

António Morais

antoniomorais.nn@gmail.com | www.linkedin.com/in/ant%C3%B3nio-morais-ab273713b | <https://github.com/AntMorais>

WORK EXPERIENCE

Vestiaire Collective

London, United Kingdom

ML Engineer

October, 2024 – Current

Working for the Seller Experience Team in the Pricing Recommendation and Chat Moderation projects.

- Migrated the Pricing Recommendation database from **Redis** to **PostgreSQL**, with automated vacuuming and **Airflow** DAGs to clean unnecessary data. Created alerts on **Datadog** to monitor the status of the database.
- Improved efficiency of **ElasticSearch** query, resulting in a latency reduction of 25% for the price recommendation endpoint.
- Migrated Chat Moderation **BERT** model from AWS S3 to **MLflow**

Marks and Spencer

London, United Kingdom

MLOps Engineer

July, 2022 – September, 2024

Collaborated with an 8-person data science team to automate, monitor, and deploy an offer allocation system into production.

- Led the migration of model training code from **AzureML** to **Databricks**. Thoroughly analyzed costs and performance across both platforms, presenting findings to management. Achieved a **3x faster and 10x cheaper** model training pipeline by using **Ray on Databricks** to train hundreds of **LightGBM** models in parallel. Leveraged hyperparameter tuning with **Ray Tune** and **Optuna** for optimized model performance.
- Implemented **MLFlow** (with **Ray** integration) for comprehensive model logging and experiment tracking, delivering data-driven insights into performance trends across hundreds of nested training runs.
- Used **Azure Data Factory** for orchestration and **Spark** for data preprocessing and batch inference
- Deployed two containerized applications on **Azure Kubernetes Service**: a **Streamlit** dashboard for visualising data insights using the Databricks API, and a **ChromaDB** Proof-of-Concept (Client/Server) for internal LLM apps.
- Made several improvements to the CI/CD. Cut dependency installation time in the CI/CD pipeline from 40 seconds to 10 seconds by introducing caching. Boosted efficiency with multi-campaign delivery, enabling simultaneous execution of up to 3 campaigns for a faster offer allocation process. Developed Support Branches to isolate production fixes from the main development branch, ensuring stability in business critical campaign deliveries. Implemented proactive error detection in CI/CD pipelines, using automated tests to replicate the code workflow and catch issues early.
- Developed a CI/CD pipeline to track DORA metrics (Lead Time for Changes, Deployment Frequency) in real-time using **GitHub** and **New Relic APIs**, providing actionable insights into development efficiency.
- Led team efforts for robust **unit, integration and system** testing with **Pytest**. Went from having no tests to running automated tests in every PR, preventing production pipeline failures.
- Championed coding best practices among data scientists and drastically expanded documentation (including a Sphinx-generated API reference page).

Axom

Basel, Switzerland

Machine Learning Engineer - Internship

October, 2021 – March, 2022

Developed computer vision solutions for large pharmaceutical companies. Performed data preprocessing, trained models, and communicated with clients to define project requirements.

- Leveraged object detection models for counting syringes and vial inspection using **YOLOV5** and **PyTorch**, improving quality control processes. Deployed the models in edge devices with **Docker** and **Docker Compose**.
- Enhanced syringe detection accuracy in complex images by implementing custom object detection techniques, such as Polygon Object Detection (<https://github.com/XinzeLee/PolygonObjectDetection>) and Sliced Inference (<https://github.com/obss/sahi>)
- Applied image augmentation techniques (e.g., rotations, flipping, color and brightness variations) to increase dataset size and robustness of the model, especially in varying conditions (e.g. low light, blurry).

Further Experience

- Awarded Master's Thesis Scholarship, University of Coimbra (Jan - Sept 2021). Researched CNN fault-tolerance in safety-critical systems. Implemented state-of-the-art (Ranger (<https://github.com/DependableSystemsLab/Ranger>)) fault-tolerance method in PyTorch C++ and compared with novel technique.
- Student Researcher, University of Coimbra (Oct - Dec 2020). Enhanced object detection model explainability for forest fire prevention using LIME explainable AI library within the Darknet framework. Source code available on my GitHub page.

EDUCATION

University of Coimbra

Coimbra, Portugal

Bachelor's in Computer Science and Master's in AI (Thesis grade: 18/20)

September, 2016 – September, 2021

SKILLS

Languages: Portuguese (Native), English (C2 Proficiency)