

# Joel Chacón Castillo

[✉ joelchaconcastillo@gmail.com](mailto:joelchaconcastillo@gmail.com) | [+52 442 124 1056](tel:+524421241056)

[GitHub](https://github.com/joelchaconcastillo) | [LinkedIn](https://linkedin.com/in/joel-chacon-castillo)

[Portfolio](#)

## Summary

Generative- and LLM-focused engineer with production experience building retrieval-augmented systems, model evaluation suites, and LLM integration into product APIs. Combines ML research background with software engineering best practices to deliver reliable, auditable model-backed features.

## Skills

**Models & Libraries:** LLaMA, OpenAI APIs, HuggingFace, PyTorch, LoRA

**RAG & Retrieval:** Embeddings, FAISS, chunking, vector pipelines

**ML Ops:** Model deployment, evaluation (factuality/hallucination), monitoring, autoscaling

**Infra & Backend:** FastAPI, Kubernetes, Docker, Redis, OpenSearch, SQL

## Work Experience

### Orion Innovation, Mexico (Contractor)

Nov 2025 – Present

#### Full-stack Gen AI Engineer

- Built end-to-end GenAI prototypes and production integrations: retrieval, embedding stores, rerankers, and LLM orchestration layers.
- Implemented evaluation and monitoring for LLM outputs (hallucination detection, grounding checks, latency/cost tracking).
- Containerized inference services with autoscaling and safe-rollout policies; collaborated with security and privacy teams.
- Delivered internal tools to standardize prompt templates and few-shot context management across teams.

### Oracle, Mexico

Sept 2023 – Jun 2025

#### Senior Software Developer

- Developed backend services and cloud infrastructure to support ML and DS workloads.
- Led prototypes for LLM-enabled developer tooling and defined requirements for productionization.
- Improved developer workflows by introducing CI/CD and automated tests for model-backed services.
- Refactored data-science codebases and coordinated model deployment processes.

### The University of Texas at Dallas, United States

April 2023 – Jul 2023

#### Postdoctoral Research Associate

- Researched graph-based prediction tasks and implemented reproducible experimental code.
- Publication: “Revisiting Link Prediction with the Dowker Complex”. [https://link.springer.com/chapter/10.1007/978-981-97-2253-2\\_33](https://link.springer.com/chapter/10.1007/978-981-97-2253-2_33)

### The University of Texas at Dallas (Remote)

Nov 2022 – Mar 2023

#### Research Collaborator — Wildfire Prediction

- Implemented spatio-temporal models and feature engineering pipelines for wildfire forecasting experiments.
- Publication and code: <https://www.computer.org/csdl/proceedings-article/bigdata/2024/10825303/23yk9W9fR8A&https://zenodo.org/records/7952192>

### Center for Research in Mathematics (CIMAT), Mexico

Aug 2014 – Jun 2015

#### Full-Stack Developer

- Built web systems to support administrative and research workflows; implemented front-end and back-end code.

### National Laboratory of Genomics for Biodiversity (LANGEBIO), Mexico

Aug 2013 – Jul 2014

#### Intern Software Engineer

- Developed pipelines and visualizations for sequence data processing, enabling efficient analysis.

## Education

### Center for Research in Mathematics (CIMAT), Mexico

Jan 2018 – Jan 2023

#### Ph.D. in Mathematics

Focus: Applied Mathematics, Machine Learning, Scientific Computing

### Center for Research in Mathematics (CIMAT), Mexico

Aug 2015 – Dec 2017

#### M.Sc. in Computer Science

Focus: Algorithms, Data Science, Machine Learning

### Instituto Tecnológico Superior de Irapuato, Mexico

Aug 2009 – Jun 2014

#### B.Eng. in Software Engineering

Focus: Software Engineering, Databases, Web Development

#### Specializations

- Meta React Specialization — Meta / Coursera (2025)
- NLP Specialization — DeepLearning.AI (2024)
- Machine Learning Specialization — DeepLearning.AI (2022)

## Projects

### PyTestGenAI

Aug 2025 – Present

AI-driven REST API Test Generation Framework

- Modular Python framework to generate tests for REST APIs using LLMs; Dockerized and CI-ready.
- Releases and testing artifacts: <https://test.pypi.org/project/PyTestGenIA/>

### agentic-codegen

Sept 2025 – Present

Python-first AI Code Generation Pipeline

- Multi-stage pipeline combining LLMs, static analysis, and sandboxed validation for safer code generation.

### LeetCode Practice

Jul 2025 – Nov 2025

Algorithm practice

- Extensive DSA practice; profile: <https://leetcode.com/u/joelchacon/>

## Certifications

- AI Pair Programming with GitHub Copilot — LinkedIn, Dec 2025
- Generative AI for Software Development — DeepLearning.AI, Oct 2025
- Fundamentals of Building AI Agents — IBM, Dec 2025
- Meta React Specialization — Meta, Oct 2025
- OCI AI Foundations Associate — Oracle, Aug 2024

## Awards

- Meta Hacker Cup 2025 — Round 3 Qualifier, Top 174 Worldwide — Certificate: <https://www.facebook.com/codingcompetition/hacker-cup/2025/certificate/408394896783829>
- Talk: Diversity to Improve the Effectiveness of Evolutionary Algorithms — CIMAT, Oct 2024