

YEH CHAN YOO

South Korea | yehchanwoo.github.io

SKILLS

- **Relevant Coding Languages:** Python (numpy, pandas, pytorch, scikit-learn, transformers, etc.), R, SQL, Git, Shell, C/C++, JavaScript (React, Node.js), Kotlin
- **Relevant Programs/Platforms:** Docker, AWS, Google Cloud Platform (including Vertex AI), Microsoft Azure, Jupyter Notebook, RStudio, GitHub, Android Studio

EDUCATION

University of Washington - Seattle (UW)

March 2026

Master of Science in Statistics - Advanced Methods and Data Analysis

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

University of California - Berkeley (UCB)

December 2023

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

- **Relevant Coursework:** Causal Inference (R); Data, Inference, and Decisions (Python); Introduction to Artificial Intelligence (Python); Modern Statistical Prediction and Machine Learning (R); Principles and Techniques of Data Science (Python, SQL)

PROFESSIONAL EXPERIENCE

Allen Institute for Brain Science • Visiting Student (Machine Learning Engineer) • Seattle, WA

September 2025 - March 2026

- Engineered, trained, and validated Bayesian foundational models with PyTorch-based deep learning module for large-scale single-cell RNA-seq analysis on 500+ GB of genomic data
- Managed codebase integrity, Git-based version control, Shell scripting, Docker containerization, and deployment/monitoring workflows; conducted testing on AWS and institutional HPC clusters
- Contributed as second author to the scVIP preprint, a generative framework integrating transcriptional profiles and phenotypic markers via cell-type-aware multi-instance learning & meant for submission to scvi-tools

scVIP preprint paper: <https://www.biorxiv.org/content/10.64898/2026.04.20.717759v1>

scVIP model codebase: <https://codeocean.allenneuraldynamics.org/capsule/4344243/tree/v3>

Mindful Conversion • Data Scientist • New York, NY

January 2024 - August 2024

- Led and managed cross-functional & client-facing data science efforts from inception to deployment as the primary data scientist for the entrepreneurial marketing analytics agency and startup Mindful Conversion
- Developed and maintained a data pipeline for the agency's predictive SEO product Kixelly using Python as well as the startup's data and machine learning infrastructure on Microsoft Azure, and Google Cloud
- Mined, cleaned, and analyzed web data for Kixelly on ~100,000 keywords and URLs using the data pipeline
- Built, scaled, and evaluated LLMs for Kixelly's webpage ranking tasks
- Analyzed and evaluated large and complex marketing datasets using Python and SQL to identify and resolve complex challenges with marketing data and content strategies for the agency and its clients
- Authored 5+ data-driven reports and presentations that utilized data visualization and strong communication skills to provide insights on marketing strategies 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 LLMs for automatic short answer grading of 200+ student answers to mathematics and statistics questions
- Achieved 75% test accuracy with a fine-tuned distilled RoBERTa model using PyTorch

PROJECTS

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

October 2024 - September 2025

- Led software development efforts with rapid prototyping and iterative experimentation on an Android app (Kotlin) and a web app (React/Node.js) that utilized Google and OpenAI's AI models to generate emotional messages and stickers for kids
- Designed and implemented a backend API integration layer for OpenAI and Google's models, including prompt engineering, response parsing, and real-time content safety detection and filtering, to ensure responsible and ethical AI use
- Contributed to a broader HCI research initiative investigating how children aged 6–10 develop social-emotional skills in digital environments, with the applications deployed for real-world user studies involving child participants

Deep Learning for Digital Pathology: Tumor Detection Using the PCam Dataset

March 2026

- Collaborated on a final project to automate tumor detection in histopathological scans with computer vision models for my Computational Biology course in UW's Allen School of Computer Science & Engineering
- Engineered a custom local VGG-style CNN model & fine-tuned a ResNet18 model on the cloud

Link to our paper: https://yehchanwoo.github.io/attachments/2026/CSE_527_Final_Report_PCam.pdf

Replication of Rec-R1

June 2025

- Reproduced and critiqued results from a preprint publication on Rec-R1, a reinforcement learning framework integrating LLMs with recommendation systems for sequential recommendation and product search, for the final team project in the Advanced Machine Learning course at the University of Washington Allen School of Computer Science & Engineering

Link to our replication paper: https://yehchanwoo.github.io/attachments/2025/CSE_493S_599S_Final_Project.pdf

Link to our project Github repository: https://github.com/yehchanwoo/Rec-R1_magic