# Harsh Shroff

**Machine Learning Engineer** — Open to Relocation

667-464-5255 harshrofff@gmail.com

linkedin.com/in/harshroff/ github.com/HarshShroff/ harshshroff.github.io/

## PROFESSIONAL SUMMARY

**Machine Learning Engineer** with hands-on experience building and deploying scalable ML systems and LLM-powered applications. AWS Machine Learning Associate Certified with proven track record delivering production AI systems achieving **60% efficiency gains** and **92% accuracy**. Expertise in end-to-end ML pipelines, cloud infrastructure, and transforming research into maintainable production code. Published researcher ready to drive ML innovation at technology companies.

## CORE COMPETENCIES

**Machine Learning & AI:** Generative AI, LLMs, RAG, LangChain, PyTorch, TensorFlow, Computer Vision, YOLOv8, OpenCV, NLP, Scikit-learn, Keras

**Deep Learning Specializations:** Neural Networks, CNNs, RNNs, Transformers, Fine-tuning, RLHF, Anomaly Detection, Reinforcement Learning, Time Series

**Software Engineering:** Python, C++, Java, SQL, FastAPI, Flask, Git, Docker, Kubernetes, CI/CD, API Development, Microservices, Test-Driven Development

**Cloud & MLOps:** AWS, SageMaker, Lambda, Bedrock, Textract, EC2, S3, MLflow, Weights & Biases, Model Monitoring, A/B Testing

**Data Engineering:** Pandas, NumPy, Spark, Hadoop, Feature Engineering, Data Pipelines, ETL, PostgreSQL, MongoDB, Redis

**Deployment & Tools:** Streamlit, Gradio, Jupyter, CUDA, Edge Computing, NVIDIA Jetson, Model Serving, Performance Optimization

## PROFESSIONAL EXPERIENCE

### AI/ML Researcher — Mar 2023 – Present

*UMBC Center for Real-time Distributed Sensing and Autonomy*

- **Led 5-engineer team** developing distributed ML systems for multi-modal sensor data fusion under US Army Research Lab funding, achieving **20% performance improvement** in real-time optimization systems. *Python, PyTorch, TensorFlow, Distributed Training*
- Designed production monitoring dashboards and real-time inference pipelines enabling autonomous systems deployment with computer vision and sensor fusion capabilities. Demo Video *Dash, Flask, OpenCV, Model Serving*
- Engineered production-ready multi-modal ML systems integrating advanced perception algorithms for next-generation AI applications using deep learning frameworks and edge computing platforms. *Computer Vision, Real-time Systems, NVIDIA Jetson*

### ML Engineer - Production AI Systems (Contract) — Jun 2024 – May 2025

*VITG Corp., Halethorpe, MD*

- Architected and deployed production **LLM automation system** on AWS reducing candidate screening time by **60%** for 200+ employee organization, processing 2,500+ regulatory documents with 40% efficiency improvement. *Python, AWS Lambda, Claude, Llama, Streamlit*
- Built commercial conversational AI chatbot using Claude 3.5 and AWS Textract with scalable ETL pipelines and RESTful APIs for real-time geospatial data processing and regulatory compliance automation. *FastAPI, PostgreSQL, Docker, AWS Textract*

### Data Scientist - Computer Vision — Aug 2023 – Dec 2023

*The Conservation Fund, Shepherdstown, WV*

- Developed production-ready **computer vision system** achieving **92% accuracy** using YOLOv8 and OpenCV, deployed on edge devices for real-time quality assessment with end-to-end ML pipeline from data collection to model serving. *Computer Vision, Edge Deployment, MLOps, Raspberry Pi*
- Co-authored peer-reviewed research demonstrating commercial impact of deep learning applications, enabling commercial deployment of AI quality control systems with automated reporting capabilities. *Research Publication, Production Deployment, Technical Writing*

## KEY TECHNICAL PROJECTS

**Enterprise Document Intelligence Platform**

*AWS, Textract, S3, Lambda, Python, LLM APIs, Streamlit*

- Built scalable document processing system automating extraction from 2,500+ PDFs using LLM-powered OCR pipelines and cloud infrastructure with real-time API endpoints and automated compliance workflows.
- Implemented interactive dashboards enabling stakeholders to query and visualize insights from unstructured data with automated reporting and decision-support capabilities.

**Multi-Modal Recommendation Engine** (GitHub)

*Python, OpenAI API, Computer Vision, PostgreSQL, A/B Testing*

- Engineered hybrid recommendation system combining structured data analysis with computer vision, reducing user bounce rate by **20%** and increasing engagement by **40%** through personalized AI-driven recommendations.
- Implemented scalable data pipelines and A/B testing framework for continuous model improvement and performance monitoring with real-time analytics and user feedback integration.

**AI-Powered Medical Diagnosis System** (GitHub)

*Python, TensorFlow, Computer Vision, Flask, Gradio*

- Designed end-to-end ML web application for medical image classification achieving high diagnostic accuracy using deep learning and computer vision techniques with automated feature extraction.
- Deployed production-ready inference pipeline with interactive web interface, enabling real-time medical assessment and automated reporting capabilities for healthcare applications.

**Real-Time Analytics Dashboard** (GitHub)

*Python, Dash, Machine Learning, Statistical Analysis, Plotly*

- Built interactive ML-powered analytics platform with real-time data ingestion, model inference, and visualization capabilities for performance optimization and predictive insights.
- Integrated predictive analytics improving performance outcomes by 20% through data-driven insights, automated reporting, and statistical modeling techniques.

**Gesture-Based Control System** (GitHub)

*Python, OpenCV, Real-time Processing, Edge Computing, NVIDIA Jetson*

- Developed real-time gesture recognition system using computer vision and machine learning for intuitive device control and human-computer interaction with low-latency inference.
- Implemented efficient edge computing solution enabling responsive control in resource-constrained environments with optimized model deployment on embedded hardware.

## EDUCATION

**Master of Science, Data Science** **Aug 2022 – May 2024**

*University of Maryland Baltimore County (UMBC) —* ***GPA: 3.8/4.0***

- **Relevant Coursework:** Machine Learning, Deep Learning, NLP, Computer Vision, Big Data Systems, Statistical Analysis, Predictive Modeling

**Bachelor of Engineering, Electronics & Communication** **Jun 2019 – May 2022**

*Gujarat Technological University —* ***GPA: 3.8/4.0***

- **Relevant Coursework:** Machine Learning, Python Programming, Data Structures, Algorithms, Embedded Systems

## CERTIFICATIONS & PUBLICATIONS

### Professional Certifications

- **AWS Machine Learning Associate Certification** – View Credential – Active through Apr 2028
- **AWS Machine Learning Foundations** – Core ML concepts, cloud deployment, and AWS service integrations
- **NVIDIA Deep Learning Institute** – Transformer-Based NLP Applications, GPU-Accelerated Computing

### Publications & Research

- **Ranjan, R.**, **Shroff, H.**, et al. (2024). "FilletCam AI: Precision color profiling using deep learning." *Journal of Agriculture and Food Research*. DOI: 10.1016/j.jafr.2024.101461
- **Trivedi, K.**, **Shroff, H.** (2021). "Mosquito identification using ML on embedded systems." *IEEE ITU Kaleidoscope Conference*. DOI: 10.23919/ITUK53220.2021.9662116