SWAROOP CHAND J C

AI PLATFORM & AGENT SYSTEMS ENGINEER

Bengaluru, India | swaroopchandrashekar@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

AI platform and machine learning engineer with 2+ years of experience building production-grade enterprise RAG, agentic workflows, knowledge-ingestion systems, and cloud-native ML services. At ResMed, built and hardened core components of a self-service conversational AI platform used to create 150+ assistants, including 2 customer-facing deployments and 10+ active internal bots, reducing end-user deployment time from weeks to minutes. Hands-on across hybrid retrieval, agent registries, MCP, LLMOps, FastAPI microservices, Kubernetes, AWS, OpenSearch, PostgreSQL/pgvector, evaluation, observability, and production reliability.

CORE EXPERTISE

AI Platforms & Agents: RAG, Agentic AI, Multi-Agent Systems, Agent Registry, MCP, LLMOps, LangGraph, LangChain, AI Evaluation, Human-in-the-Loop Workflows

Retrieval & Knowledge Systems: Hybrid Search, Dense/Sparse Retrieval, Cross-Encoder Reranking, Semantic Chunking, Metadata Filtering, OpenSearch, pgvector, Vector Databases, Multimodal Ingestion

Cloud & Software Engineering: Python, FastAPI, REST APIs, Microservices, Docker, Kubernetes, AWS Bedrock, S3, RDS/PostgreSQL, Redis, GitHub Actions, OpenTelemetry, TrueFoundry

Machine Learning: PyTorch, TensorFlow, Transformers, CNNs, RNNs, XGBoost, Random Forest, H2O AutoML, Scikit-learn, Pandas, NumPy

PROFESSIONAL EXPERIENCE

RESMED INDIA | Bengaluru, India **Jun 2025 - Present**

Associate Machine Learning Engineer - AI Platform & Agent Systems **May 2026 - Present**

- Implemented an Agent Hub and AWS AgentCore-backed registry lifecycle for agent registration, discovery, updates, retirement, and endpoint-based execution, enabling independently deployed agents to run through a unified platform while preserving built-in execution paths.
- Built and deployed a standalone FastAPI/ReAct web-research agent with SSE execution traces, API-key authentication, and web-search/page-reading tools; diagnosed cluster-egress blocking and added a three-backend fallback that restored grounded, source-cited responses.
- Developed a BMAD-based multi-agent pilot spanning planning, development, review, deployment, and incident RCA; integrated multi-model routing, codebase context, Jira workflows, partial-failure handling, conflict detection, and ranked root-cause hypotheses.
- Consolidated approximately 14 onboarding capabilities into a unified FastAPI backend and validated production workflows across Okta authentication, RAG chat, feedback, notifications, FAQ approvals, access requests, and re-ingestion.
- Refactored an internal ResNet18 leak-detection application into independent inference and review services with S3-first staging, IRSA, health/metrics endpoints, and failure-safe image retention; deployed and validated the flow in nonproduction.

Data Scientist - Enterprise RAG & Knowledge Platforms **Jun 2025 - May 2026**

- Built and hardened core components of a self-service enterprise RAG platform used to create 150+ assistants, including 2 customer-facing deployments and 10+ active internal bots; reduced end-user chatbot deployment time from weeks to minutes.
- Engineered hybrid retrieval using dense/sparse search, cross-encoder reranking, semantic chunking, per-source Top-K controls, metadata filters, and configurable vector-distance strategies.
- Expanded enterprise knowledge coverage across Confluence, SharePoint, GitHub, Workday, PDFs, tables, images, and videos through source-specific and multimodal ingestion pipelines.
- Reduced document-onboarding time by 31-40% in measured PDF workloads through bounded parallel processing and embedding-pipeline optimization (5 documents: 10 to 6 minutes; 40 documents: 118 to 81 minutes).
- Built an OAuth2 Workday connector for approximately 1,145 help articles with pagination and incremental synchronization.
- Added SHA-256 change detection and stale-record deletion to keep vector stores aligned with new, updated, and removed content.
- Implemented GitHub ingestion with GitHub Actions, repository-isolated S3 paths, and deletion propagation.
- Improved Confluence ingestion through recursive page traversal, structured table extraction, and source-URL traceability.
- Delivered a human-in-the-loop FAQ correction workflow with bot-owner approvals, idempotent queueing, S3 persistence, automatic re-embedding, OpenSearch updates, and owner notifications.
- Hardened the workflow in nonproduction by resolving schema, environment, duplicate-document, and deployment-alignment issues.

WATTWORKER INC. | Machine Learning Research and Developer **2023 - 2024**

- Developed RNN, CNN, XGBoost, Random Forest, and H2O AutoML models for HVAC power-consumption forecasting using noisy sensor data from 3 commercial buildings across the UK and US; improved reported model performance from 78% to 94% through hyperparameter optimization.
- Built physics-informed thermodynamic models using physical equations and curve fitting to estimate indoor-temperature dynamics, and presented control and operational insights to stakeholders.
- Developed supervised-learning models for EV battery remaining useful life and range prediction, and analyzed GPS telemetry for route, trip, and operational patterns.

Earlier Experience: Intern, Karunadu Technologies Pvt. Ltd. | Aug 2022 - Sep 2022

SELECTED PROJECT

On-Device Multi-Agent Assistant | Apple MLX, LangGraph, MCP

- Built a fully local ReAct-based assistant with 9 specialized agents orchestrated through LangGraph.
- Integrated Apple MLX and MCP tools for private on-device inference without cloud dependency.

EDUCATION

Master of Engineering (M.Eng.), Artificial Intelligence and Machine Learning **2026**
Manipal School of Information Sciences, MAHE | CGPA: 9.08

Bachelor of Engineering (B.E.) **2023**
Sapthagiri College of Engineering, Bengaluru