# Yeo Jin Jung

(+1) 872-904-5718 yeojinjung@uchicago.edu yeojin-jung126 yeojin-jung

## Summary

PhD candidate in Statistics specializing in high-dimensional theory and its integration with large-scale ML systems. My expertise lies in two areas: extracting low-rank structure (PCA/SVD, topic models) and developing distribution-free uncertainty quantification via conformal prediction to develop trustworthy AI systems. I utilize these frameworks to solve complex data problems, including optimizing data selection for efficient training of large generative models.

## Education

**The University of Chicago** Chicago, IL
*Ph.D. in Statistics (4th year,* ***available Jun.-Sep. 2026, graduation Jun. 2027****)* Sep. 2022 - Present

**Yonsei University** Seoul, South Korea
*M.A. in Statistics and Data Science, GPA: 4.20/4.30* Mar. 2020 - Feb, 2022

**Yonsei University** Seoul, South Korea
*B.A. in Applied Statistics, GPA: 4.02/4.30, Class Rank: 5/81* Mar. 2016 – Feb. 2020

- **Awards:** Highest Honors (Spring 2019); Honors (Spring 2016, 2018)

## Research Experience

**SIGNAL Lab (The University of Chicago)** Chicago, IL
*Research Assistant (Advisor: Prof. Claire Donnat)* Sep. 2023 - Present

- Implemented a data selection framework for efficient LLM pretraining, using gradient-based importance signals to score high-value training data points and reduce computation time.
- Developed a scalable conditional conformal prediction framework with a piecewise-linear path-tracing solver for kernel quantile regression, achieving 20–50× faster inference than standard solvers; released as a Python package.
- Contributed to a conformal data augmentation pipeline that filters synthetic data with confidence sets, improving downstream performance of LLMs, computer vision systems, and fraud detection models by 2-6% across benchmarks.
- Designed a graph-constrained iterative SVD algorithm for spatially structured data, establishing high-probability error analysis and applying to spatial omics datasets for tumor microenvironment discovery; released as a Python package.

**Yonsei University** Seoul, South Korea
*Research Assistant (Advisor: Prof. Jaewoo Park)* Nov. 2020 - Aug. 2022

- Proposed a fast Bayesian spatio-temporal model with functional data, including inference for functional covariates through simultaneous credible intervals. Validated the method on malaria incidence and U.S. COVID-19 datasets.
- Extended autologistic models to identify underlying network structure to predict psychological path of depression symptoms.

## Technical Skills

**Languages:** Python, R, SQL (basic)
**Frameworks & Platforms:** Pytorch, Tensorflow, Scikit-learn, Transformers, Linux, Slurm, HPC clusters, CUDA, Git
**Specialties:** Conformal prediction, active learning, curriculum learning, deep learning, optimization, kernel methods, high-dimensional theory, topic modeling

## Publications

[1] SpeedCP: Fast Kernel-based Conditional Conformal Prediction [Link] [Github]
**Jung, Y. J.**, Liu, Y., Wu, Z., Jeong, S., Donnat, C. (2025), *Under review,* **ICLR 2025**

- Invited talk at Royal Statistical Society International Conference 2025 "Conditional Conformal Prediction On Large Network Models with High-dimensional Data"

[2] Filtering with Confidence: When Data Augmentation Meets Conformal Prediction [Link]
Wu, Z., Jeong, S., Liu, Y., **Jung, Y. J.**, Donnat, C. (2025), *Under review,* **ICLR 2025**

[3] Graph Topic Modeling for Documents with Spatial or Covariate Dependencies [Link] [Github]
**Jung, Y. J.**, Donnat, C. (2025), *Under review,* **JRSS-B**

- **2nd place Poster Award at SNAB 2024**; Poster presentation at APS Global Physics Summit 2025

[4] Fast Bayesian Functional Regression for Non-Gaussian Spatial Data [Link]
Kang, H., **Jung, Y. J.**, Park, J. (2024)
*Bayesian Analysis. 19(2): 407-438*
- Poster presentation at Bernoulli-IMS 2021

[5] Mortality Prediction of Patients in Intensive Care Units Using Machine Learning Algorithms Based on Electronic Health Records [Link]
Choi, M. H., Kim, D., Choi, E. J., ..., **Jung, Y. J.**, Jeong, S. H. (2021)
*Scientific Reports, 12(1), 7180.*

## Presentations

Invited talk at Royal Statistical Society International Conference 09/04/2025

Poster at APS Global Physics Summit 03/17/2025

Poster at Statistical Network Analysis and Beyond (SNAB) 06/16/2024

Poster at Bernoulli-IMS 10th World Congress in Probability and Statistics 07/21/2021

## Awards and Honors

Senior Consultant Winner, University of Chicago 2025

**Poster Award at SNAB 2024 - 2nd place** 2024

**Student Research Poster Day Award - 2nd place**, University of Chicago 2024

Consulting Cup Winner, University of Chicago 2023

Excellent Paper Award at 2021 Brain Korea Academic Conference, Yonsei University 2021

Graduate (Bachelor's-Master's Combined Program) Scholarships, Yonsei University 2020,2021

Chief of National Information Society Agency Award, 2020 NIA Big Data Contest 2020

Brain Korea 21 Scholarships, National Research Foundation of Korea 2020

Semester Highest Honors Scholarship, Yonsei University 2019

Honors Scholarships, Yonsei University 2016, 2018

## Teaching Assistantship

**STAT 24620/32950**: Multivariate Statistical Analysis: Applications and Techniques Spring 2025

**STAT 27400/37400**: Nonparametric Inference Winter 2024, 2025

**STAT 24400**: Statistical Theory and Methods I Autumn, Winter 2023

**STAT 24500**: Statistical Theory and Methods II Spring 2023

## Leadership and Extracurriculars

**Statistical Consulting Center** Yonsei University, University of Chicago
*Consultant* Sep. 2021 - May. 2025
- Provided statistical advice on the appropriate use of data and analysis to researchers in the biomedical, social sciences, and engineering fields.

**Expanded Statistics Club** Yonsei University
*Vice President, Project team leader* Sep. 2018 - Jun. 2019
- Covered advanced theories in *Bayesian Statistics, Dimension Reduction*, and *Statistical Machine Learning* and presented data-driven solutions with applications to real-world data.
- Led 30+ members to actively share statistical insights in an academic session each week.