

# ELANGO M

+91 90038 24868 • muruganelango2003@gmail.com • GitHub • LinkedIn • Portfolio

## PROFESSIONAL SUMMARY

AI/ML Engineer specializing in building scalable, production-ready intelligent systems that transform complex data into impactful solutions. Experienced in end-to-end machine learning pipelines, model deployment, and applied AI across computer vision, NLP, and generative AI domains.

## EXPERIENCE

<b>IIT Madras Dept. of Data Science &amp; Artificial Intelligence</b> <i>AI/ML Systems Engineer (Project Associate)</i>	Oct 2025 Present Chennai, India
<ul style="list-style-type: none"><li>Architecting AI-driven screening and learning-support systems, translating research concepts into deployable ML applications.</li><li>Designing end-to-end ML pipelines including preprocessing, feature engineering, training, and optimization.</li><li>Building classification models using structured and behavioral datasets for screening support and learning guidance.</li><li>Deploying models as FastAPI-based services for real-time inference and application integration.</li><li>Performing model validation, hyperparameter tuning, and error analysis to improve prediction reliability.</li></ul>	Sep 2024 Sep 2025 Coimbatore, India

## PROJECTS

<b>GenAI Knowledge Assistant (RAG System)</b>	Python LangChain ChromaDB OpenAI
<ul style="list-style-type: none"><li>Engineered a Retrieval-Augmented Generation (RAG) pipeline for semantic document search and intelligent Q&amp;A.</li><li>Implemented advanced document chunking strategies and vector embeddings to improve retrieval accuracy by 30%.</li><li>Designed document ingestion, preprocessing, and embedding storage workflows for scalable knowledge indexing.</li><li>Optimized prompt templates and response validation to minimize hallucinations and improve answer relevance.</li><li>Built context-aware query handling to deliver consistent responses across diverse document types.</li></ul>	Sep 2024 Sep 2025 Coimbatore, India
<b>Real-Time Accident Detection System</b>	Python TensorFlow OpenCV Deep Learning
<ul style="list-style-type: none"><li>Designed a CNN-based computer vision system to detect traffic accidents from live CCTV streams with low-latency inference.</li><li>Developed custom preprocessing modules for motion analysis and temporal feature extraction in video frames.</li><li>Tuned model architecture and hyperparameters to improve detection reliability under varying lighting and traffic conditions.</li><li>Optimized inference pipeline for near real-time performance and deployment feasibility.</li><li>Integrated an automated alert system for real-time emergency notifications via REST API triggers.</li></ul>	Sep 2024 Sep 2025 Coimbatore, India

## Scalable ML Model Inference API

*FastAPI Docker Redis Scikit-learn*

- Built a production-ready API for serving multiple ML models with high concurrency support.
- Designed modular model loading and inference pipelines to support scalable prediction workflows.
- Standardized structured output formats and implemented data validation using Pydantic for robust client-side integration.
- Optimized request handling and caching strategies using Redis to reduce latency.
- Developed containerized deployment setup using Docker for consistent environment reproducibility.

## TECHNICAL SKILLS

---

**Programming:** Python (NumPy, Pandas), Java, SQL, PySpark

**Machine Learning:** TensorFlow, PyTorch, Scikit-learn, Classification, Feature Engineering, CV, NLP

**Generative AI:** LLM APIs, LangChain, RAG, Embeddings, Vector Databases (ChromaDB, Pinecone)

**Backend & Deployment:** FastAPI, REST APIs, Model Serving, Docker, Redis, Inference Optimization

**Cloud & Tools:** Azure, Git, GitHub Actions, Linux, Jupyter, Postman

## EDUCATION

---

### KPR Institute of Engineering and Technology

2020 2024

*B.Tech Artificial Intelligence and Data Science (CGPA: 8.34)*

Coimbatore, India