Yashwanth Erukulla

erukullayashwanth2005@gmail.com | 9989591921 | linkedin | github | portfolio | blog

EDUCATION

Vellore Institute of Technology   Vellore, India
*B.Tech in Computer Science and Engineering (Blockchain Technology) - **CGPA: 9.37/10***   *2023 – 2027*

EXPERIENCE

Interact   August 2024 – Present
***AI/ML Intern** website*   *LlamaIndex, Groq, Instructor, Python*
- Architected and operationalized a scalable hackathon judging utility capable of evaluating 1000+ repositories in parallel.
- Developed a data-driven review workflow reducing judges' assessment time by 30% while improving accuracy.
- Integrated LlamaIndex for data pipelining and utilized Groq to accelerate inference speed.

Vellore Institute of Technology - Under review at JMCS (Q1 Journal)   April 2024 – Present
UG Researcher, Team Lead, First Author - under Dr. Anjaneyulu GSGN   *Cryptography, Research, Technical Writing*
- Leading research initiative on "A Novel Secured Cryptosystem using Pseudo Random Number Generator with Laplace Transform based on Auto Generated Key," designing the algorithm, authoring the paper.

PROJECTS

GPT-2 from Scratch   2024
GPT-2 model built from scratch github   *PyTorch, NumPy, Tiktoken*
- Engineered core components including Multi-Head Self-Attention, Layer Normalization among others from scratch.
- Implemented efficient token processing pipeline achieving successful pre-training on 5,000+ token dataset.

DDPM from Scratch   2025
AI image generation model made from scratch github   *PyTorch, UNet, Diffusion Models*
- Architected a complete Denoising Diffusion Probabilistic Model implementing forward and reverse diffusion processes.
- Designed custom UNet backbone inspired by Stable Diffusion and trained the model on MNIST.
- Optimized inference pipeline achieving 25-30 iterations per second on CPU-only environments.

LTM & RAG Implementation for LLMs   2024
Long-Term Memory and RAG for Large Language Models github   *Langchain, ChromaDB, Sentence-Transformers, Hugging Face*
- Developed persistent memory architecture enabling inter-session context retention for conversational AI agents.
- Implemented vector-based retrieval system with semantic chunking for efficient document contextualization.
- Optimized Mistral-7B through 4-bit quantization, reducing model size by 75% while maintaining similar response quality.

SortIQ - Yantra Central Hack-25 Finalist   2025
AI-Powered waste categorization, tracking and analytics system github   *Detectron2, PyTorch, OpenCV, FastAPI*
- Created instance segmentation solution using DeepLabv3+ architecture for real-time waste categorization.
- Achieved near-SOTA mean Intersection over Union (mIoU) score of 52.49% on Zero-Waste dataset.
- Constructed custom object tracking algorithm based on segment properties for continuous detection across video frames.

Tech Barista - Technica-24 Hackathon — Best Freshers Award   2024
Tool for streamlining intern and employee onboarding github   *Langchain, Next.js, Django*
- Built an intelligent code analysis system that extracts architectural patterns, synthesizes implementation details, and generates comprehensive functionality summaries with an integrated technical QA capability.
- Devised scalable data processing pipeline for code analysis handling repositories up to 5,000+ lines of code.

ACHIEVEMENTS

Academic Merit Award for the year 2024-25   2024-2025
Awarded for securing Academic rank 6 of CSE(spec. in Blockchain Technology) branch   *Vellore Institute of Technology*

Academic Merit Award for the year 2023-24   2023-2024
Awarded for securing Academic rank 2 of CSE(spec. in Blockchain Technology) branch   *Vellore Institute of Technology*

Finalist - Yantra Central Hack - 25   2025
Finalist at VIT's largest hackathon for developing SortIQ   *Vellore Institute of Technology*

Best Freshers Award – Technica'24 PAN India Hackathon   2024
Won among 2,600+ participants for developing Tech Barista   *Indian Society for Technical Education - VIT*

TECHNICAL SKILLS

AI/ML Technologies: Deep Learning, PyTorch, Transformer Models, Huggingface Transformers, Langchain, LlamaIndex, Machine Learning, Scikit-Learn, NumPy, Pandas, Matplotlib, Seaborn, chromaDB, Faiss, Gradio, BitsAndBytes, Detectron2, OpenCV, LLMs, Diffusion Models (DDPM, SD), GANs, VAEs, Retrieval-Augmented Generation (RAG)

Other Technologies: Neo4j, PIL, Selenium, Git, Linux, Instructor Library, FastAPI, RestAPI, Groq API, Python, C, C++, Java, Javascript, HTML, CSS, React.js

Research Skills: Cryptography, Mathematical Modeling, Symbolic Computation, SymPy, Algorithm Design, Technical Writing