

# Kolagani Pavan Kalyan

+91 8985229562 | [kolagani.pavankalyan2003@gmail.com](mailto:kolagani.pavankalyan2003@gmail.com) | [linkedin.com/in/pavankalyan-kolagani](https://linkedin.com/in/pavankalyan-kolagani) | [github.com/Cintu07](https://github.com/Cintu07)

## EDUCATION

---

### Amrita Sai Institute of Science and Technology

Bachelor of Technology in Electronics & Communication Engineering

Paritala, India

Aug 2022 – Mar 2026

## EXPERIENCE

---

### Founding Engineer

Cortex AI

Nov 2025 – Present

Remote

- Architected production backend (8000+ LOC) for an AI phone receptionist serving medical clinics – real-time voice pipeline with Twilio Media Streams, OpenAI Realtime API, and ElevenLabs v3 TTS
- Designed multi-tenant system with independent phone routing, business hours, provider matching, and structured appointment capture via Prisma/PostgreSQL
- Built Redis-backed call routing with Lua-scripted atomic slot reservation and horizontal scaling via router/worker architecture

### Freelance Full Stack Developer

Various Clients (NDA)

Dec 2024 – Present

Remote

- Built production frontends with plug-and-play backend integration for business clients, and delivered custom AI tools and automation projects for college students and early-stage teams
- Set up and deployed AI systems for local businesses including voice agents, chatbots, and workflow automation

### Core Contributor

DXLander (Open Source)

Aug 2025 – Present

Remote

- Accelerated CI/CD pipeline execution by 40% through GitHub Actions caching and parallelized test stages; built E2E test suite with Playwright covering 90+ critical user paths
- Contributed backend features and reviewed PRs across the monorepo, improving code quality and shipping velocity for the core team

## PROJECTS

---

### ARES | Python, Ollama, OpenRouter, Monte Carlo Tree Search

2025

- Built inference-time reasoning framework (`pip install ares-reasoning`) improving LLM accuracy from 27% to 80%+ via multi-candidate generation, rule-based verification, and confidence-weighted scoring
- Implemented MCTS-based search (v2) with selection, expansion, simulation, and backpropagation for step-level reasoning exploration beyond greedy beam search

### ROUND Neural Architecture | Python, PyTorch, Riemannian Geometry

2025

- Studied novel neural architecture by Lexideck Research (US) that replaces RNN gating with topological phase accumulation on a circle, achieving 100% accuracy on 16-bit parity where GRU gets 50%
- Reproduced and ran 7+ benchmark harnesses (parity, graph cycles, Dyck-2 recursion, semantic algebra, QA, permutations) to validate claims on discrete-reasoning tasks with `hidden_size=32`

### Talos | C, eBPF/XDP, Go, Netlink

2025

- Engineered a kernel-bypass Layer 4 firewall processing packets at XDP hook point with near-zero overhead, supporting dynamic rule management through userspace Go controller via Netlink and eBPF map updates

## TECHNICAL SKILLS

---

**Languages:** C, C++, Go, Rust, Python, TypeScript, JavaScript, SQL

**AI & Applied ML:** OpenAI API (Realtime + Chat), LangChain, LangGraph, RAG pipelines, agentic workflows, prompt engineering, inference-time reasoning (MCTS, beam search), PyTorch, Whisper ASR, ElevenLabs TTS

**Backend & Infra:** Node.js, Fastify, Express, PostgreSQL, Redis, Prisma, Docker, Twilio Media Streams, WebSocket, GitHub Actions

**Frontend:** React, Next.js, Tailwind CSS, Svelte, Framer Motion