Alejandro Sanchez Ferrer

Alicante, Spain • alejandro_sanchez_ferrer@hotmail.com • +34 618 014 101
LinkedIn • GitHub • Portfolio

PROFESSIONAL SUMMARY

- **Senior AI Solutions Architect** and **PhD researcher** with 5+ years delivering **Computer Vision** and **Deep Learning** solutions for **urban infrastructure, public transport, traffic management**, and **environmental safety**.
- Hands-on expertise across the **VLM fine-tuning lifecycle** (SFT, LoRA, QLoRA) and production **optimization** (TensorRT, containerized inference) for edge, cloud, and on-prem deployments.
- Proven track record building and deploying **AI-powered solutions at city and regional scale** across **Spain, Europe, and the United States**: real-time surveillance for metro systems, traffic sign compliance, wildfire early warning, and wildlife preservation across **28+ production deployments**.
- **Customer-facing technical expert**: effective in partnering with **Sales, Product, and Engineering** stakeholders to translate mission goals into active AI-enabled plans and successful deliveries.
- Deeply familiar with the **NVIDIA AI stack** (NeMo, TensorRT, Triton, DeepStream) and active technical contributor through **GitHub repositories**, **LinkedIn publications**, and an **interactive portfolio**.

SKILLS

Multimodal AI: Vision-Language Models (VLMs), Multimodal Reasoning, Transformers, Physical AI, Retrieval-Augmented Generation (RAG)
Deep Learning Frameworks: PyTorch, TensorFlow, FastAPI, NVIDIA NeMo, TensorRT, Triton Inference Server
Model Tuning & Optimization: Supervised Fine-Tuning, LoRA, QLoRA, CUDA, Inference Benchmarking, Synthetic Data
System Integration: Python, C++14/17/20, ROS, Docker, Real-Time Pipelines
Cloud & MLOps: Azure ML, AKS/EKS/GKE, CI/CD Pipelines, Automated Testing, Custom Monitoring Dashboards
Deployment: Edge (Jetson Nano/Orin), On-Premise GPU Servers, Cloud-Native Inference
Languages: Spanish (Native), **Catalan** (Native), **English** (C1), **French** (A2)

PROFESSIONAL EXPERIENCE

Minsait — Senior AI Solutions Architect & Technical Lead Jun 2021–Present
Madrid, Spain

- **Defined the technical roadmap and methodology** for up to **5 engineering teams** (Backend, Data, Deployment), establishing best practices for development and testing (unit, integration, E2E) to ensure production reliability.
- Led the **design and deployment of 8+ Computer Vision products** across urban, transport, energy, and environmental sectors in **Spain, Europe, and the US**, delivering AI solutions from PoC to production at scale.
- Architected **scalable, low-latency pipelines** for 24/7 vision inference; **reduced inference latency by over 80%** (from +1s to sub-200ms) by resolving GPU bottlenecks via multithreading and optimized resource orchestration.
- **Led technical engagement** with enterprise clients, building **PoCs and demonstrations**, gathering customer feedback, and advising on successful solution delivery.

Key projects with urban and public-sector impact:

– **Mobeye** — Real-time person tracking and **tailgating detection** system embedded in metro access gates for a major transport operator, enhancing **public transport safety and access control**.
– **Traffic Control Copilot** — AI-powered traffic sign detection, tracking, classification, and automated **compliance verification** against construction blueprints. Deployed on **Azure cloud** with Computer Vision and Generative AI for a leading infrastructure company in Texas.
– **FishEye** — Real-time person monitoring using ceiling-mounted fisheye cameras for a major telecom retailer, extracting operational KPIs (foot traffic, wait times, heatmaps) to **improve service delivery and citizen experience**.
– **GridEye** — Compact, solar-powered **early wildfire detection** system using Transformer models optimized with TensorRT for **Jetson Nano/Orin Nano**. Deployed in **20+ devices** across Spain for **public safety and environmental protection**.
– **WindEye** — Multi-sensor detection system integrating **vision and radar** to classify endangered bird species and trigger automatic turbine shutdown. Deployed in **8+ wind farms** across Spain, supporting **environmental regulation compliance**.
– **Q-SPAM** — Integrated a **quantized VLM** into a quadruped robot for real-time query-based reasoning, tracking, and behavior recognition, demonstrating **Physical AI** in edge-constrained environments.
– Additional deployed solutions include **VDA** (automated vehicle damage assessment with fine-tuned multimodal models for fraud prevention) and **BESOS** (real-time emissions monitoring for power plants using Computer Vision).

IFFE Business School — Adjunct Professor (Master in AI) Oct 2024–Jul 2025
A Coruña, Spain

– Instructed graduate-level course on "Principles of Robotic Autonomy," covering perception, navigation, computer vision, and multimodal AI architectures.
– **Developed and delivered technical presentations** and engaging educational materials on complex AI topics to a diverse graduate-level audience.

uWare Robotics — Robotics Engineer (Intern) Jan 2021–May 2021
Alicante, Spain

– Developed and optimized real-time control modules in **Python and ROS** for autonomous underwater vehicles (AUVs), operating under significant hardware constraints.

Embention — Software Verification Engineer (Intern) Jun 2019–Aug 2019
Alicante, Spain

– Implemented and verified safety-critical drone flight control algorithms in **C/C++14/17** on a real-time Linux-based embedded system.

EDUCATION

PhD in Computer Vision, University of Alicante 2023–Present
Research on synthetic data generation and diffusion models for robust vision systems in complex underwater environments.

MSc in Automation & Robotics, University of Alicante 2022–2023
Awarded: Honors proposal for best MSc Final Project.

BSc in Robotic Engineering, University of Alicante 2016–2021
Awarded: Honors proposal for best Final Degree Project.

Exchange Program, Arizona State University — Mechatronics/Robotics/Automation 2019–2020

CERTIFICATIONS

Build Next-Gen Agents with Vision Language Models — NVIDIA 2025
Navigating NIM Agent Workflows for Multimodal Retrieval — NVIDIA 2025

PUBLICATIONS, AWARDS & HACKATHONS

An Experimental Study on Marine Debris Detection — Pattern Recognition Letters (Elsevier, 2023)
The CleanSea Set: A Benchmark Corpus for Underwater Debris Detection — Springer LNCS (IbPRIA 2022)
Best Paper Award, IbPRIA (2022) – For work on the CleanSea Set dataset.
OceanGuard AI (Google Gemma3n Hackathon) — Real-time multimodal app for marine debris detection using a **fine-tuned Gemma3n VLM**, adapted for underwater imagery and edge deployment.