

Harishraj Udaya Bhaskar

339-216-7090 | uharishraj@gmail.com | [linkedin.com/in/uharishraj](https://www.linkedin.com/in/uharishraj) | github.com/hatish2001

SUMMARY

AI renegade like Tony Stark but with less ego and more PyTorch who's built systems that outsmart regulators, slashed costs like a budget ninja, and turned chatbots into customer-whispering geniuses—ready to unleash hell on your next big project.

PROFESSIONAL EXPERIENCE

AI Software Engineer

CustomStacks

Mar. 2025 – Present

Los Angeles, CA

- Built AI-powered compliance system that identifies mislabeled consumer products containing carcinogens, enabling enforcement organizations to file 150+ legal cases protecting public health.
- Architected and deployed production multi-agent AI pipeline on AWS that automated regulatory analysis, reducing manual review time from 40 hours to 2 hours per product cycle and increasing analysis capacity by 400%.
- Led full-stack development from research to production deployment, implementing async backend, CI/CD automation, and monitoring infrastructure that supports 50+ concurrent compliance investigations.

AI Software Engineer

Factoryspace.ai

Jan. 2025 – Mar. 2025

Boston, MA

- Developed and deployed AI-powered customer support system using fine-tuned Mistral-7B LLM that handles 10,000+ natural language queries daily about order status, product specifications, and manufacturing capabilities - reducing support ticket volume by 70% and response time from 2 hours to 30 seconds.
- Architected and deployed custom fine-tuned language model (Mistral-7B) using advanced ML techniques (model distillation, QLoRA, chain-of-thought reasoning) that delivers GPT-4-level responses at fraction of the cost on high-volume workload.

Data Science Intern

American Fidelity

Jul. 2023 – Jan. 2024

Oklahoma City, OK

- Built and deployed AI-powered customer service chatbot for insurance company that handles policy inquiries across 10,000+ internal documents, reducing support ticket volume by 45% and enabling 24/7 automated customer support using GPT-4.
- Architected production RAG pipeline with document chunking, semantic search, and vector embeddings deployed on Azure with RESTful API integration - processing 5,000+ monthly customer queries with 90% answer accuracy and 2 second response time.
- Developed machine learning models (XGBoost, regression) for SBA disaster loan validation and amount prediction used by insurance agents, improving claim processing speed by 50% and reducing manual underwriting time from 3 hours to 30 minutes.

Software Engineer

CGI Inc.

Mar. 2021 – Aug. 2022

Remote

- Built AI-powered case resolution recommendation system for HR operations that suggests solutions to employee complaints, requests, and issues based on 50,000+ historical cases - reducing average case resolution time by 50% and successful first-time resolutions by 23% using collaborative filtering at 50ms latency.
- Developed OCR and NLP pipeline processing 10,000+ HR documents (resumes, contracts, benefits forms) with 98% extraction accuracy, reducing manual review time by 75% and enabling automated candidate screening for recruiting platform.
- Engineered MLOps infrastructure with Airflow, Docker, and Kubernetes for automated data ingestion, model retraining, and monitoring across 15+ production models - reduced deployment time from weeks to hours and enabled continuous model improvement.

RESEARCH & PUBLICATIONS

Conversion of Audioless Video to Speech Using AV-HuBERT Algorithm

International Conference on Intelligent and Advanced Computing Systems (ICIACS), Springer

May 2024

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, TypeScript

ML/AI Frameworks: PyTorch, TensorFlow, Scikit-learn, XGBoost, Hugging Face Transformers

LLMs & NLP: GPT-4, Mistral, DeepSeek, Fine-tuning (LoRA, QLoRA, PEFT), RAG, Prompt Engineering, spaCy, NLTK

Agent Frameworks: LangChain, LangGraph, Multi-Agent Orchestration

Cloud & Infrastructure: AWS (EC2, RDS, S3, SageMaker), Azure (Databricks, App Service), Docker, Kubernetes

MLOps: Airflow, MLflow, CI/CD Pipelines, Model Monitoring, Feature Stores

APIs & Web: FastAPI, Flask, REST API, GraphQL, WebSockets

EDUCATION

Northeastern University

Master of Science in Artificial Intelligence

Boston, MA

Sep. 2022 – Dec. 2024