

Diogo de Bastos

Lisbon, Portugal | diogodebastos.com | diogodebastos18@gmail.com | linkedin.com/in/diogodebastos | github.com/diogodebastos | [Google Scholar](https://scholar.google.com/citations?user=diogodebastos)

[try talking with my CV](#)

Summary

Lead AI/ML Engineer and Manager in EY-Parthenon's AI Hub, with a Ph.D. in Physics from the CMS experiment at CERN and hands-on machine learning since 2018 across research and industry. I build and run the firm's Model Context Protocol (MCP) server and production AI agents, and I work directly with clients, from pitch demos and scoping to onsite delivery. Before EY-Parthenon, I spent nearly four years at EDP shipping forecasting and trading-optimisation models into production and building LLM assistants for trading and risk teams. I pair rigorous statistics (causal inference, Bayesian modelling) with production engineering, and I build on Cloudflare's developer platform (Workers, Workers AI, Durable Objects, D1) in my own time.

Work experience

EY-Parthenon

Lisbon, 2026/05 - present | Manager, Lead AI/ML Engineer, AI Hub (Europe West)

- Hands-on technical lead in EY-Parthenon's Europe West AI Hub, building agentic systems, LLM integrations and data pipelines at platform level: infrastructure that other teams and client applications depend on.
- Built and rolled out the firm's Model Context Protocol (MCP) server, taking it from prototype to a supported internal service with 30 users: 31 production tools (governed SQL data retrieval, document and knowledge search, financial-model automation, deliverable generation) served over streamable HTTP with OAuth / Entra ID single sign-on, used from Claude, Cursor, Codex and GitHub Copilot.
- Run the server on Databricks Apps with Unity Catalog SQL warehouses, Vector Search, AI Gateway embeddings and MLflow usage tracing; IDE clients connect through an authenticated MCP proxy.
- Client-facing on engagements with a global automotive manufacturer and a multinational consumer-goods company, including onsite work: scoping requirements and presenting solutions to technical and business stakeholders.
- Pre-sales: build pitch demos and proofs of concept, present them in client pitches, and equip other EY consultants to take AI offerings to their clients.
- Designed and deployed production agents that automate research, document analysis, and client deliverables, including tool use, retrieval, evaluation harnesses, and guardrails for reliability and cost control.
- Applied causal inference (difference-in-differences, synthetic control, uplift modelling) and Bayesian models for sparse-data estimation and forecasting, giving clients defensible effect estimates with credible intervals rather than point estimates.
- Tech lead for AI-first teams of 3-5 engineers: technical direction, code and model review, interviewing, onboarding and mentoring, and delivery from scoping to production handover.

EDP

Lisbon, 2025/09 - 2026/05 | Market Modeling Senior Specialist, Global Energy Management Trading Modelling

- Applied machine learning to systematic trading: production code, data pipelines, model deployment and evaluation, and the design of new trading strategies.

Lisbon, 2022/09 - 2025/08 | Senior Data Scientist & Quantitative Analyst

- Reduced short-term price forecast error from 15% to 5% by combining LSTM, CNN, and Temporal Fusion Transformer models implemented in Python with PyTorch and SQL-backed data pipelines.
- Designed portfolio optimisation and algorithmic trading frameworks that improved gross margin by 30%, integrating ML outputs into decision workflows for trading and risk teams.

- Migrated analytics from MATLAB and R to a Python ecosystem, formalising code efficiency, scalability, and reproducibility for globally distributed stakeholders.
- Deployed predictive services and analytical applications on Databricks to meet availability, monitoring, and scalability requirements.
- Built LLM-powered assistants using Azure OpenAI, Llama 2, Retrieval Augmented Generation, Weaviate, and LangChain to automate report generation and data interpretation, including PoCs for report automation and research summarisation. Cut manual review time by >40%.
- Mentored colleagues on MLOps practices, supervised a master's thesis on deep learning for price forecasting, and acted as product lead for ML, optimisation, and trading initiatives.

Qold

Coimbra, 2016/10 - 2017/06 | Co-founder, Hardware & Business Development

- Co-founded a cold-chain IoT startup spun out of my M.Sc. thesis R&D at [Whitesmith](#), managing product strategy, client acquisition, and data-driven validation of market needs.
- Built sensor-to-cloud monitoring across 24 devices, integrating wireless communications (2.1 GHz/433 MHz), embedded prototyping (Arduino, Raspberry Pi), and analytics (SQLite, MATLAB).
- Designed low-power PCBs with Altium and coordinated with engineers and customers to deliver reliable hardware deployments.

[jeKnowledge](#)

Coimbra, 2015/02 - 2016/02 | Chief Executive and Innovation Officer

- Expanded the non-profit tech organisation from 10 to 30 members by redesigning recruitment and leadership development.
- Directed rebranding, launched a summer academy focused on programming and hardware, and created networking events linking academia with startups.
- Oversaw hackathons and community programs, honing stakeholder communication and execution under tight timelines.

Coimbra, 2014/01 - 2015/01 | Communications Manager

Research

[CERN](#)

Lisbon, 2018/01 - 2023/07 | Ph.D. Researcher, CMS Experiment (LIP Lisbon)

- Researched high energy physics and quantum machine learning in the CMS experiment: advanced supersymmetry searches by building C++ frameworks, boosted decision trees (ROOT TMVA), and TensorFlow neural networks to classify proton-proton collision data.
 - Performed Bayesian statistical analyses to set stringent limits on top squark pair production cross sections at low masses, managing class imbalance and detector noise.
 - Experimented with Variational Quantum Classifiers and Quantum GANs (Python, PennyLane, Qiskit) to accelerate Monte Carlo data augmentation.
 - Took CMS computing-operations shifts, monitoring grid jobs and data transfers across the Worldwide LHC Computing Grid, ticketing issues and following incidents through to resolution, and supported the Lisbon WLCG Tier-2 site.
 - Tested and calibrated ASICs for the CMS Timing Detector upgrade.
 - Mentored junior researchers and organised outreach events to promote STEM education.
-

Projects

AI-native builds, shipped with Claude Code on Cloudflare's developer platform.

- [binding-doctor](#): open-source CLI and MCP server that reconciles Cloudflare bindings across code, Wrangler config, and the live account (D1, KV, R2, Queues, Vectorize), with an idempotent diff / plan / apply loop over the Cloudflare API.
- [AI Bot Activity](#): 3D globe of AI-crawler traffic by country from the Cloudflare Radar API, with a Pages Functions proxy that keeps the API token server-side and caches responses.
- [Velvet Blum](#): a social network with no algorithm and no ads. Workers API on D1, KV, and R2; Durable Objects (SQLite) with WebSocket fan-out; cron triggers; end-to-end encrypted DMs (X25519, Argon2id, libsodium).
- [LLM Circuits](#): Workers AI playground that wires LLMs like circuit components, with AI Gateway caching, rate limiting, and per-call cost telemetry.
- [FisicAI](#): open-source agentic harness for high energy physics that searches INSPIRE-HEP and arXiv, downloads published HEPData likelihoods, and reruns pyhf inference to reproduce LHC results.
- [edge-cavai](#): personal site with a CV chatbot and blog on Cloudflare Workers (Hono, TypeScript), deployed by GitHub Actions with a pre-rendered PDF CV.

Education

- Ph.D. in Physics, Instituto Superior Técnico, Universidade de Lisboa. Thesis: "Search for top squarks in the four-body decay mode with single lepton final states in proton-proton collisions at the Large Hadron Collider" (Pass with Distinction)
- M.Sc. in Physics Engineering, University of Coimbra. Thesis: "Automated monitoring and diagnosis of cold chains" (18/20)

Publications

- 2023/06/12 - The CMS Collaboration, "Search for top squarks in the four-body decay mode with single lepton final states in proton-proton collisions at the Large Hadron Collider" [JHEP06\(2023\)060](#)
- Bastos, D. "Using Variational Quantum Algorithms and Quantum Generative Adversarial Networks for Supersymmetry in High-Energy Physics" (internal paper)
- The CMS Collaboration, "Experimental characterization of the BTL Front-end Board based on TOFHIR1" (internal note)

Skills

- Cloudflare developer platform: Workers, Workers AI, AI Gateway, Pages and Pages Functions, D1, KV, R2, Durable Objects, Queues, Cron Triggers, Radar API, Wrangler
- AI & LLMs: agentic systems, Model Context Protocol (MCP, FastMCP), Retrieval Augmented Generation, evaluation harnesses and guardrails, LangChain, Azure OpenAI, Llama 2, Weaviate, Claude Code, Codex, Cursor, speech-to-text, translation, text-to-speech
- Machine learning & statistics: PyTorch, TensorFlow, Keras, scikit-learn, time-series forecasting (LSTM, CNN, Temporal Fusion Transformer), boosted decision trees, genetic programming, causal inference, Bayesian analysis, quadratic programming, risk control
- Data & platforms: Databricks (Apps, Unity Catalog, Vector Search, MLflow), Azure, SQL, MySQL, Oracle, Pandas, NumPy, Matplotlib, Seaborn, data pipelines, MLOps
- Programming: Python, SQL, C++, TypeScript / JavaScript, Bash, R, MATLAB
- Delivery & communication: client scoping and onsite delivery, pre-sales demos and pitches, CI/CD with GitHub Actions, technical writing, conference speaking, mentoring
- Languages: Portuguese (native), English (fluent), Spanish (professional working proficiency)

Conferences

- 2024/09/13 - Speaker: "Predicting Spain power price with Deep Learning", 11th Annual Electricity Price Forecasting And Modelling Forum
- 2022/06/29 - Speaker on behalf of the CMS collaboration: "Searches for top squarks in compressed scenarios with the CMS experiment", The XXIX International Conference on Supersymmetry and Unification of Fundamental Interactions, Ioannina, Greece
- 2020/06/26 - Speaker: "High-performance timing detector for the HL-LHC Upgrade of the CMS experiment at CERN", 6th IDPASC /LIP PhD student Workshop, remote
- 2020/03/09 - Seminar: "Presentation to the CAT", Lisbon, Portugal
- 2020/02/15 - Speaker: "CMS: Searching for stop", Braga, Portugal
- 2019/09/25 - Invited speaker: "Distributed Computing at the CMS Experiment: From the point of view of a physicist", IBERGRID 2019, Santiago de Compostela, Spain
- 2019/07/01 - Speaker: "Search for the SUpersYmmetric partner of the top quark at the LHC with a multivariate approach", 5th IDPASC /LIP PhD student Workshop, Braga, Portugal
- 2019/04/06 - Invited mentor: "IST Masterclasses", Lisbon, Portugal
- 2019/02/11 - Invited speaker: "Particles - from the Universe to the lab", Lisbon, Portugal
- 2017/10/26 - Organizer of [LISBON.AI](#), a conference dedicated to Artificial Intelligence for 150 engineers
- 2015/07/23 - Organizer of [Summer JADE Meeting](#), a 4-day international congress for 300+ Entrepreneurs