

ADITYA SURYAVANSHI

Computer Engineering Student · Terminal Bench & Agent Evaluation

✉ aditya.suryavanshi.dev@gmail.com

📍 Pune, India

🌐 [linkedin.com/in/aditya-suryavanshi](https://www.linkedin.com/in/aditya-suryavanshi)

☎ +91 8262996104

🔗 freakyady.netlify.app

🐙 github.com/FreakyAdy

👤 Summary

Second-year Computer Engineering student with hands-on experience authoring accepted Terminal Bench-style benchmark tasks through Parsewave's independent contributor program. Built Dockerized Linux debugging scenarios, deterministic verifiers, and completed paid peer reviews — not as an employee, intern, or embedded contractor. Seeking to contribute more deeply to agent evaluation and benchmark quality at Parsewave.

👜 Experience

Freelance Software Developer, Freelance Software Developer 01/2024 – Present

- Built and deployed client full-stack web applications using React, Node.js, and PostgreSQL.
- Containerized environments with Docker and set up CI/CD via Vercel and Netlify.

🎓 Education

B.E. in Computer Engineering, 2029 | Pune, India
Dhole Patil College of Engineering

🧠 Skills

Languages

Python, Go, TypeScript,
JavaScript, C/C++, SQL

Benchmarking & Systems

Docker, Linux, pytest, Git,
deterministic test design,
containerized eval
environments

AI & ML

PyTorch, Gymnasium, PyTorch
Geometric, Unsloth, OpenCV

Web

React, Node.js, Express,
PostgreSQL, REST APIs, GitHub
Actions

🧩 BENCHMARK CONTRIBUTIONS

Terminal Bench Contributor (Independent), 05/2026 – Present
Parsewave Contributor Program

- Per-task compensation for accepted submissions
- Independently authored and submitted terminal-style coding benchmark tasks; accepted contributions paid on a per-task basis after automated quality gates.
- Built `metric_segment_store_repair`: Docker-packaged columnar storage repair scenario with interacting defects in encoding, persistence, and mutation logic, plus pytest verifiers.
- Designed durable-state failure modes in `append/merge` paths that pass same-instance checks but fail on fresh-instance replay, requiring multi-step Linux debugging.
- Authored `encrypted_backup_recovery`: reproducible OpenSSL recovery chain requiring long multi-command agent trajectories with deterministic oracle validation.

- Completed paid benchmark task reviews evaluating instruction clarity, oracle correctness, test rigor, and reproducibility of other contributors' submissions.

SELECTED PROJECTS

TrafficFlowRL, Gymnasium city traffic simulator with shared Actor-Critic PPO and GAE
github.com/FreakyAdy/trafficflowrl

GridMind-RL, OpenEnv building-energy simulation; Meta × PyTorch OpenEnv Top 800 of 72,000+ teams

RL-UCO, Exploring offline RL for LLVM pass selection with GNNs on program dependence graphs (in active development)
github.com/FreakyAdy/RL-for-Universal-Compiler-Optimization

NeuroBoard, AI-assisted learning canvas; InnovationXPO 2026 National Finalist
github.com/FreakyAdy/NeuroBoard

AureX, Academic workflow app with QR attendance; built in 36-hour hackathon as first-year team lead

Achievements

Meta × PyTorch OpenEnv Hackathon

Top 800 of 72,000+ teams (Top 1%) with GridMind-RL

InnovationXPO 2026

National Finalist for NeuroBoard at a national-level project competition

First-Year Hackathon Team Lead

Led the only first-year team to build AureX in a 36-hour competitive hackathon