






# Davis Jacob Thomas

Machine Learning Engineer · AI Platform

 davisjacobthomas@gmail.com  
 +34 637 122 406  
 linkedin.com/in/davis-jacob-thomas  
 djtom98.github.io  
 Barcelona, Spain | Visa: HQP

## PROFILE

MLE with 6 years of experience designing, building and operating production-grade ML platforms, MLOps pipelines, and LLM/agent systems. Author and maintainer of multiple internal Python libraries spanning ML utilities, GenAI, Knowledge Graphs, and data access, developed with open-source best practices (semantic versioning, CI/CD, conventional commits, test coverage). Proven track record in Kubernetes-native deployments, observability, model serving, and RAG architectures across scientific and enterprise environments. Passionate about leveraging AI for public good.

## PROFESSIONAL EXPERIENCE

### Machine Learning Engineer · AI Platform | *ALLY Labs*

May 2024 – Present

- ▶ Maintain and improve the organisation's shared ML/AI platform on Kubernetes (EKS), eliminating per-team infrastructure duplication and cutting model deployment lead time with shared GHA CI/CD workflows; cluster and resource management with Helm, FluxCD and Terragrunt.
- ▶ Engineered a standardised FastAPI model-serving framework adopted org-wide - factory pattern, request/response middleware, health checks, Datadog APM tracing, and multi-tenant MLflow model loading with Locust load tests; led full migration off the legacy serving stack.
- ▶ Authored and maintain an internal scikit-learn-compatible ML utilities library covering MRMR/SHAP feature selection, Optuna hyperparameter tuning, statistical drift detection, and MLflow lifecycle management; enforces governance with 80%+ test coverage, pre-commit hooks, and model versioning across teams.
- ▶ Designed a Knowledge Graph platform from greenfield to production on Neo4j: NLP-driven entity extraction, matching and merging pipelines, GraphRAG, LLM exploration agents, Pydantic data models, source/validator layers, and access surface.
- ▶ Built and maintain a shared GenAI library (embeddings, LLMs, vector DBs, Langfuse integration) and a unified data access layer (PostgreSQL, S3, Azure Blob, OpenSearch, Slack, DuckDB/DuckLake), enabling 10+ services to share a single tested data surface.
- ▶ Delivered production LLM and agent systems with measurable quality gates: hybrid + vector RAG on OpenSearch with layout-aware document understanding and production evals; real-time ReAct agents via PydanticAI + MCP servers; a unified LLM gateway (OpenAI, Bedrock) with quota management and failovers.
- ▶ Instrumented end-to-end MLOps observability: OpenTelemetry traces, Langfuse LLM cost and latency dashboards, drift monitoring.
- ▶ Built semantically enriched dbt feature layers with full lineage, Alembic data contracts, and Airflow orchestration.
- ▶ Delivered a full LLM fine-tuning pipeline: mining and LLM-cleaning production chatbot interactions from PostgreSQL/ClickHouse, LoRA fine-tuning on, GGUF quantization, and artifact publishing to S3 culminating in a purpose-built autocomplete model served via a vLLM inference engine (tensor-parallel, configurable GPU memory utilisation) behind a FastAPI endpoint on Kubernetes.

### Data Scientist · Projects Officer | *Institute for Capacity Development, IMF*

Nov 2023 – Apr 2024 | Apr 2025 – Jun 2025

- ▶ Quantified causal effects of IMF interventions on member-state conflict under a structural causal framework; built a RAG pipeline over MONA policy documents to accelerate evidence retrieval for economists.
- ▶ Designed a Human-in-the-Loop annotation pipeline (Label Studio + Few-Shot Learning) that reduced manual labelling effort while maintaining research-grade label quality; a reproducible scientific ML workflow from data to model.

### Data Analyst | *Gameloft, Barcelona*

Jul 2023 – Dec 2023

- ▶ Shipped funnel dashboards tracking CTR, DAU and RPU across titles; designed A/B testing frameworks for seasonal campaigns that measurably lifted player retention and monetisation metrics.

## Graduate Researcher | [Novartis & Barcelona School of Economics](#)

Apr 2023 – Jul 2023

- ▶ Conceived and implemented a novel inductive GraphSAGE variant (GNN/GCN/GAT) to learn embeddings from Knowledge Graphs extracted from earnings-call transcripts; used the embeddings to estimate cumulative abnormal returns in the 30-day post-event window.

## Data Engineer | [IQVIA](#)

Apr 2021 – Aug 2022

- ▶ Replaced a legacy Informatica pipeline with Spark SQL-based automated report generation, cutting processing time significantly; built ELT pipelines in Snowflake integrating pharmaceutical sales data with SAP (Star Schema); implemented GDPR-compliant data anonymisation for HCPs and onboarded new market data models in Reltio MDM.

## EDUCATION

---

### MSc Data Science | [Barcelona School of Economics](#)

2022 – 2023 | GPA 8.62 / 10

*Causal Inference · Bayesian Statistics · Deep Learning (CNN, Transformers, Knowledge Distillation) · NLP / LLMs · Graph Neural Networks · Probabilistic Inference · Network Dynamics*

### B.Tech Mechanical Engineering | [Mar Athanasius College of Engineering, Kerala](#)

2017 – 2021 | CGPA 9.13 / 10

## TECHNICAL SKILLS

---

<b>ML / AI</b>	PyTorch, Scikit-learn, Hugging Face, Knowledge Distillation, GNN (GraphSAGE, GCN, GAT), SHAP, Optuna
<b>LLM &amp; Agents</b>	PydanticAI, LangChain, MCP / FastMCP, RAG (OpenSearch hybrid+vector), Knowledge Graphs (Neo4j), Langfuse
<b>MLOps</b>	MLflow (registry, serving, migration), OpenTelemetry, Langfuse, Drift detection, Model lifecycle mgmt
<b>Cloud Native</b>	Kubernetes (EKS), Helm, FluxCD, Terragrunt, Docker, AWS (Bedrock, Textract, S3), GitHub Actions, ECR
<b>Data Platform</b>	dbt (lineage, contracts), Airflow, Snowflake, PostgreSQL, Redshift, DuckDB, Alembic, DVC
<b>Model Serving</b>	FastAPI (factory pattern, middleware, health checks), Datadog APM, Locust load testing, MLflow multi-tenant registry
<b>Languages</b>	Python (MITx certified), R, SQL, Bash; English (fluent), Spanish (learning)
<b>Open Source</b>	Active in upstream LLM/agent ecosystem; GenAI Opentelemetry SIG; evaluation of emerging tools before platform adoption

## SELECTED ACHIEVEMENTS

---

- ▶ TEDxMACE 2021: Licensee & Organiser: led a 35-person team to deliver a national conference with speakers across India.
- ▶ SAE (Society of Automotive Engineers) Six Sigma Convention: National-level winner; applied DMAIC methodology to a complex case study.
- ▶ SAE Autonomous Vehicle Challenge: National-level winner; developed ADAS-based vehicle/pedestrian detection in MATLAB.