Zohaib Khan

+1 (571) 524-4448 | zohaib.khan3502@gmail.com | Personal Website/Blog | LinkedIn | GitHub

Enthusiastic and driven AI researcher transitioning into engineering roles, with a strong foundation in Machine Learning and Applied AI. Eager to learn and apply skills to build innovative, scalable and practical solutions in real-world settings.

Experience

Research Associate Jan. 2022 – Present

Lahore University of Management Sciences

- Developed an advanced mechanism to enhance Group Query Attention (GQA) in Transformers, making the static head grouping more dynamic based off L_2 norms of Keys, leading up to an 8% improvement in Image Classification performance (WIP Paper, PyTorch Code).
- Designed methodologies and experiments to probe LLMs for social biases, leveraging synthetic data and analyzing win rates of outputs tied to social groups (Paper).
- Architected large-scale inference pipelines with translation and human-in-the-loop components, and evaluated several LLMs (including OpenAI's GPT family and Meta's Llama-2 family) for fact-checking and multilingual generalization on a large scraped dataset over multiple prompting strategies.
- Managed logistics, supervised student teams on research projects, and worked on inter-university collaborations as a Lab Lead for CSaLT at LUMS.
- Currently supervising projects on researching improvements to Knowledge Distillation and Domain Generalization, leveraging principles from Adversarial Training and Gradient Harmonization.

Fatima Fellow Sep. 2024 – Present

Fatima Fellowship

- Collaborated with undergraduate and graduate students around the world to investigate Multilingual Redteaming for LLMs.
- Implemented inference pipelines for multilingual LLMs, analyzing rates of compliance with potentially hazardous prompts and linking with retrieval mechanisms to study safety alignment.
- Manually scraped and cleaned tweets to curate a dataset related to culturally specific entities to evaluate how culturally-sensitive model completions were in filling masked prompts.

Head Teaching Assistant Sep. 2023 – Dec. 2024

Lahore University of Management Sciences

- Led as the Head TA for graduate courses, including Machine Learning, Speech and Language Processing, and Advanced Topics in Machine Learning, supporting diverse cohorts of up to 220 students.
- Co-designed and launched two innovative graduate courses, introducing cutting-edge topics such as large language models, domain generalization, and transfer learning.
- Managed teams of up to 12 TAs, ensuring smooth course delivery through effective coordination, mentoring, and development of high-quality instructional materials.

Machine Learning Engineer May 2023 – Sep 2023

ISSM.ai

- Designed and implemented a robust pipeline integrating OCR and QR code detection for accurate invoice number extraction from diverse receipt formats.
- Comprehensively studied and rewrote the docTR repository in PyTorch to train custom Text Detection and Recognition models on a proprietary dataset of over 15,000 annotated images, enhancing system adaptability.
- Achieved a 5x reduction in latency by optimizing final models for Intel CPUs using ONNX and OpenVINO, delivering processing speeds of 350ms per image without sacrificing output quality.

Deep Learning Intern June 2022 – March 2023

CodeSlash

- Developed an Object Tracking system leveraging YOLOv5 and StrongSORT to monitor and track cattle in satellite imagery, achieving a notable mean Average Precision (mAP) of 0.542.
- Optimized a Face Recognition system for CPUs using OpenVINO, MongoDB, FastAPI, and Docker, leading to a 3x improvement over the original implementation.

Education

Lahore University of Management Sciences Lahore, Pakistan

BS in Computer Science, CGPA: 3.82, SCGPA: 3.94 (High Distinction) *Aug. 2020 – May 2024*

Skills

Programming Languages and Tools: Python, C, C++, Linux, Excel, Docker
Machine Learning: PyTorch, TensorFlow, JAX, Keras, Transformers, LangChain
Data Engineering: Spark, Iceberg, Airflow, Pandas
Web: HTML/CSS/JavaScript, React, Node.js, MongoDB, Express, FastAPI, MySQL
Soft Skills: Clear Communication, Leadership, Team Management and Collaboration, Conflict Resolution
Other: Quantization/Pruning, Finetuning, Safety Alignment, Domain Generalization