Akash Pawar

Jersey City, NJ | akashpawar9619@gmail.com | (551) - 328 - 5269 | LinkedIn | Github

PROFESSIONAL EXPERIENCE

Rajiv Gandhi Institute of Technology, Mumbai, India
Machine Learning Project Lead — August 2022 - March 2023

- Designed a mental health **chatbot** using a two-stage **RoBERTa(110M)** classification system for depression detection.
- Refined RoBERTa-base for initial depression detection achieving **98.5% accuracy** and **98.4% F1-Score**.
- Tweaked RoBERTa-base for mental health categorization achieving **98.2% accuracy** outperforming traditional ML models and baseline models by **~8%** and **~30%** respectively.
- Integrated **BlenderBot-400M** for context-aware response generation in general conversations, unrelated to depression.
- Built a curated **response database** covering **80 different intents** for accurate mental health support.
- Engineered an efficient context management system that **preserved conversational coherence** by maintaining a 3-turn sliding window within RoBERTa's 512-token limit, enabling **scalable deployment** on Discord.

Stevens Institute of Technology, Hoboken, New Jersey

Deepseek for Advanced Mathematical Reasoning — Feb 2025

- Fine-tuned **DeepSeek-R1-Distill-Qwen-1.5B** using Unsloth optimization framework and LoRA adaptation techniques.
- Programmed two data preprocessing pipelines using **pandas** as well as **filter/map** functions, one suitable for machines with lower RAM (~16GB) and another for higher RAM machines.
- Enhanced mathematical reasoning capabilities, enabling systematic step-by-step solutions, even outperforming **Claude-3.5 Sonnet** on a reasoning task, where the base model failed.
- Optimized training pipeline to achieve **2x faster inference speed** while reducing trainable parameters from **1.5B to 18.5M (98.8% reduction)** through LoRA adaptation.

Protein Subcellular Localization Predictor using ESM2 — Feb 2025

- Developed a protein subcellular localization predictor using **Meta AI's ESM2-3B**, achieving **84.79% top-3** and **92.09% top-5** accuracy across 12 cellular locations.
- Boosted training efficiency by **~1.8x** using **mixed precision** and **gradient checkpointing** with A100, reducing memory footprint by **~50%** and enabling stable training.
- Structured an end-to-end pipeline, integrating **data acquisition** via UniProt's API and **robust evaluation workflows**.
- Demonstrated expertise in **protein bioinformatics** and deep learning for **biological data**.

Dependency Chain Analysis — Dec 2024

- Processed **500,000+ relationships across 442,275 nodes** in a large-scale dependency graph using **Neo4j** and **NetworkX**, extracting key structural patterns in **under 25 seconds**.
- Engineered **5 topological features** and **20 semantic attributes**, including centralities, local risk ratio, scope, and security metrics, contributing **~80%** to classifier feature importance and **enabling 100% accuracy** in critical node identification.
- Implemented **Node2Vec embeddings (128-dim)** combined with handcrafted features, achieving **100% precision, recall, and F1-score on 88,455 nodes using Random Forest**, compared to a baseline F1-score of 29% without custom features.

RAG for Document QA — Dec 2024

- Constructed a document QA system using **LangChain**, **Chroma vector store**, and **OpenAI embeddings**.
- Built a **React** frontend and **FastAPI** backend with comprehensive **PyTest** coverage spanning **6 major test cases**.
- Deployed the system on **AWS** using **Docker** after achieving **100% precision on retrieval evaluation.**

EDUCATION

Stevens Institute of Technology, Hoboken, NJ — September 2023 - December 2024
Master of Science, Machine Learning — Cumulative GPA: **3.93/4**
Notable Coursework: DL/ML, NLP, Computer Vision, LLMs, Text Mining, Statistical Machine Learning.

Mumbai University, Mumbai, India — August 2019 - May 2023
Bachelor of Engineering, Computer Science — Cumulative GPA: **9.07/10**
Notable Coursework: Data Science, Tableau, Linux, Big Data Analytics, Cloud Computing, Distributed Computing.

SKILLS

- **Programming & Data Manipulation/Visualization:** Python, SQL, pandas, numpy, seaborn, Tableau, Plotly, matplotlib, NetworkX, Neo4j, PIL, DuckDB, Dask, OpenCV, Streamlit
- **ML Frameworks/Tools:** PyTorch, Tensorflow, keras, scikit-learn, XGBoost, LightGBM, Hugging Face, MLX, stable-baselines, Wandb, MLflow
- **NLP:** OpenAI, LangChain, Unsloth, NLTK, Chroma, spaCy
- **MLOps & Cloud Platforms:** Docker, AWS (S3, Lambda, SageMaker), Flask, Git, Gradio and FastAPI