Journal of Quantitative Description: Digital Media 4(2024), 1–85 DOI: 10.51685/jqd.2024.007

Detecting Misinformation: Identifying False News Spread by Political Leaders in the Global South

VALERIE WIRTSCHAFTER

Brookings Institution, USA

FREDERICO BATISTA PEREIRA

University of North Carolina at Charlotte, USA

NATÁLIA S. BUENO

Emory University, USA

NARA PAVÃO

Federal University of Pernambuco, Brazil

JOÃO PEDRO OLIVEIRA DOS SANTOS

Independent Researcher, Brazil

FELIPE NUNES

Federal University of Minas Gerais, Brazil

We provide and examine an approach for detecting false stories that circulate as text and without hyperlinks, which are commonly found in the Global South. Our text-based approach relies on a combination of false stories identified by fact-checkers, supervised learning methods, natural language processing, and human review. We contrast our approach with the established domain-based and with Facebook's URL approaches by applying them in the case of Brazilian

Wirtschafter: valerie.wirtschafter@gmail.com
Batista Pereira: fbatist1@uncc.edu
Date submitted: 2024-01-18

political leaders. The results show that sharing false news by politicians is a rare event: less than 1% of political leaders' social media posts contain misinformation. However, we find little overlap across the approaches. The text-based approach leads to different conclusions about which politicians share misinformation and the type of false content shared, while demographic and political predictors of misinformation-sharing behavior are typically similar across approaches. Our approach produces fewer false positives than other approaches and only a small number of false negatives. Our results show that the text-based approach is an important complement to the dominant approaches as it is more effective at detecting false news.

Keywords: Misinformation, political leaders, political behavior, Global South

Misinformation is a growing concern in policy circles and public discourse. Political misinformation, commonly understood to include both false news (Ross et al., 2021) and politically biased or misleading content that is not undeniably false (Ross et al., 2021; Pennycook and Rand, 2019), has been associated with undesirable outcomes such as disruptive elections, political unrest, violence, and ethnic strife, among others (Tandoc Jr. et al., 2018; Jolley and Douglas, 2014; Einstein and Glick, 2013; Nyhan, 2018; Imhoff et al., 2021; Vegetti and Littvay, 2020; Bennett and Livingston, 2018). A critical step in describing, explaining, and combating misinformation is researchers' ability to detect it. While several approaches have been developed for use in the United States and other developed countries, researchers ability to detect misinformation is still limited (Guess and Lyons, 2020).

In this paper, we offer a text-based approach to detect misinformation using unstructured forms of text. This approach differs from several approaches that have relied on identifying low-quality/hyperpartisan websites (Guess et al., 2020b; Allcott et al., 2019), fact-checking (Allcott and Gentzkow, 2017; Resende et al., 2019a; Mosleh et al., 2021; Mosleh and Rand, 2021), systematic searches based on keywords (Wirtschafter and Meserole, 2022), machine learning methods (Padmanabhan, 2021), perceptual hashing and manual coding of images (Yang et al., 2023), and investigative journalistic work (Bovet and Makse, 2019)

to detect misinformation. Critically, the text-based approach we present here does not rely on the existence of a hyperlink to a false story or a story that contains misinformation.

We validate this text-based approach by measuring the prevalence of misinformation in political leaders' social media posts in Brazil, an example of a developing country in which 75% of all the content shared by elected politicians on social media in Brazil from 2018 to 2020 did not contain references to external domains.[1] We contrast this text-based approach with other approaches to compare their performance at detecting posts containing misinformation and identify what characteristics predict misinformation-sharing behavior.

The contributions are twofold. First, by comparing the text-based approach to existing approaches, the results indicate that what is identified as misinformation is, to a large extent, approach-dependent. There is little overlap in the posts identified as containing misinformation among the existing approaches, and between those and the text-based approach. Furthermore, the content identified in each approach is distinct: the text-based approach captures posts containing falsehoods flagged by fact-checkers, while the dominant approaches tend to detect biased or hyper-partisan content, rather than blatantly false content. In other words, the dominant approaches miss an important share of the false content in political leaders' social media posts, and tend to tag content that is hyperpartisan/low quality rather than strictly false. Second, our findings indicate that, contrary to common belief, political leaders rarely share misinformation. The predictors for sharing misinformation are similar across the text-based approach and the dominant approach, with a few relevant exceptions. Yet, regardless of the measurement approach, many of the predictors for misinformation-sharing behavior in the mass public found in the existing literature are not relevant predictors for the behavior of political leaders in our data.

Existing approaches

In the most recent work on the dissemination of misinformation, the primary mode of detecting misinformation relies on a "domain-based" approach. Focusing on ordinary social media users in the United States, this approach identifies misinformation by whether links shared by users, websites visited by users, or search engine referrals come from a list

[1]This includes posts across Twitter, Facebook, and Instagram for 945 politicians.

of domains deemed as "fake" or "hyperpartisan/low-quality" news sites (Lin et al., 2022; Osmundsen et al., 2021; Pennycook and Rand, 2019; Grinberg et al., 2019; Guess et al., 2020a, 2019). The domain approach assumes the existence and relevance of these false news/low-quality websites and that they are the main sources of false information. As a result, the analysis of this content is primarily at the publisher-level. It also assumes that publishers identified as low-quality rarely post information that is true or high quality, which is inaccurate (Stewart et al., 2021). Notably, there exist applications which rely on fact-checked stories and therefore identify users who shared debunked stories rather than domains (Mosleh et al., 2021). These applications identify the sharing of falsehoods directly, yet they still rely on the assumption that users primarily share falsehoods through hyperlinks to debunked stories.

Another common approach to map the spread of false news is to rely on Facebook's massive URLs dataset, which contains information on engagement with URLs classified by Facebook and partner fact-checking organizations as containing false and misleading information (Allen et al., 2021). Putting aside concerns about the extent to which differentially private noise can lead to potential biases in measures of engagement with these URLs, this method of detecting false news is also largely based on the existence and relevance of domains as the main sources of false information. Social media companies often flag posts containing links to misinformation, and the content tagged by companies can be used to measure the prevalence of misinformation in social media (Guess et al., 2021; Théro and Vincent, 2022).

Neither the domain approach nor the URL-based approach will be able to fully detect false or low quality content if it does not reference an external hyperlink. Yet misinformation can also be structured around content that is plain text or a combination of non-textual features (images and videos) with plain text (Yang et al., 2023; Resende et al., 2019b; Machado et al., 2019). This type of misinformation can take the form of rumors with unknown sources or simply (false) textual narratives that accompany videos and images, for example. The existence of these unstructured texts with wide circulation online is likely a result of many factors, one of which is perhaps the popularity of private instant messaging apps like WhatsApp. These apps allow the creation of large groups of users for information sharing and do not have a "newsfeed," a feature that facilitates

the spread of misinformation through URLs (Rossini et al., 2021). However, unstructured text and media are also common features of social media newsfeeds. Overall, false content that circulates across social media in developing countries is not frequently linked to online sources, publishers, or websites (Pasquetto et al., 2020). In over four million posts in our data, 75% did not include a hyperlink to an external domain.

As a consequence, domain-based approaches may fail to capture misinformation in social media that is unaccompanied by hyperlinks to websites. The Facebook-URL approach may also fall short in detecting false information as it is constrained by social media companies' policies about misinformation, their incentives in publicizing and denouncing it, the designated resources regarding identifying misinformation, and the privacy concerns that prevent the analysis of uncensored user data (Allen et al., 2021).

The text-based approach

Here, we propose an alternative approach to detect misinformation that we developed based on texts containing rumors and false stories, using supervised learning methods, natural language processing, and human review of posts. We develop our approach using novel data from Brazil. The dataset contains 4,050 rumors that circulated in Brazil between 2018 and 2020. Importantly, we obtained the plain text of the rumors as they circulated across social media and popular instant messaging applications. If the rumors circulated solely as images and videos, we have the text embedded in the images (if any text was embedded) and, sometimes, we also have transcriptions of parts of the video. Most rumors, though, are largely composed of plain text, and many rumors are a combination of different media, including text (Peng et al., 2023). We then create a dataset of rumors and of rumor-free politicians' posts across three social media platforms to train a classification model that generates predictions about posts likely to contain misinformation. Using the classification model and text-similarity measures to find the set of posts "most likely" to contain misinformation, we then rely on human review of posts to detect the presence of misinformation in about 4 million posts from Facebook, Twitter, and Instagram by 945 politicians in Brazil between 2018 and 2020. In sum, similar to recent applications that aim to match users to specific debunked claims (Mosleh et al., 2021), our approach allows us to match shared user content to a larger set of false claims even if the user is not sharing

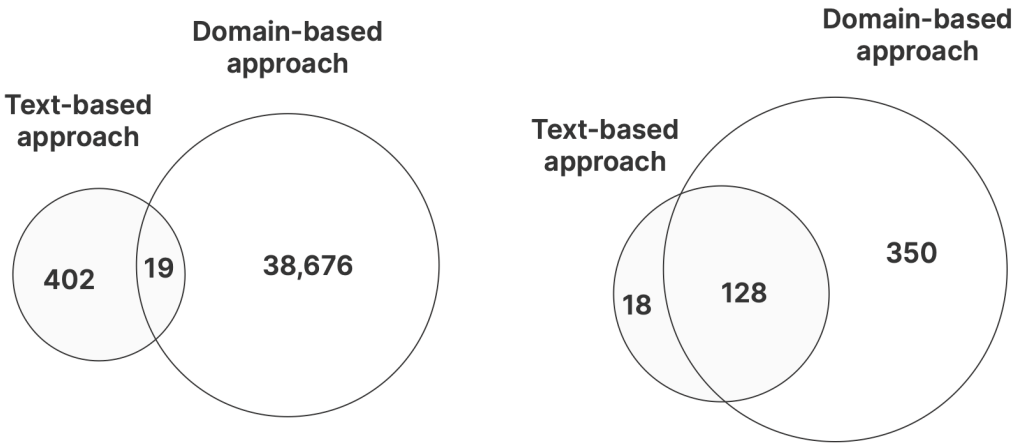
identical content.

Our results indicate that political leaders rarely share misinformation. Each of the three approaches—text-based, domain-based, and Facebook-URL—find that less than 1% of all posts contain misinformation. The proportion of politicians' posts that contain misinformation is broadly consistent with previous research analyzing political leaders in the United States, United Kingdom, and Germany (Lasser et al., 2022).

Yet, we find significant variation across these three detection approaches examined. The number of posts containing misinformation can range from 50 to 38,695, and the percentage of politicians who have shared misinformation can also range from 1.9% (18) to 50.5% (478). There is little overlap between these approaches in terms of which political leaders and posts are identified as sharing misinformation. Given overall low rates of sharing false content based the Facebook-URL detection approach, subsequent analyses primarily compare the text-based approach to the domain approach. For example, out of the 39,097 posts identified using the text- or the domain-based approach, only 19 posts were identified by both approaches (see Figure 1a), and out of the 496 politicians identified as sharing misinformation in either the text-based or the domain-based approach, 128 were identified in both (see Figure 1b).[2] Furthermore, these two approaches capture distinct content: the domain approach misses most of the false content identified by the text-based approach and does not appear to capture false content in general. Instead, it detects posts containing misleading and hyperpartisan content. By contrast, the text-based approach captures content fact-checked as false, but may miss content that is hyperpartisan, but still true, or that has not been reviewed by external sources for its veracity (see Figure 2). We systematically assess these trade-offs in the Discussion section.

To further illustrate these distinctions, we also explored whether the two different detection approaches – the domain approach and the text approach – identified different types of content. We evaluated the content of the posts flagged as false by the text-based

[2]Note that 83% of all posts tagged by the text-based approach do not contain a reference to an external URL. Of those that do reference an external URL, 4.5% are also captured in the domain-based approach. The remainder contain reference to an external URL but were not flagged by the Facebook URL domain list. The domains shared in these posts reference mainstream media outlets, like *Veja*, *Folha de São Paulo*, or *Estadão*, as well as party websites and personal blogs.



(a) Posts **(b)** Politicians

Figure 1. Overlap in posts and politicians identified as sharing false content by detection approach.

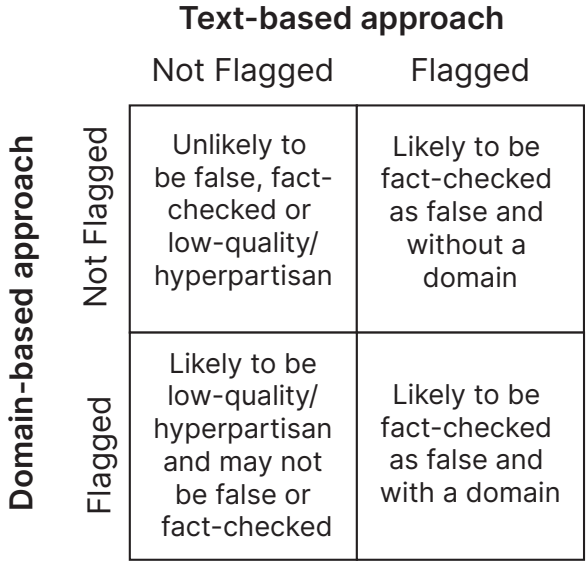
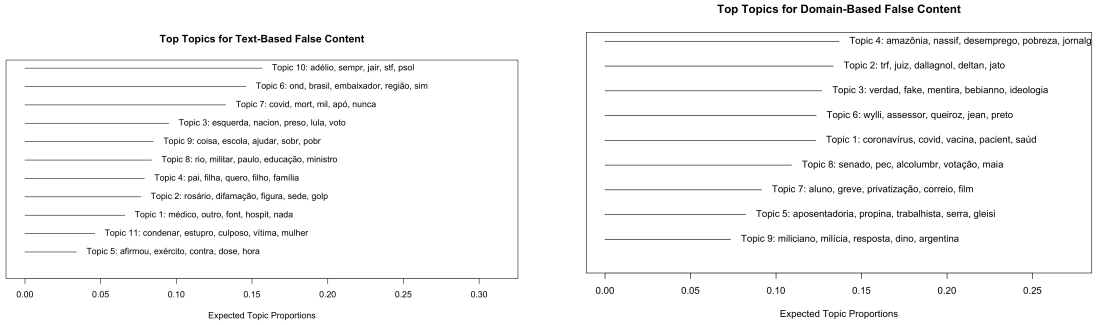


Figure 2. Overview of the types of content captured by each detection approach.

approach and the domain-based approach using a structural topic model with document-level covariates for the year of the post and platform (Roberts et al., 2014). We include additional diagnostics and topic-related information in the Appendix Section A.7. Figure 3 highlights the different topics flagged by each detection approach, focusing on words that are unique to each topic based on their frequency and exclusivity. There is overlap between the approaches, as both detect the Covid-19 epidemic and issues related to public safety. At the same time, the text-based approach (Figure 3a) seems to pick up on certain important moments – such as President Jair Bolsonaro's stabbing during the 2018 election or the Pope's statements interpreted as related to Lula and the rosary from the Vatican in 2020 – that the domain-based approach does not detect. In addition, the domain-based approach (Figure 3b) seems to pick up on polarizing, yet sometimes factually accurate moments, including the Lava Jato corruption probe or the presidential campaign in 2018.



(a) Text-Based Approach **(b)** Domain-Based Approach

Figure 3. Topics flagged by Structural Topic Models.

While the posts and politicians identified as containing misinformation are dependent on the measurement approach, results show that the characteristics that predict sharing misinformation at the politician level hardly vary based on the approach used to measure misinformation. In our models, many of the predictors for misinformation-sharing behavior identified in previous studies of ordinary social media users are not relevant for political leaders in this context. Overall, the built-in assumptions in different approaches to detecting misinformation matter in determining what posts count as containing misinformation, what content is tagged as misinformation, who shares misinformation, and how common misinformation is in political leaders' social media discourse (although all rates suggest low

prevalence). However, they matter less in describing what types of leaders, in terms of demographics and ideology, share misinformation.

We examine politicians' posts for two main reasons. Substantively, political leaders, and in particular high ranking politicians such as the ones included in our analysis, seek to and often succeed in influencing voters and public opinion in general (Broockman and Butler, 2017; Lenz, 2013; Gabel and Scheve, 2007). As a consequence, political leaders who disseminate false information could erode trust and support for democratic norms, beyond fueling beliefs in falsehoods. And, while politician-backed misinformation may not constitute the majority of misinformation online, evidence suggests that misinformation spread by elites may account for a major part of social media engagement with misinformation (Brennen et al., 2020; Mosleh and Rand, 2021). Methodologically, because politicians are public figures, their posts across all main platforms are public. This use of the text-based approach is particularly important for the study of misinformation in the Global South because it allows us to examine the prevalence of false content on Facebook and Instagram, which tends to be considerably more popular globally than Twitter (Kemp, 2021).[3] Finally, because of publicly available electoral and administrative data, we have the full set of elected and appointed political leaders, and we can select from an extensive sampling frame, which is often difficult to do with general population data collection.

Data

We rely on three different types of data: (1) social media content from Facebook, Instagram, and Twitter posts of political figures between January 2018 and December 2020; (2) false content circulating in Brazil between 2018 and 2020 according to Brazil's main fact-checking websites and the Facebook URLs Dataset; and (3) politicians' demographic data, collected from Brazil's Superior Electoral Court (TSE) electoral data repository, existing datasets from other studies (Zucco Jr et al., 2021), and through manual collection.

To collect politicians' social media posts, we created a comprehensive list of political leaders in Brazil, along with their available social media handles. Leaders include the presi-

[3]According to data from 2021, approximately 2.7 billion people use Facebook and 1.2 billion use Instagram worldwide, as compared to approximately 353 million that use Twitter.

dent, vice-president, cabinet members, governors, vice-governors, federal deputies, senators, and candidates for mayoral office in all state capitals in 2020 (N = 945). Using these social media accounts, we collected leaders' public social media posts from both CrowdTangle (Facebook and Instagram) and the Twitter Academic API. In total, we collected just over 4 million social media posts (1.5 million from Facebook, 1.1 million from Instagram and 1.5 million from Twitter).

To collect false stories circulating across Brazil from 2018 to 2020, we scraped fact-checked stories from the six main fact-checking agencies in Brazil (Boatos, E-farsas, Fato ou Fake, Lupa, Aos Fatos, and Projeto Comprova). To collect the content of rumors, we used data from one of the sources, *Boatos* (boatos.org). Unlike many other fact-checking agencies in Brazil, *Boatos* includes the exact content fact-checked by the organization, as opposed to a summary of that content written by the fact-checking organization. For our text-based approach to detecting misinformation, obtaining the content of the false stories is critical, as we rely on the substance (full text) of the false stories to detect posts with misinformation and eliminate politicians' social media posts that are unlikely to be false. In total, we compiled the original text of 4,050 false stories from *Boatos*.

Although recent research in the U.S. has found a high degree of consistency in ratings across fact-checking organizations (Lee et al., 2023), relying on a single fact-checking agency could lead to poor coverage of the unknown set of false stories that circulate in Brazil. We examine this possibility by comparing fact checks across other five main fact-checking agencies. We find that *Boatos* provides the widest coverage between 2018 and 2020. From a random sample of 200 *Boatos* posts, we found 72 false stories that also appear in at least one of the five other main fact checking websites in Brazil. The other 119 were only found in *Boatos*, and the remaining 9 were inconclusive because the selected story was in fact a collection of false stories or the stories were too ambiguous so we could not determine if they were reviewed by other fact-checking agencies.

Boatos differs from the other agencies as it only publishes stories it found to be false, whereas most other agencies publish all their checked stories, regardless of adjudicating them as true or false. This is not a problem for creating a dataset of false stories that we can use to detect political leaders' posts containing misinformation. Yet part of the

discrepancies between the other agencies and *Boatos* comes from stories that were checked by these agencies and found to be not false. Based on a random sample of 200 stories from the other five fact-checking agencies, we find that about half of those (104) were not in *Boatos*– stories are described differently in each fact-checking agency, so two research assistants working independently reviewed each story because there may be disagreement about whether or not the agencies were, in fact, reviewing the same content. Importantly, none of the stories that were examined by *Boatos* and the other agencies were adjudicated differently by these agencies. Out of the 104 stories not in *Boatos*, 63% (66) were classified as false. For the remaining 38 stories, they have a combination of true stories (16), sets of campaign promises and platform statements, stories that were about the fact-checking agencies themselves (not checked content), and stories that were not included in *Boatos*. While we cannot definitely rule out that *Boatos* provides systematically different coverage of rumors relative to other agencies, the evidence suggests that it has wide coverage, relevant overlap in coverage and the exact same adjudication as other agencies.

We utilize the Facebook URLs Dataset to create a list of domains and URLs that have been found to be associated with misinformation. This measure is designed to approximate the domain approach employed in research conducted in the United States, in the United Kingdom, and Germany (Lasser et al., 2022). Given that, to the best of our knowledge, there is no publicly available, systematic catalog of low-quality domains and URLs in the Brazilian context (for example, NewsGuard is not available for Brazil's market), we compile a list of all the URLs from Facebook's URL dataset that were shared in Brazil, flagged as false by third-party fact-checkers and were available in Facebook's URL dataset, which is partially limited to URLs that have been shared at least 100 times. We utilize both the exact URL from the dataset (N = 365) and the root domain (N = 228, excluding the common domains such as youtube.com, twitter.com, yahoo.com, and a main newspaper) from this list. While this list is surely incomplete, it is larger than and overlaps with the list independently compiled by the Global Disinformation Index (Index, 2021) that maps the risk of disinformation in Brazilian media market.[4]

[4]We obtained data from the Global Disinformation Index's (GDI) 2020 analysis of Brazilian domains. We include additional analyses using this data in the Appendix Section A.13 – results are largely consistent. We thank GDI for sharing their data.

Finally, to compare the predictors of misinformation sharing behavior using different detection approaches, we collect demographic data on politicians' party affiliation, electoral alignment (i.e., their party's coalition for presidential office), age, sex, education, and, as a measure for political experience, we identify whether political leaders had run for office in the past six electoral cycles or if they had been appointed for political office (for politicians in appointed office prior to 2018). Using federal legislature survey data from Zucco Jr et al. (2021), we estimate the ideology in 2018 for each politician based on their party affiliation. For cabinet members who do not belong to any party, we assign them the same ideology as the president's party ideology in 2018. Admittedly, this is an imperfect measure of ideology because the party's aggregate ideology is based on surveys of federal legislators and our list of political leaders includes other offices. Furthermore, given Brazil's multi-party and highly fragmented party system, not all parties in our dataset are included in Zucco and Power's data; we discuss different coding schemes for the ideology variable in the Appendix Section A.10.

Methods

In our text-based approach to detect misinformation using the content of rumors, we utilize supervised machine learning with the objective of eliminating a large number of posts that were unlikely to contain misinformation. We provide additional details on our text pre-processing steps for both the politicians' post data and the *Boatos* data in the Appendix Section A.4.

As a first step, we used stratified random sampling on the year to sample 4,050 Facebook posts without any false stories from 23 verified, well-established, professional Brazilian news outlets identified by Reuters to match the number of unique *Boatos* texts in our dataset. In order to be included in our model, any post needed to be over 10 unique words. We then trained a Naive Bayes classification model (90.86%-93.33% accuracy on a test set) with an alpha of .1, after conducting a grid search to evaluate the optimal value for this hyperparameter, on a subset of these posts. Model performance was similar across other types of classifiers, including a Deep Neural Network model (88.37%-91.66% accuracy); and an Elastic Net model (88.80%-92.28% accuracy). Since all models showed similar performance, we utilize Naive Bayes, and the F1 score for our model is 0.926, indicating

both high precision and high recall. We include the precision-recall curve, receiver operating characteristic (ROC) curve and confusion matrix in the Appendix Section A.5.

To assess where classification errors occurred at this stage, we reviewed 145 randomly sampled headlines. We found that 78% of the articles selected from reputable news sources were misclassified as false when they deployed more sensationalist rhetoric, used clickbait, and otherwise did not resemble a news article headline. Where the classification model identified fake posts from *Boatos* as "real," we found that 80% of the false content read more like a news articles or did not use sensationalist language. As a next step, we applied this model to over 4 million political leaders' posts across all platforms and selected posts that had a predicted probability of less than .1 of containing false information for each platform (Facebook, Instagram, and Twitter). This process allowed us to compile a dataset of actual social media posts from politicians that were very unlikely to contain false content.

Using stratified random sampling by political office, we then re-sampled 2,000 posts from politicians without any false stories, as determined by the initial classification model, for each platform. Two Portuguese-speaking research assistants, working independently, also verified, against our complete dataset of rumors, that these posts did not contain false stories. Based on this dataset of posts without false stories and the dataset of 4,050 false stories, we trained a Naive Bayes model for each platform (Facebook, Instagram, and Twitter), to predict whether a post contains a false story.

We applied these three models to the larger dataset of about 4 million social media posts total in an effort to eliminate posts that were unlikely to contain false content. We trained separate models due to possible linguistic or stylistic differences of posted content across platforms.[5] Across the models, our F1 score ranged from .940 to .954, and the prediction accuracy on a test set varied between 92.40% and 96.07%. We eliminated any post that had a predicted probability of less than .9 of containing false stories.[6] This left

[5]The accuracy on a test set for the Facebook model was 94.03%-96.07%. The F1 score was .954. The accuracy on a test set for the Instagram model was 92.40%-94.83%. The F1 score was .940. The accuracy on a test set for the Twitter model was 93.35%-95.73%. The F1 score was .953. All models use and alpha of .1. We include the precision-recall curve, ROC curve, and confusion matrix for these models in the Appendix Section A.5.

[6]Due to the volume of posts, we selected only posts with the highest predicted probability of being

us with 423,191 "suspicious" posts across all social media sites, or approximately 10.5% of our initial dataset.

To further trim this dataset, we compared the content of each of these suspicious posts with each false story scraped from *Boatos*. The goal was to identify which of the suspicious posts indeed contain false information, as defined by a fact-checker agency. To represent text numerically, we utilize a Term Frequency-Inverse Document Frequency (TF-IDF) transformation, which more heavily weights unusual words in a sentence.[7] We then calculate the cosine similarity across these numeric vectors to measure content overlap between each post and each fact checked article. The cosine similarity is a commonly used measure in text analysis to identify similar documents in a vector space based on the cosine of the angle between them (Han et al., 2011). A higher cosine represents a higher degree of similarity between documents. Following this process, we were then left with a measure of similarity between 0 and 1 for every suspicious post and every false story.

We excluded any pair of suspicious post and false story with a similarity below .4, which suggests that there was likely little to no overlap in content. We chose .4 as the cut off due to the fact that below this threshold, the quality of the matches deteriorated rapidly. Above this threshold, posts shared many common linguistic features. This further culled our sample to 4,102 total posts (from over four million), which we then reviewed manually. Overall, this process identifies politicians' posts containing falsehoods verified by fact-checkers without requiring these posts to be the exact same text as the stories used in the training models. In fact, most of the posts identified as containing false information were not direct quotes from our dataset of rumors, but rather new content that overlapped with the fact checked content.

For each post in the manual review, we relied on two Portuguese-speaking coders working separately to determine whether the social media post aligned with the false content

classified as false according to our models while still ensuring a large number of posts for evaluation through text matching procedures. Through this process, we included 128,073 Facebook posts, 61,986 Instagram posts, and 233,132 Twitter posts. In cases with a lower volume of posts, researchers could consider skipping this filtering step and instead directly applying the cosine similarity step to any post with a predicted probability >.5 of being false across the entire database.

[7] We also transformed the text using a bag-of-words approach but found the weighting of more important words in the TF-IDF vectorization process gave us better quality matches.

to which it was matched via a cosine similarity greater than .4. For any disagreements between the first and second coder and for all posts with the two coders classifying the post as containing false information, a third coder reviewed their assessment to make a final determination. In total, coders disagreed 241 times, or approximately 5.8% of all manually reviewed observations. We also took a random sample of 335 posts in which two coders agreed that they were not a match to a false story, and we had a third coder review them. We found 6 (1.8%) instances in which the third coder classified the post as a match to a false story. All of these instances were discussed among the co-authors who decided on classifying the post as containing a false story or not. An overview of the entire process, as well as the number of posts remaining at each point in the process is included in Table 1 and the flow chart in Appendix Figure 23.

Table 1. Process for Identifying False Content Using the Text-Based Approach.

Classification Step and Post Description	Total Number of Posts Left	Proportion of Original Dataset
Raw data: All politicians' posts	4,032,907	1
Naive Bayes classification model: Posts linguistically unlike real news	423,191	0.105
Cosine similarity >.4: Posts that are linguistically similar to false claims	4,083	< 0.001
Manually reviewed: Posts that are false	421	< 0.0001

To measure misinformation sharing behavior among political leaders using the domain approach and the Facebook-URL approaches, we rely on character string match procedures. For the domain approach, we identify all posts that contain a hyperlink to the root domains in Facebook URLs dataset. For the Facebook-URL approach we also identify posts, across all platforms, that contained the exact URL (as opposed to the root domain) labeled as false in Facebook URLs dataset. Due to the way URLs shared on Twitter appear in the API data, we unshortened all domains where applicable.

Results

Table 2 shows that, using the two common approaches and our own approach, posts from politicians that contain falsehoods are rare. Using our text-based approach, about 0.01% (421) of politicians' posts contain false stories. Politicians are very active users of social media, posting, on average, 6 times a day or 131 times a month. Yet only 0.01 out of 100 posts contain false information. The domain-based approach suggests a much larger share of posts containing misinformation (0.96%) and Facebook URLs approach indicates a much smaller share (0.001%). Regardless of approach, however, less than 1% of all the content shared by politicians contains misinformation, confirming that this practice is a rare event.

Table 2. Prevalence of Politicians Sharing False Content and Posts Containing False Content by Detection Approach.

Detection Approach	Total Identified	Pct. of Politicians Sharing False News	Pct. of Posts from Politicians Sharing False News
Text	421	15.4497	0.0104
Domain	38,695	50.5820	0.9595
Facebook URL	50	1.905	0.0012
n total		945	4,032,907

Posts containing misinformation, though infrequent, garner more user engagement than posts without misinformation. For example, posts containing misinformation using our text-based approach receive 10 times more engagement than posts without misinformation (in terms of median number of reactions).[8] The magnitude of the differences between posts with and without misinformation varies somewhat by measure of online engagement and approach of detection, but the pattern is clear: posts with misinformation receive more attention online than posts without misinformation. Overall, using at least one of the methods employed (text, domain, Facebook URL), we detect false content in 0.9694% of

[8]We calculate user engagement by combining all types of reactions across the three social media platforms. See the Appendix Section A.9 for more details.

posts, and these 0.9694% of posts represent about 1.4197% of all the online engagement to politicians' posts in the period analyzed. In terms of views, we also find that posts with misinformation tend to be seen (on average and in terms of medians) more often than posts without misinformation, but we do not find that posts with misinformation are seen disproportionately more relative to their size – in fact, they are seen at lower frequencies than posts without misinformation (only 0.06% of views are to posts with any misinformation).

Our text-based approach reveals that 15% (146) of politicians have shared misinformation at least once, whereas the domain approach finds that at least 50% (478) of politicians have done so. In contrast to these two approaches, Facebook's URL approach find that 1.48% (14) of politicians have shared misinformation at least once.

Furthermore, the estimates relying on text and domain approaches indicate that a larger share of politicians post misinformation compared to social media users in the general population – 8.5% in a US sample of social media users shared misinformation (Guess et al., 2020b). At the same time, estimates relying on Facebook URLs indicate that a much smaller number of politicians share misinformation compared to the general population.

The politicians who are identified as spreading misinformation also differ across the three approaches. Out of the 146 politicians who are identified as sharing false stories using the text-based approach, 12.3% (18) were not identified in the domain-based approach. Out of the 478 politicians who were classified as sharing misinformation in the domain-based approach, 73.2% (350) were not identified as sharing false content in the text-based approach (see Figure 1b). With the Facebook URL approaches, we find even greater rates of mismatch: at least 91.8% (134) of the politicians identified in the text-based approach are not identified in Facebook-URL approaches, and 33.3% (6) of the politicians identified in the URL approach were not identified in the text-based approach. Due to the reliance on website links, all politicians identified through the Facebook-URL approach were also identified via the domain approach.

Figure 4. Post with false story: Post captured by the text-based approach.

The posts identified as containing misinformation, across all social media platforms, also vary depending on the approach. The domain approach flags 38,695 posts as containing misinformation, the text-based approach identifies 421 posts, and the Facebook-URL approach identifies 50 posts. Furthermore, there is little overlap between these approaches; only 19 posts were identified by both the text-based and the domain-based approach (see Figure 1a). To further illustrate the differences between the approaches, Figure 4 shows a post captured by the text-based approach, containing a video accompanied by text falsely stating that former presidential candidate Haddad created programs named "gay kit" and "crack benefit." Figure 5 shows a post captured by the domain approach, which includes a link to a piece of (truthful) news from a domain included in our list of domains at risk of sharing misinformation. In this particular example, both posts contain links, but the element of falsehood for the post shown in Figure 4 was detected in text.

Figure 5. Post without false story: Post captured by the domain-based approach.

The content of the posts identified by the text- and domain-based approaches are also distinct. To investigate these approaches' ability to detect false stories, we randomly sampled 200 posts identified by the domain approach as containing misinformation. Two research assistants examined the stories linked in the posts and the content of the posts themselves against our full dataset of all fact-checking agencies (about 15,000 fact-checked stories). Of the 139 posts with working links, 131 (95%) stories were not checked by any fact-checking agency. Of the 8 stories checked by a fact-checking agency, 7 were deemed false or misleading and 1 story was checked and deemed true. Furthermore, only 5 of a sample of 2,516 posts containing hyperlinks to domains listed in the domain approach were fact-checked by the social media companies (based on a visual tag by the social media company).

All of the 421 posts deemed as containing false information by the text-based approach were reviewed by three researchers to determine whether these posts matched stories adjudicated as false by a fact-checking organization, making the likelihood that these posts do not contain falsehoods very low. Without this step, this approach would still detect these posts. However, it would also capture thousands of additional posts that might not contain false claims, but may be linguistically similar to or even refute the false stories in our database. Despite these precautions, there could be a sizable proportion of posts that do contain falsehoods but remain undetected by the text-based approach (false negatives). For example, it is possible that posts with lower degrees of similarity could contain false stories, since we only reviewed posts with a cosine similarity measure greater than 0.4. We examine this possibility empirically by reviewing 300 randomly selected posts with a probability of at least 0.9 containing a false story, as determined by our classification model, and a cosine similarity between 0 and 0.4, from which 3 coders identified no (0) false stories. In the subset of posts with cosine similarity greater than 0.4, the rate in which we found false stories was 10.3%. It is also possible that posts with a predicted probability of being false below our .9 threshold might still have a cosine similarity $> .4$. We also evaluate this empirically by taking a random sample of 306 posts, stratified by year and platform. Just three of 306 posts (.1%) produced a cosine match greater than .4 but less than .45. Only one of these featured a candidate potentially sharing an ambiguously false claim, which we discuss in depth in the Appendix Section A.8.

We plot the distribution of this random sample by predicted probability of being false and cosine similarity in Figure 6.

Given that posts with probability lower than 0.9 of containing a false story are even less likely to contain a false story than posts with probability higher than 0.9, the text-based measure is unlikely to underestimate to a large degree the extent to which social media posts contain falsehoods, unless those falsehoods have yet to be fact-checked externally. In short, even if the text-based measure is more conservative in detecting falsehoods by relying only on claims already fact-checked as false, the domain approach does not typically capture strictly false content, as defined by fact-checking organizations. Instead, it seems to identify hyperpartisan content from websites that have in the past shared false content, but may do so rarely or in conjunction with real – if polarizing – news. It is unable to capture the

content from the text-based approach due in part to the absence of external URLs in these posts.

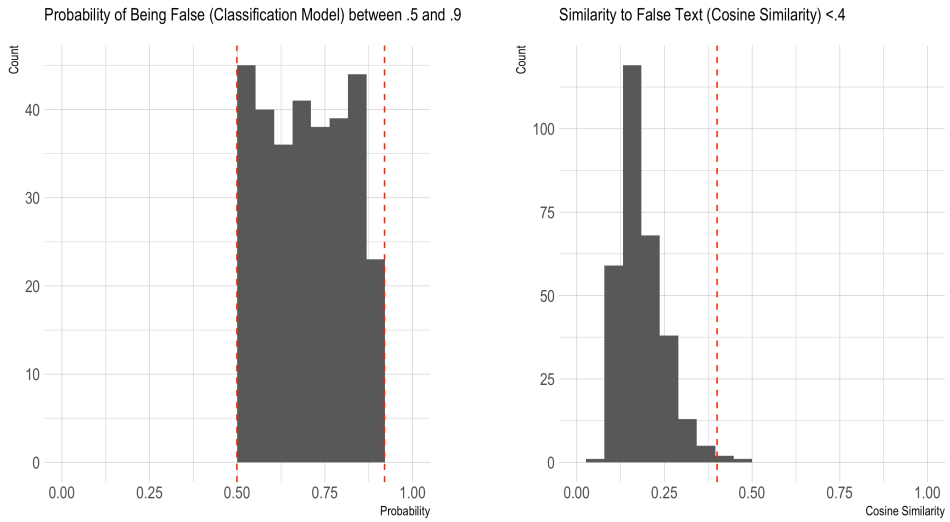


Figure 6. Distribution of Posts Classified as False with a Predicted Probability between .5 and .9 and their Highest Cosine Matches with False Claims.

Predictors of Misinformation Sharing Behavior

We further explore whether the type of politician who shares misinformation varies based on two of the three approaches to detecting of misinformation; we exclude the Facebook URL measure because only 14 politicians shared misinformation containing the exact URL in the Facebook-URL dataset. Using Ordinary Least Squares (OLS) models with robust standard errors and office fixed effects, we compare which factors explain variation in sharing false information by regressing an indicator for sharing misinformation (binary) on predictors of misinformation sharing behavior.[9] Further, we explore if these predictors can help to explain *how much* false information politicians share, using a quasipoisson model

[9]In the appendices, we also conduct this analysis using a probit and logit model. Results are consistent across the three models.

with political office fixed effects.[10] We select predictors of misinformation sharing behavior commonly found in the literature on social media users (age, sex, education, ideology, partisanship, and number of posts), and we included a few other variables that are relevant to political elites and the Brazilian context (political experience, alignment in the 2018 electoral coalitions, and political office).

In these models, political experience indicates whether a politician ran for office prior to 2018 and/or had been appointed to political office before (=1 and = 0 if not). Ideology ranges from -1 to 1 where lower numbers indicate left-leaning ideology and higher numbers indicate right-leaning ideology. We code electoral alignment as a categorical variable indicating whether the political leader belonged to a party in Bolsonaro's electoral coalition, Haddad's electoral coalition, or other (the reference group) in 2018.[11] All models contain political office fixed effects, but results are all but identical without them. We recoded all predictors to vary between 0 and 1.

Figure 7, which uses a binary outcome for whether the politician ever shared a false claim, shows consistent results on whether the same predictors explain variation in sharing behavior across the different approaches to detecting misinformation. According to both the text-based and domain approaches, politicians who were part of presidential candidates' (Bolsonaro or Haddad) electoral coalitions in the 2018 elections are more likely to share misinformation at least once compared to those who were not part of the coalitions of the main contenders for presidential office.[12]

[10] In the appendices, we also conduct this analysis using poisson and negative binomial models. Results are also fairly consistent across the three models. In addition, we use a Two-Step Heckman Model, which somewhat aligns with these other models.

[11] Jair Bolsonaro was the far-right president (2019-2022) and Fernando Haddad is a politician from the left-wing Workers' Party who was the runner-up in the 2018 presidential election.

[12] Note that, according to Table 17 in our Appendix, belonging to Haddad's coalition is not a statistically significant predictor of sharing misinformation for the repeated offenders and GDI measures of the domain-based approach – the standard errors remain largely unchanged from the main domain-based approach, but coefficients' estimates drop to about half.

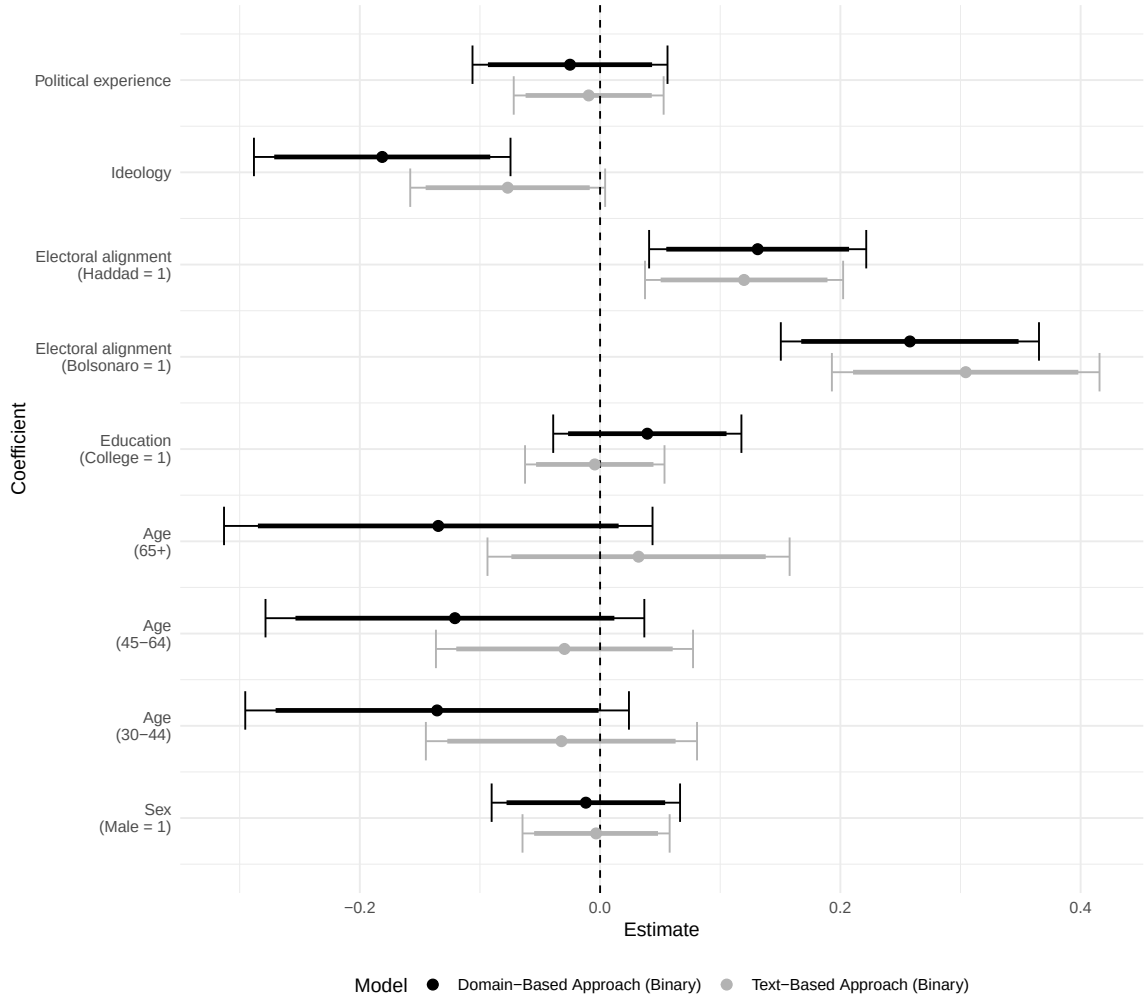


Figure 7. Predictors of Sharing Misinformation (Binary) by Different Detection Approaches.

Note. OLS model with robust standard errors includes political office fixed effects and control for the total number of posts by a politician (not shown). Bars indicate 90% and 95% confidence intervals. Different ranges in the x-axis for Figures 7 and 8 to facilitate visualization because of the different scales in the count and binary outcomes

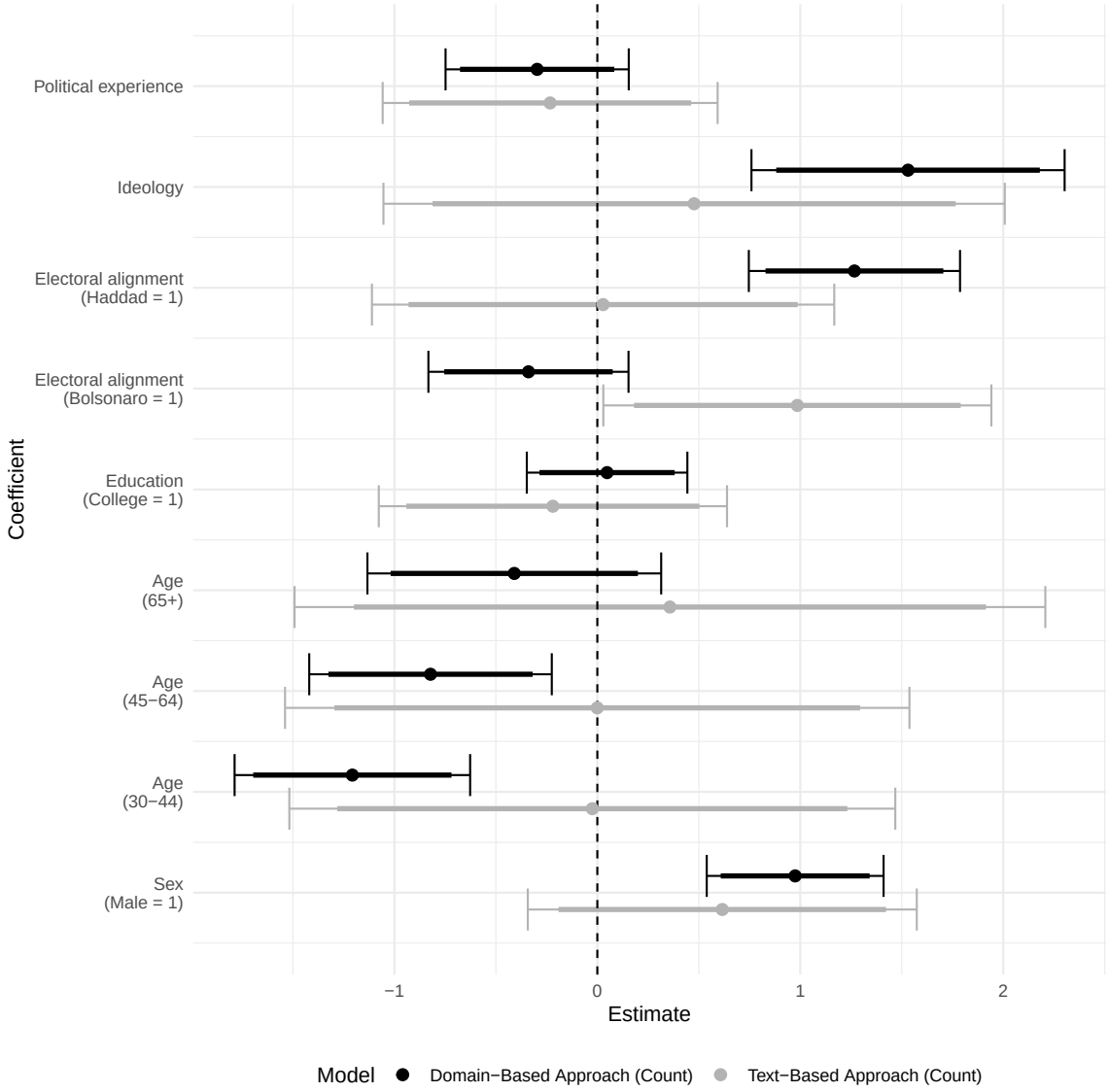


Figure 8. Predictors of Sharing Misinformation (Count) by Different Detection Approaches.

Note. Quasipoisson model includes political office fixed effects and control for the total number of posts by a politician (not shown). Bars indicate 90% and 95% confidence intervals. Different ranges in the x-axis for Figures 7 and 8 to facilitate visualization because of the different scales in the count and binary outcomes.

Yet, in Figure 8, which uses a count outcome to measure *how many* false claims a politician shared, the text approach and the domain approach have more of a contrast. For the text approach, we find that belonging to Haddad's coalition (the losing candidate) is not a predictor for number of posts with false stories, whereas it is a predictor for the domain-based approach. The only statistically significant predictor for the text approach is belonging to an electoral coalition aligned with Bolsonaro, which is not significant for the domain approach. However, caution is needed in the interpretation of these estimates, even for descriptive purposes: the uncertainty around the estimates for the text-based approach is large and the difference between the estimates for Haddad and Bolsonaro alignment is not itself statistically significant – and, regardless of statistical significance, the width of the confidence intervals suggests that the size of that difference is uncertain. That said, the uncertainty around the estimates for the domain-based approach is much narrower and suggests a starker difference between the estimates for Haddad and Bolsonaro alignment in the domain-based approach. Overall, and in light of the other models shown in the Appendix Section A.15 (poisson, negative binomial, and Heckman 2-step model) the evidence suggests a stronger positive association between belonging to Bolsonaro's coalition and number of false stories shared (identified via the text-based approach), and also a stronger positive association between belonging to Haddad's coalition and number of misinformation or hyperpartisan stories shared (identified via the domain-based approach).[13]

In terms of other predictors, we find evidence that ideology is a predictor for the domain approach across several models,[14] but the direction depends on the outcome variable: more right-leaning politicians are less likely to share misinformation at least once, but tend to share misinformation more often in the count model, as shown in Figure 8. On the other hand, right-left leaning ideology is not a robust predictor of sharing misinformation as measured in the text-approach, or when we use other operationalizations of the ideology variable (see Appendix Section A.12). But, consistent with findings from Germany (Lasser et al., 2022), Appendix Tables 15 and 16 show that less ideologically extreme politicians are less likely to share misinformation, across different measures of extremeness and misinformation. Overall, the differences between the text- and domain-based approaches in terms the

[13]Our descriptive cross-tabs and difference of means in Tables 25-26 and 38 are consistent with these interpretations.

[14]Negative binomial and Heckman are the exceptions.

strength of ideology as a predictor for sharing misinformation should be approached with some skepticism given the uncertainty around the estimates for the text-based approach and the overlap with the main estimates for the domain-based approach.[15]

Furthermore, regardless of the measure of misinformation sharing behavior, there is little evidence of age and education as predictors of misinformation sharing in our models (and the uncertainty around estimates for the text-based approach is large),[16] despite evidence from research focused on general population social media users. Yet, similar to findings from general population social media users, politicians who post more are more likely to share misinformation (coefficients not shown in Figures 7 and 8, but can be found in the Appendix Section A.11).

Overall, across different measurements of misinformation sharing behavior and models, no single predictor consistently explains variation in misinformation sharing behavior among politicians except for the (logged) number of posts by a politician. The only predictor that is systematically associated with sharing false stories, using the text-based approach, is belonging to Bolsonaro's (the winning candidate in the 2018 presidential elections) electoral coalition, but it is not a consistent predictor for other measurement approaches, such as domain. Belonging to Bolsonaro's coalition consistently predicts sharing false stories using the text-based approach, across different models, operationalizations of the dependent variable, and case-based deletion of individual politicians. This is perhaps unsurprising to observers of Brazilian politics, as several members of his coalition are under investigation for spreading false news in the 2018 elections (McCoy, 2020; Palau, 2021). In several models, and across all domain-based models, belonging to Haddad's coalition is a predictor of misinformation-sharing behavior, but the coefficients' statistical significance and magnitudes vary substantially based on how we operationalize the dependent variable (dummy, number of stories, or natural logarithmic transformation of number of stories).

[15] Tables 32-33 suggest very small differences between left- and right-leaning politicians in terms of sharing (binary) misinformation in both approaches, even though the differences are larger for in terms of number of posts for the domain approach, but they suggest left-leaning politicians share more misinformation (different from findings in Figure 8).

[16] Furthermore, there is a lot of discrepancies in these demographic predictors' estimates' magnitudes and directions depending on whether the domain-based approach uses the repeat offenders, GDI, or the main domain-based measure.

Discussion

Given the disparate results for each misinformation detection approach, but somewhat similar results regarding the type of politician who shares misinformation, these approaches may capture different, yet related, online behaviors. We found that the text-based approach measures fact-checked false (or factually inaccurate) stories more accurately than the domain-based approach. The domain approach appears to miss many instances of false content and to identify many posts without factually false stories as containing misinformation. By definition, it also excludes 75% of all posts since they do not contain reference to any external URLs.

On a fundamental level, both approaches rely on fact-checking as a way of determining what is misinformation, although the extent to which content is fact checked influences each approach differently. Given that the text-based approaches uses the fact-checked content as part of the classification model, measures of text similarity, and for the manual verification, limitations in what is selected to be fact-checked and competently verified versus what is false[17] could create problems for the text-based approach's effectiveness in detecting misinformation. The domain approach, while reliant on fact-checking content for the creation of the list of domains,[18] expands on fact-checked content by including all links to domains that had been fact checked; thereby the domain approach is less constrained by the specific choices made by fact checkers on what to verify.

The domain approach may capture highly biased content since it utilizes domains or URLs that were also tagged by fact-checking organizations working in partnership with a social media company, even if does not detect strictly false content. To evaluate to what extent the domain approach detects hyper-partisan content, our research assistants classified the randomly selected 200 news articles that were linked to the posts identified via the domain approach according to their level of "politicization/partisanship," understood as political content with a clear political/partisan slant. On average our coders found that

[17]Importantly, we make the assumption in this manuscript that we cannot determine what is true or non-true without fact-checking, although alternatives (e.g. crowd sourcing) potentially exist. Naturally, this creates implications in finding misinformation because a lot of content shared by individuals is not, in principle, amenable to be fact checked.

[18]Importantly, the list of domain created using GDI data is not strictly based on fact-checking.

in 70% of posts, either the contents of the link or the text posted in social media, were "politicized/partisan," or "clearly partisan/politicized" and less than 20% were considered "neutral."[19] While measures of partisan/political slant are somewhat subjective and likely to be influenced by the coders' perceptions about the context and the politician, they are indication that posts and links identified via the domain approach contain partisan content.

Finally, one potential limitation of the domain-approach comes from our method of developing the list of domains based on Facebook's URL-dataset. Our main analysis includes all domains,[20] which may be too expansive. In an alternative analysis, we include only domains that were found at least twice in Facebook URLs dataset (N=59 domains). As shown in the Appendix Section A.13, the domain approach, when analyzed using only "repeat offenders," still has little overlap with the text-based approach; and it further weakens the relationship between belonging to Haddad's coalition and sharing misinformation. Finally, using domains independently verified by the Global Disinformation Index (N=17 domains), coefficients are similar to the domain-based approach. Furthermore, to the best of our knowledge, there are no systematic alternative lists available, as fact-checking websites in Brazil do not publish hyperlinks to false stories that originate from websites.

Overall, differences in the type of content captured by text- and domain-based approaches likely reflect different aspects of what constitutes misinformation. Much of the public and academic discussions about misinformation conflates information that is factually false with manipulated or misleading content, and also information containing some form of hate speech and harassment (Bovet and Makse, 2019). While we neither advocate for a stricter or more encompassing definition of misinformation, nor offer a way to determine the deliberate, intentional spread of false information, falseness and hyper-partisanship represent distinct dimensions of misinformation, with different implications for democracy and accountability.

It is also important to consider the limitations of the text-based approach. The text-

[19] This excludes about 60 links that were nonworking links.

[20] With a few exceptions, including two fact-checking websites, *Aos Fatos* and *E-Farsas*, a prominent media outlet, *Folha de São Paulo*, a general news search engine, br.noticias.yahoo.com, and two social media websites, YouTube and Twitter. Except for YouTube and Twitter, each of these domains only appeared once.

based approach utilizes fact-checking as the "arbiter of truth." As such, posts that cannot be linked to stories, events, or statements verified by fact-checking agencies cannot be identified as false in the text-based approach. Although we focused on one specific organization with wide coverage, it is possible that this sample of claims could be expanded by incorporating additional fact-checking entities or even crowd sourced reviews. We also rely on supervised learning methods and natural language processing techniques to prune a large number of posts unlikely to contain misinformation, which may exclude false or misleading content that is presented without partisan rhetoric. Finally, the text-based approach still relies heavily on human review and may be less useful to detect misinformation "in real time" for large quantities of data. The domain-approach is likely more efficient, but, as noted above, it also is likely to capture more hyperpartisan than strictly false content.

Furthermore, we are constrained in our identification of false content to the stories fact-checked by external organizations, who have limited bandwidth to investigate all potential claims and their own editorial choices in selecting content to be verified. Therefore, it is possible and plausible that there is misinformation that fails to be captured by the text-based approach because it was missed by fact-checkers (an additional source of false negatives).

To empirically assess the possibility that false content was left out due to fact-checkers' limited resources and editorial choices in what to review, and to further explore more systematically the rates and sources of false positives and false negatives in both the text and domain approaches, we partnered with a fact-checking organization to review the veracity of a sample of 1,509 posts from our population of about 4 million posts.

We took three random samples of all posts, stratified by year, and asked a Brazil-based fact-checking organization[21] to review the post for false claims. Given that the sharing of false content is a rare event, the three samples included (1) one that was entirely random – or a naive sample (N=501); (2) one that was additionally stratified based on whether or

[21] We contacted *Boatos*, the same organization from which we draw our main fact-checking dataset, to conduct the review of our sample for methodological and logistical reasons. Methodologically, we wanted to keep the same fact-checking standards rather than introduce new differences that could complicate comparisons between classifications. Logistically, we had already been in touch with *Boatos* due to our questions about their fact-checking standards, which facilitated the cooperation.

not a post had been coded as false based on the domain-based classification approach – or the domain sample (N=504); and (3) one that was additionally stratified based on whether or not a post had been coded as false by the text-based classification approach – or the text sample (N=504).

The fact-checking organization then reviewed every post across these three samples and assessed whether the post was verifiable (i.e., included claims that could be fact-checked), and if so, it was true, false, exaggerated, or an unproven rumor. We compare results from this fact checker review to our different classification approaches in Table 3. In the table, dark grey cells denote false positives, or posts we coded as false (using either the text- or domain-based approach) that the fact-checking organization coded either true or not verifiable. Light grey cells denote false negatives, or posts we coded as true (or not false, by either approach) that the fact-checking organization coded as either false or a rumor, exaggeration, misleading, etc.

In the naive sample, we find high rates of true negatives – or posts that neither we nor the fact checking organization coded as false – for both the domain (95.81%) and text (96.6%) based approach. This is likely due to the low prevalence of false claims coded across the entire database of posts and the fact that this random sample was agnostic to whether or not a post had been classified as false by any detection approach. The false negatives (that we classified as true and the fact checker classified as false), which make up around three percent of the posts in both approaches, include primarily extreme exaggerations but also a few factually false posts. The false positives (that we identified as false but were rated as true or non-verifiable by the fact-checking organization) for the domain classification approach (0.8%) linked to domains in our list were primarily opinions. Overall, the naive sample suggests that both the text-based approach and the domain-based approach miss non-fact-checked false content at similar rates even though the text-based approach is more directly reliant on fact-checked content.

Table 3. Percent false positives and false negatives from sampled data for posts that were verifiable by external fact-checkers.

Sample type	Fact checker review	Classification approach			
		Text-based		Domain-based	
		False	True	False	True
Naïve *(N=501)*	True or not verifiable	-	96.6%	.8%	95.81%
	False	-	.8%	0%	.8%
	Rumor, exaggeration, misleading, etc.	-	2.4%	0%	2.4%
Domain *(N=504)*	True or not verifiable	-	93.01%	44.05%	48.4%
	False	-	1.4%	.79%	.59%
	Rumor, exaggeration, misleading, etc.	-	6.2%	6.16%	1%
Text *(N=504)*	True or not verifiable	8.3%	48.8%	0.2%	56.9%
	False	27.6%	.4%	1%	27%
	Rumor, exaggeration, misleading, etc.	14%	.8%	1.2%	13.7%

Note. Dark grey cells denote false positives, or posts we coded as false that the fact-checking organization coded either true or not verifiable. Light grey cells denote false negatives, or posts we coded as true that the fact-checking organization coded as either false or a rumor, exaggeration, misleading, etc. Cell with "-" indicate that no posts coded as false via the text-based approached were randomly selected in the naive and domain-approach samples.

In the domain sample, which stratified posts based on whether they were coded as false by the domain classification approach, we find a high rate of false positives (that we identified as false but were rated as true or non-verifiable by the fact-checking organization) but an overall low rate of false negatives (that we identified as true, but the fact-checking organization classified as false). Around 44% of posts included in this sample that we coded as false based on the presence of a suspicious domain, were coded as true or unverifiable by

the fact-checking organization. These posts largely focused on area where hyper-partisan news and opinions are prominent, including politics, social issues, and political scandals. Where we found false negatives (less than 2% of posts for the domain approach), posts included claims that were primarily exaggerations and rumors, though a few were false, including posts about COVID-19 quarantines and attempts by the Workers' Party to interfere with the federal police.

In both the naive and domain samples, we failed to capture any posts coded as false by the text-based classification approach. This is entirely based on chance and likely due the fact that the coding of posts using this strategy produced far fewer false posts than the domain-based classification method. To account for this, we created a third sample, the text sample, which stratified posts based on whether they were coded as false by the text-based classification approach. We compared our classification against the fact-checker review, and found that the false positive rate (posts we rated as false and the fact-checker as true or not verifiable) for posts dropped from 44.05% (in the domain-based approach in the domain sample) to 8.3% (in the text-based approach) or 36 percentage points, which as expected, suggests that the text-based approach is far more adept at avoiding the classification of posts that are true (or non-false) as false. The false positives that the text-based approach did produce primarily resulted from a mix of disagreement on what is verifiable by us (compared to the fact-checker), human error (usually when claims were true during a window of time and at other times were false), or due to content that was borderline true/false due to ambiguities in language and context. In terms of false negatives for the text-based sample, the text-based approach did not produce a high number of false negatives (about 1.2%), suggesting that we are not missing a substantial number of false stories by relying on a comprehensive corpus of fact-checked claims.

Across the three samples, the text-based approach produces at a minimum 1.2% and at a maximum 7.6% false negatives (posts we deemed as true and the fact checker as false) and the domain-based approached produced at a minimum 1.2% and a maximum 40.7% false negatives – primarily due to the fact that the domain approach, by definition, excludes any post that does not include reference to an external URL, or nearly three quarters of all posts in the data set.

Given the trade offs between false positives and false negatives, we view the text-based approach as an important addition to classification processes seeking to document the spread of false or misleading claims.

Conclusion

Determining who shares and what is shared as misinformation is highly dependent on the measurement approaches. Still, these different approaches to detecting misinformation consistently indicate that misinformation sharing behavior is a rare activity for political leaders in Brazil. The overall picture is that while up to one half of political leaders posted at least one piece of misinformation, it makes up a small part of their online activity on social media. We also find that political variables, such as electoral alignment with Bolsonaro or Haddad and ideology, are more systematically associated with sharing misinformation, but these predictors vary by measurement approach and operationalization of the outcome variable. Among these predictors, electoral alignment with Bolsonaro is a more robust predictor of sharing false stories when using the text-based approach.

Overall, the text- and domain-based approaches capture different, yet complementary, aspects of misinformation. The content identified by the text-based approach is more likely to contain falsehoods (as in factually false content), while the domain-based approach misses much of the false content, but appears to detect highly biased, hyper-partisan content, which may in some cases be partially true.

This paper focuses on a case of a developing country, but the type of misinformation identified using the text-based approach could potentially be found in developed and developing nations alike. The circulation of unstructured forms of text, via YouTube and podcasts, are widely popular in developed nations and can contribute to the dissemination of misinformation (Wirtschafter and Meserole, 2022; Serrano et al., 2020). It is clear, then, that current detection approaches that rely on the presence of a domain can lead to neglect of these prominent vectors of misinformation.[22] A text-as-data approach could help to fill

[22] A search of Google Scholar for "misinformation Facebook" or "misinformation Twitter" yields nearly double the amount of results as "misinformation YouTube" and nearly 13 times the number of results as "misinformation podcasts," which in the United States reach a monthly audience on par with Instagram.

this research void and expand to other prominent parts of the information ecosystem where domain-based content is absent.

Our findings suggest several future avenues for research. From a technical perspective, rapid advances in natural language processing are opening up avenues for better identification of false claims without relying on domains as a proxy for misinformation. Yet, these approaches will likely still be dependent on external fact checks and require some level of human oversight, particularly as models remain error prone. From a substantive perspective, although this paper focuses on political leaders, the text-based approach can be easily extended to ordinary users. Yet, we believe it to be especially important to examine the consequences of misinformation when shared or endorsed by political leaders, as they have greater potential to influence citizens' opinions (Flynn et al., 2017; Mosleh and Rand, 2021). The abundance of data on political elites may make the study of their behavior particularly promising, and detecting and predicting misinformation-sharing behavior is the first step in examining whether or not politicians' endorsements of misinformation exacerbate polarization, unrest, and beliefs in falsehoods. Finally, research should also move towards better understanding the consequences of misinformation by exploring whether people react differently to false or hyper-partisan content.

Acknowledgments

We thank audiences at University of Rochester and UNCC for their comments. We also thank Scott Abramson, Simon Chauchard, Emily Gade, Horacio Larreguy, Brendan Nyhan, and Tiago Ventura for excellent feedback. We are grateful for GDI for generously sharing their data with us, and for Boatos.org and aosfatos.org for answering our inquiries about their fact-checking process. We also benefited from Social Science One and CrowdTangle, which allowed us to access Facebook (Meta's) data via partnerships between industry and academia and to access to Twitter API when it was available to researchers without a charge. We also thank our research assistants (Leticia, Cecília, Eduarda, Gabriela, Thiago, Julio, Renata, Gendson, Debora, Caio, and Adriano) located in several places in Brazil who were essential in helping us to classify and categorize different parts of the data used in this project.

Funding Information

We thank Meta and FAPEMIG for funding this research.

References

Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives*, 31(2):211–235.

Allcott, H., Gentzkow, M., and Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2):2053168019848554.

Allen, J., Mobius, M., Rothschild, D. M., and Watts, D. J. (2021). Research note: Examining potential bias in large-scale censored data. *Harvard Kennedy School Misinformation Review*.

Bennett, L. W. and Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2):122–139.

Bovet, A. and Makse, H. A. (2019). Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 10(1):1–14.

Brennen, J. S., Simon, F. M., Howard, P. N., and Kleis, N. R. (2020). Types, sources, and claims of covid-19 misinformation.

Broockman, D. E. and Butler, D. M. (2017). The causal effects of elite position-taking on voter attitudes: Field experiments with elite communication. *American Journal of Political Science*, 61(1):208–221.

Carnes, N. and Lupu, N. (2015). Rethinking the comparative perspective on class and representation: Evidence from latin america. *American Journal of Political Science*, 59(1):1–18.

Einstein, K. L. and Glick, D. M. (2013). Scandals, conspiracies and the vicious cycle of cynicism. In *Annual Meeting of the American Political Science Association*.

Flynn, D., Nyhan, B., and Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38:127–150.

Gabel, M. and Scheve, K. (2007). Estimating the effect of elite communications on public opinion using instrumental variables. *American Journal of Political Science*, 51(4):1013–1028.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on twitter during the 2016 u.s. presidential election. *Science*, 363(6425):374–378.

Guess, A., Aslett, K., Tucker, J., Bonneau, R., and Nagler, J. (2021). Cracking open the news feed: Exploring what us facebook users see and share with large-scale platform data. *Journal of Quantitative Description: Digital Media*, 1.

Guess, A., Nagler, J., and Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on facebook. *Science Advances*, 5(1).

Guess, A. M., Lockett, D., Lyons, B., Montgomery, J., and Nyhan, B. (2020a). "fake news" may have limited effects beyond increasing beliefs in false claims. *Harvard Kennedy School (HKS) Misinformation Review*, 1(1):472–480.

Guess, A. M. and Lyons, B. A. (2020). Misinformation, disinformation, and online propaganda. *Social media and democracy: The state of the field, prospects for reform*, pages 10–33.

Guess, A. M., Nyhan, B., and Reifler, J. (2020b). Exposure to untrustworthy websites in the 2016 us election. *Nature human behaviour*, 4(5):472–480.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Imhoff, R., Dieterle, L., and Lamberty, P. (2021). Resolving the puzzle of conspiracy worldview and political activism: Belief in secret plots decreases normative but increases nonnormative political engagement. *Social Psychological and Personality Science*, 12(1):71–79.

Index, D. (2021). The online news market in brazil. https://www.disinformationindex.org/country-studies/2021-8-31-the-online-news-market-in-brazil/. Accessed on April 25, 2023.

Jolley, D. and Douglas, K. M. (2014). The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint. *British Journal of Psychology*, 105(1):35–56.

Kemp, S. (2021). Digital 2021: Global overview report - datareportal – global digital insights.

Lasser, J., Aroyehun, S. T., Simchon, A., Carrella, F., Garcia, D., and Lewandowsky, S. (2022). Social media sharing of low-quality news sources by political elites. *PNAS Nexus*, 1(4). pgac186.

Lee, S., Xiong, A., Seo, H., and Lee, D. (2023). "fact-checking" fact checkers: A data-driven approach. *Harvard Kennedy School Misinformation Review*.

Lenz, G. S. (2013). *Follow the leader?: how voters respond to politicians' policies and performance*. University of Chicago Press.

Lin, H., Lasser, J., Lewandowsky, S., Cole, R., Gully, A., Rand, D., and Pennycook, G. (2022). High level of agreement across different news domain quality ratings.

Machado, C., Kira, B., Narayanan, V., Kollanyi, B., and Howard, P. N. (2019). A study of misinformation in whatsapp groups with a focus on the brazilian presidential elections. In *Proceedings of The Web Conference*, WWW'19, San Francisco, USA.

McCoy, T. (2020). An investigation into fake news targets brazil's bolsonaros, and critics fear a constitutional crisis.

Mosleh, M., Martel, C., Eckles, D., and Rand, D. (2021). Perverse downstream consequences of debunking: Being corrected by another user for posting false political news increases subsequent sharing of low quality, partisan, and toxic content in a twitter field experiment. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Mosleh, M. and Rand, D. G. (2021). Falsehood in, falsehood out: Measuring exposure to elite misinformation on twitter.

Nyhan, B. (2018). *How Misinformation and Polarization Affect American Democracy*. Available at SSRN: https://ssrn.com/abstract=3144139.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American Political Science Review*, 115(3):999–1015.

Padmanabhan, D. (2021). *Data Science for Fake News*, chapter On Unsupervised Methods for Fake News Detection. Springer Nature Switzerland.

Palau, M. (2021). Inside brazil's dangerous battle over fake news.

Pasquetto, I., Jahani, E., Baranovsky, A., and Baum, M. (2020). Understanding misinformation on mobile instant messengers (mims) in developing countries.

Peng, Y., Lu, Y., and Shen, C. (2023). An agenda for studying credibility perceptions of visual misinformation. *Political Communication*, 40(2):225–237.

Pennycook, G. and Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7):2521–2526.

Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., and Benevenuto, F. (2019a). (mis) information dissemination in whatsapp: Gathering, analyzing and countermeasures. In *The World Wide Web Conference*, pages 818–828.

Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., and Benevenuto, F. (2019b). (mis)information dissemination in whatsapp: Gathering, analyzing and countermeasures. In *Proceedings of The Web Conference*, WWW'19.

Roberts, M. E., Stewart, B. M., and Tingley, D. (2019). Stm: An R package for structural topic models. *J. Stat. Softw.*, 91(2).

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., and Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American journal of political science*, 58(4):1064–1082.

Ross, R. M., Rand, D. G., and Pennycook, G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. *Judgment and Decision Making*, 16(2):484–504.

Rossini, P., Strommer-Galey, J., Baptista, E., and Oliveira, V. (2021). Dysfunctional information sharing on whatsapp and facebook: The role of political talk, cross-cutting exposure and social corrections. *New Media & Society*, 23(8):2430–2451.

Serrano, J. C. M., Papakyriakopoulos, O., and Hegelich, S. (2020). Nlp-based feature extraction for the detection of covid-19 misinformation videos on youtube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.

Stewart, A. J., Arechar, A. A., Rand, D. G., and Plotkin, J. B. (2021). The distorting effects of producer strategies: Why engagement does not reliably reveal consumer preferences for misinformation. *arXiv preprint arXiv:2108.13687*.

Tandoc Jr., E. C., Lim, Z. W., and Ling, R. (2018). Defining "fake news". *Digital Journalism*, 6(2):137–153.

Théro, H. and Vincent, E. M. (2022). Investigating facebook's interventions against accounts that repeatedly share misinformation. *Information Processing & Management*, 59(2):102804.

Vegetti, F. and Littvay, L. (2020). Belief in conspiracy theories, aggression, and attitudes towards political violence.

Wirtschafter, V. and Meserole, C. (2022). Prominent political podcasters played key role in spreading the 'big lie'.

Yang, Y., Davis, T., and Hindman, M. (2023). Visual misinformation on facebook. *Journal of Communication*, page jqac051.

Zucco Jr, C., Power, T. J., et al. (2021). Fragmentation without cleavages? endogenous fractionalization in the brazilian party system. *Comparative Politics*, 53(3):477–500.

Appendix

A.1 Data Sources

1. Checked Rumors/Stories: scraped by the authors
2. Facebook Fact-checked links: access via Social Science One partnership
3. Facebook and Instagram data: CrowdTangle
4. Twitter data: Twitter API (all shortened URLs were unshortened)
5. Ideology data: from Zucco and Power (2021)
6. Politicians' social media ids: directly collected by the authors (only verified or double checked ids were used)
7. Electoral information: from TSE and CEPESP

Data we are able to share and all code required to replicate analyses are available within the Harvard Dataverse Network, at https://doi.org/10.7910/DVN/EQL5E4.

A.2 Examples of detected posts

Figure 9. Posts identified via the text approach.

Figure 10. Posts identified via the domain approach.

A.3 Sources and Media of False Stories

Boatos indicates whether false stories came from social media (Facebook, Instagram, Twitter, YouTube), WhatsApp, traditional media/TV, and the internet – and residual "undetermined" category. These categories are not mutually exclusive and most have multiple sources (for example, social media and WhatsApp). Also, while we have 4,050 unique versions of stories from *Boatos*, we 3,847 have unique stories (several stories have more than one version, due to different wording, differences in details, etc.).

About 48% of stories came from social media, 24% from WhatsApp, and 38% from the internet – these are the largest categories. TV and traditional media account for less than 0.1% of stories. Regarding social media, the plurality of stories come from unspecified social media applications ("redes sociais"), and then the majority of false stories from specified social media applications comes from Facebook.

In terms of media, most stories (about 95%) come with some type of text, but several are accompanied by at least another type of media (many with several types), such as video (29%), image (24%), and audio (5%).

We took a random sample 30 stories from Boatos and included here print screens to show how these stories are shown in their website: https://drive.google.com/file/d/1I_LtsoSt-g8fR4aUy4F6tDzTZ5EBLXqa/view?usp=drive_link

Table 4. Types of Sources for False Stories.

	Percentages
Social Media	47.56
WhatsApp	23.75
Internet	37.90
TV	0.03
Traditional Media	0.03
Social Media (unspecified)	25.99
Facebook	20.82
Instagram	0.63
Twitter	1.50
YouTube	0.99
Total	3,847

Table 5. Types of Media for False Stories.

	Percentages
Text	95.99
Video	28.91
Image	24.15
Audio	5.18
Total	3,847

A.4 Text Pre-Processing

To process the text data from *Boatos* for the classification models, we cleaned the text by (1) removing URLs; (2) making all text lowercase; (3) removing non-alphanumeric symbols; (4) removing stopwords; (5) removing accented characters; (6) removing any single letter words; (7) trimming extra white spaces; (8) stemming words; and (9) removing observations with fewer than ten words. For politicians' social media posts, we first combined all text columns together (e.g., link text, title, image description, and message columns, where applicable) and then completed the same steps. We also ran a version of the classification models without cleaning the text, which performed equally well due to the signal provided by certain words being capitalized, abbreviated, etc.

After running the classification models to identify hyperpartisan posts amongst the politicians' social media content, we further processed the data to prepare it for the cosine similarity step. We vectorized the text data in two different ways: (1) using a Bag of Words approach, which weights all words in a document equally; and (2) using a Term Frequency-Inverse Document Frequency approach, which prioritizes and more heavily weights unusual words within posts and discounts words that may be common across posts. We reviewed cosine similarity matches for both approaches and found the matches to be more substantively aligned using the TF-IDF transformation, primarily due to the downweighting of more common words. As a result, we prioritized these matches for manual review.

Importantly, we did not remove duplicate posts across platforms (as politicians may, sometime, repeat the same post across platforms). We view each post as a unique piece of information in its own right (targeted, potentially, to different audiences), even if the message is repeated. As a result, we conduct most of our analyses at the post-level as opposed to the claim level. That said, in a review of duplicates across our dataset we found one post flagged by our text-based approach shared twice by the same politician on the same platform, one week apart. We also found 58 posts flagged by the domain-based approach shared at different points in time, sometimes by as many as four different politicians. Given our unit of analysis and the low frequency of duplicate posts using each misinformation identification strategy, we are less concerned about duplicates driving our findings.

A.5 Classification Models

In total, we ran four classification models to classify false content. The first model allowed us to get a baseline dataset of true social media posts from politicians, which we then used to identify only the most likely posts to contain false content across all three social media sites (425,949 posts total). Below we include a number of diagnostics for each of these models.

- **Confusion Matrix**: The confusion matrix tells us how well our model performed on our test set, given the training data. The matrix classifies True Negatives (top left quadrant), False Positives (Type I error, top right quadrant), False Negatives (Type II error, bottom left quadrant), and True Positive (bottom right quadrant).
- **ROC Curve**: The ROC curve, or the receiver operating characteristic curve, plots the true positive rate vs. the false positive rate, where the true positive rate represents the number of true positives divided by the sum of true positives and false negatives and the false positive rate represents the number of false positives divided by the sum of false positives and true negatives. The ROC curve tells us the ability of our model to differentiate between positive and negative classifications. The dashed line $y = x$ represents the reference line for a model that generates false positives and true positives at the same rate.
- **Precision-Recall Curve**: The precision-recall curve is another classification model diagnostic. Precision quantifies the ability of the model to correctly classify positives and is calculated as the number of True Positives divided by the sum of True and False positives. Recall quantifies the number positive classifications made, out of all possible positive classifications. It is calculated as the number of True Positives divided by the sum of True Positives and False Negatives. In both cases, a score of 1 represents perfect recall or perfect precision. Thus the precision-recall curve tells us how well our model performs at different thresholds against a no-skill classifier, represented by a dashed horizontal line.

Baseline Classification Model

The first model, fact checked stories vs. real stories, helped us get a baseline model to classify true content, which we then applied to out of sample posts from Facebook, Twitter and Instagram.

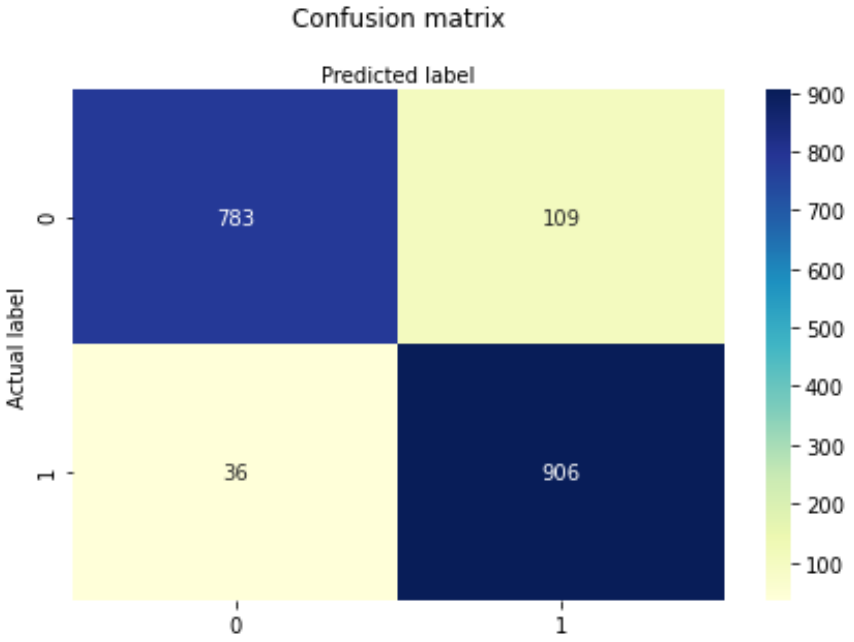


Figure 11. Confusion Matrix for Baseline Model.

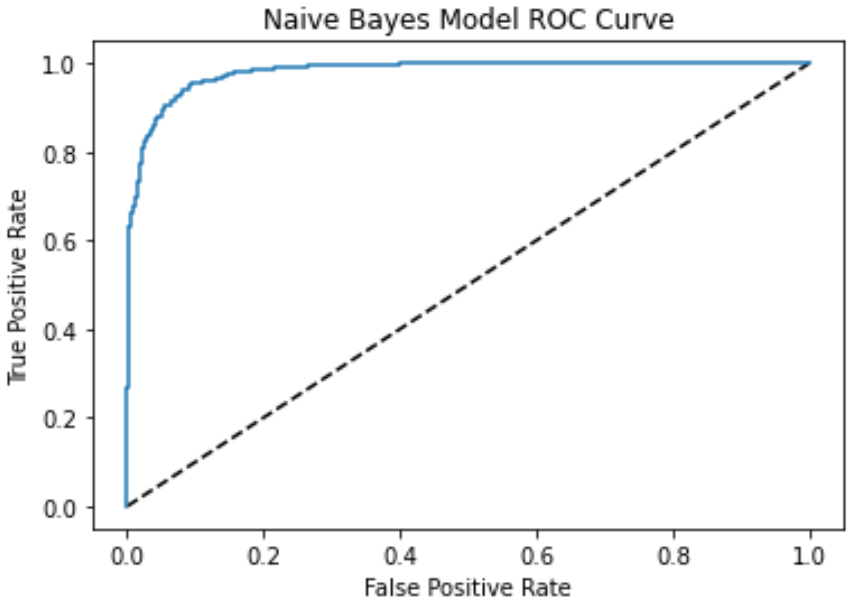


Figure 12. ROC for Baseline Model.

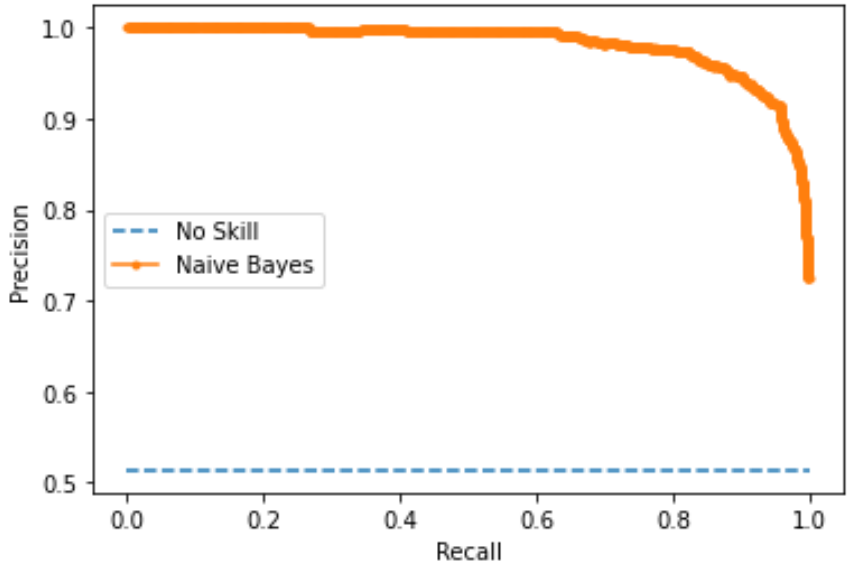


Figure 13. Precision Recall Curve for Baseline Model.

Facebook Classification Model

Using the first model, we classified all Facebook posts as to whether or not they were sharing false content. We then ran a classification model for fact checked stories vs. Facebook posts that the original model told us had a predicted probability of <.1 of being classified as false (or were very likely to be true).

Confusion Matrix

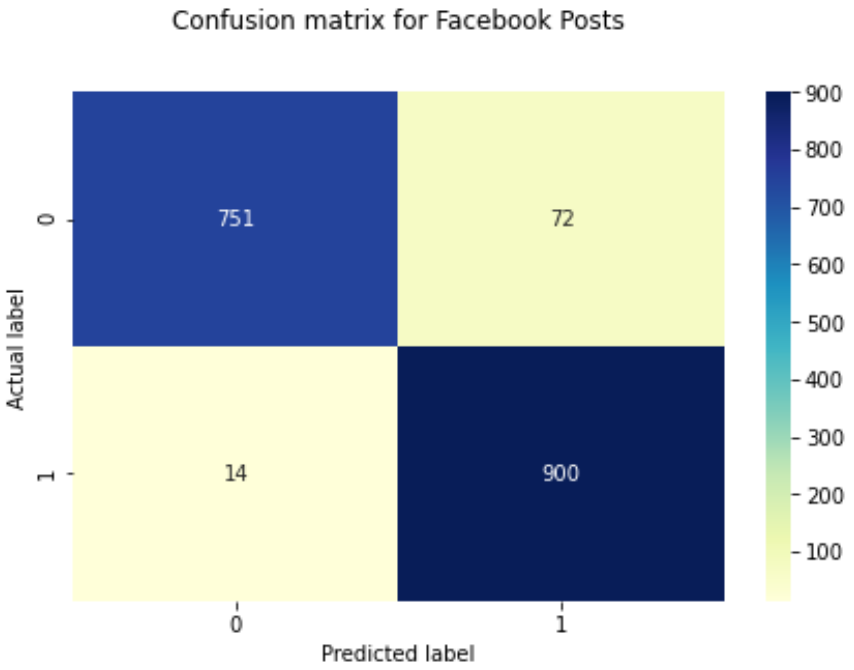


Figure 14. Confusion Matrix for Facebook Model.

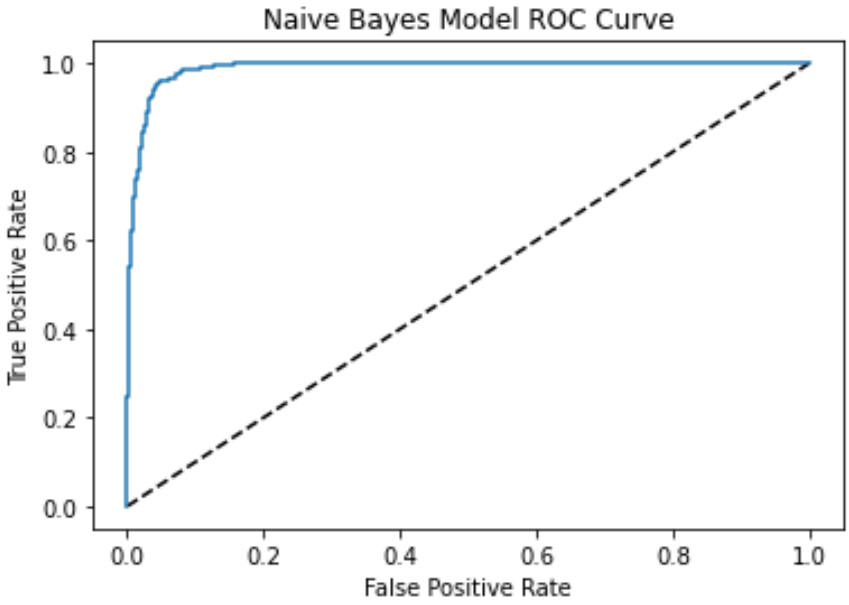


Figure 15. ROC for Facebook Model.

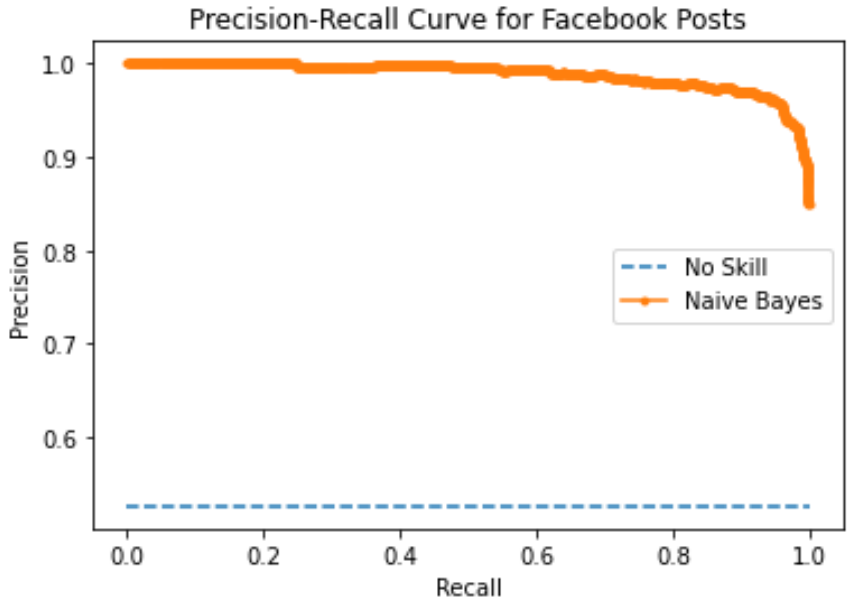


Figure 16. Precision-Recall Curve for Facebook Model.

Twitter Classification Model

Using the first model, we classified all Twitter posts as to whether or not they were sharing false content. We then ran a classification model for fact checked stories vs. Twitter posts that the original model told us had a predicted probability of <.1 of being classified as false (or were very likely to be true).

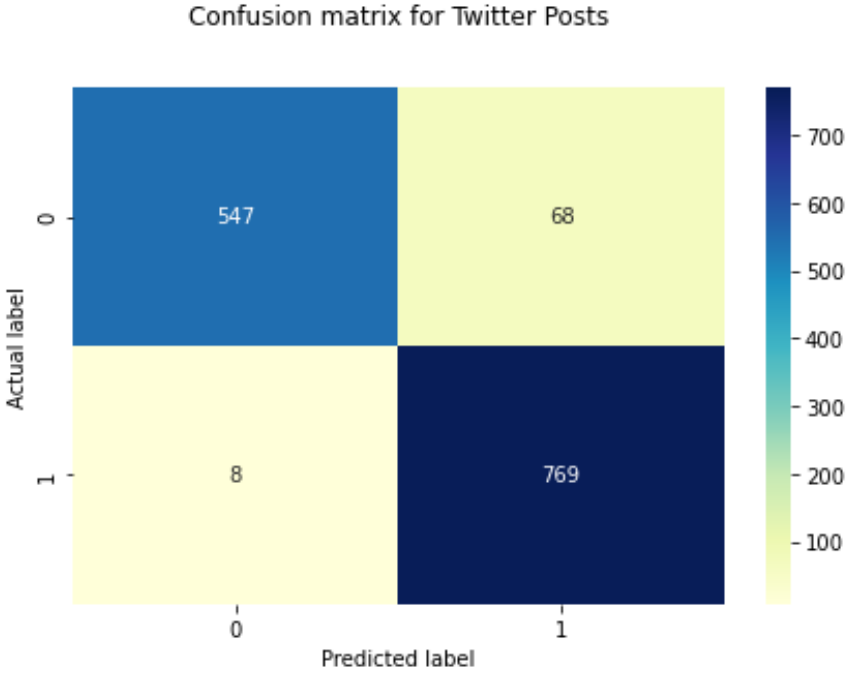


Figure 17. Confusion Matrix for Twitter Model.

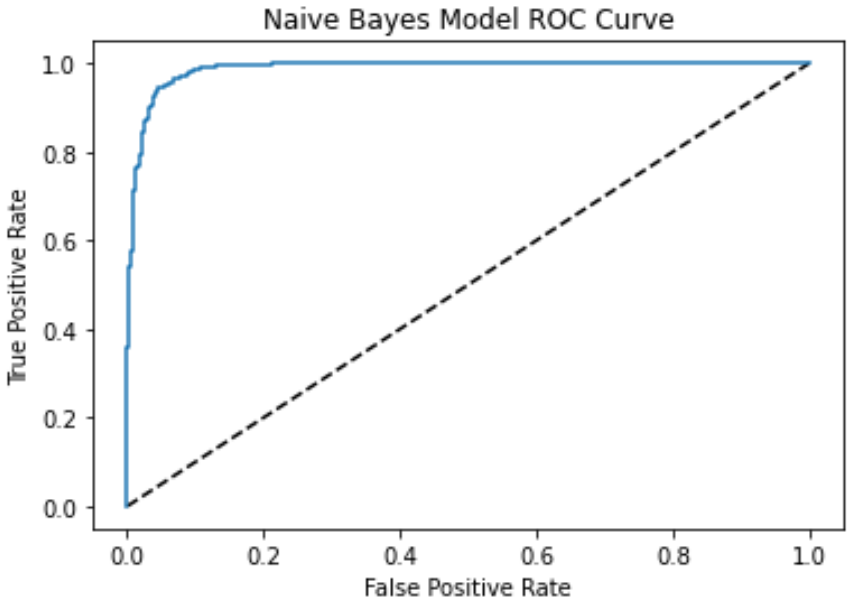


Figure 18. ROC for Twitter Model.

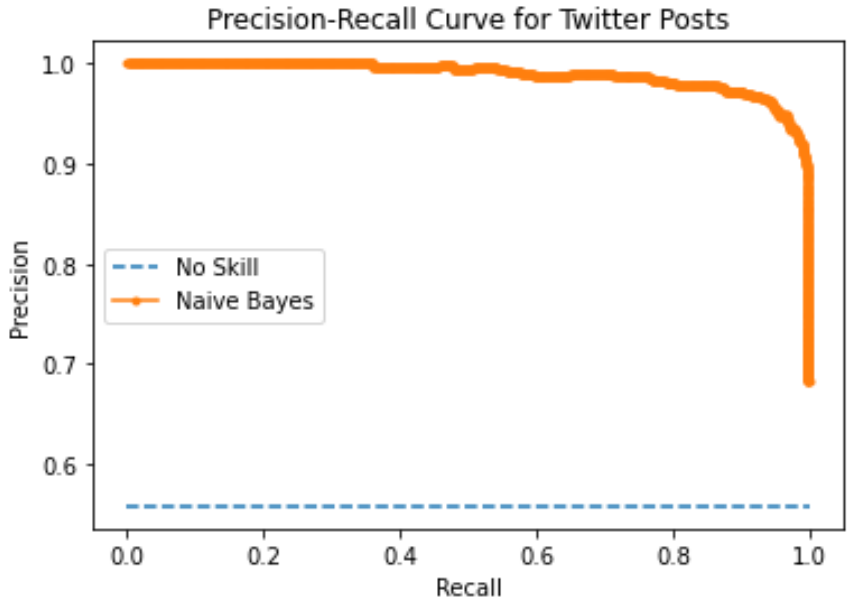


Figure 19. Precision-Recall Curve for Twitter Model.

Instagram Classification Model

Using the first model, we classified all Instagram posts as to whether or not they were sharing false content. We then ran a classification model for fact checked stories vs. Instagram posts that the original model told us had a predicted probability of <.1 of being classified as false (or were very likely to be true).

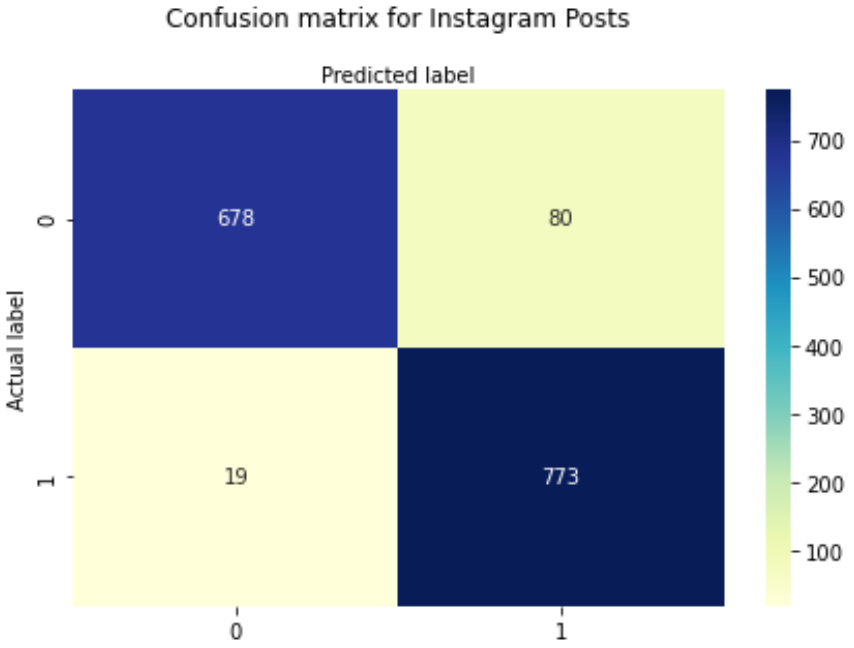


Figure 20. Confusion Matrix for Instagram Model.

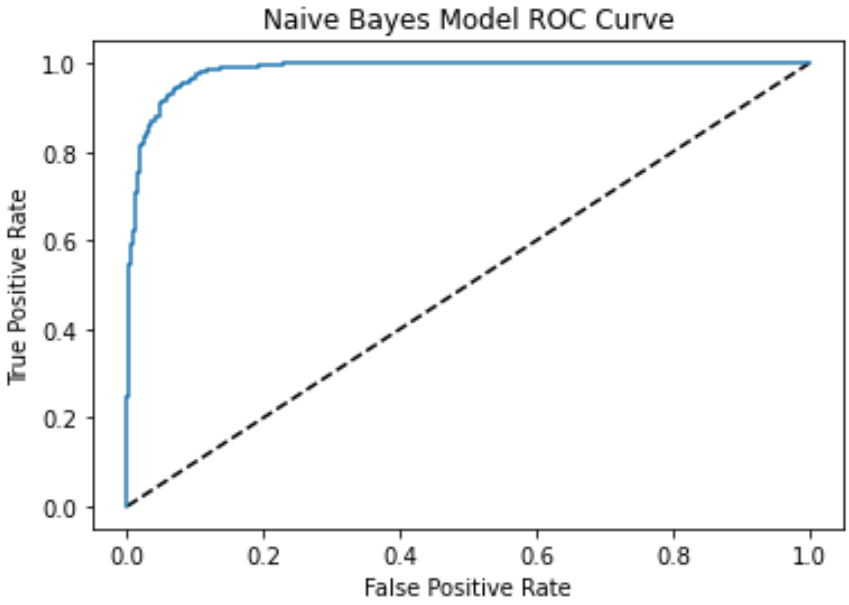


Figure 21. ROC for Instagram Model.

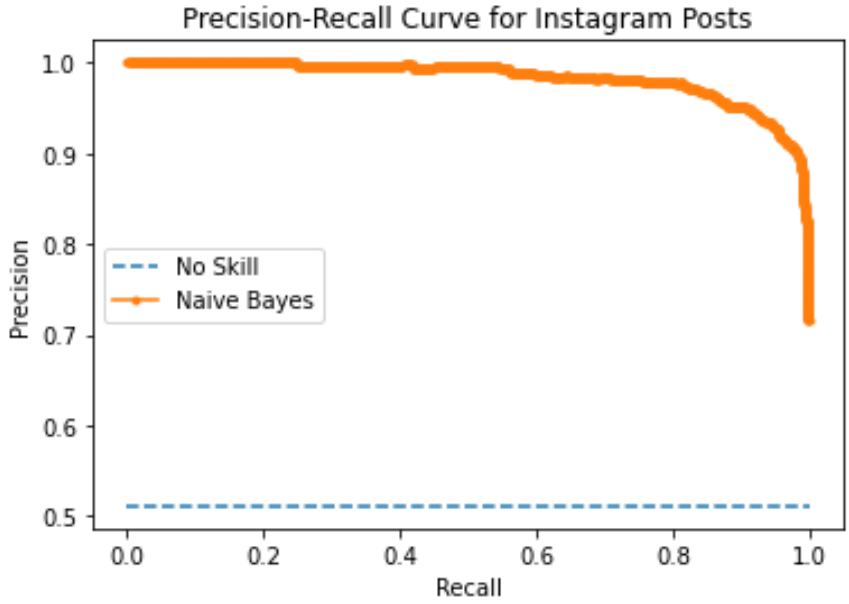


Figure 22. Precision-Recall Curve for Instagram Model.

A.6 Diagram of Text Approach

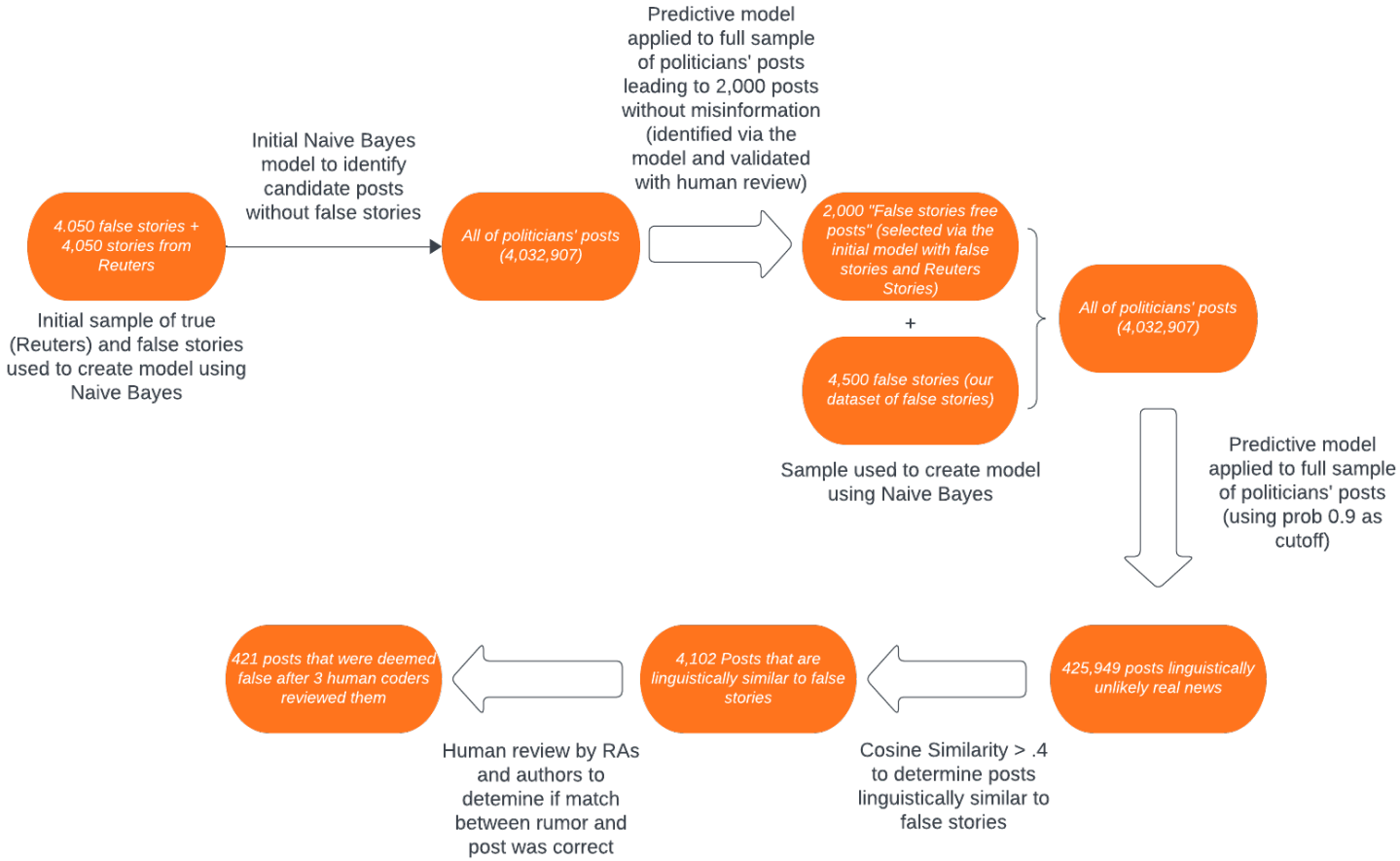


Figure 23. Diagram Illustrating the Text Approach Process.

A.7 Structural Topic Models

To identify the appropriate number of topics for both the text-based corpus and the domain-based corpus, we use the `searchK()` function in the `stm` package (Roberts et al., 2019). While there is no "correct" number of topics, we assess how the models perform at various topics thresholds ranging from 3 to 15, based on a number of common diagnostics. In assessing these diagnostics, we seek to minimize the residuals and maximize the held-out likelihood, while balancing both exclusivity of words to topics and semantic coherence. Figures 24 and 25 plot these diagnostics for our text-based approach. Based on these outputs, we select 11 topics for the text-based approach, as this number for K seems to have a good balance between average semantic coherence and exclusivity, have a high held-out likelihood value, and a low residuals value.

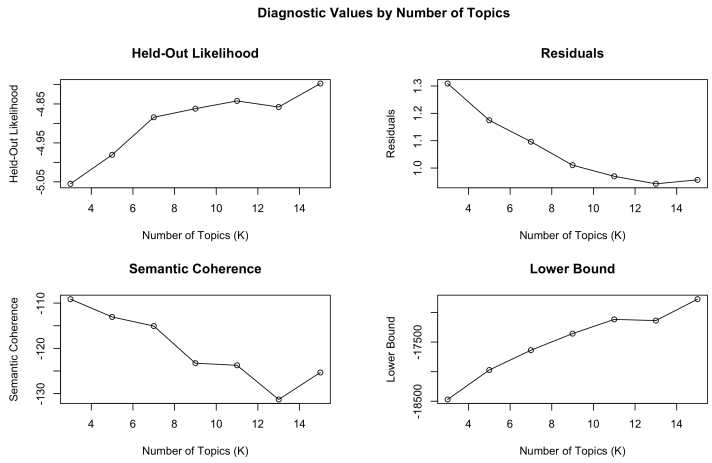


Figure 24. Diagnostics for Structural Topic Model for Posts Identified with the Text-Based Approach.

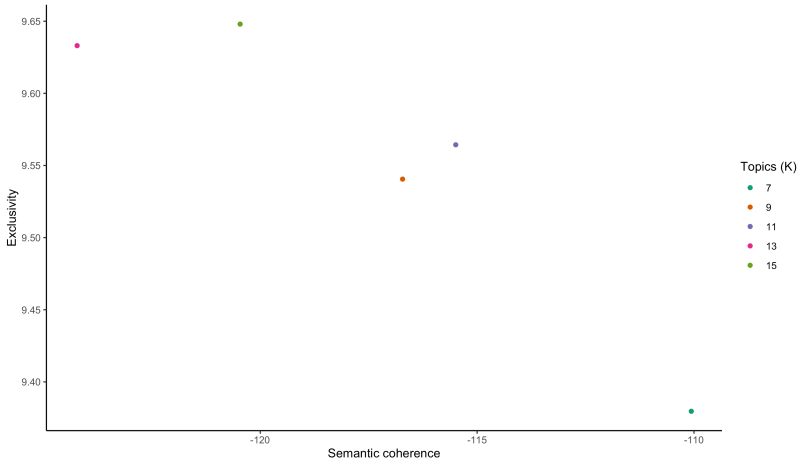


Figure 25. Comparing Average Exclusivity and Average Semantic Coherence at Different Ks for Posts Identified with the Text-Based Approach.

Figures 26 and 27 plot these diagnostics for the domain-based approach.

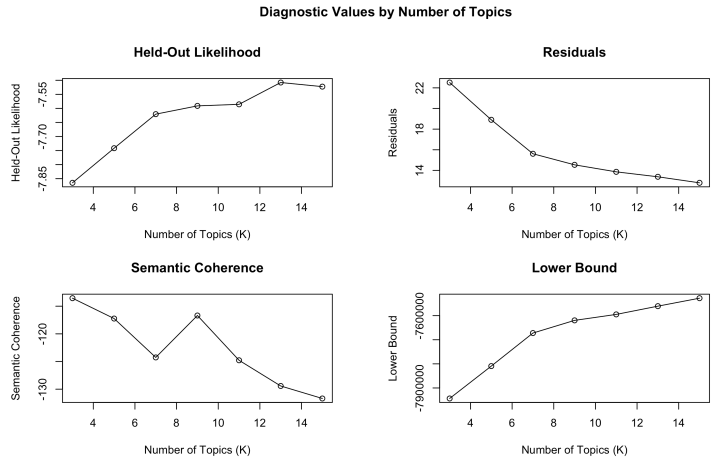


Figure 26. Diagnostics for Structural Topic Model for Posts Identified with the Domain-Based Approach.

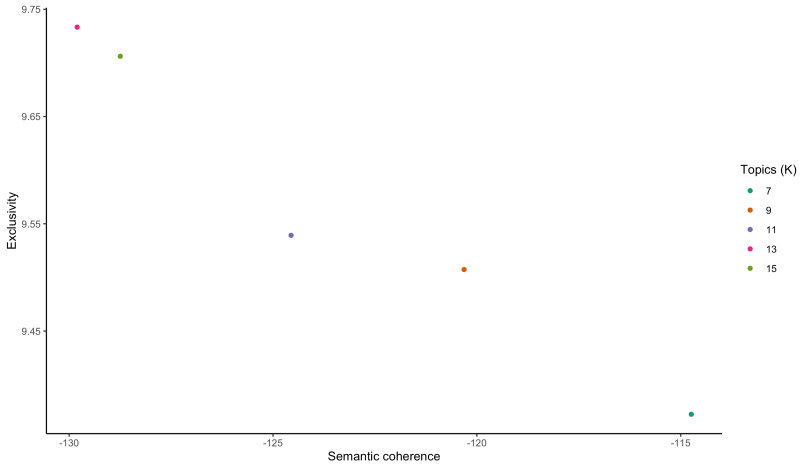


Figure 27. Comparing Average Exclusivity and Average Semantic Coherence at Different Ks for Posts Identified with the Domain-Based Approach.

Based on these outputs, we selected 9 topics for the domain-based approach for the same reasons as detailed above. We also include an overview of the top 10 frequent and exclusive words per topic for the text-based approach (Table 6) and the domain-based approach (Table 7).

Table 6. Top Frequent and Exclusive Words by Topic for Text-Based Detection Approach.

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
médico	rosário	esquerda	pai	afirmou	ond	covid	rio	coisa	adélio	condenar
outro	difamação	nacion	filha	exército	brasil	mort	militar	escola	sempr	estupro
font	figura	preso	quero	contra	embaixador	mil	paulo	ajudar	jair	culposo
hospit	sede	lula	filho	dose	região	após	educação	sobr	stf	vítima
nada	golp	voto	família	hora	sim	nunca	ministro	pobr	psol	mulher
todo	papa	assessor	casa	melhor	pra	globo	janeiro	campo	bolsonaro	estuprador
então	terço	qualquer	mim	uso	notícia	dia	maior	tão	lixo	justiça
saúd	francisco	vida	pode	enviou	apena	número	ainda	disso	governador	hoje
grand	mídia	sendo	dentro	vai	mundo	diz	vai	criança	deputado	toda
tempo	pública	brasileiro	sabe	maio	todo	pessoa	bandido	vamo	podemo	novo

Table 7. Top Frequent and Exclusive Words by Topic for Domain-Based Detection Approach.

V1	V2	V3	V4	V5	V6	V7	V8	V9
coronavírus	trf	verdad	amazônia	aposentadoria	wylli	aluno	senado	miliciano
covid	juiz	fake	nassif	propina	assessor	greve	pec	milícia
vacina	dallagnol	mentira	desemprego	trabalhista	queiroz	privatização	alcolumbr	resposta
pacient	deltan	bebianno	pobreza	serra	jean	correio	votação	dino
saúd	jato	ideologia	jornalggn	gleisi	preto	film	maia	argentina
pandemia	lula	lobo	desmatamento	hoffmann	fabrício	caminhoneiro	maduro	embaixada
cloroquina	lava	falar	ibop	previdência	facada	estat	ccj	eduardo
hidroxicloroquina	inácio	whatsapp	ibg	sindic	witzel	morador	kataguiri	celso
vírus	sergio	conservador	pib	bolívia	neymar	carro	foro	planalto
test	sítio	psicóloga	américa	reforma	maria	petrobrá	kim	macron

A.8 False Negatives in the Classification Model

Figure 28. Example of a potential false negative with a predicted probability between .5 and .9 of being false.

- Link to fact checking story relevant to this post: https://www.boatos.org/politica/pessoas-maiores-de-60-anos-votar-7h-e-10h-dia-15-11.html
- Link to true story explaining context: https://www1.folha.uol.com.br/poder/2020/09/entenda-como-sera-a-eleicao-durante-a-pandemia-do-novo-coronavirus.shtml

This is an example of a post that we classify as true because it had a predicted probability between .5 and .9 of being false so it was not further examined using cosine

similarity and human review. While one could make the point that the post is incorrect because it says that the time between 7 to 10 AM is "exceptionally" for individuals older than 60, the writing in the post is, in our assessment, more unclear than false. It is false that the period for 7 to 10 AM is reserved (or exclusive) to an older public. Rather anyone can vote in that period but individuals older than 60 have a priority line between 7 and 10 AM. Another point that it's ambiguous: the posts says to "remember to bring" a pen (in addition to a mask, ID, and a voting ID). The reality is that individuals are not required to bring a pen although it was recommended during elections that took place during Covid-19. The post does not make the statement that voting is only for people over 60 between 7 and 10 AM or that individuals are required to bring a pen, yet it's ambiguous (and badly written) enough that individuals could infer that. In summary, this is one example of the types of posts with a predicted probability between .5 and .9 of being false.

A.9 Online Engagement with Posts with and without Misinformation

Online engagement is measured via the sum of all types of user interactions across all platforms (all types of reactions on Facebook, including shares/comments, plus like/love on Instagram, and likes/re-tweet/quote/comment on Twitter). In Tables 8-11, we compare the average, median, and sum of engagement to posts with and without misinformation across the three main approaches (text, domain, Facebook URL).

Table 8. Posts with Misinformation (Text) and Online Engagement.

Misinfo Types	Median Likes	Avg. Likes	Median Reactions	Avg. Reactions	Median Views	Avg. Views
No text	79	1298.36	120	1795.64	0	3544.33
Text	596	4890.14	1182	8957.71	0	24535.56

Table 9. Posts with Misinformation (Domain) and Online Engagement.

Misinfo Types	Median Likes	Avg. Likes	Median Reactions	Avg. Reactions	Median Views	Avg. Views
No domain	78	1297.03	119	1788.95	0	3580.62
Domain	255	1473.95	537	2563.63	0	26.76

Table 10. Posts with Misinformation (FB URL) and Online Engagement.

Misinfo Types	Median Likes	Avg. Likes	Median Reactions	Avg. Reactions	Median Views	Avg. Views
No FBURL	79.00	1298.74	120	1796.39	0	3546.56
FBURL	78.50	707.50	205.50	1775.22	0	0

Table 11. Posts with Misinformation (any measure) and Online Engagement.

Misinfo Types	Median Likes	Avg. Likes	Median Reactions	Avg. Reactions	Median Views	Avg. Views
None	78	1296.66	119	1788.22	0	3578.39
Any	258	1510.22	542	2630.71	0	290.68

A.10 Measuring Ideology & Electoral Alignment

We measure electoral alignment based on the officials' partisan affiliation in the 2018 elections and their parties' membership in the official electoral coalitions. We create a three-level variable indicating Bolsonaro, Haddad, and other electoral alignment. Officials were part of Bolsonaro's electoral coalition if they belonged to either PSL or PRTB, to Haddad's coalition if they belonged to PT, PC do B, or PROS. All officials who belonged to any other party were classified in "other alignment" (our reference group in the regressions). For ministers who did not belong to any party at the time of the election, we assigned them as belonging to Bolsonaro's electoral coalitions.

We measure ideology based on Zucco Jr et al. (2021) Brazilian Legislative Survey 2017 round data. They develop estimates of the ideological positions of the main Brazilian political parties in a one-dimensional space using 20 questions regarding stated preferences from interviews with federal legislators in which higher values indicate "more conservative positions" on each questions. Their data includes 20 parties. Power and Zucco's measures have been widely used in the literature and we believe this is the best available measure for left-right

Yet, in Brazil's multi-partisan, fragmented, and somewhat volatile party system, there are challenges with any measure of ideology at the party level. First, not all parties in our sample of politicians is included in their dataset: the politicians in our sample belong to 37 parties and 20 parties are included in Power and Zucco. Notably, Bolsonaro's PSL is not included in their data because it only had one legislator in 2018, by the time the survey was conducted. Another reason why we have more parties in our sample than Power and Zucco is because we include politicians from different tiers of office (state, local, and federal) whereas Power and Zucco include only federal legislators. To deal with this issues of missing data, we made the following assumptions:

- PSL and NOVO are assigned the same score as DEM (the most "conservative" party in their dataset)
- UP, PSTU, PCO are assigned the same score as PSOL (the most "leftist" party in their dataset)

- Still, we are left with 12 parties (and 87 out of 945 politicians in our sample) without any measure of ideology (the parties are: Avante, Cidadania, DC, Patriota, PHS, PMB, PMN, PPL, PRB, PRP, PRTB, PTC).

In our main model specifications, we use Power and Zucco's continuous estimates of ideology (and inputted them for PSL/NOVO/UP/PSTU/PCO). We also constructed several categorical measures of ideology: split by the median level of ideology and by the 0.9 quantile (to create indicators of extremeness). All results are substantively similar to the one using Power and Zucco's continuous estimates.

Alternatively, we considered using individual-level measures of ideology using roll call votes, but issues of comparability across types of political offices are difficult since an important share of our sample of politicians does not cast votes in legislatures. Furthermore, even if all did, we would also be making difficult assumptions about our ability to measure ideology based on strategic votes in the legislature - see Carnes and Lupu (2015) for a discussion.

A.11 Tables for Figures 7 and 8

Table 12. Table for Figure 7 - Predictors of Sharing Misinformation (Binary) by Different Detection Approaches.

	Text-based (Binary)	Domain-based (Binary)
Sex (Male = 1)	−0.00	−0.01
	(0.03)	(0.04)
Age (30-44)	−0.03	−0.14
	(0.06)	(0.08)
Age (45-64)	−0.03	−0.12
	(0.05)	(0.08)
Age (65+)	0.03	−0.13
	(0.06)	(0.09)
Education (College =1)	−0.00	0.04
	(0.03)	(0.04)
Electoral alignment (Bolsonaro = 1)	0.30***	0.26***
	(0.06)	(0.05)
Electoral alignment (Haddad = 1)	0.12**	0.13**
	(0.04)	(0.05)
Ideology	−0.08	−0.18***
	(0.04)	(0.05)
Political Experience	−0.01	−0.03
	(0.03)	(0.04)
Number of posts (logged)	1.07***	2.01***
	(0.12)	(0.15)
R^2	0.22	0.29
Adj. R^2	0.20	0.28
Num. obs.	858	858
RMSE	0.32	0.42

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 13. Table for Figure 8 - Predictors of Sharing Misinformation (Count) by Different Detection Approaches.

	Text-based (Count)	Domain-based (Count)
(Intercept)	-15.25^{***}	-18.06^{***}
	(2.50)	(1.24)
Sex (Male = 1)	0.62	0.97^{***}
	(0.49)	(0.22)
Age (30-44)	-0.02	-1.21^{***}
	(0.76)	(0.30)
Age (45-64)	-0.00	-0.82^{**}
	(0.78)	(0.30)
Age (65+)	0.36	-0.41
	(0.94)	(0.37)
Education (College = 1)	-0.22	0.05
	(0.44)	(0.20)
Electoral alignment (Bolsonaro = 1)	0.99^{*}	-0.34
	(0.49)	(0.25)
Electoral alignment (Haddad = 1)	0.03	1.27^{***}
	(0.58)	(0.27)
Ideology	0.48	1.53^{***}
	(0.78)	(0.39)
Political Experience	-0.23	-0.30
	(0.42)	(0.23)
Number of posts (logged)	16.58^{***}	24.04^{***}
	(2.35)	(1.15)
Deviance	792.38	37085.28
Num. obs.	858	858

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

A.12 Additional Analyses for Operationalizations of Ideology

Table 14. Table for Figure 7 - Predictors of Sharing Misinformation (Binary) by Different Detection Approaches (ideology binary).

	Text-based (Binary)	Domain-based (Binary)
Sex (Male = 1)	−0.01	−0.03
	(0.03)	(0.04)
Age (30-44)	−0.04	−0.15
	(0.06)	(0.08)
Age (45-64)	−0.03	−0.13
	(0.06)	(0.08)
Age (65+)	0.03	−0.14
	(0.06)	(0.09)
Education (College =1)	−0.01	0.04
	(0.03)	(0.04)
Electoral alignment (Bolsonaro = 1)	0.29***	0.23***
	(0.06)	(0.05)
Electoral alignment (Haddad = 1)	0.14***	0.17***
	(0.04)	(0.04)
Ideology (Right-Wing = 1)	−0.02	−0.07
	(0.02)	(0.04)
Political Experience	−0.01	−0.02
	(0.03)	(0.04)
Number of posts (logged)	1.07***	2.00***
	(0.12)	(0.15)
R^2	0.22	0.29
Adj. R^2	0.20	0.27
Num. obs.	858	858
RMSE	0.32	0.43

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models are OLS models with robust standard errors and political office fixed effects.

Table 15. Table for Figure 7 - Predictors of Sharing Misinformation (Binary) by Different Detection Approaches (ideology extreme).

	Text-based (Binary)	Domain-based (Binary)
Sex (Male = 1)	−0.00	−0.02
	(0.03)	(0.04)
Age (30-44)	−0.03	−0.15
	(0.05)	(0.08)
Age (45-64)	−0.03	−0.13
	(0.05)	(0.08)
Age (65+)	0.03	−0.15
	(0.06)	(0.09)
Education (College = 1)	−0.01	0.04
	(0.03)	(0.04)
Electoral alignment (Bolsonaro = 1)	0.21***	0.12*
	(0.06)	(0.06)
Electoral alignment (Haddad = 1)	0.10*	0.15**
	(0.04)	(0.05)
Ideology (Not extreme = 1)	−0.10***	−0.11**
	(0.03)	(0.04)
Political Experience	−0.00	−0.01
	(0.03)	(0.04)
Number of posts (logged)	1.05***	1.97***
	(0.12)	(0.15)
R^2	0.23	0.29
Adj. R^2	0.21	0.28
Num. obs.	858	858
RMSE	0.32	0.42

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models are OLS models with robust standard errors and political office fixed effects.

Table 16. Table for Figure 7 - Predictors of Sharing Misinformation (Binary) by Different Detection Approaches (ideology extreme, left and right).

	Text-based (Binary)	Domain-based (Binary)
Sex (Male = 1)	0.00	−0.01
	(0.03)	(0.04)
Age (30-44)	−0.02	−0.13
	(0.05)	(0.08)
Age (45-64)	−0.02	−0.12
	(0.05)	(0.08)
Age (65+)	0.03	−0.14
	(0.06)	(0.09)
Education (College = 1)	−0.00	0.04
	(0.03)	(0.04)
Electoral alignment (Bolsonaro = 1)	0.24***	0.20**
	(0.07)	(0.07)
Electoral alignment (Haddad = 1)	0.06	0.08
	(0.04)	(0.05)
Ideology (Extreme Right = 1)	−0.08	−0.18*
	(0.06)	(0.07)
Ideology (Not Extreme = 1)	−0.14***	−0.19***
	(0.04)	(0.04)
Political Experience	−0.01	−0.02
	(0.03)	(0.04)
Number of posts (logged)	1.06***	1.99***
	(0.12)	(0.15)
R^2	0.23	0.30
Adj. R^2	0.21	0.28
Num. obs.	858	858
RMSE	0.32	0.42

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models are OLS models with robust standard errors and political office fixed effects.

A.13 Identification of URLs ("repeat offenders")

Overlap between "Repeat Offenders" and Text-Based Approach

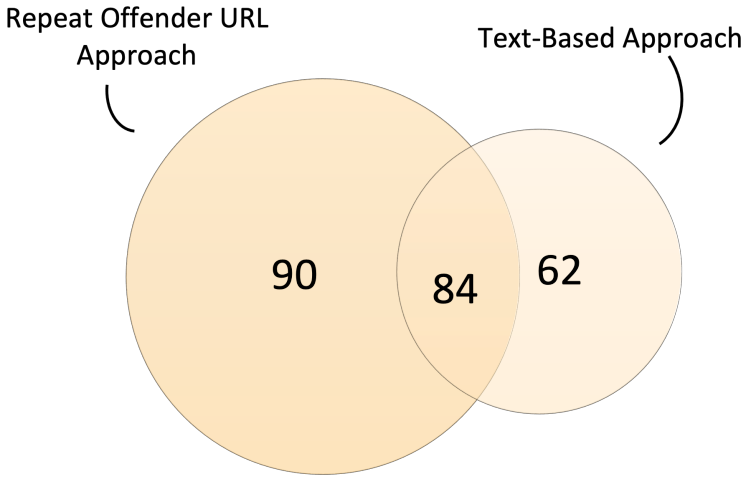


Figure 29. Number of politicians flagged as sharing misinformation using the "Repeat Offenders" URL approach, the Text-Based approach and both approaches.

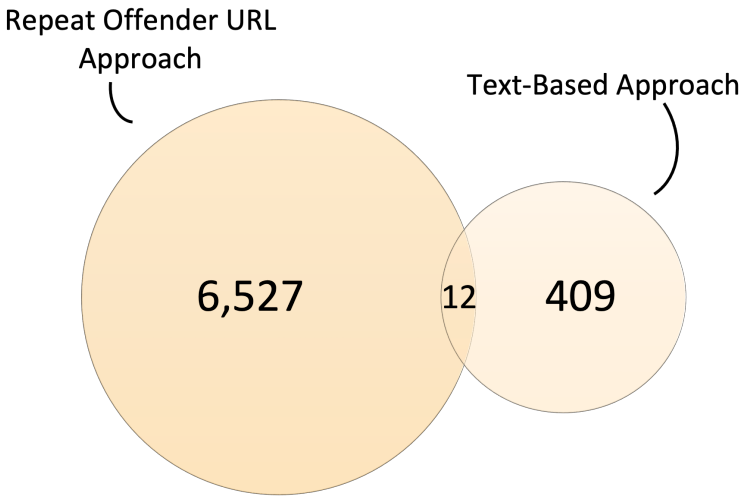


Figure 30. Number of posts flagged as sharing misinformation using the "Repeat Offenders" URL approach, the Text-Based approach and both approaches.

Results with repeat offenders and Global Disinformation Index (GDI) list

Table 17. Table for Figure 7 - Predictors of Sharing Misinformation (Binary) by Different Detection Approaches.

	Text-based	Domain-based	Domain-based (repeated offenders)	GDI
Sex (Male = 1)	−0.00	−0.01	0.00	−0.04
	(0.03)	(0.04)	(0.03)	(0.04)
Age (30-44)	−0.03	−0.14	0.05	0.03
	(0.06)	(0.08)	(0.05)	(0.07)
Age (45-64)	−0.03	−0.12	0.06	0.03
	(0.05)	(0.08)	(0.05)	(0.07)
Age (65+)	0.03	−0.13	0.04	0.02
	(0.06)	(0.09)	(0.06)	(0.08)
Education (College = 1)	−0.00	0.04	−0.07*	0.06
	(0.03)	(0.04)	(0.03)	(0.04)
Electoral alignment (Bolsonaro = 1)	0.30***	0.26***	0.39***	0.28***
	(0.06)	(0.05)	(0.06)	(0.05)
Electoral alignment (Haddad = 1)	0.12**	0.13**	0.07	0.08
	(0.04)	(0.05)	(0.04)	(0.04)
Ideology	−0.08	−0.18***	0.03	−0.30***
	(0.04)	(0.05)	(0.04)	(0.05)
Political Experience	−0.01	−0.03	−0.02	−0.12**
	(0.03)	(0.04)	(0.03)	(0.04)
Number of posts (logged)	1.07***	2.01***	1.16***	1.87***
	(0.12)	(0.15)	(0.13)	(0.14)
R^2	0.22	0.29	0.24	0.32
Adj. R^2	0.20	0.28	0.23	0.31
Num. obs.	858	858	858	858
RMSE	0.32	0.42	0.34	0.42

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models are OLS models with robust standard errors and political office fixed effects.

Table 18. Table for Figure 8 - Predictors of Sharing Misinformation (Count) by Different Detection Approaches.

	Text-based	Domain-based	Domain-based (repeated offenders)	GDI
(Intercept)	-15.25^{***}	-18.06^{***}	-33.46^{***}	-16.40^{***}
	(2.50)	(1.24)	(5.73)	(0.96)
Sex (Male = 1)	0.62	0.97^{***}	2.42	0.87^{***}
	(0.49)	(0.22)	(1.26)	(0.15)
Age (30-44)	−0.02	-1.21^{***}	2.63	0.35
	(0.76)	(0.30)	(1.89)	(0.40)
Age (45-64)	−0.00	-0.82^{**}	3.60	0.63
	(0.78)	(0.30)	(1.92)	(0.40)
Age (65+)	0.36	−0.41	2.92	1.12^{**}
	(0.94)	(0.37)	(2.33)	(0.42)
Education (College = 1)	−0.22	0.05	0.62	0.33^{*}
	(0.44)	(0.20)	(0.82)	(0.17)
Electoral alignment (Bolsonaro = 1)	0.99^{*}	−0.34	−0.29	0.63^{**}
	(0.49)	(0.25)	(0.65)	(0.23)
Electoral alignment (Haddad = 1)	0.03	1.27^{***}	−2.06	1.04^{***}
	(0.58)	(0.27)	(1.51)	(0.16)
Ideology	0.48	1.53^{***}	2.84	−0.31
	(0.78)	(0.39)	(1.48)	(0.27)
Political Experience	−0.23	−0.30	−0.60	−0.30
	(0.42)	(0.23)	(0.60)	(0.19)
Number of posts (logged)	16.58^{***}	24.04^{***}	31.33^{***}	21.25^{***}
	(2.35)	(1.15)	(4.65)	(0.87)
Deviance	792.38	37085.28	17417.97	32426.96
Num. obs.	858	858	858	858

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models are OLS models with robust standard errors and political office fixed effects.

A.14 Additional Binary Outcome Model Specifications

Text-based

Table 19. Additional model specifications for text-based binary outcome compared to the OLS model with robust standard errors.

	OLS Model	Logit Model	Probit Model
Sex (Male = 1)	−0.00	0.07	0.16
	(0.03)	(0.17)	(0.32)
Age (30-44)	−0.03	−0.02	−0.16
	(0.06)	(0.41)	(0.76)
Age (45-64)	−0.03	0.05	−0.04
	(0.05)	(0.41)	(0.76)
Age (65+)	0.03	0.48	0.73
	(0.06)	(0.44)	(0.82)
Education (College = 1)	−0.00	−0.05	−0.03
	(0.03)	(0.17)	(0.32)
Electoral alignment (Bolsonaro = 1)	0.30***	1.02***	1.78***
	(0.06)	(0.21)	(0.38)
Electoral alignment (Haddad = 1)	0.12**	0.49*	0.96**
	(0.04)	(0.20)	(0.36)
Ideology	−0.08	0.07	0.43
	(0.04)	(0.28)	(0.52)
Political Experience	−0.01	−0.13	−0.20
	(0.03)	(0.17)	(0.32)
Number of posts (logged)	1.07***	7.98***	15.45***
	(0.12)	(0.88)	(1.74)
R^2	0.22		
Adj. R^2	0.20		
Num. obs.	858	858	858
RMSE	0.32		
AIC		551.32	546.24
BIC		632.14	627.07
Log Likelihood		−258.66	−256.12
Deviance		517.32	512.24

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models use political office fixed effects.

Domain-based

Table 20. Additional model specifications for domain-based binary outcome compared to the OLS model with robust standard errors.

	OLS Model	Logit Model	Probit Model
Sex (Male = 1)	−0.01	−0.01	−0.03
	(0.04)	(0.14)	(0.24)
Age (30-44)	−0.14	−0.60	−0.99
	(0.08)	(0.31)	(0.53)
Age (45-64)	−0.12	−0.53	−0.86
	(0.08)	(0.31)	(0.52)
Age (65+)	−0.13	−0.57	−0.96
	(0.09)	(0.34)	(0.57)
Education (College = 1)	0.04	0.10	0.18
	(0.04)	(0.14)	(0.23)
Electoral alignment (Bolsonaro = 1)	0.26***	0.85***	1.43***
	(0.05)	(0.20)	(0.34)
Electoral alignment (Haddad = 1)	0.13**	0.54**	0.86**
	(0.05)	(0.18)	(0.32)
Ideology	−0.18***	−0.54**	−0.89*
	(0.05)	(0.20)	(0.35)
Political Experience	−0.03	−0.11	−0.19
	(0.04)	(0.14)	(0.23)
Number of posts (logged)	2.01***	7.89***	13.43***
	(0.15)	(0.63)	(1.14)
R^2	0.29		
Adj. R^2	0.28		
Num. obs.	858	858	858
RMSE	0.42		
AIC		904.19	905.18
BIC		985.02	986.01
Log Likelihood		−435.10	−435.59
Deviance		870.19	871.18

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models use political office fixed effects.

A.15 Additional Count Outcome Model Specifications

Text-based

Table 21. Additional model specifications for text-based count outcome compared to the quasi-poisson model.

	Quasi-poisson	Poisson	Neg. Binomial
(Intercept)	-15.25^{***}	-15.25^{***}	-13.35^{***}
	(2.50)	(0.85)	(1.38)
Sex (Male = 1)	0.62	0.62^{***}	0.51
	(0.49)	(0.17)	(0.28)
Age (30-44)	−0.02	−0.02	0.26
	(0.76)	(0.26)	(0.63)
Age (45-64)	−0.00	−0.00	0.36
	(0.78)	(0.27)	(0.63)
Age (65+)	0.36	0.36	0.82
	(0.94)	(0.32)	(0.69)
Education (College = 1)	−0.22	−0.22	−0.48
	(0.44)	(0.15)	(0.26)
Electoral alignment (Bolsonaro = 1)	0.99^{*}	0.99^{***}	1.31^{***}
	(0.49)	(0.17)	(0.32)
Electoral alignment (Haddad = 1)	0.03	0.03	−0.13
	(0.58)	(0.20)	(0.32)
Ideology	0.48	0.48	−0.29
	(0.78)	(0.26)	(0.42)
Political Experience	−0.23	−0.23	−0.22
	(0.42)	(0.14)	(0.27)
Number of posts (logged)	16.58^{***}	16.58^{***}	15.18^{***}
	(2.35)	(0.80)	(1.39)
AIC		1158.35	999.36
BIC		1239.18	1084.95
Log Likelihood		−562.17	−481.68
Deviance	792.38	792.38	379.27
Num. obs.	858	858	858

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models use political office fixed effects.

Domain-based

Table 22. Additional model specifications for domain-based count outcome compared to the quasi-poisson model.

	Quasi-poisson	Poisson	Neg. Binomial
(Intercept)	-18.06^{***}	-18.06^{***}	-11.34^{***}
	(1.24)	(0.11)	(0.81)
Sex (Male = 1)	0.97^{***}	0.97^{***}	0.65^{**}
	(0.22)	(0.02)	(0.20)
Age (30-44)	-1.21^{***}	-1.21^{***}	-0.70
	(0.30)	(0.03)	(0.44)
Age (45-64)	-0.82^{**}	-0.82^{***}	-0.51
	(0.30)	(0.03)	(0.43)
Age (65+)	-0.41	-0.41^{***}	-0.46
	(0.37)	(0.03)	(0.48)
Education (College = 1)	0.05	0.05^{**}	-0.54^{**}
	(0.20)	(0.02)	(0.19)
Electoral alignment (Bolsonaro = 1)	-0.34	-0.34^{***}	0.79^{**}
	(0.25)	(0.02)	(0.27)
Electoral alignment (Haddad = 1)	1.27^{***}	1.27^{***}	1.20^{***}
	(0.27)	(0.02)	(0.24)
Ideology	1.53^{***}	1.53^{***}	0.04
	(0.39)	(0.04)	(0.30)
Political Experience	-0.30	-0.30^{***}	-0.32
	(0.23)	(0.02)	(0.20)
Number of posts (logged)	24.04^{***}	24.04^{***}	17.84^{***}
	(1.15)	(0.10)	(0.87)
AIC		38877.15	4298.47
BIC		38957.98	4384.06
Log Likelihood		-19421.58	-2131.24
Deviance	37085.28	37085.28	779.78
Num. obs.	858	858	858

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. All models use political office fixed effects.

Two-Step Heckman Model

Table 23. Two-Step Heckman Model with text and domain outcomes.

	Text-based	Domain-based
O: Intercept	−113.06*	−3573.40***
	(51.70)	(774.74)
O: Sex (Male = 1)	2.35	81.63
	(2.15)	(71.47)
O: Age (30-44)	−0.80	−677.12***
	(5.23)	(166.89)
O: Age (45-64)	−0.38	−609.05***
	(5.26)	(161.66)
O: Age (65+)	4.57	−598.07***
	(6.48)	(178.06)
O: Education (College = 1)	−1.17	−8.41
	(2.08)	(73.13)
O: Electoral alignment (Bolsonaro = 1)	12.00*	249.45*
	(6.06)	(121.06)
O: Electoral alignment (Haddad = 1)	4.21	281.76**
	(3.60)	(96.42)
O: Ideology	1.27	−38.36
	(3.19)	(113.78)
O: Political Experience	−1.64	−110.03
	(2.18)	(71.18)
O: Number of Posts (logged)	112.73*	4733.21***
	(48.75)	(905.54)
Inverse Mills Ratio	16.42*	874.40***
	(7.43)	(195.93)
sigma	13.82	680.71
rho	1.19	1.28
Num. obs.	857	857
Censored	729	419
Observed	128	438

Note. $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Only outcome equation shown. President and vice-president excluded.

A.16 Cross Tabs for Binary Outcomes by Predictor

Table 24. False post (text) by politician's age.

	Not False (Text)	False (Text)	Sum
22-29	28(84.8%)	5(15.2%)	33
30-44	237(84.6%)	43(15.4%)	280
45-64	436(84.5%)	80(15.5%)	516
65+	98(84.5%)	18(15.5%)	116
Sum	799	146	945

Table 25. False post (domain) by politician's age.

	Not False (Domain)	False (Domain)	Sum
22-29	16(48.5%)	17(51.5%)	33
30-44	136(48.6%)	144(51.4%)	280
45-64	252(48.8%)	264(51.2%)	516
65+	63(54.3%)	53(45.7%)	116
Sum	467	478	945

Table 26. False post (text) by politician's coalition.

	Not False (Text)	False (Text)	Sum
Bolsonaro coalition	58(61.1%)	37(38.9%)	95
Haddad coalition	87(68%)	41(32%)	128
Other	654(90.6%)	68(9.4%)	722
Sum	799	146	945

Table 27. False post (domain) by politician's coalition.

	Not False (Domain)	False (Domain)	Sum
Bolsonaro coalition	32(33.7%)	63(66.3%)	95
Haddad Coalition	28(21.9%)	100(78.1%)	128
Other	407(56.4%)	315(43.6%)	722
Sum	467	478	945

Table 28. False post (text) by politician's education.

	Not False (Text)	False (Text)	Sum
No higher educ.	137(84%)	26(16%)	163
Higher education	662(84.7%)	120(15.3%)	782
Sum	799	146	945

Table 29. False post (domain) by politician's education.

	Not False (Domain)	False (Domain)	Sum
No higher educ.	90(55.2%)	73(44.8%)	163
Higher education	377(48.2%)	405(51.8%)	782
Sum	467	478	945

Table 30. False post (text) by politician's experience.

	Not False (Text)	False (Text)	Sum
Lower Experience	159(81.5%)	36(18.5%)	195
Higher experience	640(85.3%)	110(14.7%)	750
Sum	799	146	945

Table 31. False post (domain) by politician's experience.

	Not False (Domain)	False (Domain)	Sum
Lower Experience	92(47.2%)	103(52.8%)	195
Higher experience	375(50%)	375(50%)	750
Sum	467	478	945

Table 32. False post (text) by politician's ideology.

	Not False (Text)	False (Text)	Sum
Left	317(83.2%)	64(16.8%)	381
Right	412(86.4%)	65(13.6%)	477
Sum	729	129	858

Table 33. False post (domain) by politician's ideology.

	Not False (Domain)	False (Domain)	Sum
Left	165(43.3%)	216(56.7%)	381
Right	254(53.2%)	223(46.8%)	477
Sum	419	439	858

Table 34. False post (text) by politician's office.

	Not False (Text)	False (Text)	Sum
Mayoral candidate	226(86.3%)	36(13.7%)	262
Federal Deputy	430(82.2%)	93(17.8%)	523
Governor	26(96.3%)	1(3.7%)	27
Minister	20(87%)	3(13%)	23
President	0(0%)	1(100%)	1
Senator	71(87.7%)	10(12.3%)	81
Vice-governor	25(92.6%)	2(7.4%)	27
Vice-president	1(100%)	0(0%)	1
Sum	799	146	945

Table 35. False post (domain) by politician's office.

	Not False (Domain)	False (Domain)	Sum
Mayoral candidate	144(55%)	118(45%)	262
Federal Deputy	249(47.6%)	274(52.4%)	523
Governor	19(70.4%)	8(29.6%)	27
Minister	9(39.1%)	14(60.9%)	23
President	0(0%)	1(100%)	1
Senator	27(33.3%)	54(66.7%)	81
Vice-governor	18(66.7%)	9(33.3%)	27
Vice-president	1(100%)	0(0%)	1
Sum	467	478	945

Table 36. False post (text) by politician's sex.

	Not False (Text)	False (Text)	Sum
Female	122(82.4%)	26(17.6%)	148
Male	677(84.9%)	120(15.1%)	797
Sum	799	146	945

Table 37. False post (domain) by politician's sex.

	Not False (Domain)	False (Domain)	Sum
Female	66(44.6%)	82(55.4%)	148
Male	401(50.3%)	396(49.7%)	797
Sum	467	478	945

A.17 Means for Count Outcomes by Predictor

Table 38. Average false posts (text and domain) by age.

	22-29 (N=33)		30-44 (N=280)		45-64 (N=516)		65+ (N=116)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	0.8	2.3	0.5	2.1	0.4	1.8	0.3	0.9
False stories (domain)	202.1	1052.4	33.5	257.6	38.0	199.7	26.2	117.6

Table 39. Average false posts (text and domain) by coalition.

	Bolsonaro (N=95)		Haddad (N=128)		Other (N=722)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	1.4	3.6	0.6	1.2	0.3	1.5
False stories (domain)	38.7	152.7	117.2	316.3	27.7	291.4

Table 40. Average false posts (text and domain) by education.

	No higher		Higher educ.	
	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	0.5	1.5	0.4	1.9
False stories (domain)	58.5	478.6	37.3	225.8

Table 41. Average false posts (text and domain) by experience.

	Lower Exp.		Higher Exp.	
	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	0.7	2.9	0.4	1.4
False stories (domain)	67.5	526.2	34.0	175.5

Table 42. Average false posts (text and domain) by ideology.

	Left (N=381)		Right (N=477)	
	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	0.4	1.1	0.4	1.9
False stories (domain)	51.0	229.9	27.6	287.6

Table 43. Average false posts (text and domain) by sex.

	Female (N=148)		Male (N=797)	
	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	0.6	2.6	0.4	1.7
False stories (domain)	30.7	134.2	42.9	305.5

Table 44. Average false posts (text and domain) by office.

	Mayor (N=262)		Fed. Dep. (N=523)		Gov. (N=27)		Min. (N=23)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	0.3	1.3	0.6	2.2	0.0	0.2	0.2	0.6
False stories (domain)	28.2	259.6	52.9	331.1	11.2	38.6	8.4	27.0

	Pres. (N=1)		Sen. (N=81)		Vice-gov. (N=27)		Vice-pres. (N=1)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
False stories (text)	4.0	-	0.4	1.4	0.1	0.4	0.0	-
False stories (domain)	35.0	-	37.8	153.6	2.5	5.6	0.0	-