HEMANT SUDARSHAN

Bengaluru, India

Email: hemanthsudarshan2002@gmail.com | collabwithhemantgenai@gmail.com

Phone: +91-8985837147

linkedin.com/in/hemant-sudarshan-01633928a | github.com/HemantSudarshan | huggingface.co/Hemantxai

CAREER OBJECTIVE

Recent Computer Science graduate passionate about AI/ML engineering, seeking entry-level opportunities to apply hands-on experience in Large Language Models, RAG systems, and production AI deployments. Built 6+ AI projects including live deployed systems with real users. Eager to contribute to innovative AI teams while continuing to learn and grow in areas like LLMOps, multi-agent systems, and scalable AI infrastructure. Strong academic foundation with 1 patent and 2 research publications.

EDUCATION

Bachelor of Technology in Computer Science and Technology

Dayananda Sagar University, Bengaluru

2021 - 2025 | First Class

Relevant Coursework: Machine Learning, Deep Learning, Natural Language Processing, Data Structures and Algorithms, Database Management Systems, Operating Systems, Computer Networks

Higher Secondary Certificate

Narayana Junior College | CGPA: 9.12/10

TECHNICAL SKILLS

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

AI/ML Frameworks and Libraries: PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, LangChain, LlamaIndex, LangGraph

Large Language Models: GPT-4, Claude, Llama, Gemini, Groq, Prompt Engineering

Vector Databases and RAG: Weaviate, Qdrant, Pinecone, ChromaDB, Retrieval-Augmented Generation

Web Development and APIs: FastAPI, Flask, REST APIs, Streamlit, Gradio, React

Databases: MongoDB, PostgreSQL, SQLite, Redis, SQLAlchemy

DevOps and Tools: Docker, Git, GitHub, GitHub Actions, CI/CD, AWS

Specialized Skills: Natural Language Processing, Computer Vision, Audio Processing, Multimodal AI, Indic Language NLP (Hindi, Telugu, Kannada)

PROFESSIONAL EXPERIENCE

AI Operations and Localization Consultant

Pratilipi Comics, Bengaluru

August 2025 - Present

- Working on AI-driven localization initiatives for Indic language content (Hindi, Telugu, Kannada)
- Building GenAI tools using GPT-4 and Claude for automated content translation and localization
- Implementing prompt engineering techniques to improve translation quality and context awareness
- Contributing to workflow optimization that reduced manual translation time by 70 percent
- Learning and applying LLMOps best practices including prompt versioning and quality evaluation

Localization Operations and GenAI Intern

Pratilipi Comics, Bengaluru

February 2025 - August 2025

- Developed GenAI tools for multi-language content localization and translation
- Applied prompt engineering techniques to improve translation accuracy for domain-specific content
- Collaborated with cross-functional teams to integrate AI tools into existing workflows
- Helped reduce translation time significantly through automation

Machine Learning Intern

AppiVa Software Private Limited, Bengaluru

August 2023 - November 2023

- Built data preprocessing and feature engineering pipelines for machine learning projects
- Developed machine learning models and performed hyperparameter tuning
- Gained hands-on experience with model development, testing, and deployment workflows

KEY PROJECTS

Agentic Inventory Restocking Service - Multi-Agent AI System

Technologies: LangGraph, FastAPI, MongoDB, Gemini AI, Groq, Redis, Telegram Bot API, Docker

GitHub: github.com/HemantSudarshan/Agentic-Inventory-Restocking-Service

Live Demo: agentic-inventory-restocking-servic-production.up.railway.app

- Built autonomous AI agent system that analyzes inventory data and recommends restocking orders
- Implemented 3-step workflow using LangGraph: Data Retrieval, AI Reasoning, Action Generation
- Integrated dual LLM model (Gemini and Groq) with automatic fallback for reliability
- Developed confidence scoring system (0-100 percent) with human-in-loop approval via Telegram bot
- Created multi-channel notification system (Telegram, Slack, Web Dashboard) for real-time alerts
- Deployed using Docker with FastAPI backend, MongoDB database, and Redis caching
- Applied rate limiting and authentication for production-ready deployment

Compliance-GPT - AI-Powered Compliance Assistant with RAG

Technologies: FastAPI, Weaviate Vector Database, Groq LLM, Docker, BM25, GitHub Actions

GitHub: github.com/HemantSudarshan/Compliance-GPT

Live Demo: huggingface.co/spaces/Hemantxai/compliance-gpt

- Built RAG (Retrieval-Augmented Generation) system for compliance questions on GDPR, CCPA, PCI-DSS
- Implemented hybrid search using BM25 keyword matching and semantic search with 1987+ indexed chunks
- Developed query expansion feature to improve search results by generating related terms
- Ensured zero-hallucination by providing source citations with filename and page number for every answer
- Integrated security features including rate limiting (30 requests per minute), authentication, and CORS protection
- Added web search fallback using DuckDuckGo API for queries not covered in local knowledge base
- Deployed containerized application using Docker with FastAPI backend and Weaviate vector database
- Implemented CI/CD pipeline with GitHub Actions and 80+ automated tests

AudioRAG Enterprise - Audio Intelligence Platform with RAG

Technologies: AssemblyAI, Qdrant Vector Database, SambaNova LLM, Redis, PostgreSQL, FastAPI, Docker Compose

GitHub: github.com/HemantSudarshan/AudioRaG_FTSambanova

- Developed audio analysis platform with transcription, speaker identification, and conversational search
- Integrated AssemblyAI API for speech-to-text transcription with speaker diarization
- Built RAG pipeline: audio to transcription to text chunking to embeddings to vector storage in Qdrant
- Implemented semantic search over audio content using BGE-Large embeddings (1024 dimensions)
- Added SambaNova LLM for generating contextual answers from retrieved audio chunks
- Created multi-tenant system with JWT authentication and role-based access control

- Used 3-layer architecture: Redis for caching, PostgreSQL for metadata and audit logs, Qdrant for vector search
- Deployed with Docker Compose orchestrating multiple services (FastAPI, Qdrant, Redis, PostgreSQL)

ACHIEVEMENTS

- Deployed 6+ AI projects with live demos on platforms like Hugging Face and Railway.app
- Built RAG systems for Indic language localization serving users across Hindi, Telugu, and Kannada
- Reduced manual work time from 20+ minutes to 2 seconds in compliance research through automation
- Built multi-agent systems with autonomous decision-making capabilities using LangGraph
- Implemented production-ready features: authentication, rate limiting, caching, error handling
- Optimized API costs by 70 percent through intelligent caching strategies

PATENTS AND PUBLICATIONS

Patent Filed:

A System for Providing Security Using a Plurality of Factors for IoT Gadgets
Indian Patent Office, Application No. 202341040746

Conference Paper:

Discovering Insights into Heart Health: A Survey of Data Mining and Machine Learning Methods

Presented at International Conference on Computer and Intelligent Communication Technologies (ICCICCT 2023), NICHE

Research Paper:

An Examination of Different Deep Learning Models for Transparent Identification of Tomato Leaf Diseases

Focus: Computer Vision, Agricultural AI, Transfer Learning

LANGUAGES

English (Fluent), Hindi (Fluent), Telugu (Fluent), Kannada (Fluent), Japanese (Intermediate)