JANNET AKANKSHA EKKA

Applied AI Engineer · LLM Agents · RAG · AI Evaluation · Agent Security

jannetekka96@gmail.com | +91 90788 02572 | linkedin.com/in/jannet-akanksha-ekka | github.com/JannetEkka | jannetekka.github.io/DSProjects | Kolkata, India · remote or relocation

PROFESSIONAL SUMMARY

I build LLM agents and the tests that show whether they work. I spent four years at Deloitte leading QA automation and working directly with clients in US, UK and India hours, then finished Rank 1 in the Post Graduate Program in AI/ML at UT Austin McCombs. Since January 2025 I have built a patent-pending AI trading agent on my own, directing AI coding agents like Claude Code and reviewing every merge myself. Between 12 and 18 October 2026 I submitted four working hackathon entries, each on a different stack: a graph database, an open LLM running offline, a native Android app and an on-chain contract. The lesson I trust most: measure an agent before you believe it. When I graded my six trading agents on independent data, all six called direction at a coin flip, so I rebuilt the system around that result.

TECHNICAL SKILLS

GenAI & Agents	Google ADK, Model Context Protocol (servers written from scratch), A2A SDK, Gemini (Vertex AI and AI Studio), Apertus 1.5 on llama.cpp (offline), Imagen, OpenServ SDK, multi-agent orchestration, tool calling, prompt engineering, structured output, RAG (chunking, BM25, reciprocal-rank fusion, cross-encoder reranking), graph-backed agents on FalkorDB (Cypher tools, label propagation), Hugging Face
AI-Native Delivery	Claude Code, spec-driven development, agent-directed implementation, pull-request review and merge control, production log triage, defect isolation
AI Evaluation	Eval harnesses (recall@k, MRR, macro AUC, paired bootstrap), backtests against a simple baseline, grounding checks on quotes, dates and amounts, counterfactual faithfulness tests, adversarial agent testing by tool-call interception, overfitting gates (CPCV, Deflated Sharpe, PBO, FDR), replay through production code, liveness checks on production logs, pytest, Selenium, test strategy
ML & Quant	scikit-learn, TensorFlow, Keras, PyTorch, CatBoost, Optuna (TPE), contextual bandits, Kelly sizing, time-series, NLP, computer vision (ResNet, Fast R-CNN, MRI CNNs), NVIDIA cuDF / RAPIDS
Cloud & Infra	Oracle Cloud, Cloudflare Workers, Render; GCP: Compute Engine, Cloud Run, Cloud Functions, BigQuery (+ vector search), AlloyDB & pgvector, Firestore, Secret Manager, Vertex AI; AWS: S3, EC2, RDS, SageMaker, Textract, Comprehend; Docker (including offline internal networks), systemd, GitHub Actions
Engineering	**Python and SQL are the languages I take into a coding interview.** TypeScript and JavaScript I have shipped and read fluently, and I would be ramping in them rather than leading. Kotlin and Jetpack Compose for one shipped Android app. FastAPI, Flask, Streamlit, Django, React, Next.js, Postgres, REST and WebSocket APIs

FLAGSHIP PROJECT

Smart Money Trading (SMT): explainable AI trading agent · Founder & sole engineer

Jan 2025 – Present

Patent pending: Indian provisional application No. 202631090789, filed myself in July 2026

Live: smt-weex-trading-bot.jannet-ekka.workers.dev · *Public layer:* github.com/JannetEkka/smt-apac · WEEX feature

- Built it alone by specifying the work and directing AI coding agents; nothing merges without my review. **87,000 lines of Python, 3,100+ tests, 368 merged pull requests**, and a register of **680 measured findings** that decides what ships.
- Graded my own multi-agent committee (six persona agents and a learned Judge) on independent market data. **All six called direction at a coin flip.** Rebuilt on that result: written rules with replayed records open every trade, and the agents became advisors graded on every call, earning a veto only after 30 proven warnings.
- Every explanation passes a **counterfactual test** before a user sees it: flip the advisor's input and re-run the decision; if the decision doesn't change, the reason is dropped. The learning loop retunes only through a **CPCV + Deflated Sharpe + PBO + FDR gate**, which still holds a forecaster that scored AUC 0.723 on a BTC forward test in shadow.
- Tested **1,800 strategy variants, 36 million simulated trades** from 2020 to 2026 at six exchanges' published fees, replayed through the production code. Nothing holding a position under four hours cleared its fees. The crash-buy rule that did: 312 trades, 73% won, +4.76% average after fees, positive in every year.
- Audited 98 AI-written merges line by line and found two failures the green test suite missed: an order-format bug that had **rejected 1,137 live orders over three months**, and a memory freeze that kept the daemon down 60% of the time for ten days. A fix now counts as done only when the production log shows it.
- Runs 24/7 and places real orders on WEEX's exchange-granted simulated account (about $10K). When the GCP billing account closed, I **moved the live system to a zero-cost stack in two days** (Oracle Cloud free-tier VM, Cloudflare Workers, Google AI Studio's free tier) and it re-adopted its three open positions from committed state. Any substituted default logs a warning and a liveness flag, a rule that has caught four multi-week data bugs.
- Built **SMT World**, the product for users: a Three.js/WebGL world on Cloudflare Workers where people watch the agent trade and ask it why, answered by Gemini from the live book.

Stack: Python, pandas, Optuna, scikit-learn, CatBoost, Gemini, Oracle Cloud, Cloudflare Workers, BigQuery, systemd, pytest, GitHub Actions, Three.js

SELECTED AI PROJECTS

Lore: a company brain for open-source projects · Graph Hacks, FalkorDB × WeMakeDevs

Oct 2026

Live: lore-gh5y.onrender.com · *Code:* github.com/JannetEkka/lore · Demo video · Blog post

- Reads a project's GitHub history into a FalkorDB graph: who wrote and reviewed which pull requests, which files they changed, which issues they fixed. A Gemini agent answers “who knows this part of the code?” through **12 Cypher tools**, and every answer cites the PRs behind it.
- Backtested on 300 real FalkorDB PRs: a real reviewer in its top three **90%** of the time, against 88% for “the most frequent reviewer”. On the single top pick that baseline wins, and the write-up says so.
- Issue triage that drafts a reply without posting it, team detection by label propagation over who reviews whom, per-user memory graphs, GitHub sign-in and an MCP server. Built and deployed in four days on free tiers, demoed on FalkorDB's own GitHub org.

Klarbrief and Chithi: official letters explained, every point checked · Hack Apertus; Meta AI Developer Hackathon

Oct – Nov 2026

Code: github.com/JannetEkka/klarbrief · *Chithi:* chithi-oprj.onrender.com

- Apertus 1.5 8B, Switzerland's open LLM, explains an official letter in the reader's language. Before any point is shown, a deterministic check finds its quoted words in the letter and verifies its date or CHF amount. Points that fail are listed apart, never shown as fact.
- Runs fully offline: llama.cpp on an internal Docker network, a 4-core CPU, no GPU, no internet at runtime.
- Measured on 48 explanations of 28 synthetic letters: 48/48 in the reader's language, 28 of 29 real deadlines shown, **0 wrong CHF amounts shown**. The model offered 7 wrong dates and the check stopped 6. 34 tests and a 5-page technical report.
- **Chithi** (new code, Meta hackathon): the same check for Indian bills and notices, from a phone photo, in Bengali, Hindi or English, on SAM 3.1, Muse Spark and Muse Voice Transcribe. Each value must also be in a second, separate reading of the photo, and the reader sees the piece of their photo where the letter says it. 28 live explanations, 16 from folded, shadowed photos: **0 of 114 dates and amounts shown wrong**; 53 of 53 planted errors held back.

Wiretrap: adversarial security testing for AI agents · Gen AI Academy APAC, Cohort 3 *2026*

Live: wiretrap.onrender.com · *Code:* github.com/JannetEkka/wiretrap

- A security tester that checks which tools an agent actually called. Most scanners grade the agent's reply. This one reads the tool-call log.
- Every tool the agent declares is cloned as a dead copy with the same name and JSON Schema that records the call and returns fake data. Attacks run in separate Gemini conversations and a judge model reads the log.
- Caught a Google codelab agent saying "I don't have access or authorization to modify configuration files" after it had already run `ls -la` through its shell tool. Text scanners pass that; Wiretrap flags it with the arguments.
- A run with nothing classified reports **INCONCLUSIVE**, never HARDENED. Approved fixes are re-run against the attacks that broke them, and the report says when one didn't hold. 8 attack categories, 9.4K lines of TypeScript.

SmartDesk: multi-agent assistant with evaluated RAG · Gen AI Academy APAC hackathon, top 101 of 1,500+ *2026*

github.com/JannetEkka/smartdesk · *Eval write-up:* evals/RESULTS.md

- A root **ADK orchestrator** routing to inbox, calendar and knowledge-base sub-agents on Gemini and Cloud Run. Gmail and Calendar are reached through **MCP servers I wrote** against the protocol SDK.
- Built the **RAG evaluation harness before touching retrieval**: 40 labelled questions over 120 notes, recall@k and MRR@k, paired bootstrap at 10,000 resamples. Baseline recall@5 **0.963**.
- **The harness killed the change it was built to justify.** A cross-encoder reranker won on the development embedder (MRR@10 +0.094, p=0.04), then fell to +0.028 with a CI spanning zero on the production embedder, at the cost of 190 MB of PyTorch and search going from 29 ms to 2,382 ms. The baseline shipped.

November 2026 builds · Hermes Agent · NEAR · Mozilla Otari · PayPal · Catalyst *Nov 2026*

- **Lights Out** (repo), a Hermes Agent plugin that Nous Research merged into the official Hermes plugin catalog: every billable AWS create goes to Hermes's own approval prompt with its price, and a receipt lists what is still running. On 10 tasks × 3 arms, 0 creates ran without a person's yes, against 13 and 14 without it; a clean-up prompt alone also cleaned up well, and the README says so.
- **Aftershock** (NEAR Hackathon, page): an event study of 74 news spikes against ordinary days. Usage moved after 13.5% of them, against 16.5% of ordinary days, so the thesis says headlines are not evidence of usage. A second model family and a recomputation check every claim; 14 of 20 kept.
- **Countersign** (Mozilla Otari, app) finishes my 2024 sentiment project: a claim in a model-written brief shows only if a second model, a sandbox recomputation and a guardrail confirm it. On HaluEval it caught 41 of 60 planted errors and wrongly struck 19% of faithful sentences, and the page says both.
- **Pitch In** (PayPal AI): pledges held with PayPal and paid only when code rules and a photo check pass; a live sandbox run reconciled at 0 mismatches. **Orbit** (Catalyst, one-day solo build): top 8.2%, 33rd of 391 scored entries, not placed.

Two more October 2026 hackathon entries · Solana Mobile and Monad Metropolis *Oct 2026*

- **SMT World: Clock In** (repo): a native Android app in Kotlin and Jetpack Compose, built in six days. Users call the market once a day against my trading agent; both calls land on Solana devnet in one signed memo through Mobile Wallet Adapter and are graded on real prices four hours later.
- **SMT Receipts** (repo): the agent's calls and the positions it actually holds are sealed on Monad testnet by a verified, ownerless contract, then graded against RedStone-signed prices net of a 12 bp fee. A track record nobody can edit, me included.

Smart Money Tracker: multi-agent blockchain analytics · Winner, Best DeFi Application *Mar 2025*

- Multi-agent whale tracking across Ethereum, Polygon and BNB Chain on the OpenServ SDK, with Alchemy webhooks for sub-second monitoring and confidence-scored detection of accumulation, distribution and coordinated moves. SMT grew out of it.

Applied ML Portfolio · 13 end-to-end projects, UT Austin McCombs PGP *2024 – 2025*

- ResNet50 vehicle classification with Fast R-CNN detection over 16,185 images and 196 classes; churn prediction at 81.5%; semiconductor yield from 591 sensor signals; financial-news sentiment with Mistral-7B; SQL analytics. All at jannetekka.github.io/DSProjects.

PROFESSIONAL EXPERIENCE

JTech Solutions (pre-incorporation) · Founder & AI Engineer *Jan 2025 – Present · Kolkata*

- Building SMT and SMT World (above) as the company's first product, while a full-time family carer and alongside the PGP in AI/ML.

Deloitte Consulting LLP · Test Automation Engineer (Analyst) *Aug 2019 – Jan 2023 · Bangalore*

- **Led a 6-member QA engineering team** on automated testing for Fortune 500 clients including AT&T Consumer Connect, Hewlett Packard Enterprise and The Dental Supply Company.
- **Client-facing throughout:** ran demos, working sessions and requirement calls with client teams in US, UK and India hours, and covered weekend on-call.
- Built Selenium and Katalon Studio automation that **improved test execution efficiency by 83%**; Jenkins-based reporting cut manual analysis time by **75%**.
- Wrote Python analysis and Tableau dashboards tracking quality across **343 business components**, cutting test failures by 25%; analysed **50,000+ SAP Hybris transactions** to isolate performance bottlenecks.
- Joined as a QA automation intern in Jan 2019 (North Dakota Health and Human Services website, POEDI XML testing); converted to analyst after six months. 2 Spot Awards and an Applause Award.

Tech: Selenium, Katalon Studio, Python, JavaScript, Jenkins, Salesforce, SAP Hybris, Tableau, JIRA, Git

AutoKorrekt · Lead Frontend Developer (EdTech startup MVP) *May 2023 – Jan 2024*

- Led the frontend of an AI answer-evaluation platform for teachers: PDF processing with coordinate-based extraction over **AWS Textract**, a bilingual English/German feedback interface on Amazon Comprehend, and the ML data flow across S3, RDS and SageMaker; scaled to **1,000+ concurrent submissions**.

EDUCATION

Post Graduate Program in AI & Machine Learning · UT Austin McCombs & Great Lakes *Mar 2025*

Rank 1 in batch · GPA 4.09/5

Bachelor of Technology, Information Technology · KIIT University, Bhubaneswar *2015 – 2019*

CGPA 7.17/10

ACHIEVEMENTS & CERTIFICATIONS

- **Provisional patent filed**, India, App. No. 202631090789 (Jul 2026), covering the SMT decision and verification architecture.
- **WEEX AI Wars I: #1 on the preliminary leaderboard** of 230+ teams, then one of 37 finalists in the live AI trading competition; featured in WEEX's coverage and two live AMAs.
- **Winner, Best DeFi Application**, OpenServ × Hack2skill hackathon; runner-up, Best Website Application.
- **Top 101 of 1,500+ teams**, Google Cloud Gen AI Academy APAC 2026 Cohort 1 hackathon (SmartDesk).
- **Google Cloud Gen AI Academy** 2025 (ID 2025H2S04GENAI-A01917) and APAC 2026 Cohorts 1, 2 and 3; Gen AI Exchange Hackathon 2025; Agentic AI Day; Asha AI Hackathon 2025.
- **CPSAT**, Certified Professional Selenium Automation Testing, 2022. All credentials on Credly and Google Skills.