From Words to Actions: U.S.-China Political Incident Forecasting Through U.S. Official Statement Analysis

Alisha Bi, Ziyu Shu

December 2025

ABSTRACT

Great Power Competition and global shifts toward populism have changed the tone and intent of official diplomatic rhetoric. Furthermore, post-Cold War general stability in U.S. foreign relations is declining, in part motivated by dynamicism in the U.S.–China relationship. In this study, we explore to what extent official U.S. diplomatic statements about the People's Republic of China predict or bring about subsequent bilateral events. Focusing on bilateral events data and official White House statements from 2009 to 2025, we applied lexicon-based sentiment analysis to statements and examined their relationship with U.S–China events using time-series methods. Our diagnostic tests indicated no statistically significant predictive or Granger causal relationships. Limitations highlighted by these results are textual and event feature richness, and the inherent difficulty of establishing causality in international relations. Further research or analysis might address these limitations by mitigating potentially confounding data through filtering events and statements by relevance and significance, though this would require investigation of significance metrics and whether White House statements are the most relevant to diplomatic relations in the era of social media.

Keywords: Data Science, Sentiment Analysis, U.S.–China Relations, Diplomacy, Event Prediction

INTRODUCTION

In 2011, the Obama administration announced its landmark "Pivot to Asia," understood to be emblematic of the broader U.S. foreign policy shift in focus away from the post-9/11 Middle East and toward China's growing influence (Shambaugh 2013). Until then, the U.S. had neither experienced nor acknowledged a near-peer adversary since the collapse of the Soviet Union and the end of the Cold War. Accordingly, the new era of global exertion of China's economic, technological, and military capability is referred to as Great Power Competition. Reframing the U.S.–China relationship as a rivalry and the contemporaneous global shift toward populism have eroded post–Cold War relative stability, which in turn has influenced the tone and intention of U.S. official diplomatic statements (Lieberthal 2011). Official rhetoric has therefore taken on a renewed importance as both a foreign policy tool and potential predictor of diplomatic and bilateral outcomes. Taken together, it is particularly salient to explore the employment and effect of official diplomatic rhetoric on global outcomes in the context of the U.S.–China relationship, which will likely dictate international trends for years to come.

This study examines whether and to what extent official U.S. White House statements explicitly concerning China between the first Obama administration and the Biden administration are predictive of subsequent U.S.–China events, excluding the second Trump administration given its relatively short tenure at the time of this study's completion. Using a corpus of official statements, a coded timeline of significant bilateral events, and a large-scale, machine-coded event dataset, we apply text-as-data-based sentiment analysis to measure variation in diplomatic rhetorical tone over time. We use time-series methods to assess whether these rhetorical measures are predictive of or exhibit Granger causal relationships with our U.S.–China events dataset, employing the coded timeline as primarily a visual and empirically analytical tool for watershed or consequential bilateral events, such as the Coronavirus pandemic.

Our findings indicate no statistically significant predictive or Granger causal relationship between the sentiment of official White House statements and subsequent bilateral events. Accordingly, while shifts in rhetoric often coincide with periods of heightened tension or cooperation, they do not systematically predict future incidents.

1 DATA

This report utilizes data from three sources: the U.S. White House statements archive, the Global Database of Events, Language, and Tone (GDELT), and the Council on Foreign Relations' Timeline of U.S.–China Relations (Council on Foreign Relations 2025).

1.1 Sources and Acquisition

U.S. White House statements archive. The White House maintains public, online archives (Obama White House Archives 2017) (Trump White House Archives 2021) (Biden White House Archives 2025) of previous administrations' official statements, including titles, full texts, dates, and offices. Not all White House statements derive from the President as the Vice President also plays a prominent role within both foreign and domestic politics. The White House statements archives are not inclusive of the President's or Vice President's social media statements.

To get the statements, we applied govinfo API twice. The first step is fetching all package ID from the website archives from 2009-01-20, when Obama took office, to 2025-11-30. The 16929 return from query contains: “package Id”, identifying each document; “title”; “dateIssued”; “collectionCode”;“download”; “resultLink”.

The second step is to get text by items. Noted that historical archives may have different storage method: files after 2010 are othen PDFs, while the old files are usually htmls. Using the ‘download’ item from the results above, we dealt with three scenarios: If there is a txtLink in the download, we download HTML and convert it to text; If there is a PDF link, then download PDF and convert it to text; If neither exists, throw an error and skip. In the end, we successfully construct the corpus with “packageId”, “title”, “dateIssued”, and plain “text” ready for the further analysis.

Global Database of Events, Language, and Tone. GDELT is a large-scale, machine-coded dataset (Leetaru and Schrodt 2025) that records global events by parsing news articles using automated natural language processing techniques. These events are then classified according to Conflict and Mediation Event Observations ontology. This dataset provides temporal markers, actors (nations and individuals), targets, and tone and conflict measures. Importantly, it does not identify or distinguish events as text, and did not begin collecting originating article URLs until 2013.

The GDELT dataset is accessible online through Google BigQuery, though in its entirety, it generates roughly three quarters of a terabyte of data per day. It was therefore impractical and impossible to download in its entirety, so we filtered it in BigQuery before downloading the resulting dataset for further analysis. Our aim was to pull only high-signal U.S.–China interactions within this study’s temporal scope. In order to do so, we filtered out event root codes that classifying events as rhetorical to avoid duplicating our independent U.S. statements data. To elucidate: filtering out public statements or consultations form our GDELT date ensured that we were not predicting statements with themselves. We then restricted actors and targets to the U.S. and China only, such that events had either the U.S. listed as the first actor and China listed as the second, or China listed as the first actor and the U.S. listed as the second. This ensured our dataset did not include events with irrelevant countries or where the U.S. or China was acting upon itself. We also filtered out all events with only one publishing source, thereby removing news aggregates or newsletters that are largely redundant and analytically unhelpful for the purposes of this study. Finally, we only pulled events between January 20, 2009—President Obama’s inauguration—and August 1, 2025—roughly six months after President Biden departed office and therefore our imposed limitation for the predictive capacity of his White House statements on diplomatic outcomes.

Council on Foreign Relations Timeline of U.S.–China Relations. The Council on Foreign Relations maintains a timeline of U.S.–China relations, intended to capture significant events or shifts in the bilateral relationship since the inception of the People’s Republic of China in 1949. Each event is given a brief title, paragraph-long summary, and attributed to a date or date range. It reflects the Washington consensus of the most critical bilateral events—good, bad, and neutral—between both nations. Given the relative brevity of this timeline and the contents therein, temporally relevant sections were copied and pasted manually from the Council on Foreign Relations’ website for further cleaning and analysis.

1.3 Limitations

The White House statements archives, as mentioned, do not capture social media or other executive-branch department statements. This may be particularly relevant for the Trump administration, which carries out much of its most visible diplomatic rhetoric on social media. Furthermore and particularly during the Coronavirus pandemic, other executive departments such as the Health and Human Services Secretary and State Secretary were also issuing statements directed at China. However, the White House statements database was sufficiently robust and large in scale for an analysis. Furthermore, differences in diplomatic tone and strategy were visible across each administration's statement data. Public diplomatic shifts in rhetoric were thus captured meaningfully in our model and analysis, if not in their exhaustive entirety. For this reason, we opted to proceed with this study with a scope focused primarily on the President and Vice President's officially published statements.

GDELT's lack of textual identification or description of its events and the nature of global, cross-linguistic machine-automated data collection resulted in some events being so closely related as to be indistinguishable from one another. Because GDELT only began recording the originating URL of its event scraping in 2013, there was no systematic way for us to clean the data such that each event was isolated as a single observation. Separately, while GDELT's mechanism for recording country-level actors like the U.S. and China is strong, its ability to parse individual-level actors such as President Obama versus Microsoft is significantly weaker. We were therefore unable to filter out events that were not overtly diplomatic, though it is worth noting that even events that are ostensibly limited to the private sector such as Microsoft divesting from Chinese supply chains are, in reality, inextricable from broader diplomatic and geopolitical trends. These limitations therefore do not strongly impact our analysis, though they could impair finer, more microscopic analyses that were nonetheless outside of the scope of our proposed research. Finally, the nature of GDELT's data collection revolving around media mentions and articles may in some cases be indicative of popular interest rather than diplomatic significance. Overall, however, GDELT data matches well with the Council on Foreign Relations' timeline in that bilateral lows and highs in the relationship are generally well reflected.

2 METHODS

2.1 Data Cleaning

We implemented pre-processing and cleaning on statements text. We removed the unnessary notes in the text, like trailing metadata "DCPD Number", header "Administration of . . . ", weird all-caps chunks, etc. We also imported NLTK package to remove stopwords. By setting dictionary of China terms, we filtered 2459 presidential statements which are directly mentioned China.

GDELT data did not require further cleaning given how robust the BigQuery filtering functions were and how standardized the collected events data inherently are. The Council on Foreign Relations' Timeline data was manually reformatted as a dataframe in R for visualization purposes. In order to capture impact, OpenAI's GPT-5.1 was prompted to classify each event as positive, negative, or neutral in the context of the U.S.–China relationship. These classifications were manually checked for accuracy prior to population within the data.

2.2 Sentiment Analysis

It took a while for us to looking for appropriate method to extract sentiment features. The first attempt was applying Sentiment Intensity Analyzer from NLTK via VADER, polarity_scores functions would return a float for sentiment strength based on the input text. Positive values are positive valence, negative value are negative valence. By tokenization and settiing hedge and assertive wordlists, we tried to identify which entities are mentioned in the text, the hedge and assertive densities, how many times the documents mentioned China, and sentiment scores. We didn't apply other dimensions besides China mentions count in the end, becuase the limitation of Sentiment Intensity Analyzer is not suitable for policy-tone official statements. The score is systematically closer to 1, which is pretty positive and lack of informative variance.

As the final result shows, we eventually employed the FinBERT, a pre-trained BERT model trained by finance and policy corpus, to extract the sentiment features. Considering the tariffs topic is popular and repeatedly emerged between U.S. and China, we expacted the FinBERT could perform better in this task.

The model will give softmax outputs for three labels: positive, negative or neutral. Based what model returned, we calculated the continuous sentiment scores ranging from -1 to1 rather than using three categories, otherwise the very large amount of neutral tags would disturb the analysis.

Additonally, we also calculated adjusted sentiments using an entropy-based certainty measure. Specifically, we compute the Shannon entropy of the predicted class probabilities for positive, neutral, and negative sentiment. Entropy captures the degree of uncertainty in the model's prediction: it is maximized when the probability mass is evenly distributed across classes and minimized when one class dominates. We normalize this entropy by its theoretical maximum for a three-class distribution and define certainty as one minus the normalized entropy, yielding a bounded confidence weight between 0 and 1. The final sentiment measure is obtained by multiplying the original sentiment score by this certainty weight, thereby down-weighting predictions characterized by high uncertainty while preserving the magnitude of confident predictions.

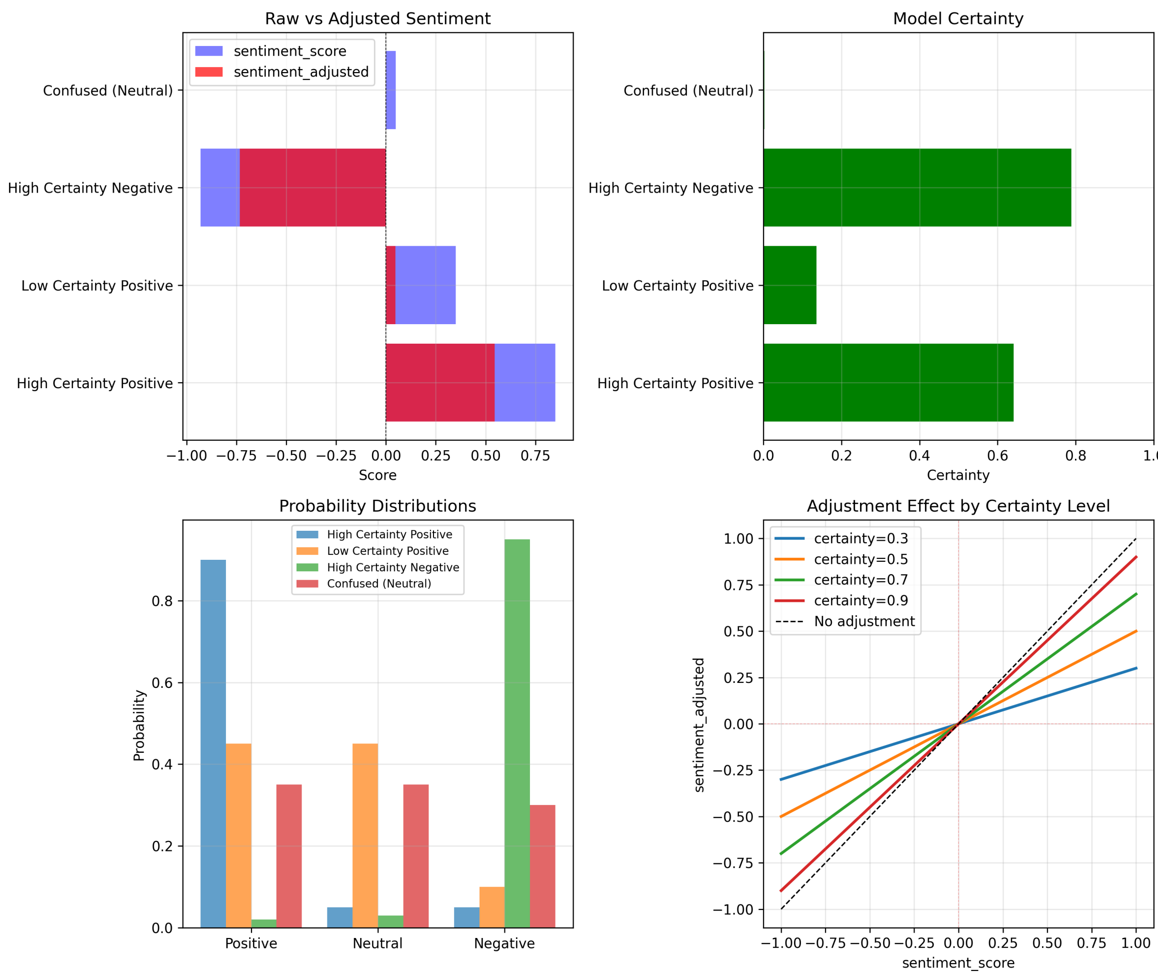

Figure 1: Adjusted Sentiment Scores Statistics

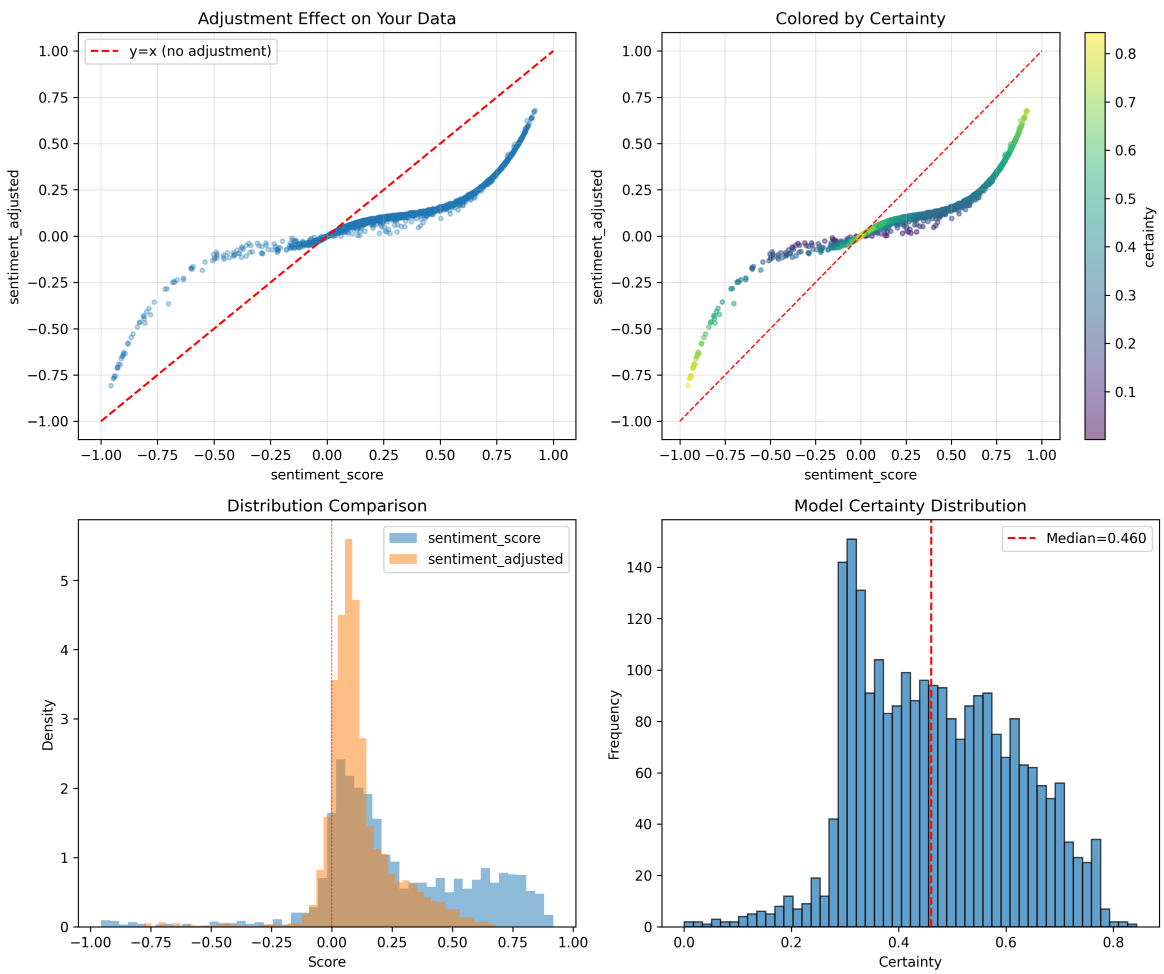

Figure 2: Adjusted Sentiment Scores Statistics

3 ANALYSIS AND MODELS

In this section, we analyze whether variation in official White House statement rhetoric about and toward China is associated with U.S.–China event outcomes through statistical analysis and predictive models. We assess both administration-level heterogeneity and the dynamic statistical relationship between statement sentiment and event conflict measures. Finally, we use our scraped White House statements text-as-data to attempt bilateral event forecasting, combining descriptive analysis with machine-learning methods.

3.1 Data Visualization

Statement Sentiment by Administation. Figures 1 and 2 visualize the sentiment patterns of White House statements across four presidential administrations: the first Obama administration (2009-2013), the second Obama administration (2013-2017), the first Trump administration (2017-2021), and the Biden administration (2021-2025), overlaid for visual support by events data. Figure 1 displays statement sentiment and event conflict data individually whereas for visual clarity, Figure 2 aggregates this data on a weekly basis.

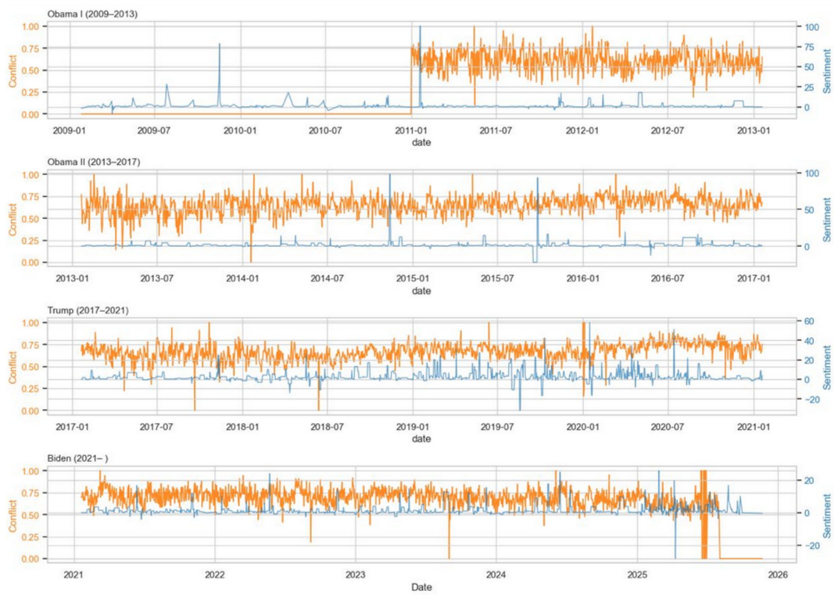

Figure 3: Statement sentiment and event conflict intensity by presidential administration

Figure 4: Aggregated weekly statement sentiment and event conflict intensity by presidential administration

There are several features to note about the statement sentiment data. First, the both Obama administrations exhibit the greatest rhetorical stability in that sentiment scores strongly cluster around neutral and positive values even in the aftermath of its 2011 Pivot to Asia. Although there is a decline and stronger variance in sentiment in the second Obama administration, we do not observe a shift toward strong sentiment volatility until the first Trump administration, which abates but continues into the Biden administration. It is noteworthy, however, that event conflict volatility begins in the second Obama administration, an initial indicator of White House statements' low predictive power with respect to diplomatic outcomes.

Event Tone by Administration. Figures 3 and 4 focus on the GDELT-derived U.S.-China event data. Figure 3 focuses on individual raw event mensions and their tones, whereas Figure 4 weighs tone by mention by multiplying event mentions by average tone, aggregated weekly to capture broader trends and reduce visual noise.

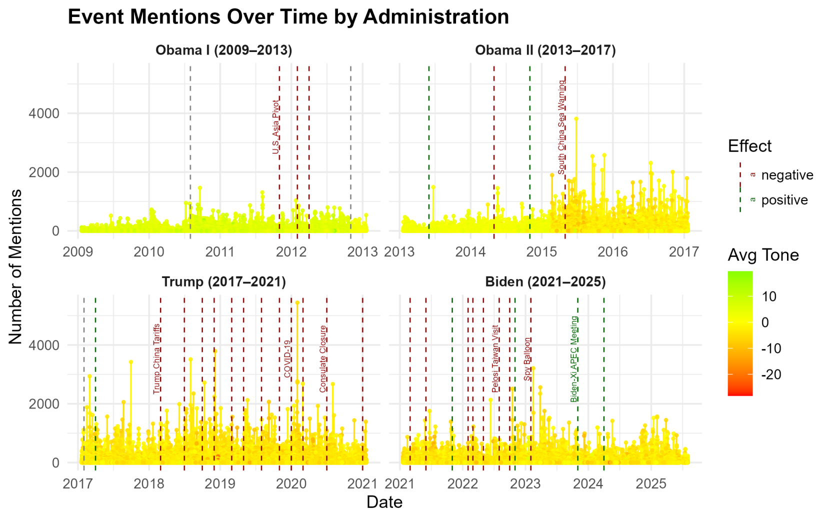

Figure 5: Event tone and mentions by date and administration

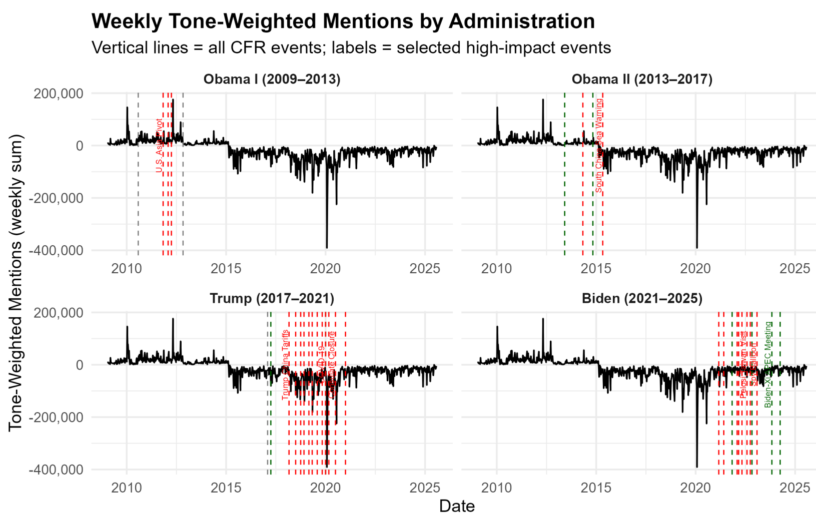

Figure 6: Weighted event-level tones by weekly mean

Notably, the first Obama administration corresponded with the highest level of stability across events and the only administration in which events trended between neutral and positive. This stability dips significantly before the onset of the U.S. involvement within the South China Sea dispute. The first Trump administration again displays the most volatility and negativity, during which the bilateral relationship experiences its lowest point around the onset of the Coronavirus pandemic within the United States. Volatility subsequently eases to pre-Trump levels with the Biden administration, though events are still generally more negative than neutral or positive.

3.2 Models

The relationship between statements sentiments and event conflict share are two-ways, intuitively: we can't simply assume the statements are necessarily prior than the event happening, we can't ensure there isn't responsive statements to happened events either. Thus, we chose the Vector Autoregression model(VAR).

A VAR model is a multivariate time-series framework in which each variable is modeled as a function of its own past values and the past values of all other variables in the system. Unlike single-equation regressions that impose a unidirectional causal structure, VARs allow for dynamic feedback and interdependence among variables over time. Each equation in a VAR is estimated by OLS, but interpretation focuses on the system's dynamic properties rather than individual coefficients.

In particular, impulse response functions (IRFs) and forecast error variance decompositions are used to trace how shocks to one variable affect the evolution of others across future periods. As a result, VAR models are well suited for studying temporal dynamics in our time series data and predictive relationships in settings where theory does not clearly specify causal direction.

3.3 Results

Although we estimate a VAR model to explore the dynamic relationship between official sentiment and conflict events, the results suggest that the two series exhibit little systematic interaction over time. Lagged sentiment terms (we chose lag = 7) do not significantly predict future conflict, and the impulse response functions remain close to zero, indicating no meaningful Granger-causal effect from language to violence (or vice versa). Thus, while contemporaneous correlations may exist, simple time-series regressions do not provide evidence of a causal impact of official sentiment on subsequent conflict dynamics.

The IRF patterns demonstrate that conflict events are driven primarily by their own past values, showing strong autoregressive dynamics. Shocks to official sentiment have no meaningful impact on the trajectory of conflict events. Conversely, conflict shocks produce only weak and temporary changes in sentiment, with effects quickly returning to zero.

Taken together, the dynamic evidence suggests little to no directional causal influence between sentiment and conflict, in either direction, within this VAR framework. The relationship is dominated by each variable's internal time-series structure rather than cross-variable effects.

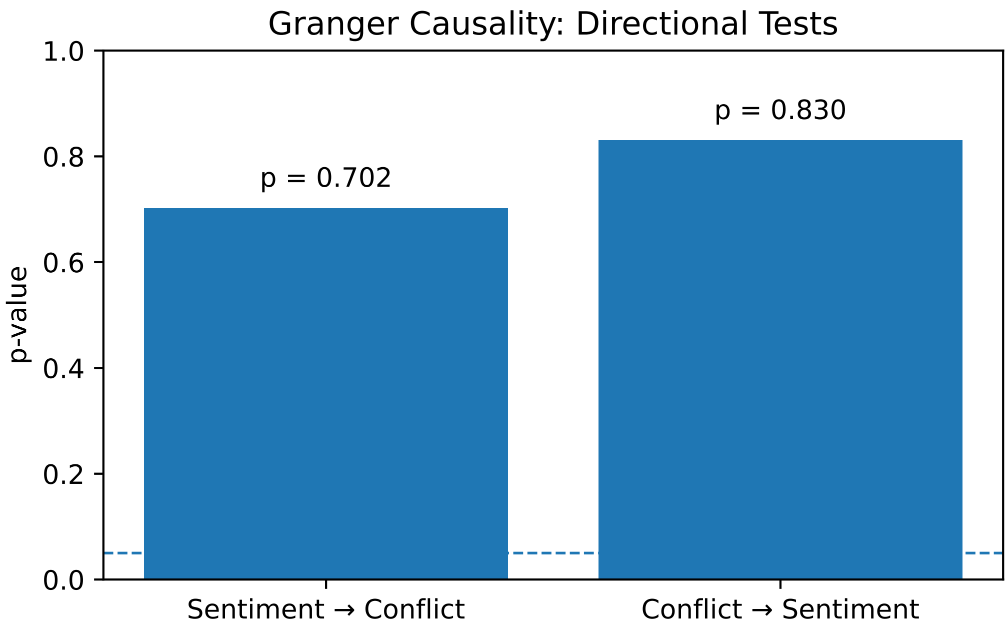

Figure 7: Granger causality tests (weekly VAR): p-values for each direction. Both p-values exceed 0.05, so we fail to reject the null of no Granger causality in either direction.

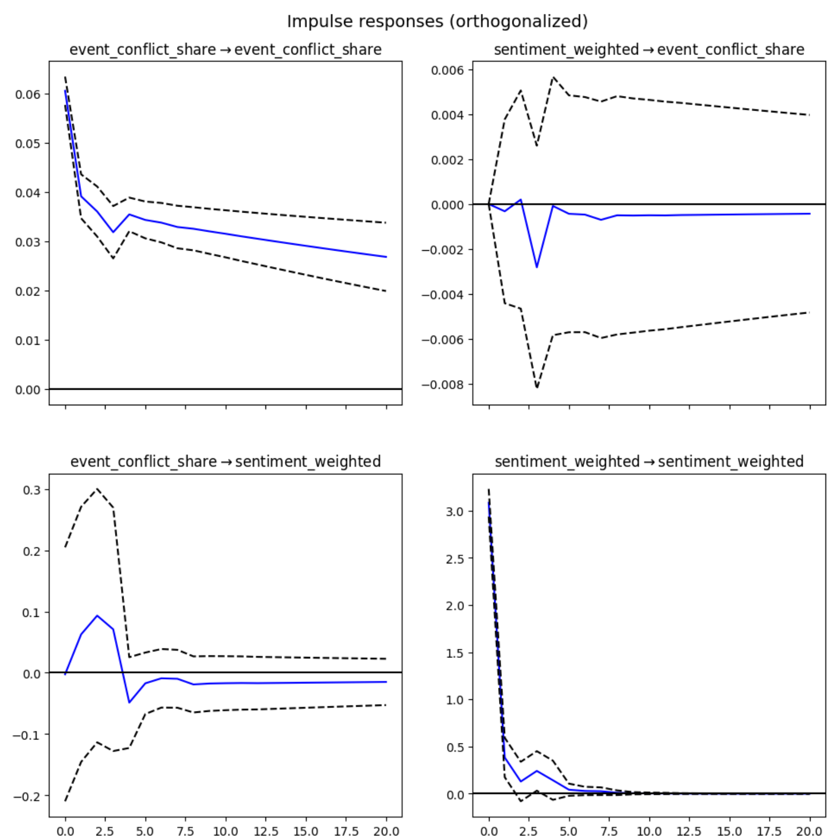

Figure 8: Impulse Response Functions (IRFs) for VAR

4 DISCUSSION

4.1 Conclusion

Ultimately, we derived a null result from both our statistical analyses and predictive models for whether White House statements have a forecasting capacity for U.S.–China bilateral event outcomes. Across visual, regression, and causality tests, we find no significant evidence that statement sentiment systematically predicts inflections in the U.S.–China relationship. These results are surprising in that diplomatic signalling through official statements is understood to be of paramount importance in the foreign policy toolkit. These results conflict with this consensus, supporting instead that international events are neither predictable nor intentionally brought about by even arguably the most empowered diplomat worldwide, the U.S. president.

According to our results, White House statement sentiment does not even reliably predict near-term U.S.–China event outcomes, which we might otherwise expect them to given how frequently official statements are coordinated bilaterally and in anticipation of events that the public does not yet have awareness of but that White House officials know will soon break in the media. We expected that the GDELT event data might capture this dynamic, though our findings indicate otherwise.

It is important to note that the absence of a predictive relationship from this study does not imply that White House statements are ineffectual or unimpactful to outcomes. It may indicate that, contrary to popular belief, White House diplomatic statements are not intended to shape future outcomes, but to justify or elucidate events that have already come to pass. Our null result may therefore indicate that observable shifts in sentiments expressed within White House statements are more directed at a domestic audience to enlighten about present foreign policy and circumstances.

4.2 Limitations

From the methodology level, looking back, we admitted that our existing pipeline of aggregating high frequency time series sentiment scores and event data are immature. We already had the intuition of using SBERT to construct 384 dimenstions of document-levl embeddings, which are then used as input features for a supervised sentiment classification (or regression) model. But since the limitation of time and workload, we failed to run through the alternative pipeline and explore more possibilities with the data we had.

In addition to data limitations discussed earlier, the results of our analyses and model tests indicate several conceptual and practical research constraints. First and most conceptually, causal identification within international relations is difficult to establish even with access to robust datasets. Even with lag structures built into models to account for temporal delays in reaction and to attempt forecasting, international relations are highly impacted by largely unpredictable and unmeasurable geopolitical factors that are difficult to measure or control for.

As an example, no White House statement prior to the outbreak of the Coronavirus could have predicted the onset of a pandemic.

In that vein, sentiment analysis may yet be insufficient in capturing diplomatic strategic ambiguity and subtle signalling. Official statements—in particular, those published by the White House and attributed to the U.S. President—are subject to complex interdepartmental legal review and calibrated with different intentions, the sum of which is perhaps not adequately captured within a compressed sentiment analysis. Furthermore, this study did not seek to contextualize statement sentiment as a predictive measure, an endeavor that also poses strong conceptual and practical difficulties.

Finally, the opaque nature of diplomatic relations makes near-contemporaneous statistical analyses and predictive modeling difficult because bilateral events that are considered salient within the media and to the public are not necessarily the most diplomatically relevant. While we inevitably capture relationship highs like presidential visits or lows like pandemics, we were not able to fully account for less visible or even classified but nonetheless significant bilateral events and outcomes.

4.3 Further Research

Our results indicate many avenues for potential further exploration. Namely, future sentiment analyses of political language as sensitive as diplomatic statements may need to be trained only on these statements and somehow inclusive of contextualization. For example, a lukewarm statement during a period of extreme volatility in the bilateral relationship may actually require positive sentiment coding in that it should be interpreted as a positive, stabilizing diplomatic overture. Beyond the scope of this study but also worthy of further research is the contribution of non-Presidential political actors' statements to bilateral relationship trends. Members of Congress and prominent private sector executives publish their own official statements and actions about the U.S.-China relationship that result in concrete diplomatic outcomes, such as then-Speaker Pelosi's 2022 visit to Taiwan, the aftermath of which resulted in overt threats from China's Foreign Minister (Yeung, Picheta, and Trimble 2022).

Finally, future research might explore domestic outcomes of official statements that are ostensibly diplomatic and foreign-facing. Given our null result and that these statements are likely consumed primarily by U.S. rather than foreign citizens, sentiment analysis my be better employed as predictive of U.S. attitudes toward China or even presidential approval.

REFERENCES

Biden White House Archives. 2025. "Briefing Room: Statements & Releases (Archived)." 2025. https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/.

Council on Foreign Relations. 2025. "Timeline: U.s.-China Relations." 2025. https://www.cfr.org/timeline/us-china-relations.

Leetaru, Kalev, and Philip A. Schrodt. 2025. "GDELT: Global Data on Events, Location, and Tone." Proceedings of the ISA Annual Convention. https://www.gdeltproject.org/.

Lieberthal, Kenneth G. 2011. "The American Pivot to Asia." Brookings Institution. December 21, 2011. https://www.brookings.edu/articles/the-american-pivot-to-asia/.

Obama White House Archives. 2017. "Briefing Room: Statements and Releases (Archived)." 2017. https://obamawhitehouse.archives.gov/briefing-room/statements-and-releases.

Shambaugh, David. 2013. "Assessing the US 'Pivot' to Asia." *Strategic Studies Quarterly* 7 (2): 10–19. http://www.jstor.org/stable/26270763.

Trump White House Archives. 2021. "Briefings and Statements (Archived)." 2021. https://trumpwhitehouse.archives.gov/briefings-statements/.

Yeung, Jessie, Rob Picheta, and Megan Trimble. 2022. "Pelosi Taiwan Trip Fuels u.s.–China Tensions." August 4, 2022. https://www.cnn.com/asia/live-news/pelosi-taiwan-trip-us-china-08-03-22-intl-hnk/.