PRATHAM SARAF

India · prathammsaraf@gmail.com · +91-8982597809 · GitHub ↗ · Kaggle ↗ · Portfolio ↗ · Linkedin ↗

EDUCATION

Indian Institute of Information Technology   Bhopal
BTech Computer Science   December 2021 - Present

WORK EXPERIENCE

TheReliable.ai   Jan 2024 - June 2024
Founding Machine Learning Intern

- Designed and implemented a multi-agent system on **AWS with LangGraph** for natural language database interactions, achieving over **80% query success rate with 100% accuracy**.
- Integrated large language models like **Mistral 7B, LLaMA 70B, and Stable LM 3B** into a multi-agent system for natural language understanding, query generation, and insight derivation, optimizing performance and cost-effectiveness through model selection and **LoRA finetuning**

LancerNinja   May 2023 - October 2023
Internship Certificate ↗

- Led development of end-to-end AI solutions as Machine Learning Intern, leveraging **OpenAI**, **LLamaIndex** , **LangChain** Incorporating **CromaDB** , **PineconeDB** to provide Vector database solutions. Including Image Generation Models , RAG pipelines .
- Developed Automated free form document detection to create an advanced RAG pipeline utilizing reflexive agents, which were utilized in healthcare domain effectively **providing 95% accurate detection rates**

Indian Institute of Technology, Madras   July 2023 - Jan 2024
Research Intern under Prof (Dr.) Rupesh Nasre

- Engineered a graph convolutional network model in **StarPlat** by crafting highly optimized functions for forward and backward passes; significantly improved model efficiency and reduced **computation time by 25%**
- Researched academic papers on utilizing StarPlat DSL for efficient parallelization of graph neural networks on multi-core CPU and GPU systems

Red Positive   November 2022 - January 2023
Machine Learning Intern

- Collected a total of **370 thousand** sentences in 37 different dialects of India with approx **10 thousand** sentence in each language to build a language detection system
- Used **35 Indian languages** for document translation model which took advantage of parallel corpora . Speech detection system for local Indian dialects which has an accuracy of over **90%**

Acciolbis   September 2022 - November 2022
Internship Certificate ↗ LOR ↗

- Improved document retrieval time to under 5 seconds through semantic searching and scoring, and optimized memory storage in **Haystack systems**.
- Employed **GCP** to infer models and evaluated various open source text to image models and their variants.

Indian Institute of Information Technology, Bhopal   November 2022 - August 2023
Research Intern under Prof (Dr.) Bhupendra Singh Kirar

- Developed and implemented Neural Networks for the research paper, which involved utilizing convolutional neural networks and transfer learning techniques.
- Analyzed the performance of the classifier using various evaluation metrics, including accuracy and specificity, and presented the findings in the research paper.

PROJECTS

MultiModalGNN-SentimentAnalysis   April 2023 - June 2023
Colab Notebook ↗

- Developed a MultiModalGNN-SentimentAnalysis model using **Graph Attention Mechanism, BERT, VGG16, and Graph Neural Network** to perform sentiment analysis on a multimodal dataset, creating a graph with **2,060,892 edges** and **39,109 nodes**, and achieved **60% accuracy** by leveraging an ensemble of neural networks for accurate sentiment classification.

  Skills Used: Python, NetworkX, DGL, OpenAI, GPT-3,Prompt engineering

SKILLS

Programming Languages:	Python, C++ ,C, R, Mojo
Libraries / Frameworks:	FastAPI , Tensorflow , Pytorch , Pytoch Geometric , OPEN-AI GYM,
	DGL , NetworkX, Dask , Pandas
Cloud Platforms:	AWS , GCP , Azure
Soft Skills:	Leadership, Communication Skills, Organised
Langues:	German , English