Abdullah Hashmat

Web Page • abdullah.hashmat02@gmail.com • LinkedIn • GitHub • +92 3035551871

EDUCATION

Lahore University of Management Science - *BS Computer Science* *Aug 2021-May 2025*

Relevant Coursework: *CS6304* Advanced Machine Learning, CS*5302* Speech Processing with Generative AI, CS*6303* Large Language Models, *CS331* Artificial Intelligence, *CS487* Cloud Development, *CS437* Deep Learning, *CS331* Principles and Techniques of Data Science

RESEARCH EXPERIENCE

Research Assistant – *German Research Center for Artificial Intelligence (DFKI) & RPTU Kaiserslautern* *July 2025– Present*

- Developed a multi-representation knowledge-distillation pipeline that compresses chemical granularity encoded by SMILES-trained BERT/RoBERTa/DeBERTa teachers (atom, functional-group, branch levels) into a single student model for molecular property prediction.
- Adapted S4, Mamba, and Hyena state-space models with SMILES aware byte pair encoding tokenizers and structure aware embeddings to improve chemical sequence modeling, achieving SOTA on a 22-task ADMET benchmark.

Research Assistant – Centre for Urban Informatics, Technology, *and Policy (CITY) Lab – LUMS* *Jan 2025 – May 2025*

- Developed confidence-aware knowledge distillation loss functions to balance shape and texture biases in image classification, enhancing OOD generalization and outperforming single/multi-teacher baselines.

Research Assistant – ***AI in Healthcare Initiative (AIHI) Lab – LUMS*** *May 2024 – May 2025*

- Worked on Sehat Sanjha, an AI-powered physician assistant, leveraging reinforcement learning and multimodal inputs (speech, text, patient imaging), and developed representation-learning frameworks with code-switching and prompt-based transcription techniques to unify diverse data streams for real-time triage and clinical decision support.
- Facilitated backend development using React, Redux, and Lambda functions, and integrated **100+** refined test cases
- Enhanced RL model performance through feedback loops, integrating physician patient interaction data to refine decision making policies for improved accuracy and reliability.

Research Assistant – *Center for Speech and Language Technologies (CSaLT) Lab – LUMS* *Sept 2023 – May 2025*

- Co-first authored PakBBQ (EMNLP 2025 Main Conference, A*), the first culturally adapted bias benchmark for Urdu/English with 17,180 QA pairs across 8 Pakistani socio-cultural dimensions
- Led a 4-member research team on LLM reasoning using sparse autoencoders and expanded the PakBBQ dataset to additional regional languages, covering underrepresented linguistic communities.
- Researched **10+** Large Language Models (LLMs) for low-resource languages, focusing on vulnerabilities to bias and jail breaking, resulting in several lab publications.

TEACHING EXPERIENCE

Teaching Assistant – *AI600 Machine Learning (Masters in AI)* *Jan 2025 – May 2025*
Teaching Assistant – *CS5302 Foundations of Generative AI* *Jan 2025 – May 2025*
Teaching Assistant – *CS535 Machine Learning* *Sept 2024 – Dec 2024*
Teaching Assistant – *CS100 Computational Problem Solving* *Sept 2023 – Dec 2023*

PUBLICATIONS

Abdullah Hashmat, Muhammad Arham Mirza, Agha Ali Raza, *PakBBQ: A Culturally Adapted Bias Benchmark for QA*. In the Main Conference of Empirical Methods in Natural Language Processing: EMNLP 2025, Suzhou, China, November 5-9, 2025. (Core: A*, H5-index: 218)

PROFESSIONAL EXPERIENCE

Delivery Hero (FoodPanda) – ***Sales Performance Intern*** *Jun 2024 – Aug 2024*

- Conducted large-scale EDA on vendor onboarding and QC data to identify trends and bottlenecks, streamlining the self-signup process and reducing QC rejection rate by **15%.**
- Developed an agentic workflow using Python and Google APIs to automatically pull daily data, generate invoices, and distribute them to vendors via WhatsApp, reducing manual effort and process time by **35%.**
- Automated data integration from Google Drive to Sheets by building ELT pipelines, enabling systematic analysis and increasing hourly throughput by **25%.**

Pakistan Television Network – *Data Science Intern* *Jun 2023 – Jul 2023*

- Revamped and enhanced PTV World's ticker headline system using machine learning models for text similarity, resulting in **20%** reduction in headline overheads
- Developed a predictive analysis model at Pakistan Television Network using Data Science techniques, enhancing viewer engagement metrics through data-driven content strategies by **10%**

RESEARCH PROJECTS

Multi-Representation SMILES-Based Molecular Property Prediction Model *July 2025 – Present*

- **Role:** Lead Researcher. Supervisor **Dr. Nabeel Asim (RPTU, DFKI)**
- **Research Question:** How can multi-granular molecular representations be leveraged to train robust models and distill them into a single student for downstream property prediction?
- **Objective:** To develop a unified student model integrating atom, functional-group, and branch level knowledge from

pretrained transformer models to achieve state-of-the-art molecular property prediction on ADMET 22 dataset.

- **Methodology:** Trained three specialized transformer models (BERT, RoBERTa, DeBERTa) on tens to hundreds of millions of SMILES strings using distinct tokenizers: atom-based for fine-grained structure, functional-group-based for chemical functionality, and branch based for molecular topology. Applied Masked Language Modeling (MLM) and FitNet knowledge distillation to a unified BPE based student model. Incorporated dynamic gating mechanisms to adaptively weight teacher contributions per chemical substructure, ensuring comprehensive multi granular representation for downstream molecular property prediction.
- **Outcome:** Achieved consistently strong performance across the 22-task ADMET benchmark, outperforming existing models that typically excel on only a subset of tasks.

PakBBQ: A Culturally Adapted Bias Benchmark for QA Link *Sept 2024 – May 2025*

- **Role:** Primary Researcher. Supervisor **Dr. Agha Ali Raza (LUMS)**
- **Research Question:** To what extent do multilingual LLMs exhibit social biases in low-resource languages and culturally contextual settings?
- **Objective:** To design and evaluate a culturally contextualized benchmark (PakBBQ) for measuring social bias in LLMs within the Pakistani socio linguistic and regional landscape, addressing the limitations of Western centric datasets like BBQ for evaluating biases in LLMs.
- **Methodology:** Constructed 17,180 QA pairs across 214 templates in English and Urdu, covering 8 socio-cultural bias categories (Age, Disability, Language Formality, Gender, Physical Appearance, Regional, Religion, SES). Adapted the BBQ benchmark using Direct Transfer (DT), Target Modification (TM), Newly Added (NA), and Sample Removed (SR) strategies. Extensively web scraped new Pakistani specific biases and validated annotations via human annotators. Benchmarked 6 multilingual LLMs (GPT-4.1, GPT-4.1 Mini/Nano, Gemini 2.0 Flash/Flash Lite, DeepSeek-V3) under zero-shot settings using cyclic prompting and majority voting across ambiguous and disambiguated prompts.
- **Results:** Published at **EMNLP** Main Conference 2025. Gemini-2.0 Flash models achieved top accuracy (English 88%, Urdu 81%), while the Newly Added templates challenged LLMs with low accuracies (Urdu 50–68%). Disambiguated prompts improved accuracy by ~12 points, and negative framing reduced stereotypical responses. Bias analysis revealed stronger counter-bias in Urdu, with Gemini models scored -1 on all disambiguated categories, exposing cross-linguistic disparities and limitations of existing multilingual LLMs in socio-cultural contexts.

Confidence Aware Multi-Teacher Distillation for OOD Generalization Link *Sept 2024 – Mar 2025*

- **Role:** Primary Researcher, Supervisor **Dr. Muhammad Tahir (LUMS)**
- **Research Question:** How can we mitigate inductive biases of ML models by training a student that balances multiple image cues (shape and texture)?
- **Objective:** Improve out of distribution (OOD) generalization by mitigating cue specific biases (texture/shape) in image classification through a knowledge distillation framework, to be used in medical imaging and image classification datasets.
- **Methodology:** Employed entropy-based confidence weighting to ensemble logits from shape and texture biased teacher models (ViT, VGG16). Trained a ViT-Small student via KL divergence loss against ensembled logits weighted by teacher confidence. Propagated teacher biases using a weighted dynamic method to produce a balanced, OOD-generalizable student model. Evaluated across multiple architectures (VGG16/11, ViT-Base/Small, ResNet50, DenseNet) and datasets (ImageNet-1k, Stylized ImageNet, Animals10, Canny Edge).
- **Results:** Achieved balanced bias (Shape: 0.49, Texture: 0.51) on the student model, outperforming single teacher and naive multi teacher KD baselines in bias mitigation. Demonstrated that confidence-aware KD improves OOD generalization, while highlighting potential for further enhancement via structural knowledge distillation in loss function.

PERSONAL PROJECTS

AI-Driven Stock Insights Platform with AWS & Claude Chatbot Link *Sept 2024 – Dec 2024*

AI-powered stock insights platform with AWS backend and Claude chatbot for real time stock market analysis

- Built and deployed full-stack real-time stock analysis platform using AWS Lambda, S3, ECS, and Neon Postgres
- Integrated Claude Haiku based chatbot with custom prompt chains for natural language queries and investment insights
- Achieved **<200ms** latency and auto-scaling to millions of users via serverless architecture and ECS Fargate

AI for All – Evaluating Transformer and LLM Models in Low Resourced Settings Link *Sept 2024 – Dec 2024*

Evaluated LLM robustness and safety in low-resource languages via QA and adversarial testing

- Benchmarked XLM-RoBERTa & LLaMA-3B on QA in Urdu/Sindhi/Pashto: Urdu strong (**EM 0.72/F1 0.72**), Sindhi/Pashto collapsed (**Pashto EM 0.28/F1 0.46**)
- Demonstrated model fragility by applying adversarial perturbations, leading to near zero performance in Sindhi and Pashto, highlighting poor robustness in low resource settings
- Executed jailbreaks on LLaMA-3B: **68.5%** success in English vs **>40%** in Urdu/Sindhi/Pashto, yet highest harm severity (avg 4.3) in low-resource Pashto/Sindhi due to weaker safety alignment

End-End Data Science Chatbot using Generative AI Link *Jan 2024 – May 2024*

AI chatbot for EDA, data cleaning, and multi-format analysis

- Developed an AI chatbot for EDA, data cleaning, and statistical analysis, improving accessibility for non-technical users.
- Fine-tuned GPT-3.5 to handle **10+** data formats, deployed on Hugging Face via Gradio, with custom prompts and validation reducing hallucinations by **30%.**
- Enabled agentic function calling to run regression, prediction, and ML workflows directly on user datasets.

GenVidea – Ai Generated Video Software *Jan 2024 – May 2024*

End-to-end AI video creation and transcription tool

- Collaborated with an AI startup to develop software for auto video and image generation from text inputs

- Built end-to-end text-to-video pipeline, integrated Stable Diffusion + Runway Gen-2 APIs for image/video generation and Whisper for transcription
- Designed Figma UI/UX and shipped production full-stack text-to-video app (Astro JS + TypeScript frontend, PostgreSQL on GCP backend), improved page load speed **15%** and mobile responsiveness **2.8×**

TECHNICAL SKILLS & INTERESTS

Languages: English, Urdu
Technical Skills: Python, MERN, Java, C++/C, MATLAB, Haskell, React, LaTeX, Pytorch, AWS, Tensorflow
Interests: Formula 1, Golf, Horse Riding, Applied Machine Learning, Data Science

LEADERSHIP EXPERIENCE AND CO-CURRICULAR ACTIVITIES

TechStars Startup Weekend, Lahore - *Convenor* *Jul 2022 – Mar 2023*

- Executed Startup Weekend Lahore 2023 in collaboration with National Incubation Center Lahore and Google for Startups, led workshops and partnered with top industry leaders for startup guidance and judging
- Managed a 54-hour event for over **40** startups and **150+** attendees across Pakistan, overlooking a team of 60 and securing **800k+** PKR in sponsorships

SPADES - *Assistant Director (Socials and Operations)* *Sept 2022 – May 2023*

- Led the organisation of PSIFI's flagship event, coordinating social function and concerts for over **1,400** attendees
- Devised creative event ideas, managed itineraries, bookings, and secured necessary clearances and approvals

RECOMMENDERS

Dr. Agha Ali Raza PhD. LTI, Carnegie Mellon University
Associate Professor of Computer Science
Lahore University of Management Sciences (LUMS), Pakistan
Lab: Center for Speech and Language Technologies (CSaLT)
Email: agha.ali.raza@lums.edu.pk
My Senior Project Advisor, TA-ship and Directed Research Project professor @ LUMS from 2023-2025

Dr. Muhmmad Tahir Ph.D. Electrical Engineering, Politecnico di Torino Italy
Associate Professor of Electrical Engineering
Lahore University of Management Sciences (LUMS), Pakistan
Lab: Centre for Urban Informatics, Technology, and Policy (CITY)
Email: tahir@lums.edu.pk
My Independent Research Project (Confidence Aware KD) Advisor @ LUMS from 2024-2025

Dr. Basmaa Ali MBA, MIT
Resident Scientist
Lahore University of Management Sciences (LUMS), Pakistan,
Clinical Instructor
Harvard Medical School
Lab: AI in Healthcare Initiative (AIHI)
Email: basmaa.ali@lums.edu.pk
My supervisor at AIHI Lab, while working on AI Patient-Physician assistant @ LUMS from 2024-2025