

# YEH CHAN YOO

Seattle, WA | [yehchanwoo@gmail.com](mailto:yehchanwoo@gmail.com) | [linkedin.com/in/yehchan-yoo](https://www.linkedin.com/in/yehchan-yoo) | [yehchanwoo.github.io](https://github.com/yehchanwoo)

## SUMMARY

Master of Science graduate in Statistics (University of Washington) with a strong foundation in machine learning engineering, scalable data pipelines, and advanced statistical analysis. Expertise in developing and validating Python-based deep learning and Bayesian models for large-scale data (500 GB+), engineering ML/data pipelines on cloud platforms (AWS, Azure, GCP), and applying NLP techniques (RoBERTa, LLMs). Proven ability to lead projects, manage codebase integrity, and translate complex data insights into high-impact solutions across interdisciplinary research and industry settings. Eager to leverage these skills to drive innovation in data science within both industry and academic settings.

## SKILLS

- **Relevant Scripting/Programming Languages:** Python (including its data science libraries -- numpy, pandas, and scikit-learn, etc.), R (including base R and tidyverse), SQL, Kotlin, React/Node.js, HTML, CSS, JavaScript, C/C++, Git, Shell, Slurm
- **Relevant Platforms:** Jupyter Notebook/JupyterLab, RStudio, Microsoft Office (including Excel), Google Analytics, Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, GitHub, Android Studio, Tableau

## EDUCATION

### University of Washington - Seattle (UW)

March 2026

Master of Science in Statistics - Advanced Methods and Data Analysis

- **Relevant Coursework:** Advanced Machine Learning (Python), Advanced Regression Methods for Independent Data (R), Advanced Regression Methods for Dependent Data (R), Applied Regression (R), Applied Statistics Capstone (Python), Design and Analysis of Experiments (R), Machine Learning (Python), Statistical Computing (C/C++)

### University of California - Berkeley

December 2023

Bachelor of Arts in Statistics and Political Economy with a Minor in Data Science

- **Relevant Coursework:** Casual Inference (R); Concepts in Computing with Data (R, Git); Concepts of Statistics (R); Data, Inference, and Decisions (Python); Introduction to Artificial Intelligence (AI) (Python); Introduction to Finance

## PROFESSIONAL EXPERIENCE

### Allen Institute • Visiting Student (Machine Learning Engineer) • Seattle, WA

September 2025 - March 2026

- Engineered, trained, and validated Python-based deep learning-based Bayesian models for large-scale single-cell analysis on 500 GB+ of data
- Managed codebase integrity and documentation with version control on GitHub and testing on AWS, AWS-based platforms, and other high-performance computing clusters

### University of Washington iSchool • HCI Researcher/Technical Lead • Seattle, WA

October 2024 - September 2025

- Led the software engineering effort to develop an Android app product (Kotlin) and a web app product (React/Node.js) for generating emotional messages and stickers through API integration with OpenAI and Google's multimodal AI models
- Contributed to the technical portions of the paper by selecting the right AI models and ensuring that the AI models used would provide safe and applicable results for the research
- Part of a wider research project to test how children aged 6-10 can develop social-emotional skills in a digitalized world in a team of HCI researchers; submitted paper as the third author to the Association for Computing Machinery (ACM)

### Mindful Conversion • Data Scientist • New York, NY

January 2024 - August 2024

- Served as the primary data scientist for the marketing data analytics agency and startup Mindful Conversion, providing leadership and management for the agency's data science projects in a cross-functional team
- Developed and maintained a data engineering and ML pipeline for the agency's predictive SEO product Kixely using Python, Microsoft Azure, and Google Cloud
- Performed data mining and data manipulation on web data from ~100,000 keywords and URLs while engineering and evaluating LLMs to develop predictive models for webpage ranking
- Analyzed and evaluated large data sets using Python and SQL to identify and improve marketing data quality and content strategies for the agency and its clients
- Authored several data-driven reports and presentations that utilized data visualization, technical expertise, domain knowledge, and strong verbal and written communication skills to present data-driven marketing insights to the agency and its clients

### UC Berkeley School of Education • Student (Research) Assistant • Berkeley, CA

June 2023 - December 2023

- Researched, coded, and experimented with deep learning-based natural language processing (NLP) models for automatic short answer grading of 200+ student answers to mathematics and statistics questions
- Achieved 75% test accuracy on autograding tasks with a fine-tuned distilled Transformers-based LLM (RoBERTa), developed with PyTorch
- Assisted with the experimental design and writing of my lab's statistical assessments

### UC Berkeley School of Education • Undergraduate Research Apprentice • Berkeley, CA

January 2023 - May 2023

- Graded and performed statistical analysis on written answers to experimental statistics assessment questions from more than 500 California high school and middle school students using Excel and R, helping researchers refine assessment methodology

### SK Wyverns • Intern • Incheon, South Korea

June 2018 - August 2018

- Examined and visualized patterns in thousands of customer records from the South Korean baseball team SK Wyvern's database using Google Analytics, Python, SQL, and Tableau to analyze app and website usage patterns
- Built a sentiment analysis NLP model using Python to identify emotions in user feedback