

Lucas González Fiz

AI / Machine Learning Engineer | Computer Vision & LLM Systems

Ourense, Spain | +34 677 228 205 | lucasglezfiz04@gmail.com
linkedin.com/in/lucas-gonzalez-fiz | github.com/LucachuTW | lucasfiz.dev

Summary

Artificial Intelligence graduate building production-oriented computer-vision, LLM/RAG and applied-ML systems for autonomous racing and healthcare. Experienced in rigorous evaluation, latency benchmarking, governed retrieval and deployment on NVIDIA hardware using Python, PyTorch, YOLO, LangGraph, FastAPI and Docker.

Education

B.Sc. in Artificial Intelligence (240 ECTS) Graduated Jul 2026
University of Vigo – ESEI Ourense, Spain
Selected coursework: Machine Learning I–II, Bio-inspired Machine Learning, Computer Vision, Natural Language Processing, Information Retrieval, Big Data and Distributed Systems.
Bachelor’s thesis (12 ECTS): Designed a reproducible multi-seed benchmark comparing PPO, A2C, DQN and RecurrentPPO in partially observable MiniGrid environments; evaluated memory, intrinsic motivation and curriculum learning.

Activities & Awards

FSUK AI 2026: 12th of 41 overall at Silverstone; second participation | **HackUDC 2026:** Inditex Tech Challenge podium.

Skills

Programming & ML: Python, SQL, PyTorch, scikit-learn, Hugging Face, Stable-Baselines3
Computer Vision: YOLO, OpenCV, SAM, ByteTrack, pose estimation, tracking, TensorRT
LLM Systems: RAG, GraphRAG, LangGraph, MCP, Qdrant, Neo4j, reranking
Engineering: FastAPI, Docker, Kubernetes, Linux, Git, ROS, NATS, NVIDIA Jetson

Experience

AI Software Developer – Autonomous Systems Nov 2024 – Present
Auria Formula Student AI Team Ourense, Spain

- Develop real-time perception software on a 1/10-scale autonomous racing platform for the Auria Formula Student AI team, working in an international Scrum environment.
- Train and benchmark pretrained and scratch YOLO cone detectors, applying structured pruning and TensorRT-oriented export for embedded deployment.
- Integrate camera and perception components into reproducible Linux, Docker and Git workflows, with ROS/NATS services prepared for NVIDIA Jetson deployment.

AI Engineer Intern Mar 2026 – Jun 2026
MicroPort CRM Ourense, Spain
Contracted through Cardiovascular Gallega

- Built an LLM/RAG pipeline to normalize heterogeneous clinical tables and map terminology across medical data sources, combining retrieval, prompting and validation stages.
- Developed and evaluated a machine-learning model for detecting pacemaker-lead malfunctions to support clinical review workflows.

Selected Projects

Sentinel Vision | *PyTorch, YOLO26-Pose, SAM 2.1, ByteTrack, TensorRT*
Built an observable real-time fall-detection service combining pose estimation, tracking, selective segmentation and temporal modelling; achieved **0.90 macro-F1**, **27.98 FPS** and **49 ms p95** latency on an RTX 2070.

Airlock Agents | *LangGraph, MCP, PostgreSQL, Ollama, FastAPI, Kubernetes*
Implemented role-scoped tools, read-only SQL access, persistent human approval gates and end-to-end audit traces; added offline regression gates over documents and **541,909** retail transactions.

Strata | *GraphRAG, Qdrant, Neo4j, BGE-M3, Qwen3, LangGraph*
Enforced clearance filters inside vector and graph retrieval before generation; combined dense and BM25 search, graph expansion and reranking, improving exact-ID recall from **0.79 to 1.00**.

Partial Observability in Deep RL | *Gymnasium, Stable-Baselines3, PPO, A2C, DQN, RecurrentPPO*
Compared memory, state representation, intrinsic motivation and curriculum learning under partial observability; curriculum learning was more reliable than direct training, while added memory or context did not consistently improve performance.

Languages

Spanish – Native | English – B2 CEFR