

ARSHRAJ RANDHAWA

Bengaluru, Karnataka · +91 9457894900 · arshraj.randhawa@gmail.com
linkedin.com/in/arshraj · DOB: 16-06-1994

PROFESSIONAL SUMMARY

Generative AI & NLP Specialist 5+ Years Experience in AI & Machine Learning and 3.5+ years specializing in NLP, machine learning, and LLM fine-tuning. Skilled in developing and deploying Generative AI solutions, leveraging LLMs, Retrieval-Augmented Generation (RAG), and vector databases to optimize business processes. Expertise in model fine-tuning (LoRA, PEFT), cloud-based AI deployments (Docker, Kubeflow), and cross-functional collaboration to drive AI-driven decision-making and automation

WORK EXPERIENCE

Manager – Global Analytics

Concentrix

May 2023 – Present

Bengaluru, Karnataka

- Led the development of NLP models for the travel industry, boosting customer engagement by 20% and reducing support response times by 15%
- Built LLM-powered NLP solutions using LangChain, Retrieval-Augmented Generation (RAG), and vector databases like Quadrant, enabling intelligent question-answering and search
- Applied advanced prompt engineering techniques (CoT, ReAct, zero-shot, few-shot) to improve LLM response accuracy and contextual understanding
- Developed fine-tuned RoBERTa models for text classification, leveraging LoRA and PEFT techniques to optimize performance while reducing computational overhead.
- Fostered collaboration in interdisciplinary teams and drove innovation in NLP projects, applying advanced NLP techniques to solve complex problems and deliver impactful business results
- Created Dockerized NLP applications and deployed them as Kubeflow pipelines, automating model training, inference, and monitoring for production use cases
- Utilized Google Cloud Platform (GCP) and Vertex AI to design, develop, and deploy generative AI projects, leveraging GCP's scalable infrastructure to support robust model deployment and resource management

Deputy Manager – Global Analytics

Concentrix

May 2021 – May 2023

Bengaluru, Karnataka

- Implemented topic modeling using BERTopic with SBERT embeddings, leveraging their contextual representations to extract meaningful themes from text data, facilitating insights and enhancing understanding of key topics within large datasets.
- Led multifaceted projects encompassing text summarization, language translation, PII removal employing NER, leveraging advanced NLP techniques to analyze unstructured text data and derive actionable business insights, thereby driving informed decision-making and strategic initiatives within the organization
- Designed and implemented an anomaly detection system for time series data utilizing Facebook Prophet, integrated with SMTP library in Python to automatically alert relevant authorities via email upon anomaly detection, ensuring timely response and proactive management of critical events

Analyst – Intern

Covance

Dec 2017 – May 2019

Bengaluru, Karnataka

- Collaborated with cross-functional teams, including statisticians and data managers, to develop and implement statistical analysis plans (SAPs).
- Conducted data analysis for multiple clinical trials, ensuring adherence to protocol specifications and regulatory requirements.
- Developed and validated SAS programs to generate tables, listings, and figures for clinical study reports.

PROJECTS

Fine-Tuning Large Language Models (LLMs) with PEFT for the Travel Industry | Python, HuggingFace

- Developed and fine-tuned multiple LLMs (RoBERTa, GPT) using LoRA and PEFT techniques, optimizing sentiment analysis and text classification for large-scale, multilingual customer feedback data (reviews, chat logs, emails, phone transcripts).
- Designed a scalable, end-to-end pipeline integrating data preprocessing, fine-tuning, model evaluation, and inference, enabling efficient training and deployment of LLM-powered sentiment analysis models.
- Reduced computational overhead while maintaining high accuracy, leveraging parameter-efficient fine-tuning (LoRA, PEFT) to enhance model adaptability across diverse customer interaction channels
- Automated sentiment analysis and customer feedback insights to drive data-driven decision-making, improving customer service quality and operational efficiency in the travel industry

Retrieval-Augmented Question Answering System | Streamlit, LangChain, Kubeflow

- Developed a retrieval-augmented generation (RAG) system using LangChain, leveraging OpenAI embeddings and models for intelligent question answering
- Integrated Quadrant for vector search, enabling efficient retrieval of relevant documents for context-aware responses
- Containerized the application with Docker and deployed it as a Kubeflow pipeline, ensuring scalability, automation, and seamless model deployment

Topic Modeling Web App with Streamlit, Docker, and Kubeflow | BERTopic, LDA

- Developed an interactive Streamlit-based web app for topic modeling, implementing LDA (Scikit-learn) and BERTopic for automated topic extraction from text data
- Containerized the application using Docker and deployed it as a Kubeflow pipeline, enabling scalable and automated topic modeling workflows
- Designed a user-friendly UI for customizable hyperparameters, allowing users to fine-tune LDA and BERTopic models and visualize topic distributions dynamically

Hate Speech Detection and Moderation with GenAI | OpenAI

- Developed an AI-powered hate speech detection and moderation system using OpenAI models, including the Omni moderation model, to analyze and filter user-generated content
- Deployed the service using FastAPI, enabling real-time content moderation through scalable API endpoints

ACHIEVEMENTS

Data & Analytics CX Solution Star - Award

April 2023

Insights & Consulting Star - Award

Sept 2022

SAS Certified Base Programmer for SAS 9 - Certification Exam Score - 97%

Sept 2017

EDUCATION

Narsee Monjee Institute of Management Studies (NMIMS)

June 2019 – May 2021

PGDM Business Analytics : CGPA 3.22

Mumbai, Maharashtra

Dr. A. P. J. Abdul Kalam Technical University

July 2012 – June 2016

A Bachelor of Technology in Electronics and Communication : Percentage 64.44

Noida, UP

SKILLS

Language: Python, SQL, SAS, R

ML and DL: TensorFlow, PyTorch, Scikit-learn, Prophet

NLP and LLM: Generative AI, LLM fine-tuning, parameter-efficient methods (PEFT, LoRA), SBERT, Spacy, NLTK, Gensim, HuggingFace, OpenAI, Regex, LangChain, Agentic AI , Quadrant, FAISS, Streamlit, BERTopic

Cloud Platforms: MLOps , GCP, Docker, Kubernetes, Kubeflow, Git, GitHub, CI/CD pipelines