Last Updated on 3rd October 2025

Nikhila Chitta

608.867.3033 | nikhilakanchinadam@gmail.com | Sunnyvale, CA

EDUCATION

UNC CHARLOTTE
Masters in Computer Science
Grad. Aug 2025 | Charlotte, NC

JNTU KAKINADA
BS in Electrical Engineering
Grad. May 2017 | Kakinada, India

LINKS

Github: **nkanch23**
LinkedIn: **Nikhila C**

COURSEWORK

GRADUATE
Algorithms & Data Structures
Artificial Intelligence
Machine Learning
Natural Language Processing
Database Systems
Software Systems Design & Impl.
Data Mining
Mobile Application Development
Computer Networks
Visual Analytics

SKILLS

Languages
Java • SQL • Python • Javascript • C
C++ • PySpark • Linux/Shell/Bash

Cloud
AWS (EC2/ECS • S3 • Redshift
Glue • Quicksight • Sagemaker
Dynamo • Athena)

GCP (BigQuery • GCS • VertexAI)

Generative AI
GPT • Claude • Gemini • LangChain
LangGraph • Pinecone • FAISS
OpenAI Agents • OpenRouter
MCP • PEFT/LORA • KV Routing

AI/ML
NLP • NER • Linear models • RAG
ElasticSearch • Embeddings • BERT
Information Retrieval (IR) • Word2Vec
Clustering • SHAP • Fairness & Bias

Frameworks
FastAPI • PyTorch • Hadoop • Ray
Airflow • Kafka • Angular • K8s
Docker • Nvidia Dynamo • Redis

EXPERIENCE

ACCENTURE | Senior Software Engineer
Mar 2018 - Jun 2022 | Hyderabad, India

- Engineered and deployed end-to-end data pipelines using PySpark, Airflow, and Kafka, reducing ETL runtime by 40% and improving data readiness for business reporting.
- Developed microservices with FastAPI, Docker, and Kubernetes, enabling modular deployments and improving scalability across client environments.
- Automated ML training and deployment pipelines with AWS SageMaker and Vertex AI, enabling continuous retraining and monitoring of models.
- Built NLP pipelines for entity recognition, embeddings, and semantic search using BERT, ElasticSearch, and Word2Vec, applied to finance and healthcare domains.
- Mentored junior engineers, participated in system design reviews, bridging the gap between concept ideation and production deployment.

SELECTED PROJECTS

Multi-Agent Conversational AI System | Python, OpenAI, LLMs, Orchestration, System Design

Architected a production-grade multi-agent IVR system with dynamic policy routing and hierarchical intent classification. Implemented orchestrator-fulfillment agent pattern enabling context-aware handoffs and scalable policy management. Designed intent disambiguation logic with L1/L2/L3 classification hierarchy, reducing ambiguous queries by 40% through strategic clarification flows. System demonstrates advanced prompt engineering, stateful conversation management, and modular architecture supporting extensible policy definitions for enterprise customer service automation. Link

Hybrid RAG Document QA System | Python, FAISS, LLMs, NLP, Vector Databases, Gemini

Built production-grade retrieval-augmented generation (RAG) system combining dense vector search (FAISS) with sparse TF-IDF retrieval for document question answering. Implemented semantic reranking pipeline that improved retrieval precision by re-scoring candidates using cross-encoder similarity. Designed answer verification framework validating LLM outputs against source documents for factual accuracy and claim verification. System features persistent vector index storage, intelligent document chunking with sentence boundary detection, and batch embedding generation. Demonstrates end-to-end RAG pipeline from PDF processing through verified response generation with source attribution. Link

AvatarCLIP: Zero-Shot Text-Driven 3D Avatar Generation | Python, PyTorch, GPU, CLIP, VAE

Developed a zero-shot pipeline for text-to-3D avatar generation and animation by integrating CLIP-based vision–language supervision with VAE latent modeling, NeuS volume rendering, and SMPL body representations. Implemented motion synthesis with transformers and clustering techniques to achieve realistic avatar geometry and animations from natural language prompts. Link

PATENT

- Generative AI based Systems and Methods for Scalable Document search via Virtual Assistants (Chatbot)