

Humberto Lomeu

AI Engineer | 📍 York, UK

✉ hcslomeu@gmail.com 📞 +44 7988 757081 🔗 [linkedin.com/in/humbertolomeu](https://www.linkedin.com/in/humbertolomeu) 🌐 humbertolomeu.com

Professional Summary

Hands-on AI Engineer with 9+ years of experience designing, building, and delivering production Machine Learning, Deep Learning, and Generative AI systems for enterprise clients in financial services, insurance, healthcare, and renewable energy. Specialises in agentic LLM applications using LangGraph, LangChain, and Retrieval-Augmented Generation (RAG) over vector databases, integrating with OpenAI and Anthropic APIs and deploying via Python, FastAPI, and AWS. Combines strong end-to-end engineering (CI/CD, testing, containerisation, observability) with direct client delivery and consulting experience from EY and a Brazilian AI startup, where I led cross-functional teams of up to 18 and mentored five engineers into technical leadership. MSc in Computer Science (Deep Learning specialisation). Based in York, UK with full right to work; flexible for hybrid working and travel.

Technical Skills

AI and Machine Learning: Large Language Models (LLMs), Generative AI, Retrieval-Augmented Generation (RAG), Agentic AI, Prompt Engineering, OpenAI API, Anthropic API, LangChain, LangGraph, LlamaIndex, Hugging Face Transformers, PyTorch, Deep Learning, NLP, Feature Engineering, Model Evaluation, MLOps

Vector Databases and Observability: Qdrant, Pinecone, FAISS, pgvector, LangSmith, LangFuse, LogFire

Software and API Engineering: Python, FastAPI, Pydantic, httpx, Pytest, Ruff, SQL, PostgreSQL, REST API development, Git, GitHub Actions, CI/CD, Docker, uv, Poetry, Test-Driven Development

Cloud and Data Engineering: AWS (EC2, S3, Lambda, SQS, Bedrock, Lightsail), Supabase, Apache Airflow, dbt, ETL/ELT pipelines, Medallion Architecture, Databricks, PySpark, HPC (Linux)

AI-Assisted Development: Claude Code (custom Skills, Hooks, MCP servers), Codex, Gemini CLI, Cursor, spec-driven development, multi-agent workflows

Professional Experience

Deep Learning Researcher, COPPETEC Foundation / TotalEnergies Mar 2024 – Feb 2026
Brazil (Remote)

- Developed short-term offshore wind speed forecasting models (ConvLSTM, PredRNN) in PyTorch, reducing inference time from 18 hours to under 10 minutes (~100x speedup) versus the incumbent numerical weather model
- Engineered end-to-end ML pipelines (data ingestion, tensor preprocessing, training with hyperparameter tuning, evaluation) on HPC infrastructure, processing WRF numerical weather model outputs and SAR satellite radar data
- Applied software engineering best practices throughout – Git version control, modular architecture, reproducible experiments, and clear documentation – enabling handover to the client's internal team
- Presented research findings directly to TotalEnergies stakeholders, translating model performance into actionable renewable energy investment decisions

Data Science and AI Manager, Kunumi (AI Consultancy) Oct 2020 – Jul 2024
Brazil (Remote)

- Led applied AI consulting delivery for enterprise clients across Healthcare, Retail, and Infrastructure, owning solutions end-to-end and managing cross-functional teams of up to 18 engineers, data scientists, and project coordinators
- Owned a portfolio of three top-tier Health Insurance providers, delivering 10+ Machine Learning and Deep Learning projects that consistently generated multi-million-dollar annual revenue – reaching up to 40% of company revenue at peak
- Built production AI solutions hands-on – conversational AI, clustering, predictive modelling, and regression pipelines – in Python, SQL, AWS, and Databricks; designed solution architectures and presented them to senior client stakeholders
- Contributed to company-level decisions alongside C-level executives: OKR and KPI definition, post-M&A restructuring using team topologies, and recruiting (sourced and hired ~30% of the engineering team)
- Mentored five junior engineers into technical leadership roles

Senior Consultant, Data and Analytics, EY (Ernst & Young) Jul 2017 – Oct 2020
Rio de Janeiro, Brazil

- Delivered Machine Learning and advanced analytics projects for highly regulated Insurance, Banking, and Pension Fund clients under Agile delivery, in client-facing consulting teams
- Built an NLP text classification pipeline (SpaCy) with a multidisciplinary team including legal counsel at a major Brazilian pension fund, unlocking \$250 million in assets previously frozen in judicial accounts
- Developed PySpark and Databricks data pipelines processing customer transactions for a Brazilian retail bank serving 110 million customers; built executive Power BI dashboards consumed by C-level stakeholders
- Led a data privacy compliance workstream for an international investment bank in Brazil, aligning engineering processes with Brazil's LGPD (equivalent to GDPR) ahead of the enforcement deadline

Selected Publications

Lomeu, H., Evsukoff, A. G., Assad, L. P. F. (2025). Deep Learning for Offshore Wind Intensity Forecasting in the South Atlantic: A ConvLSTM Approach Using ERA5 Data. *CILAMCE 2025: XLVI Ibero-Latin American Congress on Computational Methods in Engineering*. DOI: 10.55592/cilamce2025.v5i.14019 – **sole first author**

Zuin, G., Lomeu, H., Barros, G., Barbosa, M., Lima, G., Vançan, N. (2023). A Modified Louvain Approach for Medical Community Detection using Geographic Data. *IEEE CBMS 2023*, L'Aquila, Italy. DOI: 10.1109/CBMS58004.2023.00206

Zuin, G., Parreiras, L., Melo, L., Barros, G., Lomeu, H., Melo, B. (2023). An Ensemble Approach for Inconsistency Detection in Medical Bills: A Case Study. *IEEE CBMS 2023*, L'Aquila, Italy. DOI: 10.1109/CBMS58004.2023.00281 – billing anomaly ranking system increased identified-anomaly rate **2.4x**

Selected Projects

End-to-end portfolio projects (in active development) – Python, FastAPI, AWS, GenAI components. Live links and architecture write-ups at humbertolomeu.com:

- **alpha-whale** – end-to-end AWS portfolio project demonstrating production engineering patterns
- **tfl-monitor** – end-to-end AWS portfolio project applied to Transport for London open data

Education

MSc Computer Science , COPPE/UFRJ, Rio de Janeiro, Brazil	2024 – 2026
Dissertation: Deep Learning for Wind Speed Forecasting – A Case Study in the South Atlantic with ERA5 Reanalysis Data	
BSc Civil Engineering , Universidade Federal de Juiz de Fora, Brazil	2011 – 2017
Exchange Programme: Nottingham Trent University, UK (2014 – 2015)	

Languages

Portuguese (Native), English (Fluent), Spanish (Intermediate)