

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

I am a social and data scientist with over 5 years of experience using machine learning, natural language processing, and causal inference to solve problems across various fields, effectively communicating technical concepts to diverse audiences.

PROGRAMMING LANGUAGES

Python (advanced): transformers, JAX, TensorFlow, PyTorch, scikit-learn, dash, LangChain, spaCy
R (intermediate), **SQL** (intermediate), **Java** (beginner), **JavaScript** (beginner)

EXPERIENCE

- **ETH Zürich - Data Science and Economics Group** Zürich, Switzerland
Researcher, Member of ETH AI Center *Sept. 2023 - Present*
 - **LLMs in Economics**: Co-authored an entry on the Palgrave Handbook of Economics and Language on large language models and their applications in economics ([Working paper](#) - [GitHub](#)).
 - **Recommender system (Swiss Government)**: Partnered with the Swiss Government to 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.
 - **Explainable AI**: Developing a taxonomy of explainable machine learning techniques alongside a [Python library](#) that implements many of them. In addition, we are running a field experiment to test how they affect people's perception of automated decisions.
 - **Teaching**: Supported the development of three different courses with over 40 students ([GitHub](#)).
- **Machine Learning Consultant** Remote
Selected projects *June 2020 - Present*
 - **Data extraction with LLMs (Inter-American Development Bank)**: Developed LLM-based tools for the IADB to identify and categorize structural reforms in Latin American countries from articles in newspapers.
 - **Text retrievers for market research (NODO Research)**: Accomplished a 50% reduction in the time required for qualitative data analysis by developing custom retrieval systems to interact with the transcripts of market research interviews ([GitHub template code](#)).
 - **Automated transcription generation (NODO Research)**: Achieved a 95% reduction in audio transcription time and a 20% cost reduction by developing an automatic transcription pipeline leveraging WhisperAI ([Hugging Face demo](#)).
 - **Master classes (IESE, CEMFI, Colombian Central Bank)**: Taught master classes in multiple institutions focusing on modern machine learning techniques for working with textual and unstructured data (e.g. IESE 2024: [GitHub](#) - [Website](#)).
 - **Income inequality visualization (Universidad de los Andes)**: Increased in-class engagement of undergraduate economics students by creating a tool to interact with household income data and explore the distributional consequences of different shocks ([Demo](#)).
- **University College London - Barcelona School of Economics** Remote
Researcher in NLP, Machine Learning, and Economics *March 2021 - Aug. 2023*
 - **Remote work prediction from job postings (Lightcast)**: 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 data is currently used by over 50 researchers and internally by the company ([Webpage](#) - [Media - Model](#)).
 - **Text analysis methods in Economics (Annual Review of Economics)**: Demonstrated techniques for algorithmic text analysis as part of the Annual Review of Economics. The associated repository has already garnered 90 stars ([Article](#) - [GitHub](#)).

- **Characterizing firms in a trade network:** 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 ([GitHub](#) - [Poster](#) - [Blog post](#)).

- **Instiglio**

Bogotá - Colombia

Data Analyst

Jan. 2018 - Aug. 2019

- **Model development (USAID):** Led the design of predictive models 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.
- **Data analysis:** Collected, cleaned, analyzed, and visualized historical data of all the company's projects to produce relevant insights for the Board's annual meeting.
- **Business development:** 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.
- **Achievements:** Got promoted from intern to analyst level 2 in a year.

EDUCATION

- **Barcelona School of Economics**

Barcelona, Spain

Master Program in Data Science; GPA: 8.57/10.0

Sept. 2019 – Sept. 2020

- **Relevant courses:** Machine Learning (A+), Text Mining (A+), Data Warehousing and Business Intelligence (A), Reinforcement Learning (A+), Data Visualization (A-)

- **Universidad de los Andes**

Bogotá - Colombia

Bachelor's Degree in Economics; GPA: 4.44/5.0

July 2013 – Dec. 2017

SKILLS

- **Technical skills:** Causal inference, machine learning, deep learning, natural language processing, large language models (chatbots, pre-training, fine-tuning, RAG, prompting), vector databases, computer vision, algorithmic fairness, data visualization, explainable machine learning
- **Tools and platforms:** Git, GCP, AWS, Docker, Hugging Face, Linux, Heroku, CSS, HTML
- **Soft skills:** Communication, teaching, public speaking, collaboration, problem-solving, creativity, adaptability
- **Languages:** Spanish (native), English (working proficiency), French (intermediate)