# Jerush Vijay

Bengaluru, Karnataka, India | vijayjerush4687@gmail.com | github.com/jrush-vj | linkedin.com/in/jerush-vijay

## Summary

3rd-year B.E. Artificial Intelligence & Machine Learning undergraduate focused on infrastructure and AI inference. Built and operates a bare-metal Proxmox homelab with containerised services and GPU-accelerated local LLM serving, and won two hackathons in 2026 building AI-driven security tooling. Seeking **DevOps, MLOps, SRE, or AI Inference** internships.

## Technical Skills

- **Languages:** Python, Java, Bash
- **AI/ML & Inference:** PyTorch, NumPy, llama.cpp, llama-swap, CUDA, GGUF quantisation, Mixture-of-Experts (MoE) inference benchmarking
- **DevOps & Infrastructure:** Linux administration, Proxmox VE (LXC), Docker, Portainer, Cloudflare Tunnel, Git, GitHub Actions, CI/CD
- **Backend & Web:** FastAPI, Flask, REST APIs, WebSockets, WebRTC

## Projects

**Proxmox Homelab & Local MoE LLM Inference** | *Proxmox VE / LXC / Docker / CUDA / llama.cpp* Ongoing

- Built a bare-metal Proxmox VE server with 5 role-separated LXC containers, sharing an NVIDIA RTX 4060 into the AI container via cgroup device rules; compiled llama.cpp with CUDA behind llama-swap to serve Qwen3.6-35B-A3B (35B-parameter MoE) on an 8 GB GPU with CPU expert offloading and 128K context
- Benchmarked 3 MoE models across 3 llama.cpp forks, sweeping expert offload, batch sizes, KV-cache quantisation, and flash attention; mapped throughput vs. context length and root-caused a fork-specific CUDA crash at high context depth

**Self-Hosted Multiplayer UNO** | *FastAPI / WebSockets / WebRTC / Docker* uno.jrushvj.dev

- Built a real-time multiplayer UNO game for up to 6 players with FastAPI and WebSockets (private rooms, reconnect-safe sessions), plus peer-to-peer WebRTC voice/video via a Cloudflare TURN server; containerised and self-hosted on a personal domain

**RSVP Reader** | *Tauri / React / Python* Windows / Linux

- Built a cross-platform speed-reading desktop app that presents PDF text one word at a time (RSVP) with chapter detection and saved progress; automated x64/ARM64 builds and GitHub Releases with GitHub Actions

## Experience

**IoT & Robotics Intern / Machine Learning Intern** Jun 2026 – Aug 2026
*Corizo Edutech Private Limited* *Remote*

- Completed a dual-track internship covering IoT & robotics (circuit design, embedded systems) and machine learning fundamentals; finished 1st in the IoT & Robotics cohort

**Technical Team Member – Internet of Things** Sep 2025 – Present
*Yantra, The IoT Club, REVA University* *Bengaluru, Karnataka*

- Contribute to club-led IoT builds and technical initiatives as part of the core technical team

**Event Coordinator & Host** Jan 2025
*ISRO "Space on Wheels" Exhibition, REVA University* *Bengaluru, Karnataka*

- Hosted the exhibition and coordinated first-semester class representatives across departments from planning through execution

## Achievements

- **Winner, Best Open Source AI Agent Track** – Hacktober Fest, KrowdKraft Oct 2026
  Built in 5 hours (team of 2) an AI security gateway that detects prompt, SQL, and log injection and diverts attackers into a honeypot, where an open-source LLM analyses their behaviour as threat intelligence
- **Winner** – Hackurity 2026, 24-hour Cybersecurity Hackathon, Cybersecurity Club, REVA University Oct 2026
  Won with Team Firewall Crew on the team's first cybersecurity project
- **Winner, IoT & Robotics** – Carnival of Champions, Corizo Edutech Jul 2026
  Ranked 1st in the cohort on weekly assessments, mentor evaluations, and project work

## Education

**REVA University** Bengaluru, Karnataka
*B.E. in Artificial Intelligence & Machine Learning | CGPA: 8.13/10* *2024 – 2028 (Present)*

**Kendriya Vidyalaya Rail Wheel Factory** Bengaluru, Karnataka
*Senior Secondary Education (CBSE)* *2012 – 2024*