

Ayyappa Swamy Thati

Artificial Intelligence Engineer & System Design

Hyderabad, India | thatiayyappaswami@gmail.com | +91 8639813003 | [GitHub](#) | [LinkedIn](#) | [website](#)

SUMMARY

Backend & AI Engineer specializing in the architecture, deployment, and security of scalable Generative AI applications and multi-agent systems. Proven track record of designing high-performance analytics engines (Poseidon) and automated market research platforms that power public-scale, consumer data processing. Expertise in leveraging LLMs, dynamic LangGraph workflows, and DuckDB to drastically reduce manual research workflows while increasing analytical accuracy. Highly proficient in managing complete cloud lifecycles via GCP, deploying to Cloud Run, and enforcing robust VPS network security across custom Ubuntu server environments.

PROFESSIONAL EXPERIENCE

HAI Cub | [link](#)

May 2023 - July 2024

Intern · Part-time

Amrita WNA, Amritapuri

- HAI Club @ Amrita is research club of undergrad & graduate students who have come together to work on exciting cutting-edge research projects in Healthcare AI.
- Our primary focus is on exploratory and deployable research in AI models for creating better patient outcomes and making it easier for doctors in day-to-day clinical practise.

Amrita Wireless Network Applications

Sept 2023 - Dec 2023

Intern · Part-time

Amrita WNA, Amritapuri

- Using ML algorithm for disaster management, building a Unmanned Ariel Vehicle that detects damaged and near damage building during natural calamities and sends search and rescue of people.
- Pioneered a multi-task CNN deep learning architecture that predicts object instances, human poses, instance masks and tracks people end-to-end.

Jupiter Meta

Dec 2024 - Present

Artificial Intelligence & Machine Learning Engineer · Full-Time

- *Backend Architect for Public-Scale Platform*: Independently designed and engineered the complete backend for an adaptive AI market research platform (Data-Insighter), successfully scaling the system to support a mass public client base for consumer research.
- *High-Performance AI Engine*: Built a dynamic, multi-agent analytical pipeline using FastAPI, LangGraph, Google Gemini, and DuckDB, enabling instantaneous natural language querying and automated report generation.
- *Workflow Automation & Accuracy*: Developed an LLM-driven conversational agent that automates consumer survey generation and integrates complex data payloads directly into the backend. This system replaced manual workflows, reducing overall research cycle time by 95%.
- *End-to-End Cloud & Security Management*: Fully manage GCP infrastructure (Cloud Run, Kubernetes Engine), database scaling, and deployment lifecycles. Administer custom VPS network security, harden and optimize Ubuntu server instances, and implement IP-based fraud prevention protocols.

CONFERENCES AND WORKSHOPS

AI in Healthcare

Apr 2023

Amrita School of Engineering, Amritapuri

- Led hands-on experiential learning sessions, mentoring over 40 students in guiding them in practical exercises and adeptly troubleshooting technical issues. Organised workshop materials, technology, and resources to create an optimal learning environment for all participants across 3 days' workshop.

Amrita Blockchain Club

Jan 2024

Amrita School of Engineering, Amritapuri

- We started a blockchain club in campus to educate students about blockchain and evolution of it. Encourage students to show interest in the domain.

Retailers Association of India (RAI) Annual Event | Technical Presenter (Gold Sponsor)

Feb 2026

Jio retail center, Mumbai

- Delivered live technical pitches of the Data-Insighter AI platform to enterprise executives, demonstrating its automated survey creation and instant DuckDB analytics capabilities to successfully drive engagement and secure B2B leads.

SKILLS

- **Programming Languages:** Python, Go
- **Machine Learning & AI:** Supervised Learning(Linear Regression, Polynomial Regression, Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes, Decision Tree Classifier, Random Forest), Unsupervised Learning(K-Means, Time Series Analysis, Gaussian Mixture Models, DBSCAN), Deep Learning(Artificial Neural Networks (ANN), Gradient Descent, Activation Functions (Sigmoid, ReLU, SoftMax), Backpropagation, Forward Propagation, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Generative Adversarial Networks (GAN), Autoencoders, Transfer Learning), Reinforcement Learning(SARSA, DQN, SAC), GenAI(Large Language Models (LLMs), Generative AI SDKs(Gemini, OpenAI, Anthropic).)
- **Cloud Computing & DevOps:** Google Cloud Platform (GCP APIs, Cloud Run, Kubernetes Engine, VM Instances), Docker, Kubernetes, VPS security, google market place tools.
- **Web Technologies & Data Processing:** FastAPI, Web Scraping (aiohttp, bs4), Data Visualization (e.g., Matplotlib, Seaborn, Plotly), Pandas, NumPy.
- **Databases:** MongoDB, Postgres, MySQL, Vector Databases (Chroma, Milvus, Zillis), Neo4J & falkor Graph database.
- **Recent Development Work:** Advanced RAG Architectures, Deep Reinforcement Learning, Multi-Agent Systems, User Personas using Knowledge Graphs.

EDUCATION

Bachelor of Technology in Artificial Intelligence Engineering <ul style="list-style-type: none">• Amrita Vishwa Vidyapeetham, Amritapuri, Kerala	Oct 2021 - June 2025
Sri Chaitanya Educational Institution <ul style="list-style-type: none">• Sri Chaitanya Institutions, Kukatpally.	June 2009 - April 2021

PROJECTS

Leafy github Machine Learning	Feb 2023 - Jun 2023
Python, Pandas, NumPy, Tensorflow, Pytorch, Weights and Bias <ul style="list-style-type: none">• Advanced AI Architecture: Developed a state-of-the-art diagnostic engine using a custom Residual CNN integrated with CBAM Attention Modules and Stochastic Depth to achieve high-precision plant disease classification.• Production-Ready Training Pipeline: Implemented a multi-phase training strategy (progressive resizing) with RandAugment, Effective Number of Samples (ENS) class weighting, and OneCycleLR for robust model convergence.• Full-Stack ML Lifecycle: Managed the end-to-end workflow including data preprocessing, automated hyperparameter tuning via LR Finder, and experiment tracking using Weights & Biases.• AgTech Business Application: Engineered a scalable solution for real-time crop pathology, providing a framework for SaaS-based agricultural APIs to reduce pesticide use and improve global food security.	
Secure Remote Password link Cryptography	Nov 2023 - Jan 2024
Go, SHA256 Hashing, 2048-bit Modular Exponentiation, Sophie Germain Primes <ul style="list-style-type: none">• Zero-Knowledge Authentication: Engineered a system where the server authenticates users without ever receiving or storing their actual passwords, neutralizing risks from database leaks.• Cryptographic Security: Implemented SHA256 hashing and 2048-bit modular exponentiation using Sophie Germain primes to ensure resistance against man-in-the-middle and dictionary attacks.• Performance Optimization: Developed a multi-threaded prime number generator in Go that utilizes all CPU cores to efficiently handle high-bit-size cryptographic computations.• Business Integration: Designed a modular architecture suitable for securing IoT devices, financial platforms, or internal enterprise gateways requiring trustless identity verification.	
Pocket Sage link GenAI Application	Aug 2024 - May 2025
Python, Langchain, FastAPI, Assembly AI, Neo4J(graphRAG), <ul style="list-style-type: none">• Automated Clinical Documentation: Implemented an AI-driven pipeline using AssemblyAI and Gemini to transcribe medical appointments and automatically generate structured SOAP notes and clinical summaries in JSON format.• Fraud Detection: By mapping appointments and doctor interactions in Neo4j, insurers can detect irregular patterns or duplicate billing across facilities more effectively.• Scalable Hybrid Architecture: Engineered a dual-database system using MongoDB for session persistence and Neo4j Graph Database to model complex doctor-patient relationships and appointment history.	

JM Prism Search @ JupiterMeta | GenAI Application

Jan 2025 - June 2025

Python, FastAPI, Google Gemini API, GCP (Cloud Run, Kubernetes Engine, Vector DB), aiohttp, bs4, SentenceTransformers, ChromaManager, Crawl4ai

- Developed an advanced AI-powered search engine that scrapes the web to provide contextualized answers, similar to Perplexity.
- Engineered a multi-agent system for intelligent query generation from user input, utilizing Google Custom Search for URL retrieval and crawl4ai for efficient web content scraping.
- Implemented sophisticated data ingestion pipelines with multiple chunking strategies and stored processed data in a Milvus vector database with HSNW strategy.
- Designed and optimized a Advanced Retrieval Augmented Generation (RAG) algorithm to surface the most relevant information, leveraging Google Gemini API for generating coherent and accurate answers.

Hercules @ JupiterMeta | Market Research using GenAI

May 2025 - Ongoing

Python, FastAPI, Google Gemini API, Google Cloud Run

- Automated Survey Generation: Leverages Google Gemini AI to transform high-level research goals into structured, professional surveys, significantly reducing the manual effort for product developers.
- Seamless Deployment Ecosystem: Acts as the bridge between survey creators and the SuperJ respondent app, allowing businesses to deploy surveys directly to a target audience and collect real-world data.
- Enterprise-Grade Scalability: Built with a high-performance asynchronous architecture featuring robust rate-limiting, connection pooling, and structured JSON logging for reliable production operations.
- Data-Driven Market Insights: Integrated with an analytics engine to process raw respondent data from SuperJ, providing startups and FMCG companies with actionable feedback for product development.

Poseidon @ JupiterMeta | Data Science using GenAI orchestration

May 2025 - Ongoing

Python, FastAPI, Google Gemini API, Google Virtual Server, OLAP DB(duckDB)

- Architected an AI-driven analytics platform using Python, FastAPI, and LangGraph, enabling users to generate automated, executive-ready market research reports through natural language queries.
- Engineered a dynamic multi-agent workflow powered by Google Gemini to autonomously classify user intent, generate analytical paths on the fly, and select optimal statistical methodologies without relying on rigid templates.
- Optimized system performance and scalability by implementing a high-concurrency DuckDB SQL engine, replacing memory-heavy Pandas dataframes to achieve sub-second query responses on large datasets.
- Drove business value for non-technical stakeholders by building a context-aware visualization engine that embeds charts inline with data narratives, successfully reducing research cycle times by 50% and improving analytical accuracy by 25%.

ADDITIONAL INFORMATION

- **Languages:** English, Hindi, Telugu.
- **Certifications:** Introduction to Deep Learning in Python | datacamp.
- **Awards/Activities:** Second prize in Project contest at Convergence 2k25 @ VNRVJEIT.