

Sebastian Cepeda

Santiago, Chile | sebastian.cepeda.fuentealba@gmail.com | [linkedin.com/in/sebastiancepeda](https://www.linkedin.com/in/sebastiancepeda) | sebacepeda.com

Senior ML Engineer with an MSc in Electrical Engineering and 10+ years building and deploying production ML systems. Specialized in deep learning, computer vision, LLMs, and generative AI across healthcare, finance, retail, and security domains. Proven track record of shipping models that improve detection accuracy, reduce churn, and automate workflows at scale. Experienced across the full ML lifecycle: data pipelines, model training, MLOps, and cloud deployment.

Technical Skills

- **ML / AI:** PyTorch, TensorFlow, Keras, scikit-learn, Generative AI, LLMs, Diffusion Models, Deep Learning, NLP, Computer Vision, Time Series, ONNX, SHAP.
- **MLOps & Engineering:** CI/CD, model deployment, model monitoring, data drift detection, Docker, Kubernetes, Python, Bash, Java, C/C++, SQL.
- **Data:** Pandas, Dask, Spark (PySpark), PostgreSQL, MySQL, ClickHouse, feature engineering, Vowpal Wabbit.
- **Cloud:** AWS (SageMaker, RDS, Glue, Lambda, Serverless), Azure (PySpark, SQL Notebooks, VMs), DigitalOcean.
- **Computer Vision:** CNN, SfM, Stereo Vision, OpenCV, Point Cloud Library, LIDAR, 3D Point Cloud Segmentation, YOLO, Kalman Filter.

Professional Experience

Tecla / Incyte

October 2024 – March 2026

Sr. Machine Learning Engineer

- Researching and developing generative AI models (diffusion) for synthetic tabular data in a clinical research context.
- Built an LLM-powered embedding pipeline over medical literature search results; applied clustering to surface actionable insights and statistics.
- Designed prompt engineering workflows with Claude (Anthropic) to automate generation of clinical study documentation, significantly reducing manual effort.
- *Stack: PyTorch, scikit-learn, pandas, Python, Bash, Docker, AWS VMs.*

Intuition Machines

July 2021 – July 2024

Machine Learning Engineer

- Built an end-to-end MLOps pipeline for training and deploying models on 2M+ data points, including CI/CD, automated testing, and model versioning.
- Developed a CNN model for motion-based threat detection, improving malicious activity detection by 20%; optimized for real-time inference using ONNX and C runtimes.
- Improved production customer-specific models, boosting recall from 0.50 to 0.78 with no increase in false positive rate.
- Implemented SHAP-based model explainability and automated alerting for data drift and performance degradation in production.
- *Stack: TensorFlow, PyTorch, scikit-learn, pandas, Vowpal Wabbit, Python, SQL, JavaScript, Bash, Kubernetes, Docker, Azure VMs.*

Consorcio

March 2020 – July 2021

Senior Data Scientist

- Developed and deployed a churn prediction model trained on 15M data points, integrated with Salesforce to power marketing and retention campaigns.
- Built and deployed a recommender system that automated marketing campaign execution, reducing manual segmentation effort.
- Implemented NLP and sentiment analysis pipeline (text + speech-to-text) for customer relationship analysis.
- *Stack: AWS, PostgreSQL, scikit-learn, pandas, Python.*

Cencosud

December 2018 – February 2020

Data Scientist

- Developed production sales forecasting models for retail logistics across millions of SKUs for paris.cl and Colombian supermarkets; scaled data processing with Dask and PySpark.
- Built a real-time pedestrian detection and tracking system (CNN + Kalman filter) to monitor customer foot traffic across stores.

- *Stack: Azure, scikit-learn, pandas, Python, Dask, PySpark, TensorFlow, OpenCV, YOLO.*

WoodtechMS

April 2015 – December 2018

Software Engineer – Computer Vision

- Developed embedded computer vision software for industrial material measurement, deployed to customers globally.
- Owned the full software lifecycle of an industrial access control product, from design to production release.
- Mentored junior engineers and interns on signal processing, Kalman filters, and regression models.
- *Stack: Java, C++, Windows, stereo vision, 3D reconstruction, point cloud modeling.*

Education

MSc Electrical Engineering, Universidad de Chile	<i>2014 – 2016</i>
Minor in Computer Science, Universidad de Chile	<i>2011 – 2013</i>
BSc Electrical Engineering, Universidad de Chile	<i>2008 – 2016</i>

Languages

Spanish: Native. **English:** Fluent (reading, writing, speaking).

Projects & Internships

Python FAQ Chatbot *2025 – 2025*

Personal Project

RAG-based conversational chatbot over Python's official documentation. Built with FAISS vector search, sentence-transformers embeddings, and LLaMA 3 via Ollama; served via Streamlit with Docker Compose.

Biomedical Engineering Lab, Universidad de Chile *March 2014 – December 2015*

MSc Thesis

Automated blood vessel segmentation in retinal images using computer vision and image processing techniques.

Woodtech *April 2013 – February 2014*

Internship

Developed C++ algorithms for industrial wood measurement via 3D point cloud segmentation.

Sixbell *January 2013 – February 2013*

Internship

Built an IVR application in Java for Android.

AMTC Lab, Universidad de Chile *May 2011 – October 2011*

Research Assistant

QA engineering for an autonomous vehicle software project.