

Esteban Castiblanco

Data Scientist | ML-AI Engineer

✉ esteban.castiblanco24@gmail.com ☎ 3165892575 📍 Colombia 🔗 jesteban247.github.io/Jesteban247
🌐 linkedin.com/in/esteban-castiblanco-36b537230 🐙 github.com/Jesteban247

PROFILE

Data Scientist and ML-AI Engineer with expertise in statistics, programming, and data analysis. Skilled in developing and deploying advanced machine learning models, including deep learning, NLP, and computer vision. Passionate about solving real-world problems with data-driven approaches, building scalable AI systems, and staying updated with emerging AI technologies.

PROFESSIONAL EXPERIENCE

IBM, ML-AI Engineer 01/2025 – present

- Building and deploying advanced AI solutions, including the development of a production-ready chatbot using LangChain, integrating memory, tools, and FastAPI endpoints for seamless user interaction.
- Leveraging high-performance computing clusters with NVIDIA GPUs to accelerate inference and training of open-source models from Hugging Face, focusing on NLP and ML tasks that benefit from parallelization and GPU optimization.
- Responsible for designing scalable architectures, managing model deployment pipelines, and collaborating with cross-functional teams to deliver robust AI-driven applications.
- Gained hands-on experience in optimizing model performance, handling large-scale data, and implementing best practices for MLOps in enterprise environments.

EDUCATION

Master's in Artificial Intelligence, Javeriana University 01/2025 – 12/2026

Bachelor's in Data Science, Javeriana University 01/2022 – 12/2025

CERTIFICATES

IBM Machine Learning Specialist - Advanced 🔗 | Fundamentals of Accelerated Computing with CUDA Python 🔗 |
Machine Learning Specialization 🔗 | TensorFlow Developer Certificate 🔗

PROJECTS

Generative AI

- RAG Implementation: Developed a system that combines information retrieval with generative models to improve response accuracy by retrieving relevant data before generating answers.
- Adapted pre-trained models for specialized tasks by fine-tuning them on domain-specific data, improving their performance for custom use cases.
- Built AI agents capable of performing autonomous tasks, such as decision-making and problem-solving, using multi-agent frameworks.
- Currently developing a small-scale language model, focusing on optimizing its architecture and training techniques for efficient resource utilization and functionality.

Computer Vision

- Developed a deep learning system to track player movements and events during tennis matches, using video footage to provide real-time insights and performance metrics.
- Built a model to classify CT scans for medical diagnostics, focusing on detecting anomalies in medical imaging, improving accuracy, and aiding in early diagnosis.
- Currently working on an AI tool for soccer match analysis, aiming to detect key events and track player performance using advanced image recognition techniques.

Machine Learning Model Development

- Developed and deployed machine learning models for various applications, focusing on classification, regression, and clustering.
- Optimized models to enhance accuracy and scalability in production environments, ensuring seamless integration and high performance.

SKILLS

Generative AI — LLMs, VLMs, Image generation, Audio, Fine-tuning, RAG, LangChain, **NLP** — Embeddings, Text classification, NER, Summarization, Translation, **Data Science** — Data visualization, Feature engineering, EDA, Statistical modeling, Time series, **MLOps & Deployment** — CI/CD, APIs, Docker, Model monitoring, Pipeline automation, **Cloud (Learning)** — GCP, Azure, AWS, Cloud ML workflows, **HPC** — NVIDIA GPUs, Distributed training, Parallel computing