Francielle Vargas, MSc and PhD in Natural Language Processing

francielleavargas@usp.br
https://franciellevargas.github.io/

Main Skills   Natural Language Processing, Computational Social Science, Machine Learning, Responsible AI / Explainability and Interpretability and Fairness.

Education   *PhD, Computer Science and Computational Mathematics*
University of São Paulo
Socially Responsible and Explainable Automated Fact-Checking and Hate Speech Detection.
Artificial Intelligence, Natural Language Processing, Machine Learning
2019 - 2024

MSc, Computer Science and Computational Mathematics
University of São Paulo
Semantic Clustering of Aspects for Opinion Mining.
Artificial Intelligence, Natural Language Processing, Machine Learning
2015 - 2017

BA, Linguistics
Federal University of Minas Gerais
2010 - 2014

BS, Information Systems
The Pontifical Catholic University of Minas Gerais
2006 - 2009*

Awards & Honors   Google Latin America Research Awards (LARA 2024)
Diversity and Inclusion Grant Award (EMNLP 2024)
Diversity and Inclusion Grant Award (NAACL 2024)
Outstanding Academic Project and Honorable Mention (UFMG 2013)
Outstanding Academic Project and Honorable Mention (UFMG 2012)

Invited Speaker   2024: [Invited Researcher] Computational Social Science Lab
University of Southern California (USC). Los Angeles, California
Title: *Fighting Misinformation and Radicalism: Socially Responsible and Explainable Fact-Checking and Hate Speech Detection.* see

2024: [Keynote Speaker] Conference cum Conclave on Emerging trends in Journalistic and Media Practices
DG Vaishnav College (DDGDV). Chennai, India
Title: *Predicting Sentence-Level News Source Reliability for Fact-Checking.* see

2023: [Invited Speaker] Conference on Harmful Online Communication
Leibniz Institute for the Social Sciences (GESIS), Cologne, Germany
Title: *Countering Harmful Online Communication in Brazil: Predicting Fine-Grained Factuality of News and Offensive Context of Social Media Comment.* see

---

Publications　　　**2025**

MFTCXplain: A Benchmark Dataset for Explainable Moral Sentiment Detection
Francielle Vargas, Jackson Trager, Matteo Guida, Diego Alves, Ameeta Agrawal, Morteza Dehghan.
under review

HateBRXplain: A Benchmark Dataset with Human-Annotated Rationales for Explainable Hate Speech Detection in Brazilian Portuguese
Isadora Salles, Francielle Vargas, Fabrício Benevenuto.
under review

Discourse Annotation Guideline for Low-Resource Languages
Francielle Vargas, Wolfgang Schmeisser-Nieto, Zohar Rabinovich, Thiago A.S. Pardo, Fabrício Benevenuto.
Natural Language Engineering Journal. to appear

2024

Context-Aware and Expert Data Resources for Brazilian Portuguese Hate Speech Detection
Francielle Vargas, Isabelle Carvalho, Thiago Pardo, Fabrício Benevenuto
Natural Language Engineering Journal. see

Improving Explainable Fact-Checking via Sentence-Level Factual Reasoning
Francielle Vargas, Isadora Salles, Diego Alves, Ameeta Agrawal, Thiago Pardo, Fabrício Benevenuto.
7th Fact Extraction and VERification Workshop (FEVER @ EMNLP 2024). see

HausaHate: An Expert Annotated Corpus for Hausa Hate Speech Detection
Francielle Vargas, Samuel Guimarães, Shamsuddeen H. Muhammad, Diego Alves, Ibrahim Said Ahmad, Idris Abdulmumin, Diallo Mohamed, Thiago Pardo, Fabrício Benevenuto.
8th Workshop on Online Abuse and Harms (WOAH @ NAACL 2024). see

Extended Multimodal Hate Speech Event Detection During Russia-Ukraine Crisis - Shared Task at CASE 2024
Surendrabikram Thapa, Kritesh Rauniyar, Farhan Jafri, Hariram Veeramani, Raghav Jain, Sandesh Jain, Francielle Vargas, Ali Hürriyetoğlu, Usman Naseem.
Workshop Proceedings of the 14th Recent Advances in Natural Language Processing (CASE @ RANLP 2024). see

2023

Predicting Sentence-Level Factuality of News and Bias of Media Outlets
Francielle Vargas, Kokil Jaidka, Thiago A.S. Pardo, Fabrício Benevenuto.
Recent Advances in Natural Language Processing (RANLP 2023). see

Socially Responsible Hate Speech Detection: Can Classifiers Reflect Social Stereotypes?
Francielle Vargas, Isabelle Carvalho, Ali Hürriyetoğlu, Thiago A.S. Pardo, Fabrício Benevenuto.
Recent Advances in Natural Language Processing (RANLP 2023). see

NoHateBrazil: A Brazilian Portuguese Text Offensiveness Analysis System
Francielle Vargas, Isabelle Carvalho, Wolfgang Schmeisser-Nieto, Fabrício Benevenuto, Thiago A.S. Pardo.
Recent Advances in Natural Language Processing (RANLP 2023). see

Multimodal Hate Speech Event Detection - Shared Task 4, CASE 2023
Surendrabikram Thapa, Farhan Jafr, Ali Hürriyetoğlu, Francielle Vargas, Roy Ka-Wei Le, Usman Naseem.
Workshop Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (CASE @ EACL 2023). see

2022

HateBR: A Large Expert Annotated Corpus of Brazilian Instagram Comments for Abusive Language Detection
Francielle Vargas, Isabelle Carvalho, Fabiana R. Góes, Thiago A.S. Pardo, Fabrício Benevenuto.
13th Conference on Language Resources and Evaluation (LREC 2022). see

Rhetorical Structure Approach for Online Deception Detection: A Survey
Francielle Vargas, Jonas D'Alessandro, Zohar Rabinovich, Fabrício Benevenuto, Thiago A.S. Pardo.
13th Conference on Language Resources and Evaluation (LREC 2022). pp.5906-5915.

Studying Dishonest Intentions in Texts
Francielle Vargas, Thiago A. S. Pardo.
Deceptive AI. Springer: Communications in Computer and Information Science. see

Extended Multilingual Protest News Detection - Shared Task 1, CASE 2021 and 2022
Ali Hürriyetoğlu, Osman Mutlu, Fırat Duruşan, Onur Uca, Alaeddin Gürel, Benjamin J. Radford, Yaoyao Dai, Hansi Hettiarachchi, Niklas Stoehr, Tadashi Nomoto, Milena Slavcheva, Francielle Vargas, Aaqib Javid, Aaqib Javid, Erdem Yörük.
Workshop Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (CASE @ EMNLP 2022). see

2021

Contextual-Lexicon Approach for Abusive Language Detection
Vargas, F., Fabiana R. Góes, Isabelle Carvalho, Fabrício Benevenuto, Thiago A.S. Pardo.
Recent Advances in Natural Language Processing (RANLP 2021). see

Towards Discourse-Aware Models for Multilingual Fake News Detection
Francielle Vargas, Fabrício Benevenuto, Thiago A.S. Pardo.
Recent Advances in Natural Language Processing (RANLP 2021). see

Implicit Opinion Aspect Clues in Portuguese Texts: Analysis and Categorization
Mateus T. Machado, Thiago A. S. Pardo, Evandro E. S. Ruiz, Ariani Di Felippo, Francielle Vargas.
15th International Conference on the Computational Processing of Portuguese (PROPOR 2021). see

Towards A Diverse, Inclusive, Accessible and Equitable AAAI International Conference on Web and Social Media (ICWSM)
Jason R.C. Nurse, Francielle Vargas, Naeemul Hassan, Dakuo Wang, Panagiotis Andriotis, Amira Ghenai, Kokil Jaidka, Eni Mustafaraj, Kenneth Joseph and Brooke Foucault Welles
Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media (ICWSM 2021). pp. 01-12. Held Online. see

2020 and before

Linguistic Rules for Fine-Grained Opinion Extraction
Francielle Vargas, Thiago A. S. Pardo.
Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media

(SocialSens @ ICWSM 2020). see

Identifying Fine-Grained Opinion and Classifying Polarity on Coronavirus Pandemic
Francielle Vargas, Santos, R.S.S., F., Rocha, P.R.
9th Brazilian Conference on Intelligent Systems (BRACIS 2020). see

Aspect Clustering Methods for Sentiment Analysis
Vargas, F., Pardo, T.A.S.
13th International Conference on the Computational Processing of Portuguese (PROPOR 2018). see

The Coreference Annotation of the CSTNews Corpus
Thiago A. S. Pardo, Jorge Baptista, Magali S. Duran, Maria das Graças Nunes, Fernando Nóbrega, Sandra Aluísio, Ariani Di Felippo, Eloize Seno, Raphael R. Silva, Rafael Anchieta, Henrico Brum, Márcio Dias, Rafael Martins, Erick Maziero, Jackson Souza, Francielle Vargas
2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval @ SEPLN 2017). pp. 102-112. Murcia, Spain. see

Organizing and Program Committee	**Organizing Committee**

1. Co-organizer. *WOAH 2015.* see
2. Data(set) Co-chair. *ICWSM 2023.* see
3. Accessibility Chair. *ICWSM 2022.* see
4. Diversity, Equity, and Inclusion Chair. *ICWSM 2021.* see
5. Co-organizer. 6th CASE@*RANLP 2023.* see
6. Co-organizer. 5th CASE@*EMNLP 2022.* see

Program Committee

- *EMNLP* (2022) (2023) (2024)
- *NAACL* (2022) (2024)
- *ACL* (2023)
- *LREC* (2024)
- *COLING* (2025)
- *ICWSM* (2025)
- *CIKM* (2024)(2023)
- *HICSS*(2023)
- *WOAH* (NAACL-HLT 2022) (ACL 2023) (NAACL 2024)
- *FEVER* (EACL 2023) (EMLP 2024)
- *CODI* (COLING 2022) (ACL 2023) (EACL 2024)
- *ARGMINING-CMNA* (COMMA 2022) (COLING 2022) (ICLP 2023)
- *SoCal NLP* (2022)

Experience	2024-2024: *Google PhD Fellowship*, Federal University of Minas Gerais, Brazil
	2024-2024: *Visiting Researcher*, University of Southern California, USA
	2023-2023: *Invited Researcher*, Leibniz Institute for the Social Sciences, Germany
	2021-2022: *Data Scientist*, Cisco Webex, USA
	2015-2023: *Graduate Researcher*, University of São Paulo, Brazil
	2014-2015: *System Analyst*, Unisys, Brazil
	2010-2014: *Research Assistant*, Federal University of Minas Gerais, Brazil

Short Curses	2024: *2nd Mexican NLP Summer School*, NAACL 2024, Mexico. see
	2023: *2nd Summer School on Deep Learning in NLP*, RANLP 2024, Bulgaria. see
	2022: *2nd Advanced NLP School*, Université Grenoble Alpes, France. see
	2021: *4th Advanced School in Big Data Analysis*, ICMC-USP, Brazil. see
	2020: *10th Lisbon Machine Learning School*, INESC, Portugal. see
	2015: *4th Summer School in Computer Science*, Federal University of Minas Gerais, BR. see
	2014: *Software Engineering (MBA)*, Pontifical Catholic University of Minas Gerais, BR. see

Data Resources	HateBRXplain: A Benchmark Dataset with Human-Annotated Rationales for Explainable Hate Speech Detection
	MFTCXplain: A Benchmark Dataset for Explainable Moral Sentiment Detection
	HateBR: A Benchmark Expert Hate Speech Dataset of Brazilian Instagram comments
	FactNews: A Benchmark Dataset for Sentence-Level Factuality and Media Bias Prediction
	HausaHate: A Benchmark Expert Dataset for Hausa Hate Speech Detection
	OPCovidBR: A Dataset of Covid-19 Tweets For Aspect-Based Sentiment Analysis
	AspectPRO: A Dataset of Consumer Reviews for Aspect-Based Sentiment Analysis
	PRO: A Set of Taxonomies for Aspect-Based Sentiment Analysis
	MOL: A Context-Aware Multilingual Offensive Lexicon

Methods (patents)	B+M: Contextualized Bag-of-Words with Feature Saliency for Explainable HS Detection
	SSA: A Counterfactual Explanation Measure to Assess Social Bias in Hate Speech Classifiers
	SELFAR: Sentence-Level Factual Reasoning for Explanable Fact-Checking

Computer Programs	FACTual: A Fact-Checking and News Source Reliability Estimation System
	NoHateBrazil: A Brazilian Portuguese Text Offensiveness Analysis System
	OPCluster-PT: A System for Extraction and Clustering of Opinions

Coding	Python, NLTK, spaCy, Gensim, Keras, Tensorflow, PyTorch, SQL, RDF(s), OWL, Unix.

Conference Participation	ACL 2020, ACL 2023, EMNLP 2020, EMNLP 2024, LREC 2020, LREC 2022, RANLP 2021, RANLP 2023, NAACL 2024, ICWSM 2021,

Recommendation Letters	Norah Abokhodair. *Microsoft Research.* letter
	Ameeta Agrawal. *Portland State University, USA.* letter
	Carlos Ramisch. *LIS–CNRS–Aix Marseille Université, France.* letter
	Fabricio Benevenuto. *Federal University of Minas Gerais, Brazil.* letter