

Juan Manuel Pérez

Email juanmanuelzrp@gmail.com || jmperez@dc.uba.ar

Homepage <http://jmperez.ar>

Education

Ph.D. in Computer Science, Universidad de Buenos Aires (2022)

Advisors: Franco Luque and Agustín Gravano.

Thesis: “Techniques and resources for hate speech detection in social networks”

M.Sc. in Computer Science, FCEyN, UBA, (2016). PGA: 9.32.

Advisor: Agustín Gravano

Thesis: “Metrics of acoustic-prosodic entrainment in speakers and their relationship with social features of dialogues.”

Research employment

Postdoctoral researcher (2022-), Computer Science Institute (ICC), UBA-CONICET.

Ph.D. Scholar (2016-2022) Computer Science Institute (ICC), UBA-CONICET.

Publications

L. Domenech, **J.M. Pérez**, G. Rosati, D. Kozłowski. *Gender biases and hate speech: promoters and targets in the Argentinean political context*. Under review at PLOS ONE.

J.M. Pérez, M. Rajngewerc, Juan Carlos Giudici, Damián Ariel Furman, Franco Luque, Laura Alonso Alemany and María Vanina Martínez. *psentimiento: A python toolkit for sentiment analysis and SocialNLP tasks*. Under review at Language Resources and Evaluation Journal.

J.M. Pérez, F. Luque, D. Zayat, M. Kondratzky, A. Moro, P. Serrati, ... A. Gravano and V. Cotik. *Assessing the impact of contextual information in hate speech detection*. IEEE Access (2023)

F.A. Bernal, T.A. Salgueiro, A. Brzostowski, E.R. Zapata, A. Carames, **J.M. Pérez**, D. Furman, M. Graziano, P.N. Fernández Larrosa. *Top-down modulation impairs priming susceptibility in complex decision-making with social implications*. Scientific Reports, (2022)

J. M. Pérez, T.A Salgueiro, D. Furman, P.N. Fernández Larrosa. *A Spanish dataset for Targeted Sentiment Analysis of political headlines*. Electronic Journal of SADIO (EJS), (2022)

J. M. Pérez, D.E. Aleman, S.N. Kalinowski, A. Gravano. *Exploiting user-frequency information for mining regionalisms in Argentinian Spanish from Twitter*. Procesamiento del Lenguaje Natural(2022)

J.M. Pérez, D.A. Furman, L. Alonso Alemany, and F. Luque. *RoBERTuito: a pre-trained language model for social media text in Spanish*, LREC, 2022.

J.M. Pérez, A. Arango, F. Luque. *ANDES at SemEval-2020 Task 12: A jointly-trained BERT multilingual model for offensive language detection*, Proceedings of the 14th International Workshop on Semantic Evaluation. 2020.

V. Cotik, **J.M. Pérez**, F. Luque. *Window Classifiers and Conditional Random Fields for Medical Report De-Identification*, Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019).

J.M. Pérez, F. Luque. *Atalaya at SemEval 2019 Task 5: Robust Embeddings for Tweet Classification*. Proceedings of the 13th International Workshop on Semantic Evaluation. 2019.

F. Luque, **J.M. Pérez**. *Atalaya at TASS 2018: Sentiment analysis with tweet embeddings and data augmentation* Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN (TASS 2018).

P. Brusco, **J.M. Pérez**, A. Gravano. *Cross-linguistic study of the production of turn-taking cues in American English and Argentine Spanish*, Interspeech (2017)

J. Kleinhans, M. Farrús, A. Gravano, **J.M. Pérez**, C. Lai, L. Wanne. *Using Prosody to Classify Discourse Relations* Interspeech (2017)

J.M. Pérez, R. H. Gálvez, and A. Gravano. *Disentrainment may be a positive thing: A novel measure of unsigned acoustic-prosodic synchrony, and its relation to speaker engagement*. Interspeech 2016.

Teaching experience

Adjunct Professor (2022-present), Master in Data Science, Universidad de San Andrés. Courses: Natural Language Processing, Supervised Learning, Thesis Seminar.

Visiting Professor (2023) Faculty of Social Sciences, Universidad de Buenos Aires. Graduate course: Introduction to Computational Social Science

Graduate Teaching Assistant (2016-2021) Computer Science Department, FCEyN, Universidad de Buenos Aires. Courses: Numerical Methods, Programming Paradigms, Automata Theory, Introduction to Computer Science.

Student Teaching Assistant (2013) Computer Science Department, FCEyN, UBA. Courses: Numerical Methods, Theory of Languages

Professional Experience

Freelance Data Scientist (2020-present) Worked in NLP projects for companies in Argentina and in the US.

Software Engineer, Altoros (August 2014 - April 2016) Web Development in Ruby+JS; Scrum Master.

Software Engineer, Devartis (February 2013 - August 2014) Web Development in Python+Ruby+JS

Developer, Manas (May 2012, July 2012) Worked on a refactoring of an e-commerce system.

Developer, Metrogames (April 2010 - December 2011) Game development on iPhone and Facebook.

Developer, gRantAire (February 2009, November 2009). Web development.

Grant funding

PortADa: Port Arrival Data. Grant HORIZON-MSCA-2022-SE-01 (2023-2026) Horizon Europe. Role: Collaborator

PIUBA: Interdisciplinary project on social marginalizations, UBA (2020-2023). Role: Collaborator

Google TPU Research Cloud, *Google Cloud Education Grant.* Used to train *RoBERTuito* model

Google Cloud Education Grant, *Google Cloud Education Grant.* Used to train *RoBERTuito* model

Technology Transfer Agreements

Virtual Assistants using LLMs, CONICET-EDS Connection. Prototyping of a virtual assistants using LLMs (mainly OpenAI's GPT-3.5). Agreement no. IF-2023-91838154-APN-GVT#CONICET

Academic Service

Academic management

Member of Department Council, Computer Science Department, FCEyN, UBA (2020-2022)

Member of Department Council, Licenciatura in Data Science, FCEyN, UBA (2020-2022)

Member of Faculty Council, FCEyN, UBA, (2018-2022)

Thesis examiner

Juan Horgan, Master in Management + Analytics, Universidad Torcuato di Tella (2023)

Martín Pastorino, Master in Data Science, Universidad de San Andrés (2023)

Nicolás Mastropasqua, Master in Computer Science, Faculty of Exact and Natural Sciences (FCEyN), UBA (2023)

Gonzalo Sánchez Cano, Master in Computer Science, FCEyN, UBA (2022)

Marcelo Sancinetti, Master in Computer Science, FCEyN, UBA (2022)

Leandro Matayoshi, Master in Computer Science, FCEyN, UBA (2018)

Reviewer

Reviewer, European Chapter of the Association for Computational Linguistics (EACL), 2023

Program Committee, First workshop of Cross-Cultural Considerations in NLP (C3NLP), EACL, 2023

Program Chair, Simposio Argentino de Inteligencia Artificial (ASAI), 2023

Thesis advisor

Clara Villalba, Master in Data Mining, FCEyN, UBA (2023-present)

Melina Bava, Master in Management+Analytics, Universidad Torcuato di Tella (2023-present)

Paula Luvini, Master in Data Science, Universidad de San Andrés (2023-2024)

Juan José Iguarán Fernández, Master in Data Mining, FCEyN, UBA . Thesis: *Emotion detection in Presidential Election in Colombia* (2023)

Juan Carlos Giudici, Master in Computer Science, FCEyN, UBA. Thesis: *Domain adaptation for opinion mining tasks.* (2023)

Joaquín López Calvo, Master in Communication Sciences, Faculty of Social Sciences, UBA. Dissertation *An analysis of the association between toxicity and visits and interactions in humorous videos on YouTube* (2022)

Damian E. Aleman, Master in Computer Science, FCEyN, UBA. *Towards a computational method to detect contrastive lexicon*, co-advised with Agustín Gravano and Santiago Kalinowski (2017)

Awards

Thesis award 2016, Departamento de Computación, FCEN, UBA (for MSc thesis)

Last update: February 26, 2024