

Daniel Montilla Navas

EXPERIENCED AI ENGINEER | AGENTIC AI & LLM SPECIALIST

Abu Dhabi, UAE

☎ (+971) (0)50 341 8578 | ✉ danielmontillanavas@gmail.com | 🌐 danimontilla.substack.com | 📺 daniel-montilla-navas

Summary

AI Engineer with 7+ years of experience building end-to-end ML/AI solutions from ideation to production deployment and monitoring. Expertise in modern agentic frameworks including LangChain, with deep understanding of LLMs for natural language processing applications, embeddings, prompt engineering, and fine-tuning techniques. Strong software engineering foundations including Python, PyTorch, CI/CD, and production API development with proven ability to collaborate across product, design, and engineering teams. Demonstrated track record mentoring junior engineers, championing model quality and evaluation practices, and incrementally shipping user-facing AI features that deliver measurable impact. Clear technical communicator passionate about applying cutting-edge AI to transform knowledge discovery and research workflows.

Technical Expertise

LLM & Agentic AI	LangChain, Agno, OpenAI/Claude models, Transformers, RAG and multi-agent systems, prompt engineering
Backend & API	Python, JavaScript / Go (non-commercial), FastAPI, REST/gRPC, async, microservices
ML, AI & MLOps	TensorFlow, PyTorch, scikit-learn, ONNX, MLflow, DVC, Prefect, Airflow
DevOps	AWS, GCP, Kubernetes, Docker, Terraform, CI/CD Pipelines
Data	ElasticSearch, Neo4J, MongoDB, PostgreSQL, DataBricks, vector databases, embeddings
Languages	Spanish (native), English (fluent), French (fluent)

Work Experience

Liquidity Capital Ltd

Abu Dhabi, UAE

SR MACHINE LEARNING ENGINEER

Sep. 2024 - Current

- Built production LLM-powered conversational platforms using Agno framework and LangChain, processing 10K+ daily user interactions for B2C financial services applications
- Developed autonomous AI agents for document processing and automated report generation, reducing manual processing time by 75% through intelligent multi-agent orchestration
- Implemented RAG architectures integrating diverse data sources (ElasticSearch, Neo4J, MongoDB) with vector embeddings for enhanced information retrieval and threat intelligence
- Established technical leadership standards for production-ready AI code including DevOps best practices, code reviews, and cross-team collaboration with backend, frontend, and Product teams
- Optimized system performance using LangChain parallelism to reduce chatbot response times from 40s to 10s, and improved report generation from 10min to 3min through efficient concurrency implementation

Newtral [6-month contract]

Madrid, Spain

MLOps & DATA LEAD

Feb. 2024 - Aug. 2024

- Refactored LLM-based data pipelines and integrated with Prefect orchestration, reducing ETL processing times by 40% through improved workflow automation, processing 50k+ records daily
- Enhanced security infrastructure by migrating from .env files to AWS Secrets Manager for credential management, eliminating public access vulnerabilities
- Deployed production LLM inference pipeline using miniLM and mdeberta models with ONNX optimization and TritonServer, enabling seamless CPU/GPU switching for 5x cost optimization and 10x time optimization

Correcto

Madrid, Spain

HEAD OF AI & CORE BACKEND SERVICES (5TH HIRE) (ROLE ENDED DUE TO COMPANY LAYOFFS)

Jan. 2023 - Feb. 2024

- Architected greenfield AI platform from zero using Python, FastAPI, and LLMs, serving 40K+ users with 99.9% uptime through microservices architecture
- Built end-to-end MLOps pipeline with Kubernetes, GCP, and VertexAI for automated model training, testing, and deployment of NLP models, reducing time to production by 70% and enabling Data Scientists to take full ownership
- Led cross-functional engineering organization (5 engineers) including architecture roadmap planning, technology strategy, and performance optimization while partnering with C-suite to define product direction
- Implemented real-time data processing using Airflow orchestration and stream processing for continuous model monitoring and drift detection

Intelygenz

TECH LEAD & AI SOLUTIONS ARCHITECT

Madrid, Spain

Oct. 2021 - Dec. 2022

- Delivered client-facing AI architecture solutions from conception to production deployment on enterprise Kubernetes clusters, collaborating directly with business stakeholders to architect custom platforms
- Developed production NLP services using transformer models, ONNX optimization, and gRPC APIs deployed on enterprise Kubernetes clusters
- Implemented distributed model inference systems handling 100K+ daily predictions with Docker containerization and load balancing
- Led Machine Learning Engineer career development program, establishing technical standards and mentoring protocols company-wide
- Established Python coding standards and ran internal technical learning programs

Previous Roles

Madrid, Spain, London, UK & Paris,

France

SENIOR AI ENGINEER POSITIONS AT PLAIN CONCEPTS, GAIA, ANNALECT, TANDEM BANK, TILLER SYSTEMS

Sep. 2018 - Apr. 2021

- 3rd hire at Gaia startup - Built prediction algorithms using distributed computing frameworks and advanced statistical modelling in early-stage environment
- Built large-scale prediction algorithms processing billions of records using distributed computing frameworks and advanced statistical modelling
- Developed fraud detection systems reducing operational overhead by 85% through innovative ML approaches and real-time processing capabilities
- Created merchant classification algorithms improving computational efficiency by 50% using optimized feature engineering and model architecture

Community & Teaching

MLOps Community

CO-ORGANISER

Madrid, Spain

Mar. 2023 - Aug. 2024

- Organize monthly technical meetups for 200+ ML practitioners
- Host technical talks on LLM deployment, agentic AI, and production ML systems

Mioti

MASTERS LECTURER

Madrid, Spain

Mar. 2022 - Aug. 2024

- Developed curriculum: Machine Learning 101 (12h), MLOps (6h), DevOps (12h)
- Created hands-on workshops on production ML lifecycle and deployment strategies

Media & Speaking

- MLOps Community Event: "MLOps at an Early-stage start-up" [YouTube]
- Podcast Guest: "NLP in Production: From Research to Real-World Applications" [Spotify]
- Roundtable Panelist: "Reunión Anual CITT | Humanidades digitales y tecnologías" [YouTube]

Education

The University of Edinburgh

MSC IN BUSINESS ANALYTICS

Edinburgh, UK

Sep. 2017 - Aug. 2018

- First class, Distinction
- Awarded two scholarships to pursue an industry project for my dissertation

The University of Edinburgh

MENG IN MECHANICAL ENGINEERING WITH MANAGEMENT

Edinburgh, UK

Sep. 2012 - May 2015

- Upper Second Class, Honours

University of Malaga

BENG IN MECHANICAL ENGINEERING

Malaga, Spain

Sep. 2008 - May 2012

- Completed an Erasmus year at the University of Edinburgh, which led to my transfer to continue my studies there

Key Differentiators

- **Open-source contributions** to LLM orchestration and agent development frameworks with active community involvement
- **Production deployment** of 15+ LLM-powered applications with combined 500K+ monthly active users
- **Performance optimization** specialist achieving 5x cost reduction through model optimization and infrastructure tuning
- **Full-stack AI** platform development from architecture to deployment