

HARSH MANKODIYA

• +1 (602)-517-6750 • hankodi@asu.edu • linkedin.com/in/harshmankodiya • github.com/hmankodiya

Education

Arizona State University

August 2023 - May 2025

Master of Science, Computer Science: GPA 3.70

Tempe, USA

Courses: NLP, Statistical Learning, Artificial Intelligence, Data Mining

Institute of Technology, Nirma University

August 2019 - May 2023

Bachelor of Technology, Computer Science Engineering

Ahmedabad, India

Courses: Machine Learning, Deep Learning, Data Structures, Linear Algebra, Calculus, Probability and Statistics

Professional Experience

Cellino Biotech

May 2024 - August 2024

Machine Learning Intern

Cambridge, USA

- Developed a proof of concept for a central embedding model for patch selection, anomaly detection, cell segmentation and cell classification.
- Fine-tuned **DinoV2** using **Vision Transformer** based heads for downstream segmentation tasks, achieving average F1-Score of **82%**. Utilized **Weights & Biases** for experiment tracking, artifact logging and hyperparameter sweep.
- Performed embeddings decomposition using **t-SNE** and **PCA** and employed **GMM clustering** to perform zero-shot cell artifact detection.
- Integrated **GCP** API calls with **PyTorch Dataset** utilities to streamline **Zarr** to **Tensor** conversion. Added a local caching mechanism improving throughput.
- Automated the retrieval of artifact metadata from a **PostgreSQL** database and integrated it into the pipeline for validating clustering efficacy and artifact detection workflows.
- Containerized the inference pipeline with **Docker**, enabling real-time data processing and easy integration with cloud-based services.

Lens Lab, Arizona State University

August 2023 - May 2024

Research Assistant

Tempe, USA

- Integrated **eXplainable AI** techniques with autonomous vehicle agents to enhance post-hoc explainability within **Carla**, and **Gymnasium** simulation environments.
- Trained **Proximal Policy Optimization**, using **StableBaselines3**, incorporating **VAE**-based feature extraction to process image streams.
- Utilized the pre-trained **CLIP** models to generate **zero-shot** segmentation masks, enabling efficient concept sampling across multiple policy rollouts.
- Led the project, culminating in a publication at the **NeurIPS 2024 SATA Workshop**, focusing on integrating explainability into decision-making processes for autonomous robotic systems.

Bosch

January 2023 - May 2023

Research Intern

Bangalore, India

- Formulated working principal for **GradCAM** based **Knowledge Distillation** algorithm for **image segmentation** models.
- Utilized **PyTorch Lightning** to automate **data-processing**, **model training**, **evaluation**, and **inference**. Integrated **MLFlow** for experiment tracking and model registry.
- Trained SegNet and U-Net segmentation models on **NVIDIA DGX A100** systems, achieving high relative **IoU** scores exceeding **85%** across multiple datasets.

Relevant Projects

Multilingual Sentiment Classification using LLMs | *Python, PyTorch, HuggingFace*

Dec 2024

- Conducted **PEFT** on **Llama2-7B**, utilizing **Quantized Low-Rank Adaptation (Q-LoRA)** to achieve **4-bit quantization**, reducing trainable parameters by approximately **0.60%**.
- Fine-tuned **Llama2-7B** on just **2%** of a multilingual sentiment dataset spanning 12 languages, categorized into three classes: positive, neutral, and negative.
- Witnessed **30%** boost in **test AUC** and a **20%** increase in **test accuracy**.
- Performed a comparative analysis by fine-tuning **GPT2** and **BERT**, highlighting their relative performance.

Technical Skills

Languages

- Python, C++, Shell, C, MATLAB, SQL, JavaScript, JAVA, Docker, Git

ML Frameworks

- PyTorch, Lightning, Jax, TensorFlow, scikit-learn, LangChain, Stable-Baseline3, HuggingFace, Keras, ONNX, Gym, Flax, Einops, XGBoost

Python Libraries

- NumPy, SciPy, Pandas, Albumentations, OpenCV, Pillow, ImageIO, Zarr, Dask, Seaborn, Matplotlib, Plotly, W&B, MLFlow, PySpark

ML Techniques

- LLMs, Text Classification, RAG, Knowledge Distillation, Reinforcement Learning, CLIP, Image Captioning, Image Classification, Image Segmentation, VAE, GANs, Style Transfer, GradCAM, TCAV