

Devi Eswar Kumar Damerla

devieswar79@gmail.com | linkedin.com/in/devi-eswar-kumar-damerla | devieswar.github.io

PROFESSIONAL SUMMARY

Full-Stack AI Engineer with 5+ years of experience architecting scalable production systems. Expert in building Agentic AI applications and end-to-end solutions using FastAPI, NestJS, and React, with deep specialization in LLMs, RAG, and Knowledge Graph architectures. Proficient in cloud-native MLOps (AWS SageMaker, Bedrock, ECS), microservices, and DevOps. Proven track record of delivering measurable business impact through technical innovation.

PROFESSIONAL EXPERIENCE

Software Engineer - AI <i>AI for ALL LLC (AI OWL)</i>	Nov 2024 – Present <i>Columbus, OH</i>
<ul style="list-style-type: none">Lead Engineer for Flash AI, a CJIS-compliant evidence intelligence platform using RAG and multi-agent systems for secure law enforcement multimedia analysis.Architected multi-agent orchestration layer using MCP (Model Context Protocol) over JSON-RPC/SSE, facilitating autonomous agent collaboration and complex escalation workflows.Designed transformer fine-tuning pipelines using LoRA and QLoRA for legal domain adaptation, boosting semantic retrieval precision by 35% while slashing inference costs by 40%.Delivered sub-100ms hybrid search across 100K+ documents by integrating Neo4j knowledge graphs and Weaviate vector indices within AWS SageMaker microservices.Orchestrated asynchronous multimodal ingestion pipelines using WhisperX, OpenCV, and FFmpeg for audio-video transcription and scene understanding, accelerating processing time by 60% and enabling Vision-Language Model capabilities.Established Terraform MLOps pipelines for ECS Fargate and SageMaker, ensuring continuous model evaluation, immutable versioning, and strict audit compliance.Mentored junior engineers on RAG architecture design, LLM prompt engineering, and vector database schema optimization; contributed to AI product roadmap and strategic integration initiatives.	
Software Engineer <i>Airbus Group</i>	Aug 2021 – Jul 2023 <i>Bengaluru, India</i>
<ul style="list-style-type: none">Boosted aircraft CAD model loading performance by 75% via a NestJS/Angular iPaaS microservice with optimized S3 caching strategies.Cut annual software costs by \$120K+ (60%) by replacing CATIA dependencies with a custom MEAN stack solution featuring WebSocket model streaming.Implemented OAuth SSO and standardized API documentation using Swagger/OpenAPI.Automated design quality checks and optimized SQL queries from 3+ minutes to under 60 seconds, accelerating data processing for 200+ engineering workflows.Configured CI/CD pipelines using Jenkins on AWS ECS Docker containers with SonarQube integration for code quality gates; executed smooth MariaDB-to-MySQL migrations with zero downtime.	
Junior Data Scientist <i>MCR Web Solutions</i>	Aug 2019 – Jun 2021 <i>Andhra Pradesh, India</i>
<ul style="list-style-type: none">Architected optimized CNN models for image denoising, deblurring, and low-light enhancement using TensorFlow, achieving 83% structural similarity (SSIM) improvement through hyperparameter tuning and data augmentation.Engineered automated ML pipeline integrating preprocessing, training, and inference with Python, TensorFlow, and AWS Lambda, cutting processing latency by 60% and enabling on-demand image enhancement at scale.Deployed models as auto-scaling serverless microservices on AWS Lambda/Docker.Identified critical noise patterns in large-scale datasets to drive feature extraction strategy.	
Machine Learning Engineer Intern <i>TheSmartBridge Private Limited</i>	May 2019 – Jul 2019 <i>Hyderabad, India</i>
<ul style="list-style-type: none">Created machine learning model for chronic kidney disease diagnosis achieving 95% accuracy using Naive Bayes, decision trees, and random forest algorithms; deployed as Django web application with user-friendly interface.	

EDUCATION

Master of Engineering in Computer Science <i>University of Cincinnati, College of Engineering and Applied Science</i>	Apr 2025 <i>Cincinnati, OH</i>
<ul style="list-style-type: none">GPA: 3.67/4.0Relevant Coursework: Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Cloud Computing, Software Engineering	

PROJECTS

ArthaNethra — AI Financial Risk Investigator | *Python, FastAPI, Neo4j, Weaviate, Angular* Oct 2024 – Nov 2024

- **Financial AI Hackathon Finalist** — Engineered knowledge graph-native financial investigation platform mapping entities across 10-Ks, contracts, and invoices.
- Designed dual-database architecture (Neo4j + Weaviate) enabling complex cross-document queries beyond traditional RAG capabilities.
- Constructed hybrid extraction pipeline (LandingAI + Claude) achieving 99% accuracy at 80% lower cost than pure LLM approaches.
- Built explainable AI chatbot with clickable PDF citations and interactive Sigma.js graph visualizations.
- Deployed full-stack application via Docker Compose; reduced M&A due diligence time from 200+ hours to under 2 hours (90% reduction).

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

- Led MERN stack development of an image narration tool, achieving 79.7% object recognition accuracy via TensorFlow and Hugging Face Transformers.
- Optimized model performance via transfer learning on large-scale datasets using pre-trained vision-language models.

TECHNICAL SKILLS

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

AI/ML Technologies: LLMs, RAG, Knowledge Graphs, Multi-Agent Systems, LoRA/QLoRA, Prompt Engineering, PyTorch, TensorFlow, scikit-learn, Hugging Face, LangChain, LlamaIndex, Pandas, NumPy, OpenCV

Databases: Neo4j, Weaviate, Pinecone, PostgreSQL, MySQL, MongoDB, Redis, Elasticsearch

Cloud & DevOps: AWS (SageMaker, Bedrock, Lambda, ECS Fargate, S3), Docker, Kubernetes, Jenkins, CI/CD, Git

Frameworks: FastAPI, Flask, Django, Node.js, NestJS, Angular, React, Spring Boot, ASP.NET Core

Development Practices: Agile/Scrum, Microservices, RESTful APIs, TDD, MLOps, Model Governance

CERTIFICATIONS

AWS Certified Solutions Architect – Associate

Deep Learning Specialization — DeepLearning.AI (Coursera)