Tashfeen Raza

AI Engineer (Agentic AI / LLM Systems)

razatashfeen045@gmail.com +92 307 5286602 LinkedIn GitHub Portfolio Islamabad/Rawalpindi, Pakistan

Summary

AI Engineer (BSAI) with hands-on experience focused on agentic AI systems: LLM orchestration, RAG, and production deployments using FastAPI. Built Gmail security automation and voice agents integrating Gemini/Groq models, vector databases (ChromaDB/FAISS), and third-party tools/APIs. Strong focus on reliability via stress-test prompts, structured outputs, and failure-mode analysis.

Education

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

Experience

LLM Integration & Product Enhancement Engineer (Contract) — Dassoft, NASTP, Rawalpindi Oct 2025 – Dec 2025

- Enhanced an AI-powered cybersecurity product by refining LLM prompts and workflows for accurate, executive-ready reports.
- Implemented new AI-driven modules aligned with business needs; collaborated on usability, scalability, and performance improvements.
- Designed stress-test prompts (edge cases, ambiguity, multi-step reasoning) and documented failure modes to improve consistency.

AI/ML Intern — Siberkoza (Startup), NASTP, Islamabad Jul 2025 – Aug 2025

- Built SOC automation assistants and retrieval-based Q&A workflows; integrated outputs into backend APIs.
- Worked with locally hosted models and domain data; contributed to evaluation/monitoring loops for production readiness.

Projects

AI Voice Agent — Multi-Model RAG Platform

- Built voice assistant with STT (Groq Whisper) and TTS (gTTS); orchestrates Gemini, Kimi (Groq), and Llama (Groq) for side-by-side responses.
- Implemented RAG with ChromaDB vector store + embeddings; async execution for parallel model calls and faster response times.

PhishGuard — AI-Powered Phishing Detection (Chrome Extension + FastAPI)

- Built a Gmail phishing detection system combining rules, threat intel (VirusTotal, AlienVault OTX), and Gemini AI.
- Chrome Extension (Manifest V3) + FastAPI backend with live monitoring, 0–100 risk scoring, alerts, and scan history.

Smart Extract — Surveillance Clip Extraction & Severity (Final Year Project) *Code available upon request*

- Offline tool to detect accident/fire events, classify severity (Low/High), and auto-extract buffered evidence clips.
- Ensemble (VGG19/InceptionV3/ResNet50) with temporal voting and confidence thresholding for robust detection.

Technical Skills

Agentic/LLM: LangChain, LangGraph, RAG, prompt engineering, tool orchestration, structured outputs
Vector DB: ChromaDB, FAISS **Models/Platforms:** Gemini, Groq (Whisper + hosted LLMs)
Backend/APIs: Python, FastAPI, Flask, REST APIs **Integrations:** Gmail API, OAuth2, VirusTotal, AlienVault OTX
Tools: Git, Docker (basic), Streamlit

Certifications

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