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

Lorena Barberia, Pedro Schmalz and **Norton Trevisan Roman**
University of São Paulo (USP)
{lorenabarberia, pedrosantanaschmalz, norton}@usp.br

Belinda Lombard
University of Birmingham
bxl388@student.bham.ac.uk

Tatiane C. M. Sousa
University of the State of Rio de Janeiro (UERJ)
tatiane.sousa@uerj.br

Abstract

This paper details the development of a corpus with posts in Brazilian Portuguese published by Brazilian political elites on X (formerly Twitter) regarding COVID-19 vaccines. The corpus consists of 9,045 posts annotated for relevance, stance and sentiment towards COVID-19 vaccines and vaccination during the first three years of the COVID-19 pandemic. Nine annotators, working in three groups, classified these features in messages posted between 2020 and 2022 by local political elites. The annotators underwent extensive training, and weekly meetings were conducted to ensure intra-group annotation consistency. The analysis revealed fair to moderate inter-annotator agreement (Average Krippendorf's alpha of 0.94 for relevance, 0.67 for sentiment and 0.70 for stance). This work makes four significant contributions to the literature. First, it addresses the scarcity of corpora in Brazilian Portuguese, particularly on COVID-19 or vaccines in general. Second, it provides a reliable annotation scheme for sentiment and stance classification, distinguishing both tasks, thereby improving classification precision. Third, it offers a corpus annotated with stance and sentiment according to this scheme, demonstrating how these tasks differ and how conflating them may lead to inconsistencies in corpus construction, as a results of confounding these phenomena — a recurring issue in NLP research beyond studies focusing on vaccines. And fourth, this annotated corpus may serve as the gold standard for fine-tuning and evaluating supervised machine learning models for relevance, sentiment and stance analysis of X posts on similar domains.

1 Introduction

Social media platforms, such as X (formerly Twitter), play a crucial role in analyzing public discourse on policy issues, particularly due to their widespread adoption by both the general public and politicians (Shogan, 2010; Cook, 2016; Pacheco et al., 2023). Due to its prominence in public discourse and widespread adoption, X has become an essential tool for monitoring public opinion (e.g. Somasundaran and Wiebe, 2009; Walker et al., 2012; Bar-Haim et al., 2017; Addawood et al., 2017).

However, the massive volume of user-generated data makes manual analysis impractical, underscoring the need for annotated corpora to improve accuracy in tasks like filtering relevant content, sentiment analysis and stance detection. Annotated datasets not only allow for the training of algorithms to recognize subtle patterns but also enable benchmarking, validation, and adaptation to emerging topics like vaccine hesitancy. For public health applications, such as tracking COVID-19 discourse, rigorously labeled corpora are indispensable.

One approach to analyzing social media discussions is sentiment analysis, a subfield of Natural Language Processing (NLP) that aims to automatically classify the emotional tone of a given text (Pang and Lee, 2008; Liu, 2010). Over the years, it has played a crucial role in examining public discourse across diverse domains, including politics (e.g. Barberá and Rivero, 2014), consumer behavior (e.g. Asur and Huberman, 2010), financial markets (e.g. Bollen et al., 2011), and health-related discussions, such as the COVID-19 pandemic and vaccines (e.g. Boon-Itt and Skunkan, 2020; Naseem et al., 2021; Ainley et al., 2021; Slobodin et al., 2022). In sentiment analysis, texts are typically classified as expressing positive, negative, or neutral sentiment.

However, sentiment analysis is not well-suited to capture the stance individuals take when reacting to policy questions. Instead, stance detection studies in NLP are directed at identifying an author's position regarding a specific proposition or

Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities, pages 365–376
May 3-4, 2025 ©2025 Association for Computational Linguistics

predefined target (Mohammad et al., 2016, 2017; Kuçuk and Can, 2020; AlDayel and Magdy, 2021). Unlike sentiment analysis, which assesses the overall sentiment of a text, stance detection determines the opinion expressed toward a particular entity or topic, with or without sentiment demonstrations. Typically, it categorizes documents as favorable, unfavorable, or neutral in relation to the target. This approach has been applied in various contexts, including product reviews (e.g. Wang et al., 2019), political debates (e.g. Somasundaran and Wiebe, 2009; Anand et al., 2011; Walker et al., 2012; Augenstein et al., 2016; Bar-Haim et al., 2017), and fake news detection (e.g. Lillie and Middelboe, 2019). Distinguishing between these tasks is essential for a refined analysis of opinion and tone in online debates, including discussions on X (Mohammad et al., 2016, 2017)

Despite the importance of sentiment and stance classification, resources for these tasks remain scarce in languages other than English, particularly for Brazilian Portuguese (Pereira, 2020; De Melo and Figueiredo, 2021; Won and Fernandes, 2022; Hervieux et al., 2024; Pavan and Paraboni, 2024). In this context, the development of an annotated corpus in Portuguese, focusing on posts about COVID-19 vaccines, constitutes a significant contribution. Such a corpus facilitates the automatic filtering and classification of stance and sentiment, improving our understanding of social media discussions and offering researchers a valuable resource to further explore this topic. However, our approach departs from conventional frameworks in sentiment analysis and stance detection by introducing a third classification: "*unclear*". This category is distinct from "*neutral*" - widely adopted in existing research - as it accounts for instances where content neither adopts a neutral stance nor conveys any discernible sentiment or position. Specifically, we classify cases as "unclear" when ambiguity or insufficient context makes it impossible to determine a definitive stance or sentiment. Rather than interpreting such cases as an absence of stance, we frame them as uncertain or indeterminate.

To bridge this gap, we introduce a curated and annotated corpus of posts concerning COVID-19 vaccines and vaccination in Portuguese, that measures both sentiment and stance classification. We present a *corpus* of X posts[1] posted by Brazilian Political Elites (*i.e.*, candidates officially endorsed by their parties in local elections) between 2020 and 2022. The *corpus* is annotated for relevance, stance and sentiment. In total, 9,045 posts were annotated for relevance. Out of these, 5,937 posts (65,64%) were classified as relevant and received annotations for stance and sentiment.

The remainder of this article is organized as follows. Section 2 reviews related research on stance detection, sentiment analysis, and studies concerning COVID-19 vaccine discourse. Section 3 details the collection and filtering of the *corpus* of posts, as well as the annotation guidelines and protocol. Section 4 presents and discusses the annotation results and the final corpus. Finally, Section 5 outlines our conclusions and directions for future research.

2 Related Work

During the COVID-19 pandemic, researchers employed NLP methods to investigate shifts in public discourse during critical phases such as lockdown measures, vaccine distribution, and policy changes using both unsupervised and supervised machine learning techniques (e.g. Zou et al., 2020; Ainley et al., 2021; Hu et al., 2021; De Sousa and Becker, 2021; De Melo and Figueiredo, 2021; Liu and Liu, 2021; Alhuzali et al., 2022; Slobodin et al., 2022). By harnessing X data, these studies systematically analyze real-time emotional fluctuations and shifts in public narratives, thereby illuminating collective behavioral patterns and societal dynamics. However, while automated tools like VADER[2] and Textblob are widely employed for sentiment classification in the COVID-19 domain (e.g. Zou et al., 2020; Liu and Liu, 2021; Hu et al., 2021; De Sousa and Becker, 2021; Alhuzali et al., 2022; Slobodin et al., 2022; Thakur, 2023), there remains a notable scarcity of human-annotated corpora specifically tailored to vaccine-related discourse or other COVID-19 topics (positive examples are Ainley et al., 2021; Naseem et al., 2021; Qorib et al., 2023), especially to languages other than English.

To a lesser extent, and somewhat surprisingly, stance detection studies have been pivotal in analyzing public attitudes toward vaccines, lockdown measures, and government responses during the COVID-19 pandemic. Barberia et al.'s (2025) conducted a systematic review of research employ-

[1]The *corpus* is available to interested readers on GitHub, under CC BY-NC-SA 4.0. `https://github.com/NUPRAM/CoViD-Pol`.

[2]Valence Aware Dictionary and Sentiment Reasoner

ing sentiment analysis or stance detection to study discourse towards COVID-19 vaccines and vaccination spread on X. From 1 January 2020 to 31 December 2023, 51 peer-reviewed studies were identified using supervised machine learning to assess COVID-19 vaccine discourse through stance detection or sentiment analysis on Twitter/X. Of this total, only 23.5% were stance detection studies.

Many studies rely on datasets annotated for general sentiment rather than target-specific stances, conflating emotional tone with positional alignment. For instance, while tools like VADER and TextBlob present good performance at sentiment polarity detection, they often fail to disentangle implicit stances toward subtler targets (e.g., vaccine brands like Pfizer vs. AstraZeneca). Furthermore, datasets annotated with stance are rarer on COVID-19-related topics (Hou et al., 2022) and languages other than English (e.g. Won and Fernandes, 2022; Pavan and Paraboni, 2024), and this is even more so in the case of COVID-19 vaccines.

We contribute to both literatures by developing a curated corpus of manual annotations for relevance, sentiment (overall emotional tone) and stance (position toward COVID-19 vaccines). To the best of our knowledge, this is the first manually annotated dataset in Brazilian Portuguese that contains both stance and sentiment in the domain of COVID-19 vaccines and ensures that these analyses are applied to domain-relevant posts only.

3 Material and Methods

To construct this corpus, we collected posts from 2020 to 2022 posted by candidates running for mayor in all Brazilian capitals during the 2020 municipal elections. X profiles were selected based on the candidates' registration and certification by the Brazilian Superior Electoral Court (TSE). Out of the 295 candidates running for mayor in the 26 state capitals, we identified existing accounts for 258 (87.5%). Among these, 89 (30.17% out of 295) profiles were inactive during the analyzed period[3], and 35 (12% out of 295) accounts did not publish content relevant to our research topic.

Our final sample consisted of 143 (48.5%) mayoral candidates. We utilized then X's REST API[4] to collect posts published by these accounts during 2020 and 2021. However, due to changes in the API usage policies in 2022, we employed *twscrape*[5], a dedicated Python package designed for collecting X data. To refine the dataset, we filtered all posts published by these candidates using a keyword-based selection process[6]. As highlighted by Barbera et al.'s (2020), this method is preferable to alternative approaches, such as subjective categorization, as it provides researchers with greater control, ensures reproducibility, and can be adapted for use across different media platforms.

The set of keywords used in this study was developed in several test trials based on observations of spelling variations, term frequency, and usage. Orthographic and spelling issues were addressed after a preliminary analysis of common variations used by X users. As an additional measure, we also accounted for capitalization of terms. Subsequently, we identified posts that, however containing keyword terms, did not refer to COVID-19 vaccines or vaccination. After the filtering by keywords, we randomly sampled 3,015 posts per year for manual annotation.

Table 1 presents the total posts retrieved by year, the remaining posts after keyword filtering, and the random sample that was annotated for each year.

Year	Total	Filtered	Sample
2020	232,014	6,048	3,015
2021	174,638	21,477	3,015
2022	110,490	3,275	3,015
Total	517,142	30,847	9,045

Table 1: Posts Retrieved and Final Sample for Annotation

3.1 Annotation

Following the development of detailed rules for each category of relevance, stance and sentiment, a codebook was used to train the annotation team and vaccine hesitancy literature was shared with the annotation team to improve the understanding of the complexity of vaccination attitudes and emotions. The research team was also trained on the differences in context for the three years under anal-

[3]The activity status of these X accounts was manually verified by a team of coders to ensure they were professional candidate profiles with recent posts.

[4]Documentation available at: https://developer.x.com/en/docs/x-api.

[5]Documentation available at: https://github.com/vladkens/twscrape

[6]List of keywords available at appendix A. Full list of keywords are available at Github together with the Corpus and the research Codebook. https://github.com/NUPRAM/CoViD-Pol.

ysis (e.g., 2020 as a year without vaccines, 2021 as the onset of adult and adolescent vaccination, and 2022 as the onset of child and infant immunization against SARS-CoV-2). The annotation of the corpus was performed in 61 rounds. For each round, a random sample of 200 posts was classified based on using a database specifically created for this project in which anonymized posts were presented removing information about the author, date, or images associated with the message. The classification had two stages. In the first stage, the classification of posts as relevant or not relevant was performed by a group of three annotators. Following annotation, conflicts were reviewed by three senior researchers. Once relevance conflicts were resolved, posts classified as relevant to the study were further annotated by stance and sentiment type. These last two tasks were performed independently by three annotators for each classification. Similar to relevance, conflicts were reviewed by research supervisors.

The annotators followed strict guidelines in each annotation task, and weekly meetings occurred to ensure agreement and consistency throughout 61 rounds. Once the entire sample of posts for a given year was completed, meetings were held to train annotators on specific issues in the incoming sample for the next year. To measure the reliability of the annotations, we used the Krippendorfs' alpha score to calculate inter-annotator agreement (IAA) for each round prior to review discrepancies. As inconsistencies were detected, training sessions were conducted using the detailed rule guidelines published in the codebook and resolved through discussion among the annotators, with the input of a supervisor. This methodology was devised so as to improve the quality of the resulting dataset.

3.1.1 Relevance Annotation Procedure

At first, the posts were annotated as *relevant* or *irrelevant* by a group of three coders. Relevant publications were those that talked directly about COVID-19 vaccines and vaccination. We decided to also include posts that discussed treatments that were available in the period before and after the introduction of vaccination that were believed to reduce the severity of infection and posts that discussed vaccines more generally to capture generalized disposition towards vaccination that implicitly influences COVID-19 vaccine discourse. Irrelevant posts where those that either did not discuss COVID-19 vaccines (e.g. updates on number of cases and deaths, social isolation measures such as lockdowns, etc.) or used vaccination as metaphor to discuss another topic. Of the 9,045 posts annotated, 5,937 (65.64%) were classified as related to COVID-19 vaccines. As an example,[7] the following tweet was considered irrelevant:

- **Irrelevant**: "The @minsaude[8] updates the situation of #coronavirus in Brazil - 04/18: 36,599 cases and 2,347 deaths. Find more information on the platform: #COVID19."

While this tweet mentions COVID-19 directly, it is not about COVID-19 vaccines or vaccination, nor other vaccines or alternative treatments. Therefore, it is classified as irrelevant. The following example is a relevant tweet:

- **Relevant**: "@jairbolsonaro[9] was once a 'lion' against Anvisa[10] when it came to MEDICATIONS WITHOUT EFFICACY. Today, he DISDAINS Pfizer, which is the vaccine adopted in dozens of countries."

This tweet was identified as relevant as it contains keywords, such as Pfizer and also refers to "MEDICATIONS WITHOUT EFFICACY."

3.1.2 Stance Annotation Procedure

Only relevant posts were classified for stance. The unit of analysis was the individual tweet, and each post was categorized into one of three stance categories: *favorable*, *unclear*, or *unfavorable* toward COVID-19 vaccines. The "unclear" class was adopted to capture vaccine hesitancy, which is considered an important policy position towards immunization and preferable to classifying those not adopting a specific position as "neutral".

Posts classified as *unfavorable* included those expressing skepticism about vaccine efficacy or using derogatory terms, such as referring to Coronavac as *Vachina* (a portmanteau of "vaccine" and "China"), or contested vaccine mandates. *Favorable* posts were those that praised vaccines or celebrated their authorization by regulatory authorities and administration. *Unclear* posts were those that lacked a discernible stance on COVID-19 vaccines, or talked about alternative treatments (e.g. hydroxychloroquine, azithromycin) or other vaccines (e.g.

[7]Translated from Brazilian Portuguese by the authors.
[8]Profile of the Brazilian Ministry of Health
[9]The profile of Brazil's former president, Jair Bolsonaro
[10]Brazilian National Health Surveillance Agency.

the flu vaccine, H1N1, etc.) without specifically mentioning COVID-19 vaccines. As an example, the following tweet was classified as *favorable*:

- **Favorable**: "What an important day! Isa got vaccinated, I'm so happy! Let's protect our children! The vaccine is the guarantee of safety for the return to school. Vaccines save lives."

In this tweet the author does not explicitly mention COVID-19. However, the tweet is considered favorable as the vaccination of children is argued to be a necessary pre-condition to safe return to on-site schooling. Furthermore, the tweet celebrates the vaccination of a child, rendering it a favorable tweet. Here we have an unfavorable tweet:

- **Unfavorable**: "MAYOR MANDATES THE VACHINA! ANOTHER ACTION BY THE DICTATOR OF FLORIANÓPOLIS! Anyone needing help to refuse can join my Telegram channel, where I've posted a document with technical and legal grounds. Link in the channel."

Not only does the author refer to the Coronavac Vaccine as "Vachina", they also oppose mandatory vaccination, calling the mayor a dictator for such policy, and providing legal argument for vaccine refusal. For all that, this tweet is classified as unfavorable. Finally, the following tweet is classified as unclear.

- **Unclear**: "Bolsonaro insists on joking about serious matters. As if it weren't enough to recommend the use of Chloroquine for COVID-19 treatment, he now becomes proof of its ineffectiveness!"

In this tweet, the author criticizes the behavior of then Brazilian President, Jair Bolsonaro. They oppose his stance favoring alternative treatments, in this case the usage of hydroxychloroquine for COVID-19 treatment. As there is no direct mention of COVID-19 vaccines, it's stance is unclear.[11]

3.1.3 Sentiment Annotation Procedure

Along with stance, we also annotated the overall sentiment of the publications. However, differently from stance, sentiment was coded in relation to the overall emotions manifested in the posts, not in relation to COVID-19 vaccines. Messages were classified by their overall sentiment, and divided in three classes: posts that elicited positive emotions such as hope, admiration, gratitude, or a positive emotional state were classified as *positive*. Messages eliciting emotions such as pessimism, fear, or overall negative emotional state were classified as *negative*. Finally, posts where it was not possible to infer neither emotional states were classified as *unclear*. The following tweet exemplifies a positive sentiment tweet:

- **Positive**: "Certainly, in 2022, we will (all properly vaccinated) be able to celebrate life and our culture with all the intensity we deserve."

The sentiment of this tweet is positive because it expresses hope and optimism about the future. The sentence suggests that, after vaccination, people will be able to engage in celebrations, something that was limited during the COVID-19 pandemic. It bears noting that this tweet is also favorable regarding stance as it welcomes the arrival of vaccines. The next tweet was classified as negative.

- **Negative**: "We need to create a great flame of mobilization from our pain, anguish, and melancholy that these times have caused us. Vaccines! Food on the table! Bolsonaro out!"

This tweet is negative because it expresses frustration and dissatisfaction with the current situation. The call for "Vaccines! Food on the table! Bolsonaro out!" highlights unmet needs and a desire for change, reflecting discontent and urgency. On an alternative note, the tweet was classified as favorable towards the vaccine. This example demonstrates the importance of differentiating between stance and sentiment, and how mixing both concepts could impact inference and the accuracy of a corpus. Lastly, an example of an unclear tweet with respect to sentiment is:

- **Unclear**: "Mexico, Chile, and Argentina will be the first to vaccinate in Latin America."

This tweet is unclear due to the fact that the author does not express emotions clearly. Is not possible to infer if their emotions are of frustration due to other countries getting vaccinated first, or if they are just reporting some news. So, the tweet is classified as unclear.

[11] For a more complete discussion on the Unclear category, we refer the interested reader to our codebook, available at https://github.com/NUPRAM/CoViD-Pol/blob/main/Codebook%20v1.0.pdf.

4 Results and Discussion

Along the 61 annotation rounds, inter-annotator agreement, in terms of Krippendorff's alpha, was calculated for Relevance, Sentiment and Stance, as shown in Figure 1. Following each round, group supervisors conducted meetings with the annotation teams to address any questions or issues raised by the annotators.

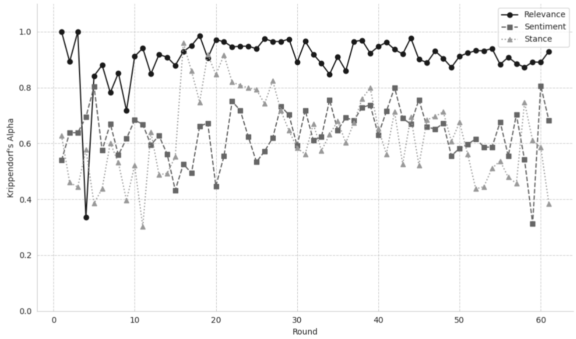

Figure 1: Krippendorf's Alpha over Rounds

As it turns out, Relevance achieved the highest overall agreement between annotators, with an average alpha score of 0.94 and minimal variation across rounds. This task benefited from a more balanced dataset, fewer conflicts among annotators, and a larger total number of observations. Both Sentiment and Stance attained moderate agreement. The average alpha was 0.67 for Sentiment and 0.70 for Stance. However, both tasks showed significant variability along the rounds. This variability can be attributed to the imbalance between classes at the stance and sentiment classification. As a result, there was greater disagreement in the content being analyzed in each round, especially for the minority categories (unfavorable and unclear).

Table 2 shows the class distribution in the relevant portion of the corpus (N=5,397) for stance and sentiment. The majority class in sentiment is the positive class, with 46.8% of the 5.937 posts belonging to it. Nearly an equivalent share of posts were negative (46.6%), and only 6.6% were identified as unclear sentiment. Thus, sentiments are mostly expressed in discourse and rarely are messages identified as expressing ambivalence.

For stance, 78.6% of the three-year sample were classified as favorable towards COVID-19 vaccines. However, there are 17.4% posts were an opinion towards vaccines to protect against SARS-CoV-2 are not self-evident based on the message. An example of discourse that is unclear are publications that express an opinion supporting alternative treatments (e.g. Hydroxy-chloroquine) without expressing an explicit position on COVID-19 vaccines. In 2020, with the uncertainty of vaccines, many politicians expressed opinions favoring the use of medications. Strikingly in contrast to related-work, where unfavorable postures seem to be quite common (*e.g.* Cheatham et al., 2022; Hwang et al., 2022; Zaidi et al., 2023), only 4% of the annotated sample could be found as having an unfavorable stance towards COVID-19.

Task	Class	Total	Percentage
	Positive	2,776	46.8%
Sentiment	Unclear	389	6.6%
	Negative	2761	46.6%
	Favorable	4,645	78.6%
Stance	Unclear	1,030	17.4%
	Unfavorable	234	4.0%

Table 2: Distribution of Classes (2020-2022)

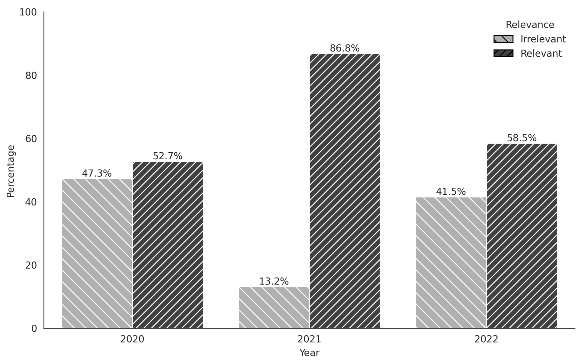

Figure 2: Percentage of relevant Posts in the samples from 2020, 2021 and 2022

There was a sharp increase in messages expressing opinions and emotions regarding COVID-19 vaccines in the year in which vaccines became available (2021), with 86.8% of posts being classified as relevant. Figure 2 shows the percentage of messages classified as relevant in 2020, 2021 and 2022. As the figure illustrates, the proportion of relevant posts in 2020 and 2022 are more similar (52.7% and 58.4%, respectively), which helps explain why the overall sample is fairly balanced between classes, with a slightly larger share of relevant posts. Overall, 65.64% of the posts were classified as relevant.

Figure 3 shows the distribution of stance categories along the years. The favorable class is majority for all years with its percentage increasing from 59.7% in 2020 to 86.3% in 2021 and 84.2% in 2022. The unclear class comes second with 33.8% in the first year, and decreasing to 12.1% in 2021 and 10.7% in 2022 .As can be noticed in the figure, the proportion on messages not expressing a favorable or unfavorable position drops with COVID-19 vaccines becoming available to Brazilians (1° Semester of 2021).

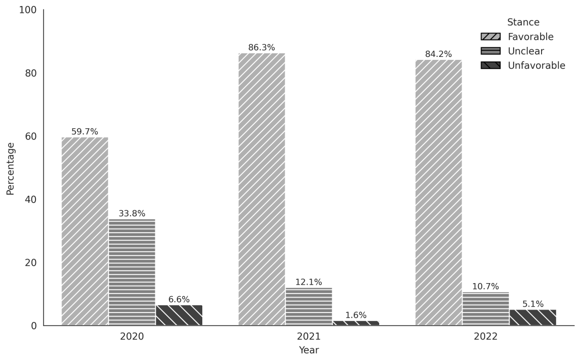

Figure 3: Distribution of Stance in 2020, 2021, 2022

The unfavorable class is minority in all years. In contrast to the uptick in favorable posts in 2021, there is a sharp decrease in unfavorable tweets during this year. These messages comprised 6.6% of the dataset in 2020, 1.6% in 2021, with an uptick to 5.1% in 2022.

Lastly, Figure 4 shows the proportion of sentiment categories along the three years. In 2020, 59.6% of the posts were interpreted as negative, 35.1% expressed positive feelings and 5.2% were neither positive or negative, which we refer to as unclear. As it turns out, there is an inversion in the year vaccines are introduced (2021) when compared to its preceding year, whereby the positive class becomes the majority class (48.8%) and there is a decrease in negative posts (now 44.2%), and unclear remains the minority with 7.1% of the publications. With the arrival of vaccines and the loosening of social distancing restrictions, there remained a relatively limited supply of vaccines, a slow rollout and the largest number of COVID-19 deaths during the Omicron and Delta waves. In this context, the proportion of negative posts keeps similar along the years as politicians continue to express frustration, and sadness. In 2022, positive posts were 54.4% of the sample, negative were 38.6% and unclear were 7%.

A final point to note is the stark differences between stance and sentiment distributions. Stance

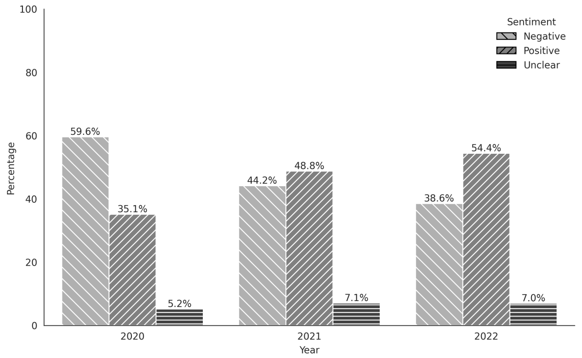

Figure 4: Distribution of Sentiment in 2020, 2021 and 2022

presents a very unbalanced distribution with most posts being favorable towards COVID-19 vaccination and vaccines. Unfavorable posts are a rare occurrence, but there are a considerable percentage of messages that are not clear on the position towards COVID-19 vaccines. On the other hand, for sentiment, the dataset is more balanced between negative and positive posts. It is rare for local political elites to fail to use emotions in their discourse towards COVID-19 vaccines.

4.1 What and How they say it

Our corpus clearly shows that stance and sentiment should be defined as distinct categories in the annotation process. Furthermore, protocols should be adopted to ensure that the corpus does not confound sentiment analysis and stance detection. It is not the case that discourse favorable to vaccines is always positive, nor is it the case that messages against vaccines are always manifest with negative emotions. Failing to separately measure both instruments can generate problems both for the corpus and to some automatic tagger trained on it. As an example, consider the following tweet:

- Let's keep the mobilization up to see if this irresponsible and incompetent government starts moving and works! #VaccineNow.

Per the annotation guidelines, the tweet is favorable, but its overall sentiment is negative because it expresses the author's frustration with the government. Besides the obfuscation of the debate by improperly grouping stance and sentiments, there are also significant discrepancies once unclear is defined as opposed to neutrality. For example:

- Everything you need to know about the COVID-19 vaccine: [link]

In most studies reviewed by Barberia et al.'s (2025), the tweet would be considered "neutral". In our opinion, this source of mis-classification is a result of the confusion between stance and sentiment. In our classification, the message has a favorable opinion towards COVID-19 vaccines. However, the sentiment is unclear since no sentiment can be inferred from this tweet due to insufficient information.[12]

Taking neutral as a proper category in stance detection can be especially problematic. In the literature, we identified studies which consider neutrality as a fairly common occurrence. There are some messages which could be interpreted as impartial. For example:

- "Covid-19: Vaccines arrive in Brazil this Saturday for the start of testing"

Some studies might classify this post as neutral (Barberia et al., 2025). However, context matters. A local politician is disseminating information about the arrival of COVID-19 vaccines in a polarized moment in Brazil. The discourse is not a neutral opinion. In this study, this tweet is classified as favorable towards COVID-19 vaccines.

Table 3 shows the cross-tabulation between stance and sentiment in our corpus (from 2020 to 2022). Figures were not found to be independent, as determined by a χ^2 test of independence[13], indicating there might be an overall association between Stance and Sentiment, aligned with the common sense intuition about these categories. That, however, does by no means imply both categories are the same, with problems arising where this association disappears, as a result of context and the circumstances involved.

	Stance		
Sentiment	Favorable	Unclear	Unfavorable
Positive	2,530	211	22
Unclear	242	119	21
Negative	1,870	692	191

Table 3: Cross-Tabulations between Stance and Sentiment Classes (2020-2022)

The table also shows that conflating stance and sentiment classes (e.g., favorable and positive, unfavorable and negative, etc.) can lead to measurement error. Of all favorable posts, 54.5% (2,530 of 4,642 posts) were also classified as positive. This aligns with the usual interpretation of the literature that sentiment indicates the stance towards a predefined target. However, 40.3% of favorable posts are negative in sentiment (1,870 messages). On polarized issues, such as COVID-19 vaccines, it is quite frequent to observe negative emotions by those favorable to vaccination.

Our research protocol separates stance to solely capture the author's position on the topic, while sentiment captures the emotional tone of the message. If the rules had been confused by assuming that positive sentiment in posts was directly correlated with a favorable stance toward COVID-19 vaccinations, our study would have a significant share of measurement error. Sentiments are polar opposite of position-taking in some cases, such as when users express approval of vaccine availability while criticising the government.

Similarly, there are many messages where an author employs a positive tone to express an unfavorable position towards vaccines. In our corpora, 81.6% of posts with an unfavorable stance (191 of 234) are associated with a negative sentiment. Whereas 18.4% of unfavorable posts do not display negative emotions. In other words, in nearly 1 out of 5 cases, a user expresses opposition to COVID-19 vaccines using positive or unclear emotional sentiment. Moreover, only 191 messages out of 2,753 negative messages (6.9%) are unfavorable towards vaccination. Thus, there is clearly a need to separate the classification of stance from sentiment, as the emotional tone may not always align with the author's position regarding and entity or topic.

The unclear category for both stance and sentiment reveals the complexity of interpreting social media content, author's positions and orientation in a given subject. An unclear opinion may or may not have unclear sentiments. For instance, 211 posts with unclear stance have positive sentiment, and 692 with unclear stance have negative sentiment. These cases highlight the challenges in the creation of a corpus and the importance of clear annotation guidelines to differentiate between stance and sentiment.

5 Conclusions and Future Work

This study detailed the development of an annotated corpus of posts in Brazilian Portuguese, posted by Brazilian political elites focusing on

[12]For a more complete discussion: https://github.com/NUPRAM/CoViD-Pol/blob/main/Codebook%20v1.0.pdf.

[13]$\chi^2(n = 5898, df = 4) = 532.43, p \ll 0.001$

COVID-19 vaccines and vaccination. The corpus was first classified according to each post's relevance to the topic. Relevant posts were then further annotated with respect to stance (favorable, unfavorable, and unclear) and sentiment (positive, negative, and unclear). The creation of this corpus addresses significant gaps in the literature, due to the scarcity of resources in Brazilian Portuguese and the lack of curated datasets in this language related to vaccines, particularly in the context of COVID-19. Furthermore, the study presents a reliable annotation scheme that distinguishes between sentiment analysis and stance detection. The analysis of the annotated corpus provides evidence that measurement error can occur due to two problems. (i) If a relevance rule is not applied, scholars may be annotating data that are not specific to the topic even if keywords are present; and (ii) when sentiment and stance tasks are not separately considered, class conflation may introduces bias.

The annotation process involved nine annotators, divided into three groups, who analyzed 9,045 posts published between 2020 and 2022. The process included extensive training, weekly meetings and supervision to resolve conflicts to ensure consistency. The analysis of annotation the results revealed fair to moderate agreement among annotators, as indicated by a 0.94 overall Krippendorf's alpha of for relevance, 0.67 for sentiment and 0.70 for stance. This annotated corpus can serve as a gold standard for training and, amongst other things, evaluating machine learning models. In future research, we are planning to explore whether the patterns reported in this study also apply to discourse on COVID-19 vaccines when specifically focusing on children and adolescents, or vulnerable populations, such as the elderly and those who have other chronic illnesses. We also plan to use this corpus to further expand the classification of COVID-19 vaccine discourse to national political elites and other X messages on the same domain. The annotated corpus is publicly available on GitHub under a Creative Commons license (CC BY-NC-SA 4.0)[14], ensuring that future research can build upon this work while respecting the ethical standards for data sharing.

Limitations

This study has several limitations that should be considered. First, the corpus is limited to posts written in Brazilian Portuguese, which may affect the applicability of the findings to other languages or dialects, especially considering the unique vocabulary and expressions of this language. The analysis is also confined to a specific time period (2020 to 2022), limiting insights into the evolving discourse on COVID-19 vaccines beyond this window. Furthermore, the reliance on manual annotation introduces potential biases and inconsistencies, despite the efforts to ensure reliability. These limitations highlight opportunities for future research, including more scalable methods for multi-language and multi-domain sentiment and stance analysis.

Ethics Statement

This study was conducted in accordance with ethical guidelines for research involving publicly available data. All posts included in the corpus were publicly posted on X (formerly Twitter) and were not retrieved from private accounts or behind any paywalls by individuals who had registered their candidacy to mayoral elections in the 26 state capitals of Brazil. The authors of posts included in this research were anonymized to ensure compliance with ethical standards and data protection regulations, particularly the Brazilian General Data Protection Law (*LGPD - Lei Geral de Proteção de Dados*). However, mentions within the posts of authorities, individuals, and public figures were retained to enable the potential classification of positioning and sentiment, as such information may provide crucial context. The authors are committed to maintaining transparency and respect for privacy in the presentation and analysis of the data.

Acknowledgments

This research project was made possible by grant support from the José Luiz Egydio Setúbal Foundation (JLES). In addition, the authors acknowledge support from the Center for Artificial Intelligence of the University of São Paulo (C4AI - http://c4ai.inova.usp.br/) and its supporting grants from the São Paulo Research Foundation (FAPESP grant #2019/07665-4) and the IBM Corporation. The project was also supported by the Ministry of Science, Technology and Innovation, with resources of Law N. 8.248, of October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex and published as Residence in TIC 13, DOU 01245.010222/2022-44.

We are deeply thankful to all the students and

[14]Available at https://github.com/NUPRAM/CoViD-Pol

researchers at the NUPRAM[15] research center, situated at the University of São Paulo, who contributed to this project, whose dedication and hard work were essential to its success. Special thanks go to Angelo Masiero, Dara Aparecida Vilela Pinto, Eliane de Santana Firmino, Evelyn Rosa, Fernanda de Almeida Silva, Gustavo Fernandes de Paula, Leticia Machado, Lian Pelser, Luiz Henrique da Silva Batista, Nina Lombard, Sabrina de Almeida Santos, Thales David Domingues Aparecido, and the additional researchers who contributed to the first version of the corpus: Andre Garibe, Vinícius Bello Pereira, Isabel Seelaender Costa Rosa, and Rebeca de Jesus Carvalho, whose efforts and commitment were invaluable. This work would not have been possible without their collaboration and enthusiasm.

References

Abdulrahman Addawood, Jens Schneider, and Mohamad Bashir. 2017. Stance classification of twitter debates. In *Proceedings of the 8th International Conference on Social Media & Society - #SMSociety17*. ACM Press.

Esther Ainley, Cara Witwicki, Amy Tallett, and Chris Graham. 2021. Using twitter comments to understand people's experiences of uk health care during the covid-19 pandemic: Thematic and sentiment analysis. *Journal of Medical Internet Research*, 23(10):e31101.

Abeer AlDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management*.

Hassan Alhuzali, Tianlin Zhang, and Sophia Ananiadou. 2022. Emotions and topics expressed on twitter during the covid-19 pandemic in the united kingdom: Comparative geolocation and text mining analysis. *Journal of Medical Internet Research*.

Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 1–9, Portland, Oregon. Association for Computational Linguistics.

Sitaram Asur and Bernardo Huberman. 2010. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 492–499, Toronto, Canada. IEEE.

[15] *Núcleo de Políticos, Redes Sociais e Aprendizado de Máquina*, or Center for Politics, Social Networks, and Machine Learning.

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.

Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. 2017. Stance classification of context-dependent claims. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 251–261, Valencia, Spain. Association for Computational Linguistics.

Pablo Barbera, Amber Boydstun, Steven Linn, Robert McMahon, and Jonathan Nagler. 2020. Automated text classification of news articles: A practical guide. *Political Analysis*, 29(1):19–42.

Lorena G. Barberia, Belinda Sousa Lombard, Norton Trevisan Roman, and Tatiane C. M. Sousa. 2025. Clarifying misconceptions in COVID-19 vaccine sentiment and stance analysis and their implications for vaccine hesitancy mitigation: A systematic review. ArXiv Pre-Print.

Pablo Barberá and Gonzalo Rivero. 2014. Understanding the political representativeness of twitter users. *Social Science Computer Review*, 33(6):712–729.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2:1–8.

Sakon Boon-Itt and Yacine Skunkan. 2020. Public perception of the covid-19 pandemic on twitter: Sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance*, 6(4):e21978.

Susan Cheatham, Per E. Kummervold, Lorenza Parisi, Barbara Lanfranchi, Ileana Croci, Francesca Comunello, Maria Cristina Rota, Antonietta Filia, Alberto Eugenio Tozzi, Caterina Rizzo, and Francesco Gesualdo. 2022. Understanding the vaccine stance of Italian tweets and addressing language changes through the COVID-19 pandemic: Development and validation of a machine learning model. *Frontiers in Public Health*, 10:948880.

James M. Cook. 2016. Twitter adoption in u.s. legislatures. In *Proceedings of the 7th 2016 International Conference on Social Media & Society - SMSociety '16*. ACM Press.

Tiago De Melo and Carlos M. S. Figueiredo. 2021. Comparing news articles and tweets about covid-19 in brazil: Sentiment analysis and topic modeling approach. *JMIR Public Health and Surveillance*.

Andre Mediote De Sousa and Karin Becker. 2021. Pro/anti-vaxxers in brazil: A temporal analysis of covid vaccination stance in twitter. In *Anais do IX Symposium on Knowledge Discovery, Mining and*

Learning (KDMiLe 2021). Sociedade Brasileira de Computação - SBC.

Natalie Hervieux, Peiran Yao, Susan Brown, and Denilson Barbosa. 2024. Language resources from prominent born-digital humanities texts are still needed in the age of LLMs. In *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, pages 85–104, Miami, USA. Association for Computational Linguistics.

Yanfang Hou, Peter van der Putten, and Suzan Verberne. 2022. The covmis-stance dataset: Stance detection on twitter for covid-19 misinformation.

Tao Hu, Siqin Wang, Wei Luo, Mengxi Zhang, Xiao Huang, Yingwei Yan, Regina Liu, Kelly Ly, Viraj Kacker, Bing She, and Zhenlong Li. 2021. Revealing public opinion towards covid-19 vaccines with twitter data in the united states: Spatiotemporal perspective. *Journal of Medical Internet Research*.

Juwon Hwang, Min-Hsin Su, Xiaoya Jiang, Ruixue Lian, Arina Tveleneva, and Dhavan Shah. 2022. Vaccine discourse during the onset of the COVID-19 pandemic: Topical structure and source patterns informing efforts to combat vaccine hesitancy. *PLOS ONE*.

Dilek Kuçuk and Fazli Can. 2020. Stance detection. *ACM Computing Surveys*.

Alexandra E. Lillie and Esben R. Middelboe. 2019. Fake news detection using stance classification: A survey (version 1). *arXiv*.

Bing Liu. 2010. Sentiment analysis and subjectivity. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing*. Chapman and Hall/CRC; 2nd edition (February 22, 2010).

Siru Liu and Jialin Liu. 2021. Public attitudes toward covid-19 vaccines on english-language twitter: A sentiment analysis. *Vaccine*.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. A dataset for detecting stance in tweets. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3945–3952, Portorož, Slovenia. European Language Resources Association (ELRA).

Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Transactions on Internet Technology*, 17(3):1–23.

Usman Naseem, Imran Razzak, Matloob Khushi, Peter W. Eklund, and Jinman Kim. 2021. Covidsenti: A large-scale benchmark twitter data set for covid-19 sentiment analysis. *IEEE Transactions on Computational Social Systems*, 8(4):976–988.

Luis Jorge Orcasitas Pacheco, Elen Cristina Geraldes, and Georgete Medleg Rodrigues. 2023. Parlamentarios en twitter: Una revisión de la literatura. *Documentación de las Ciencias de la Información*.

Bo Pang and Lillian Lee. 2008. *Opinion Mining and Sentiment Analysis*, volume 2. Now Publishers Inc.

Matheus Camasmie Pavan and Ivandré Paraboni. 2024. A benchmark for portuguese zero-shot stance detection. *Journal of the Brazilian Computer Society*.

Denilson Alves Pereira. 2020. A survey of sentiment analysis in the portuguese language. *Artificial Intelligence Review*.

M Qorib, T Oladunni, M Denis, E Ososanya, and P Cotae. 2023. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on covid-19 vaccination twitter dataset. *Expert Systems with Applications*, 212:118715. Epub 2022 Sep 5. PMID: 36092862; PMCID: PMC9443617.

Colleen J. Shogan. 2010. Blackberries, tweets, and youtube: Technology and the future of communicating with congress. *PS: Political Science & Politics*.

Ortal Slobodin, Ilia Plochotnikov, Idan-Chaim Cohen, Aviad Elyashar, Odeya Cohen, and Rami Puzis. 2022. Global and local trends affecting the experience of us and uk healthcare professionals during covid-19: Twitter text analysis. *International Journal of Environmental Research and Public Health*, 19(11):6895.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 226–234, Suntec, Singapore. Association for Computational Linguistics.

Nirmalya Thakur. 2023. Sentiment analysis and text analysis of the public discourse on twitter about covid-19 and mpox. *Big Data and Cognitive Computing*, 7(2):116.

Marilyn A. Walker, Pranav Anand, Robert Abbott, Julie E. F. Tree, Craig Martell, and Jon King. 2012. That is your evidence?: Classifying stance in online political debate. *Decision Support Systems*, 53(4):719–729.

Rui Wang, Deyu Zhou, Mingmin Jiang, Jiasheng Si, and Yang Yang. 2019. A survey on opinion mining: From stance to product aspect. *IEEE Access*, 7:41101–41124.

M. Won and J. Fernandes. 2022. Ss-pt: A stance and sentiment data set from portuguese quoted tweets. In V. Pinheiro et al., editors, *Computational Processing of the Portuguese Language. PROPOR 2022*, volume 13208 of *Lecture Notes in Computer Science*. Springer, Cham.

Z Zaidi, M Ye, F Samon, A Jama, B Gopalakrishnan, C Gu, S Karunasekera, J Evans, and Y Kashima. 2023. Topics in antivax and provax discourse: Year-long synoptic study of COVID-19 vaccine tweets. *Journal of Medical Internet Research*, 25:e45069.

Canruo Zou, Xueting Wang, Zidian Xie, and Dongmei Li. 2020. Public reactions towards the covid-19 pandemic on twitter in the united kingdom and the united states. *preprint*.

A List of Keywords

All regex patterns account for case variations and common misspellings (137 total terms). Full technical specifications available in supplementary materials.

- **Vaccines and Vaccination** (Terms related to vaccines and mandatory vaccination): `[Vv]accin*`, `[Vv]assina*`, `[Vv]accination`, `[Vv]asina`, `[Ii]mmunization`, `[Ii]mmunisa*`, `[Dd]ose`, `[Dd]oze`, `[Rr]einforcement`, `[Ii]mmunobiological`, `[Oo]bligation`, `[Oo]bligar`.

- **COVID-19 Vaccines and Laboratories** (Manufacturers and brands): CoronaVac: `[Cc]orona[Vv]ac`, `[Cc]ova[Xx]in`, `[Cc]omuna[Vv]ac`, `[Ss]inovac`; AstraZeneca: `[Aa]stra[Zz]eneca*`, `[Oo]xford*`, `[Vv]axzvria`; Pfizer: `[Pp]fizer*`, `[Bb]iontech*`, `[Cc]omirnaty`; Moderna: `[Mm]oderna*`, `mRNA-1273`, `CX-024414`; Sputnik: `[Ss]putnik*`, `[Gg]amaleya*`; Janssen: `[Jj]ans?en*`, `Ad26.COV2.S`; Covaxin: `[Cc]ovaxin`, `[Bb]harat [Bb]iotech`; Novavax: `[Nn]ovavax*`, `NVX-CoV2373`; Sinopharm: `[Ss]inopharm*`, `BIBP`; Others: `[Bb]utantan*`, `[Ff]iocruz*`.

- **Geography** (Country associations): `[Vv]achina*`, `[Vv]accine [Cc]hina`, `[Vv]accina [Bb]ritannica`, `[Vv]accine [Rr]ussia`, `[Vv]achin@da`.

- **Adverse Effects** (Reported side effects): `[Aa]naphylaxis`, `[Mm]yocarditis`, `[Tt]hrombosis`, `[Aa]utism`, `[Pp]aralysis`, `[Ss]troke`, `[Dd]eath`, `[Mm]enstrual disorders`, `[Hh]eart pain`, `[Gg]uillain-Barré syndrome`, `[Cc]ancer`, `[Mm]iscarriage`, `[Aa]utoimmune disease`.

- **COVID-19 Treatments** (Discussed therapies): `[Ee]arly [Tt]reatment`, `[Cc]hloroquine*`, `[Ii]vermectin*`, `[Oo]zone therapy`, `[Vv]itamin D`, `[Cc]onvalescent [Pp]lasma`, `[Pp]axlovid`, `[Dd]examethasone`, `[Kk]it [Cc]ovid`.

- **Political Terms** (Bolsonaro-related vocabulary): `[Dd]oriavac*` (anti-Doria vaccine rhetoric), `[Gg]uinea pig`, `[Aa]lligator` (pejorative term), `[Vv]accine-China` (sinophobic rhetoric).