Abdullah Hashmat

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

RESEARCH INTERESTS

Multilingual & Culturally Grounded NLP • LLM Alignment & Safety Evaluation • Self-Evolving & Tool-Augmented LLM Agents • AI for Scientific Discovery (Molecular Design)

EDUCATION

Lahore University of Management Sciences (LUMS) – *BS Computer Science* *Aug 2021 – May 2025*

Relevant Coursework: Advanced Machine Learning, Large Language Models, Speech Processing with Generative AI, Deep Learning, Artificial Intelligence, Cloud Development

PUBLICATIONS

- **Pak3H: Evaluating the Cost of Cultural Mismatch in LLM Alignment** — *EMNLP 2026 (Oral Presentation-Main)* Link
- **PakBBQ: A Culturally Adapted Bias Benchmark for QA** — *EMNLP 2025 (Main)* Link
- **IRTIQA: Evolving in English, Evaluated in Urdu — Does Agent Self-Evolution Transfer Across Languages?** — *Committed, ARR* Link
- **PakDecept: Benchmarking LLM Deception in Urdu** — *Under Review* Link
- **Rethinking the Architecture of Chemical Language Models: ADMET Property Prediction with State Space Models** — *Under Review*
- **MolScientist: Agentic Inverse Molecular Design With Scientific Experience** — *Under Review*

RESEARCH EXPERIENCE

Center for Speech and Language Technologies, LUMS & SocialNLP Lab, Macquarie University *Sep 2024 – Present*

NLP Researcher (Team Lead) | Supervised by Dr. Agha Ali Raza (LUMS) & Dr. Usman Naseem (Macquarie University)

- Leading a research team on culturally aware, **self-evolving LLM agents** that improve reasoning and tool use while mitigating Western-centric alignment bias
- First-authored **PakBBQ** (EMNLP 2025; **17,180** QA pairs, 8 bias categories) and **Pak3H** (EMNLP 2026; **28K** human-validated samples), revealing up to **~12-point** cross-lingual bias shifts and systematic degradation across helpfulness, harmlessness, and honesty in frontier LLMs
- Introduced **IRTIQA**, showing that agent self-evolution learned in English fails to transfer to Urdu and code-switched use (persistent **12–26-point** gaps on reasoning and long-horizon tasks), and **PakDecept**, extending LLM deception evaluation to Urdu

Max Planck Institute for Software Systems (MPI-SWS) *Apr 2026 – Aug 2026*

Research Fellow | Supervised by Dr. Abhilasha Ravichander

- Building **MolScientist**, a tool-augmented LLM agent for inverse molecular design with dynamic retrieval and a scientific experience memory that learns from feedback across design episodes
- Integrated RDKit, molecular docking, and QSAR oracles in a closed generate–evaluate–refine loop for multi-objective optimization, improving validity, diversity, and binding affinity of candidates

German Research Center for Artificial Intelligence (DFKI) & RPTU *Jul 2025 – Mar 2026*

Lead Researcher | Supervised by Dr. Nabeel Asim

- Designed a **multi-granular chemical language model** fusing atom-, functional-group-, and branch-level SMILES representations for molecular property prediction
- Pre-trained BERT, RoBERTa, DeBERTa, and **state space models** on a **500M+** compound corpus (150+ properties, distributed multi-GPU) and distilled them into one student (MLM + FitNet, dynamic gating), achieving **state-of-the-art on all 22 ADMET tasks**

Center of Urban Informatics, Technology and Policy (CITI), LUMS Link *Sep 2024 – Mar 2025*

Primary Researcher | Supervised by Dr. Muhammad Tahir

- Proposed confidence-aware multi-teacher **knowledge distillation**, entropy-weighting shape-biased (ViT) and texture-biased (VGG16) teachers to mitigate inductive bias and improve **OOD generalization**
- Achieved balanced cue reliance (shape 0.49 / texture 0.51) and **12%** lower bias, outperforming single-teacher and naive KD on ImageNet-1k and Stylized-ImageNet

INDUSTRY EXPERIENCE

Darcheeni (AI Healthcare Startup), LUMS – *MLOps Engineer* *May 2024 – May 2025*

- Led backend for a production **AI clinical triage assistant** combining speech recognition (ASR), RL-based decision models, and LLM response generation
- Improved ASR via fine-tuning and LLM-based post-processing; deployed **autoscaling serverless inference** on AWS (Lambda, DynamoDB, Auth0) for low-latency, reliable serving

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

- Built an **agentic Python automation pipeline** (Google APIs) for daily data ingestion, invoice generation, and WhatsApp delivery to vendors, cutting process time by **35%**
- Ran large-scale EDA on vendor onboarding and QC data to locate funnel bottlenecks, reducing QC rejections by **15%**

Pakistan Television Network (PTV) – *Data Science Intern* *Jun 2023 – Jul 2023*

- Deployed NLP **text-similarity models** to deduplicate live ticker headlines (**20%** less redundancy) and built a predictive content-analytics model that lifted viewer engagement by **10%**

TEACHING EXPERIENCE (LUMS)

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

SELECTED PROJECTS

- **AI for All – LLMs in Low-Resource Settings** Link – Benchmarked open-source LLMs on Urdu, Sindhi, and Pashto; adversarial jailbreak red-teaming reached up to **40%** attack success on Llama-3B
- **AI-Driven Stock Insights Platform** Link – Real-time analytics on AWS (Lambda, S3, ECS Fargate) and Neon Postgres with an LLM chatbot for investment queries; **<200ms** latency, serverless autoscaling
- **Bias Mitigation in Deep Learning Vision Models** Link – Multi-teacher distillation (CNN texture + ViT shape teachers) with space adapters for heterogeneous architectures; **12%** less bias on cue-conflict datasets under distribution shift
- **Agentic Data Science Chatbot** Link – Fine-tuned Llama 3 agent for automated EDA, statistics, and data cleaning across **10+** file formats; deployed on Hugging Face (Gradio)

TECHNICAL SKILLS

ML & LLMs: PyTorch, Hugging Face Transformers, LLM fine-tuning, knowledge distillation, agentic AI / tool calling, LLM evaluation & benchmarking, RDKit
MLOps & Data: AWS (Lambda, S3, ECS/Fargate, DynamoDB), Docker, serverless model deployment, distributed multi-GPU training, ETL & web scraping, PostgreSQL
Programming: Python, C/C++, Java, SQL, JavaScript, React/MERN, MATLAB, Haskell, LaTeX

LEADERSHIP

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

- Led a **60-member** team with National Incubation Center and Google for Startups to run a 54-hour event for **40+** startups and **150+** attendees; secured **800K+ PKR** in sponsorship

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

- Organised PSIFI's flagship social events and concerts for **1,400+** attendees