

HARSHA VARDHAN YELLELA

United States | +1-248-497-9965 | harsha.yellela@gmail.com | har5ha.in | [LinkedIn](#) | [GitHub](#)

SUMMARY

AI Automation Engineer with hands-on experience building **multi-agent systems**, **MCP-based tool integrations**, and **LLM-powered automation pipelines** for production environments. Designed agentic workflows using **CrewAI**, **LangChain**, and **n8n** with **Amazon Bedrock** and **function calling**, achieving **70% reduction in manual processes**. Built **94 serverless functions** integrating **10+ third-party APIs** and **8-microservice MLOps platforms** with auto-retraining loops. Skilled in **RAG pipelines**, **structured output generation**, and **end-to-end automation** from data ingestion to deployment.

EXPERIENCE

Graduate Research Assistant – Agentic AI <i>Lawrence Technological University</i>	Jan 2025 – Dec 2025 <i>Southfield, MI</i>
<ul style="list-style-type: none">Built and benchmarked no-code (n8n) vs. coded multi-agent systems (CrewAI + LangChain) with MCP tool integration for automated research workflows and intelligent decision-making.Deployed persistent MCP agent services on AWS Fargate and Amazon EKS, integrating OpenSearch Serverless for semantic search and RAG-based automation.Designed hybrid pipelines combining Amazon Bedrock foundation models with function calling and custom tool integrations, achieving 70% reduction in manual process time.	
Infor India Pvt. Ltd. <i>LN Technical Consultant</i>	Apr 2022 – Dec 2023 <i>Hyderabad, India</i>
<ul style="list-style-type: none">Developed modular, production-ready automation tools for global clients (Ferrari, Boeing, Triumph) by extending Infor LN ERP workflows with custom business logic.Integrated Infor ION process flows with AWS S3, Lambda, and API Gateway for asynchronous file transfer and event-driven automation, reducing batch processing delays by ~40%.Containerized business logic services using Docker and orchestrated with Infor ION for automated enterprise workflows across manufacturing and finance modules.	

TECHNICAL SKILLS

<ul style="list-style-type: none">Languages: Python, Go, TypeScript, JavaScript, SQL, BashAI Agents & Orchestration: CrewAI, LangChain, LangGraph, n8n MCP (Model Context Protocol) Function Calling, Structured OutputLLMs & Platforms: GPT-4, Gemini, Claude API, Llama, Qwen Amazon Bedrock QLoRA Fine-tuning (PEFT) Ollama, vLLMRAG & Vector DBs: Qdrant, pgvector, OpenSearch Embeddings, BM25, Cross-Encoder Reranking Semantic Search PipelinesCloud & DevOps: AWS (Lambda, Fargate, EKS, SageMaker, API Gateway) Docker, Kubernetes Terraform, GitHub ActionsBackend & Integration: FastAPI, Flask, Gin (Go) Stripe, Twilio, DocuSign, QuickBooks Event-Driven Architectures	
---	--

PROJECTS

AgenticAI – Multi-Agent Newsletter Automation <i>CrewAI, LangChain, n8n, FastAPI, AWS Lambda</i> GitHub	Apr 2025
<ul style="list-style-type: none">Designed and deployed newsletter automation using n8n (no-code) and CrewAI + LangChain (code-first) with structured output parsing, benchmarking both across latency, reliability, and developer speed.Built multi-agent coordination with function calling, custom FastAPI backends, and deployed on AWS Lambda for scalable, event-driven execution.Achieved 70% reduction in manual workload and provided actionable insights into visual vs. programmatic agent orchestration tradeoffs.	
CRE Research Agent – AI-Powered Literature Review <i>LangGraph, FastAPI, vLLM, pgvector, Qdrant, Elasticsearch, Redis</i>	Nov 2024 – Present
<ul style="list-style-type: none">Built automated literature review system using multi-step RAG pipeline: query expansion → retrieval → reranking → summarization, discovering ~20 relevant papers per query.Implemented LangGraph-style state machine orchestration with dual vector databases (Qdrant + pgvector), BM25 sparse retrieval, and BGE cross-encoder reranking.Integrated Semantic Scholar, ArXiv, and Crossref APIs for automated paper discovery with inline citation tracking and bibliography generation.	
N8N Workflow Automation – Self-Hosted AI Platform <i>n8n, OpenAI, SerpAPI, Tavily AI, Ollama, Qdrant, Docker</i> GitHub	Sep 2024
<ul style="list-style-type: none">Built newsletter automation workflow with AI agents integrating SerpAPI for trending topic discovery and Tavily AI for research.Deployed self-hosted AI starter kit with Docker Compose, configuring Qdrant vector database for semantic search and Ollama (Llama 3.2) for local LLM inference.Designed end-to-end automated pipelines from topic discovery to structured content generation with zero manual intervention.	

Lambda Microservices – Enterprise Serverless Integration Platform

2024

AWS Lambda, DynamoDB, API Gateway, Terraform, CloudFormation | Stripe, Twilio, DocuSign, QuickBooks

- Developed **94 AWS Lambda functions** for a complete SaaS platform with **automated deployment pipelines** using Git change detection.
- Integrated **10+ third-party services**: **Stripe Connect** (vendor onboarding/payments), **DocuSign** (e-signature workflows), **Twilio** (communications), **QuickBooks** (accounting sync), **EagleView** (aerial imagery).
- Built **event-driven architectures** with **API Gateway** routing and **Terraform/CloudFormation IaC** for repeatable infrastructure provisioning.

ML Sentiment Feedback Loop – Automated MLOps Platform

Nov 2024 – Present

AWS (ECS Fargate, SageMaker, S3), Terraform, GitHub Actions, Docker | GitHub

- Built **8-microservice MLOps architecture** with **automated feedback loops**: Inference → Feedback → Evaluation → Retraining → Deployment, requiring **zero manual intervention**.
- Implemented **auto-retraining triggers** with **SageMaker**, automated **model versioning/registry**, and **SNS-based notifications** for pipeline events.
- Configured **GitHub Actions CI/CD** with **Terraform IaC** for fully automated infrastructure provisioning and container deployments to **ECS Fargate**.

AI Resume Builder (Resumade.in) – Multi-LLM Automation

Jun 2025 – Present

FastAPI, AWS Lambda, DynamoDB, GPT-4, Gemini, DeepSeek, GitHub Actions

- Architected **serverless automation pipeline** using **FastAPI + AWS Lambda** with **multi-LLM orchestration** (GPT-4, Gemini, DeepSeek) and **structured JSON output** for intelligent resume generation.
- Built **real-time web scraping + AI optimization pipelines** using **BeautifulSoup** and **LinkedIn parsers**, reducing manual resume tailoring effort by **80%**.
- Deployed with **95%+ uptime**, **sub-2s PDF/Word generation**, and **GitHub Actions CI/CD** for automated testing and deployment.

EDUCATION

Lawrence Technological University

Dec 2025

Master of Science in Computer Science · GPA: 3.6/4.0

Southfield, MI

- Relevant Coursework: **Agentic AI Research, Machine Learning, Natural Language Processing, Artificial Intelligence, Intelligent Robotics (ROS)**

Geethanjali College of Engineering & Technology

Aug 2022

Bachelor of Technology in Computer Science & Engineering · GPA: 7.5/10 (~3.0/4.0)

Hyderabad, India

- Relevant Coursework: **Deep Learning & Python, Machine Learning Foundations, Software Engineering, Internet of Things**

ACHIEVEMENTS

- Selected for **Amazon Nova AI Challenge: Trusted AI Track** (2025)
- Built production SaaS platform with 94 Lambda functions and 10+ third-party integrations serving real customers (2024)
- Participated in **RSNA Pneumonia Detection Challenge**; ranked in upper quartile with VGG19 transfer learning model (2024)
- Gold Medalist in Indian National Mathematical Olympiad (INMO) (2012)