

DIVYANSHU PUROHIT

7037730830 | divyanshu.25bkiet@gmail.com | LinkedIn | GitHub | Portfolio
Dehradun, Uttarakhand

PROFESSIONAL SUMMARY

AI Engineer with hands-on experience in LLM applications, multimodal AI, microservices, and Docker-based deployments. Experienced in building production-ready AI pipelines, retrieval systems, model training, and scalable backend development. Strong foundation in Python, ML/DL, embeddings, and end-to-end system design.

EDUCATION

BK Birla Institute of Engineering & Technology Pilani, Rajasthan
Bachelor of Technology in Computer Science Engineering – CGPA: 8.2 Aug 2021 – May 2025

EXPERIENCE

AI Intern Oct 2025 – Present
GoWide Technologies Pvt. Ltd

- Worked on multiple LLM-driven client projects, designing AI pipelines, embeddings workflows, and structured reasoning systems
- Built NRRC AI involving document parsing, multi-stage inference, retrieval logic, and tokenizer-level optimization
- Developed microservices using FastAPI & Docker for indexing, search, inference, and system orchestration
- Built an internal prototype AI chatbot with intent detection, memory, embeddings, and RAG architecture
- Managed production deployments using Docker, CI-ready service structures, and container-based rollouts

AI Intern Apr 2025 – Jul 2025
Anantagani Industries

- Led large-scale data labeling for plastic classification improving data consistency and training performance
- Developed and trained CNN-based classification model using TensorFlow/Keras
- Integrated camera-based real-time detection pipeline using OpenCV
- Delivered a scalable system that improved operational plastic-waste identification workflows

Data Analyst Intern Apr 2024 – Aug 2024
Tranity

- Performed real-time analysis for client systems enabling key business insights
- Cleaned and standardized large datasets ensuring high-quality analytics
- Developed interactive dashboards in Power BI for KPI tracking and trend discovery

PROJECTS

NRRC AI | *GoWide Technologies*

- Built multi-stage reasoning engine handling document parsing, embeddings, structured outputs, and retrieval
- Designed microservices for inference, indexing, and workflow orchestration

Prototype AI Chatbot | *GoWide Technologies*

- Developed enterprise-ready chatbot with intent detection, RAG, embeddings search, and memory
- Dockerized and deployed using FastAPI microservice architecture

Face Emotion Detection System

- Built real-time emotion detection using CNN + OpenCV with seven emotion classes
- Implemented dataset labeling & ML workflows for training and evaluation

Offline Article Finder AI

- Built FAISS + BM25 hybrid research retrieval system using Sentence Transformer embeddings
- Developed FastAPI backend for local inference

TECHNICAL SKILLS

Languages: Python, Basic Java, SQL

Machine Learning: LLMs, LMMs, RAG, Embeddings, Tokenization, CNNs, TensorFlow/Keras, Scikit-learn

Tools & Frameworks: FastAPI, Docker, OpenCV, FAISS, BM25, Power BI, Git

Other Skills: Data preprocessing, microservices, pipeline development, deployment workflows