JAI P. DALVI

jaidalvi17@gmail.com♦ (412) 287-0763 ♦https://www.linkedin.com/in/jaipdalvi/

EDUCATION

Carnegie Mellon University — Pittsburgh, PA
Master of Science in Electrical and Computer Engineering | GPA: 3.75/4.0 — Dec 2025
Courseswork: Machine Learning in Production, Large Language Models:Methods and Applications, Introduction to Machine Learning for Engineers, Advanced Analytics and Machine Learning for Semiconductor Industry, Deep Generative Modelling, Advanced Natural Language Processing, Introduction to Deep Learning

Sardar Patel Institute of Technology — Mumbai, India
Bachelor of Science in Electronics and Telecommunication | GPA: 9.36/10.0 — June 2022

SKILLS

Programming Languages: Python, C, SQL
Technical Knowledge/Skills: PyTorch, scikit-learn, pandas, SQL, Git, TensorFlow, Artificial Intelligence, Machine Learning, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), LangChain, Streamlit, FAISS, CI/CD, Docker, Kubernetes
Publications: DDoS Attack Detection using Artificial Neural Network and Multi-Class Classification of Different Region Pop Songs

RESEARCH EXPERIENCE

Carnegie Mellon University — Pittsburgh, PA
Student Researcher / Graduate Research Assistant

LLM-RAG System for Engineering Drawing — Jan 2025 - Dec 2025
- Created a LLM-based RAG system for Misumi, to assist in retrieving and refining information for engineering drawings, automating corrections and improving quotation finalization

ML/CV-Based Human-Activity Tracking And Recognition — May 2024 - Dec 2024
- Built a human pose estimation system using YOLOv7 for object and keypoint detection to monitor construction-site safety

WORK EXPERIENCE

Value Buddy, Inc. — Louisville, KY
AI Engineer — May 2025 - Aug 2025
- Designed and deployed a Retrieval-Augmented Generation (RAG) agent on Google Cloud Platform for financial valuation reports, while applying strategic thinking to align technical solutions with business and client objectives
- Built scalable retrieval pipelines using LangChain, FAISS, and Cohere reranking, supporting 100+ valuation documents with low-latency responses, with reproducible evaluation metrics, monitoring, and debugging

Ernst & Young LLP — Mumbai, India
Senior Analyst — Sept 2022 - Dec 2023
- Built analytical dashboards and machine learning models on financial and operational data, enabling 30%+ reporting efficiency gains and data-driven decision making
- Analyzed clients' requirements and difficulties in the operational workflow and established as a proficient advisor

University at Buffalo — Buffalo, NY
Research Intern — June 2022 - Aug 2022
- Trained and researched GAN-based multimodal models (StyleGAN, CycleGAN, Progressive GAN) on folklore based datasets

Hindustan Petroleum Corporation Limited — Mumbai, India
Machine Learning Intern — Jan 2022 - May 2022
- Built a full-stack AI chatbot to help users get the value of Notional Coastal Freight (NCF) through interactive conversations

Bitgenie Technologies Private Limited — Mumbai, India
Software Developer — Nov 2021 - Feb 2022
- Generated NFT artwork using DALL-E and GPT-3 from single text prompts for digital marketplaces

ACADEMIC PROJECTS

Movie Based Recommendation System — Aug 2025 - Dec 2025
- Deployed a production-ready Two-Tower recommendation system for a Netflix-like platform (1M+ users, 27K movies), with a Flask API (<600 ms), Kafka-based telemetry, DVC-versioned pipelines, and CI/CD (Jenkins, pytest, 90%+ coverage)
- Containerized and operated the system using Docker & Kubernetes, enabling A/B (canary) testing, monitoring, and experimentation to analyze performance, fairness, and business impact

Temporal Reasoning in MiniRAG — May 2025
- Improved MiniRAG with temporal reasoning via graph retrieval and re-ranking, boosting accuracy and F1

DDoS Attack Detection using Artificial Neural Network — May 2022
- Detected types of DDoS attacks and parameters causing DDoS threat using Deep Learning techniques

Multi-Class Classification of Different Region Pop Songs using Spotify Database — Nov 2021
- Analyzed Indian tribal song to identify factors rendering a song popular and targeting specific language audience