Debasish Dhal

 Hyderabad | 📞 +91 8917548570 | [E-mail](#) | [LinkedIn](#) | [GitHub](#) | [HuggingFace](#) | [Scholar](#) | Joining Time : 1 month

PROFILE SUMMARY

A versatile and results-driven professional with hand-on experience in web development, natural language processing (NLP), retrieval-augmented generation (RAG), and fine-tuning large language models (LLMs).

EDUCATION

National Institute of Science Education and Research, Bhubaneswar, India	CGPA: 6.96/10.0
Master of Science in Physics, Minor in Mathematics	*August 2017-May 2023*
National Institute of Technology, Rourkela, India	CGPA: 7.97/10
B.Tech in Chemical Engineering (Voluntarily left after one year)	*2016-2017*
Ravenshaw College, Cuttack, India	Percentage: 87.33%
Higher Secondary Education in Science	*2014 - 2016*

EXPERIENCE

AI Engineer | GenAI Freelancer — May 2024 - Dec. 2024

Generative AI, LangChain, LangGraph, RAG, FastAPI, React, GraphQL, MongoDB

- Built ShopIntel, an AI-powered e-commerce ChatBot for Shopify stores using **RAG**. Demo ↗
- Optimized RAG pipeline using **AI Agents** and **Elastic Search**.
- Customized ChatBots for clients in under five minutes, enhancing customer engagement.

AI Researcher, OdiaGenAI Team — September 2023 - Present

LangChain, LangGraph, RAG, LLM, Fine-tuning, React, ChromaDB

- Developed OdiaLlama, a **fine-tuned** Llama model for the Odia language, using **LoRA** technique. Demo ↗
- Developed *Acharya*, a **RAG**-based AI-tutor, aimed at helping Hindi-medium students. Demo ↗
- Fine-tuned various families of LLMs (Llama, Mistral, Qwen, Gemma etc.) for Indic languages. Models ↗
- Curated, pre-processed and shipped various datasets for LLM training. Gen-AI blogs written by me are here ↗.

Intern, OdiaGenAI Team — June 2023 - August 2023

LLM, Data scraping

- Worked on LLM fine-tuning for Indic languages
- Participated in Workshop on Asian Translation, 2023 ↗ for Multimodal tasks from English to Indic languages.

PROJECTS

Master's Thesis (Code) | *Python, Machine Learning, Data Visualization, FORTRAN* June 2022-June 2023

- Developed and fine-tuned ML models to predict cloud parameters i.e. Cloudy/Clear discrimination, Cloud top height, etc. from radiometric data of INSAT-3DR satellite
- Collocated 5 TB of INSAT-3DR and CLOUDSAT data to create a dataset correlating cloud parameters with radiometric data. Data accumulates 50 downloads per month. Data ↗
- Developed ML-pipelines to use developed models for generation of our own maps for cloud parameters

Series of Pet Projects | *YouTube API, Python, Regex*

- Developed a web-app to calculate statistics of a YouTube playlist. Demo ↗
- Developed a web-app to help people with pronunciation of various languages, e.g. Polish, Turkish etc. Demo ↗

TECHNICAL SKILLS

Languages: Python, C++, JavaScript/TypeScript, GraphQL, Fortran, MATLAB
Data Science and Machine Learning: ML Algorithms (Regression, Classification, Clustering), Natural Language Processing (**NLP**), Retrieval-Augmented Generation (**RAG**), Large Language Models (**LLM**s), Time-Series Analysis
Databases: MySQL, PostgreSQL, MongoDB
Tools: Apache PySpark, HuggingFace, Postman, JIRA, Linux, Wandb, LaTeX
Data Analysis and visualization: Numpy, Pandas, Matplotlib, Seaborn, Power BI
Cloud and DevOps Tools: AWS (EC2, S3), Docker, GCP, Azure, Git, Github, CI/CD
Frameworks: WebDev - (FastAPI, React), Gen-AI - (LangChain, LangGraph, LlamaIndex, RAGAS)

Publications

OdiaGenAI's Participation at WAT2023 ↗ 2023

- Developed a fine-tuned NLLB-200 LLM for English→(Hindi, Bengali and Malayalam) text translation, which claimed the top spot in both evaluation and challenge sets.
- Developed a combined system of object detection (with ResNet-101) + machine translation (with NLLB-600) for English→(Hindi, Bengali and Malayalam) multimodal translation task.

Llama2-finetuned LLM for Odia language utilizing domain knowledge instruction set ↗ 2023

- Developed a niche dataset for Odia language capturing cultural and linguistic intricacies.
- Fine-tuned Llama2-7B model using LORA technique on the said dataset.

Conferences and Workshops

Odisha AI ML Conference 2023 October 2023

- Delivered a talk at the Odias in AI ML Conference 2023 ↗, representing Team OdiaGenAI. Watch the session here ↗ (timestamped).
- Discussed about Automatic Speech Recognition, LLM fine-tuning and Synthetic Data generation.

Generative AI and LLM workshop December 2023

- Delivered talks at Generative AI and LLM workshop ↗ in two sessions. Watch the sessions here ↗, and here ↗.
- Discussed about LLM Inference, Quantization, Deployment and synthetic data generation.

Three-Day Generative AI Workshop May 2024

- Presented a session on Generative AI demo ↗, and volunteered in the workshop.

Miscellaneous

- Featured by **Meta** in 2023, Advancing AI for Indic Languages with OdiaGenAI ↗
- Featured by Analytics India Magazine in 2023, Top 8 Vernacular Language Models Based on Llama 2 ↗