Talha Abrar

✉ talhaabrar11@hotmail.com | in ⏻

EDUCATION

Lahore University of Management Sciences Lahore, Pakistan
Bachelors in Computer Science *Aug. 2020 – Present*

- **Relevant Courses**: Topics in Computer and Network Security, Network Secuirty, Operating Systems, Deep Learning, Machine Learning

Lahore Grammar School Johar Town Lahore, Pakistan
GCSE O and A Level *Aug. 2015 – May 2020*

RESEARCH

Anomaly Detection via Provenance July 2023 – Present
Lahore University of Management Sciences *Lahore, Pakistan*

Goal: Devise a system that uses provenance data to show it does better accuracy anomaly detection than raw audit logs

Collaborators: Dr. Thomas Pasquier (University of British Columbia), Dr. Ashish Gehani (SRI International), Dr. David Eyers (University of Otago), Dr. Xueyuan Michael Han (Wake Forest University), Dr. Fareed Zaffar (Lahore University of Management Sciences)

- Compared two state-of-the-art anomaly detection systems: *ShadeWatcher*, which is a provenance-based system and *AirTag*, which is an event-log based system.
- Reverse engineered *ShadeWatcher* from its description in a paper.
- Introduced a novel, hybrid strategy that combined event-log data with provenance, to mitigate interference from inter-system variations for comparisons.
- Created a mutated dataset for *AirTag*.
- Worked with *SPADE*, a handy tool for capturing and querying provenance data – something we used to visualize the *ShadeWatcher* dataset

Audio Deepfake Detection Aug. 2022 – May 2023
Lahore University of Management Sciences *Lahore, Pakistan*

Goal: Use a novel dataset to create and detect Urdu Audio Deepfakes

Collaborators: Dr. Agha Ali Raza (Lahore University of Management Sciences), Dr. Ihsan Ayyub Qazi (Lahore University of Management Sciences)

- Involved in the Deepfake creation part of this Facebook-funded project. Created a novel Urdu dataset for the purpose of training large Text-to-Speech (TTS) models. Collaborated with multiple seniors to collect speaker data on this curated dataset.
- Fine-tuned and trained two end-to-end Deep Learning models, coquiTTS and Tecatron on collected Urdu speaker data. Generated realistic deepfakes from it.
- Managed the human evaluation process post-training, soliciting feedback on distinguishing human and deepfake audio for model improvement.

SELECTED PROJECTS

Early Warning System for Forest Fires | *Python, Pytorch, BeautifulSoup, LaTeX* Jan 2023 – May 2023
- Did a literature review on the use of CNNs for detecting forest fires.
- Led a group of 3 people in developing a CNN model for image classification in PyTorch.
- Wrote a project paper on our implementation.

MediNav | *MongoDB, Figma, Javascript, Git* Jan 2023 – May 2023
- Designed a MERN Stack Web App from scratch

- The app aims to improve ambulance response times by connecting existing systems of traffic redirection, hospitals' emergency services, and accident reporting.
- Designed the interface on Figma, used Trello Boards for Scrum meeting and Postman for automated API testing.

Twitter Authorship Attribution | *Python, BERT Embeddings*   Sep. 2022 – Dec. 2022
- Implemented a system for Authorship Attribution of tweets.
- Combined various ML models including KNNs, NN, and different Ensemble methods like Bagging and Boosting.
- Addressed a range of theoretical problems including understanding why certain models performed better than others, and also provided recommendations for enhancing the feature extraction and data processing methods.

EXPERIENCE

Research Intern   June 2022 – August 2022
CSaLT Lab, Lahore University of Management Sciences   *Lahore, Pakistan*
- Designed a comprehensive phonetically rich Urdu dataset specifically tailored for training deep learning TTS models.
- Acquired proficiency in utilizing Praat for recording text within this dataset and lead an effort for gathering speaker data from external sources.
- Conducted data cleaning procedures, standardizing all audio samples to a consistent frequency, eliminating erroneous recordings, and requesting speakers to resend missing samples.

PROGRAMMING SKILLS

Languages: Python, C/C++, JavaScript, Haskell
Frameworks: React, Node.js, PyTorch
Developer Tools: Git, Docker, Vagrant

ACHIEVEMENTS

Dean's Honor List (2021, 2022, 2023): Placed on the DHL at LUMS for academic achievement in all 3 years

Merit Scholarship (2019, 2020): Awarded **100%** Merit scholarship for both years of A levels

ARSSDC (2018): Reached **regional (Asian) semifinals** of ARSSDC, a space settlement design competition

IYMC (2018): Was a **finalist** at the Internation Youth Math Challenge, one of the biggest math competitions

NASA Space Camp (2017): Secured **1st place** at the Arduino competition, Kennedy Space Centre