# Yash Watwani

Bengaluru · yashwatwani28@gmail.com · linkedin.com/in/yashwatwani · github.com/yashwatwani · yashwatwani.com

## SUMMARY

Lead Data Scientist and AI engineer with 8 years building production ML, deep learning and generative AI systems across marketing, banking, insurance and restaurant tech. At Pyvot I build LLM products end to end: agent design, evaluation, cost optimisation, privacy and AWS deployment. Before that at Brainlabs I built an agentic marketing-mix-modelling platform for Google, Bayesian MMMs and forecasting, all shipped to production.

## EXPERIENCE

**Lead Data Scientist · Pyvot** **Sep 2026 – now**

- **Pia, Mynt's AI analyst (owner, end to end).** Restaurant owners ask about payouts in 5 languages and get real figures with an auto-chosen chart. LangGraph on FastAPI + PostgreSQL (AWS): triage, a self-correcting tool-calling SQL writer, and an embedding shortcut over ~1,300 common questions.
- **Hallucination control.** Every number, including derived % changes and totals, is verified against the query results: unverified answers 46% → 4%. Nightly grading on 109 real questions, ~2,960 automated tests, a 324-question model-comparison bank.
- **Cost and reliability.** ~95% cheaper per step on Claude Haiku, one-hour prompt caching, one interface over 6 LLM vendors with per-step failover, OpenTelemetry tracing.
- **Security.** Tenant isolation inside PostgreSQL, a SQL allowlist parser, and a prompt-injection screen that caught 85/85 attacks.
- **Ethos Menu AI.** Reads a restaurant's menu (PDF, photo or sheet), re-engineers it against a 90-rule rulebook and returns upload-ready Zomato and Petpooja files.
- **Product analytics.** First-party, privacy-safe event tracking for the website, dashboard and app on AWS ECS + Aurora + Metabase.

**Senior Data Scientist · Brainlabs** **Mar 2023 – Jul 2026**

- **Agentic MMM platform for Google.** Multi-agent system on Google ADK + Vertex AI that runs marketing-mix workflows end to end on Cloud Run.
- **Bayesian marketing mix models** in PyMC and Meridian for FMCG and beauty brands: channel ROI with uncertainty, response curves, budget scenarios.
- **Enrolment forecasting** on Databricks with LightGBM/CatBoost ensembles, beating budget baselines on WMAPE.
- **Automated weekly reporting** from 7 BigQuery tables to KPI PDFs across 4 markets, replacing a manual process.

**Data Scientist · AmyGB.ai** **Jul 2021 – Mar 2023**

- LSTM email classifiers for banking and insurance, and a German-language extraction pipeline.
- Document intelligence: blur detection, document clustering, spaCy + AWS Textract key-value extraction.

**Data Scientist · Atos** **Aug 2018 – Jul 2021**

- Signature detection and verification with Mask-RCNN and Siamese networks for bank fraud teams.
- A COVID-19 research search engine over 77,000+ papers with BERT and USE semantic search.
- County-level US surgical-supply demand forecasting during the pandemic peak.

## SKILLS

| | |
|---|---|
| **GenAI / LLMs** | Agents, text-to-SQL, tool calling, RAG, LLM evals, prompt caching, prompt-injection defence |
| **LLM frameworks** | LangGraph, LangChain, Google ADK, LlamaIndex, Arize Phoenix, LangSmith |
| **ML / AI** | Deep learning, NLP, computer vision (YOLO, Mask-RCNN, SAM), forecasting, marketing mix modelling |
| **Statistics** | Bayesian modelling, regression, clustering, cohort analysis, backtesting |
| **ML frameworks** | PyTorch, TensorFlow, Keras, scikit-learn, LightGBM, CatBoost, XGBoost, PyMC, Meridian, Hugging Face |
| **Code / APIs** | Python, SQL, TypeScript, FastAPI, Flask, Node.js |
| **Data** | PostgreSQL, BigQuery, Metabase, Pinecone, ChromaDB, FAISS, Milvus |
| **Cloud / MLOps** | AWS, GCP, Databricks, Azure, Docker, GitHub Actions, MLflow, OpenTelemetry |

## EDUCATION

**B.E. Computer Science · BMSIT, Bengaluru** **2014 – 2018**