

MAROŠ JANČO

Senior AI / ML Engineer

marosjanco@gmail.com • +421 951 454 277 • [linkedin.com/in/marosjanco](https://www.linkedin.com/in/marosjanco) • github.com/marosjanco • Slovakia (remote, UK/EU work history)

PROFILE

Generalist AI engineer with 9+ years shipping applied ML, NLP and LLM products end-to-end — currently leading prompts, evals, RAG and agentic systems at Upheal. Strong product, customer and design instincts from owning features and a company.

EXPERIENCE

Senior AI Engineer, Upheal, New York (remote)

Nov 2023 – Present

Best Startup Award. AI clinical-documentation platform for therapists.

- Own prompts, models and quality for AI-generated progress notes (core product) across 250+ documentation sections; shipped 100+ production releases across multiple model generations without quality regression. Mentored new hires; reported AI roadmap and cost trade-offs directly to founders.
- Lowered AI costs by 50% while improving quality. Built an LLM-as-judge eval framework on Langfuse (datasets, eval runs, trace-level flow debugging) plus an A/B-testing pipeline; used to gate releases, catch regressions and shorten the prompt iteration cycle.
- Productionised RAG, prompt orchestration and agentic patterns across Gemini, Claude, GPT-* and Llama; benchmarked Vertex AI, AWS Bedrock, Azure OpenAI and Anthropic continuously, routing traffic by use case rather than committing to a single provider.
- Automated a Claude-Agent-SDK customer-support agent (internal-tool integration, hardened prompts and guardrails).
- Ran LLM observability and MLOps (Langfuse, Grafana, BetterStack, AWS CloudWatch, SNS, SQS, Azure, Google Cloud, Vertex AI, Superset); lifted team velocity by integrating Claude Code into daily workflow, improving CI/CD, and running internal sessions on modern RAG / prompting / agentic best practices.

AI Product Consultant, ATEN Consult, London

Aug – Oct 2023

- Led ideation and planning of an expert-validated knowledge-graph platform; reshaped business model, scope, financials and technical architecture for a successful UK Smart Funding submission after 12 months of prior poor-quality work.

Senior Data Scientist, Cervest, London

Jun 2021 – Jun 2023

Global Impact 50 Award 2023. Physical climate risk for enterprise assets.

- Built from scratch the first physical-climate-impact platform assessing combined-hazard damages on assets and portfolios worldwide; shipped a financial-impact model predicting stock-price shocks from adverse climate events.
- Led a global wildfire-damage prediction product end-to-end and built a company → subsidiary → asset framework for SP500 and FTSE100 firms; optimised distributed pipelines over TBs of geospatial, satellite and financial data (PySpark, Databricks, Geopandas, Xarray, k8s).

Machine Learning Researcher, BNP Paribas, London

Jun 2016 – Jun 2021

Global investment-bank R&D lab.

- Built a history-augmented collaborative-filtering recommender for client document recommendations (Amazon S3, Spark, Dask); weekly chatbot recommendations save thousands of clients hours of search across thousands of documents.
- Designed and shipped an NLP entity-extraction system from scratch — combined Random Forests, GBT, Kneser-Ney n-gram LMs, Naive Bayes and Word2Vec into one interpretable, accurate classifier; later scaled with LSTM- and BERT-based architectures (Keras, PyTorch). A dozen+ resulting NER models now automate chat-to-price FX execution worldwide on a daily basis.

SELECTED SIDE PROJECTS

lexomat.sk — solo-founded AI legal-research chat for Slovak and Czech law, indexing millions of documents; drove product, design, agentic AI flows, pricing, GDPR and infra end-to-end (one contractor on delivery). Stack: Next.js / TypeScript, Python / FastAPI, Supabase + PGroonga, Vertex AI.

PayToEat.sk — co-founded and ran a food-delivery app (2019); led product, ops and growth across Slovak towns. Pre-AI era; the business still serves customers today in one Slovak town.

SKILLS

AI / LLMs	Prompt engineering, RAG, agentic systems, LLM-as-judge & evals, A/B testing; Langfuse; Gemini, Claude (Agent SDK), GPT-*, Llama; Vertex AI, Bedrock, Azure OpenAI, Anthropic API; LlamaIndex, Pydantic, tool use
Languages	Python, TypeScript, SQL, Bash; familiar with Julia, R, C, Fortran, Dart
Backend / Frontend	FastAPI, Next.js, React, Node.js, Supabase / Postgres (PGroonga), PySpark, Pandas, Polars, PyTorch, scikit-learn, XGBoost, LightGBM, Spacy, NLTK, Pytest
Infra & Tooling	AWS, GCP, Docker, Kubernetes, GitHub Actions, Argo, DVC, UV / Poetry; Grafana, BetterStack, CloudWatch; Git, Claude Code, Cursor, Databricks, Linear / Jira, Figma, Agile, $\LaTeX$

EDUCATION

MSc Machine Learning, University College London

2016 – 2017

Highlights: Applied ML (92%), Statistical NLP (80%), Numerical Optimisation (84%). Databricks Certified ML Associate (2023).

BSc (Hons) Mathematics with Statistics for Finance, First Class, Imperial College London

2013 – 2016

Languages: English, Slovak, Czech (fluent); German (beginner). **Interests:** investing, entrepreneurship, hiking, badminton, chess, photography.