

 Don't want to read? [Chat with my CV!](#)

EXPERIENCE

• Data Science/ML Consultant - Freelance

Nov. 2024 - Present (Selected clients. References available upon request)

- **Inter-American Development Bank:** Developed a methodology that uses Large Language Models for the bank to identify and categorize structural reforms in Latin American countries from naturally occurring data (i.e. articles in newspapers), facilitating their prioritization of loan assignments.
- **NODO Research:** Achieved a 95% reduction in audio transcription time and a 20% cost reduction by developing an automated transcription pipeline leveraging WhisperAI and LLMs, which streamlined the processing of market research interviews. Additionally, reduced the time required for qualitative data analysis by 50% through the creation of custom retrieval systems to interact with the transcriptions, enhancing efficiency in extracting insights.
- **IESE, CEMFI, Colombian Central Bank:** Developed and taught multiple master classes on modern machine learning techniques for working with textual and unstructured data (e.g. IESE 2024: [GitHub](#) - [Website](#)).

• Data Scientist - ETH Zurich (Economics and Data Science Group)

Sept. 2023 - Nov. 2024

- **SECO - Swiss Government:** Develop a recommender system for the public employment service, achieving higher user satisfaction ratings than the previous system. The system utilized methods from the algorithmic fairness literature to prevent discrimination and included explanations for each recommendation.
- Co-authored an entry on the Palgrave Handbook of Economics and Language on large language models and their applications in economics ([Working paper](#) - [GitHub](#)).
- Supported the development of three different courses with an average of 30 students ([GitHub](#)).

• Data Science Researcher - University College London

March 2021 - Aug. 2023

- Partnered with Lightcast to develop a language model (fine-tuned distilBert) to predict the presence of remote work on over 1 billion job postings. Designed the annotation process to collect the labels for finetuning and the production pipeline to continuously generate predictions. The paper has been cited over 60 times and the data is currently used by over 800 researchers and internally by the company ([Webpage](#) - [Media](#) - [Model](#)).
- Demonstrated techniques for algorithmic text analysis as part of an article for the Annual Review of Economics. The associated [GitHub repository](#) has over 100 stars, and the [article](#) has been cited over 100 times.
- Designed and developed a SQL database (PostgreSQL) to store and efficiently query the content of over 500M news articles from Dow Jones Newswires.
- Developed a method to characterize firms within a trade network using low-dimensional numeric representations (embeddings), enabling the creation of new measures of firm similarity and market competition. Presented a [poster](#) at CESifo Venice Summer Institute 2022 ([GitHub](#) - [Blog](#)).

• Data Analyst - Instiglio

Jan. 2018 - Aug. 2019

- **USAID:** Led the design of predictive models, utilizing regression analysis techniques, to forecast the performance of educational programs in Liberia. Supported the design of an innovative financing mechanism for a five-year educational project in the country.
- Collected, cleaned, analyzed, and visualized historical data of all the company's projects to produce relevant insights for the Board's annual meeting.
- Supported the business development efforts of the company by writing concept notes and formal proposals to prospective clients. Two of these efforts led to important projects for the company.

EDUCATION

- **Ph.D. in Data Science and Economics - ETH Zurich**
Sept. 2023 - Nov. 2024 (Voluntarily concluded)
- **M.Sc. in Data Science - Barcelona School of Economics**
Sept. 2019 - Sept. 2020 (GPA: 8.57/10.0)
 - **Relevant courses:** Machine Learning, Text Mining, Data Warehousing and Business Intelligence, Reinforcement Learning, Data Visualization
- **B.Sc. in Economics - Universidad de los Andes**
July 2013 - Dec. 2017 (GPA: 4.44/5.0)
 - **Relevant courses:** Advanced Econometrics, Game Theory, Microeconomics III, Analysis of Armed Conflict, Empirical Applications in Political Economy

SKILLS

- **Programming languages:** Python (transformers, JAX, TensorFlow, PyTorch, scikit-learn, dash, LangChain, LangGraph, spaCy, PySpark), SQL, R, Stata, MATLAB, Java (beginner), JavaScript (beginner)
- **Tools and platforms:** Git, GCP, AWS, Docker, Hugging Face, Linux, Bash/Shell, Heroku, CSS, HTML, PowerBI, Spark
- **Languages:** Spanish (native), English (working proficiency), French (intermediate)