Harshita Dooja Poojary

harshita1.poojary@gmail.com | +1(213)-234-8695 | in harshitapoojary

Solution-focused engineer with 4.5 years of experience in Python development, ML pipelines, and system integration, engaging stakeholders, translating complex concepts into actionable solutions, and optimizing client workflows through analytics and scalable platforms.

EDUCATION

University of Southern California, Los Angeles *Aug 2023 – May 2025*

Master's in Computer Science (Coursework-*Algorithms, Deep Learning, Artificial Intelligence, Machine Learning, Natural Language Processing*) **GPA: 3.81**

University of Mumbai, India *July 2015 – May 2019*

Bachelor's in Computer Science Engineering (Coursework- *Software Engineering, Distributed Databases, Web Technologies*) **GPA: 3.83**

SKILLS

Core Programming Languages: Python

Other Programming Languages: JavaScript, C++, Java, HTML/CSS, Scala

Frameworks & Libraries: Node.js, AngularJs, PySpark, PyTorch, Keras, TensorFlow, OpenCV, Streamlit

Databases: MongoDB, Redis, MySQL, ElasticSearch, Milvus, Hadoop, DynamoDB

Tools & Platforms: Docker, Kubernetes, AWS, Azure, Jenkins, GitHub, GitLab, JIRA, Postman, Swagger, PowerBI, Tableau, Google Analytics

WORK EXPERIENCE

AI Intern *July 2025 – Present*

Deep Defense Solutions (Remote) USA

- **Developing** ML pipelines in Python to detect false incident reports for the community management system and taking action on the incident.

Data Science Intern *Jun 2024 – Aug 2024*

Zero-True (Remote) New York, USA

- **Collaborated** with founders and client teams to design SQL/Python-integrated dashboards, presenting insights through interactive demos and advising on best practices for operational efficiency.
- **Implemented 5 Vuetify UI components** that reduced client reporting errors by 15% as measured by accuracy metrics across accounts.
- **Optimized GPU inference** efficiency by 30%, enabling scalable deployment of intensive workflows under constrained resources..

Student Researcher *Jan 2024 – Aug 2024*

University of Southern California Los Angeles, USA

- Under **Prof. Seon Kim**, analyzed **object detection and classification models** (YOLOv5/YOLOv8) on 10k Streetview imagery from Google Images and in-house dataset using **data mining** for tent detection and counting, optimizing performance in **noisy urban conditions** using geospatial metadata.
- **Executed geospatial analysis at scale** by coding grid-based segmentation logic to quantify encampment density across **502.7 square miles** of Los Angeles, generating multi-intensity **heatmaps** using latitude-longitude clustering **across five intensity levels**.

Machine Learning Engineer *Dec 2021 – Jul 2023*

Reliance Jio Mumbai, India

- **Engineered and optimized** RetinaFace detection models using **TensorRT FP16** precision on **T4 and A100 GPUs** using **C++ and Python**, reducing face registration time by **40%** and enabling real-time inference at scale across Jio's parallel distributed clusters.
- **Accelerated** deployment velocity by 50% as measured by rollout metrics by automating CI/CD pipelines in Python with Docker, Kubernetes, GitLab, and JIRA for three ML systems—JioFace, Video Motion Detection, Number Plate Recognition.
- **Delivered spoof detection** with **90% accuracy** as measured by production benchmarks by implementing ResNet/CDCNet models and scaling to **20+ sites**.
- **Directed 4 interns** to curate datasets and validate models, ensuring reproducibility and deployment readiness across production environments.

Software Engineer *Jun 2019 – Dec 2021*

Reliance Jio Mumbai, India

- **As a part of the world's largest mobile data company, collaborated** cross-functionally to build scalable microservices and REST APIs serving personalized ML-driven recommendations for 100M+ users, enhancing engagement by 20% using large-scale LLM models.
- **Designed** and **optimized** large-scale ingestion pipelines/templates for multi-device content delivery by **crafting** advanced SQL queries across fragmented relational datasets, improving throughput and aggregation performance.
- Analyzed user behaviors across 10 campaigns for marketing **personalized advertisements, leveraging regression and time-series models to identify engagement patterns** and optimize targeting strategies, which improved click-through rates and campaign ROI.
- **Advised** product managers and frontend teams on API integration and content delivery strategies, **reducing latency by 15%** as measured by system monitoring tools and **boosting user engagement by 12%** through streamlined backend-to-frontend data flows.

Software Engineer *Jan 2019 – Jun 2019*

TaksyKraft (Remote) Hyderabad, India

- **Developed** engagement prediction APIs that improved ROI by 10% as measured by advertiser adoption by delivering real-time targeting solutions for 10+ clients.
- **Engineered** data pipelines and Tableau dashboards, reducing reporting delays by 40% as measured by turnaround times across enterprise accounts.

PROJECTS

Unified Semantic Space for Multimodal Retrieval (PyTorch, Python, VectorDB)

- Analyzed multimodal LLMs, including **LLama-3,** to answer questions over frames for MIT videos and using **LangDB** Vector Databases to store frame embeddings and retrieve user queries.
- Implemented a user interface using **Streamlit** to integrate ML models, enabling users to upload videos and query the system interactively.

Analyzing NLP techniques on Low-Resource Languages(PyTorch)

- **Conducted idiom classification and metaphor classification** in the low-resource **Konkani** language using BERT-based transformers, investigating performance-preserving **attention head pruning.**

Comparison of Model Performance with Knowledge Distillation and Quantization (PyTorch, Python)

- **Fine-tuned LLama-3 model** with 4-bit quantization using **logit distillation** with LoRa (Low-Rank Adaptation).
- Computed benchmark against **GSM8k dataset with 8-shot learning** to analyze the performance against the original model with 8B parameters.