# M Rakesh Reddy

**Senior Machine Learning Engineer at Emergence.AI, Bengaluru**

+91 9543143574 | m.rakeshreddy95@gmail.com | Linkedin | Github

## Technical Skills

**Languages**: Python
**Technologies**: Agentic AI (Vision and LLM based), Generative AI ( Computer vision, Natural Language Processing/Understanding), Deep Learning, Machine Learning, Edge, Distributed, Accelerated Computing ( GPUs, TPUs, NPU), Data Analytics
**Frameworks**: PyTorch, Tensorflow, OpenCV, Scikit-Learn, CUDA, cuDNN, TFLite, ONNX

## Experience

**Senior Machine Learning Engineer** August. 2024 – Present
*Emergence.AI* *Bengaluru, India*

### Agent-E (Web AI Agent – Research & Applied)

- Designed and implemented Agent-E, a web navigation agent leveraging LLMs, and benchmarked its performance on WebVoyager and WebArena.
- Achieved 89% accuracy on WebVoyager compared to human annotators by developing a generative validation framework for self-evaluation.
- Spearheaded the development of a vision-LLM hybrid framework, improving the agent's adaptability to complex UI structures and increasing task completion rates by 30%.
- Built a test automation framework for API and UI validation to ensure Agent-E's correctness in staging environments.
- Researched long-term memory for web agents, improving retention and recall across multi-step interactions.
- Fine-tuned and optimized Gemini 1.5 Flash and 2.0 Flash to extract critical interactive attributes from the DOM, boosting extraction accuracy by 25% compared to pretrained models.

**Chief Engineer** Oct. 2022 – August. 2024
*Oppo-OnePlus R&D Center* *Hyderabad, India*

### Personalized video editing (Pre-Research & Applied) - Computer Vision

- Built the pipeline utilizing BLIP-Diffusion for image guidance and Tokenflow for text-guided editing. Interpolated frames and applied super-resolution on the pixels for better quality
- Developed attention injection from DDIM inversion to reverse diffusion for motion consistency and to preserve spatial relation which enhanced the generated videos by 15-20 %
- Optimized the pipeline using local/global token merging/un-merging by using multiple frames together thus reducing the latency from > 5 mins to <1.5 mins for a 60-sec video editing.
- Evaluated the pipeline on DDIM, Euler, UniPCMultiStep, and DPMSolverSDE schedulers.

### Smart Image Matting - Computer Vision

- Flashcut, a novel object extraction architecture using Swin-T backbone inspired from Pyramidal architecture (InSPyReNet), achieving state-of-the-art performance (MAE 0.015) on embedded devices.
- Compressed the model by 95% for on-device deployment, reaching millions of users through ColorOS 13 on Oppo & OnePlus handsets

### Demoiréing, Deshadowing - Computer Vision

- Developed AADNet, a mobile moire removal network achieving state-of-the-art PSNR (23.6) on open-source ESDNet dataset with focal frequency loss and attention between low-level features. Paper
- Applied QAT ( Quantize aware training) and reduced the model size to 15MB suitable for mobile deployment.

### Text Summarization - NLP/ LLMs

- Fine-tuned models like Llama, Mistral and MoE for Chinese and English summarizing using Peft with QLora.
- GGUF format conversion for Ollama and DPO optimization for models.
- RAG for the latest bug reports using Codellama and MoE using Langchain and Llamaindex

*Others*

- POC solutions such as PPT Graphic recognition including various shapes and chart data extraction.
- Working on Image super-resolution, Face restoration. Updated multiple operators in support of MediaTek and Snapdragon processors and deployed by keeping the model consistent.
- Model compression using NAS, Pruning, Knowledge-distillation, quantization, Low-rank factorization. Conversion to ONNX, tflite.

**Senior Machine Learning Engineer** Jan. 2021 – Oct. 2022
*Quantiphi Analytics Solutions* *Bangalore, India*

- Computer vision for safety by working on deep learning modules such as Object detection, Image segmentation, Object tracking, Pose tracking and Action recognition/classification, achieved fewer than 10 false positives over 10-hour video streams across 22 cameras.
- Neural style transfer to generate realistic positive samples from the synthetic images, matched with real samples by 78%.
- AWS Sagemaker to develop and docker images for running the pipelines, hosted all the models on Triton GPU inference servers.
- Quantized the models to lower bit integers and optimized layers for low latency of <500ms.
- Computer vision for video analytics by summarizing the media context to see the impact of viewership. Implemented Emotion recognition, Speech and Natural Language understanding on GCP and improved the analytics reports by 37%.
- Computer vision for sports analytics by implementing player detection/ tracking, OCR, Similarity matching with 89% precision and 92% recall in assigning the right track.
- POC solutions such as density of people/sq.ft, the velocity of people for western railways. De-duplication of the video contents in meta storage(reduced storage cost by 40%), drone detection, unauthorized intrusion detection by person re-identification/ Kalman filtering, deep sort and action recognition with pose/frame. Warehouse object counting and monitoring.

**Techincal Analyst** Aug. 2019 - Jan. 2021
*Oracle India Private, Ltd* *Bangalore, India*

- Search engine for the internal repository with KNN, inverted indices, cosine similarity and ELK.
- Object detection/recognition model to produce information from the hanging boards to get the details/stock listing.
- Automation of EBS logs using custom BERT models to take action on the failed program in production to reduce the manual intervention by 80%.

**Associate Analyst** Sep. 2017 - Aug. 2019
*Oracle India Private, Ltd* *Bangalore, India*

- Time-Series models to predict the incoming orders per store of OfficeDepot and implemented it in ERP Oracle APPS in the client environment.
- R&D SOTA techniques in the field of ML, DL and implemented the architectures from the research papers with modifications.

## Education

**Dr. MGR Educational and Research University** Chennai, India
*Bachelor of Technology - Computer Science Engineering* *2013 - 2017*

- Achievements: Gold medalist.

## Blogs & Research Papers

Posture tracking for keystroke controlling with OpenCV for gaming Medium
Ultrasound nerve segmentation Medium
AADNet: Attention aware Demoiréing Network arXiv

## Certifications

Applied-AI certification for Machine Learning Link
NVIDIA - Fundamentals of Accelerated Computing with C/C++ Link

## Interests

Space Engineering | Physics | Blogging