

DAVID JEREMIE ANAND

Applied AI Engineer | Agentic Systems & RAG | LLM Evaluation

New Delhi, India | +91 70118 90891 | jerryanand10@gmail.com

davidjeremieanand.com | linkedin.com/in/david-jeremie-anand | github.com/Jeremie2002-sudo

PROFILE

Builds AI systems, then proves they work: **four shipped AI projects** across agentic triage, RAG, tool calling, and LLM evaluation – each with its own **eval harness and CI suite**, three in production on Render. Two years in Genpact operations supply the instincts those systems automate: SLA pressure, escalation judgment, root-cause discipline. Backed by a PGP from Masters' Union.

PROFESSIONAL EXPERIENCE

Genpact

Aug 2024 – Aug 2026

Associate and Subject Matter Expert – Operations

New Delhi, India

- Analysed workflows handling **100,000+ cases a year** and proposed process changes behind an **18%** cut in turnaround time.
- Performed structured root-cause analysis of recurring defects and SLA deviations, tracking issues through Buganizer and validating corrective actions; operational defects fell **30%** with minimal SLA breaches.
- Automated management reports and dashboards in Excel and Power BI, removing **25%** of recurring reporting effort and saving about **100 FTE hours annually**.
- Partnered with Risk, Compliance, and Quality stakeholders on controls and workflow improvements supporting a **22% productivity uplift** across a ~20-person operations team.

TECHNICAL PROJECTS

FirstLine – AI Ops Triage Agent with Self-Scoring Evals

GitHub · Live

Python, FastAPI, Ollama/Groq, pytest, GitHub Actions

- Built an agentic first line of support that works a full ops shift – classifying root causes across 15 incident archetypes, clustering cross-service incidents, tracking SLA clocks, escalating judgment calls, and drafting RCAs – over a seeded simulator with hidden labeled ground truth.
- Designed an eval harness that grades the agent against that truth (triage accuracy, missed POs, escalation precision); live Llama 3.2 runs surfaced **100% escalation recall at 21% precision** – safety by crying wolf; shipped with a **768-test CI suite** and a live triage dashboard.

Synapse – Agentic RAG Document Assistant

GitHub · Live

Python, FastAPI, FAISS, Ollama/Groq, pytest, GitHub Actions

- Engineered a hybrid retrieval pipeline (BM25 + FAISS fused via Reciprocal Rank Fusion, LLM listwise reranking) feeding a 4-stage agent pipeline (planner, retriever, synthesizer, critic) that grounds answers in evidence and self-critiques unsupported claims.
- Added an observability UI and an automated eval harness for retrieval and faithfulness scoring, plus a pluggable model backend (local Ollama / hosted Groq); backed by a **31-test CI suite**, deployed on Render.

Ledgr – AI-Powered Finance Manager

GitHub · Live

Python, Flask, SQLite, OpenAI/Ollama, Docker

- Shipped a full-stack finance tracker with dashboards, budget tracking, and CSV import/export, plus AI transaction categorisation (concurrent LLM calls with keyword fallback) that works fully offline.
- Designed a tool-calling assistant that queries real transaction data instead of relying on LLM arithmetic, with a pluggable OpenAI/Ollama backend; covered by a **58-test CI suite**, Dockerised, deployed via Render Blueprint.

PromptDrift – LLM Prompt Regression Testing

GitHub

Python, asyncio, Ollama/Groq, pytest, GitHub Actions

- Wrote a CLI that snapshots prompt behaviour into a committed baseline and fails CI when a change degrades it, sampling every case repeatedly so genuine regressions are separated from model non-determinism.
- Tiered assertions run deterministic checks before paid LLM-as-judge calls, cutting evaluation cost; guarded by a **157-test CI suite** across Python 3.10–3.13, running fully offline through a deterministic mock backend.

INTERNSHIPS

Bhagwan Parshuram Institute of Technology, GGSIPU

Jul 2023 – Aug 2023

Data Analytics Intern

New Delhi, India Certificate

- Analysed institutional datasets in Python and Excel and built KPI dashboards and automated reports, cutting repetitive manual effort by **25%**.

Defence Research and Development Organisation (DRDO)

Jul 2022 – Sep 2022

Research and Development Intern

India Certificate

- Supported development, validation, and signal analysis of an embedded wireless sensing prototype, improving reliability and reducing errors.

Agixury

Sep 2020 – Oct 2020

Web Development Intern

Remote Certificate

- Developed responsive web interfaces from business requirements and improved usability across desktop and mobile layouts.

EDUCATION

Masters' Union

Aug 2026 – Present

PGP in Applied AI and Agentic Systems

Gurugram, India

Bhagwan Parshuram Institute of Technology, GGSIPU

2020 – 2024

B.Tech in Electronics and Communication Engineering | CGPA: 8.53/10.0

New Delhi, India

Faith Academy Senior Secondary School, CBSE

2018 – 2020

Class XII: 92.8% | Class X: 94.2%

New Delhi, India

TECHNICAL SKILLS

Programming and Data

Python, SQL, SQLite, FastAPI, Flask, REST APIs, NumPy, asyncio, data analysis, exploratory data analysis

Applied AI

Machine learning, generative AI, agentic AI, multi-agent orchestration, large language models, embeddings, prompt engineering, tool calling (function calling), guardrails, Retrieval-Augmented Generation (RAG)

Retrieval, Evaluation and Cloud

Hybrid retrieval (BM25 + FAISS), reciprocal rank fusion, LLM reranking, vector databases, LLM evaluation (LLM-as-a-judge), faithfulness scoring, eval harness design, prompt regression testing, observability, LLMops, GenAIOps, model APIs (OpenAI, Ollama, Groq), Microsoft Azure AI, Snowflake, Render

Analytics and Development Tools

Power BI, Tableau, Advanced Excel, Git, GitHub Actions, CI/CD, Docker, pytest, test automation

Delivery and Problem Solving

Root-cause analysis, process improvement, stakeholder communication, customer-facing delivery

SELECTED CERTIFICATIONS

Professional Certificates: Snowflake Generative AI | Microsoft Azure AI Essentials | Astronomer Data Engineering Foundations

AI Engineering: Become an AI Engineer | GenAIOps Foundations | Generative AI in Cloud Computing

LLMs and RAG: Fine-Tuning for LLMs | Advanced LLMs with RAG | Vector Databases and RAG | AI Ecosystem for Developers