

Yuvanesh Sankar

☎ 8072714770 ✉ yuvanesh.skv@gmail.com 🌐 Yuvanesh 🔄 Yuvanesh 🐦 Scriptosis 📁 Portfolio

EDUCATION

Vellore Institute of Technology (VIT)

Bachelor of Technology in Electronics and Communication Engineering (CGPA: 8.77)

Vellore, India

Aug 2022 – Sept 2026

EXPERIENCE

L&T Technology Services (LTTS)

May 2025 – Jun 2025

AI & Embedded Systems Intern

On-site

- Deployed a PyTorch-based DNN model on a Renesas RX24U microcontroller for real-time motor imbalance detection and enabled intelligent control shutdown via embedded AI inference.
- Integrated the DNN model using the embedded AI chip on the microcontroller for converting the .pt file to embedded c file.

JOL Energy

Jul 2023 – Jan 2025

Chief Strategy Officer (CSO)

On-site

- Managed a diverse cross-functional team spanning developers, embedded engineers, marketing, production, legal, and creative roles to align technical and non-technical efforts.
- Led 6+ market researchers to drive strategic decisions, increasing campaign success by 25% and contributing to 15% YoY growth.
- Integrated AI/ML into automation systems, saving 124 hours/month, cutting errors by 27%, and reducing development cycle from 18 to 12 days.
- Strengthened organizational growth by fostering key partnerships, expanding networks, and supporting funding opportunities that advanced both financial stability and market reach.

Collab Junction

Jun 2024 – Dec 2024

Marketing & AI SEO Engineer

Hybrid

- Developed an AI-powered SEO engine using LLMs and keyword analytics to optimize content structure, boosting visibility and engagement.
- Automated event management tools using ML and generative AI, boosting user engagement by 22% and saving 35 hours/month.

PUBLICATIONS & RESEARCH

FrugalSOT - Frugal Search Over The Models

2025

First Author - Accepted for Presentation at ICICISconf2025 (IEEE)

[Website](#)

- **Conference:** ICICISconf2025, Track 04: Artificial & Computational Intelligence (IEEE)
- Developed resource-aware adaptive model selection system for edge NLP inference using complexity classification (prompt length, NER density, syntactic structure) achieving 21.34% inference time reduction with only 2.68% relevance loss on Raspberry Pi 5.
- Implemented adaptive thresholding with exponential moving averages ($\alpha=0.2$) and cosine similarity-based quality assess(low→mid→high→fallback) based on relevance score thresholds derived from 250 prompt-response pairs.
- Eliminated cloud dependency through complete on-device processing with memory-aware selection (<8GB vs >8GB RAM), enhancing privacy and reducing latency for IoT, autonomous systems, and privacy-sensitive applications.

PROJECTS

AlphaDesign - F1 Aerodynamics Optimizer | Python, RL, Genetic Algorithms, Numpy-STL, Trimesh | [GitHub](#)

- Novel RL + genetic algorithm pipeline for F1 front wing optimization combining parametric CAD modeling, automated STL generation, and CFD analysis.
- Implemented value and policy networks guiding genetic mutations through numpy-stl and trimesh, achieving iterative self-improvement in aerodynamic efficiency.

AlphaStack - Universal Development Framework | Python, Google-GenerativeAI, Pydantic, Docker | [GitHub](#)

- Building universal AI-powered development stack supporting any technology framework with intelligent scaffolding and automated feedback based improvement loop.

CMARCOS -Multi-Robot Warehouse Coordinator | *Python, PyTorch, MARL(COMA), PyBullet, LiDAR* | [GitHub](#)

- A Counterfactual Multi Agent Robot Coordination System.
- This is a multi-agent reinforcement learning system that extends the Counterfactual Multi-Agent (COMA) algorithm to continuous action spaces for coordinating autonomous mobile robots (AMRs) in a dynamic warehouse environment.

PHYDRA - ISS Cargo Stowage System | *C++, FastAPI, ReactJS, Docker, Prisma* | [GitHub](#) | [Website](#)

- End-to-end cargo management system with advanced 3D bin packing algorithms, A*/Dijkstra pathfinding implemented in C++ for ISS applications.
- Achieved 95% compute reduction through optimized subprocess integration, full-stack deployment with Docker and MongoDB via Prisma ORM.

Customer Service Call Agent | *AWS EC2, Ollama, Whisper, ElevenLabs, Twilio*

- Mixture-of-experts voice AI handling 30 concurrent calls with custom load balancer, VAD, ASR, fine-tuned LLMs, TTS, and emotion detection.

HyperKuid Labs - University Project Gallery | *Next.js, AI Personas, Turborepo, Shadcn* | [GitHub](#)

- Art gallery platform featuring university projects with AI senior personas for instant guidance and community-driven project approval system.

TECHNICAL SKILLS

Programming Languages: Python, C++, JavaScript, CUDA, C, Java, Shell Scripting

AI/ML & Deep Learning: PyTorch, Transformers, Hugging Face, Unsloth, OpenCV, Whisper, YOLO, Scikit-learn, NLP, LLMs, Fine-tuning, Reinforcement Learning, MARL, COMA, Transfer Learning, Meta Learning, Genetic Algorithms

LLM Inference & Optimization: ONNX Runtime, TensorRT/TensorRT-LLM, GGUF workflows (llama.cpp), Post-training quantization (AWQ, GPTQ, 4/8-bit), FlashAttention, Speculative decoding (EAGLE, Medusa)

Web Development & APIs: React.js, Node.js, Express.js, FastAPI, MongoDB, PostgreSQL, Prisma ORM, Drizzle ORM, REST APIs, WebRTC, Tailwind CSS

Cloud & DevOps: AWS (EC2, S3), Docker, Linux/Unix, Git, Vercel, Raspberry Pi 5, Ollama, SGLang, Edge AI Deployment

CERTIFICATIONS & ACHIEVEMENTS

Machine Learning, DL, NLP Bootcamp – Krish Naik (Certificate)

Full-Stack Web Dev Bootcamp – Dr. Angela Yu (Certificate)

3rd Place – Envision 2025 – DPIIT Pitch to Win @ IIT Madras (Proof)