

Pedro Henrique de Santana Schmalz

✉ pedrosantanaschmalz@gmail.com.br

in Pedro Schmalz

🌐 Pedro Schmalz

ORCID

Lattes

Endereço: Avenida Jaguaré 325, Bloco 5, Ap. 25. Jaguaré, SP, 05346-000.

Telefone Celular: (11) 97563-4656.

Idade: 26 anos

Data de Nascimento: 14/05/1999

Perfil Profissional

- ◇ Pesquisador, analista e cientista de dados com formação em Ciências Sociais (USP), Doutorado em Ciência Política em andamento (DCP-USP) e MBA em Data Science e Analytics (ESALQ/USP). Possui experiência consolidada em manipulação de bases de dados públicas, construção de instrumentos de coleta, análise de indicadores sociais e produção de estudos técnicos aplicados a políticas públicas — incluindo consultoria para a EDUSP/USP com entrega de visualizações analíticas e recomendações estratégicas. Domínio avançado de Python e R para modelagem, visualização e automação de dados, com atuação em projetos no setor público (Instituto Butantan, Rede de Pesquisa Solidária). Perfil analítico, detalhista e orientado a resultados, com interesse em inovação e uso de inteligência artificial no setor público.

Experiência Profissional

- 2025-2026 ◇ **Consultor de Dados**, EDUSP — Editora da Universidade de São Paulo e Grupo Síntese. Elaboração e aplicação de survey na Feira do Livro da USP, incluindo desenho metodológico do instrumento de coleta. Tratamento e análise dos dados coletados, com produção de visualizações analíticas e relatório de resultados. Desenvolvimento de recomendações estratégicas para subsidiar decisões editoriais e de comunicação da editora.
- 2023-2025 ◇ **Pesquisador de Dados**, Fundação José Luiz Egydio Setúbal. Coleta e tratamento de dados do Twitter (X) de parlamentares e políticos brasileiros via API, com cobertura do período 2021-2023. Classificação de sentimento e posicionamento em relação à vacinação infantil contra COVID-19 utilizando modelos de Aprendizado de Máquina (BERT). Produção de relatórios técnicos e visualizações analíticas para subsidiar estudos sobre comunicação política e saúde pública.
- 2020-2023 ◇ **Pesquisador Associado**, Rede de Pesquisa Solidária — Políticas Públicas e Sociedade. Coleta e tratamento de dados sobre políticas públicas educacionais dos estados e capitais brasileiros durante a pandemia de COVID-19. Análise de indicadores socioeconômicos e produção de estudos técnicos comparativos e relatórios setoriais sobre a resposta governamental à crise educacional. Coautoria em publicação avaliada pela FGV/EESPClear e em artigo submetido ao *Comparative Education Review*.
- 2021-2022 ◇ **Assistente de Pesquisa**, Instituto Butantan — Projeto AVISA. Coleta sistemática de dados públicos em Diários Oficiais e outras fontes governamentais para subsidiar projeto de avaliação de impacto de políticas de vacinação. Organização, tratamento e documentação das bases de dados coletadas para uso da equipe de pesquisa.

Educação

- 2024 – 2026 ◇ **MBA em Data Science e Analytics** na Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo (ESALQ/USP).
- 2022 – ◇ **Doutorado Direto em Ciência Política** na Universidade de São Paulo, Departamento de Ciência Política (DCP-USP).
Orientadora: Professora Lorena G. Barberia, Departamento de Ciência Política.
Co-orientador: Professor Norton Trevisan Roman, Escola de Artes, Ciências e Humanidades (EACH-USP).
Título Provisório da Tese: Políticos, Twitter e Vacinas: Uso do Twitter (X) por políticos brasileiros na discussão de vacinas de COVID-19.
- 2017 – 2021 ◇ **Ciências Sociais**, Bacharelado na Universidade de São Paulo (USP), Brasil.

Publicações

Artigos Publicados

- 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.

Anais de Congressos

- 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.

Competências

- Programação ◇ **Python** avançado (Pandas, NumPy, Scikit-learn, Matplotlib, Web Scraping, APIs, automação); **R** avançado (tidyverse, ggplot2, modelagem estatística); **Stata** intermediário; familiaridade com **SQL** e bancos de dados relacionais.
- Dados e BI ◇ **Excel** avançado; **Power BI** (familiaridade); HTML e CSS para apresentação de indicadores e dashboards.
- IA e Machine Learning ◇ **Hugging Face**, **BERT**, **LDA**, **LangChain**.
- Idiomas ◇ Fluente em **Português** e **Inglês**.