DHIRENDRA CHOUDHARY

Bengaluru, India •dhirendrachoudhary@yahoo.com • +91-9999391056 • linkedin.com/in/dhirendrachoudhary

EDUCATION

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY BANGALORE (IIITB) — **Bangalore, India**
Ph.D. in Data Science and Artificial Intelligence - part-time — **2024- current**

CHRIST UNIVERSITY — **Bangalore, India**
Master of Science in Computer Science; Major in Data Science — **2020-2022**

GGSIP UNIVERSITY — **New Delhi, India**
Bachelor of Computer Application, Major in Computer Science — **2017-2020**

PROFESSIONAL EXPERIENCE

ZYCUS INFOTECH — **Bangalore, India**
Lead ML Engineer — **09/2023-Present**

- Architected & Productionized e**nd-to-end LLM pipelines** (GPT-4/3, LangChain) for automated invoice data extraction, achieving **97% accuracy**.
- Engineered & Deployed scalable **RAG systems using Knowledge Graphs (Neo4j)** and Vector DBs (Pinecone, FAISS) on AWS, improving semantic **search efficiency by 20%.**
- Executed advanced fine-tuning (**PEFT, LoRA, Quantization)** on LLaMA 3 models, optimizing performance for specific downstream tasks by 15%.
- Established robust **MLOps** workflows with **CI/CD** pipelines for automated model deployment, testing, and monitoring, **reducing deployment cycles by 30%.**
- Integrated multimodal models (**LayoutLM, DocVQA**) into document processing systems, enhancing overall data extraction **performance by 25%.**
- **Spearheaded** the development & evaluation of a multi-agent system POC (**AutoGen**) for procurement **QnA,** reducing query response time by 30% and **improving accuracy by 20%.**

IHX PRIVATE LIMITED — **Bangalore, India**
Data Scientist — **04/2022-09/2023**

- Developed the Discharge Summary Digitizer project, **fine-tuning a clinical BERT model,** Utilized NER Model and **SNOMED/ICD** for medical coding.
- Developed an innovative **multimodal model** (**LayoutLMv2**) for Lab Report Digitization, achieving 92% accuracy and delivering a 25% efficiency gain.
- Optimized **document Analysis** via f**ine-tuning Faster R-CNN/YOLO models** for document layout extraction.

DEFENCE RESEARCH AND DEVELOPMENT ORGANISATION (DRDO) — **New Delhi, India**
Research Intern — **12/2019-06/2020**

- Researched and Implemented a **custom architecture combining CNN and LSTM** for mental workload estimation using EEG datasets, achieving 84% accuracy.
- Designed an insightful dashboard using **Dash, Tableau, and Plotly**, facilitating data-driven decision-making.
- Implemented an **end-to-end pipeline from data collection to prediction,** ensuring seamless delivery of accurate results.

SKILLS & SPECIALIZATIONS

- **Languages:** Python, SQL, C, Bash
- **Core Areas:** GenAI, LLMs, NLP, Computer Vision, Deep Learning, ML, MLOps, Data Science, Prompt Engineering
- **LLMs & GenAI:** Fine-tuning (PEFT, LoRA, QLoRA), RAG, LangChain, LlamaIndex, Vector DBs (Pinecone, FAISS, Qdrant, Milvus), Multi-Agent Systems (AutoGen), LLM Evaluation, OpenAI API (GPT-4/3.5), LLaMA 3, Mistral
- **ML/DL:** PyTorch, TensorFlow, Keras, Scikit-Learn, HuggingFace, Spacy, NLTK, OpenCV, Pandas, NumPy, XGBoost, LightGBM, Transformers (BERT), CNNs (YOLO), RNN/LSTM
- **CV:** Object Detection, OCR (PaddleOCR), Document AI (LayoutLM, DocVQA), Image Classification
- **Cloud & MLOps:** AWS (SageMaker, EKS, ECS, S3), Azure ML, GCP AI, Docker, Kubernetes, CI/CD, MLflow, **Terraform**
- **Databases:** PostgreSQL, MySQL, MongoDB, Neo4j, Spark
- **Tools:** Git, PyCharm, VS Code, Jupyter, Anaconda, Label Studio, Tableau, Power BI, Plotly, Dash

PUBLICATION

- **Sign Language Recognition System:** Sign Language Recognition System (April 22, 2021). Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021.
- **Mental Workload Estimation Using EEG:** "Mental Workload Estimation Using EEG," 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Bangalore, India, 2020.