

JUAN MANUEL RUIZ MUÑOZ

ML Engineer · Computer Vision & LLMs

Alicante, Spain ◊ +34 678 23 12 12 ◊ juanmanuelruiz383@gmail.com ◊ [GitHub](#) ◊ [LinkedIn](#) ◊ juanmanuel.petrer.eu
EU citizen (no sponsorship required) ◊ Open to relocation, hybrid or remote (CET)

PROFILE

ML Engineer with a Physics and Mathematics background and 2.5+ years taking computer vision and language models from research to production. I build end-to-end systems (data, training, evaluation, deployment) for industrial defect inspection, including a LoRA-fine-tuned multimodal OCR running at 98.56% exact match. Rigorous by training: I want to understand why a model works before shipping it. Published at DCAI'25.

EXPERIENCE

Vraia Corp. 06/2025 – Present
ML Engineer · Computer Vision & LLMs Elche, Spain

- Built a **multimodal OCR** for tyre-sidewall text end to end: fine-tuned **GLM-OCR** with **LoRA** (LLaMA-Factory) on a 2×RTX 3090 server, added a post-processing correction stage and served it with **vLLM**, reaching **98.56% exact match in production**.
- Designed and trained **defect detection and segmentation** models (CNNs, Vision Transformers, YOLO, RF-DETR) for industrial tyre inspection: **0.83 mAP@50:95** on multi-class segmentation over proprietary datasets of thousands of images.
- Diagnosed and mitigated mask collapse on thin ring-shaped parts in a 7-class **RF-DETR** segmentation model (window/shutter components) by re-weighting the mask loss.
- Built a **synthetic defect generation** pipeline with generative models to augment scarce defect classes and improve detection robustness.
- MLOps**: ETL and augmentation pipelines for visual datasets; deployment and monitoring with **Docker**, **MLflow** and TensorBoard; evaluation protocols (precision, recall, F1, mAP, IoU); algorithm specifications in **C++** for integration by the client's engineering team.

AIR Institute & BISITE Research Group (University of Salamanca) 06/2024 – 06/2025
Data Scientist · AI Researcher Salamanca, Spain

- Co-designed an **LLM + RAG** architecture that builds a verified ontology incrementally to detect false claims in fiction (precision 0.98, F1 0.86); **published at DCAI'25**.
- Implemented **RAG** and semantic-retrieval pipelines (ChromaDB) and **multi-agent** systems (Ollama, DeepSeek, GPT-4o-mini), focused on reducing cost and latency.
- Delivered computer vision projects (detection, classification, segmentation) across the full ML lifecycle, working directly with clients on requirements and on adapting state-of-the-art methods to real use cases.

Miguel Hernández University – SCOLab Group 01/2024 – 07/2024
Research Intern Elche, Spain

- Developed Lagrangian methods for the automatic identification of atmospheric blocking (ERA5, HYSPLIT, DBSCAN); basis of my bachelor's thesis and of a paper in preparation.

SELECTED PROJECT

- vcalib** – calibration filters for RF-DETR under illumination shift open source
- A differentiable ~2 KB filter that corrects illumination shift *before* a frozen detector: no retraining, no labels. Recovers **35.9%** of activations on held-out real pairs and closes the AP gap on the level-2 shift benchmark (SAM 3 pseudo-ground truth); fitted in ~3.5 min on a laptop CPU.

EDUCATION

BSc in Physics – University of Alicante 2020 – 2024

- 120+ ECTS of the BSc in Mathematics taken in parallel (UA simultaneous-degree programme). SICUE exchange scholarship, University of Salamanca (2022 – 2023).

PUBLICATIONS

- DCAI'25**: *AI-Powered Ontology-Based Architecture for Misinformation Detection in Fiction Works*. [PDF]

SKILLS

LLMs & NLP	LoRA fine-tuning (LLaMA-Factory), Transformers, RAG, vLLM, Ollama, vector DBs (ChromaDB, FAISS), multi-agent
Computer Vision	PyTorch, OpenCV, CNNs, ViTs, YOLO, RF-DETR, SAM, generative models, detection, segmentation
MLOps & Code	Python, C++ (spec), R, Docker, MLflow, TensorBoard, Git, AWS, Linux, uv; statistical analysis
Languages	Spanish (native) · English (working proficiency)