

Marcelo Correia

Portugal (Remote) | marcelocorreia93@hotmail.com | EU Citizen | +351 911 881 497 | LinkedIn | GitHub

Summary

AI engineer with hands-on project experience in agentic LLM systems (LangChain, LangGraph, LangSmith) and 8 years of mechanical engineering experience in industry. Built an industrial-maintenance copilot with RAG, agents and an LLM-as-judge safety gate, evaluated on correctness, retrieval and safety.

Skills

Languages: Portuguese (Native), English (Fluent, C1), Spanish (Intermediate)

Programming: Python, SQL, Bash

ML/DL Frameworks: PyTorch, Hugging Face Transformers, Scikit-learn, Pandas, NumPy

LLM & Agentic Tools: LangChain, LangGraph, LangSmith, ChromaDB, Pinecone

Techniques: Retrieval-Augmented Generation (RAG), AI Agents & Tool Use, Embeddings, Reranking, Prompt Engineering, LLM Evaluation, Tracing & Observability, Cost & Safety Controls, Fine-tuning, Transfer Learning, Computer Vision

MLOps/Tools: FastAPI, Streamlit, Gradio, REST APIs, pytest, GitHub Actions CI, ruff, SQLite, Git

AI & Machine Learning Projects

The Factory Floor: Industrial Maintenance Copilot, GitHub 2026

- AI maintenance assistant for factory technicians: ask about an electric motor or variable-frequency drive, optionally attaching a photo of the defect, and get a diagnostic answer grounded in official Siemens documentation, with safety precautions first and the machine's repair history taken into account
- LangGraph agent routes each question to the right tool: semantic retrieval with reranking over 11 official manuals (3,679 pages, 12,301 chunks), a 368-entry fault-code index, each machine's maintenance history (simulated fleet of 20 assets, 77 events), and a computer-vision defect classifier (frozen ResNet18 features + logistic regression) feeding the agent's reasoning
- Evaluated on 37 scenarios against a single-call baseline: **75.7% root-cause accuracy; 86.5% retrieval hit-rate@5** with reranking (MRR 0.61 to 0.70); exact fault-code lookup returned the defining manual page for **100%** of fault-code queries vs. 18% for semantic search
- **LLM-as-judge safety gate** with a deterministic cross-check: safety-ordering failures fell from **91%** (baseline) to 33% with the agent and **6%** after the gate
- Built with production concerns in mind: decoupled service layer serving a FastAPI API and a Streamlit UI, SQLite audit trail with operator PIN authentication, per-request cost metering and daily spending caps
- **LangSmith tracing and observability** across 1,620 runs (0% exceptions, P50 latency 1.71 s); **183 automated tests** under GitHub Actions CI

Automated Customer Reviews: NLP Pipeline (Team project), GitHub 2026

- Pipeline turning 129,000 Amazon reviews into product insights: sentiment classification with fine-tuned DistilBERT and RoBERTa, product clustering with sentence embeddings + KMeans, and BART-generated comparative summaries, shipped as a Gradio app with separate consumer and seller views
- **82.97% Macro F1** vs. 49% for a zero-shot baseline (which scored 0% recall on the neutral class); diagnosed and fixed a silent embedding-truncation bug via chunked mean pooling

Fine-Grained Image Classification: Oxford-IIIT Pets (Team project), GitHub 2026

- As technical lead: fine-grained classification of 37 visually similar pet breeds: benchmarked a from-scratch CNN against transfer learning with ResNet18, ConvNeXt-Tiny and ViT-B/16; debugged shared-transform leakage between train and validation sets
- **94% test accuracy** with the best model (ConvNeXt-Tiny)

Experience

Project / Lead Equipment Engineer, Stellantis, Mangualde, PT June 2023 – Present
Via Expleo Group (2024 – Present); via ISQ & CTAG (2023 – 2024)

- Translate plant requirements into technical specifications for new production equipment on new vehicle programmes, and take each project from concept to line acceptance
- Analyse line performance and cycle-time data to build the case for equipment investments; evaluate suppliers and coordinate maintenance, production and quality teams through commissioning
- Built an Excel tracking tool covering every purchase order placed, used to follow up and update the status of each order

Process Engineer, Turning & Grinding, Renault Cacia, Aveiro, PT May 2018 – Aug 2020
Renault (2018 – 2020); via SEGULA Technologies (2022 – 2023) Sept 2022 – June 2023

- Built a dynamic Excel model that calculates cutting parameters for the grinding and turning centres, turning the machining calculation rules into a reusable algorithmic tool
- Owned two production lines for the JT4 and DB35 programmes and, later, grinding and turning equipment for gearbox differentials; used 3D metrology data to control part quality and drive process changes
- Wrote specifications (CDC) to make machines flexible across lines and ran technical reviews with CNC machining-centre suppliers

Earlier Engineering Roles Jan 2021 – Sept 2022

- **Project Engineer**, Finsa, Nelas (Dec 2021 – Sept 2022): bought and industrialised machines and full production lines; assessed investment proposals
- **Process Engineer**, Aquinos Group, Oliveirinha (Oct – Dec 2021): used LEAN and MTM time-study data to rebalance production lines and cycle times
- **Serial Manufacturing Engineer**, Eberspächer Group, Tondela (Jan – Oct 2021): owned two Renault production lines, tracking KPIs and implementing engineering changes end to end

Education

Ironhack, AI Engineering Bootcamp, 24 weeks (312 hours) Mar 2026 – Sept 2026

- Part-time programme taught in English, progressing from Python and classical machine learning to deep learning, transformers and LLM applications. Delivered five hands-on projects across ML, computer vision, NLP and agentic RAG systems, four of them in teams

Instituto Superior de Engenharia de Coimbra, Postgraduate Diploma in 2019 – 2020
Computer-Aided Industrial Manufacturing

- In partnership with Win Solutions & HAAS Automation

Instituto Superior de Engenharia de Coimbra, BSc in Mechanical Engineering 2012 – 2017