Divya Panchal

+1 (908)-930-9518 divya@divyapanchal.com LinkedIn Github Portfolio San Francisco, CA

Experience

AI/ML Research Assistant **June 2025 - Present**

San Francisco State University *San Francisco, CA*

- Improved course similarity accuracy by 5.3% via feature engineering and a normalized similarity metric, boosting downstream model performance
- Fine-tuned Sentence Transformer models (BGE, NV-Embed, SFR) using FSDP and triplet loss to optimize semantic embeddings for course equivalency analysis
- Deployed containerized ML microservices on AWS Lambda, improving system resilience and scalability through independent component scaling
- Developed and tested C-ID–based equivalency strategies through data-driven experiments, enhancing one-to-many rule logic

Machine Learning Engineer Intern **January 2024 – May 2024**

Bluepixel Technologies *Ahmedabad, India*

- Fine-tuned LLaMA-3 on medical data to build an AI health assistant generating personalized insights from patient records
- Designed a LangChain agent to structure model output with 99%+ data integrity for time-series ingestion
- Built an automated LLM-to-Grafana pipeline via JSON extraction and InfluxDB, visualizing real-time health trends
- Built a LangChain-based conversational system with dynamic context retention for multi-turn health queries and enforced safety guardrails

Technical Skills

- **Programming & Database Management**: Python, JavaScript, PostgreSQL, MongoDB, Vectorstores(ChromaDB, Qdrant)
- **AI/ML Frameworks & Libraries**: PyTorch, Tensorflow, Transformers, scikit-learn, Computer Vision, FastAPI
- **GenAI**: LLMs, RAGs, Langchain, Prompt Engineering, Mulit-modal AI, Model Fine-tuning(PEFT/LoRA)
- **Natural Language Processing**: Embeddings, Semantic Search, BERT, Seq2Seq Architecture, Attention Mechanisms
- **Cloud/DevOps**: Docker, Kubernetes, GCP(VertexAI Pipelines, GKE), AWS(Bedrock, Sagemaker), CI/CD(Github Actions)

Projects

VelocityLM: Custom Foundational LM | *PyTorch, Distributed Training, RMSNorm, RoPE, SwiGLU* | Live Demo | Github

- Designed and trained a decoder-only transformer from scratch, integrating RMSNorm for stability, RoPE for long-sequence handling, and SwiGLU for expressivity
- Achieved stable convergence on Falcon RefinedWeb dataset by optimizing memory and compute with gradient checkpointing and mixed precision FP16 training
- Delivered smooth loss reduction and high-quality text generation by enhancing performance with cosine annealing LR scheduler with AdamW optimizer

VisualiX: Gemini-Powered Video Editor | *Google Gemini AI, OpenCV, FastAPI, Agentic Workflows* | Live Demo | Github

- Built an AI-driven video editing platform that converted natural language prompts into OpenCV editing operations using 25+ custom video tools
- Orchestrated multi-step workflows with agentic AI + real-time progress tracking, ensuring smooth coordination between planning and execution phases
- Developed a FastAPI backend with async APIs, enabling reliable, fault-tolerant video editing operations for complex multi-stage tasks

VizWiz: Visual QA | *PyTorch, Transformers, ViT, BERT, NLP, Computer Vision, FastAPI, Docker* | Live Demo | Github

- Created a multimodal QA system using Vision Transformer (ViT) + BERT, combining image and text features for accurate contextual answers
- Engineered a robust preprocessing pipeline with patch-based encoding + attention-weighted fusion for deeper visual reasoning
- Achieved strong accuracy and provided answerability confidence scores by improving training efficiency with gradient accumulation with mixed-precision training

Education

San Francisco State University **San Francisco, CA**

Master of Science in Data Science and Artificial Intelligence (GPA - 4.0) *May 2026*

- Courses: Deep Learning, Statistical Modelling and Data Mining, PAMI, NLP, Software Engineering

Gujarat Technological University **Ahmedabad, India**

Bachelor of Technology in Artificial Intelligence and Machine Learning (GPA - 3.95) *May 2024*

- Courses: Fundamentals of AI, Image Processing, Machine Learning, Fuzzy Systems, Big Data, NLP