Josh Kim

**AI Researcher with background in Data Science and User Experience Research (Mixed-Methods)**

76000 Verano Road, Irvine, CA 92617 [kims17@uci.edu](mailto:kims17@uci.edu)  [Website](https://seungjun-data-science.github.io)  [LinkedIn](https://www.linkedin.com/in/kimseungjun)  [Github](https://github.com/Seungjun-Data-Science)  [Youtube](https://www.youtube.com/channel/UCDBi2xuCCF4qNWCWfyIBUlQ)

# EDUCATION

# University of California, Irvine | Doctor of Philosophy (PhD) in Informatics Irvine, CA | June 2027 | GPA: 3.948

# Advisor: [Dr. Sean D. Young](https://www.faculty.uci.edu/profile/?facultyId=6763) | Mixed-Methods Researcher interested in Artificial Intelligence, Machine Learning, Large Language Models (LLM), chat bots for Predictive Healthcare and serving the underserved and vulnerable populations (e.g. elderly, young adults, patients with rare disease), Responsible AI (Bias-Fairness- Accountability-Transparency of Algorithms), and Health Inequity

# Vassar College | Bachelor of Arts (B.A.) in Mathematics; Minor in Computer Science Poughkeepsie, NY | May 2020 | GPA: 3.80

# Oxford University, St. Catherine’s College | Visiting Students Program in Statistics and Economics Oxford, U.K. | Jan – Jun 2019

# PUBLICATIONS

∙ **Kim, S.**, & Young, S. (2024). LLM based AI curated Playlist more effective for decreasing stress and anxiety among college students than self-selected music? A Pilot Randomized Control Trial. (*Ongoing Study.*)

∙ **Kim, S.**, Hanley, A., Emish, M., & Young, S. (2024). A 20-Minute Mindful Jazz Intervention Decreased Chronic Pain Patients’ Pain and Anxiety 4 Weeks later: Results from a Pilot Randomized Controlled Clinical Trial. *(Under Review.)*

∙ Dong, J., **Kim, S.**, Li, C., Jin, Z., Olivares, C. I., Young, S. (2024). Machine Learning-Based Assessment of Per- and Polyfluoroalkyl Substance (PFAS) Occurrence and Fate in California Wastewater Treatment Plants. (*Under Review.*)

∙ Hassani, M., De Haro, C., Flores, L., Emish, M., **Kim, S.**,Kelani, Z., & Young, S. D. (2023). [Exploring mobility data for enhancing HIV care engagement in Black/African American and Hispanic/Latinx individuals: a longitudinal observational study protocol](https://bmjopen.bmj.com/content/13/12/e079900.abstract).

∙ Garett, R., **Kim, S.**, & Young, S. (2023). Ethical considerations for artificial intelligence applications for HIV. *Artificial Intelligence (AI)*.

∙ Flores, L., & **Kim, S**. (2023). [Addressing Bias in Artificial Intelligence for Public Health Surveillance](https://jme.bmj.com/content/early/2023/05/02/jme-2022-108875). *Journal of Medical Ethics*.

∙ **Kim, S.**, Ugarte, D. A., & Young, S. D. (2023). [Acceptability of an Online Brief Mindfulness Intervention for Sickle Cell Disease Pain](https://www.longdom.org/open-access/acceptability-of-an-online-brief-mindfulness-intervention-for-sickle-cell-disease-pain.pdf). Journal of Hematology & Thromboembolic Diseases.

∙ **Kim, S.** (2023). AI applications for POCUS in low-resource settings: A Systematic Review *(Under Review.)*

# OTHER RESEARCH EXPERIENCE NOT LISTED IN PUBLICATION

**University of California Irvine, Graduate Research Assistant Irvine, CA | Jan – March 2023**

[Adapting the HOPE Online Support Intervention to Increase Uptake of Medications for Opioid Use Disorder](https://clinicaltrials.icts.uci.edu/trial/NCT04712981)

∙ Conducted phone screening of potential patients living in rural areas of California with opioid misuse or addiction issues for a Randomized Control Trial (RCT) evaluating the effects of Peer Leader Interventions for risk reducing behavior. Used a screener script that incorporated the DSM-5 criteria.

**Institute for Social Research, University of Michigan, CJARS, Research Data Engineer Ann Arbor, MI | July 2020 – June 2022**

∙ Objective of the project was to measure the cost of domestic violence but mostly intimate partner violence (IPV) in the U.S.

∙ Wrote scripts in STATA and Python to:

∙ Identify perpetrators of domestic violence based on offense description from the Criminal Justice Administrative Records System (CJARS) database.

∙ Link these offenders with family relations and residence data from the U.S. Census Bureau and implemented logic to identify victims of Intimate Partner Violence (IPV).

**Vassar College Economics Department, Research Assistant Poughkeepsie, NY | May – Dec 2018**

∙ Worked with Dr. Esteban Argudo to describe the impact of the 2017 Repatriation Tax Holiday on loan supply by analyzing 10-K data for fiscal years 09’-17’ on 10,148 unique companies and lending behavior of multinational and domestic banks.

# TEACHING

# University of California, Irvine (UCI) Irvine, CA

# 2022 Fall, 2023 Fall—TA for ICS 139W: Critical Writing on Information Technology

# 2023 Winter—TA for INF 163: Project in the Social and Organizational Impacts of Computing (Qualitative Research Project class for seniors)

# 2023 Spring, 2024 Spring—TA for INF 174: Health Data Analytics (SQL Coding using Electronic Health Records)

# 2024 Fall—TA for INF 191B: Senior Capstone/Design Project

# 2024 Summer—TA for INF 171: Introduction to Health Informatics

# Vassar College Poughkeepsie, NY

# 2018 Spring—Python Tutor; Google Ignite-CS Program (Vassar College Chapter)

# Korean Minjok Leadership Academy (KMLA) South Korea

# 2014 Summer, 2013 Winter, 2011 Winter— English Debate Teaching Assistant; Global Leadership Program for Students (GLPS)

# EXPERIENCE

## **User Experience (UX) Research Intern, Institute for Health Improvement (IHI) Boston, MA | June – Sept 2023**

∙ Performed user research (Microsoft forms survey, Qualtrics, Jamboard, Interviews) and analyzed site use log data to better understand user needs for the organization's internal Knowledge Base (KB).

∙ Created prototypes in Figma for presentation and implemented changes on the KB based on these insights on user needs.

∙ Explored the use of Large Language Model (LLM) powered chatbot within the KB Sharepoint site using Microsoft PowerAutomate with the aim of helping users navigate through the KB and implemented a prototype.

∙ Built schema, frameworks, and system for a shared bibliography library in Zotero, a central repository of references that will be shared across employees who use these bibliographies for grant proposals, research reports, and presentations. Several of these libraries were made by topic area (e.g. Improvement Science, Health Equity, System Resilience, and Patient Safety).

∙ Performed market research on potential competitors in 5 countries that do similar work as IHI to better understand the current strengths and areas for improvement and growth of IHI in comparison to other similar institutions.

**Research Data Engineer, Institute for Social Research, University of Michigan,** [**CJARS**](https://cjars.isr.umich.edu/) **Ann Arbor, MI | July 2020 – June 2022**

∙ Created a system that Extracts, transforms, and loads (ETL) raw criminal justice data in Python and Stata.

∙ Built an automated data quality control program in Python that applies anomaly detection algorithms and custom logic.

∙ Wrote SQL queries that produces recidivism, health and labor outcome statistics of criminal offenders as part of the Django program that generates statistical reports from the first ever U.S. wide criminal justice database our lab built. This became the basis for the Justice Outcome Explorer (JOE) dashboard available [here](https://joe.cjars.org/).

**Data Science Intern, Spotify, Mobile Web Player Team New York, NY | Jun – Aug 2019**

∙ Helped the senior data scientist set up and evaluate A/B testing on the Spotify Wrapped campaign on the Mobile Web Player.

∙ Conducted user cohort analysis in Python on terabytes of user data to better understand user behavior and needs.

∙ Used anomaly detection techniques to identify deep link error and presented solutions to resolve it, resulting in a 5% reduction of affected daily active users (14.8% to 9.2%).

# PROJECTS

∙ [Can Large Language Models (LLMs) Replicate Human Counselor Responses? Dolly v.s. Human Counselors using CounselChat Data](https://github.com/Seungjun-Data-Science/Data-Science-Projects/blob/master/DollyLLM-MentalHealthCounseling/Team_JSJ_FinalProject.pdf)

∙ [Real, Fake and Sarcastic News Detection in a Class Imbalance Setting](https://github.com/Seungjun-Data-Science/Data-Science-Projects/tree/master/CS366-Computational-Linguistics-NLP-Final-Project)

∙ [Potential Gentrification Trends of New York City Airbnbs and Listing Price Prediction Using Machine Learning](https://github.com/Seungjun-Data-Science/Data-Science-Projects/tree/master/MATH301-Data-Science-Senior-Seminar-NYC-Airbnb-Analysis)

# SKILLS

# ∙ Data Analysis/Engineering, DB, Cloud: Python, R, Stata, MySQL, BigQuery, Excel, MLFlow, PySpark, Snowflake, AWS

# ∙ Visualization/BI: Matplotlib, Plotly, Seaborn, Bokeh, D3.js, R GGplot, Javascript, Tableau, PowerBI

# ∙ ML/NLP: Scikit-Learn, TensorFlow, Keras, PyTorch, XGBoost, LightGBM, CatBoost, Large Language Models (LLM), HuggingFace Transformers, Prompt Engineering, PowerAutomate, NLTK, spaCy, TextBlob, Gensim, PyText, FastText, Flair, PyTorch-NLP, BERTopic

# ∙ UX/UI: A/B Testing, Qualtrics, Microsoft Forms, Figma, User Journey Mapping, User Inteviews, Usability Testing, Focus Groups, Participatory Design, Ethnography

# AWARDS & FELLOWSHIP

Best Overall Analysis Award; [American Statistical Association (ASA) DataFest - Vassar College Chapter](https://github.com/Seungjun-Data-Science/Data-Science-Projects/tree/master/2020-Vassar-Datafest-impact-of-COVID) 2020

Santander Customer Transaction Prediction; Ranked 1354 out of 8802 (Top 16%); Kaggle 2019

Vassar College, Tananbaum Fellowship 2019

# CERTIFICATIONS

Activate to Captivate Public Speaking Program Certificate University of California, Irvine (UCI) Graduate Division | Mar 2023

Mentoring Excellence Program (MEP) Certificate University of California, Irvine (UCI) Graduate Division | Mar 2023

# ACTIVITIES & LEADERSHIP & SERVICE

I write articles about Data Science, ML, NLP, LLMs, AI, and UX; [Medium](https://joshnjuny.medium.com/) 2021-Present

English to Korean Translator (Volunteer); PSALT NK (Pray Serve Act Love Truth for North Korea) 2023-Present

International Student Engagement Chair; Associated Graduated Students (AGS) at UC Irvine Irvine, CA | 2022-2023

Mentor; Graduate Interconnect Program (GIC) at UC Irvine Irvine, CA | 2023-Present

Student Leader and Treasurer; Vassar Haiti Project (VHP) Poughkeepsie, NY | Jun 2019 – May 2020

Army Sergeant, Squad Leader, and English Translator; Korean Defense Intelligence Command Anyang, South Korea | Aug 2015 – May 2017

Volunteer; Liberty in North Korea (LINK) Vassar Chapter Poughkeepsie, NY | Sep 2014 – May 2015

Volunteer; Gwan-Ak Community Center Seoul, South Korea | Jul 2007 – Aug 2013

# HOBBIES AND INTERESTS

Cooking; Tennis; Surfing; Walking in hiking trails; Making, listening, and playing music; Checking out new cafes and food places; Watching K-dramas; Hot Tub or Spa; Travelling; Dog Care