

DAVID CARDONER

MLOps | Agentic AI & Cloud Developer | Data Scientist

✉ secaijournal@gmail.com  [LinkedIn](#)
 github.com/cardoner1993  github.com/SecAI-Journal

Senior ML/MLOps Engineer with 5+ years across the full Machine Learning lifecycle, including feature engineering, model training and evaluation, and large-scale production deployment (Azure, GCP, Kubernetes). Recent work centres on agentic AI: LLM agents built with LangGraph, CrewAI and LangChain, served through FastAPI and operated via lightweight React interfaces. MSc in Statistics & Operations Research, with peer-reviewed publications applying ML/optimization to real-world forecasting and decision problems. Strong Python/SQL foundation, hands-on with TensorFlow and PyTorch, and a strong interest in recommender systems, ranking and personalization, with a track record of turning large, messy datasets into models that drive product decisions.

EDUCATION

EXECUTIVE MASTER IN DIGITAL FORENSICS & CYBERSECURITY

ISF Corporate University / ongoing / June 2025

MASTER'S DEGREE IN STATISTICS AND OPERATIONS RESEARCH

Universitat Politècnica / 2019

DOUBLE DEGREE IN STATISTICS AND ECONOMICS

Universitat Politècnica - Universitat de Barcelona / 2017

LANGUAGE

- Spanish - Native
- Catalan - Native
- English - Professional working proficiency
- Italian - Professional working proficiency

SKILLS

Languages

- Python
- SQL
- R
- C#
- Shell
- JavaScript

Agentic & LLM Frameworks

- LangChain
- LangGraph
- LlamaIndex
- Hugging Face smolagents
- DSPy
- MCP (Model Context Protocol)
- RAG / Vector Databases

Data Engineering

- Spark
- Databricks
- Pandas

ML & Serving

- Tensorflow
- Pytorch
- Ray
- CUDA
- Triton
- vLLM

Data Science & Modelling

- Statistics
- Recommender Systems / Ranking

Backend & Frontend

- FastAPI
- REST APIs
- React

WORK EXPERIENCE

MLOps & DevOps ML Engineer at CEATAEC 09/2025 - Today

- Build agentic AI systems with LangGraph, covering tool calling, retrieval and memory, exposed as FastAPI services and deployed on Kubernetes.
- Develop internal React interfaces so analysts can run agents, inspect intermediate tool calls and flag bad outputs without going through engineering.
- Serve open-weight LLMs with vLLM on CUDA GPUs, tuning batching and quantisation to keep latency and cost under control.
- Deploy and scale machine learning models on Azure using Kubernetes, ensuring reliable and performant inference.
- Design containerized, reproducible ML workloads to support fast iteration on models and features.
- Maintain and automate DevOps and CI/CD pipelines, shortening the path from model to production.
- Apply SecDevOps practices across the CI/CD and ML deployment lifecycle, including tool-permission and prompt-injection checks for agent workloads.

Senior MLOps & Data Engineer at SCRM - Lidl International Hub 05/2022 - 09/2025

- Designed and managed Azure ML and Databricks pipelines for training and deploying Generative AI models across the full lifecycle.
- From 2023, built LLM applications with LangChain and LlamaIndex, including RAG pipelines over internal data sources backed by vector databases.
- From 2024, moved those into agentic workflows with LangGraph, served through FastAPI microservices with streaming responses.
- Built React dashboards on top of those services for prompt iteration and side-by-side evaluation of agent versions before rollout.
- Deployed self-hosted LLM inference with Triton Inference Server and vLLM on GPU nodes as a cheaper alternative to per-token API calls.
- Ran distributed ML workloads on Ray over Kubernetes, scaling training and batch inference across GPU nodes.
- Built real-time data pipelines with Databricks Structured Streaming to feed high-throughput AI applications.
- Used Terraform to define reusable infrastructure, and set up observability with OpenTelemetry, Prometheus and Grafana across AI services.

Senior MLOps Engineer at Ciklum 12/2020 - 05/2022

- Containerized cybersecurity detection models with Docker and Kubernetes for consistent deployment across environments.
- Built Azure-based ML infrastructure supporting the training and serving of cybersecurity models.
- Implemented CI/CD pipelines with Azure DevOps to automate testing and deployment of ML workloads.
- Exposed model inference through Python REST APIs consumed by internal detection tooling.
- Defined reusable infrastructure with Terraform and applied Kubernetes RBAC and security policies to protect model deployments.
- Instrumented services with OpenTelemetry and Prometheus to monitor model and infrastructure health.

Devops & Data Scientist at IREC/Eurecat 09/2018 - 12/2020

- Designed and implemented Machine Learning and Deep Learning algorithms for applied forecasting and decision-support problems.
- Performed feature engineering, feature selection and model optimization to improve performance and adaptability, later published in peer-reviewed journals.
- Trained deep learning models on CUDA GPUs, cutting experiment turnaround during model selection.
- Defined optimization models for decision support over SQL and NoSQL data stores.

- .NET

Cloud & Infrastructure

- Azure
- AWS
- GCP
- Docker
- Kubernetes
- Terraform
- ArgoCD
- Linux
- MongoDB
- Redis

Soft

- Critical Thinking
- Continuous learning
- Proactive mindset
- Time management
- Problem solving
- Detail oriented
- Communication

CERTIFICATION

Fast LLM Inference with Cerebras — *DeepLearning.AI* / July 2026

Attention in Transformers: Concepts and Code in PyTorch — *DeepLearning.AI* / July 2025

Improving Accuracy of LLM Applications — *DeepLearning.AI* / July 2025

How Diffusion Models Work — *DeepLearning.AI* / July 2025

LLMs as Operating Systems: Agent Memory — *DeepLearning.AI* / July 2025

Building Code Agents with Hugging Face smolagents — *DeepLearning.AI* / July 2025

MCP Course Unit 3: Automation in Production — *Hugging Face* / June 2025

DSPy: Build and Optimize Agentic Apps — *DeepLearning.AI* / June 2025

AI Agents Fundamentals — *Hugging Face* / May 2025

AI Advanced Agents Course — *Hugging Face* / February 2025

Web Scraping & API Development — *365 Data Science* / November 2024

Master LLM Prompt Engineering — *Udemy* / July 2024

LangChain Chat with Your Data — *DeepLearning.AI* / May 2024

ChatGPT Prompt Engineering for Developers — *DeepLearning.AI* / May 2024

Finetuning Large Language Models — *DeepLearning.AI* / May 2024

Functions, Tools and Agents with LangChain — *DeepLearning.AI* / May 2024

Build LLM Apps with LangChain.js — *DeepLearning.AI* / May 2024

Efficiently Serving LLMs — *DeepLearning.AI* / November 2023

HashiCorp Certified: Terraform Associate 2023 — *Udemy* / January 2023

Machine Learning Engineering for Production (MLOps) — *Coursera* / January 2023

Computational Neuroscience — *Coursera* / December 2022

SCIENTIFIC PUBLICATIONS

- Data analysis and management for optimal application of an advanced ML-based fault location algorithm for low voltage grids.
- Reliability Assessment of a Fault-Tolerant PV Multistring Inverter.
- How Building and District Algorithms Enhance Renewable Energy Integration in Energy Markets.
- A Practical Fault Location Algorithm With Increased Adaptability for Active Low Voltage Grids.
- Feature Selection and Optimization of a ML Fault Location Algorithm for Low Voltage Grids.
- Optimal management for innovative electric vehicle charging station with machine learning forecasting modules.