AASHISH GUPTA

+351-923-126-997 ◇ ashuguptahere@gmail.com ◇ LinkedIn ◇ Leetcode ◇ GfG ◇ Codolio ◇ Kaggle ◇ GitHub ◇ Portfolio

EXPERIENCE

Data Scientist & AI Engineer — Dec 2023 - Present
Quantum Leap — *Setúbal, Portugal*

- Leading and delivering computer vision solutions in Human Activity Recognition and Behavior Analysis to enhance retail loss prevention and identify suspicious customer behavior
- Developed and deployed an end-to-end system using Ultralytics and custom computer vision pipelines
- Orchestrated an in-house dataset with semi-automated labeling for real-world behavior detection
- Deployed optimized models on NVIDIA Jetson AGX using DeepStream and TensorRT, **reducing inference time by 40%** and enabling low-latency, high-throughput edge inference
- Achieved **10-20x (1000-2000%)** inference speedup by migrating Python-based code to C++ DeepStream pipelines optimized for NVIDIA hardware
- Automated large-scale deployment with Ansible and GitHub Actions with scalable architecture; currently running on 10+ devices with architecture scalable to 1000+ devices
- Contributed to AIOps/MLOps best practices by integrating CI/CD, remote monitoring, and model versioning for reliable, scalable deployment and lifecycle management
- Led the development of a fully localized Qwen-2.5/3 LLM model tailored for a law firm client, mentoring 2 junior engineers and owning delivery of production-ready features.
- Built a sophisticated chat interface allowing the client to interact with their data, facilitating quick identification of relevant documents and specific information within vast repositories of PDFs and text files.
- Spearheaded & successfully led initiatives in RAG & LoRA-based fine-tuning, enhancing domain-specific model efficiency
- Engineered backend solutions to streamline and analyze incoming streams, enabling real-time irregular behaviors detection
- Led the development of a fully localized Llama-3 70B LLM model tailored for a law firm client
- Built a sophisticated chat interface allowing the client to interact with their data, facilitating quick identification of relevant documents and specific information within vast repositories of PDFs and text files

Software Engineer 1 — Dec 2022 - Dec 2023
MAQ Software — *Noida, India*

- Spearheaded initiatives in Retrieval-Augmented Generation (RAG)
- Applied state-of-the-art LLMs like GPT-3.5, GPT-4, and Llama 2 for RAG applications
- Optimized a legacy system using Python, NumPy & Pandas, cutting runtime from 7 minutes to 10 seconds ($\approx 42\times$ faster)
- Deployed production-grade solutions on Azure: Function App, Web App, Cognitive Search (Index/Indexer), Blob Storage, SQL Databases, Speech Service
- Implemented a high-impact recommendation system for ∼1M users

System Engineer — Aug 2021 - Dec 2022
Tata Consultancy Services (TCS) — *Noida, India*

- Led and executed numerous Data Analysis and Data Science Projects
- Proficient in backend development in Python for robust, scalable applications

INTERNSHIPS

Data Science Intern — Jul 2021 - Aug 2022
VerSe Innovation — *Bengaluru, India*

- Engaged in challenging Computer Vision endeavors, including tasks encompassing Image and Video Classification

AI Trainer — Jan 2020 - Feb 2020
CETPA InfoTech Pvt. Ltd. — *Noida, India*

- Educated and taught college students in their 3rd and 4th years on a wide array of AI topics

Machine Learning & Deep Learning Intern — Jun 2019 - Jul 2019
CETPA InfoTech Pvt. Ltd. — *Noida, India*

- Internship Involvement in Twitter Sentiment Analysis Project

SKILLS

Technical — AI, ML, DL, GenAI, Data Science, Data Analysis, Computer Vision, Anomaly Detection, Image Encryption, Python, Git/GitHub, MySQL, T-SQL, Video Analytics, DeepStream, GStreamer, MQTT, LLM, NLP, RAG, LoRA/QLoRA fine-tuning, Vector Databases, EDA, C/C++, DSA, Linux, PowerBI, AIOps, MLOps, Docker, Prometheus, Grafana, Tailscale, Redis, AWS, Azure

Libraries — PyTorch, TensorFlow, Keras, OpenCV, scikit-learn, NumPy, Pandas, Seaborn, Plotly, NLTK, OpenAI, Streamlit, Gradio, Flask, FastAPI, LangChain, LangGraph, LlamaIndex, uv, ruff

Exposure — C#, Java, HTML, CSS, JavaScript (ECMAScript), Mojo, Rust, Zig

Languages — English, Hindi, Portuguese (beginner)

PROJECTS

- **Real-Time Sign Language Detection**: Achieved a validation accuracy of 99.348% and a minimal loss of 0.03621 on ResNet102. Dataset: ASL Alphabet
- **Human Activity Recognition**: Conducted Video Classification on two distinct massive datasets, UCF11 and UCF101. Demonstrated robust validation accuracy ranging from 90% to 95% on ResNet102 | Datasets: UCF11, UCF101
- **AI Cover Letter Generator**: Simple and intuitive web interface made using Gradio with real-time Cover Letter Generation. Usage of Llama 3.2 model from Ollama and usage of uv package manager for package dependency.

ACHIEVEMENTS

- Gate 2021 cleared with Marks: 25.21 & Gate Score: 339
- Practiced over 190+ DSA questions on various platforms

EDUCATION

Bachelor of Technology, CSE *from* Shaheed Bhagat Singh State Technical Campus (SBSSTC)   2016 - 2020

EXTRA-CURRICULAR ACTIVITIES

- E-Cell & NSS Member at SBSSTC, Ferozepur, Punjab
- System Incharge/Technician for the WebCam Feeds, Election Commission of India
- Examinar at Mayank Foundation for Painting Competition

CERTIFICATIONS

- Coding Blocks - Algo++
- Coursera - Deep Learning
- CETPA - Machine Learning
- CETPA - Deep Learning
- Microsoft - Fundamental of ML
- ISOEH - Machine Learning