

# S P DEVANANDHAN

AI / ML Engineer · LLM & Generative AI Systems

Bengaluru, India | +91 96330 20011 | devansp2020@gmail.com

linkedin.com/in/devanandhansp | github.com/Devan021 | hollerlabs.org

## PROFESSIONAL SUMMARY

AI/ML Engineer with 2+ years building production LLM, GenAI, and multi-agent systems — 40+ deployments across healthcare, insurance, SaaS, real-estate, and cybersecurity domains. Strong in LLM orchestration, RAG pipelines, conversational AI / chatbots, image processing (OCR), and FastAPI microservices deployed via cloud AI services. Apply software development best practices — version control, CI/CD, structured logging — across every deployment. Published AI researcher (ICCCNT 2025, IEEE / IIT Indore) on healthcare AI compliance and evaluation. Seeking to contribute to Accenture's Applied AI/ML Engineering team.

## CORE SKILLS & TECHNOLOGIES

**LLMs & Generative AI:** GPT (OpenAI), Claude (Anthropic), prompt engineering, fine-tuning, function-calling, multi-agent orchestration, tool use, memory systems, guardrails, structured outputs, LangChain patterns. Production-grade AI tools for enterprise use cases.

**Retrieval & NLP:** RAG architectures, FAISS, vector databases, embeddings, semantic search, intent classification, entity extraction, conversational AI / chatbots, OCR/ASR integration.

**Image Processing & Document AI:** OCR pipelines for document image extraction (insurance policies, leasing docs, claims), validation and anomaly detection on extracted fields, structured-data parsing from scanned and digital images.

**Deep Learning & ML:** Neural networks, transformer-based models, evaluation methodology (precision/recall/F1, A/B testing), exposure to PyTorch / TensorFlow ecosystems, ML lifecycle management.

**Data Preprocessing & Feature Engineering:** ETL pipelines, text normalization, document chunking, embedding generation, feature extraction for downstream LLM and classification tasks, data validation and quality checks.

**Cloud & Deployment:** Cloud AI services (OpenAI, Anthropic, AWS), Docker containerization, REST APIs, CI/CD, production-grade pipeline design, on-prem and cloud deployment with structured logging and monitoring.

**Software Development Best Practices:** FastAPI, async microservices, version control (Git), code review, fault tolerance, retry logic, drift detection, audit trails, regression testing, modular architecture.

**Data & Infra:** SQL, MySQL, ETL pipelines, JSON/CSV processing, workflow orchestration (n8n), GraphQL, Linux.

**Languages:** Python (primary), JavaScript / Node.js, Bash, SQL.

## EXPERIENCE

### HollerLabs — Co-Founder & AI/ML Engineer

Jun 2024 – Present

Bengaluru, India

AI Automation & GenAI Solutions Studio

- Designed and shipped **40+ production AI systems** — LLM agents, RAG pipelines, conversational AI bots, OCR/ASR workflows, and multi-step automation — for clients across healthcare, insurance, SaaS, real-estate, and cybersecurity; delivered ₹22L+ (\$27K+) in revenue across direct contracts and Upwork.
- Architected **multi-agent orchestration systems** integrating LLMs (GPT, Claude) with external APIs, tools, and memory stores for autonomous classification, decisioning, and task execution at production scale.
- Built **FastAPI microservices** with Dockerized deployment, structured logging, retry logic, and monitoring; implemented A/B evaluation and drift detection for LLM outputs in production environments.
- Implemented **RAG pipelines** using FAISS and vector retrieval over proprietary client corpora, improving response grounding and reducing hallucination on domain-specific queries by measurable margins.
- Established evaluation rigor — precision/recall/F1, regression testing, audit trails — across LLM deployments to support compliance-aware enterprise use cases.

### Srishti Innovative — Full-Stack + QA / ML Intern

Jun 2024 – Nov 2024

Technopark, India

- Built automation and regression-verification pipelines for ML model outputs using Python and MySQL; contributed to evaluation workflows, API testing, and reproducible model performance tracking.

### **Traboda CyberLabs — Full-Stack Developer**

Jan 2023 – Jun 2023

Amritapuri, India

- Built React / Next.js dashboards and backend integrations for cybersecurity scoring systems; developed API-driven applications and validated structured data pipelines.

## **SELECTED PRODUCTION PROJECTS**

---

### **Healthcare AI System — IEEE Research Publication** — *Three-model architecture for clinical compliance & cost-effective inference*

- Designed and evaluated a three-model AI architecture for healthcare integrating real-time medical research with cost-aware clinical compliance — published at ICCCNT 2025 (IEEE / IIT Indore).
- Contributed to ML workflow design, probability-based evaluation, and compliance-aware inference systems with audit-ready outputs.

### **Voice & Conversational AI Agents — Multi-client** — *ASR → NLU → RAG with fallback handling*

- Built voice and chat agents integrating ASR, NLU, and LLM pipelines with semantic retrieval (FAISS), dynamic prompting, conversational memory, and graceful fallback paths.
- One deployment ran 800+ outbound calls in week one with measurable conversion to qualified appointments — demonstrating real-world reliability of LLM-driven conversational systems.

### **Insurance AI Pipelines — StellaInsurance** — *Compliance-aware OCR + NLU for policy and claims*

- Built OCR → NLU → rule-based pipelines for policy extraction and claims triage, with validation, anomaly detection, and compliance-aware audit trails for every inference.

### **Lead Intelligence Automation — Pug.ai** — *LLM classification + scoring at scale*

- Built AI-driven ICP verification, lead scoring, segmentation, and enrichment combining LLM classification with rule-based logic; designed orchestration with retries and monitoring.

### **OCR Intelligence Platform — Mastracorp** — *Document extraction at subscription scale*

- Built OCR extraction and parsing workflows for property and leasing documents; implemented validation rules and reusable microservices that power a subscription product.

## **EDUCATION & RESEARCH**

---

### **B.Tech, Computer Science — Amrita Vishwa Vidyapeetham**

2021 – 2025

16 years of full-time formal education (10 + 2 + 4).

**Research Publication — ICCCNT 2025 (IEEE / IIT Indore):** "Three-Model Architecture for Optimized Healthcare AI Integrating Real-Time Medical Research and Cost-Effective Clinical Compliance." Contributed to ML workflow design, probability-based evaluation, and compliance-aware inference systems.