Tashfeen Raza

✉ razatashfeen045@gmail.com   📞 +92 307 5286602   in LinkedIn   ⌗ GitHub   📍 Rawalpindi, Pakistan

Summary

AI/ML Engineer (BSAI) skilled in building ML pipelines, deploying NLP models, and integrating LLM-based systems. Experienced with PyTorch, TensorFlow, Hugging Face, and LangChain. Deployed production models via FastAPI and designed feedback-driven learning systems for real-time applications.

Education

BSc Artificial Intelligence — National University of Modern Languages (NUML), Islamabad   Feb 2022 – Jan 2026
Relevant Coursework: Machine Learning, Deep Learning, NLP, Computer Vision, Speech Processing, Reinforcement Learning

Work Experience

LLM Integration & Product Enhancement Engineer — DASSOFT, NASTP   Oct 2025 – Dec 2025
- Enhanced AI-driven security monitoring by building NLP-based log parsing agent using LangChain and PyTorch
- Deployed ML models via FastAPI for real-time anomaly detection, improving incident response by 30%
- Designed evaluation pipelines with edge-case testing and feedback loops to refine model accuracy

AI/ML Intern — Siberkoza (Startup), NASTP   Jul 2025 – Aug 2025
- Built AI-powered SOC system to automate L1/L2 analyst tasks, reducing manual triage time by 40%
- Fine-tuned PyTorch models on domain-specific logs; integrated predictions via backend APIs
- Developed data pipelines for streaming log preprocessing, feature engineering, and model retraining

Projects

Smart Extract - AI Surveillance Clip Extraction System (Final Year Project)   *Code available upon request*
- Designed intelligent Web app to detect accidents & fire incidents from surveillance videos with severity classification (Low/High)
- Built stacked ensemble model combining VGG19, InceptionV3, ResNet50 achieving 94.68% fire detection & 88% accident detection accuracy
- Implemented temporal voting & confidence thresholding pipeline; auto-extracts event clips with 5-sec buffers using React GUI

PhishGuard - AI-Powered Phishing Detection Chrome Extension   ⌗ GitHub
- Built real-time phishing detection system combining rule-based analysis, VirusTotal, AlienVault OTX, and Google Gemini AI
- Developed Chrome extension (Manifest V3) with FastAPI backend; scans emails every 30s with 0-100 risk scoring
- Integrated Gmail API with OAuth 2.0 for secure read-only access; exports forensic reports to CSV

AI Voice Agent - Multi-Model RAG Platform   ⌗ GitHub — Live Demo
- Built enterprise voice assistant with real-time speech-to-text (Groq Whisper) and parallel LLM comparison (Gemini, Kimi, Llama 3.3)
- Implemented RAG pipeline with ChromaDB (79-chunk knowledge base) and async execution; deployed on Streamlit Cloud
- Integrated gTTS for text-to-speech output; achieved $0 monthly cost using free API tiers

Technical Skills

Languages & Tools: Python, SQL, C++, JavaScript, Git, Docker (basic), FastAPI, Flask, Streamlit, CustomTkinter
ML/DL Frameworks: PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face Transformers, LangChain, LangGraph
Specializations: Computer Vision, NLP, LLM/RAG systems, ensemble learning, model deployment, MLOps (basic)
APIs & Integrations: Gmail API, VirusTotal, AlienVault OTX, Google Gemini AI, Groq, OAuth 2.0
Databases: SQL, NoSQL, ChromaDB, FAISS (vector DB)

Certifications

- Machine Learning Specialization — Coursera (2024)
- Generative AI with RAG & LangChain — Coursera (2024)