

Aniruddha Sangule

 SanguleAKB |  sanguleAKB |  aniruddhasangule2001@gmail.com |  +91 8459358314 |  Visit

EDUCATION

Centre for Development of Advanced Computing (CDAC)

PG-Diploma in Big Data and Machine Learning; Grade: A+

Pune, Maharashtra

2024 - 2025

Shivaji University

Master of Computer Applications; CGPA: 8.74

Kolhapur, Maharashtra

June 2022 - August 2024

TECHNICAL SKILLS

Programming Languages

Python, SQL, R, Java

ML Frameworks & Libraries

TensorFlow, PyTorch, Scikit-learn, Keras, XGBoost, LightGBM

NLP & LLM Tools

LangChain, LangGraph, OpenAI API, Hugging Face Transformers, N8N, Retrieval-Augmented Generation (RAG)

Backend Development

FastAPI, Flask, Django REST

Databases & Storage

MySQL, MongoDB, Pinecone (Vector DB), ChromaDB

Cloud Platforms

AWS (SageMaker, EC2, S3, Lambda), Microsoft Azure (Azure ML, AI Foundry)

WORK EXPERIENCE

Data Scientist — HiCounselor

April 2025 - Present

- An educational technology company focused on enhancing learning experiences through AI and data-driven solutions.
- Worked on building and deploying AI agents to automate business workflows, leveraging LLMs and prompting techniques (zero-shot, few-shot, CoT, self-consistency, prompt chaining) to create modular, query-enabled agents tailored to dynamic business logic.
- Integrated LangGraph with in-house APIs and data sources to build stateful, context-aware pipelines for real-time inference and action chaining.
- Developed RAG applications using vector databases (Pinecone, ChromaDB) with reranking for improved results, along with guardrails to ensure data security.
- Created automated n8n workflows for lead generation, resume/hiring process automation, and auto-generated advertisements.

Python Developer Intern – Vishvavidya

February 2025 - April 2025

- Built and deployed scalable ML-powered APIs using FastAPI, SQLAlchemy, Redis, Docker, and advanced monitoring tools.
- Gained hands-on experience across the full development lifecycle including databases, ML model serving, authentication, testing, performance tuning, and production deployment.

KEY PROJECTS

NanoGPT (Decoder-only Transformer) — [\[Link\]](#)

2025

- Implemented a GPT-style decoder-only Transformer architecture from scratch using PyTorch, closely following the original GPT design principles
- Built custom components including token embeddings, multi-head self-attention, causal masking, feed-forward layers, and layer normalization
- Trained the model on large-scale text datasets to perform next-token prediction with stable convergence and efficient memory usage
- Implemented training optimizations such as gradient clipping, learning rate scheduling, and checkpointing for long-running experiments
- Gained hands-on experience with Transformer internals, attention mechanics, and autoregressive language modeling

Recipe Recommendation System — [\[Link\]](#)

2025

- Architected and deployed an intelligent Retrieval-Augmented Generation system for accurate enterprise information retrieval with context-aware responses
- Integrated MySQL for structured data management and Pinecone vector database for semantic search capabilities
- Implemented PII (Personally Identifiable Information) protection mechanisms to ensure data privacy and compliance
- Built agentic workflows using LangGraph for autonomous decision-making and multi-step reasoning tasks
- Achieved high accuracy in information retrieval while maintaining sub-second response times for user queries

CERTIFICATIONS

AWS Certified Machine Learning Engineer - Associate

Amazon Web Services

Validated expertise in designing, implementing, deploying, and maintaining machine learning solutions on AWS, including model training, optimization, and scalability using services such as SageMaker, Lambda, and EC2.

FastAPI for Machine Learning

CompusX

Completed CampusX's FastAPI for Machine Learning course focused on building, testing, and deploying ML-powered APIs with FastAPI, SQLAlchemy, Docker, and Redis. Gained hands-on experience integrating ML models into production-grade API systems with authentication, caching, monitoring, and performance optimization.