






Müsel Emmanuel Tabares Pardo

 muselemmanuel@hotmail.com

 (+52) 663 104 6329

 musel25

 musel.dev

 Compiègne, France


MEXICAN | DOB: 05/08/2002

About


Seeking an opportunity for a Master’s student internship at Mistral. Strong desire to show what I am capable of. Passionate about learning and solving problems. Obsessed with LLMs and mathematics.

Experience

- 9/2023 – 01/2024

 **Machine Learning Engineer** Moodme | United States

Designed and deployed real time facial emotion recognition systems using deep learning computer vision techniques. Implemented a Graph Auto-Encoder (GAE) to learn latent representations of facial landmarks for a Unity 3D Face SDK, enabling identity preserving expression synthesis and emotion classification.
- 02/2023 – 08/2023

 **Data Analyst** CEMEX | Monterrey, Mexico

Developed predictive and statistical forecasting models to assess user behavior trends and adoption metrics for CEMEXGO. Use of Snowflake and SQL for data retrieval; Built interactive Power BI dashboards to support Product Management decisions.

Education

- 09/2024 – Current

 **Erasmus Mundus Master in Advanced Machine Learning and Optimisation Systems (EMSSE)**

Universitat Politècnica de Catalunya | Barcelona, Spain (09/2024–01/2025)
Universiteti Politeknik i Tiranës | Tirana, Albania (03/2025–06/2025)
Université de technologie de Compiègne | Compiègne, France (05/2025–Current)
- 08/2023 – 12/2024

 **Specialization in Advanced Artificial Intelligence Applied to Data Science**


Monterrey Institute of Technology and Higher Education | Guadalajara, Mexico
- 08/2021 – 06/2024


 **B.S. in Computer Science and Technology**


Monterrey Institute of Technology and Higher Education | Monterrey, Mexico


Projects


- Instruction Tuning (SFT+QLoRA)


 End-to-end instruction-tuning workflow with prompt templating, tokenization, and QLoRA adapters; reproducible configs, logged runs, checkpoints, and artifacts; containerized runner.



- Preference Alignment (DPO)


 Pairwise preference data generation and curation; Direct Preference Optimization adapters trained and compared against SFT baselines with deterministic evals and refusal/guardrail checks.



- Instruction Dataset Factory


 Data pipeline for instruction/chat datasets: rule-based filtering, deduplication, de-contamination, quality scoring, augmentation, and chat-template rendering; dataset cards + stats.



- LLM Twin Collection Pipeline


 ZenML-orchestrated crawlers (GitHub/Medium/Custom) with dispatcher pattern and retries; raw documents stored in MongoDB via an ODM, with Selenium fall-backs and metadata lineage.


- RAG Feature Pipeline

 Cleaning, chunking, and embedding into Qdrant with change-data-capture snapshotting; Pydantic domain entities, OVM dispatchers, and HNSW index configuration for precision/latency tradeoffs.


- Advanced RAG Inference

 Retrieval module with query expansion, self-querying, filtered vector search, and reranking.



LLM Evaluation	<ul style="list-style-type: none"> Reproducible offline evaluations (general/task/RAG) with deterministic prompts/seeds, contamination checks, and RAG metrics (Ragas/ARES); minimal leaderboard UI and model cards.
Inference Quantization	<ul style="list-style-type: none"> Throughput/latency benchmarks across GGUF, GPTQ, and EXL2 variants; experiments with KV cache, continuous batching, and speculative decoding where supported; compact report.
SageMaker Inference	<ul style="list-style-type: none"> Hugging Face DLC deployment on AWS SageMaker; FastAPI business service calling the endpoint.
LLMOps: CI/CD/CT	<ul style="list-style-type: none"> GitHub Actions pipelines for CI/CD/CT; Dockerized ZenML pipelines on cloud backends; Comet experiment tracking, Opik prompt monitoring, alerting, and runtime guardrails.
Telemetry RAG YANG	<ul style="list-style-type: none"> End-to-end RAG indexing lab telemetry docs (Markdown/YANG/JSON) into Qdrant; context-only prompting via Ollama/OpenAI; reproducible Docker setup and CLI for index/ask.
Medical LV Segmentation	<ul style="list-style-type: none"> U-Net for left-ventricle segmentation on Stanford EchoNet-Dynamic; compared pixel-wise masks vs. landmark detection; Eval and morphology processing.
ML for Activity Recognition	<ul style="list-style-type: none"> Evaluated 18 ML algorithms with grid search on accelerometer data to classify activity levels in older adults; validated via CV and significance tests.
Vision Hydrological Forecasting	<ul style="list-style-type: none"> Deep CNN pipeline predicting river stage from imagery; EfficientNet transfer learning and multi-modal inputs (visual + hydrological) for accuracy gains.
LLM + RAG for Metaheuristics	<ul style="list-style-type: none"> Automated framework using RAG + LLMs to generate and refine metaheuristics for black-box optimisation; integrated Qwen2.5-coder, ChromaDB, and Optuna.
Leukocyte CNN + LIME	<ul style="list-style-type: none"> CNN to classify leukocytes (cancerous vs non-cancerous) with LIME explanations to highlight decision-driving regions.
Fine-tuned GPT-2 Chatbot	<ul style="list-style-type: none"> Fine-tuned GPT-2 on WhatsApp chats; regex parsing to HF Datasets; trained and deployed interactive chatbot pipeline.
NLP Web App	<ul style="list-style-type: none"> Real-time text emotion classifier with logging and analytics; built using Streamlit and standard NLP preprocessing.

Skills

LLM Training	<ul style="list-style-type: none"> PyTorch; HF Transformers/PEFT/TRL; tokenizers(SentencePiece/Tiktoken); SFT, LoRA/QLoRA, DPO; Hydra configs; HF Datasets (streaming); run tracking (W and B/MLflow/Comet).
RAG and Data	<ul style="list-style-type: none"> Qdrant/FAISS (HNSW); chunking (recursive/semantic); retrievers (self-querying/hybrid) + rerankers (bge-reranker/Cohere); decontamination/PII; CDC snapshotting; Pydantic; MongoDB+ODM; ZenML orchestration; Selenium/Playwright crawling.
MLOps	<ul style="list-style-type: none"> vLLM, TGI, llama.cpp (GGUF); quantization (GPTQ/EXL2), KV cache, continuous batching, speculative decoding; FastAPI; AWS SageMaker (HF DLC); Docker/Kubernetes; GitHub Actions (CI/CD/CT); monitoring/guardrails (Opik, OpenTelemetry); PyTorch Profiler; A/B testing.
Evaluation	<ul style="list-style-type: none"> lm-evaluation-harness; Ragas/ARES; deterministic prompts/seeds; contamination checks; RAG metrics and error analysis.

Languages

Mother Tongue	<ul style="list-style-type: none"> Spanish
English	<ul style="list-style-type: none"> Listening C2 Reading C2 Writing C1 Spoken Production C1 Spoken Interaction C1
French	<ul style="list-style-type: none"> Listening B2 Reading B2 Writing B1 Spoken Production B1 Spoken Interaction B1
German	<ul style="list-style-type: none"> Listening A2 Reading A2 Writing A1 Spoken Production A1 Spoken Interaction A1