

Deepak Thirukkumaran

Chennai, India | +91-9940211754 | thirudeepak2003@gmail.com

 LinkedIn |  GitHub |  Portfolio

Education

Shiv Nadar University, Chennai
B.Tech, Artificial Intelligence & Data Science

2021 – 2025
CGPA: 7.683

Experience

ORCA Digital Lean Solutions, *Contract – AI/ML Software Engineer*

Michigan, USA
Nov 2025 – Present

- KPI optimization framework for manufacturing: multi-tenant FastAPI backend with PostgreSQL, calculation engine for weighted scoring and cost-efficiency optimization, and budget-constrained action plan generation using greedy algorithms. 7 internal tools shipped.

CreatorOS by DRPCRD, *AI Engineer Intern*

London, England
Jan 2025 – Jul 2025

- Full-stack demographic and profanity detection platform for Instagram/TikTok using PaddleOCR, OpenCV, and custom NLP pipelines.
- Creator pricing prediction model with A/B testing and statistical validation, +15% campaign ROI. FastAPI, pgvector, Docker, GitHub Actions CI/CD.

IIT Hyderabad, *Research Intern*

Hyderabad, India
May 2024 – Sep 2024

- VirConvNet: multimodal 3D object detection framework (LiDAR + RGB) with Stochastic Voxel Discard and NRConv, reducing voxel density by 90% and 2× inference speedup.
- SE blocks and Pointwise Spatial Attention for detection robustness, +3.42% on KITTI 3D AP (Moderate) over Voxel-RCNN baseline.

Tata Consultancy Services (TCS), *Computer Vision Intern*

Chennai, India
May 2023 – Jul 2023

- YOLOv8 computer vision pipeline with OCR integration for P&ID symbol classification, 95%+ accuracy. Deployed as Dockerized Streamlit app for industrial inspections.

Sports Mechanics, *Data Analyst Intern*

Chennai, India
Dec 2023 – Feb 2024

- Worked with tools for Player performance prediction and talent evaluation. AI-powered commentary generation (Mistral-7B quantized), live match notifications via OneSignal API, milestone tracking.

Projects

MedFusion: Vision-Language Framework for Radiology  | *CV, LLMs, Healthcare*

- Multimodal radiology VQA and report generation system combining ConvNeXt, fine-tuned CLIP (ViT-32B), and Med-LLAMA with residual feature fusion. 76.3% accuracy on SLAKE, 72.8% on PMC-VQA benchmarks.

Fintech Intelligence Pipeline  | *LLMs, Fintech, Full-Stack*

- AI-powered stock analysis platform with live market data (Polygon.io), LLaMA 3.2 sentiment/risk analysis, and automated investment insights. Deployed with Flask APIs, React dashboard, PostgreSQL, and Railway CI/CD.

HITAL Curriculum Generator  | *n8n, Ollama/Mistral, Docker, Webhooks*

- Automated curriculum generation pipeline using Ollama/Mistral with constrained prompts for structured JSON output mapped to competency standards. Async human-in-the-loop approval gate via webhooks. Dockerized for Railway.

MiniGPT: Transformer Language Model from Scratch  | *Deep Learning, Transformers, NLP*

- 22M parameter GPT implementation: token embeddings, multi-head self-attention, positional encoding, feed-forward layers. Training, generation, and top-k/nucleus sampling. 40 perplexity on WikiText-2.

SAP Component Extraction Tool  | *Python, FastAPI, Regex, React*

- Automated extraction from unstructured SAP schedule files using regex pattern matching. Multi-file upload with cumulative quantity aggregation by custom date filters. Deployed on Railway with Excel export functionality.

Skills

Core ML/AI: Computer Vision, NLP, LLMs (LLaMA, Qwen, Mistral), Multimodal Systems, RAG, Agentic AI, RLHF

Frameworks & Tools: PyTorch, OpenCV, PaddleOCR, LangChain, HuggingFace, n8n, Ollama

Backend: Python, FastAPI, Flask, PostgreSQL, pgvector, SQLAlchemy

Frontend: React, TypeScript, Tailwind CSS, Recharts, ReactFlow

Infrastructure: Docker, AWS, Azure, GCP, GitHub Actions CI/CD, Railway, Vercel

Certifications

IBM AI Developer  | IBM Rag and Agentic AI  | AWS AI Practitioner 