Pulkit Jai

Jaipur, IN — hello@pulxit.com — Website — LinkedIn — GitHub

Summary

Applied AI engineer specializing in **agentic systems** and **LLM safety/evaluations**. Top-20/600+ in OpenAI red-teaming; shipped production ML and enterprise-hardened multi-agent code review with approx. 2-minute end-to-end latency.

Technical Skills

Languages: Python, TypeScript, C++, Rust
ML & LLM: PyTorch, scikit-learn, JAX; finetuning (LoRA/DoRA); evaluation harnesses (Petri)
Agents & Runtime: MCP (Model Context Protocol), FastAPI, Docker, AWS, Vercel
Data & Infra: Convex, PostgreSQL, CI/CD, observability (PostHog, Sentry)
Web: Next.js, React

Current Work

WhereIsMySerotonin.com / mann Sep 2025–Present
Building a calm, profile-aware mental health companion. 2 minutes. Better mornings.

AI Safety & Research

ARTEMIS: AI Safety Red-Teaming Platform (OpenAI Red-Teaming Challenge) Aug 2025
Top 20/600+ teams (Honorable Mention)

- Discovered a **systematic CoT reasoning manipulation** vulnerability; reproduced **100% bypass success** across multiple harmful categories via four attack vectors, validated by a dual-LLM evaluator (avg. confidence ≈ 0.85) with **88** automated tests.
- Built a **pip-installable CLI** for reproducible runs (fresh-conversation protocol, vectorized test matrix) and structured reports; emphasized evaluation coverage and failure analysis over single-number scores.
- Developed an early **bypass-detection harness**; achieved approx. 75% detection across 4 implemented attack families during internal tests; documented limitations and next-step ablations (cross-model generalization, FP cost).

Independent Research & Open Source 2025–Present

- **Transformer implementation from "Attention Is All You Need"**: Paper-faithful encoder–decoder with authentic WMT14 loaders, base/big configs, BLEU evaluation utilities, and a comprehensive test suite; includes a simple CLI for training/eval workflows.
- **DoRA (weight-decomposed LoRA) implementation**: ~**67%** parameter reduction at parity-level task performance; reproduces key *magnitude–direction correlation* findings (DoRA −0.31 vs. LoRA +0.83; full FT −0.62); memory-optimized training that fits on an RTX 3060 (6GB) with mixed precision and tests.

Work Experience

Hivel.ai, Remote — Apr 2025–Sep 2025
Applied AI Engineer

- **Multi-agent code review ("Agent Marco")**: Designed coordination for 6+ specialized agents using Python + MCP with async orchestration and shared state; delivered approx. 2-minute average end-to-end review latency for enterprise pilots (e.g., Warner Bros Discovery, Freshworks).
- **Enterprise hardening**: Implemented authN/Z, audit logging, and PII scrubbing; deployed within SOC 2/HIPAA/GDPR-aligned environments; integrated with CI and observability.
- **Ownership**: Took architecture from design to pilot integration with senior guidance; promoted from intern to engineer in 4 months.

ReWorked.ai, Bengaluru, KA — Jun 2024–Mar 2025
AI/ML Engineer (intern → full-time)

- Shipped PyTorch + scikit-learn models (solar/roofing/mortgage) with **F1 above 0.8** in production; processed **5K+ records/day** with parallel inference pipelines.
- Exposed models via FastAPI services on AWS; maintained approx. **99% uptime** with containerized deployments and monitoring/alerts.
- Modularized model packaging and API interfaces to streamline frontend integration.

Projects

SeeSpeak: Accessibility Device for the Visually Impaired — Jan 2024–July 2024

- Raspberry Pi + edge AI system with OCR + bilingual (Hindi–English) TTS; approx. **90%** accuracy on printed text and **70%** on handwriting; LLM APIs + edge-optimized for offline use.
- Undergrad final semester project/thesis

Leadership

Electronics Club, Head of R&D (2022–2023): Led workshops/competitions for **200+** students with a **15+** member team.

Education

Manipal University Jaipur, Jaipur, RJ — Jul 2024

- B.Tech., Electrical & Electronics Engineering (EEE)
- Minor in Computer Science and Machine Learning