Selected Publications

---

- [1] M. Kim, B. Brown, V. Pipiras, (2026), '[Parametric multi-fidelity Monte Carlo estimation with applications to extremes](#)', *Technometrics*, accepted.
- [2] M. Kim, K. O'Connor, V. Pipiras, T. Sapsis, (2025), '[Sampling low-fidelity outputs for estimation of high-fidelity density and its tails](#)', *SIAM/ASA J. Uncertainty Quantification*, **13**, pp. 30–62.
- [3] M. Kim, T. Wen, K. Lee, Y. Choi, (2024), '[Physics-informed reduced order model with conditional neural fields](#)', *NeurIPS 2024 Workshop on ML and Physical Sciences*.
- [4] M. Kim, V. Pipiras, A. Reed, K. Weems, (2023), '[Calibration of low-fidelity ship motion programs through regressions of high-fidelity forces](#)', *Ocean Engineering* **290**, 116321.
- [5] M. Kim, H. Oh, Y. Lim, (2023), '[Zero-Inflated Time-Series Clustering Via Ensemble Thick-Pen Transform](#)', *Journal of Classification* **40**, pp. 407–431.

## Technical Skills

---

**Tools:** Python, SQL, C++, R, PyTorch, TensorFlow   **ML:** Classifiers, rare event prediction, LLM evaluation, experimentation pipelines