

# P. SHABREEN TAJ

AI Engineer · Generative AI & RAG Systems · Full-Stack AI Developer · Computer Vision

[pshabreentaj.com](https://pshabreentaj.com) · [github.com/taj-shabreen](https://github.com/taj-shabreen) · [linkedin.com/in/shabreen-taj77](https://linkedin.com/in/shabreen-taj77) · shabreen777shabreen777@gmail.com · Bangalore, India ·

Open to Remote

## PROFESSIONAL SUMMARY

AI Engineer and Generative AI builder with a CGPA of 9.16/10, specializing in production-grade Retrieval-Augmented Generation (RAG) systems, LLM orchestration, and full-stack AI application development. Designed and shipped a multi-source RAG research platform with FastAPI backends, React frontends, vector databases (ChromaDB, FAISS), and citation-grounded LLM responses using Gemini and Groq APIs. Complemented by deep expertise in computer vision (CNN, VGG16 transfer learning, Grad-CAM explainability) and AI security (real-time anomaly detection with Isolation Forest and Random Forest). Comfortable across the full AI engineering stack — from embeddings and model architecture to deployment-ready APIs and modern frontend interfaces — and actively seeking AI Engineer, Generative AI, AI/ML, Software Engineer, and Research internships.

## TECHNICAL SKILLS

Languages	Python, Java, C, JavaScript, SQL
AI / ML Frameworks	TensorFlow, Keras, PyTorch, Scikit-learn, HuggingFace Transformers, LangChain, LangGraph
Generative AI / RAG	RAG Systems, Prompt Engineering, Vector Embeddings, ChromaDB, FAISS, Vector Databases, LLM APIs (Gemini, Groq, OpenAI)
Full-Stack & Backend	FastAPI, React, Flask, REST APIs, MongoDB Atlas, WebSockets, Async Endpoints
Computer Vision	CNN, VGG16, ResNet-50, EfficientNet, Transfer Learning, OpenCV, Grad-CAM, Image Classification, Object Detection
AI Security & MLOps	AI Security, Anomaly Detection, Intrusion Detection Systems, Docker (learning), Git, Linux, Model Serving
Web & Tools	HTML, CSS, JavaScript, Jupyter, VS Code, Google Colab
Cloud	AWS (EC2, S3) — Prompt Engineering Certified, Azure ML (learning)

## PROJECTS

### NexusIQ Research Assistant | Generative AI · RAG Systems · LLM Engineering · Full-Stack AI

Present

Stack: React · FastAPI · ChromaDB · FAISS · LangChain · Gemini & Groq APIs · Python

- ▶ Architected a production-grade multi-source Retrieval-Augmented Generation (RAG) research platform enabling natural language querying over custom document corpora with citation-grounded, hallucination-resistant answers
- ▶ Built a full PDF ingestion pipeline with semantic chunking and dual vector storage across ChromaDB and FAISS for high-accuracy semantic search and retrieval
- ▶ Implemented a multi-stage context reranking pipeline to improve answer relevance and surface the most query-relevant document chunks before LLM generation
- ▶ Engineered a React frontend with real-time query streaming and citation highlighting, deployed via a FastAPI backend with async endpoints for a complete research assistant workflow
- ▶ Integrated LLM APIs (Gemini, Groq) with prompt engineering and context-injection templates, designing the system architecture for production deployment

### Multi-Source RAG Student Learning Assistant | Generative AI · LLM Orchestration · RAG

2026

Stack: LangChain · ChromaDB · Gemini · Groq · HuggingFace Embeddings · Gradio · Python

- ▶ Built a production-grade multi-source RAG chatbot that ingests PDFs, web content, and notes, performing semantic chunking and vector embedding storage in ChromaDB for accurate retrieval
- ▶ Orchestrated multi-LLM routing across Gemini and Groq APIs via LangChain, with prompt templates using context injection and source-grounded answering to minimize hallucinations
- ▶ Deployed an interactive Gradio interface enabling real-time Q&A, document upload, and conversational memory for students

## Brain Tumor Detection System | [Deep Learning](#) · [Medical Imaging](#) · [Computer Vision](#)

2025

Stack: [Python](#) · [TensorFlow](#) · [Keras](#) · [OpenCV](#) · [Transfer Learning \(VGG16/ResNet\)](#) · [Grad-CAM](#) · [Flask](#)

- ▶ Designed and trained a multi-class CNN classifier using VGG16 transfer learning on annotated MRI brain scans to detect and classify tumor types (glioma, meningioma, pituitary, no-tumor) with 94%+ test accuracy
- ▶ Implemented fine-tuning strategies (frozen layers + custom classification head) to significantly reduce training time versus training from scratch
- ▶ Integrated Grad-CAM explainability to visualize model attention regions for improved clinical interpretability, and deployed a Flask web app for real-time MRI classification with visual explanations

## Real-Time Malicious Activity Detection Engine | [ML Security](#) · [Anomaly Detection](#) · [Cybersecurity](#)

2025

Stack: [Python](#) · [Scikit-learn \(Isolation Forest, Random Forest\)](#) · [Flask](#) · [WebSocket](#)

- ▶ Built an ML-powered intrusion detection system for IoT-empowered infrastructures using Isolation Forest and Random Forest models to flag anomalous behavior patterns indicative of malware and privilege escalation
- ▶ Implemented a real-time WebSocket dashboard for live threat visualization and alerting on streaming system events

## AI Sales Effectiveness Prediction System | [Machine Learning](#) · [Predictive Analytics](#) · [Business AI](#)

2025

Stack: [Python](#) · [Scikit-learn](#) · [Pandas](#) · [Predictive Modeling](#)

- ▶ Built an AI-powered sales effectiveness prediction system that classifies business leads using machine learning and predictive modeling, applying data analytics to score lead quality and prioritize outreach

## EDUCATION

---

### B.Tech — Computer Science & Engineering (AI & ML) | [GITAM School of Technology, Bangalore](#) 2024 - 2027 (Expected)

CGPA: 9.16/10 · Relevant Coursework: [Deep Learning](#), [Neural Networks](#), [Computer Vision](#), [Data Structures & Algorithms](#), [Statistics for ML](#)

### Diploma — Computer Engineering | [BGS Polytechnic, Bangalore](#)

2021 - 2024

CGPA: 10.0/10 (Perfect Score) · Distinction Graduate

## CERTIFICATIONS & ACHIEVEMENTS

---

- ▶ Foundations of AI and Machine Learning — Microsoft, via Coursera (2026)
- ▶ Introduction to AI — Google, via Coursera (2026)
- ▶ Essentials of Prompt Engineering — AWS Training & Certification (2026)
- ▶ Networking Basics — Cisco Networking Academy (2025)
- ▶ Introduction to Cyber Security — Infosys Springboard (2023)
- ▶ Fundamentals of Cryptography — Infosys Springboard (2023)
- ▶ Fundamentals of Information Security — Infosys Springboard (2023)
- ▶ Network Fundamentals — Infosys Springboard (2023)

## CURRENTLY BUILDING & LEARNING

---

- ▶ Advanced RAG Architectures: Parent-Document Retrieval and multi-stage reranking pipelines
- ▶ Agentic Workflows & Multi-Agent System Architecture using LangGraph
- ▶ Cloud ML: AWS deployment pipelines and model serving at scale
- ▶ MLOps: Docker containerization, CI/CD for ML, experiment tracking

◆ [Open to Remote Internships Globally](#) ◆