Moosa Hassan Alvi

alvimoosahassan@gmail.com • +92 323 1487960 • Lahore, Pakistan
moosa-hassan.github.io • github.com/Moosa-Hassan • linkedin.com/in/moosa-hassan-alvi

Research Interests

AI systems, efficient data systems, and data processing, with a focus on systems that preserve formal correctness guarantees while achieving practical, real-world performance. Broader interests include multi-agent LLM systems and agentic orchestration, explainable and trustworthy machine learning (e.g., ontology-grounded evaluation of GNN explanations), robust data quality and repair for multivariate and time-series data, and privacy & security of data systems. Motivated by building rigorous systems research with direct applicability to production-scale data infrastructure.

Education

Lahore University of Management Sciences (LUMS) Sep 2023 – May 2027 (Expected)
B.Sc. Computer Science *CGPA: 3.53/4.00* Syed Babar Ali School of Science & Engineering
Selected Coursework: Algorithms (A), Theory of Automata (A-), Foundations of AI & ML (A), Network-Centric Computing (A), Applied Cryptography (A-), Software Engineering (A-), Big Data Analytics (A), MLOps/LLMOps/Agentic AI (A-), Network Security (A-), Operating Systems (B+), Databases (B+)
Ongoing: Multi-Agent Systems, Distributed Systems

Lahore College of Arts and Science (LACAS) Aug 2021 – June 2023
A-Levels: Computer Science (A), Chemistry (A), Physics (A), Mathematics (B)

Research Experience

LogLite-Q: Exact Compressed-Domain Log Search Course Research Project (Group)
Python, C++

- Designed a q-gram sidecar architecture enabling provably exact substring search over stateful compressed logs, with formal mathematical guarantees of zero false negatives.
- Achieved a 207× geometric-mean speedup over full decompression and a 1.66× speedup over ripgrep on plaintext; developed a storage-aware variant (`qidx4`) limiting index footprint to 8.8–16.8% of the speed-first design.
- Evaluated exact performance across 16/16 query-dataset cells with Holm-corrected Wilcoxon significance ($p < 0.001$). Technical Report available.

Selected Projects

Agentic Pygame Game Generator Python, LangGraph, LangChain, OpenAI API

- Built a multi-agent orchestration framework that converts natural-language prompts into fully executable Pygame applications.
- Designed a graph of specialized agents (engineer, coder, QA, scorer) with automated human-in-the-loop self-correction loops that re-trigger generation on execution or evaluation failures until validation succeeds.

Teaching Experience

Teaching Assistant – CS 315: Theory of Automata Jan 2026 – May 2026
Lahore University of Management Sciences

- Designed and delivered tutorials on computational reductions, Turing Machines, and Rice's Theorem.
- Assisted in designing questions for assignments, quizzes, and exams.
- Graded coursework and held weekly office hours.

Teaching Assistant – CS 200: Introduction to Programming Jan 2025 – May 2025
Lahore University of Management Sciences

- Taught OOP fundamentals: classes, objects, inheritance, polymorphism, encapsulation.
- Held debugging office hours and evaluated lab sessions.

Professional Experience

Machine Learning Intern July 2026 – Present
10pearls

- Designing and implementing an end-to-end serverless AQI forecasting system that predicts air quality for the next 72 hours.
- Built the full pipeline: automated data extraction from Open-Meteo API, feature engineering (temporal, lag, and rolling features), model training, and evaluation.
- Conducted systematic experiments with multiple models (XGBoost, Random Forest, LSTM, GRU, MLP, Ridge) as required by the internship, evaluating them across different data regimes; currently focused on multi-step 72-hour prediction strategies that avoid data leakage.
- Responsible for system architecture, experimental design, and model selection; using LLMs as a productivity tool for implementation.

Cyber Security Intern July 2025 – Aug 2025
NETSOL Technologies Pakistan

- Conducted hands-on research in endpoint security, digital forensics, and incident investigation via system artifacts and Windows Event Viewer.
- Designed and implemented multiple Data Loss Prevention (DLP) proof-of-concept solutions: clipboard monitoring, data masking, process denylisting (name + hash), LLM-based document classification, OCR-based sensitive data detection, metadata analysis, and upload monitoring.

AI Engineer (Freelance) Jan 2025
COD Return Prediction

- Engineered 10+ domain-driven features from raw e-commerce customer data and trained a CatBoost classifier optimized for precision-heavy $F_{0.5}$ metric ($F_{0.5} = 0.72$, AUC-ROC$= 0.79$, accuracy$= 0.75$).

Technical Skills

Languages	Python, C++, JavaScript/TypeScript, SQL, PowerShell
ML / AI	scikit-learn, XGBoost, CatBoost, LightGBM, LangGraph, LangChain, OpenAI API, pandas, NumPy
Systems & Tools	Git, Docker, Next.js, NestJS, Firebase, Vercel, Jupyter, Streamlit
Security	Windows Event Viewer, Data Loss Prevention (DLP), endpoint monitoring, digital forensics

Awards & Honors

- 100% Merit Academic Scholarship (2022)
- Medal for Academic Excellence (2022)
- Student of the Year Award (2021)