



SREEVALLI MANDA

sreevalli.manda@gmail.com | +91 9010309688

Summary

AI/ML Engineer with 7 years of total experience, including 5 years specializing in Generative AI, Machine Learning, and NLP, with additional expertise across Avionics, Railway, and Pharma domains. Proven ability to design, develop, and deploy **scalable AI systems**, including RAG pipelines, Agentic AI frameworks, and intelligent automation workflows across cloud and enterprise environments. Strong track record of building production-grade solutions, optimizing model performance, and integrating modern MLOps practices for reliable and scalable deployments.

Highlights

- **AI & NLP Expertise:** 5 years of hands-on experience building and fine-tuning BERT-based spam detection, NER models, and QA chatbots using LLaMA & LangChain, with MLflow, DVC, and AWS SageMaker for scalable MLOps.
- **Generative AI & RAG:** Built production-ready RAG-powered chatbot APIs using AWS Bedrock, Lambda, API Gateway, and Flask with optimized latency and serverless scaling.
- **Agentic AI:** Designed end-to-end autonomous Agentic AI workflows (LangGraph, Phidata) enabling intelligent retrieval, reasoning, and multi-step task execution.
- **AI for Data Accessibility:** Built natural-language-to-SQL systems using GPT and Streamlit, improving data access for non-technical users.
- **Avionics & Radar AI:** Applied ML to avionics systems, including ANN-based fuel consumption modeling and RandomForest radar threat detection with **90%+ accuracy**.

Skills

- Generative AI
- Prompt Engineering
- AI/ML
- Data Science
- Langchain & LLMs
- MLflow
- Tensor Flow
- Python
- Scikit Learn
- AWS
- Pytorch

Career Timeline

Alten Calsoft Labs

Senior Engineer

05/2025 -Present

- Designing **Agentic AI workflows** to automate business processes based on required autonomy, scalability, and domain-specific use cases.
- Developing **iterative automation** pipelines enabling low-autonomy to high-autonomy task execution

EInfochips

Senior Engineer

06/2023 -08/2024

- Established best practices for prompt formulation by integrating **role-based, chain-of-thought, and few-shot techniques**, resulting in more coherent and contextually relevant outputs across multiple projects.
- Built and deployed a scalable RAG-powered chatbot API using Flask, AWS Bedrock, AWS Lambda, and API Gateway, ensuring serverless deployment and optimized latency.

Cyient
Software Engineer
04/2022 - 06/2023

- Designed and implemented an **Agentic AI framework** using LangGraph focusing on intelligent data retrieval, reasoning, and autonomous task execution.
- Developed and fine-tuned a **BERT-based spam detection** model for high-accuracy text classification. Utilized MLflow for experiment tracking, DVC for data versioning, and deployed on AWS SageMaker for scalable inference. Improved classification performance by leveraging transformer-based embeddings and optimized tokenization.
- Developed a **Named Entity Recognition (NER) system** using Hugging Face Transformers and a pre-trained BERT model, implementing data preprocessing, model training, and evaluation, while designing a **FastAPI** application for real-time entity recognition.
- Developed an **interactive web application** that uses OpenAI's GPT models to convert natural language queries into SQL queries, allowing users to interact with databases through conversational input, streamlining data retrieval and management for non-technical users. Integrated with MySQL and implemented in Streamlit, enhancing data accessibility and usability.
- Developed a QA chatbot using the LLaMA model from Ollama, integrated with LangChain for intelligent response generation. Built an interactive **Streamlit** app for real-time user interaction and seamless knowledge retrieval. Optimized prompt engineering and retrieval mechanisms to enhance chatbot accuracy and responsiveness.

Aeronautical Development Agency
Project Assistant-II
10/2020 - 12/2021

- Developed a **Machine learning model using a RandomForest Classifier** to classify radar signals for Radar Warning Receiver (RWR) flight data. Utilized MLflow for experiment tracking and DVC for data versioning, ensuring reproducibility and model optimization. Achieved over 90% accuracy in threat detection based on features like frequency, azimuth, elevation, and signal strength, with performance evaluated using a confusion matrix.
- Developed an ANN model to estimate aircraft fuel consumption based on factors like altitude, speed, payload, and weather conditions

Aeronautical Development Agency
Project Assistant-I
04/2018 - 09/2020

- Verification and Validation of **RADAR Warning Receiver (RWR)** software in compliance with **DO178B** (Level B) standard.
- Functional testing, HSIT and Rig testing of RWR.
- Test case preparation and Execution.
- PI Checks of LRUs and Generate of Test Reports.

Education

PhD (Pursuing), Artificial Intelligence
National Institute of Technology | Rourkela
Research Areas: Image processing and Applications

National Institute of Technology | Rourkela
Master of Technology in Signal and Image Processing

Nalanda Institute of Engineering and Technology
Bachelor of Technology in Electronics and Communication Engineering

Websites, Portfolios, Profiles

- <https://www.linkedin.com/in/sreevalli-manda-1ba707267/>
- <https://github.com/Msrevalli>

Certifications

- AI Productivity Hacks to Reimagine Your Workday and Career, LinkedIn
- Generative AI Skills for Creative Content: Opportunities, Issues, and Ethics, LinkedIn,
- Copilot in PowerPoint: From Prompt to Presentation, LinkedIn,
- Build Your Generative AI Productivity Skills with Microsoft and LinkedIn, Microsoft
- Introduction to Machine Learning, Amazon Web Services (AWS)
- Machine Learning Terminology and Process, Amazon Web Services (AWS)
- Machine Learning Plan, Amazon Web Services (AWS)
- Building Agentic AI Applications with a Problem-First Approach(Maven)