

# Harshit Sharma

8384030533    harsh078976@gmail.com    linkedin.com/in/harshit-sharma    github.com/HARSHIT071004  
harshit-ai-portfolio.vercel.app

## PROFESSIONAL SUMMARY

---

Final-year AI/ML Engineer specializing in Computer Vision, RAG Systems, LLM Engineering, and AI Agent Development. Hands-on experience building real-time AI surveillance platforms, hybrid retrieval pipelines, multi-agent systems, and open-vocabulary vision models. Deep expertise in LLMs, Prompt Engineering, Vector Search, Docker, Kubernetes, and MLOps. Microsoft Hackathon Finalist with proven ability to ship production-grade AI systems end-to-end.

## TECHNICAL SKILLS

---

**AI/LLM Engineering:** LangChain, RAG Pipelines, Hybrid Retrieval, Semantic Search, FAISS, Cross-Encoder Rerankers, HuggingFace Transformers, Prompt Engineering, Claude API, AI Agents, Multi-Agent Systems, Embeddings

**Computer Vision:** YOLOv8, YOLO-World, Grounding DINO, OpenCV, Real-time Video Analytics, Open-Vocabulary Detection, Anomaly Detection

**ML/DL Frameworks:** PyTorch, TensorFlow, Keras, Scikit-learn, XGBoost, Deep Learning, Model Optimization, Hyperparameter Tuning, Feature Engineering, SMOTE

**Backend, DevOps & Data:** NestJS, Flask, REST APIs, MongoDB, MySQL, Docker, Kubernetes, CI/CD, Linux, n8n, Pandas, NumPy, Power BI, Matplotlib, Seaborn

**Languages & Tools:** Python, SQL, C++, JavaScript, TypeScript, React, Electron, Git    (*Familiarity: MERN Stack, Claude API integration*)

## PROFESSIONAL EXPERIENCE

---

### AI Developer Intern

Dec 2025 – May 2026

*Codeacious Technologies, On-site*

- Designed and deployed a real-time AI surveillance platform with open-vocabulary anomaly detection over live video streams, removing manual review — threats defined in natural language with zero model retraining.
- Engineered a production-grade RAG pipeline (FAISS dense + sparse keyword retrieval, cross-encoder reranking, semantic chunking) for accurate, source-grounded LLM responses over large document collections.
- Built a multi-LLM agent desktop app (NemoClaw fork, TypeScript/Electron) with sandboxed agents and Claude API integration, connected to Slack, Notion, and GitHub via prompt-engineered workflows.
- Containerized full AI stack (Docker/Kubernetes), automated alert escalation via n8n, benchmarked YOLOv8/YOLO-World/Grounding DINO, and delivered NestJS/MongoDB + React monitoring dashboard.

## KEY PROJECTS

---

### AI Surveillance System

*YOLOv8, YOLO-World, Grounding DINO, NestJS, React, MongoDB*

2025

- Production-grade real-time surveillance platform with open-vocabulary anomaly detection — threats defined in plain language, zero model retraining needed.
- Multi-model pipeline (YOLOv8/YOLO-World/Grounding DINO) auto-selects by use case; alerts routed via n8n; live monitoring on React dashboard.

### Advanced RAG Pipeline with Reranking & Hybrid Retrieval

*FAISS, LangChain, Transformers*

2025

- Architected a RAG system with FAISS dense + sparse retrieval, cross-encoder reranking, and semantic chunking for source-grounded, low-hallucination LLM responses.
- Experimented with multiple retrieval strategies; reranker re-scores candidate chunks before LLM ingestion, sharply improving answer precision.

### Crime Data Classification System

*Scikit-learn, XGBoost, SMOTE*

2025

- Ensemble classifiers (LR, RF, XGBoost) with SMOTE oversampling to resolve class imbalance, significantly boosting minority-class recall.
- EDA with temporal and geographic visualizations surfaced key crime patterns that directly drove feature engineering decisions.

### Portfolio Optimization Using Deep Learning

*TensorFlow, Keras, Financial APIs*

2025

- Implemented ANNs on historical market data to learn optimal asset allocation weights balancing return vs. downside risk.
- Captured non-linear market dynamics beyond classical methods; integrated financial APIs for live data ingestion and backtesting.

## EDUCATION

---

**B.Tech in Information Technology** — CGPA: 8.48/10

Expected: 2026

Gautam Buddha University, Greater Noida    *Coursework: ML, Deep Learning, DSA, Algorithms, DBMS, Statistics*

## CERTIFICATIONS & ACHIEVEMENTS

---

**Certifications:** SQL (HackerRank), Python & Pandas (Kaggle), Power BI, Advanced Excel, Machine Learning A-Z, Statistics for Data Science

**Achievement:** Finalist — Hackathons Ignite 2025 organized by **Microsoft** at Gautam Buddha University

## ADDITIONAL INFORMATION

---

**Languages:** English (Fluent), Hindi (Native)    **Interests:** LLM Research, Computer Vision, Open-Vocabulary Detection, AI Agents, Semantic Search, MLOps, AI Safety