Muhammad Haroon

mharoon@ucdavis.edu | (530) 760 9545 | www.muhammadharoon.xyz
haroon96 mharoon_ haroon96

EXPERIENCE

Meta Jun 2025 – Sep 2025
Software Engineer Intern (PhD)
Interned at the Meta Data AI and Superintelligence Labs working on LLM-powered web agents. I designed better task generation strategies incorporating difficulty preferences and devised better metrics for evaluating task generation and task success rates.

University of California, Davis Mar 2021 - Jun 2025
Graduate Student Researcher
Conducted research on privacy and safety harms associated with online systems. I designed and evaluated automated software and machine learning models for auditing online systems.

Lahore University of Management Sciences Mar 2019 - Mar 2021
Research Associate
Worked on privacy preservation and adversarial attacks using ensemble machine learning models and natural language processing for news media analytics. Additionally, I worked on automating workflows for archiving historic content.

Techlogix Jun 2018 - Mar 2019
Software Engineer
Worked as a full-stack developer on HealthCloud - a hospital management system. I improved the nearly hour-long deployment protocol for delivering critical updates to under 10 minutes.

EDUCATION

Ph.D. in Computer Science 2021 - 2025
University of California, Davis

M.S. in Computer Science 2021 - 2023
University of California, Davis

B.S. in Computer Science 2014 - 2018
National University of Computer & Emerging Sciences, Pakistan

PUBLICATIONS

"Whose Side Are You On?" Estimating Ideology of Political and News Content Using Large Language Models and Few-shot Demonstration Selection 2025
Asia-Pacific Chapter of the Association for Computational Linguistics (AACL)
Authors: **M. Haroon**, M. Wojcieszak, A. Chhabra
[preprint]

Re-ranking Using Large Language Models for Mitigating Exposure to Harmful Content on Social Media Platforms 2025
Association for Computational Linguistics (ACL)
Authors: R. Oak, **M. Haroon**, C. Jo, M. Wojcieszak, A. Chhabra
[preprint]

Nudging the Recommendation Algorithm Increases News Consumption and Diversity on YouTube 2024
PNAS Nexus
Authors: X. Yu, **M. Haroon**, E. Menchen-Trevino, M. Wojcieszak
[paper] [website]

Auditing YouTube's Recommendation System for Ideologically Congenial, Extreme, and Problematic Recommendations 2023
Proceedings of the National Academy of Sciences (PNAS)
Authors: **M. Haroon**, M. Wojcieszak, A. Chhabra, X. Liu, P. Mohapatra, Z. Shafiq
[paper] [website] [code]

HARPO: Learning to Subvert Online Behavioral Advertising 2022
The Network and Distributed System Security Symposium (NDSS)
Authors: J. Zhang, K. Psounis, **M. Haroon**, Z. Shafiq
[paper] [code]

Avengers Ensemble! Improving Transferability of Authorship Obfuscation 2021
arXiv:2109.07028
Authors: **M. Haroon**, F. Zaffar, P. Srinivasan, Z. Shafiq
[preprint] [code]

RESEARCH TOOLS DEVELOPED

Personalized Quality News Up-ranker
One of ten finalist algorithms for the Prosocial Ranking Challenge: our algorithm boosts personalized news posts from credible and ideologically diverse news sources to make users more resilient to democratic threats.

YouTube Audit Platform
Identifies the daily top YouTube recommendations for politically left, right, and center users by automatically running several **Selenium**-based sock puppets in **Docker** containers. Web platform built using **Vue**.
[website]

HARPO
Web extension that uses a A2C reinforcement learning model backend built using **PyTorch** to obfuscate the user's online profile.
[code]

ResearchTube
Web extension for YouTube that subjects users to various algorithmic manipulations and observes changes to their content consumption patterns.
[website]

CenterTube
A DQN reinforcement learning model implemented in **PyTorch** that monitors a user's YouTube homepage and systematically manipulates user recommendations towards moderate/neutral news content.

YENET
A Convolutional Neural Network for video frame interpolation.
[code]

LUMS Digital Archive
Developed automated workflows for restoring and cataloging historic documents, and developed a platform in **AngularJS** to allow researchers easy access.
[website]

YouTube/TikTok Driver
A series of libraries for programmatically interacting with YouTube, YouTube Shorts, and TikTok implemented using **Selenium** on web and **Android Debugging Bridge (ADB)** on mobile.
[code]

INVITED TALKS & PODCASTS

Auditing Social Recommender Systems

- Politics and Computational Social Science Conference (PaCSS), Harvard University, 2022
  [slides]
- Summer Institute in Computation Social Science (SICSS), University of Pennsylvania, 2022
  [slides]
- The Backdrop - A Podcast by UC Davis
  [listen]
- UC Davis College of Engineering News
  [read]

Offensive Privacy to Counter AdTech Surveillance

- Nothing to Hide - A Data Privacy Podcast
  [listen]

SKILLS

Languages	Python, C/C++, C#, R, Java, JavaScript, SQL, LaTeX
Frameworks	Docker, Selenium, Playwright, Git, Node.js, Vue.js, AngularJS, React
Libraries	pytorch, scipy, spacy, statsmodels, beautifulsoup, pandas, scikit-learn, matplotlib, seaborn

TEACHING EXPERIENCE

ECS 152A: Computer Networks
Teacher's assistant responsible for designing and conducting weekly discussion sessions, and lectures.

AWARDS & HONORS

Finalists ($6,000) 2024
Prosocial Ranking Challenge

GGCS Summer Research Fellowship ($18,000) 2024
University of California, Davis

GGCS Summer Research Fellowship ($18,000) 2023
University of California, Davis

GGCS Spring Research Fellowship ($7,177) 2023
University of California, Davis

Dean's Honor List 2018
National University of Computer & Emerging Sciences, Pakistan

REFERENCES

Magdalena Wojcieszak
Professor of Communication
University of California, Davis
mwojcieszak@ucdavis.edu

Zubair Shafiq
Associate Professor of Computer Science
University of California, Davis
zubair@ucdavis.edu

Fareed Zaffar
Associate Professor of Computer Science
Lahore University of Management Sciences
fareed.zaffar@lums.edu.pk

Anshuman Chhabra
Assistant Professor of Computer Science
University of South Florida
anshumanc@usf.edu