

A Text-as-Data Approach to Measuring Interstate Relations Using National Security Strategies

J Andrés Gannon

September 25, 2025

This study develops and validates a novel text-as-data method for measuring interstate relations from political texts. Building on advances in natural language processing, I construct a directed dyad-year index of a state’s expressed salience (prominence) and affinity (evaluative stance) toward other states in its national security strategy. The measure is derived through a pipeline that integrates sentence-level segmentation, entity recognition, sentiment analysis, stance detection, and best–worst scaling calibration. Applied to a new corpus of 438 national security strategy documents issued by 92 countries (1962–2024), the approach produces reliable, continuous measures of interstate posture that closely track but also meaningfully diverge from traditional benchmarks such as alliance portfolios and UN voting patterns. This methodological contribution provides scholars with a replicable tool for extracting conveyed relationships from a wide range of political texts, enabling new insights into the articulation of political relationships.

1 Introduction

Who do political actors talk about and how? Many social science theories concern relational signaling, where actors convey the importance of other actors and their orientation toward them. However, this has proven difficult to measure because signals of the nature of relationships is often embedded in actors' communication with each other, motivating scholars to instead rely on indirect material behaviors that suggest certain actors are important or that the relationship with that actor is conflictual versus cooperative. This problem is particularly pronounced in international relations, where scholars rely on coarse metrics of political relevance and foreign policy alignment that have little temporal or spatial variation and that require assumptions about other motivations for those actions that may complicate their interpretation. Political relevance is assumed to be driven by geography and great power status which vary little over time, even as researchers suspect (and posit) that the international actors a states considers relevant to its foreign policies likely spans the globe in varying degrees. And although there are measures for interstate alignment like alliances and rivalries as well as measures of foreign policy similarity, methods for estimating non-reciprocal affinities as well as how states themselves communicate those alignments is lacking.

To understand how actors themselves communicate the importance of others and their orientation toward them, we need a measure that captures both the intensity and directionality of those relationships. For example, the credibility of alliance commitments relies on the protege state believing that the patron has sufficient interest in the protege's security to risk conflict for the protege's sake (Fuhrmann and Sechser 2014). To know if states communicate changing relations in real time requires identifying that change in intensity and direction using indicators that states themselves observe.

The lack of estimates for this is attributed to the difficulty in measuring how actors themselves

signal relationships, as most attention in international relations has been focusing on signaling interests and related dimensions like intention, capability, and resolve (Fearon 1994; Gartzke et al. 2017; Min 2025). And scholars have largely used measures that states themselves are unlikely to use, as behaviors like UN General Assembly votes and alliance ties serve other purposes that make it unlikely states themselves use them as proxies for relational signaling.

This paper develops and validates a novel text-as-data method for measuring interstate relations from political texts. I take advantage of state-of-the-art artificial intelligence and machine learning techniques using large language models (LLM) to measure how states communicate the salience and affinity of interstate relationships in official public documents expressly written for international audiences: National Security Strategies (NSS). The measure is derived through a pipeline that integrates sentence-level segmentation, entity recognition, sentiment analysis, stance detection, and best–worst scaling calibration. Applied to a new corpus of 438 national security strategy documents issued by 92 countries (1962–2024), the approach produces reliable, continuous measures of interstate posture that closely track but also meaningfully diverge from traditional benchmarks such as alliance portfolio similarity and UN voting patterns. Furthermore, this measure predicts interstate crises and diplomatic visits better than existing measures of foreign policy similarity, demonstrating that states themselves ‘code’ and publicly broadcast their own sense of relevance and affinity beyond what scholars currently capture with indirect proxies.

While computational text analysis is increasingly used in international relations to analyze diverse texts (eg. Baturo, Dasandi, and Mikhaylov 2017; Carson, Min, and Nuys 2024), this study’s unique contribution lies in its systematic operationalization and validation of affective salience, extracting explicitly relational cues of salience and affinity from the specific genre of strategic public state documents. Importantly, measuring salience and affinity from these documents is not fully representative, as not all states write these documents and not in all

years, and they are further limited to the context of national security. Nevertheless, this approach offers significant unique insights about how states themselves communicate interstate relationships in a way that is not possible using existing measures. Moreover, the generalizability of the methodological approach and data pipeline demonstrates how computational text analysis can reveal patterns in how actors perceive and prioritize relationships over time. A further methodological contribution is providing scholars with a replicable strategy for extracting conveyed relationships from a wide range of political texts, enabling new insights into the articulation of political relationships.

I proceed as follows: first I review the literature on salience and affinity, focusing on existing approaches in international relations where these concepts are contested, but reliable measures have emerged. Second, I explain how the NSS data are gathered, processed, structured, and made publicly available as an independent empirical contribution to the study of state communication. Next, I describe the modeling approach that produces new directed-dyad year measures of salience, affinity, and affective salience. I then validate the estimates against existing measures of political relevance and affinity. The next section applies the measure, demonstrating that it predicts future conflict and cooperation better than existing measures. The final section concludes with a discussion of ways this technique opens up new directions in the study of social signaling and suggests future work to improve and extend the approach.

2 Measurements of Interstate Ties

Initially motivated by a concern about which interstate dyads constituted the most appropriate universe of empirical cases, much ink has been spilled both theorizing and measuring what makes a state relevant to the security of another (Maoz 1996). The prevailing view takes a gravity model approach assuming the opportunity and willingness of two states to

undertake meaningful political engagement is a function of their geographic distance and material power/status.¹ Early work thought about geographic distance as some form of direct or indirect contiguity (Weede 1976; Bennett 2006), minimum distance (Gleditsch and Ward 2001), capital-to-capital distance (Lemke 1995), or co-location in the same region (Small and Singer 1982). Measures of material power and status similarly abound, with many arguing a dyad is politically relevant if it contains a great power (Organski and Kugler 1980; Singer 1988; Maoz and Russett 1993) and others using more granular measures of force projection like the Composite Index of National Capabilities (CINC) (Singer, Bremer, and Stuckey 1972) or military spending (Kydd 2000).

Broadly speaking, existing research has considered affinity as potential or actualized threats or partnerships. Despite being a bipolar scale, much work justifiably focuses on either the positive or negative dimension of affinity to hone theoretical richness.² Some scholarship has highlighted historical instances of hostility and rivalries (Thompson 2001; Terechshenko 2020; Thompson, Sakuwa, and Suhas 2021) while others identify and theorize formal alliances and similar forms of defense cooperation (Leeds et al. 2002; Gibler 2009; Kinne 2020). Given the unavoidable entanglement between positive and negative interstate relations, others have developed ways to simultaneously measure the full spectrum of affinity using dyadic peace levels ranging from serious rivalry to security community (Diehl, Goertz, and Gallegos 2021). The most common measures of international affinity conceptualize it as foreign policy similarity measured using alliance portfolios (Signorino and Ritter 1999; Chiba, Johnson, and Leeds 2015) or UN voting

¹My aim is not to rehash debates about ‘politically relevant dyads’ here, nor to provide another reformulation of what dyads share a propensity for conflictual or cooperative political interaction. For detailed evaluations of various criteria for ‘political relevance’, see in particular Lemke and Reed (2001), Benson (2005), Bennett (2006), Xiang (2010), and Braumoeller and Carson (2011). For extensions beyond the ‘political relevance’ framework, more tailored understandings include dangerous dyads (Bremer 1992), politically active dyads (Quackenbush 2006, 2024), strategic reference groups (Maoz 2010), globally relevant dyads (Lee and Thompson 2017), extra-dyadic linkages (Gartzke and Gleditsch 2022), and grievance dyads (Owsiak and Vasquez 2022). For the turn toward network-based models of political relevance see, among others, Maoz (2010), Ward, Ahlquist, and Rozenas (2013), Dorff and Ward (2013), and Cranmer and Desmarais (2016).

²This section is by no means exhaustive since, as is implied by its name, the entire field of international relations is about variation in interstate relationships.

patterns (Gartzke 1998; Bailey, Strezhnev, and Voeten 2017).

These measures are widespread for good reason; the data have been carefully collected, they often have broad geographic and temporal scope, and findings using these data are consistent with much of what scholars intuitively suspect about international interactions. But while empirically tractable, these approaches tend to assume symmetry, relative temporal stability, and/or are often binary or ordinal measures that make more nuanced distinctions between observations challenging.³ In reality, some allies may be indispensable while others are peripheral; some rivals dominate security planning while others are ignored. Given the binary nature of measures of political relevance, it is primarily used as a methodological scoping condition rather than something that calibrates the intensity of expressed affinities. Furthermore, states themselves do not produce these datasets and it is unlikely that government decisions are influenced by how interstate relations are coded by scholars. So, indirect scholastic measures of how states relate to each other captures something similar, but not identical, to how states communicates their relations to each other.

3 NSS Corpus

I define an NSS as a public text written by a state outlining aspects of their foreign policy pertaining to security and defense. These documents matter; they do guide and describe national security decisions by connecting ends and means and by conveying intended national security priorities (Caudle 2009; Ettinger 2017). Policymakers invest significant resources in their production and attend closely to those issued by others, in part because these documents

³All these shortcomings do not apply to every measure. UN voting patterns do exhibit temporal variation, but have been criticized for masking strategic abstentions (Bailey, Strezhnev, and Voeten 2017), inflating consensus (Häge and Hug 2016), and statistical difficulties accounting for chance agreement (Häge 2011). Alliances have well established observable differences in their depth, reliability, and reliance (Mattes 2012; Alley 2021; Gannon 2025). Rivalries are similarly distinguished as being “lesser” or “severe” (Diehl, Goertz, and Gallegos 2021) and in terms of their cause (Thompson, Sakuwa, and Suhas 2021).

exist to be appraised and scrutinized by foreign audiences (Snider and Nagl 2001; Mohr et al. 2013; DuMont 2019). These texts are intended to provide relational insight, functioning “as strategies for relating enemies and allies” (Burke 1941, 304). Indeed, because threat assessment itself is “at its heart, a form of storytelling,” NSS documents serve as one of the primary vehicles for such narratives (Krebs 2015, 16). These characteristics explain why, “if any words matter, the words in [national security strategy documents] do” (Chin, Skinner, and Yoo 2023, 104). The corpus can thus be thought of as a sample of public documents in which states communicate the importance of other international actors and their orientation toward them.

The first stage involved compiling 438 NSS documents (1962–2024) from 92 states via online government archives, think tank repositories, and multilingual secondary databases, building on efforts by Becker and Malesky (2017), Razeto and Jenne (2021), Becker et al. (2024), and Neal and Gardner (2025). Metadata was also collected to inventory the document name, issuing country, year of publication, and language of publication. In cases where the document was a specific document subcategory like a cybersecurity strategy or budget outlay that was also recorded.

In the second stage, I convert the pdf files to markdown and plaintext format using PyMuPDF4LLM (Smith 2023; McKie 2024) and Marker (Paruchuri 2024). This conversion process preserves formatting markers demarcating natural linguistic boundaries like headings, paragraph breaks, lists, and the location of images and tables to maintain context provided by the structure and order of the documents. The csv version of the data is thus formatted with each page of the pdf as a distinct row and columns providing the markdown text as well as identifier information about the country, year, document name, and page number. This then allows each NSS to be separated into textual segments (chunks) using an MPNet tokenizer from ‘sentence-transformers/all-mpnet-base-v2’ (Song et al. 2020).⁴

⁴The alternatives to chunking often pre-process texts, which fails to retain contextual information (Lahoti,

The result is a dataset of NSS documents in pdf, plain text, markdown, and csv format that can be viewed and downloaded in bulk or individually at (REDACTED URL). The code used to convert the documents is also provided there so that future improvements in text processing, particularly in non-English languages and of non-Latin characters, can use the existing corpus as a benchmark for measuring improvements.

The full processed and machine-readable corpus is made available so future research can examine the issue-specific documents, compare security strategies issued by different agencies within a government, and study the degree to which non-English language or non-Latin character documents systematically differ in how they are processed. But as a note of caution, the comparability of these documents across space and time must be theoretically informed by knowledge of the document’s specific context and the research question being explored (Stolberg 2012). Not all states produce these documents, and those that do often do so in some years, but not others. This could occur because a state produces these documents at regular temporal intervals like once per administration, because of strategic reasons to keep such information secret, or because a state has insufficient bureaucratic capacity to produce them (or even insufficient concern for international security). Alternatively, states may produce more than one type of NSS and may even do so in the same year.⁵

To ensure comparability across texts, I focus on the NSS for the U.S. as “the only complete whole of government national security document that the U.S. Government publishes” (Stolberg 2012, xi) and make similar choices for other states when more than one document exists for a given year. While acknowledging the significant diversity in state capacity, political systems, and bureaucratic traditions that shape these documents, this study proceeds on the

Garimella, and Gionis 2018; Burnham 2024; Overbeck et al. 2025) and has recognized biases when used with the more advanced computational models used here (Palomino et al. 2017; Denny and Spirling 2018; Miyajiwala et al. 2022).

⁵For example, the United States publishes a National Security Strategy (NSS) as its primary document, but also publishes a National Defense Strategy (NDS), National Military Strategy (NMS), Nuclear Posture Review (NPR), National Homeland Security Strategy (NHSS), and National Intelligence Strategy (NIS).

premise that when states choose to publicly issue a formal ‘National Security Strategy’ or equivalent high-level defense White Paper, they are engaging in a broadly similar *type* of strategic communication aimed, at least in part, toward international audiences. The core function of articulating security priorities and framing key relationships is assumed to be a common denominator, though the specific nuances and emphases will vary.

Consequently, the indices of salience and affinity that constitute the remainder of this paper are calculated using only English-language NSS documents that represent a ‘whole-of-government’ security strategy most analogous to the U.S. NSS; issue-specific national security documents like those concerning only cyber defense or budget outlays are excluded. The majority of NSS documents are published in English, as it often serves as a lingua franca for international diplomatic communication and states issuing an English version of their NSS likely intend it for a broad international audience. And methodologically, while multilingual LLMs are advancing (Waight et al. 2025), achieving high fidelity across languages using the computational models used here presents significant challenges, and ensuring comparable measurement quality across languages for the indices of interest is a substantial undertaking beyond the scope of this study.

4 Using NSS to Measure Salience and Affinity

For each state in the Correlates of War Project (2017) state membership list, I create a list of aliases and synonyms that include variations of the country name, capital city, and demonyms.⁶

I then use these lists to identify the chunks in the NSS documents where at least one non-self

⁶States may reference each other indirectly (e.g. ‘our northern neighbor’) or as collectives (e.g. ‘our European allies’). Given the number of texts, possible references, and ambiguity about appropriate inclusion involved in properly coding such references, I here only include direct country-to-country references. Future work using more advanced NLP techniques and area-specific knowledge could provide additional coding for country-region and country-institution references and study the conditions under which states use intentionally ambiguous references like euphemisms or amalgamations.

state is mentioned. The chunks are then pruned into sentences using spaCy’s sentence tokenizer to identify sentences where State B is mentioned as well as the sentence immediately before and after that reference (Montani et al. 2023). This process allows me to identify all mentions of each state in the NSS documents and their immediate semantic context.

Applying text-based machine learning techniques to this processed data, I produce a continuous directed dyad-year measure of salience and affinity from each issuing state toward each target state referenced in the issuing state’s NSS (n = 14,789). The measures of salience and affinity describe the target state’s prominence and the issuing state’s expressed evaluative orientation toward that target state in the issuing state’s NSS document, defined as follows:

1. **Salience:** This reflects the prominence of the target entity within a specific context indicating the degree to which the target entity occupies the issuing entity’s attention and is judged relevant to their considerations or objectives. Specifically, salience is the prominence of a target state within an issuing state’s NSS document.
2. **Affinity:** This denotes the evaluative orientation – positive, negative, or neutral – publicly expressed by a communicating entity toward a target entity. Specifically, affinity is the communicating alignment or misalignment of interests between the issuing state and the target state.

This procedure is summarized in Figure 1 and explained in the sections that follow.

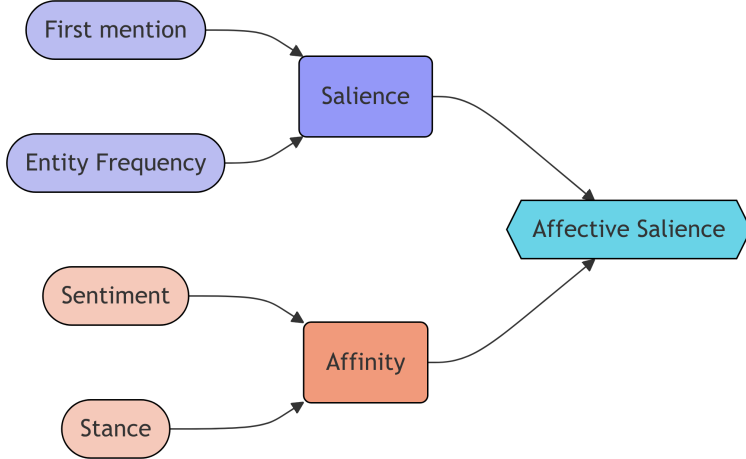


Figure 1: Components of affective salience and their measurements.

4.1 Salience

First, I measure the *salience* of a target entity in the issuing state’s NSS text chunks using two indices that are commonly used both independently and in combination in computational models of search engine optimization (SEO) and salient named entity extraction (SNEE); first-mention and entity-frequency (Dunietz and Gillick 2014; Wu et al. 2020; Grönberg 2021).⁷

First-mention is the document chunk index m_e for each target entity e where it is first mentioned. I then identify the total number of chunks N where at least one entity is mentioned and invert and normalize the measure as $S_{\text{first}} = \frac{N-m_e+1}{N}$ so that entities mentioned in the first chunk have a score of 1 and entities mentioned in the last chunk $m_e = N$ have a score strictly greater than, but approaching zero.⁸ This normalization adjusts for the fact that NSS

⁷Other salience detection models like SWAT (Ponza, Ferragina, and Piccinno 2019), SEL (Trani et al. 2018), and transformer models (Asgarieh, Thadani, and O’Hare 2024; Bullough et al. 2024) can achieved higher precision and recall, but produce only binary outputs and are very sensitive to fine-tuning, impacting replicability and external validity (Bhowmik et al. 2024).

⁸An alternate measure using $S_{\text{first}} = \frac{N-m_e}{N}$ would mean that states first mentioned in the last chunk of a text would have a first-mentioned score of 0. The use of $m_e + 1$ allows me to differentiate these states from those that are not mentioned at all, as the latter are more appropriately measured as having a salience of 0 than the former. The notation that follows assumes that non-mentioned entities are ignored and then assigned a

documents vary greatly in length and some states mention other actors frequently while others mention other states infrequently.⁹

Entity-frequency is measured as the total number of chunks f_e where entity e is mentioned. This is calculated as $S_{\text{freq}}(e) = \frac{f_e}{f_{\text{max}}}$ where f_{max} is the highest number of chunks any single entity appears in within that same document. This normalization similarly adjusts for the fact that some issuing states mention many states in their NSS and others mention relatively few. Because the data are processed in pruned chunks rather than individual words or sentences, multiple references to a country within the same contextual chunk do not bias the measure.

First-mention and entity-frequency are both bounded $x \in (0, 1]$ with higher values indicating higher salience. Salience is then measured by linearly combining both components as $S_{\text{salience}} = \alpha S_{\text{first}} + (1 - \alpha) S_{\text{freq}}$ where $\alpha = 0.5$ is a hyper-parameter equally weighing the two components and maintaining the original scale. This reflects the theoretical assumption that both initial prioritization and sustained attention contribute equally to a state’s communicated strategic focus within these documents.¹⁰ Whereas political relevance is typically operationalized by scholars as a binary property, salience here represents a continuous, directed measure of stated importance from one state toward another.

4.2 Affinity

Second, I measure *affinity*, the affective position expressed toward a target state. Although many computational text analysis models have used sentiment to measure attitude toward

score of 0 if it is a recognized COW state in that year.

⁹This approach implicitly assumes that the relative position within a document (e.g., appearing halfway through) carries similar weight regardless of absolute document length, effectively normalizing for reader attention across documents of varying sizes. While well-resourced state apparatuses might scrutinize entire documents meticulously, the attention of broader audiences or analysts in less critical dyads might indeed wane with document length. This assumption is a simplification, and alternative normalization schemes could be explored in future research.

¹⁰For robustness, I also calculate, and then similarly scale a multiplicative combination $S_{\text{salience}} = S_{\text{first}} \times S_{\text{freq}}$ which provides a higher penalty for low values on either dimension.

an entity, more recent work has demonstrated that sentiment conveys something sufficiently distinct from stance that the former cannot accurately proxy for the latter (Mohammad et al. 2016; Rauh 2018; Boukes et al. 2020; Chan et al. 2021; Küçük and Can 2021; van Atteveldt, van der Velden, and Boukes 2021; Bestvater and Monroe 2023). For example, consider the following hypothetical sentences:

1. *“The death and destruction in Ukraine is a unprecedented tragedy.”*
2. *“It is an unprecedented tragedy that individuals from both Russia and Ukraine have experience death and destruction in the ongoing conflict.”*
3. *“The war in Ukraine should end.”*

Sentences (1) and (2) both using similar emotionally charged language that would be coded similarly by sentiment analysis, but the stance toward Ukraine is positive in sentence (1) and neutral in sentence (2). Like sentence (1), sentence (3) also expressed a positive stance toward Ukraine, but less intensively so. This illustrates how emotional valence (sentiment) has difficulty capturing entity relations (stance) because it is not synonymous with support, opposition, or indifference. Similarly, measures of support, neutrality, or opposition toward a target entity (stance) produce only binary measures that do not capture the intensity of that expression (sentiment) and consequently cannot be easily aggregate across multiple text segments in a way that produces continuous values with measurement uncertainty. To remedy these limitations, I calculate affinity using a novel combination of sentiment and stance.

For *sentiment*, the evaluative language concerning a target state is processed using VADER, a well-established rule-based sentiment analyzer (Hutto and Gilbert 2014). In this context, ‘sentiment’ refers to the positive, negative, or neutral evaluative orientation of the language used, rather than an imputation of a collective emotional state to the entity issuing the document. VADER assigns an initial continuous score $x \in [-1, 1]$ for each text chunk mentioning a given target state.

Then, *stance* is measured by leveraging recent advances in few-shot learning and target-aware opinion classification to capture sentiment directed toward a distinct entity rather than the overall affect conveyed in a text chunk (Biber and Finegan 1988; Du Bois 2007; AlDayel and Magdy 2019). In the context of NSS documents, the issuing state conveys a stance toward a target state which may be distinct from the general sentiment being expressed in that particular document chunk. This form of stance detection is done using natural language inference (NLI), also known as textual entailment, which uses language models as pre-trained universal classifiers to identify the stance of a text toward a particular entity (Burnham 2024).

Stance is detected using a fine-tuned classification model defining a directional relationship between a text T and a hypothesis H about whether a human reading the text T would infer that hypothesis H is true (A. Wang et al. 2020; AlDayel and Magdy 2021; Burnham et al. 2024; Laurer et al. 2024).¹¹ In this context, the model calculates a stance classification as a Boolean response to the proposed hypothesis H that the issuing state’s stance toward the target state is positive, negative, or neutral. Given the Boolean output, the model is run separately for each hypothesis H (positive, negative, and neutral) and the classification with the highest probability is selected. These results produce a stance score $x \in \{-1, 0, 1\}$, for each target state in each chunk.

To return to the previous example, assume an NSS text chunk T states ‘The death and destruction in Ukraine is a unprecedented tragedy’. After the previous step in the data pipeline identified each target entity in the text chunk (in this case, only Ukraine), the language model is asked to classify the author’s stance toward the target entity Ukraine. The first step proposes the hypothesis H , ‘the author of this text expresses support or alliance with Ukraine’. Based

¹¹Further technical details are provided in the appendix but in brief, I employ a few-shot classification approach by fine-tuning an LLM (Political DEBATE, based on DeBERTa V3) on a small set of hand-coded document excerpts from the corpus through active learning (Burnham et al. 2024). The out-of-sample predictions for the fine-tuned model achieve a Matthews Correlation Coefficient of 0.925 and an accuracy of 0.962. For similar work using active learning to fine-tune models like RoBERTa to a specific local task, see Bonikowski, Luo, and Stuhler (2022) and Card et al. (2022).

on its reading of the text chunk, the model would classify the hypothesis as ‘true’ with a given probability. The classification is repeated with the negative and neutral hypotheses and from those three the one with the highest probability is selected.

This appropriately captures the semantic relationship between the expression and target in a directional manner composed of an evaluated recognizable objective H , an expression T containing a positive, negative or neutral evaluation, and a true/false relation that attributes the expressed valuation to the object in question (Fogel-Dror et al. 2019; Overbeck et al. 2025). Put more plainly, what this does – and sentiment analysis alone does not – is identify an expressed stance toward a target entity even when the general sentiment in that text is incongruent with the sentiment toward the specific target entity or when the general sentiment provides no information about the author’s stance toward the target entity.

The sentiment and stance detection values are then combined to produce the baseline measure of directed dyad affinity. First, I calibrate the sentiment scores for each target entity in each text chunk by aligning their valence with the predicted stance: support/alliance predictions become positive, threat/rival predictions become negative, and neutral predictions remain unchanged. This adjustment accounts for cases like the sample sentences (1) and (3) where emotionally charged words appear in contexts contrary to their typical connotation. I then scale the scores into three distinct ranges: negative stances are $x \in [-1, -0.5]$, neutral stances are $x \in (-0.5, 0.5)$, and positive stances are $x \in [0.5, 1]$. This scaling preserves the categorical stance information while allowing for differences in intensity within each category. The values are then averaged across all chunks for a given target state, producing a continuous value $x \in [-1, 1]$ reflecting both the direction and strength of the affinity.

I refine these baseline estimates of affinity by incorporating a gold-standard calibration process using Best-Worst Scaling (BWS), a survey method that produces reliable comparative rankings using human annotation (Finn and Louviere 1992; Louviere 2015; Kiritchenko and Mohammad

2016, 2017). In this BWS task, human coders were presented with randomized sets of four text chunks, each mentioning a specific target state. For each set, they were asked to identify the chunk expressing the most positive and the chunk expressing the least positive evaluative language toward the identified target state.¹² This method generates multiple pairwise comparisons, yielding reliable relative rankings of evaluative language across the corpus which are then used to calibrate the computationally derived affinity scores to the specific nuances of this unique corpus using design-based supervised learning (DSL). This method reduces potential biases produced by the initial general-purpose models and calibrates the affinity scores to the specific corpus in question (Egami et al. 2024b, 2024a).

Together, these innovative measurement strategies allow for a more comprehensive and nuanced analysis of how actors express relationships. The salience detection method provides a novel way to quantify the relative importance of different targets in an actor’s strategic discourse and the fusion of sentiment analysis with stance detection allows for a more fine-grained understanding of how actors express relationships beyond simple positive or negative categorizations.

4.3 Affective Salience

Finally, I combine the indices of salience and affinity to produce a single measure of *affective salience*: the intensity and valence of a state’s expressed posture toward another state. This measure is calculated as $S_{\text{affective}} = S_{\text{salience}} \times S_{\text{affinity}} \in [-1, 1]$ with higher values indicating higher affective salience.¹³ Put more concretely, this measure identifies the centrality conveyed about a target state *and* how the issuing state feels about the target state, with more extreme

¹²Coders were given the target state to code for the particular entry (as some text chunks mention more than one state) and instructed to make their evaluation using only information in the chunk. Further details on the BWS implementation and inter-coder reliability are provided in the Appendix.

¹³This multiplicative approach means that extremely high and low affective salience typically requires a target state to be both highly salient and strongly positive/negative in affinity. A low score on either dimension will commensurately reduce the overall affective salience.

values indicating more intense affective salience and values closer to 0 indicating the target state is relatively peripheral and neutral. This measure is particularly useful for identifying cases where a state is highly salient to another state but the affective valence of that relationship is low, or vice versa. For example, a state may mention another state frequently in its NSS documents but express negative sentiment toward that state, indicating that the relationship is characterized by intense rivalry or hostility. Conversely, a state may mention another state infrequently but express positive sentiment toward that state, indicating that the relationship is characterized by a middling degree of support or alliance.

5 Validation

5.1 Face Validity

To illustrate the face validity of the measures of salience and affinity, Figure 2 shows the scores for states mentioned in the U.S. NSS issued during the first Trump administration (2017) and the Biden administration (2022). The data are consistent with the first Trump administration’s “America First” strategy warning of competition with great powers like Russia and China and rogue states like Iran and North Korea as well as his dissatisfaction with U.S. allies in Europe (Landler and Sanger 2017). This is contrasted with President Biden’s 2022 NSS that identified US security priorities as long-term competition with China, an immediate threat from Russia, and a stronger emphasis on positive multilateral ties with traditional U.S. allies like the United Kingdom, India, and Japan (McDonald 2023). And although the affinity expressed toward Ukraine is positive in both documents, Ukraine’s salience in the 2022 NSS (0.77) is much higher than in the 2017 NSS (0.21), reflecting a shift in how the US communicated the importance of Ukraine in its national security discourse as well as the reality on U.S. involvement in the conflict.

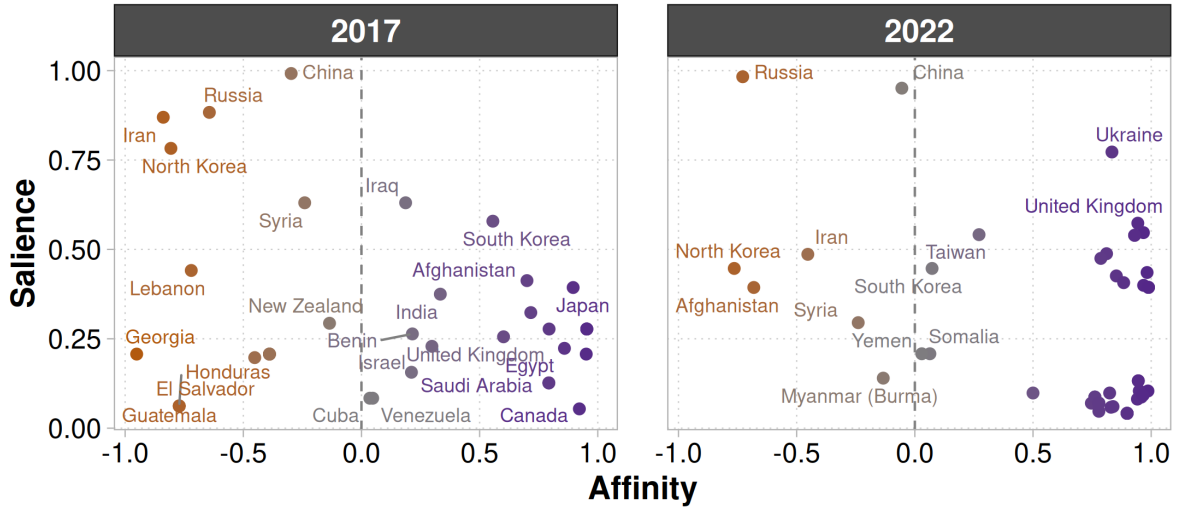


Figure 2: Affinity and saliency scores for states mentioned in the US NSS. Colors refer to each state’s value for affinity. Closely clustered states are unlabeled to simplify visual interpretation.

Figure 2 also visually demonstrates that while not strictly orthogonal, saliency and affinity measure different dimensions: states with high, medium, and low saliency scores can be found across the spectrum of affinity values.¹⁴ Notably, some states exhibit high or low affinity despite having a relatively low saliency score (e.g., values clustering near zero on the saliency y-axis but on the tails of the saliency y-axis), suggesting the US conveys strong alignment or animosity, but not much concern for the centrality of those states in its security policy.

As a second example of face validity that sheds light on the weights of the individual measurement components, Table 1 provides an illustrative example from Ukraine’s 2017 NSS of both measures as well as their inputs for the 10 highest saliency states. Not surprisingly, Russia is the most salient state in Ukraine’s NSS, as it is the second country Ukraine references (immediately after Belarus) and it is also mentioned much more frequently than any other country.

¹⁴In the full sample, the correlation between saliency and the absolute value of affinity is -0.02. The distribution of affect and saliency scores across the entire population is provided in the Appendix.

Table 1: Sample of measures for Ukraine (2017) showing the 10 countries with the highest salience. Higher values indicate higher values for all columns. The theoretical range of salience values is $[0, 1]$ and for affinity values is $[-1, 1]$.

Target Country	Salience			Affinity			Affective Salience
	First Mention	Entity Frequency	Salience	Sentiment	Stance	Affinity	
Russia	0.99	1.00	1.00	-0.02	-0.94	-0.78	-0.77
Poland	0.92	0.78	0.85	0.68	0.43	0.55	0.47
United Kingdom	0.97	0.61	0.79	0.70	0.36	0.53	0.42
United States	0.88	0.56	0.72	0.60	0.60	0.60	0.43
Lithuania	0.92	0.44	0.68	0.67	0.38	0.52	0.36
Belarus	1.00	0.28	0.64	0.51	0.40	0.45	0.29
Germany	0.77	0.44	0.61	0.31	0.50	0.40	0.25
Norway	0.86	0.33	0.60	0.88	0.83	0.86	0.51
Denmark	0.95	0.17	0.56	0.37	0.00	0.18	0.10
Netherlands	0.97	0.06	0.51	0.76	1.00	0.88	0.45

The U.S., by comparison, is in the 88th percentile of the first-mention input, and Ukraine mentions the U.S. approximately half as often as it mentions Russia. Ukraine expresses negative affinity toward Russia and positive affinity toward the U.S. and other European states with notable variation across the positive cases.

5.2 Construct Validity

Second, I validate the measures by showing their correlation with existing measures of salience and affinity. Theoretical expectations posit that states should signal more positive affinity toward states with which they share common security interests, as arrangements like defense alliances are, in part, a means to “document, advertise or encourage such an affinity” (Gartzke and Weisiger 2013, 25). I calculate the distribution of the NSS affinity index across three categorical measures of alignment: formal defense pact membership (Leeds et al. 2002), strategic rivalries (Thompson, Sakuwa, and Suhas 2021), and peace levels (Diehl, Goertz, and Gallegos 2021).

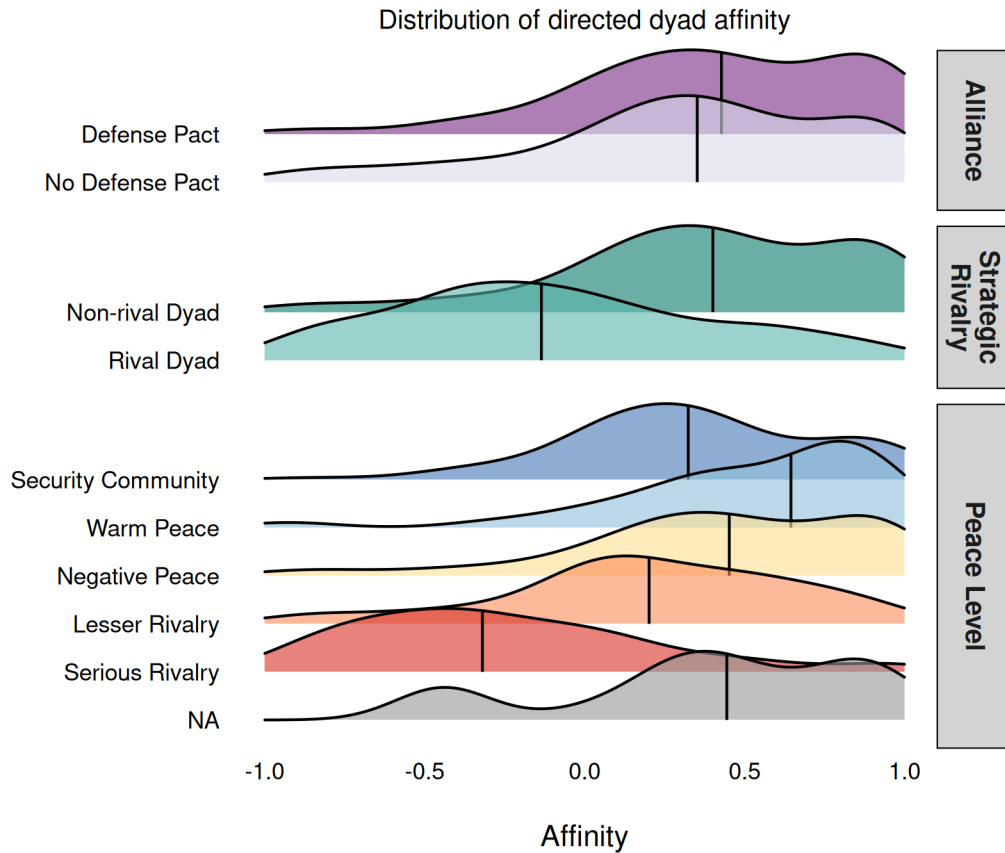


Figure 3: Distribution of the NSS affinity index across varying levels of existing measures of affinity. Vertical lines represent median values. Dyads without significant and ongoing interactions are coded in the peace level data as NA.

The results in Figure 3 show dyads in an alliance have, on average, a higher NSS affinity than dyads without a defense pact, and that dyads composed of strategic rivals have a lower NSS affinity than dyads that are not strategic rivals. Interestingly, the five ordered categories in the peace level data show a clear and consistent positive correlation between peace levels and NSS affinity *until* the highest peace level – ‘security community’ – which has a median NSS affinity higher than ‘negative peace’ but lower than ‘lesser rivalry’. Many of the security

community dyads are composed of states with shared EU membership that otherwise have little active or relevant security cooperation with one another, and may consequently have less intensely positive expressions toward each other in their NSS documents. Alternatively, deeply integrated security communities with highly institutionalized security cooperation may be largely taken for granted, lessening the perceived need to performatively emphasize positive affinity and reserving stronger positive framings for relationships where deeper cooperation is being more actively cultivated (Gannon and Kent 2021).

Next, I identify the correlation between signals of affinity in NSS documents and the most well-regarded measures of foreign policy similarity: kappa corrected s-score measures of alliance portfolio similarity and UN voting patterns.¹⁵ Conveyed affinity have a positive and statistically significant correlation with both measures (alliance portfolios = 0.14 and UN voting = 0.09).¹⁶ For comparison, the correlation between alliance and UN voting kappa corrected s-scores is 0.39. States publicly signal warmth toward others in patterns similar, but not identical to, what one would assume given similarity of defense partners and UN General Assembly voting patterns.

6 Application

I model the association between affective salience and two well-regarded indicators of higher intensity interstate engagement: the onset of international crises and the frequency of diplomatic visits. Using two different dependent variables that capture distinct types of state interaction

¹⁵I use the kappa chance-corrected versions of the s-scores (Cohen 1960; Häge 2011), but a full table of correlations using pi-corrected, binary, or weighted un-corrected measures as well as using COW rather than ATOP alliance data is provided in the appendix.

¹⁶These correlations are calculated for directed dyad-years where the salience of State B > 0. Dyads with no mention (and thus no NSS affinity score) are necessarily excluded from this specific comparison, as non-mention primarily speaks to a lack of expressed salience rather than the affective tone of a relationship that is discussed.

– one conflictual and the other cooperative – produces a symmetrical test of whether the relational logic signaled in NSS documents predicts future behavior. I use the directed dyad-year as the unit of analysis, lagging affective salience by a year. The crisis onset variable is a binary indicator drawn from the International Crisis Behavior (ICB) dataset identifying whether the issuing state and the target state are on opposing sides of a newly initiated international crisis.¹⁷ Diplomatic visits is measured as the annual count of official high-level visits between heads of state or government using the Country and Organizational Leader Travel (COLT) dataset (Moyer et al. 2025).

I fit logistic regression models to predict crisis onset and Poisson regression models to predict diplomatic visits. All models include control variables associated with each outcome in existing literature: CINC ratios of military power for crisis onset (Singer 1988), whether either state is a great power for diplomatic visits (Lebovic and Saunders 2016), and capital to capital distance and shared regime type for both (Markowitz and Fariss 2018). To benchmark the model performance, I also compare the models with affective salience to other models instead using kappa-corrected alliance portfolio and UN voting s-scores.

The results, presented in Table 2, demonstrate a statistically significant association between affective salience and future high intensity state interactions.¹⁸ Dyads with more positive affective salience are less likely to experience a subsequent crisis onset (Model 1) and engage in more high-level diplomacy (Model 4). Though relatively low-cost, these public performances convey meaningful information about how states anticipate and structure their international relationships.

In contrast, the traditional alignment measures perform unevenly. Alliance portfolio similar-

¹⁷A crisis is defined as an international event with hostile interactions and a heightened probability of military escalation (Brecher and Wilkenfeld 1997, 4–5; Brecher et al. 2023).

¹⁸Coefficients for the control variables are omitted here, but provided in the appendix, to avoid misinterpretation since confounders for the controls variables are not included in the model (Keele, Stevenson, and Elwert 2020; Dworschak 2023).

Table 2: Coefficient estimates for regression models.

	Crisis onset			Diplomatic visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Affective salience	0.006** (0.005)			1.306** (0.070)		
Alliance portfolio s-score		2.567 (1.631)			0.581** (0.050)	
UN voting s-score			0.040** (0.040)			0.037 (0.044)
Num.Obs.	8790	8790	8790	8528	8528	8528
AIC	242.8	275.2	263.1	15979.6	16185.1	16319.6
BIC	278.2	310.6	298.5	16014.9	16220.4	16354.9
RMSE	0.05	0.05	0.05	0.96	0.97	0.97
R2 tjur	0.07	0.02	0.04			
R2 Nagelkerke				0.36	0.34	0.32

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Models 1-3 are logistic regression with odds ratios. Models 4-6 are Poisson regression models. Full model results including control variables are provided in the appendix.

ity is positively associated with future diplomatic visits, but not significantly predictive of future crisis participation. And UN voting similarity is a statistically significant predictor of the likelihood of an impending international crisis, but shows no statistically significant association with future diplomatic visits. Only affective salience performs consistently in predicting both cooperative and confrontational state interactions. Additionally, the models with affective salience demonstrate higher overall model fit demonstrated by lower Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values than those relying on the more traditional measures, indicating a better balance of explanatory power and parsimony. Similarly, the root mean squared error (RMSE) values for affective salience are equivalent or marginally lower, suggesting more accurate predictions of observed values, and the model-specific R^2 value are comparatively higher which indicates improved classification performance.

Additional models provided in the appendix confirm that these results hold across alternative specifications. I include bivariate models that avoid post-treatment bias and over-adjustment (Clarke, Kenkel, and Rueda 2018; Montgomery, Nyhan, and Torres 2018; Dworschak 2023; Hünermund and Louw 2025) as well as models with country, dyad, and year fixed effects to account for unobserved spatial and temporal heterogeneity. Consistent results are found in models with alternate dependent variable measures like Militarized Interstate Dispute (MID) participation and diplomatic visits to and from the issuing country. The same is true in models using panel-corrected standard errors (Beck and Katz 1995; Chiang and Sasaki 2023), lagged dependent variables to adjust for temporal autocorrelation, and zero-inflated Poisson models to account for the excess zeros in diplomatic visits.

7 Conclusion

On the empirical and methodological front, the application of state-of-the-art computational techniques to novel text data offers a robust framework for measuring the signaling of relational hierarchies. I provide a unique form of evidence about who states want to publicly convey they are thinking about and how. Existing indicators of foreign policy similarity, formal alliance ties, and rivalries are consistent with how states themselves signal these affinities, even if the traditional indicators offer coarser granularity. And finally, novel empirical results suggest that the affective salience signaled in NSS rhetoric is a reliable indicator of future interstate behavior, particularly in the context of international crises and diplomacy.

These data also offer a valuable opportunity to explore how states justify signals of affinity, as such variation can be indicative of the perceived costs of hypocrisy (McManus and Yarhi-Milo 2017) or signals for a desire to deepen cooperation (Gannon and Kent 2021). Additionally, the data may reveal whether individual heads of state influence the tone of bureaucratic documents (Lupton 2020, 2024) and whether public opinion predates or follows from official public statements about other states (Page and Shapiro 1983; Matush 2023), thereby linking elite rhetoric and public opinion to broader patterns of international behavior. Further work could extend the analysis to incorporate internal policy documents like President’s Daily Briefs (Goldfien and Joseph 2023; Carson, Min, and Nuys 2024) or legislative debates (Kim, Londregan, and Ratkovic 2018), thereby examining the divergence between public signaling and private deliberations (McManus 2016), or audience-specific public statements like diplomatic communiqués (Connelly et al. 2021), international agreements (Baturu, Dasandi, and Mikhaylov 2017), and diplomatic speeches in forums like the UN General Assembly (Finke and Risse 2024; Mitra 2024; Turco 2024).

Beyond international relations, this research contributes to wider debates about the relation-

ship between public rhetoric and political behavior, demonstrating how computational text analysis can bridge the gap between stated intentions and observable actions across domains of political science. While I here theorize only affective salience, NSS documents and others like them contain a vast array of other information that has been analyzed in other contexts like the use (or non-use) of emotional and moral language (Hall 2015; Post 2023; Yoon 2025) and different rhetorical devices like persuasion (Walsh 2005; Krebs and Jackson 2007) and expressions of resolve (Trager 2010; McManus 2017).

The utility and consequences of LLM models to social science questions and methods is, and will likely remain, the subject of ongoing discussion (Bisbee et al. 2024; Laurer et al. 2024; Y. Wang 2024; Caballero and Jenkins 2025; Leo et al. 2025; Mens and Gallego 2025; Ornstein, Blasingame, and Truscott 2025; Thapa et al. 2025). My approach expands upon recent applications in political science of similar models for classification (Burnham et al. 2024), dictionary creation (Häffner et al. 2023), event extraction (Arslan, Munawar, and Cruz 2024), named entity recognition (Balluff, Boomgaarden, and Waldherr 2024), and stance detection (Burnham 2024). Given the exciting speed at which advances in these techniques are occurring, I have no doubt significant extensions will be feasible by the time this article is in print.

References

- AlDayel, Abeer, and Walid Magdy. 2019. "Assessing Sentiment of the Expressed Stance on Social Media." In *Social Informatics: 11th International Conference, SocInfo 2019, Doha, Qatar, November 18–21, 2019, Proceedings*, 277–86. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-030-34971-4_19.
- . 2021. "Stance Detection on Social Media: State of the Art and Trends." arXiv. <https://doi.org/10.48550/arXiv.2006.03644>.
- Alley, Joshua. 2021. "Alliance Participation, Treaty Depth, and Military Spending." *International Studies Quarterly* 65 (4): 929–43. <https://doi.org/10.1093/isq/sqab077>.
- Arslan, Muhammad, Saba Munawar, and Christophe Cruz. 2024. "Political-RAG: Using Generative AI to Extract Political Information from Media Content." *Journal of Information Technology & Politics* 0 (0): 1–16. <https://doi.org/10.1080/19331681.2024.2417263>.
- Asgarieh, Eliyar, Kapil Thadani, and Neil O'Hare. 2024. "Scalable Detection of Salient Entities in News Articles." arXiv. <https://doi.org/10.48550/arXiv.2405.20461>.
- Bailey, Michael A., Anton Strezhnev, and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* 61 (2): 430–56. <https://doi.org/10.1177/0022002715595700>.
- Balluff, Paul, Hajo G. Boomgaarden, and Annie Waldherr. 2024. "Automatically Finding Actors in Texts: A Performance Review of Multilingual Named Entity Recognition Tools." *Communication Methods and Measures* 18 (4): 371–89. <https://doi.org/10.1080/19312458.2024.2324789>.
- Baturo, Alexander, Niheer Dasandi, and Slava J. Mikhaylov. 2017. "Understanding State Preferences with Text as Data: Introducing the UN General Debate Corpus." *Research & Politics* 4 (2): 1–9. <https://doi.org/10.1177/2053168017712821>.
- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-

- Series Cross-Section Data.” *The American Political Science Review* 89 (3): 634–47. <https://doi.org/10.2307/2082979>.
- Becker, Jordan, Haemin Jee, Andreea Budeanu, Maxwell Love, and Seth Benson. 2024. “European Strategies in the Shadow of Sino-American Competition: A Text-as-Data Approach.” In *United States Military Academy, West Point*. Rochester, NY. <https://doi.org/10.2139/ssrn.4758032>.
- Becker, Jordan, and Edmund Malesky. 2017. “The Continent or the ‘Grand Large’? Strategic Culture and Operational Burden-Sharing in NATO.” *International Studies Quarterly* 61 (1): 163–80. <https://doi.org/10.1093/isq/sqw039>.
- Bennett, D. Scott. 2006. “Exploring Operationalizations of Political Relevance.” *Conflict Management and Peace Science* 23 (3): 245–61. <https://doi.org/10.1080/07388940600837748>.
- Benson, Michelle A. 2005. “The Relevance of Politically Relevant Dyads in the Study of Interdependence and Dyadic Disputes.” *Conflict Management and Peace Science* 22 (2): 113–33. <https://doi.org/10.1080/07388940590948556>.
- Bestvater, Samuel E., and Burt L. Monroe. 2023. “Sentiment Is Not Stance: Target-Aware Opinion Classification for Political Text Analysis.” *Political Analysis* 31 (2): 235–56. <https://doi.org/10/g72d6x>.
- Bhowmik, Rajarshi, Marco Ponza, Atharva Tendle, Anant Gupta, Rebecca Jiang, Xingyu Lu, Qian Zhao, and Daniel Preotiuc-Pietro. 2024. “Leveraging Contextual Information for Effective Entity Salience Detection.” arXiv. <https://doi.org/10.48550/arXiv.2309.07990>.
- Biber, Douglas, and Edward Finegan. 1988. “Adverbial Stance Types in English.” *Discourse Processes* 11 (1): 1–34. <https://doi.org/10.1080/01638538809544689>.
- Bisbee, James, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson. 2024. “Synthetic Replacements for Human Survey Data? The Perils of Large Language Models.” *Political Analysis* 32 (4): 401–16. <https://doi.org/10.1017/pan.2024.5>.
- Bonikowski, Bart, Yuchen Luo, and Oscar Stuhler. 2022. “Politics as Usual? Measuring

- Populism, Nationalism, and Authoritarianism in U.S. Presidential Campaigns (1952–2020) with Neural Language Models.” *Sociological Methods & Research* 51 (4): 1721–87. <https://doi.org/10/grzg7s>.
- Boukes, Mark, Bob van de Velde, Theo Araujo, and Rens Vliegthart. 2020. “What’s the Tone? Easy Doesn’t Do It: Analyzing Performance and Agreement Between Off-the-Shelf Sentiment Analysis Tools.” *Communication Methods and Measures* 14 (2): 83–104. <https://doi.org/10.1080/19312458.2019.1671966>.
- Braumoeller, Bear F., and Austin Carson. 2011. “Political Irrelevance, Democracy, and the Limits of Militarized Conflict.” *Journal of Conflict Resolution* 55 (2): 292–320. <https://doi.org/10.1177/0022002710393918>.
- Brecher, Michael, and Jonathan Wilkenfeld. 1997. *A Study of Crisis*. Ann Arbor: University of Michigan Press.
- Brecher, Michael, Jonathan Wilkenfeld, Kyle C. Beardsley, Patrick James, and David Quinn. 2023. “International Crisis Behavior Data Codebook.” Codebook Version 15.
- Bremer, Stuart A. 1992. “Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965.” *Journal of Conflict Resolution* 36 (2): 309–41. <https://www.jstor.org/stable/174478>.
- Bullough, Benjamin, Harrison Lundberg, Chen Hu, and Weihang Xiao. 2024. “Predicting Entity Salience in Extremely Short Documents.” In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, edited by Franck Dernoncourt, Daniel Preotiuc-Pietro, and Anastasia Shimorina, 50–64. Miami, Florida, US: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-industry.5>.
- Burke, Kenneth. 1941. *The Philosophy of Literary Form*. Berkeley: Univ of California Press.
- Burnham, Michael. 2024. “Stance Detection: A Practical Guide to Classifying Political Beliefs in Text.” *Political Science Research and Methods*, 1–18. <https://doi.org/10.1017/>

psrm.2024.35.

- Burnham, Michael, Kayla Kahn, Ryan Yank Wang, and Rachel X. Peng. 2024. “Political DEBATE: Efficient Zero-shot and Few-shot Classifiers for Political Text.” arXiv. <https://doi.org/10.48550/arXiv.2409.02078>.
- Caballero, William N., and Phillip R. Jenkins. 2025. “On Large Language Models in National Security Applications.” *Stat* 14 (2): e70057. <https://doi.org/10.1002/sta4.70057>.
- Card, Dallas, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. “Computational Analysis of 140 Years of US Political Speeches Reveals More Positive but Increasingly Polarized Framing of Immigration.” *Proceedings of the National Academy of Sciences* 119 (31): e2120510119. <https://doi.org/10/gqkxwz>.
- Carson, Austin, Eric Min, and Maya Van Nuys. 2024. “Racial Tropes in the Foreign Policy Bureaucracy: A Computational Text Analysis.” *International Organization* 78 (2): 189–223. <https://doi.org/10/gvhs28>.
- Caudle, Sharon L. 2009. “National Security Strategies: Security from What, for Whom, and by What Means.” *Journal of Homeland Security and Emergency Management* 6 (1): 1–26. <https://doi.org/10/fgnxvr>.
- Chan, Chung-hong, Joseph Bajjalieh, Loretta Auvil, Hartmut Wessler, Scott Althaus, Kasper Welbers, Wouter van Atteveldt, and Marc Jungblut. 2021. “Four Best Practices for Measuring News Sentiment Using ‘Off-the-Shelf’ Dictionaries: A Large-Scale p-Hacking Experiment.” *Computational Communication Research* 3 (1): 1–27.
- Chiang, Harold D., and Yuya Sasaki. 2023. “On Using The Two-Way Cluster-Robust Standard Errors.” arXiv. <https://doi.org/10.48550/arXiv.2301.13775>.
- Chiba, Daina, Jesse C. Johnson, and Brett Ashley Leeds. 2015. “Careful Commitments: Democratic States and Alliance Design.” *The Journal of Politics* 77 (4): 968–82. <https://doi.org/10.1086/682074>.

- Chin, John J., Kiron K. Skinner, and Clay Yoo. 2023. "Understanding National Security Strategies Through Time." *Texas National Security Review* 6 (4): 103–24. <https://doi.org/10.26153/tsw/48842>.
- Clarke, Kevin A., Brenton Kenkel, and Miguel R. Rueda. 2018. "Omitted Variables, Countervailing Effects, and the Possibility of Overadjustment." *Political Science Research and Methods* 6 (2): 343–54. <https://doi.org/10/gf5hbn>.
- Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37–46. <https://doi.org/10/dghsrr>.
- Connelly, Matthew J., Raymond Hicks, Robert Jervis, Arthur Spirling, and Clara H Suong. 2021. "Diplomatic Documents Data for International Relations: The Freedom of Information Archive Database." *Conflict Management and Peace Science* 38 (6): 762–81. <https://doi.org/10.1177/0738894220930326>.
- Correlates of War Project. 2017. "State System Membership List, V2016."
- Cranmer, Skyler J., and Bruce A. Desmarais. 2016. "A Critique of Dyadic Design." *International Studies Quarterly* 60 (2): 355–62. <https://doi.org/10.1093/isq/sqw007>.
- Denny, Matthew J., and Arthur Spirling. 2018. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26 (2): 168–89. <https://doi.org/10/gdjsqk>.
- Diehl, Paul F., Gary Goertz, and Yahve Gallegos. 2021. "Peace Data: Concept, Measurement, Patterns, and Research Agenda." *Conflict Management and Peace Science* 38 (5): 605–24. <https://doi.org/10.1177/0738894219870288>.
- Dorff, Cassy, and Michael D. Ward. 2013. "Networks, Dyads, and the Social Relations Model." *Political Science Research and Methods* 1 (2): 159–78. <https://doi.org/10.1017/psrm.2013.21>.
- Du Bois, John W. 2007. "The Stance Triangle." In *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, edited by Robert Englebretson, 139–82. Amsterdam/Philadelphia:

John Benjamins Publishing.

DuMont, Malia. 2019. “Elements of National Security Strategy.” Washington, DC: Atlantic Council.

Dunietz, Jesse, and Daniel Gillick. 2014. “A New Entity Salience Task with Millions of Training Examples.” In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2: Short Papers*, edited by Shuly Wintner, Stefan Riezler, and Sharon Goldwater, 205–9. Gothenburg, Sweden: Association for Computational Linguistics. <https://doi.org/10.3115/v1/E14-4040>.

Dworschak, Christoph. 2023. “Bias Mitigation in Empirical Peace and Conflict Studies: A Short Primer on Posttreatment Variables.” *Journal of Peace Research* 61 (3): 462–76. <https://doi.org/10.1177/00223433221145531>.

Egami, Naoki, Musashi Hinck, Brandon M. Stewart, and Hanying Wei. 2024a. “Using Large Language Model Annotations for the Social Sciences: A General Framework of Using Predicted Variables in Downstream Analyses.”

———. 2024b. “Using Imperfect Surrogates for Downstream Inference: Design-based Supervised Learning for Social Science Applications of Large Language Models.” arXiv. <https://doi.org/10.48550/arXiv.2306.04746>.

Ettinger, Aaron. 2017. “U.S. National Security Strategies: Patterns of Continuity and Change, 1987–2015.” *Comparative Strategy* 36 (2): 115–28. <https://doi.org/10.1080/01495933.2017.1317538>.

Fearon, James D. 1994. “Signaling Versus the Balance of Power and Interests An Empirical Test of a Crisis Bargaining Model.” *Journal of Conflict Resolution* 38 (2): 236–69. <https://doi.org/10.1177/0022002794038002004>.

Finke, Daniel, and Tobias Risse. 2024. “Multidimensional Conflicts over Disarmament and International Security: Analyzing Speeches in the First Committee of the UN General Assembly.” *Political Science Research and Methods*, 1–18. <https://doi.org/10.1017/psrm>.

2024.56.

- Finn, Adam, and Jordan J. Louviere. 1992. "Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety." *Journal of Public Policy & Marketing* 11 (2): 12–25.
- Fogel-Dror, Yair, Shaul R. Shenhav, Tamir Sheaffer, and Wouter Van Atteveldt. 2019. "Role-Based Association of Verbs, Actions, and Sentiments with Entities in Political Discourse." *Communication Methods and Measures* 13 (2): 69–82. <https://doi.org/10.1080/19312458.2018.1536973>.
- Fuhrmann, Matthew, and Todd S. Sechser. 2014. "Signaling Alliance Commitments: Hand-Tying and Sunk Costs in Extended Nuclear Deterrence." *American Journal of Political Science* 58 (4): 919–35. <https://doi.org/10.1111/ajps.12082>.
- Gannon, J Andrés. 2025. "Allies as Armaments: Explaining the Specialization of State Military Capabilities." *Security Studies* 34 (1): 28–58. <https://doi.org/10.1080/09636412.2025.2467140>.
- Gannon, J Andrés, and Daniel Kent. 2021. "Keeping Your Friends Close, but Acquaintances Closer: Why Weakly Allied States Make Committed Coalition Partners." *Journal of Conflict Resolution* 65 (5): 889–918. <https://doi.org/10.1177/0022002720978800>.
- Gartzke, Erik A. 1998. "Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace." *American Journal of Political Science* 42 (1): 1–27. <https://doi.org/10.2307/2991745>.
- Gartzke, Erik A., Shannon Carcelli, J Andrés Gannon, and Jiakun Jack Zhang. 2017. "Signaling in Foreign Policy." *Oxford Encyclopedia of Foreign Policy Analysis*.
- Gartzke, Erik A., and Kristian Skrede Gleditsch. 2022. "Ties That Bias in International Conflict: A Spatial Approach to Dyadic Dependence from Alliance Ties and Inbetweenness." *International Studies Quarterly* 66 (1): 1–13. <https://doi.org/10/gtvhj>.
- Gartzke, Erik A., and Alex Weisiger. 2013. "Fading Friendships: Alliances, Affinities and the

- Activation of International Identities.” *British Journal of Political Science* 43 (1): 25–52. <https://doi.org/10.1017/S0007123412000208>.
- Gibler, Douglas M. 2009. *International Military Alliances, 1648-2008*. Washington, DC: CQ Press.
- Gleditsch, Kristian Skrede, and Michael D. Ward. 2001. “Measuring Space: A Minimum-Distance Database and Applications to International Studies.” *Journal of Peace Research* 38 (6): 739–58. <https://www.jstor.org/stable/425562>.
- Goldfien, Michael A., and Michael F. Joseph. 2023. “Perceptions of Leadership Importance: Evidence from the CIA’s President’s Daily Brief.” *Security Studies* 32 (2): 205–38. <https://doi.org/10.1080/09636412.2023.2200203>.
- Grönberg, David. 2021. “Extracting Salient Named Entities from Financial News Articles.” Master’s thesis, Linköping University.
- Häffner, Sonja, Martin Hofer, Maximilian Nagl, and Julian Walterskirchen. 2023. “Introducing an Interpretable Deep Learning Approach to Domain-Specific Dictionary Creation: A Use Case for Conflict Prediction.” *Political Analysis* 31 (4): 481–99. <https://doi.org/10/gtkcqc>.
- Häge, Frank M. 2011. “Choice or Circumstance? Adjusting Measures of Foreign Policy Similarity for Chance Agreement.” *Political Analysis* 19 (3): 287–305. <https://www.jstor.org/stable/23011439>.
- Häge, Frank M., and Simon Hug. 2016. “Consensus Decisions and Similarity Measures in International Organizations.” *International Interactions* 42 (3): 503–29. <https://doi.org/10/gtv3zr>.
- Hall, Todd H. 2015. *Emotional Diplomacy: Official Emotion on the International Stage*. Ithaca: Cornell University Press.
- Hünemann, Paul, and Beyers Louw. 2025. “On the Nuisance of Control Variables in Causal Regression Analysis.” *Organizational Research Methods* 28 (1): 138–51. <https://doi.org/10.1177/10944281231219274>.

- Hutto, C., and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8 (1): 216–25. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- Keele, Luke, Randolph T. Stevenson, and Felix Elwert. 2020. "The Causal Interpretation of Estimated Associations in Regression Models." *Political Science Research and Methods* 8 (1): 1–13. <https://doi.org/10.1017/psrm.2019.31>.
- Kim, In Song, John Londregan, and Marc Ratkovic. 2018. "Estimating Spatial Preferences from Votes and Text." *Political Analysis* 26 (2): 210–29. <https://doi.org/10.1017/pan.2018.7>.
- Kinne, Brandon J. 2020. "The Defense Cooperation Agreement Dataset (DCAD)." *Journal of Conflict Resolution* 64 (4): 729–55. <https://doi.org/10.1177/0022002719857796>.
- Kiritchenko, Svetlana, and Saif M. Mohammad. 2016. "Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best–Worst Scaling." In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, edited by Kevin Knight, Ani Nenkova, and Owen Rambow, 811–17. San Diego, California: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N16-1095>.
- . 2017. "Best-Worst Scaling More Reliable Than Rating Scales: A Case Study on Sentiment Intensity Annotation." In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, edited by Regina Barzilay and Min-Yen Kan, 465–70. Vancouver, Canada: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P17-2074>.
- Krebs, Ronald R. 2015. *Narrative and the Making of US National Security*. Cambridge Studies in International Relations. Cambridge: Cambridge University Press.
- Krebs, Ronald R., and Patrick Thaddeus Jackson. 2007. "Twisting Tongues and Twisting Arms: The Power of Political Rhetoric." *European Journal of International Relations* 13

- (1): 35–66. <https://doi.org/10.1177/1354066107074284>.
- Küçük, Dilek, and Fazli Can. 2021. “Stance Detection: Concepts, Approaches, Resources, and Outstanding Issues.” In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2673–76. SIGIR ’21. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3404835.3462815>.
- Kydd, Andrew. 2000. “Arms Races and Arms Control: Modeling the Hawk Perspective.” *American Journal of Political Science* 44 (2): 228–44. <https://doi.org/10/cbfzcb>.
- Lahoti, Preethi, Kiran Garimella, and Aristides Gionis. 2018. “Joint Non-negative Matrix Factorization for Learning Ideological Leaning on Twitter.” In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 351–59. WSDM ’18. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3159652.3159669>.
- Landler, Mark, and David E. Sanger. 2017. “Trump Delivers a Mixed Message on His National Security Approach.” *The New York Times*, December, A1.
- Laurer, Moritz, Wouter van Atteveldt, Andreu Casas, and Kasper Welbers. 2024. “On Measurement Validity and Language Models: Increasing Validity and Decreasing Bias with Instructions.” *Communication Methods and Measures* 19 (1): 46–62. <https://doi.org/10.1080/19312458.2024.2378690>.
- Lebovic, James H., and Elizabeth N. Saunders. 2016. “The Diplomatic Core: The Determinants of High-Level US Diplomatic Visits, 1946–2010.” *International Studies Quarterly* 60 (1): 107–23. <https://doi.org/10.1093/isq/sqv008>.
- Lee, Michael J., and William R. Thompson. 2017. “Major Powers Vs. Global Powers: A New Measure of Global Reach and Power Projection Capacity.” *Oxford Research Encyclopedia of Politics*. <https://oxfordre.com/politics/politics/politics/display/10.1093/acrefore/9780190228637.001.0001/9780190228637-e-610>. <https://doi.org/10.1093/acrefore/9780190228637.013.610>.
- Leeds, Brett Ashley, Jeffrey Ritter, Sara Mitchell, and Andrew Long. 2002. “Alliance Treaty

- Obligations and Provisions, 1815-1944.” *International Interactions* 28 (3): 237–60. <https://doi.org/10.1080/03050620213653>.
- Lemke, Douglas. 1995. “The Tyranny of Distance: Redefining Relevant Dyads.” *International Interactions* 21 (1): 23–38. <https://doi.org/10.1080/03050629508434858>.
- Lemke, Douglas, and William Reed. 2001. “The Relevance of Politically Relevant Dyads.” *Journal of Conflict Resolution* 45 (1): 126–44. <https://doi.org/10.1177/0022002701045001006>.
- Leo, Riccardo Di, Chen Zeng, Elias Dinas, and Reda Tamtam. 2025. “Mapping (A)Ideology: A Taxonomy of European Parties Using Generative LLMs as Zero-Shot Learners.” *Political Analysis* 33 (4): 456–63. <https://doi.org/10.1017/pan.2025.7>.
- Louviere, Jordan J. 2015. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge: Cambridge University Press.
- Lupton, Danielle L. 2020. *Reputation for Resolve: How Leaders Signal Determination in International Politics*. Ithaca: Cornell University Press.
- . 2024. “Threat Construction and Coercive Credibility.” *Security Studies* 33 (3): 407–38. <https://doi.org/10.1080/09636412.2024.2307058>.
- Maoz, Zeev. 1996. *Domestic Sources of Global Change*. Ann Arbor: University of Michigan Press. <https://doi.org/10.3998/mpub.14663>.
- . 2010. *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816–2001*. Cambridge University Press.
- Maoz, Zeev, and Bruce Russett. 1993. “Normative and Structural Causes of Democratic Peace, 1946-1986.” *The American Political Science Review* 87 (3): 624.
- Markowitz, Jonathan N., and Christopher J. Fariss. 2018. “Power, Proximity, and Democracy: Geopolitical Competition in the International System.” *Journal of Peace Research* 55 (1): 78–93. <https://doi.org/10.1177/0022343317727328>.
- Mattes, Michaela. 2012. “Reputation, Symmetry, and Alliance Design.” *International Orga-*

- nization* 66 (4): 679–707. <https://doi.org/10.1017/S002081831200029X>.
- Matush, Kelly. 2023. “Harnessing Backlash: How Leaders Can Benefit from Antagonizing Foreign Actors.” *British Journal of Political Science* 53 (3): 902–18. <https://doi.org/10.1017/S0007123422000370>.
- McDonald, Gloria. 2023. “Defining National Security: A Comparison of the Trump and Biden Administration’s National Security Strategies.” Research Report. Washington DC: America First Policy Institute.
- McKie, Jorj. 2024. “Pymupdf4llm: PyMuPDF Utilities for LLM/RAG.”
- McManus, Roseanne W. 2016. “The Impact of Context on the Ability of Leaders to Signal Resolve.” *International Interactions* 43 (3): 453–79. <https://doi.org/10.1080/03050629.2016.1171764>.
- . 2017. *Statements of Resolve: Achieving Coercive Credibility in International Conflict*. Cambridge University Press.
- McManus, Roseanne W., and Keren Yarhi-Milo. 2017. “The Logic of ‘Offstage’ Signaling: Domestic Politics, Regime Type, and Major Power-Protégé Relations.” *International Organization* 71 (4): 701–33. <https://doi.org/10.1017/S0020818317000297>.
- Mens, Gaël Le, and Aina Gallego. 2025. “Positioning Political Texts with Large Language Models by Asking and Averaging.” *Political Analysis* 33 (3): 274–82. <https://doi.org/10.1017/pan.2024.29>.
- Min, Eric. 2025. *Words of War: Negotiation as a Tool of Conflict*. Cornell University Press.
- Mitra, Alessio. 2024. “Words That Matter: A Machine Learning Analysis of United Nations General Assembly Speeches and Their Influence on Aid Allocation.” *Research & Politics* 11 (3): 20531680241272626. <https://doi.org/10/gt82pb>.
- Miyajiwala, Aamir, Arnav Ladkat, Samiksha Jagadale, and Raviraj Joshi. 2022. “On Sensitivity of Deep Learning Based Text Classification Algorithms to Practical Input Perturbations.” arXiv. <https://doi.org/10.48550/arXiv.2201.00318>.

- Mohammad, Saif M., Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. "SemEval-2016 Task 6: Detecting Stance in Tweets." In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, edited by Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, and Torsten Zesch, 31–41. San Diego, California: Association for Computational Linguistics. <https://doi.org/10.18653/v1/S16-1003>.
- Mohr, John W., Robin Wagner-Pacifici, Ronald L. Breiger, and Petko Bogdanov. 2013. "Graphing the Grammar of Motives in National Security Strategies: Cultural Interpretation, Automated Text Analysis and the Drama of Global Politics." *Poetics* 41 (6): 670–700. <https://doi.org/10/gd23f6>.
- Montani, Ines, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem, and Henning Peters. 2023. "Explosion/spaCy: V3.7.2: Fixes for APIs and Requirements." Zenodo. <https://doi.org/10.5281/ZENODO.1212303>.
- Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62 (3): 760–75. <https://www.jstor.org/stable/26598780>.
- Moyer, Jonathan D, Collin J Meisel, Adam Szymanski-Burgos, Andrew C Scott, Matteo C M Casiraghi, Alexandra Kurkul, Marianne Hughes, Whitney Kettlun, Kylie X McKee, and Austin S Matthews. 2025. "When Heads of Government and State (HOGS) Fly: Introducing the Country and Organizational Leader Travel (COLT) Dataset Measuring Foreign Travel by HOGS." *International Studies Quarterly* 69 (2): sqaf013. <https://doi.org/10.1093/isq/sqaf013>.
- Neal, Andrew W., and Roy B. Gardner. 2025. "National Security and Defence Documents Dataset (1987-2024) V2.0." University of Edinburgh. School of Social and Political Science. Politics and International Relations. <https://doi.org/10.7488/ds/7904>.
- Organski, A. F. K., and Jacek Kugler. 1980. *The War Ledger*. Chicago: University of Chicago

Press.

- Ornstein, Joseph T., Elise N. Blasingame, and Jake S. Truscott. 2025. “How to Train Your Stochastic Parrot: Large Language Models for Political Texts.” *Political Science Research and Methods* 13 (2): 264–81. <https://doi.org/10.1017/psrm.2024.64>.
- Overbeck, Maximilian, Christian Baden, Tali Aharoni, Eedan Amit-Danhi, and Keren Tenenboim-Weinblatt. 2025. “Beyond Sentiment: An Algorithmic Strategy for Identifying Evaluations Within Large Text Corpora.” *Communication Methods and Measures* 19 (1): 24–45. <https://doi.org/10.1080/19312458.2023.2285783>.
- Owsiak, Andrew P, and John A Vasquez. 2022. “The Limited Scope of the Democratic Peace: What We Are Missing.” *International Studies Perspectives* 23 (2): 169–90. <https://doi.org/10.1093/isp/ekab015>.
- Page, Benjamin I., and Robert Y. Shapiro. 1983. “Effects of Public Opinion on Policy.” *The American Political Science Review* 77 (1): 175–90. <https://doi.org/10.2307/1956018>.
- Palomino, Marco A., Aditya Padmanabhan Varma, Gowriprasad Kuruba Bedala, and Aidan Connelly. 2017. “Investigating the Lack of Consensus Among Sentiment Analysis Tools.” In *Human Language Technology. Challenges for Computer Science and Linguistics: 8th Language and Technology Conference, LTC 2017, Poznań, Poland, November 17–19, 2017, Revised Selected Papers*, 58–72. Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-030-66527-2_5.
- Paruchuri, Vik. 2024. “VikParuchuri/Marker.”
- Ponza, Marco, Paolo Ferragina, and Francesco Piccinno. 2019. “Swat: A System for Detecting Salient Wikipedia Entities in Texts.” *Computational Intelligence* 35 (4): 858–90. <https://doi.org/10.1111/coin.12216>.
- Post, Abigail S. 2023. “Words Matter: The Effect of Moral Language on International Bargaining.” *International Security* 48 (1): 125–65. https://doi.org/10.1162/isec_a_00466.
- Quackenbush, Stephen L. 2006. “Identifying Opportunity for Conflict: Politically Active

- Dyads.” *Conflict Management and Peace Science* 23 (1): 37–51. <https://doi.org/10.1080/07388940500503804>.
- . 2024. “Politically Active Dyads Revisited: An Update Through 2014.” *Conflict Management and Peace Science*, October, 1–13. <https://doi.org/10/g73v38>.
- Rauh, Christian. 2018. “Validating a Sentiment Dictionary for German Political Language—a Workbench Note.” *Journal of Information Technology & Politics* 15 (4): 319–43. <https://doi.org/10.1080/19331681.2018.1485608>.
- Razeto, Sebastián Briones, and Nicole Jenne. 2021. “Security and Defence Policy Documents: A New Dataset.” *Defense & Security Analysis* 37 (3): 346–63. <https://doi.org/10.1080/14751798.2021.1959730>.
- Signorino, Curtis S., and Jeffrey M. Ritter. 1999. “Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions.” *International Studies Quarterly* 43 (1): 115–44. <https://www.jstor.org/stable/2600967>.
- Singer, J. David. 1988. “Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816–1985.” *International Interactions* 14 (2): 115–32. <https://doi.org/10.1080/03050628808434695>.
- Singer, J. David, Stuart Bremer, and John Stuckey. 1972. “Capability Distribution, Uncertainty, and Major Power War, 1820-1965.” In *Peace, War, and Numbers*, 19–48. Beverly Hills, CA: Sage Publications.
- Small, Melvin, and Joel David Singer. 1982. *Resort to Arms: International and Civil Wars, 1816-1980*. Sage Publications.
- Smith, Julian. 2023. “PyMuPDF: A High Performance Python Library for Data Extraction, Analysis, Conversion & Manipulation of PDF (and Other) Documents.”
- Snider, Don M., and John Nagl. 2001. “The National Security Strategy: Documenting Strategic Vision.” In *U.S. Army War College Guide to Strategy*, edited by Joseph R. Cerami and James F. Holcomb, 127–42. Fort Belvoir, VA: Defense Technical Information Center.

<https://doi.org/10.21236/ADA392553>.

- Song, Kaitao, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. “MPNet: Masked and Permuted Pre-training for Language Understanding.” In *Advances in Neural Information Processing Systems*, 33:16857–67. Curran Associates, Inc.
- Stolberg, Alan. 2012. “How Nation-States Craft National Security Strategy Documents.”
- Terechshenko, Zhanna. 2020. “Hot Under the Collar: A Latent Measure of Interstate Hostility.” *Journal of Peace Research* 57 (6): 764–76. <https://doi.org/10.1177/0022343320962546>.
- Thapa, Surendrabikram, Shuvam Shiwakoti, Siddhant Bikram Shah, Surabhi Adhikari, Hari-ram Veeramani, Mehwish Nasim, and Usman Naseem. 2025. “Large Language Models (LLM) in Computational Social Science: Prospects, Current State, and Challenges.” *Social Network Analysis and Mining* 15 (1): 4. <https://doi.org/10.1007/s13278-025-01428-9>.
- Thompson, William R. 2001. “Identifying Rivals and Rivalries in World Politics.” *International Studies Quarterly* 45 (4): 557–86. <https://www.jstor.org/stable/3096060>.
- Thompson, William R., Kentaro Sakuwa, and Prashant Hosur Suhas. 2021. *Analyzing Strategic Rivalries in World Politics: Types of Rivalry, Regional Variation, and Escalation/De-escalation*. Springer Nature.
- Trager, Robert F. 2010. “Diplomatic Calculus in Anarchy: How Communication Matters.” *American Political Science Review* 104 (2): 347–68. <https://doi.org/10.1017/S0003055410000158>.
- Trani, Salvatore, Claudio Lucchese, Raffaele Perego, David E. Losada, Diego Ceccarelli, and Salvatore Orlando. 2018. “SEL: A Unified Algorithm for Salient Entity Linking.” *Computational Intelligence* 34 (1): 2–29. <https://doi.org/10.1111/coin.12147>.
- Turco, Linnea R. 2024. “Speaking Volumes: Introducing the UNGA Speech Corpus.” *International Studies Quarterly* 68 (1): sqae001. <https://doi.org/10.1093/isq/sqae001>.
- van Atteveldt, Wouter, Mariken A. C. G. van der Velden, and Mark Boukes. 2021. “The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary

- Approaches, and Machine Learning Algorithms.” *Communication Methods and Measures* 15 (2): 121–40. <https://doi.org/10.1080/19312458.2020.1869198>.
- Waight, Hannah, Solomon Messing, Anton Shirikov, Margaret E. Roberts, Jonathan Nagler, Jason Greenfield, Megan A. Brown, Kevin Aslett, and Joshua A. Tucker. 2025. “Quantifying Narrative Similarity Across Languages.” *Sociological Methods & Research*, June, 00491241251340080. <https://doi.org/10.1177/00491241251340080>.
- Walsh, James I. 2005. “Persuasion in International Politics: A Rationalist Account.” *Politics & Policy* 33 (4): 642–70. <https://doi.org/10.1111/j.1747-1346.2005.tb00217.x>.
- Wang, Alex, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2020. “SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems.” arXiv. <https://doi.org/10.48550/arXiv.1905.00537>.
- Wang, Yu. 2024. “On Finetuning Large Language Models.” *Political Analysis* 32 (3): 379–83. <https://doi.org/10/g73vwd>.
- Ward, Michael D., John S. Ahlquist, and Arturas Rozenas. 2013. “Gravity’s Rainbow: A Dynamic Latent Space Model for the World Trade Network.” *Network Science* 1 (1): 95–118. <https://doi.org/10.1017/nws.2013.1>.
- Weede, Erich. 1976. “Overwhelming Preponderance as a Pacifying Condition Among Contiguous Asian Dyads, 1950-1969.” *Journal of Conflict Resolution* 20 (3): 395–411. <https://www.jstor.org/stable/173286>.
- Wu, Chuan, Evangelos Kanoulas, Maarten de Rijke, and Wei Lu. 2020. “WN-Saliency: A Corpus of News Articles with Entity Saliency Annotations.” In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, edited by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, et al., 2095–2102. Marseille, France: European Language Resources Association.
- Xiang, Jun. 2010. “Relevance as a Latent Variable in Dyadic Analysis of Conflict.” *The*

Journal of Politics 72 (2): 484–98. <https://doi.org/10.1017/S0022381609990909>.

Yoon, Hohyun. 2025. “Anger Expressions and Coercive Credibility in International Crises.”

American Journal of Political Science 0 (0): 1–17. <https://doi.org/10.1111/ajps.12937>.