

Juan Manuel Pérez

MACHINE LEARNING ENGINEER · PROFESSOR

✉ juanmanuelzrp@gmail.com | jmperez@dc.uba.ar | 🏠 <http://jmperez.ar> | 📄 [finiteautomata](#) | 📺 [perezjuanma](#)

Education

Ph.D. in Computer Science

Buenos Aires, Argentina

UNIVERSIDAD DE BUENOS AIRES

2022

- Thesis: "Techniques and resources for hate speech detection in social networks"
- Advisors: Franco Luque and Agustín Gravano.

M.Sc. in Computer Science

Buenos Aires, Argentina

UNIVERSIDAD DE BUENOS AIRES

2016

- Thesis: "Metrics of acoustic-prosodic entrainment in speakers and their relationship with social features of dialogues."
- Advisor: Agustín Gravano and Ramiro Gálvez.
- GPA: 9.32/10

Research Employment

Postdoctoral researcher

Buenos Aires, Argentina

COMPUTER SCIENCE INSTITUTE (ICC), UBA-CONICET

2022-

Advisors: María Vanina Martínez and Laura Alonso Alemany.

Ph.D. Scholar

Buenos Aires, Argentina

COMPUTER SCIENCE INSTITUTE (ICC), UBA-CONICET

2016-2022

Advisors: Franco Luque and Agustín Gravano.

Teaching experience

Master in Data Science, Universidad de San Andrés

2022-present

ADJUNCT PROFESSOR

- Courses: Natural Language Processing, Supervised Learning, Thesis Seminar.

Faculty of Social Sciences, Universidad de Buenos Aires

2023

VISITING PROFESSOR

- Graduate course: Introduction to Computational Social Science.

Computer Science Department, FCEyN, Universidad de Buenos Aires

2016-2021

GRADUATE TEACHING ASSISTANT

- Courses: Numerical Methods, Programming Paradigms, Automata Theory, Introduction to Computer Science.

Computer Science Department, FCEyN, Universidad de Buenos Aires

2013

STUDENT TEACHING ASSISTANT

- Courses: Numerical Methods, Automata Theory.

Professional Experience

Machine Learning Engineer

2024-present

ACCENTURE

- Currently working in Text2SQL project, using state-of-the-art NLP models and Knowledge Graphs for an internal project.
Skills and Tools: Python, GCP, Gemini, Neo4J, LLMOps.

Freelance Data Scientist

2020-2024

SELF-EMPLOYED & TOPTAL

- Developed a Proof of Concept (PoC) for a legal tech startup automating contract analysis. We used custom NER models, knowledge graphs and semantic search.
- Collaborated with researchers and NGOs to analyze surveys and focus groups in African countries, using advanced NLP techniques to extract insights from unstructured text.
- Created a PoC for a local company developing virtual assistants, particularly focusing on a banking client.
- Analyzed social media data for public opinion research companies.
Skills and Tools: Python, Pandas, PyTorch, huggingface's transformers, spaCy, AWS, ElasticSearch.

Software Engineer

August 2014 - April 2016

ALTOROS

- Developed an MVP for *LawAdvisor*, a legal tech platform connecting users with professional advice.
 - Developed an MVP for *CattleMatch*, a company aiming at the digitalization of the cattle industry, from auctions to simple stock management.
 - Maintained and worked in new *T4 Spatial*: GIS data integration for sewer CCTV inspection.
- Skills and Tools:** Ruby, Ruby on Rails, JavaScript, Backbone.js, PostgreSQL, AWS, Agile.

Software Engineer

February 2013 - August 2014

DEVARTIS

- Worked as backend developer for many projects, focusing on scalability, good design and performance. Key projects included:
 - *Avalancha*: E-commerce platform for white goods and electronics. We developed this platform from scratch using Spree, Ruby and Rails, and ElasticSearch.
 - *Quilmes*: Official website for a leading beer company, enhancing brand visibility and customer engagement.
 - Besides from development, I was responsible for keeping the Scrum process on track and also dealt with clients directly.
- Skills and Tools:** Django, Ruby on Rails, Spree, PostgreSQL, RabbitMQ, AWS.

Game Developer

April 2010 - December 2011

METROGAMES

- Contributed to the development of popular Facebook and iPhone games, such as *Fashion World* and *Mall Dreams*
- Skills and Tools:** : ActionScript, Objective-C, MySQL, Facebook API, iPhone SDK, Cocos 2D.

Freelance Web Developer

February 2008 - November 2012

SELF-EMPLOYED

- Delivered web development solutions for personal clients and small businesses, both locally and abroad.
 - I was engaged in the development of three early-stage social networks for clients in Spain: Petuky, Cinemavip, and SigoJoven.
- Skills and Tools:** : PHP, Ruby on Rails, jQuery, MySQL, HTML, CSS.

Publications

Gender biases and hate speech: promoters and targets in the Argentinean political context

Laia Domenech, Juan Manuel Pérez, German Rosati, Diego Kozłowski

PLOS ONE, 2025

Exploring Large Language Models for Hate Speech Detection in Rioplatense Spanish

others Juan Manuel Pérez

NAACL Findings (To be published), 2025

Identification of emotions on Twitter during the 2022 electoral process in Colombia

Juan Jose Iguaran Fernandez, Juan Manuel Perez, Germán Rosati

Memories from JAIIO 10.1 (2024) pp. 60–73. 2024

MessiRve: A Large-Scale Spanish Information Retrieval Dataset

Francisco Valentini, Viviana Cotik, Damián Furman, Ivan Bercovich, Edgar Altszyler, Juan Manuel Pérez

arXiv preprint arXiv:2409.05994 (2024). 2024

Assessing the impact of contextual information in hate speech detection

Juan Manuel Pérez, Franco M Luque, Demian Zayat, Martín Kondratzky, Agustín Moro, Pablo Santiago Serrati, Joaquín Zajac, Paula Miguel, Natalia Debandi, Agustín Gravano

IEEE Access 11 (2023) pp. 30575–30590. IEEE, 2023

psentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks

Juan Manuel Pérez, Mariela Rajngewerc, Juan Carlos Giudici, Damián A. Furman, Franco Luque, Laura Alonso Alemany, María Vanina Martínez

Under review, 2023

A Spanish dataset for Targeted Sentiment Analysis of political headlines

Juan Manuel Pérez, Emilio Recart Zapata, Tomás Alves Salgueiro, Damián Furman, Pablo Nicolás Fernández Larrosa

Electronic Journal of SADIO (EJS) 22.1 (2023) pp. 53–66. 2023

Top-down modulation impairs priming susceptibility in complex decision-making with social implications

Franco Agustín Bernal, Tomás Alves Salgueiro, Axel Brzostowski, Emilio Recart Zapata, Ayelén Carames, Juan Manuel Pérez, Damián Furman, Martín Graziano, Pablo Nicolás Fernández Larrosa

Scientific Reports 12.1 (2022) p. 17867. Nature Publishing Group UK London, 2022

Exploiting user-frequency information for mining regionalisms in Argentinian Spanish from Twitter

Juan Manuel Pérez, Damián E Aleman, Santiago N Kalinowski, Agustín Gravano

Procesamiento del Lenguaje Natural 69 (2022) pp. 51–62. 2022

RoBERTuito: a pre-trained language model for social media text in Spanish

Juan Manuel Pérez, Damián Ariel Furman, Laura Alonso Alemany, Franco M Luque

Proceedings of the Thirteenth Language Resources and Evaluation Conference, 2022

ANDES at SemEval-2020 Task 12: A jointly-trained BERT multilingual model for offensive language detection

Juan Manuel Pérez, Aymé Arango, Franco Luque

Proceedings of the 14th International Workshop on Semantic Evaluation, 2020

Window Classifiers and Conditional Random Fields for Medical Report De-Identification.

Viviana Cotik, Franco M Luque, Juan Manuel Pérez

IberLEF@ SEPLN, 2019

Atalaya at SemEval 2019 Task 5: Robust Embeddings for Tweet Classification

Juan Manuel Pérez, Franco M Luque

Proceedings of the 13th International Workshop on Semantic Evaluation, 2019

Atalaya at TASS 2018: Sentiment Analysis with Tweet Embeddings and Data Augmentation.

Franco M Luque, Juan Manuel Pérez

TASS@ SEPLN, 2018

Cross-Linguistic Study of the Production of Turn-Taking Cues in American English and Argentine Spanish.

Pablo Brusco, Juan Manuel Pérez, Agustín Gravano

Interspeech, 2017

Using prosody to classify discourse relations

Janine Kleinhans, Mireia Farrús, Agustín Gravano, Juan Manuel Pérez, Catherine Lai, Leo Wanner

Interspeech 2017, Stockholm, Sweden, 2017

Disentrainment may be a Positive Thing: A Novel Measure of Unsigned Acoustic-Prosodic Synchrony, and its Relation to Speaker Engagement.

Juan M Pérez, Ramiro H Gálvez, Agustín Gravano

Interspeech, 2016

Funding

SCHOLARSHIPS

Postdoctoral fellowship

August 2022 - August 2025

CONICET

Doctoral fellowship

April 2016 - April 2022

CONICET

Undergraduate fellowship

August 2012 - August 2014

AGENCIA NACIONAL DE PROMOCIÓN CIENTÍFICA Y TECNOLÓGICA

GRANTS

PortADa: Port Arrival Data

2023-2026

HORIZON EUROPE, EUROPEAN UNION

- Digital humanities project dedicated to construct knowledge and a database in maritime and social history, gathered from port arrival data in the 19th and 20th centuries.
 - NLP and CV techniques to extract information from printed sources.
 - The project is a collaboration from several nodes across Europe and Latin America.
- Role: Collaborator

OpenAI Researcher Access

2024

OPENAI

- Research credits on the OpenAI platform (\$1,000 granted)
 - Granted to research on automatic generation of QA and IR resources.
- Role: PI

PIUBA: Interdisciplinary projects of the Universidad de Buenos Aires

2020-2023

UNIVERSIDAD DE BUENOS AIRES

- Project dedicated to the study of social marginalization in Argentina. Particularly focused on hate speech detection in social media.
- Role: Collaborator

Google TPU Research Cloud

GOOGLE CLOUD EDUCATION GRANT

- TPU Access and GCP credits (\$2,000 granted)
 - Used to train *RoBERTuito*, a Spanish language model.
- Role: PI

TECHNOLOGY TRANSFER AGREEMENTS

Virtual Assistants using LLMs

2023

CONICET-EDS CONNECTION

IF-2023-91838154-APN-GVT#CONICET

- Development of a proof-of-concept for virtual assistants using Large Language Models (LLMs), for the EDS Connection company.
- The PoC consisted in a retrieve-and-generate assistant for customer service using an unstructured knowledge base.
- Experimented with re-rankers and dense retrievers to enhance current solution.

Academic Service

MANAGEMENT

Member of Department Council

2020-2022

DEPARTMENT OF COMPUTER SCIENCE, FACULTAD DE CS. EXACTAS, UBA

Member of Department Council

2020-2022

LICENCIATURA IN DATA SCIENCE, FACULTAD DE CS. EXACTAS, UBA

Member of Faculty Council

2018-2022

FACULTAD DE CS. EXACTAS, UBA

THESIS ADVISOR

Matías Peirano

2024-present

AUTOMATIC DETECTION OF TEXT GENERATED BY LANGUAGE MODELS

Master in Data Science, UdeSA

Paula Iglesias

2024-present

ASPECT-BASED SENTIMENT ANALYSIS ON USER COMMENTS IN SPANISH GASTRONOMIC SITES

Master in Data Science, UdeSA

Miguel Szjenblum

2024-present

LANGUAGE MODELS FOR CHECK-WORTHINESS DETECTION

Master in Data Science, UdeSA

Melina Bava

2023-present

RECOMMENDATION SYSTEMS FOR FOOD PRODUCTS

Master in Management + Analytics, UTDT

Clara Villalba

2023-present

VQA MODELS FOR VISUALLY IMPAIRED PEOPLE

Master in Data Mining, Exactas, UBA

Juan José Iguarán Fernández

2023

DETECTION OF EMOTIONS IN PRESIDENTIAL ELECTIONS IN COLOMBIA

Master in Data Mining, Exactas, UBA

Paula Luvini

2024

AI TOOLS TO FIGHT DIGITAL VIOLENCE AGAINST WOMEN

Master in Data Science, UdeSA

Juan Carlos Giudici

2023

DOMAIN ADAPTATION IN NEURAL MODELS FOR OPINION EXTRACTION TASKS

M.Sc. in Computer Science, Exactas, UBA

Joaquín López Calvo

2022

DOES TOXICITY PAY OFF? ANALYSIS OF THE ASSOCIATION BETWEEN TOXICITY AND VIEWS AND INTERACTIONS IN HUMOROUS VIDEOS ON YouTube.

B.Sc. in Communication Sciences, FSOC, UBA

Damian E. Aleman

2017

A COMPUTATIONAL METHOD TO DETECT CONTRASTIVE LEXICON

M.Sc. in Computer Science, Exactas, UBA

THESIS EXAMINER

M.Sc. in Computer Science

- Nicolás Mastropasqua (2023)
- Gonzalo Sánchez Cano (2022)
- Marcelo Sancinetti (2022)
- Leandro Matayoshi (2018)

Master in Data Mining

- Vanesa Copa

Master in Management + Analytics

- Juan Horgan

REVIEWER

European Chapter of the Association for Computational Linguistics (EACL)

REVIEWER

2023

First workshop of Cross-Cultural Considerations in NLP (C3NLP), EACL

PROGRAM COMMITTEE

2023

Simposio Argentino de Inteligencia Artificial (ASAI)

PROGRAM CHAIR

2023

Awards

Department of Computer Science, Exactas, UBA

“CONCURSO DE TESIS 2016”

2016

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Universidad Torcuato di Tella
Escuela de Negocios