Nishant Tyagi

AI Engineer · Evaluation · Runtime trust · Open source

github.com/nishanttyagi28 · linkedin.com/in/nishant-tyagi-7aa458356 · x.com/tnishant838

SUMMARY

AI engineer building production-grade agentic systems, evaluation frameworks, and secure autonomous infrastructure. Focused on AI reliability, LLM evaluation, agent safety, runtime trust systems, and open-source infrastructure that can be inspected, gated in CI, and proven at the effect boundary.

SELECTED SYSTEMS

AgentEval

PyPI · v0.3.0 · MIT

Git-native evaluation harness for multi-step LLM agents. YAML golden suites, baseline regression gates, Failure Memory, and adapters for LangGraph, CrewAI, AutoGen, and the OpenAI Agents SDK.

- Release verification: 1115 deterministic tests passed, 1 skipped; Failure Memory suite 59 passed.
- Production failure loop: redact, cluster, replay, minimize, human-approve, export golden YAML, gate CI.
- Install: pip install nishanttyagi-agenteval · github.com/nishanttyagi28/agenteval

KarmaSakshi Protocol

PyPI · v0.2.0 · MIT

Runtime trust infrastructure for consequential AI-agent actions. Cryptographic effect sealing, exactly-once execution, independent verification, TOCTOU protection, and Action Passports.

- 1049 tests passed, 90.5% coverage, mypy strict clean. 74 security invariants mapped to tests.
- Lifecycle: PROPOSE → PREPARE → SEAL → AUTHORIZE → COMMIT → VERIFY → PROVE.
- github.com/nishanttyagi28/karmasakshi-protocol

Agentic Data Analyst

Live demo · MIT

Multi-agent analysis system: quality gating, text-to-SQL, statistical tests, forecasting, AutoML, HTML reports, and ChromaDB RAG over session memory. Groq Llama 3.3. SELECT-only SQL allowlist.

PATH

Economics background
University of Delhi — systems thinking from incentives, trade-offs, and failure at the margin.

MCA, AI / ML
Graduate study in machine learning and computational systems.

Freelance engineering
Data systems, RAG backends, and client-facing agent tools.

Open-source infrastructure
AgentEval, KarmaSakshi Protocol, and Agentic Data Analyst shipped in public.

PRINCIPLES

Evaluation-first. Systems over demos. Open source as proof. Agent safety at the effect boundary. Production quality means fail closed and measure honestly.

References and source: github.com/nishanttyagi28