Francielle Vargas

francielleavargas@usp.br +55 31 999685891
[web page] [dblp] [google scholar] [linkedin] [github]

Education	*PhD, Computer Science and Computational Mathematics* University of São Paulo 2019 - 2023 *MSc, Computer Science and Computational Mathematics* University of São Paulo Thesis: Semantic Clustering of Aspects for Opinion Mining. see 2015 - 2017 *BA, Linguistics* Federal University of Minas Gerais 2010 - 2014 *BSc, Information Systems* Pontifical Catholic University of Minas Gerais 2006 - 2009

Invited speaker	Leibniz Institute for the Social Sciences (GESIS), 2023 Conference on Harmful Online Communication. Cologne, Germany

---

Research projects	2020 - 2023: *Exploring Factual Reasoning for Fact-Checking* University of São Paulo, Brazil 2022 - 2023: *Measuring Social Biases in Multilingual Masked Language Models* Sorbonne University, France 2022 - 2023: *Predicting Hate Speech for Hausa African Indigenous Language* University of São Paulo, Brazil 2020 - 2023: *Discourse-Aware Computational Resources for Fake News Detection* University of São Paulo, Brazil 2020 - 2022: *Socially Responsible Methods and Resources for Hate Speech Detection* University of São Paulo, Brazil 2022 - 2022: *Expanding Evaluation Data for the Multilingual Protest News Detection* Koç University, Turkey 2020 - 2020: *Detecting Antisemitism on Social Media* Indiana University Bloomington, United States 2019 - 2020: Aspect-Based Sentiment Analysis University of São Paulo, Brazil 2015 - 2017: Summarization for Clever Information Access University of São Paulo, Brazil

---

Organizing and program committee	**Organizing Committee**

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

Program Committee

- **ACL**
  Computational Social Science and Cultural Analytics Track (2023)
  Discourse and Pragmatics Track (2023)
- **EMNLP**
  Computational Social Science and Cultural Analytics Track (2022) (2023)
  Discourse and Pragmatics Track (2022) (2023)
- **NAACL**
  NAACL Student Research Workshop (2022)
- **ICWSM**
  Dataset and Demo Track (2023)
- **CIKM**
  Dataset and Demo Track (2023)
- **HICSS**
  Digital and Social Media Track (2023)
- **SoCal NLP**
  5th Southern California Natural Language Processing Symposium (2022)
- **WOAH**
  7th Workshop on Online Abuse and Harms (ACL 2023)
  6th Workshop on Online Abuse and Harms (NAACL-HLT 2022)
- **FEVER**
  6th Fact Extraction and VERification (EACL 2023)
- **CASE**
  5th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text (EMNLP 2022)
  4th International Workshop Challenges and Applications of Automated Extraction of Socio-political Events from Text (ACL-IJCNLP 2021).
- **CODI**
  4th Workshop on Computational Approaches to Discourse (ACL 2023)
  3rd Workshop on Computational Approaches to Discourse (COLING 2022)
- **ARGMINING**
  9th Workshop on Argument Mining (COLING 2022)
  22th Workshop on Computational Models of Natural Argument COMMA 2022)
  1st Workshop on Argumentation & Machine Learning (COMMA 2022).

Awards received	2021: **Research PhD Fellowship**, SINCH Sweden see
	2013: **Academic Relevance**, Federal University of Minas Gerais.see
	2013: **Honorable Mention**, Federal University of Minas Gerais. see
	2012: **Academic Relevance**, Federal University of Minas Gerais. see

Publications

Pre-Print

FACTual: Explainable Fact-Checking Claims with Factual Reasoning
Vargas, F., Pardo, T.A.S., Benevenuto, F.
pp.1-7. see

TEAR: A Hate Speech Dataset Of Hausa Facebook Comments
Vargas, F., Guimaraes, S., Alves, D., Hassan, S., Lamine, D. M. Benevenuto, F.
pp.1-10. see

Discourse Annotation Guideline for Low-Resource Languages
Vargas, F., Rabinovich, Z., Benevenuto, F., Pardo, T.A.S.
pp.1-56. see

2023

Context-Aware and Expert Data Resources for Brazilian Portuguese Hate Speech Detection
Vargas, F., Carvalho, I., Pardo, T.A.S., Benevenuto, F.
Natural Language Engineering Journal. pp.1-21. accepted

Socially Responsible Hate Speech Detection: Can Classifiers Reflect Social Stereotypes?
Vargas, F., Carvalho, I., Hürriyetoğlu, A., Pardo, T.A.S., Benevenuto, F.
Recent Advances in Natural Language Processing (RANLP). pp.1-10. Varna, Bulgaria.
accepted

Predicting Sentence-Level Factuality of News and Bias of Media Outlets
Vargas, F., Jaidka, K., Pardo, T.A.S., Benevenuto, F.
Recent Advances in Natural Language Processing (RANLP). pp.1-10. Varna, Bulgaria.
accepted

NoHateBrazil: A Brazilian Portuguese Text Offensiveness Analysis System
Vargas, F., Carvalho, I., Schmeisser-Nieto, W., Benevenuto, F., Pardo, T.A.S.
Recent Advances in Natural Language Processing (RANLP). pp.1-7. Varna, Bulgaria. accepted

Multimodal Hate Speech Detection
Thapa, S., Jafr, F. A., Hürriyetoğlu, A., **Vargas, F.**, Lee, R. K., Naseem, U.
6th International Workshop Challenges and Applications of Automated Extraction of Socio-Political Events from Text (RANLP). pp.1-9. Varna, Bulgaria. accepted

2022

Rhetorical Structure Approach for Online Deception Detection: A Survey
Vargas, F., D'Alessandro, J., Rabinovich, Z., Benevenuto, F., Pardo, T.A.S.
13th Conference on Language Resources and Evaluation (LREC). pp.5906-5915. Marseille, France. see

HateBR: A Large Expert Annotated Corpus of Brazilian Instagram Comments for Abusive Language Detection
Vargas, F., Carvalho, I., Góes, F.R., Pardo, T.A.S., Benevenuto, F.
13th Conference on Language Resources and Evaluation (LREC). pp.7174–7183. Marseille, France. see

Studying Dishonest Intentions in Brazilian Portuguese Texts
Vargas, F., Pardo, T.A.S.
Deceptive AI. Springer, vol 1296. pp.166–178. see

Extended Multilingual Protest News Detection
Hurriyetoglu, A., Mutlu, O., San, F. D., Uca, O., Gurel, A. S., Radford, B., Dai, Y., Hettiarachchi, H., Stoehr, N., Nomoto, T., Slavcheva, M., **Vargas, F.**, Javid, A., Beyhan, F., Yoruk, E.
5th International Workshop Challenges and Applications of Automated Extraction of Socio-Political Events from Text (EMNLP). pp.223–228. Abu Dhabi, Arab Emirates. see

2021

Contextual-Lexicon Approach for Abusive Language Detection
Vargas, F., Góes, F.R., Carvalho, I., Pardo, T.A.S., Benevenuto, F.
Recent Advances in Natural Language Processing (RANLP). pp.1442-1451. Held Online.see

Towards Discourse-Aware Models for Multilingual Fake News Detection
Vargas, F., Benevenuto, F., Pardo, T.A.S.
Recent Advances in Natural Language Processing (RANLP). pp.210-218. Held Online. see

Implicit Opinion Aspect Clues in Portuguese Texts: Analysis and Categorization
Machado, M. T., Pardo, T.A.S., Ruiz, E. E. S., Di Felippo, A., **Vargas, F.**
15th International Conference on the Computational Processing of Portuguese (PROPOR). pp.68-78. Fortaleza, Brazil. see

2020

Linguistic Rules for Fine-Grained Opinion Extraction
Vargas, F., Pardo, T.A.S.
5th International Workshop on Social Sensing: Special Edition on Narrative Analysis on Social Media (ICWSM). pp.1-6. Held Online. see

Identifying Fine-Grained Opinion and Classifying Polarity on Coronavirus Pandemic
Vargas, F., Santos, R.S.S., F., Rocha, P.R.
9th Brazilian Conference on Intelligent Systems (BRACIS). pp.511-520. Rio Grande, Brazil. see

2019 and before

Aspect Clustering Methods for Sentiment Analysis
Vargas, F., Pardo, T.A.S.
13th International Conference on the Computational Processing of Portuguese (PROPOR). pp.365-374. Canela, Brazil. see

The Coreference Annotation of the CSTNews Corpus
Pardo, T.A.S., Baptista, J., Duran, M.S., Nunes, M.G.V., Nóbrega, F.A.A., Aluísio, S.M., Di Felippo, A., Seno, E.R.M., Silva, R.R., Anchieta, R.T., Brum, H.B., Dias, M.S., Martins, R.S.O., Maziero, E.G., Souza, J.W.C., **Vargas, F.**
2nd Workshop on Evaluation of Human Language Technologies for Iberian Language (SE-PLN). pp.102-112. Murcia, Spain. see

Book published / organized	**Por que amo Llansol?** Bittencourt, A.; Costa, F. A.; **Vargas, F. A.** Publicações Viva Voz, FALE/UFMG: Belo Horizonte: 34p. ISBN 978-85-7758-227-3. pdf

Previous positions	**Academic position**
	2019/2023: **Graduate Research Assistant - Dept. of Computer Science**. University of São Paulo. see
	2021/2021:**Graduate Teaching Assistant in Neural Networks and Deep Learning**. University of São Paulo. see
	2020/2020:**Graduate Teaching Assistant - Computing Theory and Compilers**. University of São Paulo. see
	2016/2016: **Graduate Teaching Assistant in Computer Seminars**. University of São Paulo. see.
	2015/2017: **Graduate Research Assistant - Dept. of Computer Science**. University of São Paulo. see
	2011/2014: **Undergraduate Research Assistant - Dept. of Linguistics**. Federal University of Minas Gerais. see
	Industry position
	2021/2023 **Doctoral Researcher**. Sinch.
	2021/2022 **Data Scientist**. Webex-Cisco
	2014/2015: **System Analyst**. Unisys.

Conference participation	2023: Recent Advances in Natural Language Processing (RANLP). to appear
	2023: 61st Annual Meeting of the Association for Computational Linguistics (ACL). see
	2022: 13th International Conference on Language Resources and Evaluation (LREC). see
	2021: Recent Advances in Natural Language Processing (RANLP). see
	2021: 15th International AAAI Conference on Web and Social Media (ICWSM). see
	2020: 2022 Empirical Methods in Natural Language Processing (EMNLP). see
	2020: 12th International Conference on Language Resources and Evaluation (LREC). see

Mentoring	**Computer Science Undergraduate Student**: Lucas Sobral Fontes Cardoso *Web resources for an aspect identification and clustering tool in fine-grained opinion mining.* **University of São Paulo**, 2020. see

Summer school and short courses	2023: 2nd Summer School on Deep Learning in NLP, Varna, Bulgaria. see
	2022: 2nd Advanced Language Processing School, Université Grenoble Alpes, France. see
	2021: 4th Brazilian Advanced School in Big Data Analysis, USP, São Paulo, Brazil. see
	2020: 10th Lisbon Machine Learning School, INESC Lisboa, Portugal. see
	2019: Introduction on the Stars - Astrophysics, USP, São Paulo, Brazil. see

Technical-scientific resources	**FACTual**: Automated fact-checking and news credibility. web system
	OpCluster: Fine-grained opinion extraction and clustering tool. web system
	NoHateBrazil: A Brazilian Portuguese text offensiveness analysis system. web system
	OPCovidBR: Aspect-based sentiment analysis dataset of Covid-19 Tweets. dataset
	AspectPRO: Aspect-based annotated corpus of web consumer reviews dataset.
	HateBR: Large-scale abusive language dataset of Brazilian Instagram comments. dataset
	Deceiver: Multilingual discourse-annotated corpus of fake and truth news. dataset
	FactNews: Sentence-level annotated dataset to predict factually and media bias. dataset

PRO: Taxonomies for aspect-based sentiment analysis in Portuguese. lexicon
MOL: Multilingual offensive lexicon. lexicon

Recommendation Letters	**Prof. Dr. Fabricio Benevenuto.** *Computer Science Department, Federal University of Minas Gerais.* letter **Dra. Norah Abokhodair.** *Microsoft Research.* letter **Prof. Dr. Carlos Ramisch.** *Laboratoire d'Informatique et Systèmes – LIS – UMR CNRS 7020 – Aix Marseille Université.* letter
Languages	Python, SQL, RDF(s), OWL, XML, JSON, Unix/Bash.