Md Kaif Imteyaz

kaifimteyaz.k@gmail.com | kaif-imteyaz.github.io | linkedin.com/kaifimteyaz | scholar.google.com/kaifimteyaz

Summary

I am an AI/ML Research Engineer with experience in machine learning, deep learning and large language models. I work with PyTorch and Transformers on model development, fine-tuning and evaluation. Currently, I am part of a collaborative research project involving UCL, AIIMS and IIT Delhi, focused on AI-assisted dental record digitisation.

Education

Jamia Hamdard — New Delhi, India
Bachelor of Technology in Computer Science & Engineering; — 2023 – 2026

Skills

Programming: Python, JavaScript

ML & Deep Learning: PyTorch, Scikit-learn, OpenCV, CNNs, InceptionV3, Attention Mechanisms (CBAM), Transfer Learning, Fine-Tuning, Model Quantization

AI Research: LLMs, VLMs, Hugging Face, Qwen3-VL, OpenAI API, Ollama, RAG, RLHF, LoRA, QLoRA, Explainable AI (XAI), Hallucination Detection, AI Fairness & Bias, Ablation Studies, Multi-Seed Evaluation

Data & MLOps: NumPy, Pandas, Matplotlib, Seaborn, Plotly, Weights & Biases (W&B), Hypothesis Testing, Cohort Analysis

Web-dev: FastAPI, Node.js, Express.js, React.js, Next.js, Tailwind CSS, Supabase **Databases:** PostgreSQL, MySQL

Awards

- **Won 2nd Place**, AMD India Reinforcement Learning Hackathon among 200+ teams, organized by AMD, Yardi School of AI (IIT Delhi), and Unsloth AI (2026)
- **7th Rank** (out of 2,000+ participants, two-phase), AISEHack 2026 organized by ANRF in collaboration with IBM and IIT Delhi, hosted at IIIT Hyderabad (2026)

Experience

Research Engineer — Aug. 2026 – present
Working with Dr. Prateek Raj, University College London

- Working on a joint AIIMS–IIT Delhi–UCL research project applying AI and HCI to digitise paper-based dental records and improve accessibility of digital healthcare.

Research Intern(AI & Data Engineering) — Apr. 2026 – Oct. 2026
Worked with Prof. Amrita Dhillon, King's College London

- Built an end-to-end pipeline extracting structured asset and liability data from scanned Bihar government officer affidavits (pilot scope: 4 districts) using a vision-language model (Qwen3-VL),converting raster scans with no embedded text layer into a validated 78-field schema.
- Engineered DOB, PAN, and income-tax field parsing plus source-aware deduplication preserving cross-year filings and tuned 4-bit GPU inference with adaptive batching, cutting extraction time from 285s to 23s per document.
- Delivered dual CSV & JSON outputs with source-PDF links and confidence flags, achieving 98.8% valid DOB and 90.8% valid PAN extraction, with remaining gaps traced to blank or illegible source documents.

Data Science Intern — Sept. 2025 – Feb. 2026
Worked with Prof. Aaditya Dar, Independent Researcher

- Implemented a scalable data pipeline that cleaned, standardized, and structured approximately 2 million government official financial asset declaration records for downstream use.
- Designed and executed comparative experiments across BART, FLAN-T5, GPT-2, and Qwen models, evaluating performance trade-offs for structured information extraction tasks.
- Evaluated encoder architectures for entity standardization and record linkage, analyzing trade-offs in accuracy, generalization, and inference efficiency.

Research Intern — Mar. 2025 – Aug. 2025
Worked with Prof. Prabhat Barnwal, Michigan State University

- Developed the Peer Knowledge Exchange Platform to improve retrieval speed, accessibility, and user experience for government officers in Punjab, India.
- Engineered Sewa Assist, a conversational retrieval interface deployed via WhatsApp, enabling officers to access program resources through natural language queries.
- Integrated an on-site chatbot that reduced user friction and improved resource discovery through keyword-driven retrieval logic.
- Delivered the solution using Next.js with version control via GitHub for long-term maintainability and scalability.

Full Stack Developer Intern Dec. 2022 – Jan. 2023
FTK Center for Information Technology, Jamia Millia Islamia New Delhi, India

- Led full stack development of an alumni community web application serving 2,000+ users, improving engagement using HTML, CSS, JavaScript, and PHP.
- Optimized site architecture for mobile responsiveness and reduced page load times by 20%.
- Collaborated in a four-member Agile team to ship three major features within six weeks.

Research Projects

Design Theater: A Benchmark for Generative UI (Accepted at AIES 2026) | LLM Benchmarking

Developed a benchmark to measure authenticity gaps between generated reasoning traces and actual model behavior through systematic evaluation and interpretability analysis.

Episode-Aware PM2.5 Forecasting Across India | Spatiotemporal Deep Learning

Designed a spatiotemporal deep learning model for country-level PM2.5 forecasting over a 140×124 grid at 25 km resolution, predicting 16 future hours from a 10-hour lookback across 26 input channels with no future meteorological context.

CATS: Culturally Adaptive Therapeutic System | Final Undergraduate Project

Designed an agentic system for physiotherapists to generate personalized exercise videos aligned with patients' cultural backgrounds through culturally appropriate demonstrations.

Projects

Ink2Info: Digitizing Health Informatics | *Python, Next.js, FastAPI, OpenAI, Gemma 7B* | Repository 2026

- Developed a prescription digitization pipeline using OpenAI and Gemma 7B with a Fabric.js canvas for real-time handwriting extraction, directly modifying inference behavior of the underlying models for domain adaptation.
- Implemented role-based access control with HMAC-signed QR codes for secure data sharing across multiple user roles.

De-Tail: Agentic AI Search Assistant | *React, OpenAI API, Tavily API* | Repository 2025

- Engineered a retrieval-augmented generation pipeline that exposes its full step-by-step reasoning chain, integrating OpenAI Reasoner with Tavily API for real-time web retrieval.
- Designed end-to-end retrieval flow from query intake through document fetching, reranking signal generation, and response synthesis, with full transparency at each stage.

Certifications & Activities

- **IBM SkillsBuild Certifications:** Artificial Intelligence, Data Fundamentals
- **AI Summer School**, IIIT Delhi (May 2024, July 2023)