ANIQUE TAHIR

(480) 432-9104 | artahir@asu.edu | Github: aniquetahir | LinkedIn: aniquetahir | https://anique.org

SUMMARY

- Prolific scientist with skills and experience in Machine Learning, Data Mining, and Distributed ML Implementation.
- Led collaborative projects, resulting in peer-reviewed publications at top CS venues with over a hundred citations.
- Thorough prolific understanding of machine learning models, deep-learning, out-of-distribution inference, causal inference, uncertainty quantification, and distributed database systems.
- Self-motivated, problem-solving and collaborative scientist with excellent communication skills.
- Contributed to research projects which ended up as successful startups (Wherobots, CRY3CON).

TECHNICAL SKILLS

Programming Languages. Javascript/Node • Python • Java • C • C++ • C# • PHP • BASH
Frameworks. jax • pytorch • tensorflow/keras • numpy • pandas • matplotlib • scikit-learn • webpack • Apache Spark • Hadoop • React • MapReduce • selenium
Utilites. Github • Pupeteer • Docker • Nginx • qemu/kvm
LLMs Encoder Models (BERT, RoBERTa) • Decoder Models (GPT 2, Llama 1/2) • Finetuning (LoRA, Instruction tuning) • Constrained Generation (guidance, lmql) • Prompting (CoT, exemplars, RAG)

EDUCATION

- PhD, Computer Science, Arizona State University, 2019-present (expected graduation: April 2025)
- MS, Computer Science, Arizona State University, 2018
- BS, Computer Science, Lahore University of Management Sciences, 2012

EXPERIENCE

Amazon **Palo Alto, CA**
Applied Scientist Intern *May 2024 - Aug 2024*
- Implemented a LLM based reranking algorithm for PECOS. My contributions were open-sourced (amzn/pecos).
- Created a new approach for LLM efficient reranking which substantially improved efficiency by reducing memory footprint to $\sim 50\%$ and improving speed by $\sim 40\%$.

Arizona State University **Tempe, AZ**
Research Assistant *Aug 2020 - present*
- Published methods for tackling the Algorithmic Bias and Distributional Shift problems in top CS conferences and journals (ACL, CIKM, ICDE, and more).
- Led Data Mining and Machine Learning (DMML) lab as lab manager.
- Prolific skillset demonstrated by delivering research objectives funded by Merck, NSF, AFRL and ONR.

Arizona State University **Tempe, AZ**
Graduate Teaching Associate *Spring/Fall 2019, Spring/Fall 2020*
- Graduate Teaching Associate for CSE 340: Principles of Programming Languages.
- Helped students with reinforcing course material by delivering recitation lectures.
- Created a Gradescope/Canvas(LMS) based autograder for C++ projects which provides instant feedback to students regarding their progress in course projects.

Arizona State University **Tempe, AZ**
Instructor - CSE 205 *Summer 2020*
- Selected to instruct Computer Science students for the Summer semester during the COVID-19 pandemic based on previous record as a Teaching Assistant

Arizona State University **Tempe, AZ**
CIS Research Aide *May 2019 – Aug 2019*
- Successfully implemented Deep Learning models on cloud infrastructure to predict successful and unsuccessful crowd sourced projects on the internet through deep learning models for video/image analysis

Arizona State University **Tempe, AZ**
Graduate Services Assistant *May 2016 – Aug 2016*
- Delivered software automation tools which used Selenium and Mechanize to mine data from the Dark Web
- Successfully led a team tasked with creating automated systems for crawling and parsing dark web data where we evolved the system to handle a large part of the dark web
- Participated in the provisional patent related to our work and one of the first people to work on the project which later became the successful startup named *CRY3CON*

Sofizar/ConstellationCK
Software Engineer (fullstack) *Apr 2013 – Jun 2015*

SELECTED PUBLICATIONS

- Anique Tahir, Lu Cheng, and Huan Liu. "JORA: JAX Tensor-Parallel LoRA Library for Retrieval Augmented Fine-Tuning" 2024 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024).
  We demonstate our library for fine-tuning large language models which shows ∼12x improvement over HuggingFace/Deepspeed by utilizing tensor-parallelism and jit compilation.
- Tahir, Anique, et al. "Evaluating LLMs Capabilities Towards Understanding Social Dynamics" 2024 16th International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
  We evaluate LLMs abilities when presented with prompts containing social dynamics. The study involves highlighting language understanding capabilities, evaluation metrics, and directionality analysis. We also study the affects of fine-tuning and CoT prompting to improve responses.
- Anique Tahir, Lu Cheng, and Huan Liu. "Fairness through Aleatoric Uncertainty." 2023 32nd ACM International Conference on Information and Knowledge Management (CIKM).
  We address the problem of systematic bias in machine learning methods by viewing their predictions from an uncertainty dichotomization perspective. Confident predictions by generalized models are unlikely to be fair, while uncertain predictions require attention.
- Ujun Jeong, Paras Sheth, Anique Tahir, and Faisal Alatawi, H Russell Bernard, and Huan Liu. "Exploring platform migration patterns between twitter and mastodon: A user behavior study." 2024 18th International AAAI Conference on Web and Social Media (ICWSM).
  Why do users migrate between social media platforms. In this study, we do a comprehensive evaluation based on user migrations between Twitter and Mastodon. We outline social factors and preferences while also accounting for different types of individuals.
- Anique Tahir, Lu Cheng, Ruocheng Guo, and Huan Liu. "Distributional Shift Adaptation using Domain-Specific Features," 2022 IEEE International Conference on Big Data (Big Data).
  Machine Learning models trained through Empirical Risk Minimization face the issue of overfitting on the training distribution. In this work we outline a novel approach where the models learn patterns in the unlabelled production data and improvise.
- Raha Moraffah, Paras Sheth, Mansooreh Karami, Anchit Bhattacharya, Qianru Wang, Anique Tahir, Adrienne Raglin, and Huan Liu. "Causal inference for time series analysis: Problems, methods and evaluation." Knowledge and Information Systems (2021): 1-45.
  A survey around the use of causality for time series data. Observational data does not allow interventional analysis. Causal frameworks aim to study the affect of counterfactuals by making assumptions. The problem is enhanced in the setting of time-series.
- Anique Tahir, Yuhan Sun, and Mohamed Sarwat. "An Automated Framework for Explaining Facts Extracted From Mobility Datasets." 2019 20th IEEE International Conference on Mobile Data Management (MDM). IEEE, 2019.
  We propose a systematic approach for explaining observations made on big data during exploratory analysis. The explanations take the form of predicates which explain the significance of an observation.
- Jia Yu, Anique Tahir, and Mohamed Sarwat. "GeoSparkViz in action: a data system with built-in support for geospatial visualization." 2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019.
  Big Data is challenging to visualize due to the sheer amount of processing involved in aggregating information. We demonstrate a system designed to make this process convenient.
- Faisal Alatawi, et al. "A Survey on Echo Chambers on Social Media: Description, Detection and Mitigation." arXiv preprint arXiv:2112.05084 (2021).
  A survey which studies the formation of echo chambers, as well as methods to detect and simulate them. In addition, we outline different mitigation strategies presented in literature.

Preprint

- Wang, Qihan, et al. "Exploring Musical, Lyrical, and Network Dimensions of Music Sharing Among Depressed Individuals." arXiv preprint arXiv:2310.11557 (2024).
  Music can serve as an important tool to cope with feelings. In this work, we study the listening habits of depressed individuals.

Google Scholar: https://scholar.google.com/citations?user=6b8YRBQAAAAJ&hl=en&oi=ao

AWARDS AND HONORS

- **Google** Developer Expert 2024
- **SDM** 2022 Student Scholarship
- **ASU** SCAI Travel Award 2022
- **SBP-BRiMS** 2022 Travel Award

SERVICE

- Invited Talk - ASU GPU Day 2024
- Speaker - ASU AI Day 2023

- Subreviewer - IEEE BigData 2024
- PC Member - AAAI 2023 Conference, AAAI 2024 Conference
- PC Member - BESC 2023 Conference
- Volunteer - ASU Open Door 2023
- Volunteer - SBP-BRiMS 2022 Conference
- Subreviewer - SBP-BRiMS 2022 Conference
- Volunteer - WSDM 2022 Conference
- Subreviewer - SIGIR 2022
- Volunteer - KDD 2021 Conference
- Subreviewer - IEEE CogMI 2021
- Volunteer - CySIS lab - Arizona State University

MENTORSHIP

Honored to mentor the following students:

- Qihan Wang - Peking University - Student
- Ariz Chang - Arizona State University - Amazon/ConnectWise
- Manlin Zhang - Arizona State University - DBSI-Inc

PROJECTS

My publicly available open-source projects and contributions can be found here:
https://github.com/aniquetahir