

Devi Eswar Kumar Damerla

devieswar79@gmail.com | linkedin.com/in/devi-eswar-kumar-damerla | [Portfolio](#)

PROFESSIONAL SUMMARY

Full-stack Software and AI Engineer with 5+ years of experience building secure, scalable systems across web, cloud, and machine-learning environments. Progressed from Java / Spring Boot and Angular development to architecting Retrieval-Augmented Generation (RAG) and Large Language Model (LLM) platforms using FastAPI, Weaviate, and AWS SageMaker. Skilled in transformer fine-tuning (LoRA / QLoRA), vector databases, and multimodal pipelines that fuse audio, video, and text for real-time evidence analysis. Strong foundation in microservices, containerization (Docker, ECS Fargate), and CI/CD with emphasis on model governance, compliance (CJIS), and measurable performance gains.

EXPERIENCE

AI Engineer

Nov 2024 – Present

AI for ALL LLC (AI OWL)

Columbus, OH, United States

- Lead engineer for Flash AI, a CJIS-compliant evidence intelligence platform enabling secure multimedia (video, audio, document) analysis for law-enforcement and investigative use cases. Designed scalable Retrieval-Augmented Generation (RAG) and multi-agent reasoning systems powered by Large Language Models (LLMs) such as Claude 3 and Legal-BERT.
- Developed transformer fine-tuning pipelines (LoRA / QLoRA) and parameter-efficient adaptation for legal and investigative domains, improving semantic retrieval precision and reducing model inference cost by 40%.
- Integrated Weaviate vector database with AWS SageMaker and FastAPI microservices to deliver sub-100 ms hybrid search, metadata-driven filtering, and user-isolated access control across large document collections.
- Built asynchronous multimodal ingestion pipelines using WhisperX, OpenCV, and FFmpeg for audio–video transcription and scene understanding, reducing processing time by 60% and enabling Vision-Language Model (VLM) capabilities.
- Implemented MLOps and continuous model evaluation pipelines with Docker, ECS Fargate, and SageMaker, establishing versioning, model governance, and audit compliance for production workloads.
- Mentored engineers on RAG architecture, LLM prompt optimization, and vector-database schema design; contributed to AI roadmap and strategic product integration initiatives for Flash AI and educational LLM platforms.

Software Developer

May 2025 – Oct 2025

MediTechSafe Inc

West Chester, OH, United States

- Developed a secure product development platform with interactive system architecture modeling using Cytoscape.js, featuring drag-and-drop component placement, real-time network visualization, and SysML integration for component-level assessments, attack scenario simulation, and incident response planning built with Angular, Go, and PostgreSQL.
- Built a standardized risk management platform applying Go, AI/ML, and full-stack frameworks (Node.js, React, Python) with interactive Cytoscape network diagrams, real-time risk scoring visualization, and dynamic attack scenario path highlighting to enable real-time monitoring of cyber-physical systems and secure cloud deployments for IoT data analytics.

Software Engineer

Aug 2021 – Jul 2023

Airbus Group

Bengaluru, KA, India

- Improved Aircraft 2D/3D CAD Drawing model loading times by 75% using a microservices-based iPaaS application built with NestJs and Angular, leveraging AWS load balancing and S3 for optimization.
- Developed a MEAN stack application to reduce reliance on costly CATIA licenses, cutting expenses by 60%, and utilized WebSockets for efficient model loading.
- Implemented SSO via OAuth and built comprehensive API documentation using Swagger. Contributed to Agile/Scrum and SAFe methodologies, involved in all development phases, and managed version control with Git.
- Automated design quality checks, refactored LAMP applications, and optimized SQL queries under 60 seconds for faster data processing, leading to cost savings.
- Built CI/CD pipelines with Jenkins agent in an AWS ECS Docker container, integrated SonarQube for code quality checks, and facilitated smooth data migrations from MariaDB to MySQL for deployment efficiency.

Junior Data Scientist

Aug 2019 – Jun 2021

MCR Web Solutions

Andhra Pradesh, India

- Developed and optimized convolutional neural network (CNN) architectures for image denoising, deblurring, and low-light enhancement, improving structural similarity (SSIM) scores by 83%. Experimented with hyperparameter tuning and data augmentation to enhance model generalization.

- Engineered an automated ML pipeline integrating data preprocessing, model training, and inference using Python, TensorFlow, and AWS Lambda, reducing processing latency by 60% and enabling on-demand image enhancement at scale.
- Deployed trained models as microservices via Docker and RESTful APIs, leveraging AWS infrastructure for serverless execution and scalable compute allocation.
- Performed exploratory data analysis (EDA) and statistical evaluation on large-scale image datasets using NumPy, Pandas, and Matplotlib to identify noise patterns and optimize feature extraction.

Machine Learning Engineer Intern
TheSmartBridge Private Limited

May 2019 – Jul 2019
Hyderabad, India

- Developed a 95% accurate machine learning model for chronic kidney disease diagnosis using Python, employing Naive Bayes, decision trees, and random forest to enhance model precision. Successfully integrated the model into a user-friendly Django web application.

PROJECTS

Context-Driven Image Narration | *React, Node.js, Python, TensorFlow, MongoDB, AWS* Sep 2023 – Feb 2024

- Led the development of a MERN stack application integrated with a deep learning model, achieving 79.7% accuracy in object recognition using TensorFlow and Transformers. Applied LLMs for context-based captioning and narration. Built a data pipeline for preprocessing large datasets and improved model performance using transfer learning with pre-trained models.

TECHNICAL SKILLS

Programming Languages: Python, JavaScript, TypeScript, Java, Go, C#, Shell Script, SQL

AI & Machine Learning: Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Multi-Agent Systems, LoRA / QLoRA Fine-Tuning, Prompt Engineering, HuggingFace Transformers, PyTorch, TensorFlow, Scikit-Learn, Pandas, NumPy

Data & Vector Databases: Weaviate, Pinecone, PostgreSQL, MySQL, MongoDB, DynamoDB, Redis, Elasticsearch

MLOps & Cloud Infrastructure: AWS (SageMaker, ECS Fargate, ECR, Lambda, S3, CloudFront, CloudWatch), Docker, Jenkins, CI/CD, Model Governance, Versioning, Bedrock, OpenShift, Linux/Unix

Web Frameworks: FastAPI, Node.js, Angular, React, ASP.NET Core, Spring Boot, NestJS, Express.js, Django, Vue.js, Tailwind CSS

Multimodal & Media Processing: WhisperX, OpenCV, FFmpeg, Vision-Language Models (VLM), Speech-to-Text Pipelines

Collaboration & Dev Practices: Git, GitHub Actions, SonarQube, JIRA, ClickUp, Agile/Scrum, SAFe

CERTIFICATIONS

AWS Certified Solutions Architect – Associate
Deep Learning Specialization — DeepLearning.AI