

Samith Deshai Siddo

☎ +1(940)843-8866

✉ siddosamith10@gmail.com <https://www.linkedin.com/in/samith-deshai-siddo-787830353/>

Summary

AI Engineer and Full Stack Developer with 3+ years of experience designing intelligent systems, developing agentic AI frameworks, and deploying scalable cloud architectures. Skilled in Python, Java, JavaScript, React, and Node.js with hands-on expertise in LangChain, LangGraph, and Retrieval-Augmented Generation (RAG). Experienced in integrating LLMs with backend services, tool-calling workflows, and real-time voice agents to deliver autonomous, data-driven applications. Adept at optimizing system performance, balancing reliability and cost at scale, and collaborating cross-functionally to translate ambiguous product needs into measurable, production-ready solutions.

Education

- | | |
|---|------------------------------|
| University Of North Texas
<i>Master of Science in Computer Science (GPA: 3.90 / 4.00)</i> | May 2025
Denton, TX |
| <ul style="list-style-type: none">• Relevant Coursework: Analysis of Algorithms, Software Engineering, Fundamentals Of Database, Natural Language Processing, Information Retrieval, Software Development For AI | |
| BV Raju Institute of Technology
<i>Bachelor of Technology in Computer Science (GPA: 3.32 / 4.00)</i> | May 2023
Hyderabad, India |
| <ul style="list-style-type: none">• Relevant Coursework: C, Java, Data Structures and Algorithms, Operating Systems, Software Engineering, Database Management Systems, Design Patterns, Web Technologies | |

Experience

- | | |
|--|---|
| Data Color AI
<i>Forward Deployed AI Engineer</i> | Jan 2025 - Present
Dallas, TX |
| <ul style="list-style-type: none">• Built a LangChain and LangGraph-powered agentic AI workflows on AWS Bedrock to orchestrate reasoning and decision-making across unstructured enterprise data, streamlining analyst workflows and reducing manual triage time by 55%.• Built an AI Data Steward an MCP client leveraging the Reltio AgentFlow MCP Server to analyze entity conflicts, simulate match and merge outcomes, validate governance constraints, and route human-in-the-loop approvals, significantly reducing manual stewardship workload in enterprise MDM workflows.• Built DocuAssist, an Informatica AI Agent Engineering Platform-based AI agent that converts unstructured documents into governed, structured data sources through RAG-driven extraction and reasoning, reducing manual document processing and accelerating ingestion by 3 times.• Developed and deployed Datacolor Atlas Co-Pilot, a domain-specific MCP client integrating with multiple back-end servers such as Salesforce and Informatica MDM, automating contextual recommendations and reducing support ticket volume by 40%.• Implemented a Spark-based data quality validation framework on Databricks, enabling LLM-driven evaluation and automated generation of data quality rules; integrated with MLflow for continuous monitoring and feedback, enhancing data reliability and model observability across pipelines.• Built secure data integration pipelines between MCP servers and third-party clients using AWS Cognito OAuth and Bedrock AgentCore runtime, ensuring compliance with internal data governance and maintaining service reliability.• Leveraged and deployed open-source Hugging Face models (embedding model, instruction-tuned LLMs) within RAG and agentic workflows, optimizing inference latency and improving model retrieval quality across enterprise search and document processing pipelines. | |
| Club InQuizitive
<i>Full Stack Developer</i> | Aug 2021 - May 2023
Hyderabad, India |
| <ul style="list-style-type: none">• Led a team of developers in building the club's main website and several event-specific platforms from scratch.• Developed responsive web applications using React, JavaScript, and Tailwind CSS, with PostgreSQL and Firebase for backend services and data management.• Deployed and hosted websites to support real-time updates and event registrations, ensuring smooth user experience and accessibility. | |

- Worked closely with club leadership to turn ideas into engaging digital experiences that supported club activities and outreach.

Projects

Smart Mapping | *Python, RAG, Gemini, BigQuery VECTOR_SEARCH, GCP Cloud Run, Docker*

- Architected an end-to-end hybrid RAG pipeline for post-acquisition data integration, reconciling each acquired brand's idiosyncratic service codes to a parent roll-up's canonical catalog on a clear contract: rules advise, retrieval supplies candidates, the LLM decides, and confidence gates a human.
- Designed the retrieval layer to union BigQuery VECTOR_SEARCH embeddings with TF-IDF candidates so semantic paraphrases and lexical near-matches both reach a Gemini re-ranker, with text normalization, knowledge-base rules, and category pinning narrowing the candidate space before any LLM call.
- Productionized the system on GCP — Docker images in Artifact Registry, Cloud Run Jobs, GCS staging, and a normalized BigQuery audit schema with per-run token and cost tracking — onboarding 20 acquired brands and thousands of codes with zero processing errors across delivered runs.
- Built the confidence model that auto-escalates ~25–35% of rows to analyst review and auto-accepts the rest, focusing human effort where the model is uncertain and delivering an estimated 30–60× speedup over fully manual code mapping.
- Engineered a 12-month rate signal (Job/Initial/Contract usage) and an authoritative product-type prior as data-backed classifiers reaching 94–100% agreement with historical billing behavior; shipped the prior as a “strong prior + disagreement flag” rather than a hard override after error-asymmetry analysis surfaced 88 products a keyword heuristic had misclassified in a single brand; hardened by 76 pytest cases and an adversarial review over a v6→v13 release cycle.

AI Match Resolution Workbench — Agentic Data Steward for Reltio MDM | *Python, FastAPI, Azure OpenAI (GPT-5.4-mini), MCP, Azure Service Bus, Cosmos DB, React/TypeScript*

- Engineered an agentic AI data steward that automates match resolution in Reltio MDM — an LLM agent reasons over the Reltio MCP server (100+ tools) to drive merge/reject/defer decisions with human-in-the-loop review and a full audit trail.
- Scaled AI analysis horizontally with the competing-consumers pattern — Azure Service Bus feeding a stateless worker fleet on Container Apps, hardened with dead-lettering and automatic lock renewal.
- Cut LLM cost and latency through fleet-wide prompt caching (stable cache-key routing) and a one-time agent and MCP bootstrap per worker instead of per request.
- Made irreversible MDM merges idempotent and retry-safe with cross-job decision guards.

Technical Skills

Languages: Python, Java, JavaScript

Databases: PostgreSQL, BigQuery, Pinecone Vector DB, MongoDB, MySQL, Firebase, Chroma

Frameworks: React.js, Node.js, Next.js, jQuery, Flutter

AI: Retrieval Augmented Generation, Hybrid Retrieval (TF-IDF + Vector Embeddings), LLM Re-ranking, Tool Calling, LangChain, LangGraph, Model Context Protocol (MCP), Hugging Face Transformers, Human-in-the-Loop ML

Automation & RPA: UiPath (workflow automation, orchestrator concepts), Serverless automation patterns

Cloud Platforms & Big Data: AWS Bedrock, GCP (Cloud Run, BigQuery VECTOR_SEARCH, GCS, Artifact Registry), Azure, Databricks

Version Control & Tools: Git, GitHub, Jira, Docker, pytest

AI Development tools: Cursor, Claude Code

Certifications

deeplearning.ai: MCP: Build Rich-Context AI Apps (Anthropic) | LangChain for LLM Application Development | JavaScript RAG Web Apps with LlamaIndex

Udemy: The Complete 2024 Web Development Boot Camp

Achievements

Won 1st place at UNT GradInnoHack 2025 for building MediCall

Placed in the top 10 at the Google Gemini × Pipecat Hackathon at Y Combinator (YC 2025), a flagship SF Tech Week event bringing together top innovators in Voice and Agentic AI.

Launched a Flutter-based mobile app with 500+ users and a 4.5 star rating on the Play Store.

Published a research paper on the design and implementation of a product recommendation system

ieeexplore.ieee.org/document/10212422