DHANUSH G

Chennai, India |6381894626|dhanushgd0801@gmail.com|http://www.linkedin.com/in/d0801

PROFILE

AI/ML & Generative AI Engineer | I specialize in designing and implementing scalable AI solutions that leverage cutting-edge machine learning, large language models, and agent-based systems to address complex, real-world challenges. My work focuses on building intelligent automation tools that streamline workflows, predict outcomes, and improve user experiences through conversational interfaces and generative applications. With a passion for transforming raw data into actionable insights, I bridge the gap between data, automation, and decision-making. Driven by the intersection of language, logic, and learning, I strive to develop AI systems that not only automate tasks but also empower businesses with intelligent, data-driven solutions.

EDUCATION

Master of Technology in Software Engineering — June 2024
VIT, Vellore,

SKILLS

- **Programming Languages:** Python
- **API Development:** FastAPI, REST APIs
- **Machine Learning & AI:**
  - **ML Libraries:** Scikit-learn
  - **Deep Learning & Neural Networks**: TensorFlow, Keras, ANN, RNN, LSTM, GRU, Transformers
  - **Natural Language Processing (NLP):** NLTK, spaCy, Hugging Face Transformers (e.g., BERT, DistilBERT)
- **Generative AI & Large Language Models (LLMs):**
  - **Frameworks & Tooling:** LangChain, LangSmith, LangServe, Streamlit
  - **Multi-Agent Systems:** CrewAI, LangGraph (stateful orchestration)
  - **LLMs & Inference APIs**: OpenAI, LLaMA, Google Gemma, Groq, Ollama
  - **Embeddings & Retrieval:** SentenceTransformers (all-MiniLM-L6-v2), FAISS, ChromaDB
  - **RAG Techniques:** Retrieval-Augmented Generation (RAG), Hybrid RAG, Conversational RAG
  - **Fine-Tuning:** LoRA, QLoRA, PEFT, Unsloth, Lamini
  - **Graph Database:** Neo4j, Cypher Query Language
- **Model Deployment:** AWS SageMaker, AWS Bedrock, API Gateway, Docker
- **DevOps & Cloud Infrastructure:** AWS Lambda, EC2, S3, GitHub Actions
- **Version Control & Collaboration:** Git, GitHub, Jira
- **Authentication & Authorization:** OAuth2, JWT (JSON Web Token)

WORK EXPERIENCE

PGET - Renault Nissan Technology and Business Centre India — January 2024 - Present
AI/ML and Generative AI Engineer

I specialize in developing innovative AI applications, focusing on Generative AI, Large Language Models (LLMs), and intelligent automation systems. My work involves creating end-to-end solutions that leverage advanced machine learning techniques to automate processes, enhance decision-making, and improve user interactions. I collaborate with cross-functional teams to build scalable, production-ready AI tools that drive business efficiency and deliver impactful, data-driven outcomes. By continuously exploring new AI advancements, I aim to deliver cutting-edge applications that push the boundaries of intelligent automation.

CERTIFCATES

- **Internship Certificate -** Renault Nissan Technology and Business Centre.

PROJECTS

SupportMind – AI-Powered IT Support Automation System

- Designed and deployed a **stateful multi-agent** IT support system using **LangGraph** to automate ticket classification, document retrieval, and troubleshooting workflows.
- Fine-tuned the **Google Gemma 2B** model with **LoRA/QLoRA and Unsloth** for accurate IT ticket **achieving 88% accuracy**, classification, integrating **FAISS** with **MiniLM embeddings** for context-aware document retrieval.
- Integrated the Groq API with **LLaMA 3.3** to enable real-time, ultra-fast troubleshooting and automated decision-making , Developed an interactive **Streamlit** UI to facilitate seamless IT support interaction, feedback collection, and enhanced user experience.

Multi-Agent IT Assistant with Hybrid Search RAG

- Developed a Multi-Agent IT Assistant using **CrewAI** and **GEMMA 2**, incorporating **hybrid search RAG** (semantic + keyword retrieval) for answering IT queries from internal documents like SOPs and troubleshooting guides.
- Implemented **FAISS + HuggingFace Embeddings (MiniLM)** to enhance document retrieval accuracy and integrated fallback support from Wikipedia for unanswered queries.
- Leveraged **Streamlit** for a user-friendly interface, enabling PDF uploads, query entry, and API key authentication, facilitating seamless interaction with the system.
- Built with **LangChain** to manage the RAG pipeline, coordinating multiple **AI agents** for document ingestion, retrieval, answer generation, and validation, automating IT ops workflows.

Serverless Blog Generator with Meta's LLaMA 3 on AWS

- Built a serverless blog generation system using Meta's **LLaMA 3** (70B Instruct) via **AWS Bedrock** to generate blog posts from topic inputs, fully leveraging AWS-native services.
- Implemented **AWS Lambda** for running the blog generation logic, **AWS API Gateway** for handling topic input, and **Amazon S3** for storing the generated blog posts in text format.
- Enabled scalable, cost-efficient blog generation using a fully **serverless architecture** (no EC2/containers) and ensured smooth testing with **Postman** for rapid local API validation.

Jira Bug Status Prediction and Recommendation System

- Developed a deep learning model using an **Artificial Neural Network (ANN)** built with **Tensorflow & Keras**, achieving an **87% accuracy** in predicting Jira bug statuses based on bug descriptions.
- Implemented smart recommendations leveraging **TF-IDF and cosine similarity** to suggest similar bugs and recommend fixes, enhancing debugging efficiency.
- Integrated **Jira** and **GitLab**, scraping GitLab links from Jira tickets and fetching commit details, code diffs, and file changes to provide comprehensive bug insights.
- Deployed a **Dockerized** web app within the company's private network using **Flask**, providing a user-friendly interface and ensuring consistent deployment across environments.

Breast Cancer Prediction using Machine Learning

- Developed an AI diagnostic tool using the Breast Cancer Wisconsin dataset; implemented multiple classifiers including **KNN, SVC, Decision Trees, Logistic Regression, and an ANN (91% accuracy**) which outperformed others. Applied **GridSearchCV** for tuning and evaluated models using **ROC-AUC**, **confusion matrices, and classification reports**.
- Built a **Flask** web app for model selection, result visualization, and downloads; integrated **ensemble learning** (Voting Classifier) and streamlined preprocessing with **StandardScaler** and **Label Encoding**.