# Yufeng Wu

sw20@williams.edu / yfwu02@gmail.com  linkedin.com/in/yufeng-wu/  yufeng-wu.github.io

## EDUCATION

**Williams College** August 2020 - May 2024
*B.A. in Computer Science and Economics* *Williamstown, MA*

- GPA: 4.0/4.0
- Thesis: "Causal Inference With Contagion and Latent Homophily Under Full Interference" advised by Prof. Rohit Bhattacharya [PDF] [defense]
- Honors: *Valedictorian*, *Summa Cum Laude* (top 2% of class), and Highest Honors in Computer Science

## PAPERS

[P2] **Yufeng Wu**, Rohit Bhattacharya. Network Causal Effect Estimation In Graphical Models Of Contagion And Latent Confounding. *4th Conference on Causal Learning and Reasoning (CLeaR)* (Accepted) *Preprint:* arxiv.org/pdf/2411.01371

[P1] (* = equal contribution) Divij Jain*, Saatvik Kher*, Lena Liang*, **Yufeng Wu***, Ashley Zheng*, Xizhen Cai, Anna Plantinga, Elizabeth Upton. Improving and Evaluating Machine Learning Methods for Forensic Shoeprint Matching. *Journal of the Royal Statistical Society Series C: Applied Statistics* (under review). *Preprint:* arxiv.org/pdf/2405.14878

## RESEARCH EXPERIENCE

**Computer Science Thesis in Causal Inference [P2]** [arXiv] [poster] [oral] August 2023 - May 2024
*Thesis Student* (Advisor: Rohit Bhattacharya) *Williams College*

- Invented a causal effect estimation method for non-i.i.d. social network data, enabling previously unfeasible real-world studies.
- Designed an algorithm to verify assumptions required by our method, utilizing a data-driven approach.
- Conducted experiments using synthetic and real-world data, verifying the overall correctness of our approach.
- Gave an oral presentation at the 2024 American Causal Inference Conference. Submitted the paper to CLeaR 2025.

**SMALL REU in Machine Learning [P1]** [arXiv] [poster] [web app] Summer 2023
*Research Assistant* (Advisors: Xizhen Cai, Anna Plantinga, Elizabeth Upton) *Williams College*

- Evaluated robustness of machine learning methods for shoeprint matching, in the context of forensic analysis.
- Published a web app to visualize why a model trained on lab-collected prints fails to generalize to real-world scenarios, revealing the mechanism and harm of distribution shifts in key predictive features that the model relies on.
- Enhanced model generalizability through data augmentation and submitted a journal paper.

**Independent Research in Causal Machine Learning** [technical report] January 2023 - June 2023
*Research Student* (Advisor: Rohit Bhattacharya) *Williams College*

- Developed an algorithm using causal graphs to identify stable parts of ML models across source and target domains.
- Ran experiments showing that our method improves sample efficiency and accuracy during domain adaptation.

**Summer Research in Computer Architecture** [poster] Summer 2022
*Research Assistant* (Advisor: Kelly Shaw) *Williams College*

- Built a C++ tool for workload characterization, analyzing program attributes such as the number of memory accesses.
- Applied the tool to identify hardware stress points in graph mining algorithms executed on large social networks.

## WORK EXPERIENCE

**Charles River Associates** June 2024 - Present
*Analyst* (Antitrust & Competition Economics practice) *Chicago, IL*

- Conduct market competition analysis in fashion, finance, and healthcare industries using large-scale datasets.
- Create visualizations to simplify complex analyses into client deliverables to support litigation arguments.

**Williams College Computer Science Department** February 2022 - December 2022
*Teaching Assistant* *Williamstown, MA*

- Held weekly TA hours, assisted lab sessions, and graded homeworks for 2 courses:
  CSCI 136 Data Structures & Advanced Programming, and CSCI 134 Introduction to Computer Science.

**China World Academy STEM Department** August 2020 - June 2021
*Teacher* *Suzhou, China*

- Prepared my own lecture materials to teach Python to high school students with no prior coding experience.

## PROJECTS

1. Distilling Large Language Models (LLMs) for Twitter Sentiment Analysis (2023) [GitHub] [paper]
   - Distilled a sentiment classifier from an LLM, beating the baseline model trained from scratch by 2.7% in test accuracy.
2. Deep Learning for Video Popularity Prediction (2023) [GitHub] [paper]
   - Built a web-scraper to collect data from 12,000 online videos and trained a neural net to predict video popularity.
3. Impact of Chinese Infrastructure Aid on Education in Kenya: A Causal Analysis (2023) [paper]
   - Conducted difference-in-difference analysis and robustness checks using 18 years of investment and survey data.
4. Islandr App (2020) [GitHub]
   - Designed and launched a web app for 600 students, streamlining campus event exploration and management.

## TALKS & PRESENTATIONS

* indicates equal contribution.

1. Oral presentation (acceptance rate: 68/268 = 25.4%) at American Causal Inference Conference (ACIC), Seattle, WA, May 2024. **Yufeng Wu** and Rohit Bhattacharya. "Causal Inference With Contagion and Latent Homophily Under Full Interference." [slides]
2. Poster presentation at American Causal Inference Conference (ACIC), Seattle, WA, May 2024. **Yufeng Wu** and Rohit Bhattacharya. "Causal Inference With Contagion and Latent Homophily Under Full Interference." [poster]
3. Thesis presentation and defense before the Computer Science department, Williams College, MA, May 2024. [slides]
4. Oral presentation at Promoting Inclusion in Economic Research (PIER) conference, Williams College, MA, April 2024. **Yufeng Wu**. "Impact of Chinese Infrastructure Aid on Education: Evidence from Kenya." [slides]
5. Poster presentation at the New England Statistical Society (NESS)–NextGen Data Science Day, University of Connecticut, CT, October 2023. Divij Jain*, Saatvik Kher*, Lena Liang*, **Yufeng Wu***, Ashley Zheng*, Xizhen Cai, Anna Plantinga, Elizabeth Upton. "Evaluating Machine Learning Methods for Shoeprint Matching." [poster]
6. Poster presentation at Summer Research Fair, Williams College, MA, 2022. Chris Brown*, Sam Chistolini*, Fatima-zohra Guettabi*, Emma Neil*, Kelsey Richter*, **Yufeng Wu***, Kelly Shaw. "Workload Characterization of Graph Algorithms." [poster]

## MENTORSHIP EXPERIENCE

**Williams College Underrepresented Identities in CS** *Board Member* August 2022 - June 2024

- Co-led a mentorship program pairing first-year CS students with upperclassmen and alumni, provided training for both mentors and mentees, and personally advised 6 mentees on career and research opportunities.

**PoKe Project Incubator** *Founder* June 2022 - Present

- Organized original project-based learning (PBL) workshops and mentored 16 students in developing passion projects.

**PEAK English Nonprofit** *Founder* May 2020 - May 2022

- Created 20 educational videos for English learners in China, reaching over 1,800 followers and 190,000 views.

## HONORS & AWARDS

1. *Valedictorian*, Williams College, 2024
2. *Summa Cum Laude* (top 2% of class), Williams College, 2024
3. *Sigma Xi Society*, Williams College, 2024; Inducted into the Sigma Xi Society for excellence in research.
4. *Phi Beta Kappa* National Honor Society Junior Year Inductee (top 5% of class), Williams College, 2023
5. *Best Poster Award*, New England Statistical Society–NextGen Data Science Day Poster Competition, 2023

6. *Dean's List* for all semesters, Williams College, 2020-2024
7. *Davis UWC Scholarship*, 2020-2024

## SKILLS & LANGUAGE

- English (fluent), Mandarin (fluent)
- Python: Machine Learning, Causal Inference, Visualization, Web Scraping, Web Development
- Other Skills: Java, C, C++, R, STATA