Islam Kathat

Last Updated on 21st May 2026
current resume here

islamkhan.in
faiz.14a@gmail.com | +91 80008 02709 | LinkedIn | GitHub

EDUCATION

JECRC UNIVERSITY
MCA in Computer Science
July 2025 | Jaipur, Rajasthan
CGPA: 8.57

MDS UNIVERSITY
BCA in Computer Science
July 2023 | Ajmer, Rajasthan
CGPA: 7.68

SKILLS

PROGRAMMING LANGUAGES
- Java
- SQL
- Python
- JavaScript, TypeScript

ML LIBRARIES
- Scikit-learn
- TensorFlow
- PyTorch (primary)
- NumPy, Pandas, SciPy
- Transformers, PEFT, TRL
- BitsAndBytes, Accelerate

LLM & AI ENGINEERING
- Weights & Biases
- HuggingFace Hub
- Cosine LR, AdamW
- KV Cache, GQA, MoE
- Gradient Checkpointing
- NF4 4-bit, BitsAndBytes
- Mixed Precision (bf16/fp16)
- QLoRA, LoRA, PEFT, SFTTrainer

FULL-STACK ENGINEERING
- Realm, Kafka
- GitHub, GitLab
- React.js, Next.js
- Docker, Firebase, Expo
- React Native, Native Modules
- PostgreSQL, MongoDB, Redis
- Node.js, FastAPI, Flask

SOFT SKILLS
- Rapid Learning
- Public Speaking
- Problem-Solving
- Cross-Team Collaboration

CERTIFICATIONS

Responsible Generative AI (Microsoft)
Deep Learning and AI (KPMG International)
Java Certificate (HackerRank)

PROJECTS

MISTRAL-7B FINE-TUNING Fine-tuned an LLM Model | GitHub
QLoRA | PEFT | TRL | BitsAndBytes | Hugging Face | Weights & Biases
- Fine-tuned Mistral-7B-Instruct-v0.3 using QLoRA (NF4 4-bit, r=64) on 500 domain-specific instruction pairs with 0.5% of trainable parameters.
- Achieved 13% perplexity reduction vs base model baseline using cosine LR, Paged AdamW, and gradient checkpointing on Nvidia A100 80GB GPU.
- Logged training metrics via **Weights & Biases**, merged the LoRA adapter into the base model, open-sourced the final model on **Hugging Face** with a model card.

RAG SYSTEM Retrieval-Augmented Generation System | Live
FastAPI | ChromaDB | Ollama | SearXNG | Sentence Transformers
- Architected full-stack RAG application using FastAPI, React.js, ChromaDB, and LLaMA 3.1 with Ollama for local LLM inference and Grok API for production.
- Integrated SearXNG and Trafilatura in local and tavily in production with multi-format document ingestion for reducing RAG retrieval time by 17%.
- Achieved sub-600ms retrieval via Sentence Transformer embeddings, 512-char chunking with 100 overlapping chunks and cross-encoder re-ranking.

MNIST DIGIT CLASSIFIER ML Classification Models | GitHub
TensorFlow | Keras | TFLite | Scikit-learn | Seaborn | Matplotlib
- Trained MNIST digit classifiers (MLP, SelectKBest-optimized, CNN with BatchNorm) from scratch, each achieving 90%+ accuracy in prediction.
- Tuned training pipeline using EarlyStopping, ReduceLROnPlateau, ModelCheckpoint, and L2 regularization for better generalization.
- Performed pre-training EDA using Seaborn, Matplotlib, and Plotly; validated via classification report and confusion matrix to improve performance by 14%.

OTHER PROJECTS ML | App | Full Stack | Productivity Tools
- **CIFAR-10 Object Classifier**: Developed and benchmarked CNN (92%), ResNet (95%), and EfficientNetV2B0 transfer learning (97%) classification models.
- **AI Club**: Multi-AI desktop application enabling parallel LLM inference and real-time response comparison from a single prompt saving 15% of time.
- **Web Controls**: Chrome extension for dark mode and video controls.

EXPERIENCE

JSB GLOBAL INFOTECH | MERN Stack Developer | Certificate
May 2025 - May 2026 | Full Time
- Engineered React Native (Expo) applications and deployed React.js/Next.js web platforms, including admin panels and public-facing frontends.
- Architected backend microservices using Redis, Docker, AWS S3, and AI pipelines to automate scraping and rewriting automotive blogs/news content.
- Shipped web and mobile apps using REST APIs, RealmDB, and advanced MongoDB pipelines, reducing API response time by 35%.

CODEALPHA | Data Science Intern | Certificate
July 2024 - August 2024 | Remote
- Applied advanced feature engineering and selection techniques on real-world datasets, boosting ML model accuracy by 70% over the unengineered baseline.
- Enhanced model accuracy by 15% through iterative data visualization, hyperparameter tuning, and pipeline optimization across classification tasks.
- Achieved 90% precision on assigned tasks using Scikit-learn and Pandas, consistently delivering all project milestones ahead of deadlines.