

SUMMARY

AI Engineer and Software Engineer with 3+ years of combined experience building scalable enterprise applications and production AI-native systems. Hands-on expertise in Agentic AI, LangGraph, LangChain, MCP, pgvector RAG, OpenAI APIs, and FastAPI, developing intelligent workflows and enterprise AI integrations. **U.S. Work Authorization: F-1 STEM OPT (36 months of immediate, continuous work authorization; no immediate H1B transfer or sponsorship fees required).** Strong background in backend engineering, microservices, CI/CD, and cloud deployments using Python, TypeScript, Java, PostgreSQL, Docker, and GitHub Actions, with a passion for delivering production-ready AI products and automation solutions.

PROFESSIONAL EXPERIENCE

AI Engineer Intern Scale AI USA	Feb 2026 – Present
Collaborated with senior AI engineers to build and test multi-agent workflows using Python, LangGraph, and FastAPI, contributing to enterprise Agentic AI solutions that improved task orchestration efficiency by 22% during client trials. Built and integrated Model Context Protocol (MCP) connectors under senior guidance, enabling AI agents to securely query internal knowledge bases and external APIs, which reduced overall tool-latency overhead by 15%. Assisted senior engineers in optimizing Retrieval-Augmented Generation (RAG) pipelines using PostgreSQL, pgvector, and semantic embedding-based retrieval to ground LLM responses, improving context retrieval accuracy to 94% in staging deployments. Designed and tested reusable agent workflows incorporating planner, researcher, executor, and reviewer patterns under senior supervision, helping teams accelerate client-specific adaptation timelines by 30%. Supported the integration of LLM applications with enterprise services through function calling and REST APIs under senior direction, automating document analysis and achieving structured JSON outputs with high consistency. Conducted prompt engineering trials and assisted in establishing system evaluation frameworks, validating structured outputs and documenting edge cases to reduce model hallucinations by 18% in staging. Assisted with containerized development testing environments using Docker and GitHub Actions, working with the DevOps lead to automate CI/CD verification and resolve staging integration bottlenecks. Collaborated with cross-functional engineering teams to triage application performance issues and analyze system logs, helping maintain 99.9% availability for enterprise client pilot runs.	
Software Engineer Chargebee India	July 2021 – July 2024
- Designed and developed scalable backend services and RESTful APIs using Python, Java, and SQL to support subscription billing and payment workflows, processing thousands of daily transactions while improving system performance by 30% and reducing service response times. - Engineered microservices-based solutions to enhance platform scalability and maintainability, collaborating with distributed teams to implement reusable components that reduced development effort by 26% and accelerated release schedules. - Developed and optimized database queries, indexing strategies, and data access layers across PostgreSQL and MySQL environments, improving query execution efficiency by 35% and minimizing production bottlenecks affecting critical customer workflows. - Implemented CI/CD pipelines utilizing Jenkins, GitHub Actions, and Docker to automate build, testing, and deployment processes, decreasing deployment failures by 32% and reducing release preparation efforts significantly. - Partnered with frontend developers, QA teams, and business stakeholders throughout the SDLC to translate requirements into technical solutions, contributing to successful delivery of customer-facing enhancements with high stakeholder satisfaction. - Introduced automated unit testing and integration testing frameworks that increased code coverage to over 85%, improving defect detection during development cycles and reducing post-release production incidents by 27%. - Monitored application health using logging and observability tools to proactively identify performance issues, conducting root cause analysis that improved platform stability and contributed to maintaining service reliability targets. - Participated in architectural reviews, peer code reviews, and technical mentoring initiatives to uphold engineering best practices, improving code quality standards and reducing technical debt across multiple product modules.	

TECHNICAL SKILLS

- Languages:** Python, TypeScript, JavaScript, Java, SQL, C.
- AI & LLM Engineering:** OpenAI API (GPT-4o, Whisper), LLM Integration, Prompt Engineering, RAG, Context-Injection Memory, LangChain, LangGraph, MCP (Model Context Protocol), AI Agents, Multi-Agent Orchestration, Embeddings, Vector Databases, Claude Code, Groq.
- Machine Learning:** Scikit-learn, PySpark MLlib, TensorFlow, Model Evaluation, Feature Engineering, NLP, Deep Learning Fundamentals.
- Frontend:** React, Next.js 16, Tailwind CSS, Shadcn/UI, HTML, CSS.
- Backend & Security:** FastAPI, Spring Boot, Node.js, REST APIs, JWT Authentication, Supabase Auth, Row-Level Security (RLS), Stripe API, Webhooks, Microservices.
- Cloud & DevOps:** AWS (S3, EMR, Lambda, Hive), Docker, Vercel, Render, Git, GitHub Actions, Jenkins, CI/CD, Bash/Shell Scripting, MLOps.
- Databases:** PostgreSQL, Supabase, MySQL, MongoDB, Redis, Hive, Vector DBs (pgvector, ChromaDB, Pinecone, FAISS).
- Data Engineering:** PySpark, Pandas, NumPy, ETL Pipelines, Distributed Data Processing, Tableau.
- Core CS:** Data Structures, Algorithms, System Design, Object-Oriented Programming, Design Patterns, SOLID Principles, Distributed Systems, Concurrency.
- Methodologies:** Agile, Scrum, SDLC, Test-Driven Development, Code Reviews, Technical Documentation.

PROJECTS

Deployed: [unfrozen.me](#)

UNFROZEN EMBER — AI Coaching Platform

- Built and deployed a production-ready AI coaching platform UNFROZEN (unfrozen.me) powered by OpenAI GPT-4o-mini and Whisper, delivering personalized communication coaching through persistent user context and real-time feedback for active paying subscribers.
- Engineered a custom cross-session context-injection memory system (bypassing high-latency vector database lookups), securing a **97%** gross margin and driving session API operating costs down to **\$0.09**.
- Implemented robust multi-tenancy database isolation using Supabase Row-Level Security (RLS) policies with service role keys via a Next.js proxy, protecting sensitive communication transcripts.

E-commerce Churn Prediction System | Cloud Data Engineering & ML

- Built an end-to-end cloud-native ML pipeline on AWS (S3, EMR, Hive, Lambda) processing 500K+ customer records, reducing pipeline runtime by **40%** through distributed Hive query optimization and partitioning.
- Developed a churn prediction model with PySpark MLlib achieving **79.9%** accuracy, enabling targeted retention strategies projected to reduce churn by **15%**.
- Engineered automated ETL workflows and built Tableau dashboards translating model output into actionable recommendations for non-technical stakeholders.

EDUCATION

Master of Science (M.S.) in Data & Artificial Intelligence	University of North Texas, Denton, USA	Aug 2024 – May 2026
Bachelor of Technology (B.Tech.) in Electronics and Communication Engineering	SVCE, Tirupati, India	Sep 2019 – May 2023

CERTIFICATION

- AI Fluency: Framework & Foundations — Anthropic (2026)
- Claude 101 — Anthropic (2026)
- IBM Machine Learning Professional Certificate — Supervised & Unsupervised Learning, Deep Learning, ML Pipelines (2025)