




# Muhammad Haroon

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

 haroon96  mharoon\_  haroon96

## EDUCATION

<b>Ph.D. in Computer Science</b> University of California, Davis	2021 - Present
<b>M.S. in Computer Science</b> University of California, Davis	2021 - 2023
<b>B.S. in Computer Science</b> National University of Computer & Emerging Sciences, Pakistan	2014 - 2018

## EXPERIENCE

<b>Meta</b> <i>Software Engineer Intern (PhD)</i> Internship at the Data AI Labs at Meta working on AI agents for automated web crawling.	2025 - Present
<b>University of California, Davis</b> <i>Graduate Student Researcher</i> <i>Advisor: Prof. Magdalena Wojcieszak</i> Research on privacy and safety harms associated with online systems. I design and evaluate automated software techniques and machine learning models for auditing online systems.	2021 - Present
<b>Lahore University of Management Sciences</b> <i>Research Associate</i> <i>Advisor: Prof. Fareed Zaffar</i> 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.	2019 - 2021
<b>Techlogix</b> <i>Software Engineer</i> 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.	2018 - 2019

## PUBLICATIONS

<b>Whose Side Are You On? Estimating Ideology of Political and News Content Using Large Language Models and Few-shot Demonstration Selection</b> <i>arXiv:2503.20797</i> Authors: <b>M. Haroon</b> , M. Wojcieszak, A. Chhabra <a href="#">[preprint]</a>	2025
<b>Re-ranking Using Large Language Models for Mitigating Exposure to Harmful Content on Social Media Platforms</b> <i>Association for Computational Linguistics (ACL)</i> Authors: R. Oak, <b>M. Haroon</b> , C. Jo, M. Wojcieszak, A. Chhabra <a href="#">[preprint]</a>	2025
<b>Nudging the Recommendation Algorithm Increases News Consumption and Diversity on YouTube</b> <i>PNAS Nexus</i>	2024

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. [\[code\]](#)

## **OTHER PROJECTS** *(can be found on my GitHub among other projects)*

---

### **ASL Handsigns Recognition**

A Convolutional Neural Network implemented in **PyTorch** with a **Tkinter** frontend that recognizes ASL handsigns in real-time using the system webcam.

## Scholar Lens

A VSCode Extension that auto-fills BibTeX entries using Semantic Scholar's API. Overleaf version in-progress.

## Discord Stats Bot

Developed a Discord bot for video game statistics calculation that organically grew to over 200+ Discord servers.

## 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*

**Platforms/Frameworks:** *Docker, Selenium, Git, Node.js, Vue, Android Debugging Bridge (ADB)*

**Libraries:** *PyTorch, SciPy, spaCy, BeautifulSoup, Pandas, Scikit-learn, Matplotlib, Seaborn, Plotly, Tkinter*

## TEACHING EXPERIENCE

---

### ECS 152A: Computer Networks

Teacher's assistant responsible for designing and conducting weekly discussion sessions, and lectures.

## AWARDS & HONORS

---

<b>Finalists (\$6,000)</b> Prosocial Ranking Challenge	2024
<b>GGCS Summer Research Fellowship (\$18,000)</b> University of California, Davis	2024
<b>GGCS Summer Research Fellowship (\$18,000)</b> University of California, Davis	2023
<b>GGCS Spring Research Fellowship (\$7,177)</b> University of California, Davis	2023
<b>Dean's Honor List</b> National University of Computer & Emerging Sciences, Pakistan	2018

## 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