PANJUGULA ANVITA

(513) 510-8295 | panjugaa@mail.uc.edu | linkedin.com/in/panjugulaanvita | github.com/anvitaxreddy | anvitaxreddy.github.io

SUMMARY

AI/ML Engineer and Software Developer specializing in production GenAI systems — multi-agent LLM orchestration, RAG pipelines, agentic workflows, and cloud-native backends. Shipped StatVisor at P&G using **LangGraph**, **Azure OpenAI GPT-4**, **FAISS** hybrid retrieval, **MCP** observability, and async **FastAPI** with full **LLM governance**. Experienced in full-stack development, statistical modeling, and translating complex AI outputs for senior stakeholders.

SKILLS

Languages: Python, MATLAB, **R**, Java, C, C++, **TypeScript**, GoLang/Go, **R**ust
MLOps: Deep Learning, **LLMs**, Neural Networks, Statistical Modeling, Time Series, **Computer Vision**, **NLP**, Feature Engineering, Model Training, **MLflow**, Weights & Biases, LangSmith, **Hugging Face**, Stable Diffusion, ControlNet, **CUDA**, Grafana, Chainlit
GenAI: Azure OpenAI (GPT-4), **LangChain**, **LangGraph**, **LlamaIndex**, LoRA, Prompt Engineering, **Fine-Tuning**, **MCP**, **RAG**, **Agentic AI**, Transformers, Semantic Kernel, GraphQL
ML Frameworks: TensorFlow, **PyTorch**, Scikit-learn, SciPy, NumPy, **XGBoost**, Regression, **CNN**, **Reinforcement Learning**
Data & Vectors: SQL, Postgre**SQL**, ETL/ELT, Data Pipelines, Large-Scale Data Processing, Statistical Analysis, Signal Processing, **BigQuery**, **FAISS**, Weaviate, Pandas, Pinecone, **Snowflake**, MongoDB, Redshift
Backend: FastAPI, Flask, Asyncio, REST APIs, Pydantic, JSON-RPC, Model Deployment, **Kafka**, Kernel Development
Frontend: React, Vite, Tailwind CSS, **Streamlit**, JavaScript, .NET, HTML, CSS, PDF.js
Cloud / DevOps: Azure, **GCP**, **AWS** (SageMaker, EC2, EKS), **Azure** AD, JWT, Vercel, Render, **CI/CD**, DevOps, Linux, Automation, SSL/TLS
Developer Tools: Git, **Kubernetes**, **Docker**, **GitHub Actions**, VS Code, Cursor, GitHub Copilot, Postman, Jupyter, Conda, Redis, Airflow, Terraform

EXPERIENCE

Procter & Gamble — Digital Accelerator | *Cincinnati, OH* *Aug 2024 – Present*
AI Scientist — RAG & LLM Systems (R&D)

- Architected and shipped **StatVisor**, a production multi-agent AI system using **LangGraph**, LangChain, **LlamaIndex**, and **Azure OpenAI GPT-4** — engineered **function/tool calling**, **context grounding**, **BM25** sparse + dense hybrid retrieval, and OCR-based document ingestion for context-aware extraction across large enterprise datasets with 95%+ retrieval accuracy.
- Built end-to-end **LLM governance** and **observability** infrastructure using **FastAPI**, Azure Monitor, and **MCP** — enforced audit logging, cost controls, schema validation, and responsible AI compliance; reduced API downtime by 40% via real-time **Grafana** and **Chainlit** monitoring dashboards.
- Orchestrated multi-step **agentic workflows** using LangGraph, LlamaIndex, and **Semantic Kernel** with guardrails, multi-hop reasoning, and **tool-calling** agents; built **knowledge graph** retrieval pipelines using **Neo4j** and Stardog; aligned data governance with Collibra cataloging standards.
- Engineered **Snowflake Cortex** LLM, vector search, and **Cortex Analyst** pipelines for AI-native retrieval; built **async Python FastAPI** backends across BigQuery and Redshift; developed statistical REST APIs using **R Plumber**, ggplot2, dplyr, and tidyverse — presented solutions directly to senior P&G leadership.

Analytics Quad4 | *Bengaluru, India* *May 2023 – Jul 2024*
Software Engineer — AI Automation & Supply Chain

- Built and deployed **GenAI** agents and ML models in Python/R on **GCP Vertex AI** — applied **hallucination mitigation** strategies (output validation, grounding, confidence scoring) with feature engineering and **model versioning** across logistics, inventory, and supply chain planning workflows.
- Designed high-throughput Python pipelines with concurrency and load-balancing, cutting processing overhead by 35% across millions of records; built multi-threaded **Selenium** scraper processing 11K+ sources in <3 seconds at 95%+ accuracy for **warehouse resource utilization**, lane allocation, and **AI-driven scheduling** — adopted by 15+ engineers.
- Orchestrated end-to-end ML workflows using Vertex AI, **Airflow**, **MLflow**, **Docker**, **Kubernetes**, Terraform, **Spark**, and **Kafka**; built **Power BI** dashboards with DAX and semantic layer modeling for supply chain KPI tracking.

Wissen Technology | *Hyderabad, India* *Jun 2022 – Aug 2022*
Software Engineer Intern — MLOps

- Engineered end-to-end data workflows on **Azure** Cloud using **ADF** and **Airflow** — leveraged **Azure** compute and storage services with **Docker**-containerized services for scalable, production-grade data engineering and **MLOps** workloads; collaborated with cross-functional teams to implement **CI/CD** practices via **Azure** DevOps and Git, monitoring workflow **observability** and optimizing pipeline performance across the full data lifecycle.
- Developed 3+ **RESTful APIs** and 10+ frontend components using **Python**, **PostgreSQL**, HTML, CSS, and JavaScript — streamlined data pipeline workflows serving 500+ daily transactions with 99%+ uptime.

PROJECTS

Mockview https://mockview-ruby.vercel.app | *FastAPI, React/Vite, Gemini AI, Supabase, PostgreSQL, Vercel, Render*

- Full-stack AI mock interview platform — **Gemini AI** adaptive question generation across 8+ categories, automated scoring on 5 dimensions; **JWT** auth with **FastAPI**/Supabase backend; trimmed API response times by 210ms and DB query times by 75ms; deployed via **CI/CD** on Render and Vercel with <2s load times.

NeuroCache — LLM Memory & Reasoning Optimization | *LLMs, Python, FAISS, Asyncio, PyTorch*

- Designed LLM memory management layer with **context compression**, **FAISS**-backed semantic prioritization, and **fine-tuning** strategies to reduce token usage while preserving reasoning quality; **async** benchmarking engine across 10+ model configs on latency, memory, and accuracy.

EDUCATION

University of Cincinnati — M.Eng. Computer Science *Aug 2024 – Apr 2026*
Mahindra University — B.E. Computer Science *Aug 2020 – Jun 2024*

CERTIFICATIONS

Anthropic Prompt Engineering • Snowflake Generative AI Professional • NVIDIA Deep Learning • AWS • Databricks • IBM (2024–2025)