UDAY GUPTA

AI-ML Engineer

Gurugram, Haryana

udaygupta.vercel.app 3

LinkedIn – Uday Gupta 🛅



GitHub – udayG01



91-7042820522



Education

Bachelor of Technology, Artificial Intelligence and **Machine Learning**

2021 - 2025

CGPA: **9.13**

University School of Automation and Robotics, Suraimal Vihar, New Delhi (GGSIPU East Campus)

Intermediate 2021

NK Bagrodia Public School, Sector-4, Dwarka, New Delhi

Score: 96.4%

Experience

ML Intern

Jun 2024 - Aug 2024, Online

ARCedtech Solutions Pvt. Ltd

- Lead the team and bult an RAG Chatbot "Treebo" (link) educating users on 17 SDGs
- 15 SDG PDFs as RAG input, response time: 3-4s with Gemini 1.5 Flash
- LangChain, Gemini Embeddings, Kubernetes Tech Used: -
 - Vercel deployment

IVIL Intern

Elint Data Pvt. Ltd.

Aug 2023 - Oct 2023, Online

- Domains: NLP: RAG bots (2), BERT; Vision: Object Detection
- Collaborated on Quantized LLM (4,8 bit), response time accelerated: 2 mins -> 20s for RAG bot (PDFs, URLs), BERT embeddings
- Traffic Detection via YOLOv5 (mAP>90%)
- Hugging Face, Tensorflow, FastAPI, LLMs Llama2, Falcon Tech Stack: -
 - **Docker, Azure Machine Learning, YAML** (1 streamlit app)

Web Development Intern

The Letude

Jul 2022 - Aug 2022, Online

- Designed and developed 3 front-ends (login page, travel directory, memes app), assissting in a team of 4.
- Tech Stack: -HTML, CSS, JS, ReactJS, NodeJS, Git, Github
 - RestAPI, Vercel deployment, Github Pages deployement

Projects

For viewing more of my projects: links(Website, Codes)

- BookPal-Llama2 (link) a guide to query about books PDFs or URLs using RAG in LLMs
 - Quantized (4,8 bit) LLMs, HuggingFace Embeddings, 5-10 PDFs as RAG input, response time- 10-15s
 - Tech: Langchain, QLora, Azure deployment
- **DreamSound Studio (link)** Generate sounds with NST and Google's DeepDream (Inception V3)
 - Developed CNNs (5 style, 1 content layer), 300 training iters
 - o Tech: Tensorflow (CNNs), Librosa, Gradio for UI
- Brain Tumor Detection (link) Custom CNN, Transfer learning on VGG19, ResNet50 for tumor detection
 - 4 classes, 3264 images, data augmentation, dropout(0.5)
 - Tech: Tensorflow, Gradio UI

Skills

Foundational Programming

- Python, Java, JS, DSA, OOPs
- Git, GitHub

Full-Stack Development

- HTML, CSS
- **MERN**

Machine Learning

- Tensorflow, OpenCV, NLTK
- HuggingFace, ChromaDB, Pinecone
- Google Cloud Platform

Soft-Skills

- Leadership, Teamwork (See "Accomplishments" section)
- Digital literacy, Problemsolving (See "Projects")
- Communication skills (English, Hindi)

Accomplishments

Mercer Mettl Hackathon

- First runner-up
- Rs.10,000 (Amazon Credits)

Smart India Hackathon

- Team Lead
- Top-5 teams: internal hack

Google Developer Student Clubs

- ML Co-Lead
- Lead 200 participants for GCP Cloud Campaign
- Assisted 80 participants for tier-1 reward