SHAOOR MUNIR

3rd YEAR PH.D. STUDENT IN COMPUTER SCIENCE (MACHINE LEARNING, PRIVACY, TECHNOLOGY POLICY)

✉ smunir@ucdavis.edu | **in** shaoormunir | ○ shaoormunir | ⚲ shaoormunir.com

📱 +1 (530) 953-3757 | ➤ Davis, California, USA

EDUCATION

University of California, Davis	Davis, US
Doctor of Philosophy in Computer Science	*September 2021 – Present*
National University of Computer and Emerging Sciences	Lahore, PK
Bachelors in Computer Science	*August 2014 – May 2018*

EXPERIENCE

University of California, Davis	**Advisor**: Prof. Zubair Shafiq
Graduate Student Researcher	*September 2021 – Present*

- Research to evaluate existing privacy-preserving technologies and development of new techniques to counter advancements in online tracking.
- Analysis of existing technology policies and how they impact online privacy

Lahore University of Management Sciences	**Advisor:** Prof. Fareed Zaffar
Research Assistant	*November 2019 – August 2021*

- Worked in collaboration with academics from The University of Iowa and UC Davis to explore state of the art in Natural Language Generation and Natural Language Understanding
- Worked on exploring automated methods to attribute a synthetic text to its source language model
- Collaborated with external and internal resources to work on projects related to news event detection and Urdu language text corpus

Northbay Solutions	
Software Engineer	*April 2019 – November 2019*

- Worked on an internal Big Data Research Project making use of Kylo, Nifi, Apache Spark, and Amazon Web Services to create an end-to-end Data Orchestration and Transformation tool for organizations dealing with large amount of unstructured data
- Worked as part of team designing core framework with Java API endpoints for Data Lake management in Amazon Web Services

Systems Limited	
Associate Software Engineer	*July 2018 – September 2018*

- Worked as a trainee in the Machine learning and Data Science team
- Worked on a R&D project for an American bank, making use of Random Forests, SVM, Word2Vec, Doc2Vec to automatically categorize banking forms and extract useful data

TechJuice	
Editor	*October 2018 – April 2019*

- Managed a writer panel consisting of 15 writers, spread across Lahore, Islamabad and Karachi
- Worked towards improving the editorial quality, hired and trained new writers to expand the existing team

Content Writer	*November 2016 – April 2018*

- Wrote and published more than 450 articles with one of the largest technology media organizations in Pakistan
- Organized, managed, and carried out interviews for TechJuice 25 under 25 2017 and 2018 editions
- Worked closely with major companies in Pakistan like PTCL, Jazz, Telenor, and many more for strategic content production

Ongoing Research

Etrigan: Large Scale Adversarial Analysis of Elusive Bots (Ongoing research)
- Large scale analysis of bots that evade commercial bot detection services
- Application of adversarial machine learning approach to evade commercial bot detection services
- Analysis of inconsistencies in bot fingerprints to propose enhanced detection by these services

Blocking Tracking JavaScript at the Function Granularity (Conditionally Accepted at CCS 2024)
- Analysis and detection of mixed JavaScript code (scripts including both tracking and functional resources)
- Working on creating an ML model to automatically detect which methods in a mixed script are tracking and functional, to allow for easy blockage of tracking resources
- Ongoing work in collaboration with researchers from Virginia Tech

PURL: Safe and Effective Sanitization of Link Decoration (Under Submission, code)
- Analysis of the use of link decorations for tracking purposes in the face of third-party cookie blockage.
- Scope of work includes measurement of the extent of tracking, and creation of a machine learning model to help filter out tracking link decorations from functional parameters
- Ongoing work in collaboration with researchers from Imperial University and Washington University in St. Louis

Published Research

Google's Antitrust Paradox (Accepted at Vanderbilt JETLaw)
- Research to evaluate the influence of major tech companies on privacy policies
- Evaluating their privacy policies, the impact of privacy restrictions enacted by them on their competition, and categorizing violation of existing anti-competitive regulations
- Proposing new regulations to reduce this influence, curtail monopolistic practices and advocate fair competition
- Ongoing work in collaboration with researchers from Maastricht Uni

COOKIEGRAPH: Understanding and Detecting First-Party Tracking Cookies (ACM CCS 2023, code)
- Research on how third-party scripts running in a first-party context can misuse this privilege and, combined with other tracking techniques like fingerprinting and first-party identifiers, can use this privilege to track users across different sites in the presence of restriction implemented by modern web browsers on third party cookies

Through the Looking Glass: Learning to Attribute Synthetic Text (EACL 2021)
- Analysis of different classification techniques to learn how to attribute text generated by a language model, and the effect of using different parameters and fine-tuning on this attribution

Projects

Makhzan (Urdu Text Corpus)
- Helped build an open-source standardized Urdu Text corpus to help advance Natural Language Processing in Urdu language.

Writeprints (Python Package)
- A Python library to extract features from text. The extracted features can be used to fingerprint text to train attributors.

K2 | *Java, AWS* (Data Lake Solution)
- Helped develop an end-to-end Data Lake solution with a plug-and-play Java API for end users

Intelligent Mock Interviewer | *Python, Django, OpenCV* (Final Year Project, code)
- An intelligent interviewer which used OpenCV, SVMs, and Neural Networks to analyze an interview's audio and video to judge the interviewee's personality. The final assessment of the interview was based on the Big Five personality scale - one of the most commonly used scales in professional environments

Talks and Media Coverage

- **IMDEA Networks:** presented a talk on safeguarding user data from storage and exfiltration by trackers
- **The Register:** covered online tracking through link decoration and detection of these link decorations
- **Ad-Filtering Dev Summit 2023:** presented work on automatically detecting tracking link decorations
- **DataSkeptic Podcast:** discussed first-party cookie-based tracking and countermeasures
- **Ad-Filtering Dev Summit 2022:** presented an overview of the COOKIEGRAPH paper
- **EACL 2021:** presentation on attribution of text generated by language models

Technical Skills

- **Programming Languages**: Python, C/C++, Java

- Experience working with Machine Learning models, AI algorithms and Data Science libraries for data analysis

- Experience working with Amazon Web Services and working knowledge of other cloud services

References

Zubair Shafiq	University of California, Davis
Associate Professor of Computer Science	*zubair@ucdavis.edu*
Fareed Zaffar	Lahore University of Management Sciences
Associate Professor of Computer Science	*fareed.zaffar@lums.edu.pk*
Carmela Troncoso	EPFL
Tenure Track Assistant Professor	*carmela.troncoso@epfl.ch*
Sandra Siby	New York University, Abu Dhabi
Tenure Track Assitant Professor	*s.siby@imperial.ac.uk*