

AMAN JAIN

Applied Scientist | AI Engineer

amanj3335@gmail.com | +918769128326 | [amanjain.codes](#) | [linkedin.com/in/stelios357](#) | [github.com/stelios357](#)

Applied Scientist & AI Engineer specialising in production recommender systems and agentic LLM platforms. Built Cars24's search and recommendation stack serving 1M+ monthly sessions (+150% Click Recall@50, +16% buyer conversion). Currently designing a SEBI-regulated multi-agent trading system with deterministic DSL-gated execution, OPA/Rego compliance, and >90% frontier-model cost reduction. Published in Springer LNAI (JURISIN 2022) on domain-adapted legal NLP.

WORK EXPERIENCE

Founding AI Engineer

KotiLabs (Aagman AI | Pre-seed | 10-person team) · Bengaluru, KA

July 2025 - Present

Agentic AI Systems: Designed a production multi-agent AI platform from scratch - orchestration, hybrid RAG, deterministic guardrails, and a two-tier inference optimisation layer in a SEBI-regulated execution environment.

- Built an **end-to-end voice-to-action agentic pipeline**: WebSocket audio streaming → **Mastra Voice** transcription → **fine-tuned intent classifier <50ms** → DAG orchestration → JSON DSL translation → **OPA** compliance → tool execution → Explainability.
- Designed a **two-layer AI safety architecture** between LLM output and tool execution: **structural safety via a JSON DSL+ semantic safety via OPA/Rego policies**; provides a deterministic, auditable guarantee of zero hallucinated tool calls.
- **Reduced per-query inference cost 10x (\$0.10 → \$0.01)** by building a two-tier routing layer (Redis Cache + SLM intent classifier) and golden-test evaluation harness - cutting frontier-model calls by >90% without sacrificing accuracy on the golden suite.
- Built a **real-time data ingestion pipeline**: WebSocket→tick-to-OHLCV→Redis Pub/Sub event bus→ClickHouse + TimescaleDB.
- **Built a VectorBT-backed backtesting engine** with **deterministic signal evaluation**, intrabar SL/TP logic entry/exit timestamps, and Sharpe/drawdown metrics across all executions.
- **Mentored 2 engineers** on deterministic execution design, OPA policy authoring, and CI/golden-test methodology.

Data Scientist - Ranking, Personalisation & Pricing

Cars24 · Gurugram, HR

Dec 2023 - July 2025

Recommender Engine, Personalisation & Pricing: Rebuilt the core search and discovery engine for India's largest used-car marketplace - owning user-level recommendations, pricing, and similar-vehicle ranking across 1M+ monthly sessions.

- Rebuilt the user-level recommendation engine using a **Two-Tower retrieval model** with ANN (HNSW) index; **improved Click Recall@50 by 150% (12% → 30%)**, deployed to **100% of 1M+** monthly sessions with **servicing at <100ms p99 latency**.
- **Expanded personalised coverage from 65% → 100%** by **solving cold-start for 35% of new users** - using search query tokens and filter-selection impressions as implicit affinity proxies.
- **Increased buyer conversion by 16% (live A/B) and clicks per vehicle by 7% via** default recommendations.
- **Improved similar-vehicle ranking by 12% across 1M+ listing views** via a hybrid recommender (collaborative + content).
- **Improved view-to-buy intent (V2Bi) by 21%** (in live A/B) by deploying a LightGBM re-ranker; validated offline via NDCG.
- Mentored 2 junior analysts on A/B test validation, dashboarding and out-of-time evaluation.

Data Scientist

6sense · Bengaluru, KA (Remote work concurrent with M.Tech research: May 2021 - May 2023 | Full-time: May 2023 - Oct 2023)

- **Engineered a CV logo-similarity micro-service processing 400M+ pairwise comparisons** on AWS S3/Hive; **reduced false positives by 4%**, eliminating **~16M errors**.
- **Implemented BERT/RoBERTa multi-label classifier pipeline in Togylop** (internal NLP library), enabling transformer fine-tuning workflows across 3 ML teams.

RESEARCH PUBLICATION

[Comparative Study of BERT and Legal-BERT for Predicting Indian Legal Case Judgements](#)

Springer LNAI 2022

EDUCATION

The LNM Institute of Information Technology (LNMIT)

Aug 2018 - April 2023

- B.Tech + M.Tech Dual | CSE | Major:Data Science| SPI: 9.25

CORE COMPETENCIES

ML SYSTEMS: Recommendation, Pricing, Two-Tower, GBDT, Uplift Modelling, Cold-Start, A/B Testing, Feature Stores, PSI Drift.

LLM & AGENTIC AI: Multi-Agent Orchestration, RAG (BM25+HNSW+RRF), OPA/Rego, Semantic Caching, DSL-Gated Execution, LoRA/PEFT, LLM Evaluation (RAGAS), Transformer Fine-tuning, pgvector.

INFRASTRUCTURE: ClickHouse, PostgreSQL, Redis, Snowflake, AWS S3, Apache Hive, FastAPI, Docker, WebSocket Pipelines.

LANGUAGES & FRAMEWORKS: Python, SQL, PyTorch, LightGBM, Scikit-learn, HuggingFace Transformers, VectorBT.

EXPERIMENTATION: Out-of-Time Validation, Power Analysis, Bayesian A/B, MTC Correction, SRM Detection.