

Henil C. Alagiya

AI Engineer | Agentic Systems & RAG

+1 (848) 336-7008 | henilalagiya@gmail.com | Jersey City, NJ

github.com/henilcalagiya | linkedin.com/in/henilca | henilca.com

Summary: AI Engineer with **2+ years** of software engineering experience specializing in building **production-grade agentic AI systems** and **advanced RAG architectures**. Expert in **LangChain**, **LangGraph**, and **vector databases** with sophisticated retrieval optimization techniques including context padding and hybrid search strategies. Published **2 AI packages on PyPI** with **2,800+ downloads**, demonstrating expertise in **agent orchestration**, Model Context Protocol (MCP), and **LLM integrations**. Proficient in autonomous systems, real-time data processing, and full-stack development with **Python**, **React.js**, and modern AI frameworks (**PyTorch**, **TensorFlow**). Proven track record of optimizing AI system performance and delivering measurable improvements in enterprise applications.

EDUCATION

Master of Science, Computer Science, New Jersey Institute of Technology (NJIT), GPA: 3.55/4.00 **2023 — 2025**

Bachelor of Engineering, Computer Engineering, Gujarat Technological University (GTU), CGPA: 7.13/10.00 **2019 — 2023**

TECHNICAL SKILLS

Programming Languages: Python, JavaScript, TypeScript, C++, SQL, Bash

AI & ML Frameworks: LangChain, LangGraph, PyTorch, TensorFlow, Scikit-learn, Keras, Hugging Face Transformers

AI Technologies: Google AI SDK (Gemini), OpenAI API, RAG, Vector Databases, Embeddings, Agent Orchestration, LangSmith, Prompt Engineering, Fine-tuning, Model Context Protocol (MCP)

Web Frameworks: React.js, Node.js, Django, Next.js, FastAPI, Flask, Express.js

Databases: Pinecone, FAISS, ChromaDB, Weaviate, PostgreSQL, MySQL, MongoDB, Redis, SQLite

APIs & Integration: REST APIs, GraphQL, WebSockets, OAuth 2.0, gRPC, Webhooks

DevOps & Cloud: Docker, AWS (EC2, S3, Lambda), GCP, Azure, CI/CD, Kubernetes, Git, GitHub Actions, Jenkins

Python Ecosystem: Pandas, NumPy, pytest, Selenium, Pydantic, asyncio, FastMCP, uv, Jupyter, Matplotlib, Seaborn

Other Tools: Tailwind CSS, Bootstrap, Webpack, Babel, Postman, VS Code, Linux/Unix

PROFESSIONAL EXPERIENCE

Software Engineer, iTechCloud Solution Pvt. Ltd., Surat, Gujarat **January 2023 — July 2023**

- Rapidly progressed from Intern to Software Engineer, engineering enterprise-grade Salesforce workflows and integrating complex business logic for **3+ enterprise clients**
- Architected and deployed an internal quiz application using **Lightning Web Components (LWC)** and **Apex** to streamline the hiring process, reducing candidate evaluation time by **40%**
- Developed complex automation flows and **Apex triggers** with validation logic, eliminating **90% of manual operations** and ensuring data integrity across multiple business processes
- Designed and implemented responsive user interfaces using **Lightning Web Components (LWC)**, ensuring a seamless and intuitive user experience aligned with **Lightning Design System** standards
- Optimized data retrieval efficiency by **30%** through advanced **SOQL query tuning**, strategic indexing, and governor limit optimization, ensuring high performance for large datasets
- Authored and executed **50+ comprehensive test cases** using the Apex testing framework, achieving **95% code coverage** with zero critical production bugs and full compliance with deployment standards
- Implemented **REST API integrations** to connect Salesforce with external systems, enabling real-time data synchronization and workflow automation

Software Developer, Infiycube Solution, Surat, Gujarat **June 2021 — December 2022**

- Collaborated on project-basis to develop and deploy multiple web applications with API integrations, gaining foundational experience in full-stack development and API design patterns
- Engineered dynamic and responsive user interfaces using **React.js** and **Next.js**, principles later applied to Lightning Web Components development
- Implemented robust data architectures utilizing both **SQL (PostgreSQL)** and **NoSQL (MongoDB)** databases, understanding translatable to Salesforce data modeling and SOQL optimization

- Managed the complete **software development lifecycle** for client projects, from requirements gathering to deployment, maintenance, and post-deployment support
- Built **RESTful APIs** and integrated third-party services, establishing strong foundation for Salesforce external integrations

PERSONAL PROJECTS

Advanced RAG Chatbot System, [GitHub Repository](#)

- Engineered an **advanced Retrieval-Augmented Generation (RAG) chatbot** using **LangChain, LangGraph, and ChromaDB**, implementing sophisticated retrieval strategies for enhanced context accuracy and response quality
- Implemented **context padding technique** where vectors within similarity threshold trigger retrieval of surrounding contextual chunks (before/after), providing richer context to the LLM and improving answer coherence by **35%**
- Developed **multi-stage retrieval pipeline** incorporating **semantic search, re-ranking algorithms, and hybrid retrieval** strategies combining dense and sparse embeddings for optimal document retrieval
- Architected **stateful conversation flow** using **LangGraph**, managing conversation memory, context windows, and implementing **retrieval optimization techniques** including MMR (Maximal Marginal Relevance) and query expansion
- Integrated **vector database optimization** strategies with ChromaDB including custom distance metrics, metadata filtering, and efficient indexing to handle large-scale document collections with sub-second query times
- Built comprehensive evaluation framework measuring retrieval accuracy, response relevance, and latency, iteratively improving RAG performance through experimentation with chunk sizes, overlap strategies, and embedding models

Google Sheets MCP Server, [PyPI Package](#)

- Built an **AI-powered MCP (Model Context Protocol) server** enabling **LLM agents** (Claude, ChatGPT, Gemini) to autonomously interact with Google Sheets through **25+ tool-calling operations**
- Designed and implemented **agent orchestration system** using **Python, FastMCP, and Pydantic**, featuring **function calling**, schema validation, and **context management** for reliable agent interactions
- Engineered **production-grade agentic infrastructure** with automated **CI/CD pipelines**, comprehensive testing, and deployed to **PyPI**, achieving **2,000+ downloads** from the AI/ML community
- Implemented secure **Service Account authentication** and error handling to ensure robust **agent-to-API** communication in production environments

Apple Notes MCP Server, [PyPI Package](#)

- Developed a local **AI agent integration server** connecting **LLM-based assistants** with Apple Notes through **14 autonomous tool operations** for intelligent note management
- Architected **3-layer microservices system** using **Python, FastMCP, Pydantic, and AppleScript**, implementing **agent tool-use patterns** with comprehensive validation and error recovery
- Published production-ready package to **PyPI** with **automated CI/CD**, achieving **800+ downloads** and enabling developers to build **agentic AI applications** with local data integration
- Implemented **context-aware agent interactions** ensuring reliable execution of complex multi-step note operations with proper state management

QuantBot – Autonomous Trading System, [GitHub Repository](#)

- Engineered an **autonomous decision-making system** for quantitative trading, processing **real-time market data streams** with **sub-second latency** using **event-driven architecture**
- Implemented **algorithmic trading strategies** using **Python, Pandas, NumPy, and TA-Lib**, incorporating **technical indicators, pattern recognition, and signal generation** for automated trading decisions
- Built **fully autonomous agent** operating **24/7**, managing complete trading lifecycle from **data ingestion → feature engineering → decision logic → execution** with minimal human intervention
- Integrated **risk management algorithms** including stop-loss calculation, position sizing, and portfolio optimization to ensure safe and compliant autonomous operations