

GAURAV BHARDWAJ

Jamshedpur, Jharkhand, India

+91 6204220436 Email LinkedIn GitHub Kaggle

Work Experience

Outlier

Aug 2024 – Oct 2024

AI Trainer

Remote - San Francisco, CA

- Developed detailed annotations for over 500 data points used in training AI models, identifying key areas for optimization that reduced output inaccuracies by 30%.
- Streamlined the AI-generated content review process through structured feedback loops, improving response clarity and consistency across automated interactions.
- Collaborated with cross-functional teams to ensure model outcomes aligned with business requirements, contributing actionable data insights for the business team.

Education

INDIAN INSTITUTE OF TECHNOLOGY, DHANBAD

Dec 2021 - May 2025

Bachelor of Technology (B. Tech) in Electrical Engineering

Dhanbad, India

RAJENDRA VIDYALAYA, JAMSHEDPUR

Class-10 ICSE (95.8 %)

2017-2019

Class-12 ICSE (93.75 %)

2019-2021

Projects

Sentiment Analyser | Github | Python, Scikit-Learn, Pytorch, beautifulsoup

- Developed and implemented a robust sentiment analysis model using **LSTM**, **BERT**, & **NLTK**, processing data from 60K+ tweets to provide actionable insights that influenced trading strategies for 25 high-profile stocks.
- Integrated advanced data-driven sentiment signals into portfolio risk analytics, enhancing the predictive accuracy and enabling more informed investment decisions across various asset classes.

STEM Solver | Python, FastAPI, pytorch

- Developed OpenELM-450M base model for STEM/math tasks(8GB vram), achieving **2.36%** higher accuracy than OLMo with **2x** fewer pre-training tokens.
- Strong on ARC-c(42.24), MMLU(26.76), Hellaswag(73.28) benchmarks.
- Enhanced performance through fine-tuning using PEFT/LoRA, synthetic data and DPO with **Metal** optimization, deploying final model via FastAPI for inference.

Real-Time Data Pipeline | Amazon EC2, kubectl

- Set up and managed Kafka and Zookeeper on Kubernetes using Helm charts, making sure the system stayed available and could handle failures.
- Packaged Python-based data producers and consumers with Docker, making them easy to run and scale across system.
- Created a real-time data pipeline with Kafka and Kubernetes to move data quickly and reliably to machine learning systems.

Achievements

- * Ranked **14th out of 5000+** participants data science competition by IITG, achieving significant performance improvements in an Faster-RCNN, trained on 1,800+ HD images to detect automobile injuries (68%). **Project Link**
- * Analyzed 10K data samples addressing loan defaults in quant research program at **JPMorgan Chase**; crafted effective strategies for risk prediction by enhancing existing models.
- * **Global rank 142** in CodeChef Starters 66.
- * **Global rank 1740** in Google Kickstart Round H 2022.
- * **All India rank 5963** in JEE Advanced 2021.

Courses and Certifications

Deep Learning Essentials with Keras, Data Visualization with Python, Operations Research, Generative AI with Large Language Models, Applied Prompt Engineering, Data Structures and Algorithms, Topics in Mathematics with applications in Finance (MIT OCW), Probability and Statistics

Technical Skills

Languages: Python, C, C++, Swift, bash

Frameworks & Libraries: AWS, GCP, Sagemaker AI, Pytorch, LangChain, Crew, Terraform, Tableau, GIT, Github

Others: Spark, Kafka, Flink, Docker, Airflow, Time Series Analysis, RAG, Prompt Engineering