

## NITHISH SURESH BABU

Lead AI Engineer · Data Scientist · AI Solutions Architect  
+1 2062268935 | nithish952001@gmail.com | linkedin.com/in/nithish-suresh-babu  
github.com/nithish-95 | portfolio.nithish.net

### SUMMARY

---

AI Engineer with 7+ years of experience building agentic AI systems, financial automation pipelines, and reusable AI accelerators for PE-backed portfolio companies and enterprise clients. End-to-end owner across discovery, architecture, and production deployment. Deep expertise in LLM orchestration, Salesforce integrations, and cloud-native data engineering.

### TECHNICAL SKILLS

---

**AI / LLM:** LangGraph, LangChain, OpenAI API, Anthropic API, HuggingFace, TensorFlow, PyTorch, XGBoost, FastText, BERT, Prophet, Scikit-learn, MLFlow

**Languages:** Python, JavaScript, React, SQL, SOQL, PySpark

**Cloud:** AWS (ECS, Lambda, S3), Azure (ADF, ACI, ACR, DevOps, AzureML), Databricks, Snowflake

**Integrations:** Salesforce, NetSuite, SAP, Microsoft Graph API, SharePoint, N8N

**Dev Tools:** Cursor, Claude Code, OpenAI Codex, Docker, Git, Azure DevOps

**Visualization:** Power BI, Tableau, Plotly, Seaborn

### EXPERIENCE

---

#### **TECH TORCH** - *Lead AI Engineer*

Jul 2025 – Present

*Remote · South Elgin, IL*

- Architected and sole-developed an AI-powered ARR/NARR Calculator replacing a manual process handled by 5–10 finance ops staff, reducing calculation time from days to seconds
- Engineered a trigger-based LangGraph pipeline that activates on Salesforce Opportunity stage change, resolves complex deal lineage (renewals, amendments, upgrades), and writes computed ARR/NARR matrices directly back to Salesforce
- Owned full-stack delivery — requirements gathering, system design, Python backend, React frontend, AWS deployment — with zero external dependencies
- Eliminated 1–3 hrs of daily manual FP&A effort across 4–5 finance team members by building a fully automated report pipeline ingesting Salesforce and NetSuite data and delivering daily outputs to SharePoint via Microsoft Graph API
- Deployed an internal Revenue Leakage Accelerator for a portfolio company, surfacing contract-vs-invoice discrepancies and quantifiable revenue leakage KPIs previously invisible to leadership
- Built a production Document Intelligence Accelerator ingesting 6+ formats (PDF, scanned PDF, DOCX, Excel, XML, HTML) with LangGraph and OpenAI, feeding structured output to downstream AI apps, data warehouses, and audit workflows
- Led AI engineering for TechTorch's Data Practice pod — drove PRD creation, architecture, and POC delivery across 3–5 internal accelerators simultaneously
- Selected by the company chairman to demo and present the Revenue Leakage tool in a client-facing webinar; consistently recognized at monthly all-hands for technical leadership
- Mentored 3 junior engineers on AI architecture patterns, LLM tooling, and production best practices
- Delivered end-to-end solution scoping, system design, and project estimates for new client engagements, translating ambiguous requirements into structured technical proposals
- Built a Salesforce CRM data cleansing and migration pipeline with automated validation, reducing manual data remediation effort for a portfolio company
- Accelerated delivery across all projects using AI-assisted development tools including Cursor, Claude Code, and OpenAI Codex

#### **SQUARE RESULTS** - *Software Development Engineer - GenAI*

Aug 2023 – July 2025

*Remote, WA*

- Architected RAG based job matching pipeline using Gemini 2.5 Flash that ingests user resume from AWS S3 and queries

- relevant job descriptions from AWS Bedrock knowledge base (a vector database), generating personalised job recommendations
- for the user. This pipeline runs periodically, and the results are stored in Supabase for fast retrieval to be shown on the user's home
- page. This approach improves match accuracy through semantic understanding of skills and job requirements over traditional text-based search.
- Automated deployment by building a GitHub Actions workflow that containerises backend services with Docker, pushes images
- to AWS ECR, deploys them to AWS App Runner, and deploys the frontend to AWS Amplify, resulting in faster releases and fewer
- errors.
- Built a secure, multi-tenant backend with FastAPI (Python) and responsive frontend interfaces with Next.js, ensuring strict data
- isolation, reducing data-leak risks, and enabling safe access for all users.
- Collaborated with stakeholders across product and data teams to demo real-time recommendation capabilities, translating technical
- AI concepts into actionable business insights for leadership.

**CLASSA.AI - Software Developer**  
Contract Remote, WA

Jul 2022 - Jul 2023

- **Re-architected a document generation workflow** for an aviation insurance platform using **Temporal** to manage a complex **30-branch decision tree** for private jet insurance policies.
- **Replaced a fragmented, database-backed state machine** with a **durable workflow model**, eliminating the need for manual state persistence and complex database triggers across a multi-step user wizard.
- **Enhanced data consistency and fault tolerance** by leveraging **Temporal's** stateful history, providing a deterministic way to resume interrupted document generations and automate policy delivery to high-net-worth clients.

## ACADEMIC PROJECTS

---

### SCALABLE CHAT APPLICATION

- **Engineered a real-time, low-latency** chat application using **Go** and **WebSockets** to support bi-directional communication across **multiple concurrent clients**. I designed a centralised "hub" to manage the lifecycle of active connections, ensuring that incoming messages are broadcasted across the participant pool with minimal delay.
- Implemented a **thread-safe** connection manager using Go's concurrency primitives **goroutines** and **channels** to handle simultaneous messaging streams without **race conditions**. By decoupling message ingestion from the broadcasting logic, I ensured that the system remains responsive even under **high-concurrency loads**, preventing blocking operations in the main execution path.
- Focused on operational stability by incorporating heartbeat monitoring and graceful cleanup of stale **WebSocket** sessions to mitigate **memory leaks** and **resource exhaustion**. I utilised a clean, modular architecture that separates the transport layer from the business logic, making the system easy to extend with features like private rooms or persistent message storage.

### CLOUD COMPUTING

#### MULTIPLAYER TIC-TAC-TOE

- Architected and deployed a real-time Tic-Tac-Toe application using Go and WebSockets on **AWS Fargate**, leveraging a **serverless container model** to eliminate infrastructure overhead. I containerised the service using a **multi-stage Docker build** to minimise the attack surface and image size, ensuring a lean production environment by separating the Go build pipeline from the final Alpine runtime.
- Designed a secure networking stack featuring an **Application Load Balancer (ALB)** and **Route 53** for **custom domain mapping and traffic orchestration**. I implemented granular IAM role separating execution permissions from task logic and configured isolated security groups to ensure least-privilege access, allowing only necessary traffic from the ALB to the containerised application.
- Established a **scalable deployment workflow** using **Amazon ECR** and **ECS**, enabling **automated lifecycle management** and **cost-efficient resource scaling**. By moving to a Fargate-based architecture, I transitioned the application from a manual setup to a production-ready environment that supports durable WebSocket connections and automatic health monitoring.

## OPEN SOURCE CONTRIBUTIONS

---

### Go library for accessing the GitHub v3 API | go-github – Contributor

- **Engineered and implemented new API methods** to manage **GitHub Enterprise App** repository installations, successfully enhancing the functionality of the widely-used library.
- Authored **production-quality** Go code and comprehensive unit tests to guarantee the reliability and stability of new library features.
- Collaborated with other maintainers through the Pull Request (PR) review process, refactoring code to align with established project conventions and Go best practices.

### Time Tracking & Project Management Platform | Chrono - Project Maintainer

- **AI-Powered Project Management Suite:** Architected a unified **Time Tracking and Kanban** platform for small teams, featuring a custom **AI Analytics layer** that ingests workflow data to generate productivity insights using multiple LLM models.
- **Modern Full-Stack Implementation:** Built a high-performance application using **Next.js 16** and **React 19**, leveraging Server Actions and Server-Side Rendering (SSR) for optimal SEO and speed.

**Data & Schema Design:** Designed a relational schema on **PostgreSQL** to manage complex relationships between users, timesheets, and project tasks, ensuring data integrity and efficient querying.

## ACHIEVEMENTS

---

- **People's Choice of the Year 2025 (Google Cloud Developers):** Designed and Developed AI Voice agent that Automates the Insurance Quoting via Phone calls

## EDUCATION

---

### University Of Michigan Graduated - Apr 2022

*Master's, Computer and Information Science, Michigan, USA*

## CERTIFICATIONS

---

IBM Data Science Professional · Intel AI Foundation Competency · Dataiku Core Designer · AWS Lambda Foundations · Applied Concurrency in Go · Microsoft Power BI Analyst (In Progress)