Mayank Kumar Pokhriyal

Jersey City, NJ · (201) 492-7298 · mayankpokhriyal96@gmail.com

[LinkedIn](#) · [GitHub](#) · [Credly](#) · [Kaggle](#)

Summary

AI Engineer and Data Scientist with 5+ years of professional experience at IBM and Kyndryl and current M.S. student in Artificial Intelligence (graduating Dec 2026). Specialized in Agentic AI systems, Retrieval-Augmented Generation (RAG), LLM-powered APIs, and end-to-end AI product development. Strong background in healthcare-adjacent predictive modeling, multimodal deep learning, cloud-deployed ML systems, and production-grade backend engineering using Python, FastAPI, Docker, and AWS.

Education

Yeshiva University (Katz School of Science & Health), Manhattan, NY — Jan 2025 – Dec 2026
Master of Science in Applied Artificial Intelligence — GPA: 3.7/4.0

Birla Institute of Technology and Science (BITS Pilani), India — Sep 2020 – Sep 2022
Master of Technology in Data Science & Engineering — GPA: 7.5/10.0

Dr. A.P.J. Abdul Kalam Technical University, India — Aug 2015 – Aug 2019
Bachelor of Technology in Computer Science and Engineering — GPA: 7.2/10.0

Technical Skills

AI & LLMs: Agentic AI, LangChain, Ollama, OpenAI API, RAG, Prompt Engineering, Embeddings, Vector Databases
Machine Learning & Deep Learning: PyTorch, TensorFlow, CNNs, RNNs, Attention, Transformers, Multimodal Learning
Backend: Python, FastAPI, Flask, REST APIs, Authentication (JWT), SQLAlchemy ORM
Frontend: React.js (component-based UI, API integration)
Databases: PostgreSQL, SQLite, FAISS, ChromaDB
Cloud & DevOps: AWS EC2, AWS S3, Docker, GitHub Actions (CI/CD), Nginx
Tools: Git, Linux, VS Code, Jupyter, Kaggle

Professional Experience

Kyndryl (formerly IBM Global Technology Services), Bangalore, India — Sep 2021 – Jan 2025
AI/ML Engineer

- Designed, trained, and deployed large-scale fraud detection and credit-risk prediction systems using TensorFlow and XGBoost for high-volume financial transaction data.
- Built and maintained production-grade RESTful ML inference services using Python and FastAPI with modular architecture and authentication.
- Automated data ingestion and transformation pipelines processing millions of records per day, significantly reducing operational latency.
- Collaborated with cloud, DevOps, and data engineering teams to containerize and deploy ML systems using Docker and CI/CD pipelines.
- Delivered explainable AI outputs and monitoring dashboards for business-critical risk systems.

IBM, Bangalore, India — Aug 2019 – Sep 2021
Data Engineer

- Developed and optimized large-scale ETL pipelines for enterprise-grade financial datasets using Python, SQL, and workflow schedulers.
- Built analytical reporting dashboards enabling insight into multi-million-dollar operational and compliance systems.
- Partnered with ML teams to productionize predictive models into business data platforms.

Selected AI & Research Projects

Agentic AI Profile & Multi-Agent Systems GitHub
Developed structured multi-agent AI systems using LangChain-style orchestration with autonomous tool usage, memory, reasoning chains, and modular agent workflows. Focused on production-ready Agentic AI patterns aligned with modern LLM ecosystems.

Multimodal Stress Detection (WESAD Dataset) GitHub
Built a healthcare-focused deep learning model (CNN–BiGRU–Attention) for physiological stress classification using wearable sensor data (ECG, EDA, EMG), achieving 94% test accuracy. Included ablation studies, confusion-matrix analysis, and research-style reporting.

LLM-Powered Sentiment Analysis API GitHub
Implemented a fully containerized FastAPI-based sentiment analysis service using LangChain and Ollama. Exposed RESTful endpoints, structured outputs, and automated CSV/PDF/DOCX exports. Deployed with Docker and integrated CI/CD via GitHub Actions.

RAG-Based Financial 10-K Analyzer GitHub
Designed an end-to-end Retrieval-Augmented Generation pipeline for large SEC 10-K filings using FAISS vector indexing, document chunking, semantic retrieval, and local LLM inference. Enabled natural-language querying over long-context financial documents.

Circadian Rhythm Optimizer (MMASH Dataset) GitHub
Developed a personalized deep learning system for circadian rhythm prediction and optimization using wearable sleep, activity, and heart-rate data. Focused on solar-synchronous timing and health-oriented behavioral recommendations.

NeuroNap Sleep Optimization System GitHub
Built a full-stack ML-driven sleep optimization application with GUI (Tkinter), user authentication, SQLite-based structured data storage, and predictive scheduling models for improved sleep quality and energy levels.

Research & Platform Experience

- Experienced in building end-to-end LLM systems using SDKs, CLI tools, and local inference servers (Ollama).

- Hands-on with vector databases (FAISS, Chroma), embeddings, and semantic search pipelines for RAG-style applications.

- Developed RESTful AI microservices with authentication, persistent storage, and modular, cloud-ready architecture (FastAPI + Docker + AWS).