HASNAT KHAN

AI / ML Engineer

Islamabad, Pakistan · +92 313 564 9135 · hk5610504@gmail.com · github.com/HasnatKhan010 · linkedin.com/in/hasnat-khan-8706222b4 · hasnatkhan.tech

SUMMARY

Final-year BS Computer Science student and AI/ML engineer who builds retrieval and machine-learning systems together with the pipelines that keep them running. Fine-tuned a transformer to 98.33% accuracy and 0.9830 macro-F1 against three baselines on one shared split; built a cross-script Roman-Urdu retrieval pipeline reaching 0.983 Recall@10, up from 0.192 on the same frozen split, on CPU alone; wrote a Rust data-cleaning engine that filters a million rows in 189 ms. Works across the whole lifecycle — ingest, train, evaluate, serve, monitor — in Python, Rust, C++, C and TypeScript, with PyTorch, scikit-learn, FastAPI, Docker and GitHub Actions. Two AI/ML internships plus Cycle 1 of Pakistan's national AI training initiative. Open to AI/ML, MLOps, data engineering and software engineering roles; available immediately, remote, hybrid or on-site.

TECHNICAL SKILLS

Languages: Python, C++, C, Rust, JavaScript, TypeScript, SQL, x86 Assembly

Machine learning: PyTorch, TensorFlow / Keras, scikit-learn, Hugging Face Transformers, NumPy, Pandas, XGBoost, Neural networks, Backpropagation, Feature engineering, Cross-validation, Hyperparameter tuning

LLMs, NLP & retrieval: Transformer fine-tuning, Sentence embeddings, Dense retrieval, BM25, Reciprocal Rank Fusion, Cross-encoder reranking, Retrieval-augmented generation, Topic modelling (LDA, NMF), Named entity recognition, spaCy, Prompt engineering, LangChain

MLOps & deployment: Docker & Docker Compose, FastAPI, Model serving, Training pipelines, Drift monitoring, Scheduled retraining, Model registry & persistence, GitHub Actions CI, pytest, Pinned environments

Data engineering & automation: Web scraping, ETL pipelines, Data cleaning at scale, Parameterised SQL, API automation, Bash scripting

Systems & security: Operating systems, Data structures & algorithms, Cryptography (RSA, AES-256, SHA-256), Digital signatures & PKI, Blockchain / proof of work, Linux (Kubuntu / KDE)

Build & ship: React, Next.js, Node.js / Express, MongoDB, MariaDB / MySQL, REST API design, Streamlit, Tailwind

EXPERIENCE

AI/ML Intern — ITSimplera Solutions · Remote · Jun 2026 – Aug 2026 · github.com/HasnatKhan010/ITSimplera_Institute_AI-ML_internship

- Fine-tuned DistilBERT to 98.33% accuracy and 0.9830 macro-F1 on a 419-article held-out test set, benchmarked against three approaches on one shared stratified 70/10/20 split.
- Ran the same DistilBERT backbone both frozen and fine-tuned to isolate where the gain came from — frozen embeddings beat TF-IDF by only +0.24 F1, attributing nearly all the advantage to fine-tuning rather than to the pre-trained representations.
- Implemented C_NPMI, C_UMass and topic-diversity coherence metrics from their mathematical definitions after gensim failed to build on Python 3.14, then used them to sweep topic count over k=4–14 under a selection rule declared before results were seen.
- Trained a Random Forest regressor to 1.213 kWh RMSE on a year of 15-minute industrial energy readings (~35,040 records), then quantified the cost of PCA compression: 11 components at 95% variance raised RMSE by 171%.
- Segmented ~9,000 credit-card customers into four behavioural profiles with K-Means, validated by elbow, silhouette and a Ward-linkage dendrogram, and mapped each segment to a retail-banking action.
- Shipped every week as a runnable app — Streamlit dashboards for weeks 1, 2 and 5, a FastAPI prediction service for week 3 — with a shared clean_text() module so training and live inference could not drift apart.

AI/ML Intern — COMSATS University Islamabad (Junaid Zaidi Library) · Islamabad, Pakistan · 2026 · github.com/HasnatKhan010/koha-opac-ai-plugin

- Shipped an AI search assistant into the Koha library OPAC as an installable plugin package, replacing traditional search forms with natural-language queries.
- Built intent detection across 14 query types — title, author, ISBN, publisher, barcode, call number, branch, language, year, subject, recommendation and compound filter searches — with fuzzy fallback that corrects misspelled queries.
- Backed it with a FastAPI service querying Koha's MariaDB schema through parameterised SQL, with per-IP rate limiting at 20 requests per 60 seconds.
- Made the widget keyboard-operable end to end: full Tab/Shift+Tab focus trapping, Escape to close, Enter to send.
- Wrote DataPrep, a Rust data-cleaning engine measured at a million rows — filtering that set in 189 ms — built in Rust specifically because the Python tooling could not hold the working set in memory.

SELECTED PROJECTS

Raabta — Python, FastAPI, React, sentence-transformers, BM25, Streamlit · github.com/HasnatKhan010/raabta
Cross-script retrieval answering Roman-Urdu questions from Urdu-script documents: Recall@10 0.983 and MRR@10 0.583, up from 0.192 / 0.101 on the same 120 frozen dev questions, over 16,352 passages, CPU-only. Multi-view query

reformulation with drift gating; three routes (Urdu BM25, multilingual dense, uroman-romanized title n-grams) fused by weighted RRF; cross-encoder reranking; and a seven-check gate that abstains with a reason rather than substituting a fact. FastAPI, React, Streamlit, GitHub Actions CI.

Prognostix — PyTorch, FastAPI, Docker, pytest · github.com/HasnatKhan010/prognostix-ml
Predictive-maintenance platform estimating remaining useful life from NASA CMAPSS turbofan telemetry, built as all five lifecycle stages in one repository: ingest, train, evaluate, serve, monitor. LSTM, GRU and attention models against a Random Forest baseline, a scripted rolling/lag feature pipeline, a FastAPI inference service, drift and degradation monitoring with alerting, and docker-compose bringing up the whole stack.

DataPrep Studio — Rust, TypeScript, Columnar engine, Windows, MARC · github.com/HasnatKhan010/dataprep-studio · github.com/HasnatKhan010/dataprep-studio/releases/latest
Windows desktop data-cleaning application with a columnar, lazy engine in Rust: filters 1,000,000 rows in 189 ms and opens a file in 3.8 ms by materialising only the rows on screen. 30+ reversible operations across CSV, TSV, JSON, Excel and MARC, every value held as text so identifiers keep their leading zeros. Shipped as an installer with published SHA-256 checksums; no upload, no account, no telemetry.

Attention Lab — PyTorch, Python, Jupyter · github.com/HasnatKhan010/Attention-Lab
Attention mechanisms implemented from tensor primitives in PyTorch and verified numerically against framework output — scaled dot-product, causal and multi-head, extending to RoPE, RMSNorm, SwiGLU and grouped-query attention, at seven documented stages per mechanism.

HousePricePredictor PK — Python, scikit-learn, BeautifulSoup, CustomTkinter · github.com/HasnatKhan010/HousePricePredictor-Pk
End-to-end property-pricing pipeline for Islamabad: scrapes live Zameen.com listings, normalises the local market's mixed units (marla, kanal, square feet), engineers features, trains a scikit-learn regressor and serves it in a CustomTkinter desktop application.

AegisChain — Python, RSA, AES-256, Blockchain · github.com/HasnatKhan010/AegisChain-Banking-System
Banking ledger in Python where every transaction is signed and hash-chained: RSA, AES-256, Diffie-Hellman, digital signatures and PKI certificate issuance implemented from the specifications, with hybrid encryption over a proof-of-work chain so history edits are detectable.

Virtual File System — C, Systems programming, Python · github.com/HasnatKhan010/Virtual-File-System-With-the-manual-System-calls-with-memory-with-basic-shell-implementation
Virtual file system in C with a hand-written POSIX-style syscall layer, its own block manager and memory allocator, and an interactive shell for navigating the virtual disk.

NovaCare — React 18, Node.js, MongoDB, Docker, Tailwind · github.com/HasnatKhan010/NovaCare-Hospital-System
Hospital management system and patient portal on the MERN stack (React 18, Node.js, Express, MongoDB), containerised with Docker, with staff and patient roles separated in the API layer over one backend.

EDUCATION

BS Computer Science — COMSATS University Islamabad · Islamabad, Pakistan · 2024 – Present · Currently in the 6th semester

Relevant coursework: Artificial Intelligence; Machine Learning; Database Systems; Advanced Database Management Systems; Programming Fundamentals; Object-Oriented Programming; Data Structures & Algorithms; Design & Analysis of Algorithms; Software Engineering; Web Technologies; Operating Systems; Computer Organization & Assembly Language; Information Security; Computer Networks; Introduction to ICT; Discrete Structures; Linear Algebra; Probability & Statistics; Calculus & Analytical Geometry

FSc Pre-Medical — Tameer-i-Wattan Public College · Pakistan · 2022 – 2024 · 1024/1100 (93.1%)

Matriculation — Al Quran Beacon School · Pakistan · 2020 – 2022 · 968/1100 (88.0%) · Grades 9 and 10

CERTIFICATIONS

IBM Machine Learning Professional Certificate — IBM · 2026 · 6 courses · Credential ID N33DFX3MDLFD · Supervised and unsupervised learning, deep learning and reinforcement learning, through to a capstone · coursera.org/share/20aa00ac9bfb497072cb168de6a0b4e7

Mathematics for Machine Learning and Data Science — DeepLearning.AI · 2026 · 3 courses · Credential ID MKOA07D3TMA6 · Linear algebra, calculus and probability — the maths under gradient descent and probabilistic modelling · coursera.org/share/cad9f1738a338e286bd97213512ec307

ACT AI — Awareness, Competency & Tools Training for AI — National AI Training Initiative, Pakistan · 2026 · 8-week programme, Cycle 1 · Credential ID ACTAI-C1-2026-00194-EZMB · Government of Pakistan initiative under the PM's Youth Programme with HEC and NAVTTC, delivered by AI SkillBridge · actai.aiskillbridge.pk/v/YAVFR8KV2JRXEF2L

Component courses, each certified individually: Exploratory Data Analysis for Machine Learning; Supervised Machine Learning: Regression; Supervised Machine Learning: Classification; Unsupervised Machine Learning; Deep Learning and Reinforcement Learning; Machine Learning Capstone; Linear Algebra for Machine Learning and Data Science; Calculus for Machine Learning and Data Science; Probability & Statistics for Machine Learning & Data Science

Also: Crash Course on Python (Google); Our Earth's Future (American Museum of Natural History)