Methodology

Data collection

This Pew Research Center analysis examines all English-language tweets posted between Jan. 1 and Oct. 24, 2022, by accounts affiliated with candidates running for office in 2022 at the federal, state and local level. Researchers obtained from Ballotpedia a list of candidates running for office in 2022 which included demographic characteristics (party affiliation, district, etc.) and any affiliated Twitter handles. This list included known personal or campaign Twitter handles for all 2022 U.S. congressional candidates; all 2022 executive, legislative and judicial statewide candidates; and all 2022 municipal candidates in the 100 most populous U.S. cities, according to the 2010 decennial census. As this analysis was interested in candidates, any incumbents who were not up for reelection in 2022 are not included.

Next, researchers collected tweets shared from Jan. 1 to Oct. 24 from all public and valid (as of Aug. 29, 2022) Twitter handles listed in the dataset using the Twitter Search API. The collected tweets were subsequently filtered to those written in English using the Twitter API's flag for language, resulting in a total of nearly 3.4 million tweets from 7,720 candidates which were used to generate the findings in the analysis.

Measurement

Researchers used several methods to label each tweet included in the analysis for the following characteristics:

- Mention or discussion of 16 substantive topics
- Sentiment of text (positive, negative, neutral)
- Mention of specific groups or figures:
    - major political parties (Democrat, Republican)
    - national party figures (members of the 117th congress)
    - President Joe Biden
    - former President Donald Trump
    - Candidate's opponent in race

Identifying the substantive topic areas in candidates' tweets

To identify the substantive topics mentioned in candidate's tweets, researchers first developed a list of substantive topics that candidates were likely to mention in their tweets. In order to code each of the nearly 3.4 million tweets, researchers used a series of machine learning models. This process is described below.

First, researchers pulled a random sample of 250 tweets from a similar population (tweets from seated members of the 117th Congress posted in 2022), labeled their topical content manually, and developed a codebook of common substantive topic areas, based in part on the issue areas the Center has been included on its regular issue surveys of registered voters.

Next, researchers developed a suite of machine learning classifiers to automatically identify these topics in unseen tweets. This task, used a transfer learning approach, in which models initially trained on natural language inference (NLI) were adapted to classification on unseen texts and categories (an approach known as zero-shot classification).[1] We used bart-large-mnli, a large Transformer-based language model as our topic classifier, and the model was configured to make predictions for each topic independently so that any given tweet could have multiple topic labels applied to it if appropriate.

To develop prompts for the model and validate its predictions for each of the substantive topic areas, two researchers iterated on the following steps:

- Apply classifier to a sample of tweets
- Manually review the predicted labels in the sample for each topic separately
- Adjust the model prompt and classification probability threshold as needed to improve predictions

This process was repeated for each topic until the zero-shot model's predictions reached acceptable levels of performance (per-topic accuracy of 95% or better and both F1 score and Cohen's Kappa statistics of 0.75 or better). To effectively evaluate the model's ability to predict low-prevalence topics, performance metrics were calculated on stratified samples for each topic, then weighted according to each topic's prevalence.

As a final check, a third researcher manually reviewed a random sample of labeled tweets using the finalized models, giving a binary response for whether the model's predictions were appropriate. Overall accuracy across all substantive topics was 96%.

The final list of topics included in this analysis, along with the model prompt and performance metrics are listed in the table below.

[1] NLI models usually perform the task of "logical entailment," where they are presented with a pair of documents – a "premise" and a "hypothesis" – and asked to determine whether or not the hypothesis follows logically from the information included in the premise text. NLI models can be adapted to serve as zero-shot topic classifiers by reframing the labels as a series of "true or false" questions to be applied to the text. For example, if the model is shown a premise such as "Joe Biden is the president of the United States" along with the hypothesis "This text is about politics," it will return a true or false value which indicates the topic classification. (See Yin, Hay, and Roth (2019) for more details.)

Predicting substantive topics

Topic name	Description	Model prompt used	Performance
Abortion	Tweets about abortion and the status of abortion rights. Can include expressions of "pro-life"/"pro-choice" stances, discussions of court decisions such as Roe v. Wade, Dobbs v. Jackson, organizations like Planned Parenthood.	*"abortion or pro-choice or pro-life"*	Accuracy: 0.991 Cohen's Kappa: 0.937 F1 Score: 0.943
Economy	Tweets about economics or the state of the economy. Can include discussions of inflation, consumer prices, economic policy, the stock market, or the Fed.	*"the economy"*	Accuracy: 0.956 Cohen's Kappa: 0.776 F1 Score: 0.800
Climate & environment	Tweets about climate change and the environment. Can include discussions about climate change as an issue, or references to the EPA, "Green New Deal," environmental policies or regulations, or fossil fuels.	*"climate change or environment"*	Accuracy: 0.985 Cohen's Kappa: 0.816 F1 Score: 0.824
Health care	Tweets about health care policy. Can include discussions of public health, health care costs, as well as legislation or proposed policies such as the Affordable Care Act, Medicare, "Medicare for all." Excludes references to the COVID-19 pandemic and Abortion as health care.	*"health care"*	Accuracy: 0.976 Cohen's Kappa: 0.840 F1 Score: 0.853
Race	Tweets about race and ethnicity and/or racism. Can also include references to concepts like critical race theory/CRT, or White supremacy, or movements like Black Lives Matter/BLM.	*"issues around race and ethnicity"*	Accuracy: 0.983 Cohen's Kappa: 0.786 F1 Score: 0.795
Education	Tweets about education/educational policy. Can include references to school boards, curriculum restrictions/book banning, charter schools/school choice, student loan forgiveness.	*"education or student policy"*	Accuracy: 0.994 Cohen's Kappa: 0.886 F1 Score: 0.889
Immigration	Tweets about immigration and immigration policy. Can include references to DACA/Dreamers, "open/closed borders," ICE, "build the wall."	*"immigration policy"*	Accuracy: 0.999 Cohen's Kappa: 0.985 F1 Score: 0.985
Guns or gun policy	Tweets about guns, gun rights, and gun policy. Can include references to 2nd Amendment rights, as well as discussions of mass shootings.	*"gun policy"*	Accuracy: 0.975 Cohen's Kappa: 0.881 F1 Score: 0.895
Supreme Court of the United States	Tweets about the Supreme Court. Can include references to specific justices as well as appointments, decisions, the legitimacy of the court, SCOTUS reform.	*"supreme court or SCOTUS"*	Accuracy: 0.995 Cohen's Kappa: 0.924 F1 Score: 0.927
Energy production	Tweets about energy, energy policy and energy production. Can include discussion of various energy sources and associated costs, as well as policy concepts like "energy independence."	*"energy policy or energy production"*	Accuracy: 0.984 Cohen's Kappa: 0.840 F1 Score: 0.848
Foreign policy	Tweets about U.S. foreign policy and international politics. Includes U.S. policy toward/relationship with other countries, or reactions to political or military actions of other countries.	*"foreign policy or international relations"*	Accuracy: 0.985 Cohen's Kappa: 0.908 F1 Score: 0.916
COVID-19 pandemic	Tweets about COVID-19 or social/government response to the pandemic. Includes public health PSAs, discussion of specific policies such as mask or vaccine mandates, references to case counts.	*"covid-19 or coronavirus or pandemic or COVID"*	Accuracy: 0.998 Cohen's Kappa: 0.939 F1 Score: 0.940
Violent crime	Tweets about violent crime & policies to address it. Can include references to crime in other countries. Excludes references to violence in the context of warfare or international terrorism.	*"violent crime"*	Accuracy: 0.975 Cohen's Kappa: 0.850 F1 Score: 0.864
Taxation	Tweets about taxes or tax policy. Can include references to specific tax legislation as well as ideological statements such as "tax the rich" or "taxation is theft."	*"tax policy"*	Accuracy: 0.990 Cohen's Kappa: 0.868 F1 Score: 0.873
LGBTQ+ issues & identity	Tweets about LGBTQ+ identities, issues, and rights. Can include policy areas such as marriage equality or bathroom bans, as well as more general statements of support or opposition.	*"LGBTQ rights"*	Accuracy: 0.998 Cohen's Kappa: 0.904 F1 Score: 0.905
January 6	Tweets about the Jan. 6 insurrection and related hearings.	*"january 6 or january6 or jan6"*	Accuracy: 0.986 Cohen's Kappa: 0.799 F1 Score: 0.806

PEW RESEARCH CENTER

Supplementing machine learning topic prediction with term dictionaries

In some cases, the topic classifiers missed relevant content due to missing contextual knowledge. For instance, a tweet discussing "mask mandates" should clearly be assigned the label *COVID-19 pandemic,* but because the topic classifier was not specifically trained on data related to pandemic response, it lacks the context required to make that connection and relies on the co-occurrence of other known terms to make correct predictions. To address this issue, researchers also developed small sets of issue-specific key terms that were used as an additional mode of identifying tweets about a given topic to supplement the topic labels predicted by the machine learning classifiers described above. Researchers used the predicted topic labels to subset the data by topic and then applied pointwise mutual information to the tweet texts to identify the top 50 most distinctive terms for each topic, by party. This yielded term lists for each topic that represent the distinctive, topic-specific language that candidates from each party were likely to use as they discuss those issues. These term lists were then manually reviewed by the researchers to identify a set of terms that would indicate a tweet is highly likely to be on the given topic if found in the text. This process yielded a total of 704 key terms across 16 topics.

Additional adjustments

To reduce the amount of overlap between substantive topics, any tweets that were identified as being both about *health care* and *abortion* were recoded to only *abortion,* and tweets identified as being both about *health care* and *COVID-19* were recoded to only *COVID-19 pandemic.* Similarly, researchers also adjusted labels such that tweets that were labeled *violent crime* excluded any tweets that were also labeled with *foreign policy*.

Identifying the sentiment of candidates' tweets

To measure the sentiment of candidates' tweets, researchers used another Transformer language model, this one a RoBERTa-base sentiment classifier specifically developed for sentiment identification in tweets. To preprocess the tweets for sentiment classification, all emoji characters were replaced with their text descriptions.

The performance of this sentiment classifier was evaluated against a hand-labeled set of 500 candidate tweets and achieved 98.4% accuracy with an F1 score of 0.984 and a Cohen's Kappa of 0.976.

Identifying mentions of parties, political figures and opponents in candidates' tweets

To identify mentions of the Republican or Democratic Party in tweets, researchers used the same process described in the substantive topics section, with model prompts adjusted to capture mentions of the two major parties. This model was evaluated against a sample of 500 tweets hand-labeled for party mentions, and achieved 98% accuracy with an F1 score of 0.959 and a Cohen's Kappa statistic of 0.946.

Researchers similarly used minimal term dictionaries to supplement the labels predicted by the machine learning model.

- Democratic Party: "democratic party," "democrats," "dems"
- Republican Party: "republican party," "republicans," "gop"

Additionally, researchers used a keyword dictionary consisting of the full names and Twitter handles (if applicable) of all members of the 117th Congress to identify mentions of national political figures, which were then coded as mentions of that figure's party.

To identify generic mentions of bipartisanship, Donald Trump and Joe Biden, researchers used keyword dictionaries.

- Bipartisanship: "bipartisan," "bipartisanship," "across the aisle"
- Donald Trump: "donald trump," "trump," "president trump," "realdonaldtrump," "trump administration"
- Joe Biden: "joe biden," "joseph biden," "biden," "president biden," "joebiden," "POTUS," "biden administration"

Finally, to identify mentions of opponents, researchers created candidate-specific keyword dictionaries by filtering the full list of candidates to only those competing in the same race.

Identifying agreement/disagreement with a mentioned figure or party in candidates' tweets

To identify whether a tweet mentioning a political party or figure indicates agreement or disagreement, researchers used the same machine learning approach described in the section about topic classification. Here, the model prompts were designed to both provide the model with information about the specific mentioned figure or party *and* capture agreement towards them. For example, if in the previous step, a tweet had been identified as mentioning Joe Biden, the model would be asked to evaluate the statement "This text agrees with Joe Biden." This model was evaluated against a sample of 123 tweets validated as containing mentions and hand-labeled for agreement and achieved 88.6% accuracy with an F1 score of 0.89 and a Cohen's Kappa of 0.783.

Identifying topic-specific ‘distinctive’ language for each party

To identify unique, topic-specific language that candidates use to frame the issue areas examined in this analysis, researchers examined tweets by Republicans and Democrats separately. Then, for each topic, pointwise mutual information was used to identify the 100 most distinctive terms that distinguish tweets on that topic from off-topic tweets. The terms on these lists represent the language that characterizes how candidates from each party discuss each issue. These distinctive term lists were then compared across parties to reveal instances of overlap in how Republicans and Democrats discuss issues.

This analysis was performed at the level of individual words (unigrams), but in some cases where two or more highly distinctive unigram terms commonly co-occur as compound phrases, those constituent terms have been merged into those phrases for reporting. For instance, the terms “critical,” “race” and “theory” are all individually characteristic of how Republican candidates tweet about race, but in over 90% of cases they co-occur as the phrase “critical race theory.” Where such merging has been performed, researchers first verified that the reported phrase was the most common usage of the constituent terms.