

Pedro Henrique de Santana Schmalz

✉ pedrosantanaschmalz@gmail.com.br

in Pedro Schmalz

🔗 Pedro Schmalz

🌐 pedroschmalz.github.io

ORCID

Lattes

Location: Jaguaré, São Paulo, SP, Brazil.

Phone: +55 (11) 97563-4656.

Professional Profile

- ◇ Data scientist and researcher with a background in Social Sciences (USP), an ongoing PhD in Political Science (DCP-USP), and an MBA in Data Science and Analytics (ESALQ/USP). Advanced proficiency in Python and R for machine learning, statistical modeling, and data visualization. Proven ability to deliver end-to-end data projects — from collection and processing to analytical reporting and strategic recommendations — in both public sector and academic settings. Interested in transitioning to industry roles at the intersection of data science, AI, and public policy.

Professional Experience

- 2025–2026 ◇ **Data Consultant**, EDUSP — Editora da Universidade de São Paulo and Grupo Síntese. Designed and administered a survey with 400+ respondents at the 26th USP Book Fair, including full methodological design of the data collection instrument. Processed and analyzed collected data, delivering analytical dashboards and a results report to support editorial decisions. Developed strategic recommendations adopted by the publisher's communication team.
- 2023–2025 ◇ **Data Researcher**, Fundação José Luiz Egydio Setúbal. Collected and processed a dataset of 2,500,000+ tweets from Brazilian legislators and politicians via the Twitter/X API, covering the 2021–2023 period. Fine-tuned BERT-based models for sentiment classification and stance detection on COVID-19 vaccination, achieving strong performance on a manually annotated corpus. Produced technical reports and visualizations to support research on political communication and public health, contributing to two peer-reviewed publications.
- 2020–2023 ◇ **Associate Researcher**, Rede de Pesquisa Solidária — Public Policies and Society. Collected and processed data on educational public policies across all 26 Brazilian states and capital cities during the COVID-19 pandemic. Produced comparative technical studies and sectoral reports on governmental responses to the educational crisis, informing policy recommendations. Co-authored a publication in the *Comparative Education Review* (2026) and a paper presented at NLP4DH 2025.
- 2021–2022 ◇ **Research Assistant**, Instituto Butantan — AVISA Project. Systematically collected and structured public data from Official Gazettes and governmental sources to support a vaccination policy impact evaluation project. Organized and documented datasets, enabling efficient use by a multidisciplinary research team.

Education

- 2024–2026 ◇ **MBA in Data Science and Analytics**, Luiz de Queiroz College of Agriculture, University of São Paulo (ESALQ/USP).

Education (continuação)

- 2022– ◇ **Direct PhD in Political Science**, University of São Paulo, Department of Political Science (DCP-USP).
Advisor: Professor Lorena G. Barberia, Department of Political Science.
Co-advisor: Professor Norton Trevisan Roman, School of Arts, Sciences and Humanities (EACH-USP).
Provisional Thesis Title: Politicians, Twitter and Vaccines: Use of Twitter (X) by Brazilian Politicians in the Discussion of COVID-19 Vaccines.
- 2017–2021 ◇ **Bachelor's Degree in Social Sciences**, University of São Paulo (USP), Brazil.

Publications

Published Articles

- 1 L. G. Barberia, L. G. Cantarelli, P. Schmalz e U. D. Peres, «Federative Coordination and Inequality of Provision: Brazil's Education Policy Response to the COVID-19 Pandemic,» *Comparative Education Review*, vol. 70, n.º 1, fev. de 2026. DOI: 10.1086/739193.

Conference proceedings

- 1 L. G. Barberia, P. Schmalz, N. T. Roman, B. Lombard e T. C. M. Sousa, «It's about What and How you say it: A Corpus with Stance and Sentiment Annotation for COVID-19 Vaccines Posts on X/Twitter by Brazilian Political Elites,» em *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities (NLP4DH 2025)*, 2025.
URL: <https://aclanthology.org/2025.nlp4dh-1.32/>.
- 2 L. Barberia, P. Schmalz e N. T. Roman, «When Tweets Get Viral: A Deep Learning Approach for Stance Analysis of COVID-19 Vaccines Tweets by Brazilian Political Elites,» em *Anais do Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)*, set. de 2023, pp. 104–114. DOI: 10.5753/stil.2023.233961.

Skills

- | | |
|-------------------------|--|
| Programming | ◇ Advanced Python (Pandas, NumPy, Scikit-learn, Matplotlib, Web Scraping, APIs, automation); Advanced R (tidyverse, ggplot2, statistical modeling); Intermediate Stata ; familiarity with SQL and relational databases; version control with Git/GitHub . |
| Data and BI | ◇ Advanced Excel ; Power BI (familiarity); HTML and CSS for indicator presentation and dashboards. |
| AI and Machine Learning | ◇ Hugging Face , BERT (fine-tuning), LDA, LangChain ; experience with LLM-based pipelines and prompt engineering. |
| Languages | ◇ Fluent in Portuguese and English . |