

# Siddharth

## EXPERIENCE

### AI ENGINEER

Aug 2024 - Present | India

- Designed PoC for AI-driven Revenue Cycle Management (RCM); deployed LangChain/Flowise workflows with (Pinecone/Weaviate) VectorDB for insurance document processing, reducing manual review 90%.
- Built RAG pipeline with chain-of-thought reasoning (90% accuracy) for claims adjudication; implemented human-in-the-loop validation for low-confidence cases.
- Engineered dynamic token allocation for GPT-4 endpoints (40% cost reduction), monitored via LangSmith (99.9% SLA adherence).
- Containerized AI services with Docker/K8s ; built Flask/FastAPI ETL pipelines (40% latency reduction).
- Partnered with DevOps to automate scaling for healthcare/finance clients; processed structured/unstructured insurance data.

### SDE-1

July 2022 - Oct 2023 | India

- Customized Flowise AI open source codebase to achieve an enterprise proprietary version with RBAC, multi-tenant per user data isolation and custom nodes for NIIT AI factory  
Used: (Node.js, Typescript, Python, JWT, OAuth2, React.js, MongoDB, Azure OpenAI, Microservices)
- Design and developed multilingual AI voice assistant to facilitate training simulation for Metlife Client Support, Improving interaction efficiency to 97 percent

### NIIT | INTERNSHIP

Jan 2022 - June 2022 | Gurugram, India

- Reduced learner dashboard latency by 70 percent through service decomposition and DB optimization
- Improved system security and test coverage to 90 percent using React.js, Java, Springboot, AWS, Docker

## SKILLS

- Strong MERN and Python feature development & designing skills
- Multi-Agent AI workflow orchestration using Langchain, Flowise, n8n
- Fine-tuning LLM models, vector database (Weaviate, Pinecone) , RAG , Document Data extraction, setting up ETL pipelines using Flask/FastAPI and Celery.
- Token monitoring, token usage cost optimisation and workflow response time management using Langsmith
- Understanding of tokenisation, attention mechanisms and transformer architecture
- Team player, Good Analytical, Presentation and Interpersonal Skills with proficiency in demonstration, business requirements and collaborating with cross-functional teams.
- Ability to deploy and scale AI stack using Docker, AWS Sagemaker and Kubernetes.

## EDUCATION

### DELHI

BTech INFORMATION TECHNOLOGY

July 2022 | Delhi

CGPA 8.41/10

## TECHNOLOGIES

- Core Stack:
  - React • Node • Flask • Python
  - LangChain • LlamaIndex • Flowise
  - FastAPI • Microservices • N8n
- Databases:
  - MongoDB • MySQL • Postgres
  - VectorDB (Weaviate/Pinecone)
  - Neo4j • Redis
- Infra:
  - AWS SageMaker • Azure OpenAI
  - CI/CD (GitLab)
  - Docker • Kubernetes
- Tools:
  - LangSmith • RabbitMQ
  - Websockets
- Standards:
  - REST • GraphQL • JWT
  - OAuth2 • OpenAPI

## ACHIEVEMENTS

- Participant in Google Gemini API Developer Competition (2024)
- 1st position (from 20 teams) Hackathon @IncubateIND April 2020
- Ranked 42 (out of 13522) Backend Engineering Contest @NextLevel (2024)
- Modified the Flowise(11m AI based framework) opensource code to support custom chatbot creation (2023)
- Solved 200+ problems on Leetcode and Codechef and solving more!