

TRAKSHAN MISHRA

AI/ML Engineer | LLM Applications, RAG, Agentic Workflows & AI Evaluation
Noida, India | trakshanmishra477@gmail.com | +91 7426050883 | [LinkedIn](#) | [GitHub](#)

SUMMARY

Generative AI / ML Engineer with production experience building LLM-powered applications, including an NL2SQL conversational database-querying system and a RAG evaluation pipeline benchmarked against baselines (79.6% precision). Skilled in LangChain/LangGraph agentic workflows, prompt/context engineering, and shipping high-throughput FastAPI inference endpoints (150 concurrent req/s) backed by optimized PostgreSQL pipelines.

TECHNICAL SKILLS

Data & Analytics: SQL, PostgreSQL, Data Pipelines, REST APIs, Quantitative Benchmarking & Evaluation, Metrics Definition, A/B-style Comparative Analysis

AI-Enabled Workflow Automation: LangChain, LangGraph, AI Agents, RAG, Prompt/Context Engineering, LLM Evaluation, NL2SQL, OpenAI API

Engineering & Tools: Python, TypeScript, JavaScript, React.js, FastAPI, Django REST Framework, Docker, Git, CI/CD, OAuth 2.0, Cloud Deployment, Linux

EXPERIENCE

Software Engineer / AI Agent Evaluation Contractor - Turing Aug 2026- Present

- Build and validate Python backend connectors that replicate SaaS tools and workflows for AI-agent training and evaluation.
- Create and verify realistic agent tasks and workflows, assessing correctness, realism, and task quality through structured QA and evaluation rubrics.
- Contribute to infrastructure and training data used to develop and evaluate AI agents, including failure analysis and repeatable quality checks

Software Engineer — NextGen Invent (Remote) September 2024- Jan 2025

- Architected and shipped production full-stack applications using React.js, Django REST, and FastAPI, improving API response times by 35% across multiple client deployments.
- Owned features end-to-end from design to deployment, collaborating with product and QA in an agile workflow.

AI Engineer Intern — Modern Monk Labs (Remote) June 2024 - Sep 2024

- Built high-throughput FastAPI inference endpoints serving deep learning models in production (150 concurrent req/s); optimized PostgreSQL pipelines and model-serving infrastructure, reducing latency.

Generative AI Engineer Intern — VRARMR (Remote) March 2024- June 2024

- Built NL2SQL LLM applications for conversational database querying and automated data-processing and outreach workflows with LangChain agentic pipelines, cutting recurring manual effort.

PROJECTS

DiffContext — Context Engineering for Code LLMs (Python, AST, LLM Evaluation) 2026 – Present

- Built a repository-scale context-selection engine for code LLMs: multi-file dependency-graph mapping with token-budget-aware pruning — cut prompt tokens from 35,270 to 10,747 (69.5%).
- Implemented a runtime tracer (`sys.settrace()`) generating gold-standard execution graphs — 79.6% precision on the Requests-library benchmark; running structured comparative analysis (BM25 vs. semantic retrieval vs. proposed method).

TradeTrack Pro — Fintech Dashboard (React, FastAPI, PostgreSQL, OCR, OAuth) 2025

- Built and shipped a live analytics dashboard with real-time market-data pipelines, automated tax calculation, and OCR-based data capture at 94% accuracy — a working example of turning raw data into a self-serve, stakeholder-facing tool.

Movie/TV Streaming Platform (React.js, Node.js, TMDb API, PostgreSQL) 2025

- Deployed a full-stack streaming platform with user auth and a content-delivery and ad-management back-end; in active daily use.

CERTIFICATIONS

- [SAP Certified Generative AI Developer \(Mar 2026\)](#) — benchmarked Gemini, Mistral, and GPT on SAP AI Core.

EDUCATION

B.Tech, Electronics & Communication — Jaypee Institute of Information Technology, Noida