

Muthaiya Thavamani

muthu.thavamani.03@gmail.com

AI-Native Engineer | Technical Architect

+(91) 75027 51029

AI-Native Engineer and Technical Architect with 11 years of experience, operating as a Forward Deployed Engineer - embedding with teams to take Generative AI from discovery to production. Agentic systems, RAG, and copilots that take real actions, backed by strong data engineering and predictive ML. I've shipped products from architecture through rollout across multiple early-stage startups and enterprise clients.

Skills

AI / ML Engineering	- Generative AI & LLM apps: foundation models (OpenAI, Anthropic Claude, LLaMA), RAG, prompt engineering, fine-tuning, embeddings, vector databases, pgvector, Azure AI Foundry, Copilot Studio - Agents: multi-agent, LangChain, LangGraph, MCP, tool/function calling, copilots; LLM evaluation, guardrails, Responsible AI; TensorFlow, forecasting, anomaly detection, computer vision
Data Engineering	- Serverless ETL & delta pipelines, data validation, governance & compliance - Databricks Certified Professional, medallion architecture, PySpark, Delta Lake, Teradata, Airflow
Blockchain & Web3	- Solidity, OpenZeppelin, HardHat, Foundry; DID (Privado ID) - EVM, Polygon CDK, Avalanche Subnet, Cosmos SDK, Tezos (SmartPy), Hyperledger Fabric
Languages, Frameworks & Data	- Golang, Python, JS / TS (NodeJS), Solidity, Java; - Flutter (Dart) & React Native; - FastAPI, ExpressJS, Hasura (GraphQL) - PostgreSQL, MySQL, MongoDB, DynamoDB, Cloud Firestore
Cloud & Infra	- AWS, GCP & Azure - Serverless & microservice deployment, Docker

Professional Experience

Lead AI Engineer, Contract

Sep 2025 - Present

- Forward-deployed AI delivery - embedded with clients to ship Generative-AI products from discovery to production.

- **AIDLC - AI-Ops Platform:** a LangChain / LangGraph multi-agent engine that automates the path from requirement to delivery, with prompt engineering and LLM evaluation built into the loop.
- **Governed Asset Agents (Enterprise Asset Management):** a multi-agent system that predicts, analyses, and acts on asset operations - combining anomaly detection with LLM reasoning, and wrapping autonomous actions in guardrails, Responsible-AI checks, and human-in-the-loop approval so every action stays auditable.
- **AI Copilot over MCP:** connects to any system through a secured Model Context Protocol link, using RAG over enterprise data plus tool / function calling to turn scattered tools into one conversational interface, with evaluation keeping responses grounded.
- **Predictive HVAC Engine:** an agentic copilot and predictive engine for healthcare HVAC, applying time-series forecasting and anomaly detection on live telemetry to flag failures before they happen.
- **IT Support Copilot:** a policy-driven, school-safe RAG copilot for Multi-Academy Trusts - a governed agentic workflow with a three-tier retrieval cascade (hybrid pgvector + BM25, RRF fusion) on Azure AI Foundry and Copilot Studio, grounded by content filtering and groundedness evaluation, with structured auto-escalation to IT.

Technologies: Python, Generative AI, LLMs (OpenAI, Claude, Hugging Face, LLaMA), LangChain, LangGraph, MCP, RAG, Hybrid Search (RRF), pgvector, Azure AI Foundry, Copilot Studio, Power Automate, FastAPI,

Lead Data Engineer, Greenmen Consulting

Oct 2024 - Aug 2025

- Led end-to-end data migration from S3, SQL, Databricks, and Teradata into Salesforce CRM, embedded with data owners to enforce data quality, governance, and compliance.
- Built a serverless system automating delta (incremental) loads over millions of records, and a Databricks medallion architecture (Bronze/Silver/Gold) feeding curated data to the PowerBI team.
- **BRD Engine** - Implemented this multi-agent system (LangChain / LangGraph) from ideation through production.

Technologies: Python, Databricks, PySpark, Delta Lake, Multi-Agent AI, Teradata, DynamoDB, AWS.

Product Development Head, Unibase Labs

June 2022 - Sep 2024

- Partnered with founders to take Web3 platforms from idea to production, owning architecture, decentralized identity (DID), and L1 development on Avalanche Subnet and Polygon CDK.

- **Turf Network** - led an esports GameFi ecosystem end to end; built a Privado ID identity system publishing 400K+ DID's and Solidity contracts for the Esports Engine, token sale, and node sale.
- **Instaraise (Tezos Launchpad)** - re-architected the IDO platform as serverless, leading backend services and the entire AWS infrastructure.

Technologies: Golang, Python, Solidity, Polygon CDK, Privado ID, Tezos, Cosmos SDK.

Lead Blockchain Engineer, [Coinfantasy](#)

Dec 2021 - June 2022

- Founding engineer - built smart contracts (game, stake/unstake, ERC20) and a Golang game scheduler handling high volumes of parallel events.
- Integrated Biconomy gasless transactions, Lossless protocol, and providers (Magic.link, Transak, CoinGecko, CoinMarketCap); mentored and helped build the core team.

Technologies: Golang, NodeJS, Solidity, HardHat, MySQL, MongoDB, ReactJS, AWS.,

Backend Security Engineer, [Securex AI](#)

Oct 2021 - Mar 2022

- Core Golang backend engineer - integrated cloud-security platforms (Veracode, Dome9, Qualys, Azure Defender) and AWS/GCP asset connectors via their official SDKs.
- Built a vulnerability service pulling CVE data and remediations from the NVD.

Technologies: Golang, Python, AWS, GCP, Azure, Hasura, Postgres.,

Lead Member Technical Staff, [Imstrong](#)

Dec 2019 - Nov 2021

- Built an ML fitness product using TensorFlow PoseNet - real-time pose estimation (computer vision) scoring yoga and exercise posture accuracy with live form feedback.
- Backend engineer for the video-streaming platform: Golang microservices, a self-built serverless live-streaming system (nginx-rtmp + Docker), VOD tooling, and AWS/GCP infrastructure.

Technologies: Python, TensorFlow (PoseNet / Computer Vision), Golang, NodeJS, Firestore, nginx-rtmp, AWS, GCP.

Software Engineer, Sysvine Technologies

Dec 2015 - Nov 2019

- Peloton Blockchain - ERC20 chaincode in Go on Hyperledger Fabric (Fabric Node SDK).
- Broadridge (Dojima Trade & Clear) - order/market-data services over FIX/FIXML and the NASDAQ Genium API.
- Elsevier (GeoFacets) - content & metadata extraction, OCR, and geotagging pipelines for map-based discovery.

Technologies: GoLang, Java, NodeJS, Python, Blockchain (Hyperledger), Websocket, IBM MQ
