



Safia Baloch

[LinkedIn](#) | [GitHub](#) | [Google Scholar](#) | [Home Page](#)

[Email: shahalsafia45@gmail.com](mailto:shahalsafia45@gmail.com) [Email: safia.baloch@giki.edu.pk](mailto:safia.baloch@giki.edu.pk)

RESEARCH SUMMARY

I believe meaningful research begins with asking the right questions before seeking the right answers. My research focuses on **machine learning**, **time-series modelling**, **graph-based learning**, and **interpretable AI**. My research experience includes LSTM-based time-series forecasting, and I am currently exploring graph-based pruning techniques to reduce the computational cost of large language models, alongside **mechanistic interpretability** to investigate their internal representations and computational behaviour. I am interested in developing robust, interpretable learning methods for heterogeneous and longitudinal data, particularly for modelling patient trajectories, integrating multimodal clinical information, and supporting reliable, information-driven healthcare.

RESEARCH PUBLICATIONS

* Equal contribution

[1] Hanan Shabir, Noor Fatima, **Safia Baloch**, Masroor Hussain “City Sentinel: A Unified AI-Based Smart Surveillance Framework for Real-Time Multi-Threat Detection Using Deep Learning”, **Under Review** FITs Conference 26

[2] Rahemeen Khan, **Safia Baloch***, “ A Comparison of Text Classification and Topic Modeling Methods on AG News”, **Under Review** FITs Conference 2026

[3] **Safia Baloch**, Rahemeen Khan, “Towards Human-Centered Multi-Agent Systems: Integrating Cognition, Culture, Values and Cooperation in AI Agents”, <https://arxiv.org/pdf/2606.08274> , 2026

[4] Nauman Ali Murad, **Safia Baloch***. "Governed by Agents: A Survey on the Role of Agentic AI in Future Computing Environments." *arXiv preprint, arXiv:2509.16676* , 2025

[5] Rahemeen Khan, **Safia Baloch***. "Forecasting the Stability of COVID-19 Vaccine Companies Stock Market using LSTM and Time-Series Models." *International Conference on Communication, Computing and Digital Systems (C-CODE / ICCDS 2023)*

- **UrduTokBench**: A Multidomain Tokenization Benchmark for Urdu NLP- **in preparation**

LECTURE/ TALKS

Delivered a session on “**Introduction to Artificial Intelligence**” for young female students at public colleges in Swabi, promoting AI awareness and encouraging women’s participation in technology. (5/10/16) [Link](#)

Invited to talk in “Ikhtiraa” **Women Engineering Society** event at Ghulam Ishaq Khan Institute of Engineering Sciences and Technology (2/22/25) for **Amplifying the voices of women breaking barriers in technology, industry, and leadership.**

RESEARCH & PROJECT

1. **Early Flood Warning System | WWF Pakistan & GIKI** [Link](#)
Contributing to the development of an AI-powered early flood warning system for mountainous regions of Swat. The project integrates camera-based river-level monitoring with historical weather data, real-time conditions, and weather forecasts to predict flood risks and support timely warnings for community safety.
2. **Health Atlas: AI-Driven Prescription Digitization & Disease Mapping:** [GitHub](#)
Supervised the development of an intelligent healthcare analytics platform that turns handwritten and printed medical prescriptions into structured, verified digital records, automatically extracting drug names and dosages, normalizing them against standard drug nomenclature, and checking for dangerous drug-drug interactions before they reach a patient or pharmacist.
3. **IntelliDentX: Automated Dental Lesion Detection and Root Canal Evaluation:** [GitHub](#)
Guided the development of a computer vision-based dental diagnostic system using the YOLOv11 object detection model. The solution automatically identified dental lesions from radiographic images and evaluated the quality of root canal fillings, assisting clinicians in diagnosis and treatment assessment through AI-powered image analysis.
4. **Stocks in Sight: Stock Market Forecasting Using Time Series Analysis:** [GitHub](#)
Supervised a financial analytics project that leveraged historical Pakistan Stock Exchange (PSX) data and time series forecasting models to analyze market behavior and predict stock price movements. The project involved data preprocessing, trend analysis, and predictive modeling to support investment decision-making and market insights.
5. **Lossless Compression of Vector Databases:** [GitHub](#)
Mentored undergraduate students in a CS412 Information Retrieval project titled “**Lossless Compression of Vector Databases.**” The project explored vector indexing, Approximate Nearest Neighbor Search (ANNS), IVF-based retrieval, Product Quantization, Bolt compression techniques, and evaluation of retrieval efficiency, memory usage, and search performance in large-scale vector databases.
6. **GraphReddit: Social Network Analysis and Community Detection on Reddit:** [GitHub](#)
Supervised the design and implementation of a graph-based social media analytics system that streamed Reddit data, constructed user-content interaction networks in Neo4j, and applied graph algorithms including PageRank and community detection. The project integrated sentiment analysis and network visualization techniques to study user influence, information flow, sentiment diffusion, and cross-community interactions within online discussion platforms.
7. **QuizBolt: Scalable AI-Powered Learning Platform with Cloud Tool (AWS):** [GitHub](#)
Mentored students in developing an AI-powered study platform that generates personalized quizzes, flashcards, study notes, and citation-grounded RAG-based chat from uploaded academic content using AWS, Next.js, Supabase, and OpenAI.
8. **Improved Rational Retrieval Acts for Sparse Retrieval Models:**
Mentored undergraduate students in an Information Retrieval project investigating context-aware retrieval methods, pragmatic reasoning, and RSA-based enhancements for sparse retrieval models, achieving measurable improvements in ranking quality and retrieval performance.

9. Tuberculosis Classification Using Machine Learning:

Supervised the development of a machine learning framework for multiclass tuberculosis classification using the ImageCLEF Tuberculosis dataset. The project involved medical image analysis, feature extraction, and predictive modeling to distinguish between different tuberculosis manifestations, supporting computer-aided diagnosis in healthcare.

10. Virtual Therapy for Acrophobia and Claustrophobia

Mentored the design and development of a Unity-based Virtual Reality (VR) application that provides immersive exposure therapy environments for individuals with acrophobia (fear of heights) and claustrophobia (fear of confined spaces/caves). The system utilized controlled virtual scenarios to support gradual fear reduction and therapeutic intervention.

11. Automatic Resume Screening and Ranking System

Guided the development of an intelligent resume screening system that employed feature selection techniques, candidate profiling, and machine learning algorithms to extract relevant information from resumes and automatically rank applicants based on job requirements, improving recruitment efficiency and decision-making.

EDUCATION

M.S. in Computer Science: FAST National University of Computer & Emerging Sciences Aug 2017 – Jun 2020

GPA: 3.13/4.0 | Supervisor: Prof. Dr. Jawad Ahmed Shamsi

Thesis: "Automatic Grading of Short Answers using Text Similarity Techniques"

Applied knowledge-based and corpus-based text similarity techniques on word-level representations. Constructed a novel short-answer dataset from computer science quizzes across multiple subjects.

B.S. in Computer Science: FAST National University of Computer & Emerging Sciences Aug 2012 – May 2016

Supervisor: Dr. Nadeem Kafi:

Final Year Project: HuMa: Platform for Human-Machine Computations Using Facebook API

EXPERIENCE

Lecturer: GIK Institute of Engineering Sciences and Technology, Swabi Oct 2024 – Present

Teaching

**Data Science, Cloud Computing,
DevOps Lab, Information Retrieval,
Big Data Analytics.**

Coordinator of Data Science Students Summer Internships and Curriculum Committee.

Batch Advisor for DS Final Year Students

Developed and updated course outlines, syllabi, learning outcomes, and assessment plans for undergraduate courses. Designed laboratory manuals, programming assignments, and project-based learning activities.

Lecturer: FAST National University of Computer & Emerging Sciences, Karachi Jan 2021 – Sep 2024

Courses Taught:

Operating Systems, Discrete Structures, Data Structures, & Parallel Distributed Computing.

Instructor: FAST National University of Computer & Emerging Sciences, Karachi Aug 2017 – Dec 2020

Delivered labs for Programming Fundamentals, Data Structures, Operating Systems, and ICT Tools.

SQA Engineer: OvrloD Pvt Ltd, Karachi

Oct 2016 – Jun 2017

Conducted software quality assurance for product releases.

ACADEMIC SERVICE(s)/ LEADERSHIP

Academic Reviewer

International Conference on Frontiers of Information Technology (FIT) **2026**

International Conference on Innovations in Computer Science & Software Engineering (ICONICS) **2026**

International Conference on Emerging Trends in Smart Technologies (ICETST) **2024**

International Conference on IT and Industrial Technologies (ICIT) **2024**

Academic Leading Role:

Team Member, ICPC Asia-Topi	Jan 2025 – Present
Faculty Head, Literature Club Society, FAST NUCES Karachi	Jul–Sep 2024
Faculty Head, ACM-Women Chapter, FAST NUCES Karachi	Aug 2022 – Jun 2023
Faculty Head, ACM Chapter, FAST NUCES Karachi	Aug 2021 – Jun 2022
Faculty Co-Head, Sportics Society, FAST NUCES Karachi	Aug 2020 – Jun 2021
Head, Procom Flagship Event Creative Competitions, FAST NUCES	Apr 2019
Judge, Procom Creative Competition (Sketching) FAST NUCES	Apr 2018

AWARDS/ HONORS& SCHOLARSHIPS

Outstanding Instructor Award, Data Science Department, GIKI	May 2025
Member, Pakistan Youth Parliament	Nov 2025 – Present
ML to LLM Summer Bootcamp Instructor, GIKI	Jul–Aug 2025
MS Scholarship, FAST NUCES Karachi	Aug 2017 – Jul 2020
Sindh Endowment Scholarship for BS Degree	Aug 2012 – Jun 2016
1st Prize, Inter-Board College Debate Competition, Karachi	Apr 2009
Best Performance Award, Secondary School	May 2007

CERTIFICATIONS & PROFESSIONAL DEVELOPMENT

International English Language Testing System (IELTS)	<u>6.5 Band</u>
Python for Data Science, AI & Development	<u>Coursera</u>
Mathematics for Machine Learning: Linear Algebra	<u>Imperial College London (Coursera)</u>
AI Python for Beginners	<u>DeepLearning.AI</u>
Graph Analytics for Big Data	<u>Coursera</u>
PyTorch for Deep Learning	<u>DeepLearning.AI</u>
Multi-vector Image Retrieval	<u>DeepLearning.AI</u>
Document AI: From OCR to Agentic Document Doc Extraction	<u>Deeplearning.AI</u>

AI Agents in LangGraph
The Complete Quantum Computing Course for Beginners
Introducing Multimodal Llama 3.2
Large Multimodal Model Prompting with Gemini
Knowledge Graphs for RAG
AI Agentic Design Patterns with AutoGen

[DeepLearning.AI](#)
[Coursera](#)
[DeepLearning.AI](#)
[DeepLearning.AI](#)
[DeepLearning.AI](#)
[Deeplearning.AI](#)

TECHNICAL SKILLS

Programming Languages: Python, Pytorch, Tensorflow, C, C++, Java, Shell Script, SQL.

Tools: Anaconda Jupyter, Power bi, Weka, Hadoop, Spark, AWS: EC2, S3, IAM, Docker, Kubernetes, Android Studio, Flask, Microsoft Visio, NetBeans, Dev (C++), LaTeX, Grafana, Jenkins, Terraform, Prometheus, GitHub, Neo4j, Langchain, Huggingface, BigQuery, Neo4j.

REFERENCES

Dr. Farrukh Hasan Syed

Assistant Professor, Computer Science

Institute of Business Administration (IBA), Karachi

fhsyed@iba.edu.pk

Dr. Nadeem Kafi

Assistant Professor, Computer Science

Mohammad Ali Jinnah University (MAJU), Karachi

nadeem.kafi@jinnah.edu

Prof. Dr. Jawwad Ahmed Shamsi

Dean of Computing & Professor, School of Computing

FAST National University of Computer and Emerging Sciences (NUCES), Karachi

jawwad.shamsi@nu.edu.pk

Prof. Dr. Masroor Hussain

Head of Department, Computer Engineering & Data Science

Ghulam Ishaq Khan Institute of Engineering Sciences & Technology (GIKI), Topi

hussain@giki.edu.pk