Which Issues Divide? Interpretable Ideal-Point Estimation from Political Text

Jefferson L. Leal*

August 21, 2026

Word count: 9,347

Abstract

Political actors reveal ideology through both the positions they take and the issues they emphasize. Yet, most text-based ideal point measures capture only one of these two channels. Existing approaches also impose a trade-off: unsupervised methods struggle where language is not sharply polarized, while supervised ones require human-labeled training data and ex ante knowledge of how relevant issues and policy positions map onto ideology. I develop an item-response model that jointly scales ideal points from issue salience and issue-specific stance. My approach discovers topics inductively and uses transfer learning with multilingual data from the Manifesto Project to classify stance, requiring no additional human labels. I apply the model to campaign platforms from 4,507 candidates in US House primaries and 16,830 candidates in Brazilian mayoral elections. The estimated positions align with benchmarks based on roll-call votes, campaign contributions, and expert surveys, even in the Brazilian corpus, where local campaign language is largely non-polarized. The model further recovers substantively distinct salience and stance divisions, providing an interpretable account of which issues structure ideological competition.

1 Introduction

The study of party competition, coalition formation, polarization, and policymaking, among other topics, depends on measures of the position of political actors in an ideological space. Political texts are particularly useful as inputs for ideal point estimation

*PhD Candidate, Department of Political Science, University of Rochester. jefferson.leal@rochester.edu.

because they are often available when other sources—such as roll-call votes, campaign donations, and social media networks—are not (Grimmer and Stewart, 2013). Text-as-data also provides an alternative in institutional contexts where roll-call votes are strongly influenced by confounders such as party discipline (Lauderdale and Herzog, 2016; Peterson and Spirling, 2018; Proksch and Slapin, 2010).

Measuring ideal points from text is inherently difficult because ideology is a latent construct, not directly observable. In political texts, lexical variation reflects style, context, and topics, as well as ideology and policy positions (Lauderdale and Herzog, 2016; Laver, 2001). Two traditions in the study of party competition offer theoretical accounts of how ideological differences become visible in language. Saliency theory emphasizes that political actors distinguish themselves by allocating attention to different issues, while positional approaches highlight disagreement over the alternatives proposed within those issues (Budge, 2001b; Laver, 2001). However, existing text-based scales generally operationalize ideology through only one of these channels, or through textual variation that does not map transparently onto either. I fill this gap by proposing a novel item response theory (IRT) model in which issue salience and issue-specific stance within each document are explicitly modeled as two manifestations of the same latent ideal point.

Existing automated text-based ideology scaling methods face a trade-off. Unsupervised methods, such as Wordfish (Slapin and Proksch, 2008) and document embeddings (Rheault and Cochrane, 2020), scale texts without requiring labels, but they depend on the ideological dominance assumption, meaning that they only proxy ideal points well when the main direction of textual variation is driven by ideology (Grimmer and Stewart, 2013). Often, word choices do not differ systematically across ideological poles, a situation I define as non-polarized language. Whenever that happens, unsupervised algorithms instead pick up other major sources of variation, such as language, style, topics, or regional divisions, rather than ideology (Lauderdale and Herzog, 2016; Motolinia, 2025).

Supervised approaches fix the direction of ideology by training a classifier on labeled data, and modern supervised methods based on the transformer architecture achieve high accuracy (Timoneda and Vallejo Vera, 2025; Wang, 2024), but they require extensive hu-

man annotation and prior specification of the issues and how policy positions correspond to a left–right scale in order to create appropriate coding rules (Grimmer and Stewart, 2013; Laurer et al., 2024). Zero-shot and few-shot approaches, including those using generative large language models, mitigate the need for training data,[1] but they still require ex ante specification of how relevant issues and stated positions map onto ideological divisions. Even with generative LLMs, in the absence of good coding rules the analyst is left with whatever unspecified criteria the model applies. Ironically, such requirements are most restrictive for corpora where a measure is most wanted: understudied contexts in which the issue divides are unknown and the analyst cannot easily gather training data representative of the important issues and of the rival positions on each.

I propose a new method, Fasti,[2] combining the main advantage of unsupervised scaling—its ability to inductively discover the main topics (or issues)—with transfer learning on political stance, using data from the Manifesto Project (Lehmann et al., 2025; Merz et al., 2016). The IRT model then combines the topic and stance information into a single scale. Specifically, I use contextual sentence embeddings, nonlinear dimensionality reduction, and density-based clustering to identify the main topics of each corpus, following the approach proposed by Angelov (2020) and Grootendorst (2022). Next, I fine-tune a supervised, multilingual natural language inference (NLI) transformer (Burnham et al., 2026; Laurer et al., 2024) with data from the Manifesto Project to classify the political stance of each sentence into one of three categories: left, right, or neutral. Finally, I use an IRT model to estimate a single latent ideal point per candidate that explains both topic emphasis and mean stance per topic. Importantly, the model does not receive party identifiers as an explicit input, and ideological differences emerge from the content itself.

The measure is designed around three principles. First, ideology measures are most useful analytically when they are low-dimensional. Low dimensionality is not a mere convenience for empirical research but a product of electoral competition: parties have

[1] A classifier is zero-shot when it assigns labels without domain-specific training, and few-shot when it sees only a handful of labeled examples per label. Generative LLMs and natural language inference transformers can be effectively used as zero- or few-shot classifiers (Burnham et al., 2026; Laurer et al., 2024; Le Mens and Gallego, 2025).

[2] Fasti stands for From Attention, Stance, and Topics to Ideal points. I am developing a software package to implement it.

a rational incentive to simplify a myriad of policy issues and alternatives into a few coherent, roughly ordered ideological labels that voters can easily understand (Downs, 1957; Hinich and Munger, 1996). Second, any measure should be interpretable, ideally in terms of issues and policy divisions that characterize political competition (Budge, 2001b; Grimmer and Stewart, 2013; Humphreys and Garry, 2000; Laver, 2001). By design, my method recovers issue-level parameters that distinguish ideological divisions in terms of salience or stance. Third, results should be robust to discretionary analyst choices, since scales that swing with preprocessing decisions are untrustworthy (Denny and Spirling, 2018).

I apply the method in two contrasting contexts to show how it works under different degrees of language polarization. The first is a corpus of 4,507 campaign platforms from United States House primaries (2018–2022) (Porter et al., 2025), where language is very polarized (Case, 2025; Meisels, 2026). The second corpus consists of 16,830 Brazilian 2020 mayoral campaign platforms (Tribunal Superior Eleitoral, 2020), composed mostly of non-polarized promises of public services. Earlier text scalings of the latter recover weak party orderings (Pereira, 2021; Salles, 2019; Salles and Guarnieri, 2019), sometimes read as evidence that local politics in Brazil is only weakly organized by ideology (Salles, 2021). Nonetheless, surveys of politicians, experts, and voters recover an ideological ordering of Brazilian parties, at least at the national level (Bolognesi et al., 2023; Carroll and Meireles, 2024; Desai and Frey, 2023; Power and Zucco, 2009). Such left–right divisions are consistent with redistributive preferences (Power and Zucco, 2012). Even at the local level, salient issues in campaign platforms and budget allocation by mayors follow the lines of party ideology (Desai et al., 2026; Marques et al., 2026; Pereira, 2021). Therefore, the Brazilian corpus is a hard case for any text-based ideology measure: it shows whether the method recovers the left–right ordering of political parties under non-polarized language.

For both corpora, the estimates have strong construct validity. In the US corpus, the issues salient to the left are labor, minorities and human rights, housing, the environment, and several public services, while the right emphasizes political institutions, civil liberties, macroeconomics, and international affairs. The most divisive issues in terms of stance are

immigration, energy, values, and macroeconomic policy. This pattern is consistent with previous work on text-based ideology scaling of campaign platforms (Meisels, 2026) and issue ownership (Cummins, 2010; Fagan, 2021; Hayes, 2005; Wright et al., 2022). Across Brazilian platforms, the left stresses a similar set of issues—minorities and human rights, social assistance, and various public services—but, unlike in the US, also underscores values and worldview. The right foregrounds technology, infrastructure, urban planning, macroeconomics, and public safety. The top three stance-divisive issues in the Brazilian corpus are business, values and worldview, and partnerships.

I then assess convergent validity against benchmark measures. In the US corpus, the estimates correlate with roll-call scores at 0.85 and with donor-based scores at 0.81; in Brazil, estimated party means correlate with expert placements at 0.86. I document that my method correlates more strongly with these benchmark measures than do alternative models based on bag-of-words or embeddings. I show that traditional unsupervised bag-of-words scaling is unstable under non-polarized language: Wordfish estimates for the Brazilian corpus are highly sensitive to preprocessing choices. Transfer learning of stance from the Manifesto Project allows Fasti to isolate the ideological direction from other sources of textual variation, while retaining flexibility in how issue salience relates to ideal points.

One potential objection is that generative large language models (LLMs) make a new method for ideal point measurement unnecessary since they produce scales of policy positions that correlate with expert assessments (Benoit et al., 2026). I apply the ask-and-average procedure of Le Mens and Gallego (2025) to a sample of 1,000 platforms from each corpus, using two different OpenAI models. The LLM and IRT estimates have broadly comparable convergent validity against the external benchmarks. However, the cost of using LLMs quickly adds up (Laurer et al., 2024), and there is no off-the-shelf solution to interpret LLM estimates in terms of issue salience and stance divisions per issue. I find concerning discrepancies that cast doubt on the construct validity of LLM-based measures for this task. Minor prompt variations in whether the LLM is instructed to ignore party or name cues affect document-level scores, especially those pertaining to extreme candidates.

Additionally, I fit my joint IRT model using LLM-based topic and stance assignments and find that, for the Brazilian corpus, the LLM identifies implausible ideological divides, putting policy issues outside municipal jurisdiction, such as international affairs and immigration, among the most ideologically discriminating issues.

At the party level in the Brazilian case, Fasti also recovers expert scores better than the LLM-based measure. Performing well under non-polarized language is useful, especially when studying local or comparative politics. In much of the Global South, parties typically compete on valence issues, such as the provision of public services, instead of engaging with positional issues tied to national ideological or class cleavages (Bleck and van de Walle, 2012; Kitschelt, 2007), partly because economic inequality creates broad electoral support for redistribution and welfare (Meltzer and Richard, 1981). Even in the United States, although local politics has become more partisan and aligned with a left–right scale (Bucchianeri et al., 2021), the typical municipal arena is still dominated by debates over public service provision and spatial distributive conflicts, and policymaking coalitions often shift from issue to issue (Bucchianeri, 2020).

Future advances in NLP may improve the quality of topic discovery and stance classification.[3] Whatever those advances, an IRT framework that connects issue salience and stance will remain useful for assessing the construct validity of ideal point measures based on text. Here, I focus on deep learning approaches there are scalable, high-quality, and free to run. The topic model can be fit on a sample and applied to the rest of the corpus, allowing the pipeline to scale to millions of sentences on a single modest graphics processing unit (GPU). The stance classifier is trained entirely using transfer learning from comparative data and requires no additional human labels for a new corpus, at least in languages seen by the NLI transformer during training.

The rest of the paper proceeds as follows. Section 2 discusses how different approaches for measuring ideology from text relate to issue salience and stated positions per issue. Section 3 presents the IRT model and its estimation. Section 4 describes the two corpora

[3]For instance, recent work proposes advances in aggregating outputs from different LLMs while accounting for correlated errors in their annotations (Bouyamourn and Spirling, 2026), and in combining embeddings with LLM annotations for continuous scaling of social science constructs in texts (Ornstein, 2026).

and shows that language is more polarized in the US House corpus than in the Brazilian local context. Section 5 presents the results and discusses at length the substantive findings on which issues divide each corpus by stance and salience. Section 6 validates the estimates against the external benchmarks and compares their performance with alternative methods, including LLMs. Section 7 concludes.

2 Measuring Ideal Points from Political Text

Researchers estimate ideal points from many kinds of data. Item response theory models based on roll-call votes are the standard toolkit for scaling the positions of legislators and courts (Clinton et al., 2004; Jackman, 2001; Martin and Quinn, 2002; Poole and Rosenthal, 1997). Other data inputs for ideological scaling include social media connections (Barberá, 2015), campaign donations (Bonica, 2014), and political texts, such as party manifestos (Budge et al., 2001; Slapin and Proksch, 2008), campaign platforms (Case, 2025; Meisels, 2026), and speeches (Laver et al., 2003; Peterson and Spirling, 2018). Text is particularly useful since it exists for non-incumbents and for actors that do not vote or donate, providing clear information about ideology in institutional contexts where roll-call votes are strongly influenced by confounders such as parliamentary coalition dynamics (Peterson and Spirling, 2018).

Political texts can inform the ideal point of a politician through at least two channels: the salience of different policy issues and the stated position on each issue (Budge, 2001b). First, the attention a politician or party places on each topic signals how important that issue is to them (i.e., issue salience). Second, within each issue, politicians may differ with respect to the alternative solutions they propose (i.e., stance). Saliency theory, based on the idea that parties emphasize some policy issues more than others, justifies measuring ideal points from differences in issue emphasis in political texts (Budge, 2015; Laver, 2001). Salience differences are informative about ideal points because politicians allocate a finite budget of attention across issues, and that distribution signals government priorities (Budge and Hofferbert, 1990; Fagan, 2021; Humphreys and Garry, 2000). Moreover,

parties often avoid taking opposing sides on the same issue and instead compete by giving attention to the issues on which their reputation is strongest (Budge, 2001b, 2015; Petrocik, 1996; Wagner and Meyer, 2014). One of the reasons for the lack of confrontational position-taking is that some policy areas are valence or one-positional issues, which represent electoral consensus (Bawn et al., 2012; Budge, 1982; Jones et al., 2009; Robertson, 1976; Stokes, 1963). Importantly, for such uncontroversial issues, relative differences in salience become especially informative of the ideological priorities of different candidates (Laver, 2001). Conversely, strategic omission of certain topics is also informative about the latent political position of a politician.

Models based on word counts implicitly rely on saliency theory; the frequencies of words reflect the issues that authors care about and, consequently, their political positions (Sältzer, 2022). Wordfish is an unsupervised item-response model that estimates a one-dimensional latent scale from word counts (Slapin and Proksch, 2008). Wordscores is a supervised analogue that uses scored reference texts and compares the relative frequency of words in the reference texts to the frequency of words in the target texts in order to scale ideal points (Laver et al., 2003), but it does not model how sharply each word discriminates ideology (Slapin and Proksch, 2008). Both Wordfish and Wordscores are approximations of a solution to a correspondence analysis on the document–term matrix (Lowe, 2008, 2016), and political scientists have applied correspondence analysis for new corpora in order to scale ideal points (Sältzer, 2022). Additionally, all such approaches depend on the assumption of ideological dominance, meaning that the main axis of variation in word counts is due to ideological differences between authors (Grimmer and Stewart, 2013).

Even authors sympathetic to the salience approach recognize that some issues cannot be analyzed without taking into account the deep substantive disagreements between different political actors (Budge, 2001a; Laver, 2001). In certain contexts, most candidates concentrate attention on the same set of issues, a phenomenon called issue convergence by Sigelman and Buell (2004), making salience uninformative about ideal points. For instance, a sizable share of US Senate speeches corresponds to procedural or symbolic

statements (Quinn et al., 2010). In general, on issues where opposed sides both care intensely, issue emphasis carries no information about the stated position at all (Lowe et al., 2011; Stokes, 1963). A measure of ideal points should therefore identify stated positions against the overwhelming competing sources of variation in text, such as language, style, and topics (Lauderdale and Herzog, 2016).

This reasoning justifies what has been called the confrontational or positional approach, i.e., measuring ideal points based on the stated position per issue (Budge, 2001a; Laver, 2001). Laver and Garry (2000) create a dictionary of words associated with opposing positions on a set of issues, and use the difference between shares of words in each pole as a measure of the position per issue. Supervised classifiers can be used to detect issue-specific stance (Bestvater and Monroe, 2023; Laurer et al., 2024; Nikolaev et al., 2023; Stewart and Zhukov, 2009). Recent papers propose using natural language inference (NLI) classifiers to classify the stance of sentences or paragraphs (Burnham et al., 2026; Laurer et al., 2024). NLI classifiers have the advantage of requiring a relatively small training set, because they leverage the knowledge of a base model pre-trained on general hypothesis entailment tasks. Alternatively, Le Mens and Gallego (2025) propose directly prompting large language models to score the ideological position of sentences and average them within documents, the so-called ask-and-average approach.

A third strand of papers uses deep learning to measure ideal points from political texts. Rheault and Cochrane (2020) propose measuring ideology with document embeddings, using a Doc2Vec neural network that learns a vector representation of each document by predicting its words. Measures of ideology and other features of party competition, such as opposition and government behavior, can be obtained by applying principal component analysis to the document embeddings. Text embeddings encode the representation of the meaning of documents and words in dense vector representations, and the cosine distance between two embeddings can be used to measure the semantic similarity between them (Rodriguez and Spirling, 2022). However, text embeddings capture not only information about ideology, but also about other document characteristics, such as topic, context, and style (Arora et al., 2017). To ensure that the extracted dimension reflects

ideology rather than other document characteristics, Case (2025) proposes using semantic projection (Grand et al., 2022), meaning a semi-supervised linear projection of paragraph embeddings into a unidimensional subspace defined by a set of anchor words representative of the ideological poles. Ornstein (2026) proposes a similar supervised approach, but instead of anchor words it takes a small set of pairwise comparisons between labeled documents, and scores each text along the direction in embedding space that corresponds to the construct of interest.

Some studies have incorporated information about topics based on framing (Chong and Druckman, 2007; Entman, 1993), the notion that different word choices within a topic inform a latent dimension of political disagreement. Lauderdale and Herzog (2016) propose Wordshoal, an unsupervised method that estimates a latent position per issue and then a common ideological dimension across issues. Wordshoal does not discover the issues, since the analyst provides the grouping of texts into issues, proxied in the original application by the parliamentary debate in which each speech was delivered. Vafa et al. (2020) propose the text-based ideal point model, which jointly learns authors' ideal points, topic distribution, and the ideological leaning of each word per topic.

This paper contributes a model that estimates ideal points as a unidimensional latent factor explaining both issue salience and stated positions. For each document, the model takes as inputs the count of sentences per topic and the mean stance of the sentences per topic. In order to obtain such quantities, I propose to use topic modeling on sentence embeddings, and I fine-tune an NLI classifier using labeled sentences from the Manifesto Project (Lehmann et al., 2025; Merz et al., 2016).

3 Method: Scaling Ideology from Measures of Salience and Stance

As discussed in the previous section, at least two components are relevant for the measurement of ideology from political texts: the relative importance of different issues to each candidate (i.e., salience), and the stated position each candidate takes on those issues

(i.e., stance) (Budge, 2001a; Laver, 2001). Applied researchers typically assume that the more each candidate emphasizes an issue, the more salient it is to them, but emphasis alone is silent on direction.

In order to capture both of these components, I propose an item-response model that combines information on the distribution of topics within a document with the mean stance per topic. I then model ideology as a unidimensional latent factor that generates both components simultaneously. I proceed in four steps: three machine learning steps extract the topics and stances, and a final step models the latent ideology.

Step 1: Sentence Embeddings. Split each document into sentences and use pre-trained language models to encode each sentence into a contextual embedding vector representation.

Step 2: Topic Modeling. Use dimensionality reduction and density-based clustering to group semantically similar embeddings into topics, following the pipeline proposed by Angelov (2020) and Grootendorst (2022).

Step 3: Stance Detection. Classify the stance of each sentence using a fine-tuned natural language inference (NLI) transformer. Every sentence receives a predicted label: left (-1), right ($+1$), or neutral (0).

Step 4: Item-Response Model. Scale the unidimensional ideal point using a generative item-response model. The model measures a document's ideology as a latent factor that explains both the distribution of sentences across topics and the mean stance taken on each topic. The parameters of the model can be interpreted in terms of the issues that divide stated positions (stance gap) and relative differences in how much each side of the political spectrum emphasizes each issue (salience gap).

The remainder of this section develops each part in turn. Section 3.1 states the item-response model first, since it defines the quantities the machine learning steps must deliver. Section 3.2 details the topic model behind the counts (Steps 1 and 2), and Section 3.3 details the stance classifier (Step 3). Section 3.4 presents identification and estimation.

3.1 Item-Response Model

For document i, let Y_{it} denote the count of sentences assigned to topic t (a non-negative integer), $M_i = \sum_t Y_{it}$ the total count of sentences, and $\pi_{it} \in [0, 1]$ the probability that a sentence is assigned to topic t. I model the allocation of sentences across topics as a multinomial distribution whose composition depends on a latent ideal point θ_i.

$$Y_i \sim \mathit{Multinomial}(M_i,\ \pi_i), \tag{1}$$

$$\pi_{it} = \mathit{exp}\{\eta_{it}\} \big/ \textstyle\sum_s \mathit{exp}\{\eta_{is}\}, \tag{2}$$

$$\eta_{it} = \psi_t + \rho_t\, \theta_i + \kappa_t(\theta_i^2 - 1) \tag{3}$$

The term η_{it} in Equation 3 is a score that predicts the occurrence of each topic in the document, and it is a quadratic function of the latent ideal point θ_i. The intercept ψ_t captures the topic's baseline usage in the corpus. The linear coefficient ρ_t measures the degree to which the topic's coverage depends on the candidate's ideal point. The quadratic term κ_t allows for non-monotone salience: some topics may be most intensely used by the extremes, the center, or either side.[4] Centering the quadratic at $\theta_i^2 - 1$ facilitates interpretation.[5] Equation 2 converts the scores into probabilities: the denominator normalizes π_i to sum to one.

Modeling the topic distribution is informative about the relation between ideology and issue salience, but says nothing about the position taken by the actor on any given issue. To account for this, I use data on the stance of each sentence, which takes discrete values: $+1$ if right, -1 if left, and 0 if neutral. Let $S_{it} \in [-1, 1]$ denote the mean of these scores over the corresponding Y_{it} sentences. I model mean stance as an affine function of

[4] Setting κ_t to zero results in a linear specification in which ρ_t has to accommodate possible non-monotonic patterns. In such cases, issue salience is assumed to be one-sided or roughly constant across ideology.

[5] The ideal points are centered at zero (Equation 6) and the prior of Equation 9 gives them unit variance, so the term $\theta_i^2 - 1$ averages to zero across documents. The intercept ψ_t therefore keeps its meaning as the topic's average score in the corpus, instead of absorbing part of the curvature, and κ_t measures only the departure from the linear trend. Without the centering, a topic with strong curvature would report a baseline that is not its average coverage, and shrinking κ_t toward zero would move that baseline with it.

the latent ideal point (θ_i).

$$S_{it} \sim \mathcal{N}\Big(\alpha_t + \beta_t\,\theta_i,\ \tfrac{\sigma^2}{Y_{it}}\Big) \tag{4}$$

In the equation above, the intercept α_t captures the baseline mean stance per topic, i.e., the position expected of a candidate at the center of the ideological scale. The loading β_t measures how much stance divisions covary with ideology. A positive β_t means the topic's stance moves with the ideology gradient, so candidates further to the right express positions coded further to the right, and a negative β_t means the topic's stance division runs against the coding convention.[6] A topic with β_t near zero carries no stance discrimination, and topics with large $|\beta_t|$ divide the sides by stance. This specification implies that when most politicians take the same stance on an issue, that stated position carries little information about the underlying ideology. The variance is inversely proportional to Y_{it} because S_{it} is a mean; stance is measured more precisely for a topic that a document covers in many sentences. A document with very few sentences per topic therefore carries mean stances that take their extreme values mechanically, so I exclude such documents from the fit.

The count model in Equations 1–3 and the stance model in Equation 4 together form the joint item-response model: the first ties each topic's salience to the ideal point θ_i, and the second ties each topic's stance to the same θ_i. A single latent factor generates both how much a candidate emphasizes an issue and which side they take on it.

3.2 Issue Salience using Topic Modeling

I identify the issues present in the corpus using unsupervised topic modeling, meaning that, instead of having an analyst specify the issue taxonomy in advance, the method discovers topics in a data-driven way. I follow the pipeline proposed by Angelov (2020) and

[6] The sentence codes carry a general left–right convention that is fixed before the model is fit (Section 3.3), so a reversal is a finding about the corpus, not an artifact of the scale. Trade illustrates the point: protectionism is conventionally coded left, but a populist right can be protectionist as well. The topic then takes $\beta_t < 0$, which the model absorbs as a topic-specific loading and leaves the latent scale intact.

Grootendorst (2022), by obtaining sentence embedding vectors, reducing their dimensionality, and clustering them into topics.[7] I use contextual sentence embeddings, trained to identify the semantic nuances of each word based on the surrounding context. They are obtained by passing each sentence through a pre-trained transformer model (Reimers and Gurevych, 2019) that maps each sentence to a high-dimensional vector representation.

Data in high-dimensional spaces is typically sparse, and clustering in such spaces is challenging. I therefore reduce the dimensionality of the embeddings before clustering. As proposed by Angelov (2020) and Grootendorst (2022), I use Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018), a non-linear dimensionality reduction technique that preserves the local structure of distances between data points. Then, I use Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al., 2017), a density-based clustering algorithm that does not require a pre-specified number of clusters. Each sentence is assigned to its closest cluster.

I adapt the class-based TF-IDF representation proposed by Grootendorst (2022) to identify topic keywords. I use a normalized representation for each term w in topic t.

$$W_{w,t} = \frac{f_{w,t}}{\sum_{w'} f_{w',t}} \times \log \frac{T}{n_w} \tag{5}$$

Here, $f_{w,t}$ counts occurrences of term w in the concatenated text of topic t, T is the number of topics, and n_w the number of topics whose text contains w. The first factor measures the term's importance within the topic, and the second factor discounts terms common to many topics.

For each topic, I give GPT-5.4 mini the ten leading key terms and 150 sentences of the topic, 50 drawn from each stance class, and ask for a short label. I read every label against those same sentences and correct the few that misdescribe the topic. I submit the topics that receive the same name back to the model, in one request, asking it to distinguish them.[8]

[7] Researchers typically refer to this topic modeling approach as the BERTopic method (Grootendorst, 2022). The name is confusing, because the BERTopic software package is modular and allows for several distinct implementations of embedding-based topic modeling and keyword representation.

[8] Labels are used for display and do not enter estimations. Appendix B.1 reports the constraint the schema imposes on the answer and reproduces both prompts.

Three practical issues arise. First, a corpus can be too large to cluster in one pass. I fit the pipeline on a stratified sample of documents[9] and then project the remaining documents through the fitted UMAP and HDBSCAN models. Second, HDBSCAN classifies data points in non-dense regions as outliers, and in my applications around half of the sentences fall into that category. Discarding such data would misstate topic salience, so I assign every noise sentence to the topic it belongs to with the highest probability under the fitted cluster membership model.

Finally, there exist topics that concentrate in a handful of documents, usually due to idiosyncratic vocabulary, such as proper nouns. In the item-response model, such topics act as fixed effects for the few documents that carry them, gaining large but spurious discriminative influence. I dissolve topics whose key terms spread across too few documents and reassign their sentences to the nearest surviving topics.[10]

Appendix D.6 compares performance measures of the pipeline I follow with common alternatives: Latent Dirichlet allocation, non-negative matrix factorization, structural topic model, and Dirichlet-multinomial regression (Blei et al., 2003; Lee and Seung, 1999; Mimno and McCallum, 2008; Roberts et al., 2014), each using the same number of topics for comparability. In both applications, the embedding-based topics are more coherent and more distinct than those of the four alternatives, indicating high semantic validity (Grimmer and Stewart, 2013).

3.3 Stance by Transfer Learning

The IRT model also takes as input each sentence's **stance**, the main political leaning it takes on its subject: right (+1), left (−1), or neutral (0). In a fully supervised approach, the analyst would have to provide a large number of sentences annotated from the target corpus to train a classifier. Instead, I follow a transfer learning approach, and use human-labeled text data from the Manifesto Project (Lehmann et al., 2025; Merz et al., 2016) to train a multilingual stance classifier, so that the method can apply to a new corpus

[9] In the applications, I use party as a stratum, but any categorical metadata can be used.
[10] Appendix B.1 details the implementation of the full pipeline for the present applications.

without further annotation and at reduced labeling cost.[11]

The Manifesto Project codes quasi-sentences of national party manifestos into numerous policy position categories.[12] I map each category to two dimensions (economy and society) and three stance classes: left, right, and neutral. The two dimensions correspond to the economic and social policy scales that Lowe et al. (2011) build from the Manifesto Project categories, which separate the market-versus-state divide from the divide over social order and civil rights. I extend this mapping to the full coding scheme.

On the economy, I define the left as favoring redistribution, regulation, public services, and environmental protection, while the right favors market freedom and economic growth. On the society dimension, the left favors civil rights, democratic participation, and the inclusion of minorities and immigrants, while the right favors nationalism, traditional morality, and authority. Neutral sentences take no clear position, make valence or competence claims, or address orthogonal political issues, mainly corruption and federalism or decentralization.[13]

There are known issues with low intercoder reliability in the Manifesto Project data (Mikhaylov et al., 2012). Following Burnham et al. (2026), I take a second and independent reading of every sentence from GPT-4.1 (OpenAI, 2025), and I keep only the sentences on which the two readings agree. The prompt for this task is presented in Figure B.3.[14]

For the classifier, I fine-tune a natural language inference (NLI) model. Such models read a *premise* (the sentence) and *hypotheses* and score the probability that the premise entails each hypothesis.[15] NLI transformers can be fine-tuned on a relatively small training set (Burnham et al., 2026; Laurer et al., 2024). I fine-tune such a model on the

[11]The classifier can be applied, at least, to a corpus in any of the nine languages it is trained on: Catalan, English, French, Galician, German, Italian, Portuguese, Spanish, and Swedish.

[12]I keep twelve countries in nine languages: Argentina, Brazil, Canada, France, Germany, Italy, Mexico, Portugal, Spain, Sweden, the United Kingdom, and the United States.

[13]Table B.5 in Appendix B.3 reports the mapping in full, listing every Manifesto Project category mapped to each of the six combinations of dimension and stance. I use the standard categories and the version-5 handbook subcategories, but not the Central and Eastern European extension.

[14]Large language models (LLMs) perform well in classification tasks (Gilardi et al., 2023). I consider concordance as an extra filter to ensure high-quality training data. For the purpose of training a classifier, I do not need to assume that LLMs perform better than the original human coders. I report the disagreement rates and intercoder reliability in Appendix B.3.

[15]For a full explanation of NLI classifiers, consult Laurer et al. (2024).

tasks of dimension classification, stance classification per dimension, and overall stance classification, and show that it reaches a balanced accuracy of 0.83 in the final task on a held-out test set. The exact entailment hypotheses, the training configuration, and the full held-out metrics are in Appendix B.3.

3.4 Estimation

The model detailed in Section 3.1 is only identified up to reflection, translation, and rescaling of the latent parameters. These indeterminacies are well known in item-response models (Bafumi et al., 2005; Clinton et al., 2004; Goplerud, 2019; Jackman, 2001). I impose the three constraints below to fix the latent scale of ideal points.

$$\sum_i \theta_i = 0 \tag{6}$$

$$\frac{1}{T}\sum_t \beta_t^2 = 1 \tag{7}$$

$$\bar{\theta}_{\mathcal{R}} > 0 \tag{8}$$

Equation 6 sets the location by centering the ideal points so that the average candidate ideal point is zero.[16] Equation 7 fixes the scale and prevents the topic stance loadings from overdispersing. Equation 8 solves the reflection issue by imposing the direction of the scale.

Two priors regularize the fit. The second prior in Equation 9 shrinks curvature, since topics with few observations tend to overfit, and its variance τ^2 is estimated by marginal likelihood (Morris, 1983).

[16] Centering is innocuous for relative scaling and common practice (Bafumi et al., 2005; Case, 2025; Clinton et al., 2004; Meisels, 2026; Slapin and Proksch, 2008). However, whenever one partitions the sample into any two mutually exclusive and collectively exhaustive groups, their means are mechanically constrained. Let n_g and $\bar{\theta}_g$ denote group g's size and mean position. The constraint $\sum_i \theta_i = 0$ implies that $n_1\bar{\theta}_1 + n_2\bar{\theta}_2 = 0$, hence $\bar{\theta}_2 = -(n_1/n_2)\bar{\theta}_1$. Thus, the two groups must have means with similar absolute values (up to group-size ratio) but opposite signs. In their application, Case (2025) notes that Democratic and Republican means are equidistant from the center. The symmetry, nonetheless, is largely a result of the location constraint with two similarly sized parties, not a substantive finding.

$$\theta_i \sim \mathcal{N}(0,\ 1), \qquad \kappa_t \sim \mathcal{N}(0,\ \tau^2) \tag{9}$$

The two blocks are not naturally on an equal footing. The count block enters sentence by sentence, so it would dominate the stance block. I therefore estimate the model in two stages. The first stage fits the stance component of Equation 4 alone by alternating least squares. It yields the stance residual variance $\hat{\sigma}^2$ and a provisional scale. Holding that scale fixed, I estimate three calibration quantities: an overdispersion factor d_i (Wedderburn, 1974), the prior variance τ^2 of Equation 9, and a Godambe factor c_Y that aligns the information the count block claims with the variability its score shows (Godambe, 1960; Varin et al., 2011). Appendix C.1 states each formula.

The second stage minimizes one calibrated objective jointly over the ideal points and both item blocks. Let ℓ_i^S denote document i's stance loss and ℓ_i^Y its negative multinomial log-likelihood; Appendix C.1 writes both in full.

$$\mathcal{L} \;=\; \sum_i \left[\frac{\ell_i^S}{2\hat{\sigma}^2} \;+\; \frac{\ell_i^Y}{d_i\, c_Y} \;+\; \frac{\theta_i^2}{2} \right] \;+\; \frac{1}{2\,\tau_{\mathit{eff}}^2\, c_Y} \sum_t \kappa_t^2 \tag{10}$$

The first two terms are the blocks on their calibrated weights, and the last two are the priors of Equation 9. Estimation uses block coordinate descent, which scales to large corpora. Because Equation 10 combines two overlapping descriptions of the same document, it is a composite rather than a true likelihood, and the usual inverse-information standard errors would be too small (Varin et al., 2011). I report robust standard errors from the inverse of the Godambe information $\mathcal{G}$ (Godambe, 1960; White, 1982).

$$\mathcal{G}^{-1} \;=\; H^{-1} J\, H^{-1} \tag{11}$$

The matrix H is the expected curvature of the objective and J is the covariance of its document-level scores, with documents as the independent clusters, the level at which independence is credible. Appendix C.1 derives every calibration quantity.

4 Data

I use campaign platforms in two different contexts: US congressional primaries (2018, 2020, and 2022) and Brazilian mayoral races (2020). Campaign platforms are convenient for ideal point estimation because they are written by the candidates' campaign teams, span multiple policy issues, and are available for most candidates, not only incumbents (Druckman et al., 2009; Meisels, 2026). The two applications show that the method performs well under very different conditions of language polarization.

In the US corpus, language is much more polarized and I am able to compare my estimates with the previous literature (Case, 2025; Meisels, 2026). In the Brazilian case, language is less polarized and most candidates devote a significant portion of their platform to supporting the expansion of public services. The contrast is crucial: if the method works only when language is polarized, Brazil will expose it.

United States. The US corpus comes from CampaignView (Porter et al., 2025), a database of candidates' campaign platforms from the two major parties in US House of Representatives primaries in 2018, 2020, and 2022. CampaignView separates each candidate's website into a policy platform and a biographical narrative, and I analyze only the platforms. The analysis set contains 4,507 candidate–cycle platforms, and it includes the challengers and primary losers that roll-call-based measures do not capture.[17]

Brazil. Brazilian electoral law requires every candidate for executive office to file a campaign platform with the electoral courts at the time of candidacy registration.[18] I collect 16,830 platforms filed by mayoral candidates in the 2020 municipal elections, spanning 33 parties. The platforms typically enumerate proposals across policy areas such as health, education, sanitation, and public safety.

[17]Each candidate carries a party label and a Federal Election Commission identifier, which I later use to match candidates to the external validation benchmarks.

[18]The expression for the Brazilian official campaign documents (in Portuguese, *proposta de governo*) could be translated as candidate manifestos in the terminology used in Comparative Politics. In this article, I borrow the US term *campaign platform*. The legal requirement was created by Law no. 12.034/2009, amending the electoral law (Law no. 9.504/1997). Candidates file their platforms with regional electoral courts, and the Superior Electoral Court (*Tribunal Superior Eleitoral*) publishes the corpus. I parse the texts directly from the PDF files whenever possible and use optical character recognition when necessary.

The Brazilian corpus is a demanding test for text scaling. It is an unusually large corpus, and much of the text is consensual since most candidates support the expansion of diverse public services. Appendix A.1 describes the composition of the Brazilian corpus.[19] Scaling platforms from the three Brazilian levels of executive office, Salles (2021) concludes that the ideological organization of political competition is weakest at the municipal level. However, other studies show that ideology does affect local policy. Right-wing mayors spend less on the poor than left-wing mayors in low-poverty municipalities (Desai and Frey, 2023), and partisan control shifts the composition of urban investment (Marques et al., 2026). Mass and elite survey data also show that state representatives share a common left–right dimension with voters (Carroll and Meireles, 2024).

Unit of analysis. Both the topic model and the stance classifier operate on sentence-length text segments. Text preprocessing is minimal, as appropriate for deep learning pipelines (Rheault and Cochrane, 2020). The text segmentation pipeline is described in Appendix A.2. Segmentation yields 439.3 thousand sentence-length text fragments in the US corpus and 3.10 million in the Brazilian one. Table 1 summarizes descriptive statistics for both corpora.

Missing platforms in the US corpus come mostly from candidates in uncontested primaries with little fundraising (Porter et al., 2025). In Brazil, filing is mandatory, and 86% of the mayoral candidates filed a platform.

Topics. Fitting the topic model pipeline on the whole US corpus yields 95 subtopics. The Brazilian corpus is too large to cluster efficiently in a single pass, so I fit the topic model on a stratified sample of (at most) 150 platforms per party, and project the remaining sentences through the fitted reduction and cluster membership model, resulting in 168 subtopics. I then group the discovered topics into the supertopics of my own codebook, which gives 25 classes in the United States and 26 in Brazil, and the model of Section 3

[19] A small number of the filed documents are paperwork for registering a candidacy rather than a platform, e.g., a power of attorney, a court certificate stating that nothing is on record against the candidate, or a party's convention minutes. These templates state no positions, so I use keywords to identify and remove them before estimating the model. Appendix A.1 describes this filtering process and a validation exercise.

Table 1: Data: Campaign Platforms

	United States	Brazil
Office	House	Mayor
Election cycles	2018–2022	2020
Documents	4,507	16,830
Parties	2	33
Total words (millions)	8.5	54.9
Words per document (median)	1,093	2,369
Sentences (thousands)	439.3	3,100.0
Sentences per document (median)	58	148
Words per sentence (median)	18	15

US documents are candidate–cycle policy platforms from CampaignView (Porter et al., 2025), excluding biographical narratives. Brazilian documents are the official campaign platforms of mayoral candidates. The sentence segmentation pipeline is described in Appendix A.2.

estimates over those classes. Appendix B.1 details the topic modeling pipeline, including the hyperparameters used, and Appendix B.2 lists the grouping in full. Appendix E re-estimates the model at the subtopic resolution and compares the resulting scales.[20]

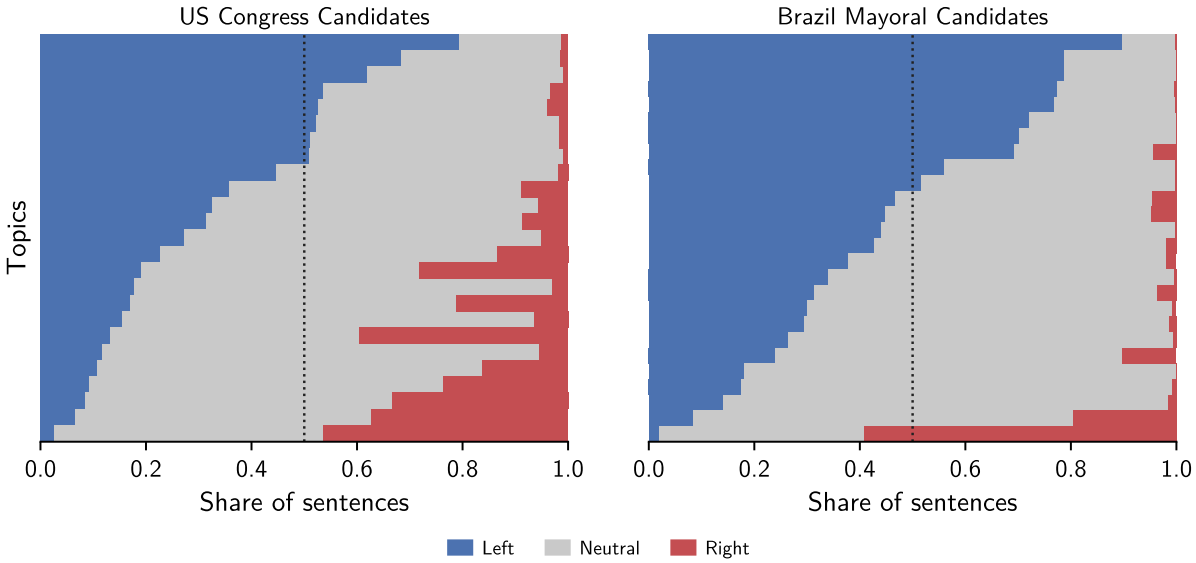
Stance and Language Polarization. Figure 1 contrasts the two corpora on language polarization. Each horizontal bar corresponds to one topic and shows how its sentences distribute across left, neutral, and right stances. In the United States (left panel), topics are usually stance-polarized. The language that national candidates from opposing parties use is very polarized, and right-wing candidates do not shy away from making their position explicit.[21]

Issue attention patterns are also distinct between parties. The correlation of coverage shares between parties is moderate (0.73). When both parties emphasize the same issue, they often take opposite positions. For instance, on immigration, Democrats typically lean left and Republicans lean right. The US corpus presents an ideal situation for measurement: stance and salience divide cleanly by ideological group.

[20]The two corpora differ in how much non-policy text they carry. CampaignView separates each website into a policy platform and a biographical narrative, so the US corpus holds little boilerplate, whereas the Brazilian platforms retain legal citations, tables, section headers, and biographies. The topic model recognizes this boilerplate as topics: 25.5% of the Brazilian sentences against 15.7% of the US corpus. I drop that class before estimating, which allows the item-response model to focus on substantive issues.

[21]Table D.2 in Appendix D.3 lists the 15 top covered topics in the US corpus, and each party's coverage rank, share, and mean stance.

Figure 1: Stance Composition of Topics



Each horizontal bar is one topic. Each bar gives the share of the topic's sentences that had their stance classified as left (blue), neutral (gray), and right (red). Topics are sorted by their left share, and the dotted line marks an even split.

In Brazilian local politics, there is much less language polarization (Figure 1, right panel). Most top issues are shared between leftist and rightist parties. The coverage shares of the two party groups correlate at 0.98.[22] Most platforms favor the expansion of public services: 92% of supertopics lean left, and the average sentence score is leftist (−0.44). Brazilian mayoral platforms are, overwhelmingly, promises to provide: more leisure, culture, and sports options; improved teachers' careers; more support for farmers.

Most parties follow the same formula, and right-wing parties often emphasize public service expansion as much as left-wing ones do. In any corpus with similarly low language polarization, a measure of ideology must recover the small deviations of each candidate's stated positions and issue emphasis from the common baseline.

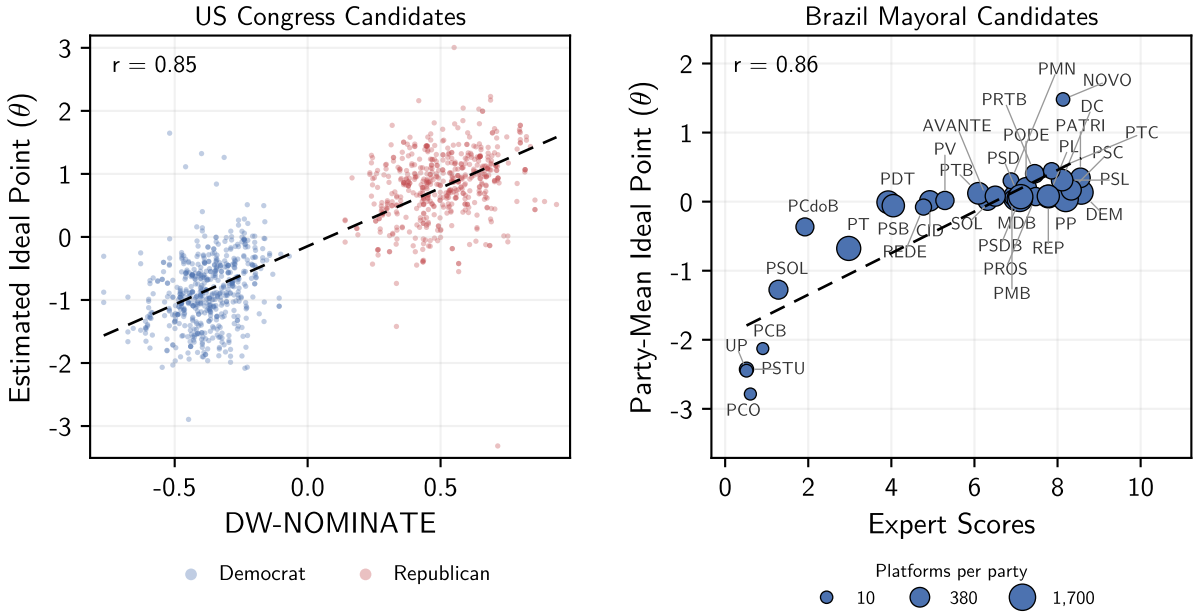
5 Results

I first show that the ideal point estimates produced by Fasti recover broad patterns of party ideological ranking in both corpora. Figure 2 plots the estimated ideal points against DW-NOMINATE (Lewis et al., 2026) in the United States and, at the party

[22]Table D.3 in Appendix D.3 lists the 15 top covered topics in the Brazil corpus, with each bloc's own coverage rank, share, and mean stance.

level, against the expert placements of Bolognesi et al. (2023) in Brazil. The correlations are 0.85 across 1,111 US platforms and 0.86 across the 33 Brazilian parties. Neither benchmark is derived from platform text. In Figure 3, I present the distributions of ideal points per party group, scaled to mean zero and unit standard deviation.[23] Table 2 presents the corresponding summary statistics.[24]

Figure 2: Estimated Ideal Points Against External Benchmarks



Left panel: each dot corresponds to a US House platform, representing its estimated ideal point θ_i against the candidate's DW-NOMINATE score (Lewis et al., 2026); Democrats in blue and Republicans in red. Platforms without a DW-NOMINATE score are excluded. *Right panel*: each dot is a Brazilian party, representing its mean ideal point against the party's expert score (Bolognesi et al., 2023). Marker size increases with the number of platforms per party, and each party is labeled by its acronym. Dashed lines are least-squares fits; r is Pearson correlation.

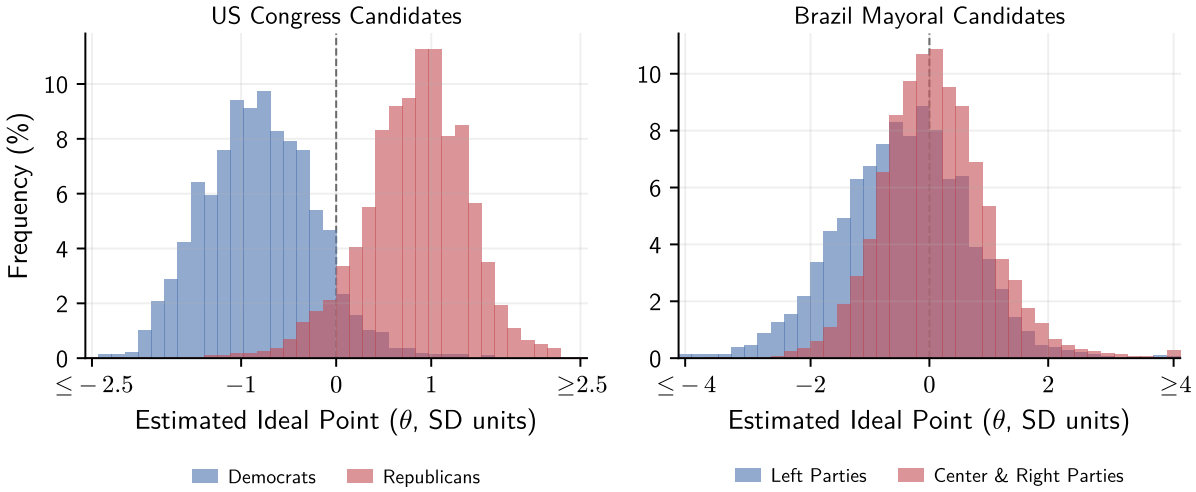
The distribution of estimated ideal points for US House candidates is bimodal, consistent with the known pattern of polarization among political elites (Bafumi and Herron, 2010; McCarty et al., 2016). Traditional measures of ideal points based on roll-call votes or campaign contributions usually result in minimal or no overlap between the two parties. Roll-call votes may overseparate parties due to the endogenous selection of which bills and issues reach the floor (Bateman et al., 2017; Clinton, 2012) or party discipline (Minozzi and Volden, 2013). Contribution-based scores may exaggerate polarization as

[23]Party group means are constrained to be similar in absolute values up to group size differences, as explained in footnote 16. Table D.4 in Appendix D.4 details party mean estimated ideal points and expert placements for the Brazilian corpus.

[24]Party group means track the respective medians closely in both corpora, indicating that the group distributions are not distorted by extreme platforms.

well, since donation decisions may be based not only on ideological proximity but also on partisan loyalty (Li, 2018; Meisels et al., 2024).[25]

Figure 3: Distributions of Estimated Ideal Points by Party Group



Distributions of the fitted ideal points θ_i. Higher values are more rightist. *Left panel*: US House candidate platforms, Democrats in blue and Republicans in red. *Right panel*: Brazilian mayoral candidate platforms, split into the leftist parties (blue) and centrist and rightist parties (red), divided by expert scores (footnote 26 lists the parties in each bloc). Positions are comparable within a panel, not across panels.

Ideal point estimates based on campaign platforms for recent US congressional elections usually show some overlap between the two parties (Case, 2025; Meisels, 2026). My estimates also show this small overlap: 0.5% of Democratic platforms sit above the Republican median and 0.8% of Republican platforms below the Democratic one; and 7.9% of Democrats lean more conservative than the mean House candidate, against 7.3% of Republicans more liberal than average.

Local elections in Brazil show much less separation (Figure 3, right panel). For instance, 34.8% of the platforms from leftist parties are situated to the right of the median candidate of the bloc of center and right-wing parties.[26] Ideology measures based on national legislator surveys indicate that, in recent years, the Brazilian parties are distancing themselves from the center (Zucco and Power, 2024). At the local level, my estimates

[25]For comparison, Figure D.1 in Appendix D.2 plots the US distributions by party under both external benchmarks, DW-NOMINATE and CFScores.

[26]Throughout the paper, I separate the Brazilian parties into two blocs by their placements in the expert survey of Bolognesi et al. (2023). The Left Parties are PSTU, UP, PCO, PCB, PSOL, PC do B, PT, PDT, and PSB, and the Center & Right Parties are the remaining 24 parties the survey places. Table D.4 in Appendix D.4 lists every party with its expert score and estimated position.

show that most candidates are centrist. This relatively nonpolarized distribution broadly resembles the national distributions of Brazilian voters and House representatives found in survey-based estimates (Carroll and Meireles, 2024). The lack of separation between party blocs is not a surprise. Even at the national level, experts place the left-wing Brazilian parties confidently, but they struggle to tell the centrist and rightist parties apart (Bolognesi et al., 2023).

Table 2: Estimated Ideal Points by Party Group

Corpus	Group	n	Mean	Median	SD
United States	Democrats	2,187	-0.209	-0.211	0.150
	Republicans	2,146	+0.213	+0.225	0.146
Brazil	Left Parties	3,299	-0.053	-0.049	0.131
	Center & Right Parties	13,375	+0.013	+0.007	0.113

Summary of the fitted ideal points θ_i within each group of documents. The corpus-wide standard deviation of θ is 0.258 in the United States and 0.119 in Brazil. Footnote 26 lists the parties in each Brazilian bloc. Group means are tied by the centering constraint of Section 3.

In the following subsections, I detail the issues that drive the estimated ideal points in each corpus, first by salience and then by stance. The units for both are supertopics, so that a reader can compare the two corpora on one scheme.[27]

5.1 United States

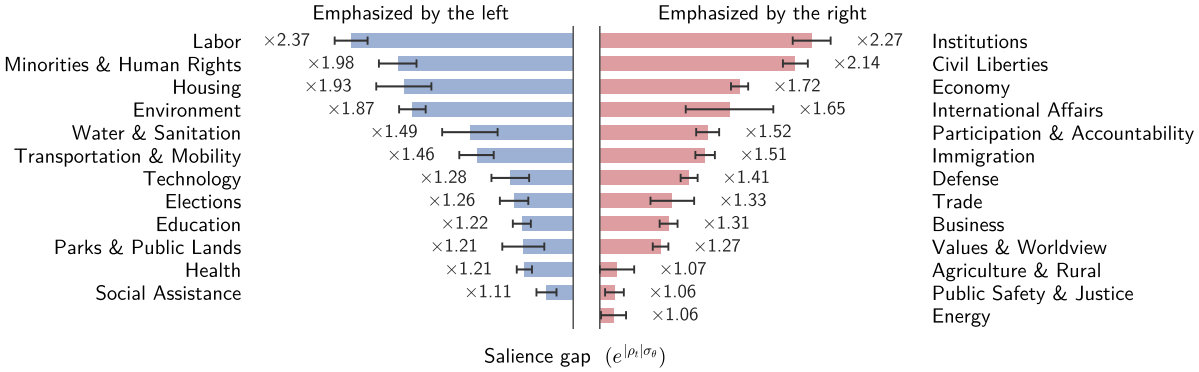
One advantage of the method proposed here is that it names the main issues dividing the political spectrum by salience and by stance. These issue-level parameters allow for an assessment of **construct validity**. Figure 4 ranks the topics for how much they differentiate the poles in the US corpus by salience. The quantity of interest is a multiplicative change in coverage per standard deviation of estimated ideology ($e^{|\rho_t|\sigma_\theta}$), i.e., the factor by which expected coverage differs between two equal-length platforms one standard deviation apart, placed symmetrically about the corpus mean.[28]

[27] Appendix B.2 lists every supertopic with the subtopics it collects, and Appendix E.1 presents issue salience and stance gaps at the subtopic resolution. Results are qualitatively consistent regardless of grouping.

[28] Appendix C.2 proves this interpretation: for comparisons placed symmetrically about the corpus mean, the coverage gap is exact for any curvature κ_t, and it equals the corpus-average change in coverage per standard deviation.

The topics that the left emphasizes most are related to labor, minorities, the environment, and public services. The right, in comparison, disproportionately emphasizes political institutions, civil liberties, the economy, and international affairs. These differences in salience also match the party reputations documented in the issue-ownership literature, which posits Democrats as the owners of welfare, healthcare, and the environment and Republicans as the party of national security, immigration, public safety, and (in some accounts) the economy (Cummins, 2010; Fagan, 2021; Hayes, 2005; Wright et al., 2022). At the subtopic resolution, the same divide appears with narrower names, the left on paid family leave, campaign finance, early childhood education, and climate change, and the right on the Second Amendment, sanctuary cities, trade with China, and critical race theory (Appendix E.1).

Figure 4: Topics Each Side Emphasizes, United States

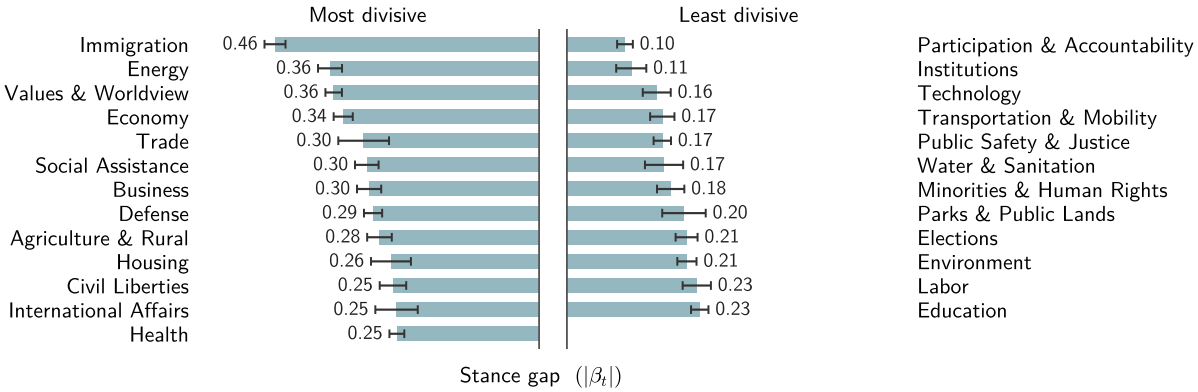


Each bar is a topic's salience gap $e^{|\rho_t|\sigma_\theta}$, the factor by which its expected coverage differs between two equal-length documents one standard deviation of ideology apart, placed symmetrically about the corpus mean; the whisker is the 95% confidence interval and the vertical line marks parity.

I also present the most divisive issues by stance: Figure 5 shows the predicted change in a topic's mean stance when the author's ideal point increases by one standard deviation ($|\beta_t|$), split between most divisive issues (left panel) and least divisive ones (right panel). Immigration divides the two sides most, followed by energy, values, and the economy. These discriminating issues are consistent with previous work on the ideological scaling of US campaign platforms (Meisels, 2026). The consensual end holds participation and accountability, political institutions, and technology.[29]

[29]Public safety is worth a note on subtopic-level stance. Criminal justice reform divides the two sides sharply, while the Second Amendment does not (Appendix E.1, Figure E.3).

Figure 5: Largest and Smallest Stance Gaps, United States

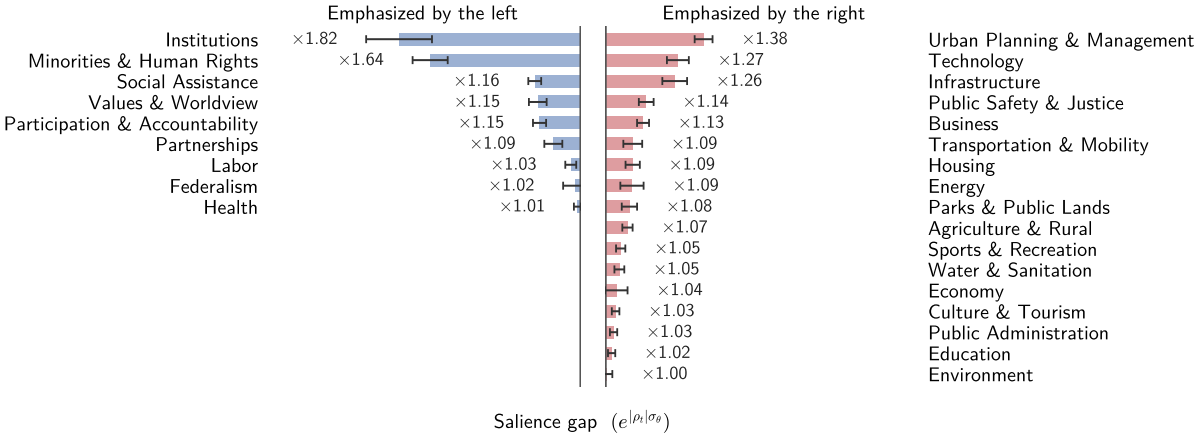


Each bar is a topic's stance gap $|\beta_t|$, the absolute change in its mean stance per standard deviation of ideal point. Whiskers are 95% confidence intervals; vertical line marks zero. Topics' mean stance lies in $[-1, 1]$. *Left Panel*: largest stance divisions; *Right Panel*: smallest stance divisions.

5.2 Brazil

Here, I discuss the topics that discriminate among local candidates in Brazil by salience and stance (Figure 6). The relative differences in salience largely reflect a left–right divide. The left emphasizes political institutions, minorities, social assistance, and values. The right emphasizes urban management, technology, and infrastructure, followed by public safety and business.

Figure 6: Topics Each Side Emphasizes, Brazil



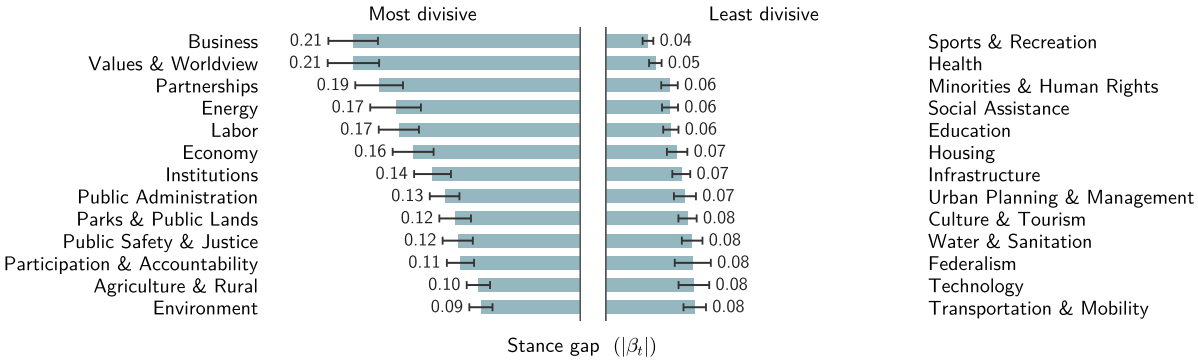
Each bar is a topic's salience gap $e^{|\rho_t|\sigma_\theta}$, the factor by which its expected coverage differs between two equal-length documents one standard deviation of ideology apart, placed symmetrically about the corpus mean; the whisker is the 95% confidence interval and the vertical line marks parity.

Figure 7 ranks topics by stance differences. Divisive issues include business, values, partnerships, energy, labor, and the economy. The consensual end holds sports and recreation, health, minorities, and several public services. Conditional on mentioning

minorities or local public services, most candidates tend to support them, regardless of ideology. As suggested by saliency theory, politicians in this context seem to avoid bringing up issues they do not own (Budge, 2001b, 2015; Petrocik, 1996; Wagner and Meyer, 2014).

Previous attempts to scale Brazilian mayoral campaign platforms produced weak results, misplacing some leftist parties on the rightist side (Pereira, 2021; Salles, 2019; Salles and Guarnieri, 2019). My estimates show the danger of using measures based exclusively on salience in contexts of non-polarized language and large within-party ideological heterogeneity. An inadequate measure invites the incorrect reading that local politics is completely inchoate and non-ideological.

Figure 7: Largest and Smallest Stance Gaps, Brazil



Each bar is a topic's stance gap $|\beta_t|$, the absolute change in its mean stance per standard deviation of ideal point. Whiskers are 95% confidence intervals; vertical line marks zero. Topics' mean stance lies in $[-1, 1]$. *Left Panel*: largest stance divisions; *Right Panel*: smallest stance divisions.

These results are reassuring on the method's capacity to measure ideal points even where language is not polarized. Small relative differences in salience and stance suffice to recover a recognizable left–right dimension.

6 Validation

In order to validate the estimated ideal points, I assess convergent validity, correlating them with benchmarks built outside the texts. I also compare them with unsupervised and language-model baselines estimated on the same corpora.[30]

[30]The two machine learning components are validated separately; the topic model in Appendix B.1 and the stance classifier in Appendix B.3.

Table 3 reports the correlation between the estimated ideal point and three external benchmarks. For the United States, I use two measures of candidate-level ideal points: CFScores (Bonica, 2014), built from campaign contributions, and DW-NOMINATE (Lewis et al., 2026), built from roll-call votes. For Brazil, I use the expert placements of Brazilian parties (Bolognesi et al., 2023). There is no measure for Brazilian local candidates, so all validation exercises are carried out at the level of party means.

Table 3: Correlation of Estimated Ideal Points With External Benchmarks

	United States						Brazil
	CFScore			DW-NOMINATE			Expert
Model	All	Dem.	Rep.	All	Dem.	Rep.	Party level
Unsupervised scaling, bag-of-word							
Correspondence analysis, unigrams	0.53	−0.17	0.13	0.70	0.10	0.40	0.64
Correspondence analysis, bigrams	0.65	−0.17	**0.26**	0.74	0.10	**0.41**	0.54
Wordfish, unigrams	0.72	−0.09	0.25	0.80	0.16	0.38	0.19
Wordfish, bigrams	0.76	−0.03	0.23	0.81	0.22	0.34	0.73
Document embeddings							
Doc2Vec + PCA, unigrams	0.19	−0.23	0.12	0.47	0.02	0.26	0.70
Doc2Vec + PCA, bigrams	0.21	−0.23	0.12	0.49	0.00	0.26	0.69
Pre-trained Embeddings + PCA	0.56	−0.08	0.23	0.57	−0.02	0.38	0.69
Large language models							
Ask and average, GPT-4o mini	**0.87**	0.17	0.20	**0.89**	0.20	0.26	0.82
Ask and average, GPT-5.4 mini	0.85	**0.22**	0.21	0.88	0.19	0.27	0.82
This paper							
Fasti	0.81	0.10	0.20	0.85	**0.29**	0.28	**0.86**
n	3,788	2,020	1,768	1,111	576	535	33

Pearson correlation between the estimated ideal point θ and each external benchmark, with n the documents behind each correlation, or the parties in Brazil. The Brazilian column compares party-mean θ with the expert placements of Bolognesi et al. (2023). The two ask-and-average rows use a sample of 1,000 platforms per corpus (Section 6). Bold marks the best model in each column, and Appendix D.5 details the baselines.

I compare the model with alternative unsupervised and zero-shot approaches in terms of convergent validity with external benchmarks.[31] Wordfish is competitive in the United States: it reaches 0.76 against CFScores and 0.81 against DW-NOMINATE. US congressional platform language is highly polarized (Section 4), so an unsupervised scale recovers most of the between-party signal. In Brazil, the same estimator is fragile. Its party-level correlation moves from 0.19 to 0.73 depending on a single preprocessing choice: whether

[31] Appendix D.5 details the preprocessing steps for the unsupervised models and the hyperparameters used for the embedding-based ones.

to merge strongly associated word pairs (bigrams).[32] This sensitivity to discretionary preprocessing decisions is precisely what Denny and Spirling (2018) caution against in unsupervised text analysis.

Correspondence analysis does not require those adjustments, but it trails the best version of Wordfish in both corpora. Surprisingly, in the applications, bag-of-words models outperform embedding-based scaling (Rheault and Cochrane, 2020). Both approaches disagree with CFScores on the direction of ideological ordering within Democrats, as evidenced by the negative correlation.

Generative Large Language Models Recent work has advocated using generative large language models (LLMs) for text scaling tasks in the social sciences (Benoit et al., 2026; Lehmann et al., 2025). I reproduce the ask-and-average approach suggested by Le Mens and Gallego (2025). The LLM receives the prompt reproduced in Appendix D.5.6 (Figure D.2) and a batch of 50 sentences. It scores each sentence from 0 (extremely left) to 100 (extremely right) and returns `NA` (i.e., missing) when a sentence carries no political content.[33] The estimated ideology is the mean of the non-missing sentence scores within a document.

I draw a sample of 1,000 platforms per corpus, stratified by party: equal split in the United States, and an equal quota across the 33 Brazilian parties, capped at the number of platforms each party filed. Every sentence of every sampled platform is then scored.[34] The LLM achieves high correlations with CFScores and DW-NOMINATE, regardless of whether the scores come from GPT-5.4 mini or GPT-4o mini. Table 4 reports the cost of the exercise. A linear extrapolation to the full estimation sample implies a cost of roughly $320.[35]

[32] Also, the model objective function does not converge unless the vocabulary is reduced and the high-frequency term counts are capped.

[33] To enhance reproducibility, I set the temperature to zero and use a fixed seed. I also use the structured-output interface of the OpenAI API (OpenAI, 2024) to impose a hard schema constraint on the response format, instead of simply prompting the model with a soft request. Each batch returns a JSON object keyed by exactly the batch's sentence identifiers, and each score is an integer from 0 to 100 or `NA`. The schema binds the generation, so the model cannot omit or add an identifier nor return an out-of-range score.

[34] 95,701 sentences in the United States and 202,646 in Brazil.

[35] The four runs were executed in parallel, with six concurrent requests in each.

In the US, Fasti's convergent validity remains comparable to that of the LLMs. However, in Brazil, Fasti achieves a higher correlation with party means, suggesting that the pre-trained LLMs might be less precise in scaling local politics in developing countries than in the US. Additionally, the ask-and-average approach returns a single number per document and does not provide any interpretable parameter, while Fasti decomposes the estimation into issue-loadings that divide the candidates by stance and by salience.

All text-based measures separate candidates within parties weakly in the US, at least in relation to CFScores and DW-NOMINATE, consistent with findings by Case (2025) and Meisels (2026).[36] Future advances in natural language processing may improve topic and stance measurement. Fasti can take whatever topic and stance measures are available and produce ideal points with high construct validity, interpretable in terms of issue salience and stance divisions. In Appendix E, I re-estimate the Fasti model using LLM-based measures of topic and stance. In the body of the paper, I focus on a scalable and cost-free approach.

Table 4: Cost, Running Time, and Validation of the Ask-and-Average Baseline

				Sample (realized)			Full corpus (projected)	
Model	Corpus	Sentences	Tokens (millions)	r[a]	Min.	Cost[b]	Hours	Cost[b]
GPT-4o mini	US	95,701	6.92	0.87	16.5	$1.11	1.3	$5
GPT-4o mini	Brazil	202,646	16.49	0.82	37.8	$2.61	9.6	$40
GPT-5.4 mini	US	95,701	6.96	0.85	15.1	$7.01	1.2	$32
GPT-5.4 mini	Brazil	202,646	16.54	0.82	30.8	$16.01	7.8	$243

Each row is one run of the ask-and-average baseline on a sample of 1,000 platforms. Cost was measured on 22 July 2026, and Min. is wall-clock time with 6 concurrent requests. The last two columns extrapolate cost and runtime linearly to the full corpus.
[a] r is the correlation against CFScores in the United States and party-level against the expert placements in Brazil.
[b] The provider also offers a discounted 24-hour batch tier at half these prices; I report full price.

[36] As Grimmer and Stewart (2013) note, agreement with an external benchmark is evidence of convergent validity, not accuracy. The goal of a text-based measure is not to replicate alternatives based on roll-call votes or campaign donations. Neither benchmark is an undisputed measure, and both are proxies for a latent, non-observable ideal point. I emphasize that Fasti is particularly attractive in terms of evaluating construct validity, because it identifies which issues divide the political spectrum by stance and salience.

Face Validity and Fasti Components In Appendix D.1 (Table D.1), I assess **face validity** for the US corpus, locating well-known US politicians: candidates with known ideological positions are placed on the correct side, and celebrated moderates sit near zero.

Table 5 clarifies the roles each component of Fasti serves. The stance component alone reaches high correlations with baseline measures in both corpora, while the correlations of the topic component are more modest, similar to those found when using bag-of-words models. Taken together, these comparisons suggest that stance helps anchor the scale to the ideological direction and isolate it from other sources of textual variation, while salience adjusts candidates' locations according to the observed differences in issue emphasis.

Table 5: Correlation With External Benchmarks by Model Component

	United States						Brazil
	CFScore			DW-NOMINATE			Expert
Model	All	Dem.	Rep.	All	Dem.	Rep.	Party level
Topic IRT	0.68	0.04	0.17	0.68	0.15	0.29	0.65
Stance IRT	0.79	0.10	0.14	0.84	0.23	0.23	0.84
Fasti	0.81	0.10	0.20	0.85	0.29	0.28	0.86
n	3,788	2,020	1,768	1,111	576	535	33

Pearson correlation between the estimated ideal point θ and each external benchmark. Topic IRT and Stance IRT fit Equations 1 and 4 alone, and Fasti fits both. The Brazilian column compares party-mean θ with the expert placements of Bolognesi et al. (2023). n counts the documents (for Brazil, the parties) behind each correlation.

7 Conclusion

Candidates convey their ideology both by the issues they emphasize and by their stated positions within each (Budge, 2001a; Laver, 2001). Most text-based measures model only one of the two channels. I propose a topic–stance item-response model that estimates ideal points using both inputs: the count of sentences per topic and the mean stance within each topic. The estimated item parameters indicate which issues divide the political

space by stance, which divide it by salience, and which carry little information about the latent dimension.

In contexts where word choice does not reflect the stated positions of political actors, unsupervised scaling typically struggles, since style and other sources of lexical variation can dominate the estimation (Grimmer and Stewart, 2013; Lauderdale and Herzog, 2016). Precisely for that reason, I apply the model to two corpora of campaign platforms, one from US House primary races and one from Brazilian mayoral elections, that differ in how polarized their language is. In the United States, where campaign language is polarized, the estimated ideal points correlate strongly with roll-call and contribution scores, and the loadings separate the issues that divide the parties by stance, such as immigration and energy, from those each side emphasizes, such as labor and international affairs, in accordance with the literature (Fagan, 2021; Meisels, 2026; Wright et al., 2022).

The Brazilian application illustrates why this distinction matters. Earlier scalings of these documents recover the left–right ordering of the parties only partially, severely misplacing some leftist parties (Pereira, 2021; Salles, 2019, 2021; Salles and Guarnieri, 2019). In this corpus, candidates usually express a positive stance toward a wide range of municipal services. Despite these challenges, Fasti orders the Brazilian parties on a left–right dimension compatible with expert surveys (Bolognesi et al., 2023). The Brazilian left at the local level emphasizes minorities, participation, and welfare, and the right focuses on the economy, infrastructure, public management, and public safety. The largest stance gaps are on business, worldview, and values.

The method has four practical advantages. First, the model yields substantive findings about which issues divide the political spectrum by salience and which by stance, so its construct validity is easy to assess, unlike that of measures based on embeddings (Rheault and Cochrane, 2020) or on prompting large language models (Le Mens and Gallego, 2025). Second, Fasti requires neither human labels nor a prior codebook of issues and ideological positions, so it is suitable for scaling corpora where these resources are unavailable. Third, the estimation is robust to discretionary preprocessing decisions (Denny and Spirling, 2018). Finally, the stance classifier is available in nine languages,

and the topic model can be fit on a sample, so the pipeline scales to millions of sentences on modest computational resources.

References

Angelov, D. (2020). Top2Vec: Distributed Representations of Topics. arXiv preprint arXiv:2008.09470.

Arora, S., Liang, Y., and Ma, T. (2017). A Simple but Tough-to-Beat Baseline for Sentence Embeddings. In *International Conference on Learning Representations (ICLR)*.

Bafumi, J., Gelman, A., Park, D. K., and Kaplan, N. (2005). Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation. *Political Analysis*, 13(2):171–187.

Bafumi, J. and Herron, M. C. (2010). Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress. *American Political Science Review*, 104(3):519–542.

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, 23(1):76–91.

Bateman, D. A., Clinton, J. D., and Lapinski, J. S. (2017). A House Divided? Roll Calls, Polarization, and Policy Differences in the U.S. House, 1877–2011. *American Journal of Political Science*, 61(3):698–714.

Bawn, K., Cohen, M., Karol, D., Masket, S., Noel, H., and Zaller, J. (2012). A Theory of Political Parties: Groups, Policy Demands and Nominations in American Politics. *Perspectives on Politics*, 10(3):571–597.

Benoit, K., De Marchi, S., Laver, C., Laver, M., and Ma, J. (2026). Using Large Language Models to Analyze Political Texts through Natural Language Understanding. *American Journal of Political Science*. Advance online publication.

Bestvater, S. E. and Monroe, B. L. (2023). Sentiment Is Not Stance: Target-Aware Opinion Classification for Political Text Analysis. *Political Analysis*, 31(2):235–256.

Bleck, J. and van de Walle, N. (2012). Valence Issues in African Elections: Navigating Uncertainty and the Weight of the Past. *Comparative Political Studies*, 46(11):1394–1421.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Bolognesi, B., Ribeiro, E., and Codato, A. (2023). A New Ideological Classification of Brazilian Political Parties. *Dados*, 66:e20210164.

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science*, 58(2):367–386.

Bouyamourn, A. and Spirling, A. (2026). Interpretable Aggregation of Correlated LLM Annotations. Working paper, Princeton University. Prepared for the 2026 Society for Political Methodology Summer Meeting.

Bucchianeri, P. (2020). Party Competition and Coalitional Stability: Evidence from American Local Government. *American Political Science Review*, 114(4):1055–1070.

Bucchianeri, P., Carney, R., Enos, R., Lakeman, A., and Malina, G. (2021). What Explains Local Policy Cleavages? Examining the Policy Preferences of Public Officials at the Municipal Level. *Social Science Quarterly*, 102(6):2752–2760.

Budge, I. (1982). Strategies, Issues, and Votes: British General Elections, 1950–1979. *Comparative Political Studies*, 15(2):171–196.

Budge, I. (2001a). Theory and Measurement of Party Policy Positions. In Budge, I., Klingemann, H.-D., Volkens, A., Bara, J., and Tanenbaum, E., editors, *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*, pages 75–92. Oxford University Press, Oxford.

Budge, I. (2001b). Validating the Manifesto Research Group Approach. In Laver, M., editor, *Estimating the Policy Positions of Political Actors*, pages 50–65. Routledge, London.

Budge, I. (2015). Issue Emphases, Saliency Theory and Issue Ownership: A Historical and Conceptual Analysis. *West European Politics*, 38(4):761–777.

Budge, I. and Hofferbert, R. I. (1990). Mandates and Policy Outputs: U.S. Party Platforms and Federal Expenditures. *American Political Science Review*, 84(1):111–131.

Budge, I., Klingemann, H.-D., Volkens, A., Bara, J., and Tanenbaum, E. (2001). *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford University Press, Oxford.

Burnham, M., Kahn, K., Wang, R. Y., and Peng, R. X. (2026). Political DEBATE: Efficient Zero-Shot and Few-Shot Classifiers for Political Text. *Political Analysis*, 34(3):329–343.

Carroll, R. and Meireles, F. (2024). Multi-level legislative representation in an inchoate party system: Mass-elite ideological congruence in Brazil. *Party Politics*, 30(1):151–165.

Case, C. R. (2025). Measuring Strategic Positioning in Congressional Elections. *Journal of Politics*. Forthcoming.

Chong, D. and Druckman, J. N. (2007). Framing Theory. *Annual Review of Political Science*, 10:103–126.

Clinton, J., Jackman, S., and Rivers, D. (2004). The Statistical Analysis of Roll Call Data. *American Political Science Review*, 98(2):355–370.

Clinton, J. D. (2012). Using Roll Call Estimates to Test Models of Politics. *Annual Review of Political Science*, 15(1):79–99.

Cummins, J. (2010). The Partisan Considerations of the President's Agenda. *Polity*, 42(3):398–422.

Denny, M. J. and Spirling, A. (2018). Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It. *Political Analysis*, 26(2):168–189.

Desai, Z. and Frey, A. (2023). Can Descriptive Representation Help the Right Win Votes from the Poor? Evidence from Brazil. *American Journal of Political Science*, 67(3):671–686.

Desai, Z., Frey, A., and Tyson, S. A. (2026). Partisan Leaning. *American Political Science Review*, 120(1):72–91.

Downs, A. (1957). *An Economic Theory of Democracy*. Harper and Row, New York.

Druckman, J. N., Kifer, M. J., and Parkin, M. (2009). Campaign Communications in U.S. Congressional Elections. *American Political Science Review*, 103(3):343–366.

Entman, R. M. (1993). Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

Fagan, E. (2021). Issue Ownership and the Priorities of Party Elites in the United States, 2004–2016. *Party Politics*, 27(1):149–160.

Gilardi, F., Alizadeh, M., and Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120.

Godambe, V. P. (1960). An Optimum Property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.

Goplerud, M. (2019). A multinomial framework for ideal point estimation. *Political Analysis*, 27(1):69–89.

Grand, G., Blank, I. A., Pereira, F., and Fedorenko, E. (2022). Semantic Projection Recovers Rich Human Knowledge of Multiple Object Features from Word Embeddings. *Nature Human Behaviour*, 6(7):975–987.

Grimmer, J. and Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Hayes, D. (2005). Candidate Qualities through a Partisan Lens: A Theory of Trait Ownership. *American Journal of Political Science*, 49(4):908–923.

Hinich, M. J. and Munger, M. C. (1996). *Ideology and the Theory of Political Choice*. University of Michigan Press, Ann Arbor.

Humphreys, M. and Garry, J. (2000). Thinking About Salience. Unpublished manuscript.

Jackman, S. (2001). Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking. *Political Analysis*, 9(3):227–241.

Jones, B. D., Larsen-Price, H., and Wilkerson, J. (2009). Representation and American Governing Institutions. *The Journal of Politics*, 71(1):277–290.

Kitschelt, H. (2007). Party Systems. In Boix, C. and Stokes, S. C., editors, *The Oxford Handbook of Comparative Politics*, pages 522–554. Oxford University Press, Oxford.

Lauderdale, B. E. and Herzog, A. (2016). Measuring Political Positions from Legislative Speech. *Political Analysis*, 24(3):374–394.

Laurer, M., van Atteveldt, W., Casas, A., and Welbers, K. (2024). Less Annotating, More Classifying: Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI. *Political Analysis*, 32(1):84–100.

Laver, M. (2001). Position and Salience in the Policies of Political Actors. In Laver, M., editor, *Estimating the Policy Positions of Political Actors*, pages 66–75. Routledge, London.

Laver, M., Benoit, K., and Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331.

Laver, M. and Garry, J. (2000). Estimating Policy Positions from Political Texts. *American Journal of Political Science*, 44(3):619–634.

Le Mens, G. and Gallego, A. (2025). Positioning Political Texts with Large Language Models by Asking and Averaging. *Political Analysis*, 33:274–282.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lehmann, P., Franzmann, S., Al-Gaddooa, D., Burst, T., Ivanusch, C., Lewandowski, J., Regel, S., Riethmüller, F., and Zehnter, L. (2025). Manifesto Corpus. Version 2025-1. Berlin: WZB Berlin Social Science Center / Göttingen: Institute for Democracy Research (IfDem).

Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., and Sonnet, L. (2026). Voteview: Congressional roll-call votes database. https://voteview.com/.

Li, Z. (2018). How Internal Constraints Shape Interest Group Activities: Evidence from Access-Seeking PACs. *American Political Science Review*, 112(4):792–808.

Lowe, W. (2008). Understanding Wordscores. *Political Analysis*, 16(4):356–371.

Lowe, W. (2016). Scaling Things We Can Count. Working paper, Princeton University.

Lowe, W., Benoit, K., Mikhaylov, S., and Laver, M. (2011). Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly*, 36(1):123–155.

Marques, E. C. L., Peres, U. D., and Fischer Armani, G. (2026). What Difference Does Politics Make in Municipal Budgets? Evidence from Brazil. *Local Government Studies*. Advance online publication.

Martin, A. D. and Quinn, K. M. (2002). Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis*, 10(2):134–153.

McCarty, N., Poole, K. T., and Rosenthal, H. (2016). *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press, 2 edition.

McInnes, L., Healy, J., and Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint arXiv:1802.03426.

Meisels, M. (2026). Candidate Positions, Responsiveness, and Returns to Extremism. *The Journal of Politics*, 88(2):681–697.

Meisels, M., Clinton, J. D., and Huber, G. A. (2024). Giving to the Extreme? Experimental Evidence on Donor Response to Candidate and District Characteristics. *British Journal of Political Science*, 54(3):851–873.

Meltzer, A. H. and Richard, S. F. (1981). A rational theory of the size of government. *Journal of Political Economy*, 89(5):914–927.

Merz, N., Regel, S., and Lewandowski, J. (2016). The Manifesto Corpus: A new resource for research on political parties and quantitative text analysis. *Research & Politics*, 3(2).

Mikhaylov, S., Laver, M., and Benoit, K. (2012). Coder Reliability and Misclassification in the Human Coding of Party Manifestos. *Political Analysis*, 20(1):78–91.

Mimno, D. and McCallum, A. (2008). Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, pages 411–418.

Minozzi, W. and Volden, C. (2013). Who Heeds the Call of the Party in Congress? *The Journal of Politics*, 75(3):787–802.

Morris, C. N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*, 78(381):47–55.

Motolinia, L. (2025). When Reelection Increases Party Unity: Evidence from Parties in Mexico. *British Journal of Political Science*, 55:e55.

Nikolaev, D., Ceron, T., and Padó, S. (2023). Multilingual Estimation of Political-Party Positioning: From Label Aggregation to Long-Input Transformers. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9497–9511, Singapore. Association for Computational Linguistics.

OpenAI (2024). Structured Outputs. https://platform.openai.com/docs/guides/structured-outputs. OpenAI API documentation.

OpenAI (2025). GPT-4.1. https://platform.openai.com/docs/models/gpt-4.1. Large language model, API snapshot `gpt-4.1`.

Ornstein, J. T. (2026). Measuring Scalar Quantities from Document Embeddings. Working paper, University of Georgia.

Pereira, L. A. (2021). *Electoral Competition and Platform Choice: A Computational Linguistics Approach Based on Brazilian Data*. Ph.D. dissertation, Insper Instituto de Ensino e Pesquisa, São Paulo.

Peterson, A. and Spirling, A. (2018). Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems. *Political Analysis*, 26(1):120–128.

Petrocik, J. R. (1996). Issue Ownership in Presidential Elections, with a 1980 Case Study. *American Journal of Political Science*, 40(3):825–850.

Poole, K. T. and Rosenthal, H. (1997). *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press, Oxford.

Porter, R., Case, C. R., and Treul, S. A. (2025). CampaignView, a database of policy platforms and biographical narratives for congressional candidates. *Scientific Data*, 12(1):1237.

Power, T. J. and Zucco, C. (2009). Estimating Ideology of Brazilian Legislative Parties, 1990–2005: A Research Communication. *Latin American Research Review*, 44(1):218–246.

Power, T. J. and Zucco, C. (2012). Elite Preferences in a Consolidating Democracy: The Brazilian Legislative Surveys, 1990–2009. *Latin American Politics and Society*, 54(4):1–27.

Proksch, S.-O. and Slapin, J. B. (2010). Position taking in European Parliament speeches. *British Journal of Political Science*, 40(3):587–611.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2010). How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science*, 54(1):209–228.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong.

Rheault, L. and Cochrane, C. (2020). Word embeddings for the analysis of ideological dimensions in political texts. *Political Analysis*, 28(1):112–133.

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

Robertson, D. (1976). *A Theory of Party Competition*. Wiley, London.

Rodriguez, P. L. and Spirling, A. (2022). Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *The Journal of Politics*, 84(1):101–115.

Salles, N. (2019). Programs and Parties: Rethinking Electoral Competition Through Analysis of Brazilian 'Grotões'. *Brazilian Political Science Review*, 13(2):e0004.

Salles, N. (2021). Ideologia e Partidos no Brasil: Reflexão e Prática a Partir dos Programas de Governo. *Revista Política Hoje*, 29(1):252–338.

Salles, N. and Guarnieri, F. (2019). Estratégia Eleitoral nos Municípios Brasileiros: Componente Programático e Alinhamento Partidário. *Revista de Sociologia e Política*, 27(72):e001.

Sältzer, M. (2022). Finding the Bird's Wings: Dimensions of Factional Conflict on Twitter. *Party Politics*, 28(1):61–70.

Sigelman, L. and Buell, Emmett H., J. (2004). Avoidance or Engagement? Issue Convergence in U.S. Presidential Campaigns, 1960–2000. *American Journal of Political Science*, 48(4):650–661.

Slapin, J. B. and Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722.

Stewart, B. M. and Zhukov, Y. M. (2009). Use of Force and Civil–Military Relations in Russia: An Automated Content Analysis. *Small Wars & Insurgencies*, 20(2):319–343.

Stokes, D. E. (1963). Spatial Models of Party Competition. *American Political Science Review*, 57(2):368–377.

Timoneda, J. C. and Vallejo Vera, S. (2025). BERT, RoBERTa, or DeBERTa? Comparing Performance Across Transformers Models in Political Science Text. *The Journal of Politics*, 87(1):347–364.

Tribunal Superior Eleitoral (2020). Candidatos: Eleições municipais 2020 (candidate microdata). Repositório de Dados Eleitorais, https://dadosabertos.tse.jus.br/.

Vafa, K., Naidu, S., and Blei, D. M. (2020). Text-Based Ideal Points. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5345–5357.

Varin, C., Reid, N., and Firth, D. (2011). An Overview of Composite Likelihood Methods. *Statistica Sinica*, 21(1):5–42.

Wagner, M. and Meyer, T. M. (2014). Which Issues do Parties Emphasise? Salience Strategies and Party Organisation in Multiparty Systems. *West European Politics*, 37(5):1019–1045.

Wang, Y. (2024). On Finetuning Large Language Models. *Political Analysis*, 32(3):379–383.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, 61(3):439–447.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25.

Wright, J. M., Clifford, S., and Simas, E. N. (2022). The Limits of Issue Ownership in a Polarized Era. *American Politics Research*, 50(5):694–706.

Zucco, C. and Power, T. J. (2024). The ideology of Brazilian parties and presidents: A research note on coalitional presidentialism under stress. *Latin American Politics and Society*, 66(1):178–188.

Appendix Contents

A Corpus and Text Processing

This first group of appendices documents how the two corpora are built and prepared: the composition of the Brazilian mayoral corpus from the universe of 2020 candidates, and the segmentation pipeline that produces the sentence unit of analysis for both corpora.

A.1 Composition of the Brazilian Mayoral Corpus

Table A.1 traces the Brazilian corpus from the universe of candidates who contested the 2020 mayoral election to the documents that enter the estimation sample, and records what is lost at each step.

Candidates file their platforms with the electoral authority alongside the paperwork of registering a candidacy, and a small number of the documents recovered under a candidate's name are that paperwork rather than a platform, i.e., a power of attorney, a court filing, a certificate stating that nothing is on record against the candidate, a filing receipt, a consent form, a party's convention minutes, a campaign consultant's workbook, a printout of the candidate's registry page, a cover page whose body never arrived. Such a document states no positions. I therefore remove these documents before estimating, rather than leaving the prior to shrink them once they are in.

A document is a platform if it proposes at least one thing for the municipality. Thinness never excludes a document: a platform of ten lines is a platform, and so is the single surviving bullet of a truncated one. A document is not a platform when it is some other kind of document altogether, or when its body is missing. Applying that boundary means detecting proposals, and in this corpus a proposal is rarely a finite clause. It is an infinitive (*ampliar o atendimento*), a deverbal noun with a complement (*construção da praça no distrito*), or several dozen of either separated by semicolons in one run-on paragraph, so counting finite verbs measures prose style rather than content. I flag candidate documents with ten rules, each keyed to one kind of filing and each conjoined with the requirement that the document propose nothing anywhere in its text. No rule uses length.

I read in full every document the rules flagged. Table A.2 lists the excluded documents by kind. Most are filings addressed to a court, a notary, or the electoral registry, and powers of attorney alone account for the majority. The screening applies to Brazil alone, since the US platforms arrive with the biographical section already separated and carry no filings of this kind.

Table A.1: Composition of the Brazilian Mayoral Corpus

Stage	Documents
Mayoral candidates, 2020 ordinary election[a]	19,380
Municipal campaign platforms collected as PDF[b]	17,380
mayoral platforms with a PDF	16,838
mayoral platforms parsed to text	16,830
after removing filings that are not platforms[c]	16,789
entering the estimation sample ($\geq$ 10 sentences)	16,674

[a] The universe of candidates who contested a mayoralty in the 2020 ordinary municipal election (Tribunal Superior Eleitoral, 2020).
[b] The collected campaign-platform PDFs. 542 were filed for a non-mayoral office and are excluded from every model.
[c] Documents filed under a candidate's name that are not campaign platforms, such as a power of attorney or party convention minutes. Appendix A.1 describes the screen.
All 4,507 US platforms belong to congressional candidates, and 4,353 enter the estimation sample under the same rule.

Table A.2: Documents Excluded as Filings Rather Than Platforms

Kind of document	Excluded	Would have been scaled
Court, notarial or registry filing	58	12
No content	8	1
Plan to be filed later	5	5
Annex of tabulated data	2	1
Party convention minutes	2	2
Campaign workbook	1	1
Extract of a longer document	1	1
Contents page and a letter	1	1
Candidate letter or CV	1	1
Shattered extraction	1	1
Total	80	26

Every document listed here was read whole and confirmed not to be a campaign platform. *Excluded* counts the documents in the whole Brazilian corpus; *Would have been scaled* counts those carrying at least ten sentences, so that the length floor would not have removed them anyway. Kinds are the screen's own categories, described in Appendix A.1.

A.2 Sentence Segmentation

The unit of analysis for the topic and ideal point models is the *sentence*. A *single* pipeline segments both corpora. The US texts are scraped from campaign websites and mix prose with bulleted lists, slogans, and section headers. The Brazilian platforms come from scanned PDFs read by optical character recognition and carry scanning artifacts. In both, a naive punctuation splitter *under*-segments undelimited bullet lists and *over*-segments on abbreviations. The pipeline builds on spaCy (Honnibal et al., 2020) and ftfy (Speer, 2019), and proceeds in eight stages. Stage 6 is the only language-specific one.

1. **Normalization.** I repair encoding artifacts with ftfy, normalize to Unicode NFKC, and insert a missing space after sentence punctuation followed by a capital letter, turning *2017.The* into *2017. The*. The substitution spares abbreviations such as *U.S.* and *Ph.D.*

2. **Boilerplate removal.** I delete any line a document repeats three or more times, and any line-opening phrase of at least five words that opens three or more lines and does not end on a short function word.

3. **Line unwrapping.** I close up a line break when the next line opens in lower case or with a digit, when the current line ends on a one- or two-letter word, or when the current line reaches the document's own measure, which I estimate as the ninetieth percentile of line length in that document. I never close up a line ending in a full stop, question mark, exclamation mark, semicolon, or colon, and I leave documents of fewer than ten lines untouched.

4. **Itemized lists.** I remove runs of four or more repeated dashes, dots, or underscores, and spans of text repeated four or more times in a row. I then split on blank lines, hard line breaks, numbered, lettered, and dash enumerators, and bullet glyphs, including arrows, check-marks, and the symbol-font characters common in exported PDFs. I also split at a semicolon when what follows is capitalized or numbered, and I cut a leading run of capitals off the text that follows it, as in *SAUDE Medidas*

de prevencao.

5. **Colon list headers.** A line of at most twelve words ending in a colon is a list header, e.g. *"I support:"*. I attach it as a prefix to each following item of at most eight words, up to the first longer item, and I emit the header once on its own. A unit opening with a bullet is an item, never a header.

6. **Sentence boundaries (language-specific).** I segment each unit with a neural network model from spaCy (Honnibal et al., 2020), the large English pipeline for the US corpus and the large Portuguese pipeline for the Brazilian one. The model sees a unit only when a full stop sits in its interior, and I keep a boundary only where the preceding text ends in a full stop, question mark, or exclamation mark. I extend the Portuguese abbreviation list with every form of six letters or fewer that appears with a trailing period at least twice in either corpus, some 12.5 thousand forms, which I read by hand, and with the nonbreaking-prefix lists of the Moses toolkit (Koehn et al., 2007). Only abbreviations that require a following word enter the list, which holds 583 Portuguese and 263 English forms.

7. **Stranded tails.** I join a unit of at most four words that opens in lower case back to the preceding sentence with a single space.

8. **Length cap.** I split sentences longer than 150 words into overlapping windows of at most 150 words with a 10-word stride.

Result. The pipeline yields 439.3 thousand sentences over 4,507 US candidate–cycle platforms and 3.10 million sentences over 16,830 Brazilian mayoral campaign platforms. Table A.3 reports the length distribution. No sentence exceeds 150 words. The segmented sentences, keyed to their document, party, and policy identifiers, are the input to the topic model of Section 3.2.

Fragment audit. I asked an AI agent running Claude Opus 5 to read 50 platforms per corpus, drawn at random, sentence by sentence and in document order, and to assign

Table A.3: Distribution of Sentence Length After Segmentation

	min	p5	p10	p25	median	p75	p90	p95	max
United States	1	4	6	11	18	25	33	39	150
Brazil	1	2	4	8	15	24	35	43	150

Sentence length in words. Percentiles are computed over all sentences of each corpus. The upper bound of 150 words holds by construction, from the length cap of Stage 8.

every sentence to one of four classes.

- **Good.** One policy statement, one heading, or one boilerplate line, carrying a single topic and a single stance.
- **Mixed.** The sentence spans two topics or two stances, so no single label is right.
- **Split.** The pipeline cut a unit and left a piece that carries no classifiable content on its own.
- **Non-prose fragment.** The sentence is not running prose, such as a table row, a dateline, a signature block, a running header, or a navigation label.

Table A.4 reports the tally. Errors are rare in both corpora, and the mixed sentences come from Brazilian documents that extraction returned with no blank-line structure.

Table A.4: Classification of Segmented Sentences in a Random Sample of Platforms

	United States	Brazil	Both
Documents	50	50	100
Sentences	4,603	7,610	12,213
Non-prose fragments	16	189	205
Mixed	0	18	18
Split	16	48	64
Good (%)	99.7	99.1	99.3

One row per class of the codebook above, over every sentence produced from 50 randomly drawn platforms per corpus. The share of good sentences excludes non-prose fragments from the denominator.

B Topic Model and Stance Classifier

This group documents the measurement pipeline: the topic model that discovers the subtopics, the codebook those topics are grouped into, and the transfer-learned classifier

that scores each sentence's stance.

B.1 Topic-Model Details

A single fixed random seed governs the topic pipeline of Section 3.2, and I record its settings here.

Embeddings. I embed each sentence with a sentence-transformer (Reimers and Gurevych, 2019). Both models return 384-dimensional vectors:

- United States: `all-MiniLM-L6-v2`
- Brazil: `paraphrase-multilingual-MiniLM-L12-v2`

I compute the embeddings once and reuse them at every downstream step.

Reduction and clustering. UMAP (McInnes et al., 2018) reduces the embeddings to 10 dimensions with 60 nearest neighbors, minimum distance 0, and Euclidean metric. HDBSCAN (McInnes et al., 2017) clusters the reduced points with minimum cluster size 400, minimum samples 25, Euclidean metric, and excess-of-mass cluster selection. HDBSCAN labels low-density sentences as noise, 47.9% of sentences in the US fit and 53.0% in the Brazilian fit. I assign each noise sentence to its highest-probability cluster under HDBSCAN's soft cluster membership, and I read a random sample of 200 reassigned sentences per corpus against the assigned topics' key terms. The pipeline runs through the BERTopic implementation (Grootendorst, 2022), with each component supplied explicitly.

Key terms. The candidate vocabulary is unigrams with corpus stop words removed, English for the US corpus and English and Portuguese for the Brazilian one, keeping terms that appear in at least two sentences and at most 95 percent of them, capped at the 50,000 most frequent. I recompute the Brazilian key terms after fitting, on the unchanged cluster assignments, under a stricter vocabulary: terms of at least three letters containing no digits, present in at least 25 training sentences and in at most half of

them. The recomputation changes only the displayed key terms, never a sentence's topic assignment.

Brazilian sample-and-project scheme. I cluster the Brazilian mayoral corpus (3.10 million sentences) from a stratified sample of 150 platforms per party (4,068 platforms; 770,770 sentences), which keeps small parties represented in the topic space, and project the remaining 2.3 million sentences through the *fitted* pipeline. The trained UMAP transform maps each sentence's embedding, and the cluster with the highest soft-membership probability receives it. No re-clustering occurs at projection time. I read a held-out set of 200 sentences drawn from 200 out-of-sample platforms against the topics' key terms.

Screening document-specific topics. Density clustering occasionally produces a topic held together by the idiosyncratic vocabulary of a handful of documents, such as form headers or one municipality's name. I screen for these with a key-term document spread, the median over a topic's top seven key terms of the number of distinct documents containing the term. I dissolve topics whose spread falls below a threshold, one percent of the corpus's platforms in the US (46 platforms) and 20 platforms in Brazil, and reassign their sentences by a majority vote of the 50 nearest neighbors in the embedding space, among sentences of surviving topics. The screen dissolves six US topics, moving 16.3 thousand sentences (3.7% of the corpus), and none in Brazil.

Topic labels. I name each surviving topic with GPT-5.4 mini, one request per topic. The model reads the topic's ten leading key terms and 150 of the topic's own sentences, 50 drawn at random from each stance class and reshuffled together; a topic with fewer than 50 sentences in a stance contributes what it has. The prompt asks for a name of two or three words, in English for both corpora, and the structured-output interface (OpenAI, 2024) caps the answer at four words. I flag any name that uses the fourth word. I set the temperature to zero and use a fixed seed. Figure B.1 reproduces the prompt.[37]

[37]The Brazilian prompt is the same text with three changes: the opening sentence names municipal platforms in Portuguese, the parenthetical examples come from that corpus, and one further instruction requires the name in English whatever the language of the sentences. Both prompts therefore name topics in English.

Two clusters that address the same subject receive the same name. I therefore run a second pass over the collisions alone, in which the topics sharing a name go to the model together in one request, each with the key terms and the sentences the first pass saw, and the model renames them so that no two carry the same name. Figure B.2 reproduces that prompt.[38] I flag a name that still collides. Finally, I read every name against the sentences the model saw and correct the few that misdescribe the topic.

Figure B.1: The Topic-Naming Prompt

```
You will be provided with the leading key terms of one topic from a topic model
and a sample of the sentences assigned to that topic, in random order. The
sentences belong to campaign platforms in English of candidates for the US
Congress. Name the topic.

Answer with a label in English, in Title Case, two or three words. Use one word
only when a single noun names the topic exactly (e.g. Immigration, Abortion,
Taxes). Use "&" or a comma to join two or three subjects that each recur (e.g.
Veterans & War; Childcare & Pre-K; Medicines, Telehealth & Vaccines).

Name every subject that runs through the sample, not only the largest one; name a
single subject only when it dominates the sample. Name the subjects themselves,
and never replace them with a broader word that covers them. A few sentences might
be assigned to the wrong topic: ignore them, and never widen the label to cover
them. Never answer with a vague or filler word, e.g. Miscellaneous, Misc, Various,
General, Generic, Other, Assorted, Mixed, Issues, Topics, Boilerplate, Related. If
the sentences are not policy but parts of the document itself, name that exactly
(e.g. Section Headers, Table Of Contents, Candidate Bios, Campaign Slogan,
Committee Record). Never put a person's name in the label.

You will only respond with a JSON object holding the label. Do not provide
explanations.
```

The prompt the model receives, together with one topic's leading key terms and the sample of its sentences. Line wrapping is set for the page; the prompt is sent as running text.

[38] The Brazilian version differs in the same three ways as the naming prompt (footnote 37).

Figure B.2: The Collision Prompt

```
You will be provided with several topics from a topic model that all received the
same name. For each topic you get its leading key terms and a sample of the
sentences assigned to it, in random order. The sentences belong to campaign
platforms in English of candidates for the US Congress. Name each of these topics
again, so that no two of them carry the same name.

Answer with one label per topic, in English, in Title Case, two or three words.
Use one word only when a single noun names the topic exactly. Use "&" or a comma
to join two or three subjects that each recur.

Name what separates each topic from the others in this set: the subject, the
policy instrument, the population served, or the part of the document the
sentences belong to. Read the samples against one another and name the difference
you find in them; never invent one. Never separate the topics by adding a number,
a letter, or a word such as Other, Further, Additional, or Related. Where two
topics do address the same subject, name each one by the aspect of that subject
its own sentences carry. Never answer with a vague or filler word, e.g.
Miscellaneous, Misc, Various, General, Generic, Other, Assorted, Mixed, Issues,
Topics, Boilerplate, Related. Never put a person's name in the label.

You will only respond with a JSON object whose keys are the topic ids given to you
and whose values are the labels. Do not provide explanations.
```

The prompt for the second pass, sent once per set of topics that the naming pass gave the same name, together with the key terms and sentences of every topic in the set. The identifiers are positional and carry no topic number. Line wrapping is set for the page; the prompt is sent as running text.

B.2 Subtopics Grouped into Supertopics

The topic model discovers subtopics, and the model of Section 3 estimates over the supertopics those subtopics are grouped into. Table B.1 and Table B.2 list every supertopic with the subtopics it collects and the share of the corpus's sentences it carries.

I assign each subtopic by hand, reading the sentences the topic holds rather than its label or its key terms. The codebook is the same in both corpora, and a class with no subtopic in a corpus carries no column there, so the United States has 25 supertopics and Brazil 26. The grouping changes only the columns of the count and stance matrices, and Appendix E compares the subtopic and supertopic estimates.

The subtopics that carry document structure, naming, and templated framing all belong to one class, which the estimates drop. That class holds 25.5% of the Brazilian sentences, against 15.7% of the US ones. I keep the residual "none of the above" class wherever it carries subtopics.

Table B.1: Subtopics Grouped into Supertopics, United States

Supertopic	%	Subtopics
Economy		
Economy	5.8	Banking Regulation; Budget Deficit; Inflation; Jobs & Economy; Tax Reform
Business	3.4	Competition & Regulation; Economic Development & Jobs; Small Business; Small Business Regulation
Labor	2.1	Minimum Wage; Paid Family Leave; Unions & Collective Bargaining; Vocational Training
Agriculture & Rural	1.8	Agriculture & Nutrition; Rural Communities
Trade	0.6	Trade Agreements
Technology	1.1	Net Neutrality & Broadband; Science, Space & Technology
Welfare Policies		
Social Assistance	2.4	Child Care & Tax Credits; Children & Youth; Income Inequality
Health	11.6	Addiction Treatment; Healthcare Reform; Mental Health & Suicide; Opioid Epidemic
Education	6.4	College Affordability; Early Childhood Education; Education Funding; Education Reform; STEM Education; School Choice & Safety; School Choice & Vouchers; Student Loan Debt; Teachers & Testing
Housing	1.5	Affordable Housing; Homelessness
Rights, Values & Identity		
Civil Liberties	2.5	Constitutional Rights; Free Speech & Censorship; Privacy & Surveillance; Religious Freedom
Minorities & Human Rights	2.7	Disability Rights; LGBTQ Rights; Racial Justice & Hate Crimes; Sexual Assault & Harassment; Tribal Sovereignty
Values & Worldview	7.1	Abortion Rights; American Dream & National Unity; Critical Race Theory; Families & Family Separation; Socialism; State, Markets & Society
Immigration	4.7	Abolish ICE; DACA & Dreamers; Dreamers & TPS; Immigration Reform; Sanctuary Cities
Territory, Environment & Infrastructure		
Environment	4.0	Air & Water Pollution; Animal Welfare; Carbon Pricing; Climate Change; Climate Change & Disaster Response; Environment & Climate; Forest Management & Wildfires; Green New Deal; Paris Climate Agreement; Waterways & Flooding
Energy	2.6	Energy Policy; Oil, Gas & Coal; Oil, Gas & Fracking
Water & Sanitation	0.6	Water Infrastructure
Transportation & Mobility	1.8	Transportation Infrastructure
Parks & Public Lands	0.8	Public Lands & Parks
Emergency & Safety		
Public Safety & Justice	7.2	Criminal Justice Reform; Gun Violence Prevention; Marijuana Legalization; Police Reform; Second Amendment Rights
Politics & Foreign Affairs		
Defense	7.5	Veterans & National Defense
Elections	2.9	Citizens United; Democracy & Elections; Gerrymandering; PAC Donations; Public Financing; Voting Rights
International Affairs	0.9	China Trade & Security; Israel-Palestinian Conflict
Participation & Accountability	1.7	Congressional Reform; Political Corruption
Institutions	0.8	Rule Of Law; Term Limits
Others		
Boilerplate†	15.7	Campaign Rhetoric; Candidate Bios & Record; Legislative Records; Representation Pledges; Section Headers & Press Releases; Short Chunks (5 Words or Fewer)

All 96 subtopics the topic model discovers in the United States corpus are listed, each assigned by hand to one supertopic of the codebook, against the subtopic's own sentences rather than its label or its key terms. The supertopic is the unit the model estimates: the columns of (Y_{it}, S_{it}) are summed within it, and the documents and the sentences are untouched. The percentage is the share of the corpus's sentences the supertopic carries. Subtopics are listed by the label assigned to them and separated by semicolons, because several labels contain a comma. †Boilerplate collects the 6 format and register subtopics, and the estimates drop it.

Table B.2: Subtopics Grouped into Supertopics, Brazil

Supertopic	%	Subtopics
Economy		
Economy	0.9	Financial Management & Debt; Property Taxes & Agriculture; Public Finance & Taxation; Public Finances & Economic Development
Business	3.2	Economic Growth & Entrepreneurship; Entrepreneurship & Jobs; Local Business Support
Labor	2.1	Cooperatives & Solidarity Economy; Economic Development & Job Training; Employment & Income; Youth Employment & Vocational Training
Agriculture & Rural	3.4	Agriculture & Rural Development; Community Gardens & Agriculture; Dairy Farming & Genetic Improvement; Fishing & Agriculture; Parks, Agriculture & Rural Producers
Technology	0.4	Digital Inclusion & Public Wi-Fi
Welfare Policies		
Social Assistance	4.7	Child Labor & Sexual Exploitation; Elderly Care & Active Aging; Family & Social Assistance; Food Security & Local Farmers; Social Assistance & Inclusion; Social Assistance & Protection; Social Assistance Centers; Social Protection & Vulnerable Populations; Youth, Children & Adolescents
Health	9.5	Ambulances & Patient Transport; Basic Health Units; COVID-19 Response; Cancer Prevention & Screening; Dengue & Endemic Disease Control; Dental Prostheses & Oral Health; Diabetes, Hypertension & Obesity; Drug Prevention & Treatment; Family Health; Family Health Teams & Coverage; Family Health Units & Hospital; Health Care Access & SUS; Health Scheduling & Records; Health Services & Facilities; Health Surveillance; Health Worker Training & Pay; Home Care & Patient Transport; Hospital Reform & Emergency Care; Maternal & Infant Health; Medical Specialties & Diagnostics; Medical Specialties & Private Health; Medical Tests & Imaging; Mental Health & Drugs; Non-Emergency Patient Transport; Pharmaceutical Assistance & Drug Prevention; Physical Rehabilitation & Drug Prevention; Public Health & Primary Care; Rural Health Care; SAMU & Emergency Care; School Health & Drugs; Surgical Procedures & Exams; Urgent Care & Emergency; Vaccination Campaigns
Education	8.7	Digital Education & School Technology; Drug Prevention & School Psychologists; Early Childhood Education; Education & Citizenship; Education & Foreign Languages; Education & Literacy; Education Policy; Education Quality Indicators; Family & School; Higher Education & Vocational Training; Inclusive Education & Teacher Training; Professional Training & Teacher Development; School Infrastructure; School Infrastructure & Full-Time Education; School Meals & Family Farming; School Transportation; School Uniforms & Supplies; Schools & Adult Education; Teacher Careers & Training
Housing	1.2	Housing & Land Regularization
Culture & Tourism	6.3	Libraries & Reading; Tourism & Culture
Sports & Recreation	4.1	Children's Playgrounds & Recreation; Sports Development
Rights, Values & Identity		
Minorities & Human Rights	2.3	Disability Inclusion; Violence Against Women
Values & Worldview	1.2	Citizenship & Governance Principles; School Security & Civic-Military Schools
Territory, Environment & Infrastructure		
Environment	2.5	Animal Protection; Environmental Education; Environmental Licensing & Conservation
Energy	0.3	Solar Energy
Water & Sanitation	2.9	Sanitation & Drainage; Waste Collection & Recycling; Water Supply & Sanitation
Infrastructure	2.4	Public Lighting; Roads & Bridges; Roads & Rural Access; Streets, Pavement & Sidewalks
Transportation & Mobility	1.6	Bicycle Infrastructure; Public Transportation; Traffic Safety & Education; Traffic Safety & Parking
Urban Planning & Management	2.3	Municipal Works & City Profile; Squares, Buildings & Upkeep; Urban Planning & Development; Urban Public Works
Parks & Public Lands	0.7	Cemeteries & Mortuary Chapels; Tree Planting & Green Areas
Emergency & Safety		
Public Safety & Justice	2.5	Drug Prevention & Anti-Violence; Public Safety & Surveillance
Politics & Foreign Affairs		
Participation & Accountability	3.4	Civic Participation & Citizenship; Education Councils & Plans; Government Transparency; Health Councils & Planning; Municipal Councils & Tutelary Council; Participatory Budget & Planning; Participatory Governance; Public Administration & Anti-Corruption; Public Administration, Transparency & Participation; School Democratic Management
Public Administration	6.6	Administration & Career Plan; Administrative Reform & Civil Service; Administrative Reform & Legislation; Digital Government & Citizen Services; Economic Development; Economic Development & Infrastructure; Procurement & Performance Indicators; Professional Training & Career Plans; Public Administration & Management; Public Servant Pay & Training; Public Servants Training & Hiring; Salaries & Payroll; Vehicle Fleet Renewal

Table B.2: Subtopics Grouped into Supertopics, Brazil *(continued)*

Supertopic	%	Subtopics
Federalism	0.3	Intergovernmental Partnerships & Funding
Institutions	0.3	Political Democracy
Partnerships	0.8	Public-Private Partnerships; Volunteering & Community Organizations
Others		
Boilerplate†	25.5	Administration & Assistance Headers; Agriculture Section Headers; Campaign Platform Introductions; Campaign Pledges & Goals; Campaign Rhetoric & Slogans; Candidate Bios & Tickets; Candidate Experience Preambles; City Vision Rhetoric; Coalitions, Sources & Titles; Contents & Budget Tables; Culture Section Headers; Dot Leaders & Fragments; Education & Introduction Headers; Environment Section Headers; Health Section Headers; Infrastructure Section Headers; Management Model Pledges; Plan Cover Pages; Plan Disclaimers & Values; Plan Framing & Mixed Proposals; Planning & Goals; Population & City Statistics; Problem Diagnosis & Critique; Programme Titles & Plan Sections; Rhetorical Fragments & Quotes; Safety & Housing Headers; Safety Section Headers; Section Headers & Names; Short Chunks (5 Words or Fewer); Social Assistance Headers; Sports Headers & Boilerplate; Table Rules & Headers

All 169 subtopics the topic model discovers in the Brazil corpus are listed, each assigned by hand to one supertopic of the codebook, against the subtopic's own sentences rather than its label or its key terms. The supertopic is the unit the model estimates: the columns of (Y_{it}, S_{it}) are summed within it, and the documents and the sentences are untouched. The percentage is the share of the corpus's sentences the supertopic carries. Subtopics are listed by the label assigned to them and separated by semicolons, because several labels contain a comma. †Boilerplate collects the 32 format and register subtopics, and the estimates drop it.

Table B.3 and Table B.4 report the ten strongest class-based TF-IDF terms of each supertopic, computed over the supertopic groups on the estimation-sample sentences. The terms line up with each supertopic's assigned label in both corpora, and no politician's name reaches the top ten of any class. Boilerplate is the one class whose terms read as generic campaign language ("congress", "fight", "district" in the United States; "governo", "plano", "candidato" in Brazil), since it also collects every sentence of five words or fewer.

Table B.3: Top 10 Keywords per Supertopic, United States

Supertopic	Top keywords
Economy	
Economy	tax, spending, budget, debt, cuts, economy, jobs, businesses, pay, government
Business	businesses, small, jobs, regulations, create, economy, district, government, economic, work
Labor	wage, minimum, unions, workers, leave, paid, hour, training, labor, work
Agriculture & Rural	farmers, farm, agriculture, food, rural, crops, ranchers, programs, production, family
Trade	trade, agreements, deals, american, manufacturing, tariffs, nafta, jobs, workers, free
Technology	internet, broadband, technology, rural, access, speed, neutrality, innovation, net, science
Welfare Policies	
Social Assistance	children, families, child, income, working, wages, unemployment, class, middle, parents
Health	health, care, healthcare, insurance, medicare, costs, drug, affordable, social, medical
Education	education, students, schools, college, teachers, public, loan, children, debt, parents
Housing	housing, homelessness, affordable, home, rent, income, new, families, communities, mortgage
Rights, Values & Identity	
Civil Liberties	rights, freedom, liberty, constitution, protect, god, government, religious, life, speech
Minorities & Human Rights	discrimination, sexual, lgbtq, equality, gender, rights, orientation, identity, act, disabilities
Values & Worldview	women, abortion, life, right, families, government, equal, reproductive, woman, parenthood
Immigration	immigration, border, illegal, citizenship, country, security, legal, law, wall, undocumented
Territory, Environment & Infrastructure	
Environment	climate, change, environmental, water, environment, clean, carbon, air, pollution, protect
Energy	energy, oil, renewable, clean, solar, fuel, gas, wind, fossil, electric
Water & Sanitation	water, drinking, flood, clean, california, dams, infrastructure, supply, projects, reservoir
Transportation & Mobility	infrastructure, transportation, roads, bridges, transit, investment, rail, highway, projects, funding
Parks & Public Lands	lands, parks, natural, beautiful, protect, national, resources, public, conservation, preserve
Emergency & Safety	
Public Safety & Justice	gun, police, law, violence, firearms, criminal, amendment, right, weapons, prison
Politics & Foreign Affairs	
Defense	veterans, military, national, service, war, served, defense, care, support, country
Elections	voting, elections, voter, campaign, money, democracy, political, rights, corporate, candidates
International Affairs	israel, china, palestinian, israeli, states, chinese, peace, united, jewish, ally
Participation & Accountability	congress, washington, politicians, elected, representatives, political, corruption, interests, party, people
Institutions	term, limits, law, constitution, court, years, senate, congress, amendment, supreme

Others	
Boilerplate†	congress, work, fight, support, legislation, rights, house, protect, bill, need

The 10 strongest class-based TF-IDF terms (c-TF-IDF) of each supertopic in the United States corpus, computed directly over the supertopic groups on the estimation-sample sentences – the same sentences and the same short-sentence rule that feed the main fit – rather than pooled from the subtopics' own key terms. Unigrams of at least three letters, no digits, stop words removed, ranked by stem so a plural or other inflected form does not split a term's score; the word shown is the most frequent surface form of its stem. †Boilerplate pools every format and register subtopic together with every sentence of five words or fewer, whatever its original topic, so its own terms are dominated by document structure rather than policy content.

Table B.4: Top 10 Keywords per Supertopic, Brazil

Supertopic	Top keywords
Economy	
Economy	iptu, fiscal, dívida, impostos, milhões, gastos, receita, arrecadação, tributária, financeira
Business	empresas, emprego, empreendedores, pequenas, desenvolvimento, comércio, renda, novas, município, econômico
Labor	trabalho, jovens, emprego, cursos, mercado, parceria, profissionalizantes, qualificação, programa, profissional
Agriculture & Rural	agricultura, produtores, produção, agricultores, familiar, agrícola, rural, produtos, rurais, pequenos
Technology	internet, praças, wifi, pontos, conectada, móvel, comunicação, gratuita, acesso, digital
Welfare Policies	
Social Assistance	social, assistência, idosos, famílias, adolescentes, cras, vulnerabilidade, programa, crianças, atendimento
Health	saúde, atendimento, médico, unidades, atenção, hospital, exames, serviços, pacientes, equipamentos
Education	escolas, educação, ensino, alunos, professores, rede, cursos, profissionais, estudantes, formação
Housing	moradias, casas, habitação, habitacionais, habitacional, construção, regularização, populares, fundiária, programa
Culture & Tourism	cultura, turismo, cultural, turístico, eventos, festas, município, música, histórico, artistas
Sports & Recreation	esporte, futebol, prática, lazer, modalidades, atletas, campeonatos, atividades, quadras, física
Rights, Values & Identity	
Minorities & Human Rights	mulheres, violência, deficiência, contra, pessoas, políticas, sexual, vítimas, negra, combate
Values & Worldview	direitos, social, escolas, humanos, segurança, todos, vida, pública, sociais, pessoas
Territory, Environment & Infrastructure	
Environment	ambiental, animais, ambiente, meio, sustentável, animal, preservação, proteção, municipal, desenvolvimento
Energy	energia, solar, elétrica, fotovoltaica, prédios, renováveis, energética, limpa, públicos, instalação
Water & Sanitation	água, resíduos, lixo, coleta, sólidos, saneamento, esgoto, seletiva, reciclagem, abastecimento
Infrastructure	estradas, iluminação, ruas, pavimentação, manutenção, pontes, vias, pública, vicinais, melhorar
Transportation & Mobility	transporte, trânsito, mobilidade, ônibus, urbana, ciclovias, cidade, sinalização, coletivo, pedestres
Urban Planning & Management	cidade, urbana, município, praças, construção, bairros, públicos, municipal, revitalização, áreas
Parks & Public Lands	arborização, árvores, cemitério, áreas, verdes, plantio, nativas, mudas, praças, ciliares
Emergency & Safety	
Public Safety & Justice	segurança, guarda, polícia, militar, civil, câmeras, monitoramento, policiais, pública, municipal
Politics & Foreign Affairs	
Participation & Accountability	transparência, conselho, pública, gestão, participação, administração, municipal, população, governo, todos
Public Administration	servidores, públicos, serviços, administração, gestão, todos, municipal, município, cargos, governo
Federalism	federal, governo, estadual, recursos, parcerias, buscar, junto, federais, projetos, estado
Institutions	política, poder, partido, cenário, governo, voto, promessas, povo, luta, democracia
Partnerships	privada, parcerias, público, município, desenvolvimento, social, apoio, projetos, entidades, associações
Others	
Boilerplate†	governo, propostas, plano, prefeito, gestão, cidade, administração, pública, candidato, municipal

The 10 strongest class-based TF-IDF terms (c-TF-IDF) of each supertopic in the Brazil corpus, computed directly over the supertopic groups on the estimation-sample sentences – the same sentences and the same short-sentence rule that feed the main fit – rather than pooled from the subtopics' own key terms. Unigrams of at least three letters, no digits, stop words removed, ranked by stem so a plural or other inflected form does not split a term's score; the word shown is the most frequent surface form of its stem. †Boilerplate pools every format and register subtopic together with every sentence of five words or fewer, whatever its original topic, so its own terms are dominated by document structure rather than policy content.

B.3 Stance-Classifier Details

I record here how I build the training labels, the agreement between the two label sources, the entailment hypotheses, the training configuration, and the held-out performance of the stance pipeline of Section 3.3.

Label construction. The pool is the Manifesto Project corpus for twelve countries (Argentina, Brazil, Canada, France, Germany, Italy, Mexico, Portugal, Spain, Sweden, the United Kingdom, and the United States) in nine languages. I map each Manifesto Project category to a dimension (Economy or Society) and a pole (left, right, or neutral) through a hand-built crosswalk, reported in Table B.5, and I exclude categories that conflate opposite poles.[39] Because the archive codes *quasi*-sentences that often continue one another, I keep only standalone full sentences: units that end in terminal punctuation, follow a unit boundary, and contain at least eight words. From this pool I draw up to 1,000 sentences per language–dimension–stance cell (48,982 sentences in total) and recode each with GPT-4.1 (snapshot `gpt-4.1`; OpenAI, 2025), which receives the sentence and a compact restatement of the same codebook: two dimensions, with the poles phrased as state against market on the economy and progressive against conservative on society, and a neutral option available on both. Figure B.3 reproduces the full coding prompt. It defines the two dimensions (D01 economy, D02 society) and five stance tags (S01 state, S02 market, S03 progressive, S04 conservative, S05 neutral), which the crosswalk collapses to the general stance used for training (S01 and S03 to left, S02 and S04 to right, S05 to neutral).

Intercoder agreement by task. Table B.6 reports agreement between the Manifesto Project labels and GPT-4.1, by task and by category within each task. The Manifesto Project coder reads each sentence inside its manifesto and codes it once, and recoding experiments report intercoder reliability for the scheme well below conventional standards (Mikhaylov et al., 2012). GPT-4.1 reads each sentence in isolation.

[39] One example is "Freedom and Human Rights" (201). While support for human rights is a leftist position, economic freedom typically belongs to the agenda of the right.

Figure B.3: The Stance-Coding Prompt Given to GPT-4.1

```
You are a political-text coder. Classify each manifesto fragment by meaning,
independent of language.

Task
- Pick only one dimension tag and one stance tag for each item.
- A sentence that only names a topic, states a problem, or gives a vague goal is
neutral within its dimension.
- Pick the dominant position and dimension if more than one appears.
- Classify the fragment based on its substantive content. Do not consider the
names of politicians or political parties for the classification.
- When judging stance, focus on the direction or position the fragment takes, not
simply topic or keyword identification. If the fragment leans a direction, tag
that stance even when it is subtle; use neutral only when it takes no side.

Dimension tags
- D01: texts about the role of the state versus the market: production,
public/private/mixed companies, regulation, taxation, public spending,
redistribution, and growth-vs-environment trade-off.
- D02: texts about political institutions, law and order, international politics,
social order, tradition, values, identity, rights, and the nation.
- Do not classify as D01 merely because a fragment mentions economic policy
instruments such as employment or spending when the main topic is law & order,
foreign policy, defense, immigration, political institutions, national identity,
religion, tradition, or civic duty.

Stance tags
- S01: The state should steer the economy, regulate, own companies, provide public
services and redistribute income/wealth to address poverty and economic inequality
and protect the environment. Favors: regulation, consumer protection; central
planning; mixed economy (tripartite bargaining, state-guided development plans);
trade protectionism; public spending; price controls; minimum wages;
nationalization and public ownership of companies; Marxist views; welfare state;
public services; public education; labor unions; progressive taxation; cash
transfers and tax breaks/deductions to the poor or the working class; higher taxes
on corporations and the wealthy; environmental protection; or sustainability.
- S02: The market should allocate resources; government should be small and
fiscally disciplined. Favors: free markets, free enterprise, entrepreneurship,
private property; business incentives and supply-side measures (tax breaks, tax
deductions or subsidies to firms and employers); free trade; lower taxation on the
wealthy and capitalists; economic growth and productivity; fiscal orthodoxy
(balanced budgets; reduction of budget deficits; less public spending); limited
welfare; fees for using public services; outsourcing of provision of public
services. Opposes: labor unions. NOTE: S02 includes market allocation, private
enterprise, deregulation, fiscal restraint, free trade, business incentives, or
reduced state provision. Public-private coordination, tripartite bargaining, or
state-led sector planning is S01, not S02.
```

Figure B.3 continues on the next page.

Figure B.3: The Stance-Coding Prompt Given to GPT-4.1 *(Continued)*

```
Stance tags (continued)
- S03: Supports human rights, equality, openness, and an inclusive, secular,
cosmopolitan society. Favors: human and civil rights (free speech, press,
assembly; pro refugee policy); democracy and citizen participation (including
separation of powers); social justice and equality (no discrimination based on
gender, age, sexuality, race or origin); immigration; multiculturalism and
cultural diversity; indigenous rights; minorities (e.g.: LGBT+, people with
disabilities); disarmament; international multilateralism; international
cooperation; peace; transitional justice (e.g., truth commissions). Opposes:
nationalism; traditional values; Christian morality; law and order policies
("tough-on-crime"); criminalization of drugs; militarism; imperialism; foreign
intervention on domestic policies; military invasions in other countries. Does not
include: personal and economic freedom.
- S04: Supports individualism, nationalism, tradition, authority, social cohesion
and the status quo. Favors: patriotism, national pride; personal and economic
freedom; individualism; traditional and Christian religious morality; law and
order ("tough-on-crime"); criminalization of drugs; strong government and
leadership (even if restricting civil liberties or human rights to defend the
nation); civic duty and national cohesion; cultural assimilation (i.e., immigrants
adopting the national culture); militarization; foreign intervention in other
countries; national sovereignty; cooperation with or amnesty for former
authoritarian elites. Opposes: immigration; internationalism; democratic rights;
special indigenous protection.
- S05: Any content with no market-vs-state or progressive-vs-conservative
direction. Includes: vague economic goals with no specific stance; centralization
or decentralization of political power and autonomy; uncontroversial goals (such
as fighting corruption/clientelism); claims of politicians' or party's valence or
competence.
```

The system instruction GPT-4.1 receives, together with a batch of manifesto sentences to label; it returns one dimension tag and one stance tag per sentence. Reproduced verbatim, except that typographic quotation marks are set as plain quotes.

The disagreement is directional. GPT-4.1 recovers 68.2% of the sentences the human coder places on the left but only 43.5% of those placed on the right, and it applies the left label more often, so only 51.1% of its left labels are confirmed by the human coder. I therefore keep for training only the sentences on which the two readings coincide, at the cost of roughly half the sample. Per-language stance agreement ranges from 49.1% (Italian) to 63.0% (Galician), with Portuguese at 49.4% and English at 53.1%, and the dimension task stays between 72% and 82% in every language.

Entailment hypotheses. Each task is phrased as a set of short hypotheses; the classifier scores whether the sentence entails each one. The exact wordings are:

- **T1 dimension:** "This text is about the state and the economy." and "This text

is about political institutions and society."

- **T2 economic pole:** "This text expresses a stance supporting state intervention and redistribution on economic issues," "...a stance supporting free markets and growth on economic issues," and "...a neutral stance on economic issues."

- **T3 social pole:** "This text expresses a {progressive, conservative, neutral} stance on social issues," one hypothesis per filler.

- **T4 general stance (production):** "This text expresses a {leftist, rightist, neutral} political stance," one hypothesis per filler.

Training configuration. The base model is a multilingual NLI classifier (`bge-m3-zeroshot-v2.0`; Laurer et al., 2024), fine-tuned with a binary entailment head. Each labeled sentence contributes two pairs per task, the matching hypothesis as an entailment and one randomly drawn alternative as a non-entailment, in the original language and, where available (82% of sentences), in English machine translation. The four tasks together yield 394,087 pairs from the 43,173 sentences that survive the concordance filter on at least one task. The split allocates 70% of manifestos to training and 15% each to validation and test, grouped by *manifesto*, so no party–election document contributes to more than one of the training, validation, and test sets; I oversample minority classes in the training set only (259,606 training pairs). Training uses maximum sequence length 192, learning rate 1.3×10^{-5} with a linear schedule and 6% warmup, weight decay 0.01, effective batch size 64, and early stopping on validation Matthews correlation (stopped just before the fourth epoch); one A100 GPU trains the model in under an hour.

Held-out performance. Table B.7 reports test-set performance by task. On the production task the classifier reaches Matthews correlation 0.75 and balanced accuracy 0.83, with per-class F1 of 0.86 (left), 0.79 (right), and 0.85 (neutral). Balanced accuracy by language ranges from 0.76 (Galician) to 0.86 (Catalan); English is 0.82 and Portuguese 0.82.

Inference on the target corpora. At inference I score only the three production hypotheses. Following the grouped decision rule of Laurer et al. (2024), I normalize the three entailment scores to probabilities $(p_{\text{left}}, p_{\text{right}}, p_{\text{neutral}})$ that sum to one and take the label with the highest, applying no threshold or post-hoc calibration. The classifier is speaker-blind, seeing the sentence text only, never the party, country, or corpus. Scoring all 3.5 million sentences of the two corpora completes in a single GPU session.

Table B.5: Manifesto Project Categories Mapped to Each Stance Cell

Pole	Manifesto Project categories (with code)
Economy dimension	
Left	Market Regulation (403), Economic Planning (404), Corporatism/Mixed Economy (405), Protectionism (406), Keynesian Demand Management (409), Controlled Economy (412), Nationalisation (413), Marxist Analysis (415), Anti-Growth and Sustainability (416), Environmental Protection (501), Welfare State Expansion (504), Education Expansion (506), Labour Groups (701)
Right	Free Market Economy (401), Incentives (402), Free Trade (407), Economic Growth (410), Economic Orthodoxy (414), Welfare State Limitation (505), Education Limitation (507), Labour Groups Negative (702)
Neutral	Economic Goals (408), Technology & Infrastructure (411), Culture (502), Agriculture & Farmers (703)
Society dimension	
Left	Anti-Imperialism (103), Military Negative (105), Peace (106), Internationalism (107), Human Rights (201.2), Democracy (202), Pre-Democratic Elites Negative and Rehabilitation (305.5–.6), Equality (503), National Way of Life Negative (602), Traditional Morality Negative (604), Law & Order Negative (605.2), Multiculturalism (607), Underprivileged Minorities (705), Non-economic Demographic Groups (706)
Right	Military (104), Internationalism Negative (109), Freedom (201.1), Democracy Negative (202.2), Strong Government (305.3), Pre-Democratic Elites Positive (305.4), National Way of Life (601), Traditional Morality (603), Law & Order (605), Civic Mindedness (606), Multiculturalism Negative (608)
Neutral	Foreign Special Relationships (101–102), European Union (108, 110), Constitutionalism (203–204), Decentralisation (301), Centralisation (302), Political Corruption (304), Political Competence (305.1–.2), Bottom-Up Activism (606.2), Middle Class & Professionals (704)

Each Manifesto Project category (Lehmann et al., 2025) is assigned by hand to one dimension and one pole, extending the economy and society scales of Lowe et al. (2011). A left pole on either dimension maps to a left sentence and a right pole to a right sentence. Where a parent code and its sub-codes share a pole, only the parent is listed. Codes follow the fifth edition of the coding instructions.

Table B.6: Agreement Between Manifesto Project Labels and GPT-4.1, by Task and Category

Task	Category	n	Recall	Precision	κ
T1: Dimension	Economic	24,598	89.5%	73.8%	0.58
	Social	24,384	68.0%	86.5%	
Overall		48,982	78.8%		
T2: Economic pole	State	7,697	71.8%	48.0%	0.27
	Market	7,051	47.2%	70.4%	
	Neutral	7,258	34.0%	42.9%	
Overall		22,006	51.5%		
T3: Social pole	Progressive	5,827	68.7%	62.8%	0.40
	Conservative	5,989	47.3%	76.3%	
	Neutral	4,770	63.3%	46.5%	
Overall		16,586	59.4%		
T4: General stance	Left	17,371	68.2%	51.1%	0.28
	Right	14,907	43.5%	66.9%	
	Neutral	16,704	44.5%	46.1%	
Overall		48,982	52.6%		

Recall is the share of sentences the human coder assigns to a category that GPT-4.1 also assigns to it; precision is the share of GPT-4.1's assignments that the human coder confirms. n counts the sentences the human coder assigns to each category, the overall row gives the raw agreement rate, and κ is Cohen's chance-corrected coefficient. T2 and T3 use the sentences both coders assign to the task's dimension, and T4 is the production task.

Table B.7: Held-Out Test Performance of the Stance Classifier by Task

Task	n	MCC	Balanced accuracy
T1: Dimension	6,176	0.88	0.94
T2: Economic pole	1,678	0.76	0.84
T3: Social pole	1,626	0.77	0.84
T4: General stance	4,172	0.75	0.83

Test sentences come from manifestos never seen in training. MCC = Matthews correlation coefficient. T4 is the production task used for inference on the target corpora.

C Ideal-Point Estimation

This group states the estimator in full and derives the interpretation of its item loadings.

C.1 Estimation Details

Sums over document–topic cells run over the active cells ($Y_{it} > 0$) throughout, and $M_i = \sum_t Y_{it}$. I state here the calibrated objective, the overdispersion correction, the Godambe calibration, the empirical-Bayes shrinkage of the curvatures, the identification transformations, and the sandwich standard errors of Section 3.4.

Stage 1: stance pilot. The stance component alone is a weighted linear factor model, and alternating least squares minimizes its weighted squared error.

$$\sum_{i,t:\, Y_{it}>0} Y_{it} \left(S_{it} - \alpha_t - \beta_t \theta_i\right)^2 \tag{12}$$

The scale θ is re-centered and re-scaled to unit variance after every sweep. The count weight Y_{it} mirrors the variance σ^2/Y_{it} of Equation 4: cells averaging more sentences are more precise. The pilot yields provisional scores θ_i^S, the item parameters (α_t, β_t), and the residual variance $\hat{\sigma}^2$, estimated from the count-weighted stance residuals. All calibration quantities below are computed at this pilot scale and then *frozen*; Stage 2 never revisits them.

Overdispersion of the counts. In both corpora, platforms repeat the same proposals across sections, so the topic counts Y_i are more variable than a multinomial with M_i independent draws. I quantify this overdispersion with a Pearson-type composition statistic per document.

$$R_i \;=\; \sum_t \frac{Y_{it}^2}{M_i\, \hat{\pi}_{it}} \;-\; M_i \tag{13}$$

The expectation of R_i is $B \equiv T - 1$ under the multinomial and $B\, d_i$ under a Dirichlet–multinomial with dispersion κ_{DM}.

$$d_i \;=\; \frac{M_i + \kappa_{DM}}{1 + \kappa_{DM}} \tag{14}$$

Matching the weighted means of R_i and M_i (denoted $\bar{r}$ and $\bar{m}$) gives the method-of-moments estimate $\hat{\kappa}_{DM} = (B\bar{m} - \bar{r})/(\bar{r} - B)$, truncated at the multinomial boundary. The count log-likelihood of document i then enters the objective divided by d_i, the standard quasi-likelihood correction (Wedderburn, 1974): d_i is the factor by which repetition inflates the count variance, so M_i/d_i acts as the document's effective number of independent sentences.

Godambe calibration of the count block. Even after the overdispersion correction, the working count likelihood may misstate its information about θ_i because the model is an approximation. Let u_i be the derivative with respect to θ_i of document i's (overdispersion-weighted) count log-likelihood at the pilot solution, and h_i its expected (Fisher) curvature. The calibration factor is the ratio of the empirical variance of the score to the expected information (Godambe, 1960; Varin et al., 2011).

$$c_Y \;=\; \frac{J}{H}, \qquad J = \sum_i w_i \,(u_i - \bar{u})^2, \qquad H = \sum_i w_i \, h_i \tag{15}$$

Dividing the count block by c_Y makes its effective curvature match the actual variability of its score, so the block contributes to θ_i the information its score carries. The stance block is the reference and keeps $c_S = 1$, since its variance $\hat{\sigma}^2$ is already estimated from its own residuals.

Empirical-Bayes shrinkage of the curvatures. The curvature κ_t is the item parameter most prone to overfitting: a thin topic can acquire a large $|\kappa_t|$ that the data do not support. I place a normal prior on the *standardized* curvature $\tilde{\kappa}_t = \kappa_t \, s_q(\theta)$, where $s_q(\theta)$ is the standard deviation of θ^2 after residualizing on $\{1, \theta\}$. Standardizing makes the prior invariant to the arbitrary latent scale: under $\theta \to s\theta$ the raw curvature scales as $1/s^2$ while s_q scales as s^2, so $\tilde{\kappa}_t$ is unchanged. The prior variance τ^2 is set by marginal likelihood via an EM iteration (Morris, 1983).

$$\tau^2 \;\leftarrow\; \frac{1}{T}\sum_t \left(\hat{\kappa}_t^2 + v_t\right) \tag{16}$$

The iteration alternates between a penalized refit of the count items at the current τ^2 and the update above, in which v_t is the posterior variance of κ_t from the inverse penalized item information. The converged τ^2 is frozen with the rest of the Stage-1 calibration.

Stage 2: the calibrated joint objective. Write each block's contribution for document i as its own loss, in the notation of Equation 10. The stance block contributes the weighted squared error.

$$\ell_i^S \;=\; \sum_{t:\,Y_{it}>0} Y_{it}\left(S_{it} - \alpha_t - \beta_t\theta_i\right)^2 \tag{17}$$

The count block contributes the negative multinomial log-likelihood with the document intercept profiled out.

$$\ell_i^Y \;=\; M_i \log \textstyle\sum_t e^{\eta_{it}} \;-\; \sum_t Y_{it}\,\eta_{it}, \qquad \eta_{it} = \psi_t + \rho_t\theta_i + \kappa_t(\theta_i^2 - 1) \tag{18}$$

Stage 2 minimizes the calibrated objective over θ and all item parameters.

$$\mathcal{L} \;=\; \sum_i w_i \left[\frac{\ell_i^S}{2\hat{\sigma}^2} \;+\; \frac{\ell_i^Y}{d_i\, c_Y} \;+\; \frac{\theta_i^2}{2}\right] \;+\; \frac{1}{2\,\tau_{\mathit{eff}}^2\, c_Y}\sum_t \kappa_t^2 \tag{19}$$

The document weights w_i are unity unless survey weights are supplied, the standard normal prior on θ_i appears as the $\theta_i^2/2$ term, and $\tau_{\mathit{eff}}^2 = \tau^2/s_q(\theta)^2$ is the raw-scale prior variance implied by the frozen standardized prior at the current latent scale. Optimization is by block coordinate descent with monotone backtracking: the stance items (α_t, β_t) solve in closed form, the count items $(\psi_t, \rho_t, \kappa_t)$ take batched Newton steps, and each θ_i takes a safeguarded one-dimensional Newton step with a grid-search fallback for the rare non-concave document. After every sweep the fit is re-identified: θ is centered, the scale is set by $\frac{1}{T}\sum_t \beta_t^2 = 1$, and the item parameters are transformed so that every fitted predictor is exactly preserved; the sign is fixed by orienting a known reference group to the right.

Standard errors. Equation 19 is a *composite likelihood* (Lindsay, 1988; Varin et al., 2011), since it combines the stance and count blocks on calibrated weights rather than multiplying them as independent factors. Under such misspecification the expected curvature H and the score covariance J no longer coincide, so the usual inverse-information formula H^{-1} understates uncertainty. The correct object is the *Godambe information* $\mathcal{G} = HJ^{-1}H$ (Godambe, 1960), whose inverse is the robust covariance.

$$\mathcal{G}^{-1} \;=\; H^{-1}JH^{-1} \tag{20}$$

This covariance is consistent for the variance of the maximizer of a misspecified objective (White, 1982) and is the standard choice for composite likelihoods (Varin et al., 2011).[40] It collapses to H^{-1} exactly when the second Bartlett identity $J = H$ holds, that is, when the objective is a correctly specified likelihood.

I use documents as clusters, since sentences within a platform are dependent. Let g_i denote document i's score contribution to the stacked item parameters at the solution, with the document scores θ_i profiled out through the corresponding cross-derivative blocks. Then $J = \sum_i g_i g_i^\top$ is the empirical score covariance and H the expected curvature of Equation 19. Because the calibration constants $(\hat{\sigma}^2, \hat{\kappa}_{DM}, c_Y, \tau^2)$ are frozen in Stage 1, they are treated as fixed features of the objective rather than estimated parameters; the reported intervals are conditional on that working calibration.

Computational cost. Embedding the sentences, fitting the topic model, labeling them for stance, and estimating the joint model take about two hours for both corpora on a single A100 GPU.

C.2 Interpreting the Salience Gap

Section 5 reads the salience gap $e^{\rho_t \sigma_\theta}$ as the multiplicative change in a topic's expected coverage per standard deviation of ideology. In that signed form the factor runs from

[40] Equation 20 is often called the sandwich estimator, after the form in which J sits between two copies of H^{-1}.

left to right, so it falls below one for a topic the left emphasizes. The figures report the side-specific magnitude $e^{|\rho_t|\sigma_\theta}$, the factor by which the side that emphasizes a topic covers it more, and each panel names that side. Throughout, σ_θ denotes the standard deviation of the estimated ideal points, unrelated to the stance residual variance σ^2 of Equation 4.

The multinomial count block admits an equivalent Poisson form that makes coverage statements direct. Let each count follow an independent Poisson law whose rate combines a document verbosity term a_i with the topic score of Equation 3.

$$Y_{it} \sim Poisson(\lambda_{it}), \qquad \lambda_{it} = exp\{a_i + \eta_{it}\} \tag{21}$$

Conditioning the count vector Y_i on its total M_i recovers Equations 1 and 2 exactly, because the factor e^{a_i} cancels from the numerator and the denominator of Equation 2; the two descriptions are observationally identical. I write $\lambda_t(\theta)$ for the expected number of sentences that a document of fixed length devotes to topic t, its predicted salience at ideology θ.

Consider two documents of equal length at positions θ and θ'. Dividing their predicted saliences cancels the verbosity term, the intercept ψ_t, and the centering constant inside $\kappa_t(\theta^2 - 1)$, all of which are common to both documents.

$$\log \frac{\lambda_t(\theta')}{\lambda_t(\theta)} \;=\; \rho_t\,(\theta' - \theta) \;+\; \kappa_t\,(\theta'^2 - \theta^2) \tag{22}$$

The gap therefore depends on where the pair sits whenever $\kappa_t \neq 0$; a one-standard-deviation move produces different coverage changes at different points of the scale. The comparison the paper reports is the symmetric one. Place the two documents symmetrically about the corpus mean, which the location constraint of Equation 6 fixes at zero, so that $\theta' = +\delta$ and $\theta = -\delta$ for some half-width $\delta > 0$. The two endpoints have equal squares, the curvature term of Equation 22 vanishes for every topic and every half-width, and the contrast reduces to its linear part.

$$\log \frac{\lambda_t(+\delta)}{\lambda_t(-\delta)} \;=\; 2\,\rho_t\,\delta \tag{23}$$

In words, along the whole family of comparisons that straddle the corpus mean, the loading ρ_t is exactly the log-coverage gap per unit of ideological distance, whatever the value of κ_t. Evaluated at one standard deviation of distance, $2\delta = \sigma_\theta$, the gap is exactly $e^{\rho_t \sigma_\theta}$, the factor by which expected coverage of topic t differs between two equal-length platforms one standard deviation apart and symmetric about the corpus mean, a typical center-left author against a typical center-right one.

Two corpus-level identities extend the reading beyond center-straddling pairs. Differentiating $\log \lambda_{it}$ with respect to θ_i gives $\rho_t + 2\kappa_t \theta_i$, and averaging over documents under the location constraint leaves ρ_t alone, so ρ_t is the average marginal effect of ideology on log predicted salience. The discrete version runs the one-standard-deviation comparison centered at each document's own position and averages the resulting log gaps, which Equation 22 evaluates in closed form.

$$\frac{1}{n}\sum_i \log \frac{\lambda_t(\theta_i + \sigma_\theta/2)}{\lambda_t(\theta_i - \sigma_\theta/2)} \;=\; \rho_t\,\sigma_\theta \;+\; 2\,\kappa_t\,\sigma_\theta\,\frac{1}{n}\sum_i \theta_i \;=\; \rho_t\,\sigma_\theta \tag{24}$$

The curvature contributions differ document by document but cancel in the aggregate because the ideal points are centered. Under the linear specification that sets κ_t to zero, Equation 22 has no quadratic term at all, and the gap equals e^{ρ_t} per unit of distance for every pair of documents, with no symmetry requirement.

The figure notes state the gap in terms of a topic's expected share, which adds the normalizing denominator of Equation 2. Write $Z(\theta)$ for that denominator. Applying Equation 23 to every topic inside the sum factors the denominator at the right endpoint as $Z(+\delta) = Z(-\delta)\sum_s \pi_s(-\delta)\, e^{2\rho_s \delta}$, so the ratio of expected shares becomes the topic's own gap over a weighted average of all gaps.

$$\frac{\pi_t(+\delta)}{\pi_t(-\delta)} \;=\; \frac{e^{2\rho_t\delta}}{\bar{G}(\delta)}, \qquad \bar{G}(\delta) \;=\; \sum_s \pi_s(-\delta)\, e^{2\rho_s\delta} \tag{25}$$

The constant $\bar{G}(\delta)$ carries no topic index. It cancels from every comparison across topics: the share of topic t relative to topic r changes by exactly $e^{2(\rho_t - \rho_r)\delta}$, and relative to a topic whose coverage does not respond to ideology, one with $\rho_r = 0$, by exactly $e^{2\rho_t\delta}$.

Every ranking and every relative-emphasis statement in Section 5 is therefore free of the normalization; only the absolute level of a single topic's share carries it.

The latent scale admits the reparameterization $\tilde{\theta}_i = c\,\theta_i$, $\tilde{\rho}_t = \rho_t/c$, $\tilde{\kappa}_t = \kappa_t/c^2$, with the leftover constants absorbed into the free intercepts ψ_t, and this leaves every η_{it}, hence the likelihood, unchanged. The loading ρ_t alone is therefore arbitrary, while the product $\rho_t\sigma_\theta$ is invariant, since $\tilde{\rho}_t\tilde{\sigma}_\theta = (\rho_t/c)(c\,\sigma_\theta) = \rho_t\sigma_\theta$. Quoting the gap per standard deviation is what makes it scale-free and comparable across the two corpora.

D Validation and Comparison

This group collects the convergent-validity checks, the face-validity placements, and the comparisons against alternative scaling and topic models.

D.1 Estimated Positions of Notable US Candidates

Table D.1 places eleven well-known US politicians on the estimated scale as a face-validity check at the candidate level. Candidates widely regarded as being on the far left, such as Pramila Jayapal, Alexandria Ocasio-Cortez, and Ilhan Omar, sit more than a full party standard deviation to the left of the Democratic mean. Lauren Boebert and Marjorie Taylor Greene, far-right politicians, sit comparably far into the Republican tail. Famous moderates fall near zero, including Henry Cuellar, the most conservative House Democrat of this period, and Brian Fitzpatrick, its most moderate Republican. The estimates are also stable across cycles. For instance, Ocasio-Cortez's three platforms span -0.42 to -0.30, and Jayapal's three sit within 0.01 of one another.

Four placements depart from public perception. First, Matt Gaetz is widely perceived as hard right, but his platform reads mid-party, consistent with his mid-party roll-call record. Second, Liz Cheney's 2022 platform lands at zero because 71% of its sentences are legislative record items (i.e., bills cosponsored and letters signed on agriculture, minerals, and public lands), a procedural register that carries little stated stance. Third, Elise Stefanik's 2018 platform lands left of zero: her early campaigns emphasized trade

with Canada, agricultural visas, and environmental protection. Finally, Jared Golden, a moderate by roll-call record, reads as a left Democrat because his platforms emphasize unions, veterans' programs, and labor protections. In all four cases, the measure summarizes what the candidates chose to say to voters, which is related to, but not the same as, how they later vote (Meisels, 2026). The within-party correlations of Section 6 quantify that distinction.

Table D.1: Estimated Positions of Notable US Candidates

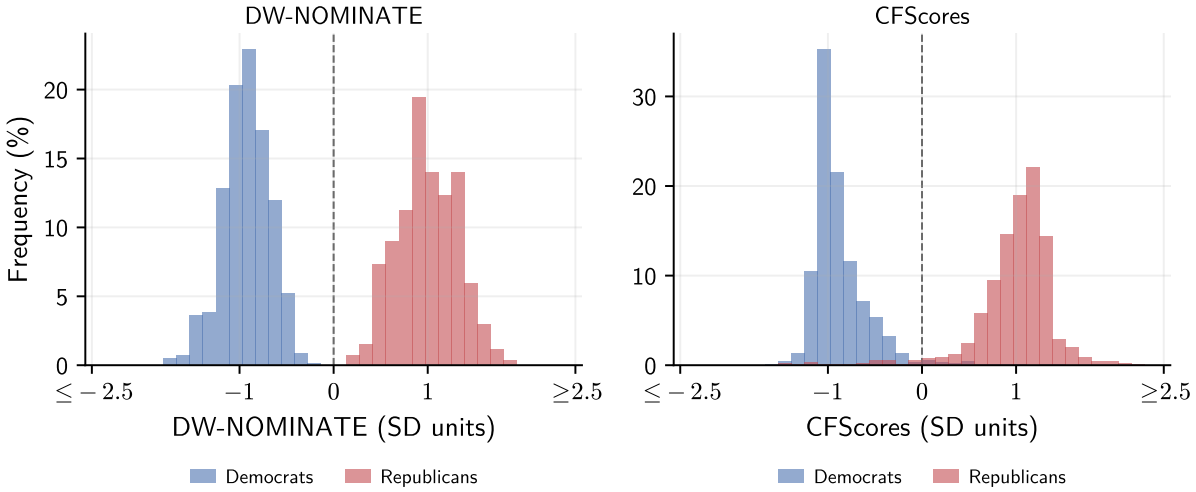
Candidate	District	Cycle	θ
Pramila Jayapal (D)	WA-7	2022	−0.37
Alexandria Ocasio-Cortez (D)	NY-14	2022	−0.49
Ilhan Omar (D)	MN-5	2022	−0.43
Jared Golden (D)	ME-2	2022	−0.28
Henry Cuellar (D)	TX-28	2020	−0.01
Brian Fitzpatrick (R)	PA-1	2020	+0.07
Liz Cheney (R)	WY-AL	2022	−0.03
Elise Stefanik (R)	NY-21	2018	+0.01
Matt Gaetz (R)	FL-1	2022	+0.25
Lauren Boebert (R)	CO-3	2022	+0.37
Marjorie Taylor Greene (R)	GA-14	2022	+0.46

Fitted ideal point of the candidate's platform in the listed primary cycle; each candidate–cycle platform is scaled as its own document. For reference, the Democratic mean is −0.21, the Republican mean +0.21, and the corpus standard deviation 0.26. Candidates were selected before inspecting their estimates; the accompanying text discusses the exceptions.

D.2 Benchmark Distributions by Party

Figure D.1 plots the positions of the US House candidates by party under the two external benchmarks of Section 6, DW-NOMINATE and CFScores. Both benchmarks show a bimodal, nearly disjoint distribution, with less overlap than the platform-based scales. Coverage differs, since DW-NOMINATE exists only for the 1,111 candidate–cycles of winners, while CFScores cover 3,769.

Figure D.1: Distributions of Benchmark Positions by Party



Distributions of DW-NOMINATE (left panel) and CFScores (right panel) for the US House candidates of the estimation sample, Democrats in blue and Republicans in red, in standard-deviation units of each benchmark. Bar height is the percent of the group's platforms in the bin; the outermost bins collect platforms beyond the axis bounds. The format matches Figure 3.

D.3 Most Covered Topics by Party Group

Table D.2 lists the fifteen most covered US topics with each party's coverage share and mean stance. Health is the most covered topic for both parties (15.1% of Democratic sentences against 11.3% of Republican ones), followed by public safety, education, and defense. The parties differ on these topics in both salience and stance. Republicans devote nearly twice as much coverage to immigration as Democrats do (7.7% against 4.3%) and take the opposite stance on it (+0.48 against −0.32); values and worldview splits +0.10 against −0.48, and the economy +0.54 against +0.05. Civil liberties ranks eighth among Republicans and twentieth among Democrats, while minorities and human rights ranks eighth among Democrats and nineteenth among Republicans. Health, education, and the environment lean left for both parties, more so for Democrats, and public safety reads right for both (+0.37 and +0.60).

Table D.3 does the same for Brazil. The most covered topics are the municipal service portfolio, health, education, public administration, and culture and tourism, and the two blocs barely differ on them. The coverage shares are nearly identical, meaning that almost every difference is one of degree. The Left Parties read more left on every topic and cover minorities and human rights more (5.1% against 2.4%), while public safety reads right

for both blocs (+0.48 and +0.60) and business is the one topic the two blocs place on opposite sides of neutrality (−0.01 against +0.15). In Brazil, therefore, the identification of the scale rests on many small differences with the same sign from a shared baseline rather than on opposed positions.

Table D.2: Most Covered Topics, United States

Topic	Democrats			Republicans		
	Rank	Share	Stance	Rank	Share	Stance
Health	1	15.1%	−0.61	1	11.3%	−0.24
Defense	4	7.6%	−0.02	2	11.1%	+0.33
Public Safety & Justice	2	9.2%	+0.37	6	7.3%	+0.60
Values & Worldview	5	7.5%	−0.48	3	10.0%	+0.10
Education	3	8.5%	−0.61	7	6.0%	−0.25
Economy	7	5.1%	+0.05	4	9.8%	+0.54
Immigration	9	4.3%	−0.32	5	7.7%	+0.48
Environment	6	5.7%	−0.67	11	3.0%	−0.40
Business	11	3.6%	+0.08	9	4.9%	+0.46
Elections	10	4.1%	−0.51	14	2.1%	−0.15
Minorities & Human Rights	8	4.4%	−0.83	19	1.2%	−0.45
Energy	14	3.1%	−0.69	10	3.0%	−0.13
Civil Liberties	20	1.6%	−0.19	8	5.3%	+0.16
Social Assistance	13	3.1%	−0.40	13	2.5%	+0.02
Labor	12	3.4%	−0.72	20	1.0%	−0.34

The fifteen most covered topics of the corpus, ordered by overall coverage. Rank is the topic's position in the group's own coverage ranking; Share is the percent of the group's sentences assigned to the topic; Stance is the mean of the group's sentence codes on the topic, on the $[-1, 1]$ scale of Section 3.3. Topics carry their assigned labels.

Table D.3: Most Covered Topics, Brazil

Topic	Left Parties			Center & Right Parties		
	Rank	Share	Stance	Rank	Share	Stance
Health	1	11.7%	−0.79	1	13.0%	−0.79
Education	2	11.3%	−0.72	2	11.7%	−0.70
Public Administration	3	9.2%	−0.41	3	8.8%	−0.31
Culture & Tourism	4	8.1%	−0.17	4	8.6%	−0.11
Social Assistance	5	6.6%	−0.80	5	6.3%	−0.78
Sports & Recreation	8	4.8%	−0.20	6	5.8%	−0.17
Participation & Accountability	6	5.2%	−0.59	8	4.4%	−0.49
Agriculture & Rural	10	4.2%	−0.49	7	4.6%	−0.42
Business	9	4.2%	−0.01	9	4.4%	+0.15
Water & Sanitation	11	3.8%	−0.74	10	4.0%	−0.71
Public Safety & Justice	13	3.0%	+0.48	11	3.5%	+0.60
Environment	12	3.2%	−0.80	12	3.4%	−0.76
Infrastructure	16	2.7%	−0.22	13	3.3%	−0.15
Urban Planning & Management	14	3.0%	−0.31	14	3.2%	−0.24
Minorities & Human Rights	7	5.1%	−0.93	16	2.4%	−0.87

The fifteen most covered topics of the corpus, ordered by overall coverage. Rank is the topic's position in the group's own coverage ranking; Share is the percent of the group's sentences assigned to the topic; Stance is the mean of the group's sentence codes on the topic, on the $[-1, 1]$ scale of Section 3.3. Topics carry their assigned labels. Footnote 26 lists the parties in each bloc.

D.4 Estimated Positions, Party by Party

Table D.4 lists the Brazilian parties in the order the expert survey places them, together with the mean, median, and dispersion of the platforms their candidates filed. The parties at the far left of the expert scale, PSTU and PSOL, hold the leftmost estimated positions, and NOVO the rightmost. The many parties the expert survey places near the center of the scale are barely distinguishable from one another, consistent with the results of Section 5.

D.5 Unsupervised and Embedding Baselines

Table 3 compares the models of this paper with several baselines. I document the four unsupervised baselines here, and Section 6 documents the language-model baseline, whose prompt Appendix D.5.6 reproduces.

All four baselines read the same documents as the models of this paper. In Brazil, that

Table D.4: Estimated Positions of Brazilian Parties, Ordered by Expert Placement

Party	Expert	n	Mean θ	Median θ	SD θ
PSTU	0.51	45	−0.290	−0.297	0.110
UP	0.51	15	−0.292	−0.299	0.053
PCO	0.61	7	−0.333	−0.334	0.002
PCB	0.91	6	−0.254	−0.252	0.117
PSOL	1.28	312	−0.152	−0.180	0.139
PC DO B	1.92	230	−0.044	−0.045	0.121
PT	2.97	1,100	−0.081	−0.092	0.122
PDT	3.93	825	−0.001	−0.007	0.111
PSB	4.05	759	−0.007	−0.007	0.101
REDE	4.77	114	−0.009	−0.014	0.116
CIDADANIA	4.93	470	+0.001	−0.005	0.103
PV	5.29	249	+0.002	+0.003	0.111
PTB	6.10	640	+0.014	+0.011	0.103
AVANTE	6.32	375	+0.002	−0.003	0.113
SOLIDARIEDADE	6.50	438	+0.009	+0.003	0.106
PMN	6.88	99	+0.036	+0.011	0.139
PMB	6.90	53	+0.007	+0.046	0.117
MDB	7.02	1,699	+0.007	+0.003	0.115
PSD	7.09	1,434	+0.005	+0.000	0.106
PSDB	7.11	1,136	+0.009	+0.005	0.101
PODE	7.24	508	+0.024	+0.016	0.108
PRTB	7.45	265	+0.048	+0.043	0.117
PROS	7.47	249	+0.009	+0.007	0.115
REPUBLICANOS	7.78	728	+0.010	+0.004	0.117
PL	7.78	834	+0.009	+0.002	0.107
PTC	7.86	128	+0.053	+0.026	0.189
PSL	8.11	627	+0.037	+0.030	0.110
DC	8.11	112	+0.039	+0.028	0.141
NOVO	8.13	27	+0.177	+0.175	0.085
PP	8.20	1,349	+0.004	−0.000	0.109
PSC	8.33	444	+0.020	+0.012	0.114
PATRIOTA	8.55	375	+0.041	+0.035	0.117
DEM	8.57	1,022	+0.016	+0.006	0.120

Parties are ordered by the expert left–right placement of Bolognesi et al. (2023), on which higher means further right. n is the number of the party's platforms in the estimation sample; mean, median, and SD summarize their ideal points θ_i. Across the 33 parties, party-mean θ and the expert score correlate at $r = 0.86$. Parties without an expert placement are estimated with the rest but omitted from this table.

means mayoral platforms only: I exclude from every model several hundred documents filed for vice-mayoral and city-council candidacies or for supplementary elections. The US corpus contains only congressional candidate platforms, so no analogous exclusion is needed.

D.5.1 Shared Preprocessing

The three baselines that read text as tokens share the same token stream. I lowercase the text, discard numbers and standalone symbols, and remove stopwords, except negation words, which survive as terms of their own, so that a platform promising not to raise taxes does not read as one promising to raise them. I also bar negations and sentence boundaries from joining any word pair, so no pair straddles two sentences.

The one discretionary choice I vary is whether to merge strongly associated word pairs into single terms. I preprocess each corpus twice: once as plain unigrams, and once with the two thousand most strongly associated pairs merged, ranked by normalized pointwise mutual information among pairs occurring at least fifty times. Nothing else differs between the two variants, so the sensitivity reported in Table 3 is attributable to this single decision.

I do not discard rare terms. In every specification, I fold them into a single out-of-vocabulary bucket, which preserves document lengths. Table D.5 reports the vocabulary each specification ends with.

Table D.5: Vocabulary Size Under Each Specification

				Word-count scaling		Embeddings
Corpus	Variant	Tokens (millions)	Types	Features	Fitted	Types
United States	Unigrams	5.1	54,778	6,590	6,590	15,245
	With bigrams	4.7	56,716	8,088	8,088	17,035
Brazil	Unigrams	40.3	230,805	14,850	8,000	26,009
	With bigrams	39.0	232,575	15,854	8,000	26,831

Tokens counts running words after non-negation stopwords are removed, and Types the distinct tokens in that stream. Features is the vocabulary the two word-count models see after stemming, bucketing, and folding rare types into an out-of-vocabulary bucket (Appendix D.5.2). Fitted reports the features that enter Wordfish. The embedding column reports the unstemmed types Doc2Vec retains.

D.5.2 Wordfish and Correspondence Analysis

Both models scale documents from word counts alone. Correspondence analysis takes the first dimension of the document–term matrix (Lowe, 2008). Wordfish (Slapin and Proksch, 2008) places documents on a latent scale through a Poisson model of term

frequencies. For these two models, I stem the vocabulary, merge accent variants, and bucket short tokens that are not acronyms, which in these corpora are overwhelmingly scanning fragments. Terms occurring at most twenty times, or in at most ten documents, fold into the out-of-vocabulary bucket.

In Brazil, Wordfish requires three additional adjustments before it produces finite estimates, because on the full vocabulary a handful of terms separate the likelihood and the Poisson rate overflows. First, I cap per-document term counts at fifty, so no single repeated term dominates a document. Second, I relax the convergence tolerance by one order of magnitude, to 10^{-6}. Finally, I reduce the vocabulary further, escalating the reduction only when a fit returns no finite estimates, which in Brazil stops at eight thousand terms. In contrast, Wordfish converges on the full US vocabulary, and correspondence analysis requires none of these adjustments in either corpus.

D.5.3 Doc2Vec

The embedding baseline follows Rheault and Cochrane (2020). I train document embeddings on the corpus itself, with no pretraining, and take the first principal component of the document vectors as the scale. I use the distributed-memory architecture with 200 dimensions, a context window of 15 words, 15 negative samples, and subsampling of frequent terms at 10^{-4}. Training runs for at most fifteen passes over the corpus and stops when the median cosine change in document vectors between successive passes falls below 0.01. Each platform is one tagged document, the level at which I define the scale.

For rare terms, I follow Case (2025), who implements this design for congressional campaign text. I mask terms occurring at most ten times into a single bucket rather than drop them, and I keep that threshold for the United States. The Brazilian corpus is an order of magnitude longer, so the same absolute count admits far thinner terms; there, I require a term to occur at least thirty times and in at least ten documents.

D.5.4 Sentence Embeddings and Principal Components

The last baseline scales documents from the same sentence embeddings the topic model reads (Appendix B.1), without learning any representation from the target corpus. I normalize each sentence embedding to unit length, average the sentence vectors within each document, and take the first principal component of the document-level matrix as the estimated position. The baseline never reads a token stream, so none of the preprocessing decisions of Appendix D.5.1 apply to it. Doc2Vec requires an architecture, a vector dimension, a context window, a negative-sampling count, a subsampling rate, a stopping rule, and a rare-term threshold, while this baseline requires only the pre-trained encoder and the pooling rule.

D.5.5 What the Baselines Show

Two results in Table 3 stand out. First, Wordfish is competitive in the United States, where it reaches 0.76 against CFScores and 0.81 against DW-NOMINATE. The language of US congressional platforms is highly polarized (Section 4), so a model that reads only word frequencies recovers most of the between-party signal without supervision.

Second, the same model is fragile in Brazil, where the single choice of whether to merge word pairs moves the party-level correlation from 0.19 to 0.73. The fit provides no diagnostic for that choice, and Denny and Spirling (2018) document the same sensitivity for unsupervised text analysis. Moreover, the recovered dimension is difficult to interpret: its poles mix ideology with proper nouns and administrative boilerplate, and in the unigram specification the highest-loading terms are largely candidate names.

The joint model avoids these problems. The supervised sentence codes anchor its scale, it separates stance from salience, and its estimates do not depend on a vocabulary threshold or a count cap.

D.5.6 The Language-Model Baseline Prompt

The ask-and-average baseline of Le Mens and Gallego (2025) scores each sentence with the prompt in Figure D.2. I fill two bracketed fields per corpus and leave the rest of the

prompt identical across corpora and models.

Figure D.2: The Sentence-Scoring Prompt

```
You will be provided with the text of multiple sentences belonging to campaign
platforms in {language} of candidates for {context}. Each sentence is labeled with
an id. Where does each sentence stand on the "left" to "right" wing scale? For
each sentence, provide your response as a score between 0 and 100 where 0 means
"Extremely left" and 100 means "Extremely right". If a sentence does not have
political content, set the score to "NA". Judge each sentence on its own text
alone. You will only respond with a JSON object with the ids as keys and the
scores as values. Return exactly one answer for every item, using the item's id as
the key. Do not provide explanations.
```

The prompt the sentence-scoring baseline receives, together with a batch of sentences. `{language}` is *English* for the United States and *Portuguese* for Brazil; `{context}` is *the US Congress* and *mayor in Brazil*. Line wrapping is set for the page; the prompt is sent as running text.

D.6 Topic Quality Against Alternative Topic Models

This appendix compares the topics of Section 3.2 with four classical topic models, latent Dirichlet allocation (LDA; Blei et al., 2003), non-negative matrix factorization (NMF; Lee and Seung, 1999), the structural topic model (STM; Roberts et al., 2014), and the Dirichlet-multinomial regression topic model (DMR; Mimno and McCallum, 2008). I score every model on coherence, which tracks human judgments of topic interpretability (Bouma, 2009; Lau et al., 2014), and on diversity, which falls when topics share key terms (Dieng et al., 2020; Grootendorst, 2022). The two measures capture what Grimmer and Stewart (2013) call the semantic validity of a topic model, the requirement that its topics form coherent groups, internally homogeneous yet distinct from one another. Coherence scores the within-topic homogeneity and diversity the across-topic distinctness.

I fit every model at a topic count comparable to the pipeline's, 108 in the United States and 158 in Brazil, on the same documents as the models of this paper. The input to the four alternatives is the bigram token stream of Appendix D.5.1 under the out-of-vocabulary thresholds of Appendix D.5.2, i.e., 12,489 unstemmed terms in the United States and 30,812 in Brazil, so the comparison introduces no new preprocessing decision. For STM, the topic-prevalence covariates are candidate, state, and party in the United States and state and party in Brazil, where each candidate files a single platform and

a candidate factor would saturate the design. For DMR, the same covariates enter the topic-prevalence prior through a log-linear link (Mimno and McCallum, 2008). I initialize LDA and DMR at random, NMF by nonnegative double singular value decomposition, and STM by deterministic spectral initialization. I fit LDA and DMR in tomotopy, NMF in scikit-learn, and STM in stm (Lee, 2022; Pedregosa et al., 2011; Roberts et al., 2019). Sentence embedding is the costly stage of the BERTopic time in Table D.6, accounting for 66.8 percent of the wall clock in the United States and 90.9 percent in Brazil.

Coherence averages the normalized pointwise mutual information over all pairs of a topic's top ten terms (Grootendorst, 2022).

$$NPMI(w_i, w_j) \;=\; \frac{\log \frac{P(w_i, w_j)}{P(w_i)\,P(w_j)}}{-\log P(w_i, w_j)} \tag{26}$$

Here, $P(w_i, w_j)$ is the probability that the two terms co-occur within a sliding window of ten tokens in the corpus, and $P(w_i)$ the probability of observing term w_i in a window. The measure is bounded between -1 and 1, and Table D.6 reports the mean across topics. Diversity is the share of unique terms across all topics' top ten.

$$Diversity \;=\; \frac{\left|\bigcup_{t=1}^{T} \mathcal{W}_t\right|}{10\,T} \tag{27}$$

Here, $\mathcal{W}_t$ is the set of the ten top terms of topic t and T the number of topics, so diversity equals one when no two topics share a key term. The few key terms of the BERTopic topics that fall outside the reference vocabulary drop from the scored lists; 99.8 percent are scored in the United States and 100.0 percent in Brazil.

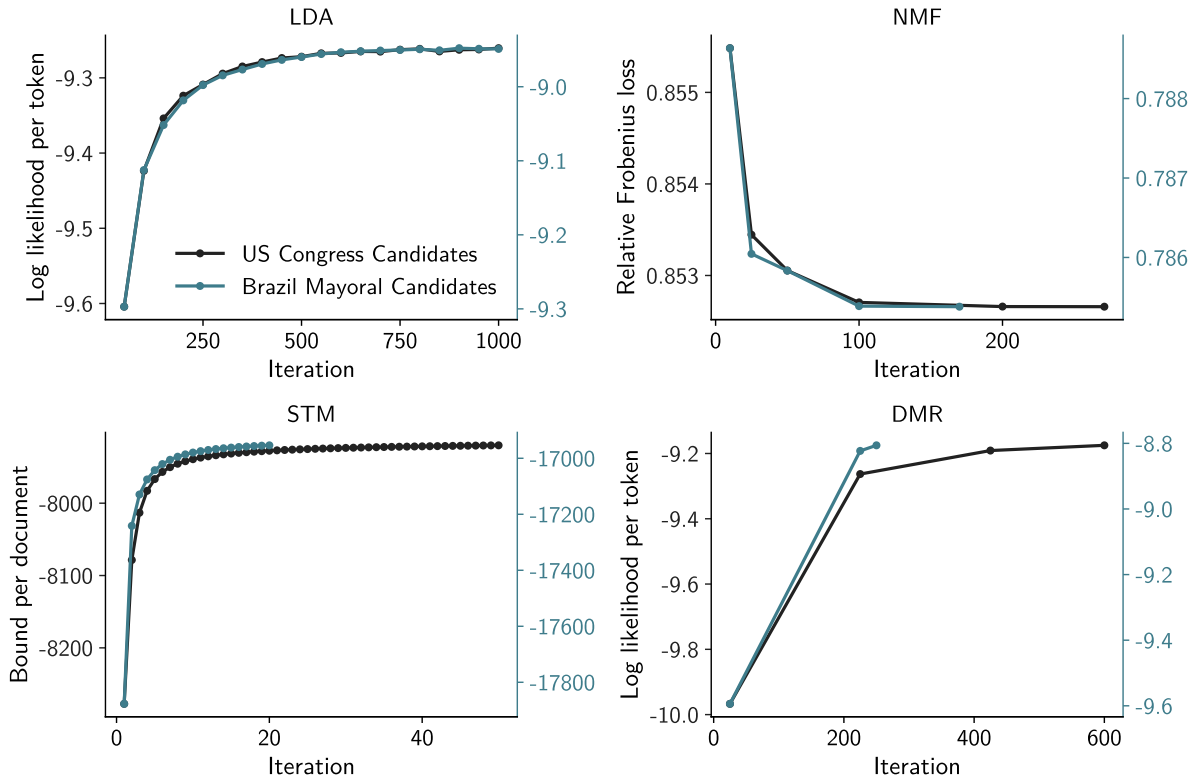
Table D.6 reports the scores. The BERTopic topics are the most coherent and the most diverse in both corpora; coherence is 0.079 against 0.061 for the best alternative in the United States and 0.115 against 0.059 in Brazil. The quality advantage does not cost time, i.e., the BERTopic wall clock is of the same order as the fastest alternative in each corpus. Figure D.3 plots each alternative's training objective by iteration; the per-sweep gains at the last recorded iteration are small fractions of the early gains in every fit.

Table D.6: Topic Quality and Runtime Against Alternative Topic Models

	US Congress Candidates				Brazil Mayoral Candidates				
Model	Coherence	Diversity	Time (min.)	Iter.	Coherence	Diversity	Time (min.)	Iter.	Hardware
BERTopic	**0.079**	**0.77**	5	–	**0.115**	**0.76**	69	–	T4/L4 GPU
LDA	0.061	0.49	6	1000	0.059	0.57	73	1000	6-core CPU
NMF	0.014	0.67	7	271	0.056	0.58	29	170	4-core CPU
STM	0.014	0.22	293	50	0.056	0.58	573	20	1-core CPU
DMR	0.036	0.49	25	600	0.048	0.64	38	250	6-core CPU

Coherence is the mean normalized pointwise mutual information over each topic's top ten terms, and diversity the share of unique terms across all topics' top ten (Grootendorst, 2022). All models are scored on the same reference corpus (Appendix D.5.1) at the corpus's final topic count. BERTopic is the topic model of Section 3.2. Time is one fit's wall clock in minutes on the listed hardware, and Iter. the number of training sweeps.

Figure D.3: Convergence of the Alternative Topic Models



Each panel shows one model's training objective on its native scale, i.e., log likelihood per token for LDA and DMR, relative Frobenius reconstruction loss for NMF, and the variational bound per document for STM. Points mark the recorded iterations. The United States follows the left ruler and Brazil the right, each tinted to match its line; the iteration counts match Table D.6.

E Alternative Topic Resolutions

The estimates of Section 5 rest on one choice the model does not make for itself, namely how finely the topic columns are cut. This group of appendices reports what changes

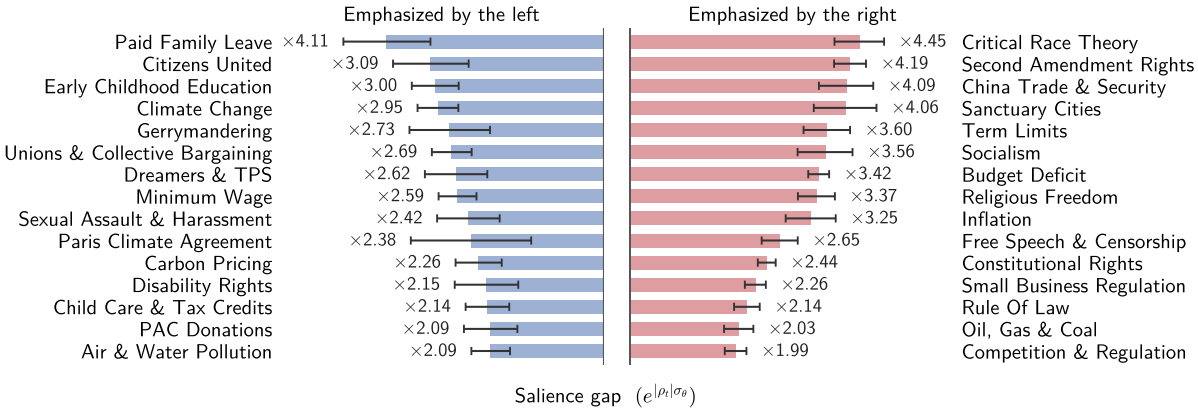
when that choice changes. I first re-estimate at the resolution the topic model itself produces, then compare the resulting scales document by document, and finally replace both the topics and the stance with a language model's own coding of the same sentences. The estimator, the documents, and the identification are the same throughout, so any difference comes from the columns. A sentence of five words or fewer carries no usable topic or stance signal on its own (a bare section-header label, a stray punctuation fragment the segmenter produced, a leaked table cell), so every such sentence joins the format and register class before estimating at every resolution reported here, whatever topic it was otherwise assigned; this reaches 7.9% of the surviving US sentences and 14.0% of the Brazilian ones.

E.1 Stance and Salience Gaps at the Subtopic Resolution

Figures E.1 to E.4 repeat the two gap figures of Section 5 over the subtopics the topic model discovers, with the format and register topics removed before estimating rather than grouped and dropped. There are too many subtopics to plot whole, so each panel keeps the fifteen strongest, and topics appearing in fewer than 50 documents are excluded.

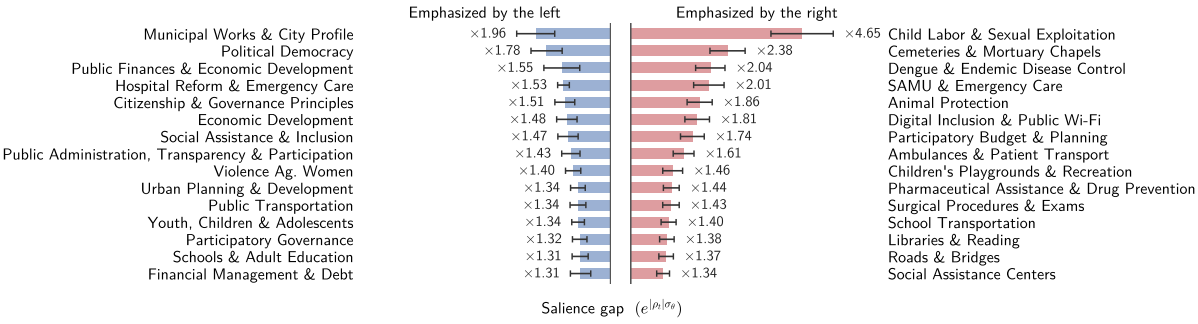
The subtopic resolution names the same divisions more narrowly. In the United States the topics the right emphasizes most are critical race theory, the Second Amendment, sanctuary cities, and trade with China, which the supertopic estimates collect into values and worldview, public safety, immigration, and international affairs; the left's are paid family leave, climate change, campaign finance, and early childhood education, collected into labor, environment, elections, and education. In Brazil the right's strongest emphases are child labor and sexual exploitation, cemeteries, and endemic disease control, and the left's are political democracy, citizenship and governance principles, and social assistance. The gain from the subtopic resolution is specificity, and the cost is that a topic carrying a few thousand sentences is estimated far less precisely than a class carrying a tenth of the corpus, which the wider confidence intervals in these figures show directly.

Figure E.1: Subtopics Each Side Emphasizes, United States



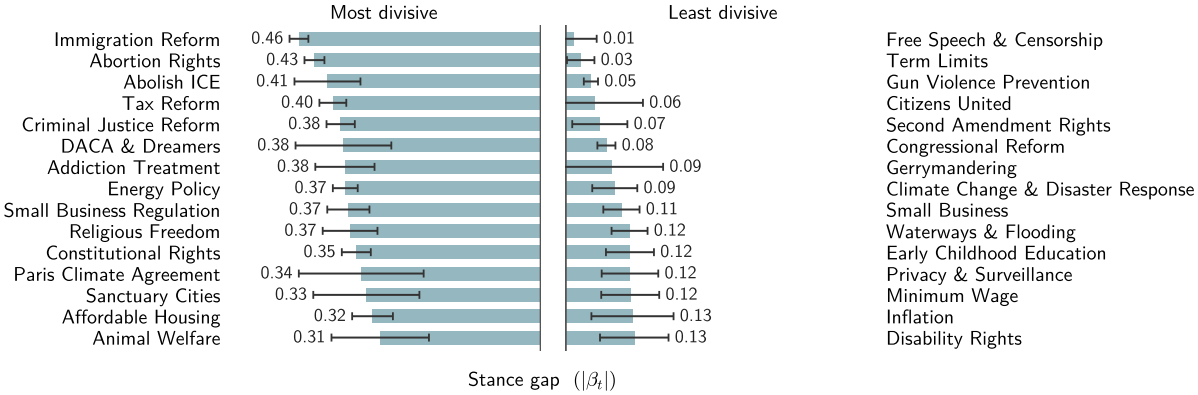
As Figure 4, estimated over the subtopics rather than the supertopics. Each bar is a topic's salience gap $e^{|\rho_t|\sigma_\theta}$ with its 95% Godambe confidence interval, and the vertical line marks parity. Each panel shows the fifteen topics most emphasized by one side. Format and register topics are removed before estimating, and topics in fewer than 50 documents are excluded.

Figure E.2: Subtopics Each Side Emphasizes, Brazil



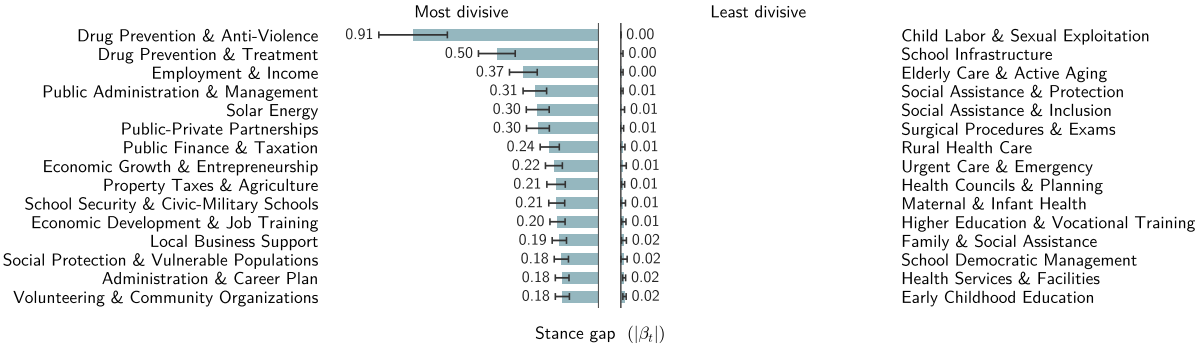
As Figure 6, estimated over the subtopics rather than the supertopics. Each bar is a topic's salience gap $e^{|\rho_t|\sigma_\theta}$ with its 95% Godambe confidence interval, and the vertical line marks parity. Each panel shows the fifteen topics most emphasized by one side. Format and register topics are removed before estimating, and topics in fewer than 50 documents are excluded.

Figure E.3: Largest and Smallest Stance Gaps at the Subtopic Resolution, United States



As Figure 5, estimated over the subtopics rather than the supertopics. The left panel shows the fifteen largest stance gaps and the right panel the fifteen smallest, on one shared axis. Each bar is a topic's stance gap $|\beta_t|$ with its 95% Godambe confidence interval, and the vertical line marks a gap of zero.

Figure E.4: Largest and Smallest Stance Gaps at the Subtopic Resolution, Brazil



As Figure 7, estimated over the subtopics rather than the supertopics. The left panel shows the fifteen largest stance gaps and the right panel the fifteen smallest, on one shared axis. Each bar is a topic's stance gap $|\beta_t|$ with its 95% Godambe confidence interval, and the vertical line marks a gap of zero.

E.2 Agreement Between the Main Estimate and the Alternatives

Figure E.5 plots the estimated ideal points of Section 5 against the subtopic estimate of Appendix E.1, document by document and colored by party group. Both estimates cover the whole estimation sample and rest on the same documents, the same sentences, and the same stance predictions, so the only difference between them is how finely the columns are cut.

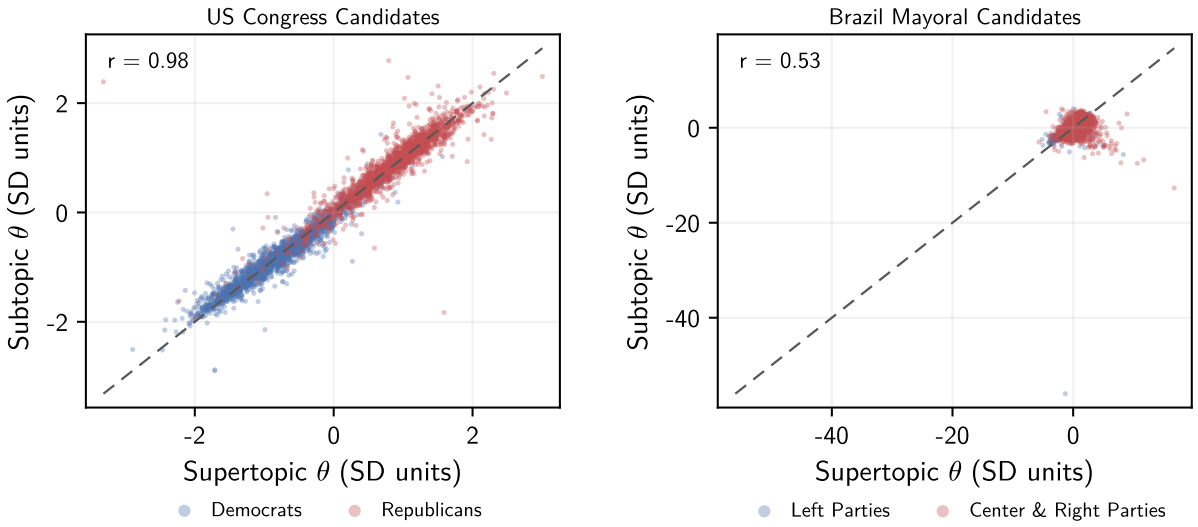
Grouping the subtopics leaves the US scale in place and moves the Brazilian one. The two estimates correlate at 0.98 in the United States and 0.53 in Brazil, and in both corpora the disagreement is spread across documents rather than concentrated in one party group. What the grouping buys is precision and legibility: a supertopic pools the sentences of several subtopics, so its loadings carry tighter intervals, and 25 classes fit on a page where 96 do not.

Figure E.6 then compares the main estimate with every scale the language model can produce: its topics and stance through the joint model, and its sentence scores averaged per document, each under both stance prompts. The two corpora behave differently. In the United States every pair correlates between 0.87 and 1.00, so nothing in this design choice matters there. In Brazil the main estimate correlates between 0.27 and 0.46 with the language-model scales, the two checkpoints correlate between 0.32 and 0.34 with each

other, and the same checkpoint under the two prompts correlates at 0.62 for GPT-5.4 mini and 0.82 for GPT-4o mini.

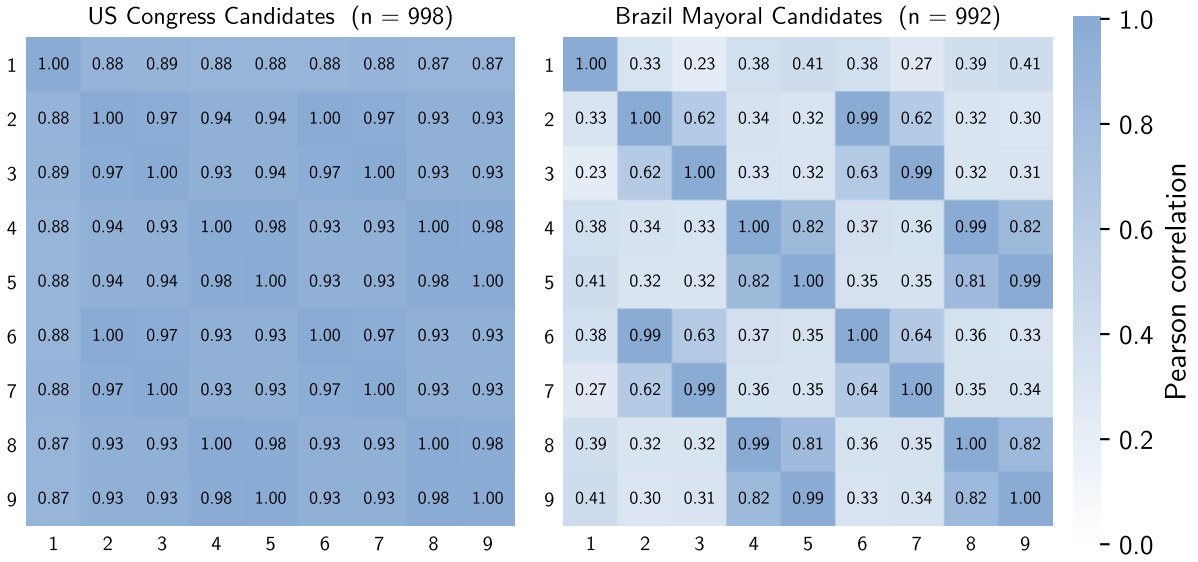
One further reading of that matrix bears on the model rather than on the language model. Within a checkpoint and a prompt, the joint fit and the plain average of the sentence scores correlate at 0.99 in Brazil. The topic columns therefore carry no information there, and the scale is the stance average alone, which is not the case in the United States or in either corpus for the unsupervised leg.

Figure E.5: Supertopic and Subtopic Estimates Compared



Each dot is one platform, its supertopic ideal point on the horizontal axis against its subtopic ideal point on the vertical, both in standard-deviation units. Left panel: US House platforms, Democrats in blue and Republicans in red. Right panel: Brazilian mayoral platforms, leftist parties in blue and centrist and rightist parties in red. The dashed line is parity and r is the Pearson correlation.

Figure E.6: Correlations Among the Estimated Scales



1. Joint IRT, supertopics
2. Joint IRT, GPT-5.4 mini
3. Joint IRT, GPT-5.4 mini, no cues
4. Joint IRT, GPT-4o mini
5. Joint IRT, GPT-4o mini, no cues
6. Ask-and-average, GPT-5.4 mini
7. Ask-and-average, GPT-5.4 mini, no cues
8. Ask-and-average, GPT-4o mini
9. Ask-and-average, GPT-4o mini, no cues

Weighted Pearson correlations among nine estimated scales, on the documents the language-model sample covers. Scale 1 is the estimate of Section 5. Scales 2 to 5 apply the same joint model to each language model's own topics and stance, and scales 6 to 9 average that model's sentence scores per document, with no topic model and no item-response model. "No cues" marks the stance run whose prompt instructs the model to ignore the names of parties, politicians, and places. Both axes carry the numbering of the key.

E.3 Topics and Stance from a Language Model

The final alternative replaces both inputs. Rather than discovering topics from the embeddings and scoring stance with the transfer-learned classifier, I ask a language model to place each sentence in one class of the same codebook and, separately, to score it from left to right. The joint model then runs unchanged on the result, so the comparison is one of inputs and not of estimators.

Cost bounds this exercise, so both stages run on the stratified sample of 1,000 documents per corpus that Section 6 describes, and every quantity carries the design weight $w_i = N_h/n_h$ of its party stratum. The stance scoring uses the prompt that instructs the model to ignore the names of parties, politicians, and places, so that proper nouns cannot carry the left–right signal on their own. I drop the format and register class before estimating, as in the main estimates, and keep the residual class. A document left with fewer than ten sentences cannot support a document-specific parameter and is dropped,

which removes 2 US and 5 Brazilian platforms.

I report GPT-5.4 mini. Both models are within a few hundredths of each other against the external benchmarks, and the choice rests on how they use the codebook: GPT-5.4 mini applies the format and register class to 8.6% of the Brazilian sentences and 1.2% of the US ones, close to the shares the topic model itself assigns, whereas the alternative model applies it to almost none and places that material in the residual class and in the policy classes instead.

The two corpora then behave differently. In the United States the language model's coding reproduces the substantive ranking: immigration and values and worldview carry the largest stance gaps under both sets of inputs, and civil liberties, institutions, and macroeconomics sit among the classes the right emphasizes. In Brazil the classes at the top of its rankings are the ones a mayor has no competence over and that the coding almost never uses. International affairs carries 0.04% of the Brazilian sentences and immigration 0.01%, and the model nonetheless places international affairs first among the classes the right emphasizes and immigration first among those the left emphasizes.

In Brazil, therefore, the language model's coding does not describe municipal politics. The classes it puts at the top of both rankings lie outside municipal competence and carry almost none of the corpus, and the two checkpoints place them on opposite sides: international affairs is the class GPT-5.4 mini reads as most emphasized by the right and GPT-4o mini as second-most emphasized by the left. The resulting estimates track the expert benchmark at the party level while disagreeing with each other, and with the estimates of Section 5, document by document. A correlation against an external benchmark is therefore necessary but not sufficient evidence that a scale measures what it claims to, and the issues a model says divide a corpus are the check that separates the two.

Appendix References

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Bolognesi, B., Ribeiro, E., and Codato, A. (2023). A New Ideological Classification of Brazilian Political Parties. *Dados*, 66:e20210164.

Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the Biennial GSCL Conference*, pages 31–40.

Case, C. R. (2025). Measuring Strategic Positioning in Congressional Elections. *Journal of Politics*. Forthcoming.

Denny, M. J. and Spirling, A. (2018). Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It. *Political Analysis*, 26(2):168–189.

Dieng, A. B., Ruiz, F. J. R., and Blei, D. M. (2020). Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

Godambe, V. P. (1960). An Optimum Property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.

Grimmer, J. and Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, A. (2020). spaCy: Industrial-strength natural language processing in Python. https://spacy.io.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic.

Lau, J. H., Newman, D., and Baldwin, T. (2014). Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 530–539.

Laurer, M., van Atteveldt, W., Casas, A., and Welbers, K. (2024). Less Annotating, More Classifying: Addressing the Data Scarcity Issue of Supervised Machine Learning with Deep Transfer Learning and BERT-NLI. *Political Analysis*, 32(1):84–100.

Le Mens, G. and Gallego, A. (2025). Positioning Political Texts with Large Language Models by Asking and Averaging. *Political Analysis*, 33:274–282.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lee, M. (2022). bab2min/tomotopy. https://doi.org/10.5281/zenodo.6868418. Version 0.13.0.

Lehmann, P., Franzmann, S., Al-Gaddooa, D., Burst, T., Ivanusch, C., Lewandowski, J., Regel, S., Riethmüller, F., and Zehnter, L. (2025). Manifesto Corpus. Version 2025-1. Berlin: WZB Berlin Social Science Center / Göttingen: Institute for Democracy Research (IfDem).

Lindsay, B. G. (1988). Composite Likelihood Methods. In Prabhu, N. U., editor, *Statistical Inference from Stochastic Processes*, volume 80 of *Contemporary Mathematics*, pages 221–239. American Mathematical Society.

Lowe, W. (2008). Understanding Wordscores. *Political Analysis*, 16(4):356–371.

Lowe, W., Benoit, K., Mikhaylov, S., and Laver, M. (2011). Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly*, 36(1):123–155.

McInnes, L., Healy, J., and Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint arXiv:1802.03426.

Meisels, M. (2026). Candidate Positions, Responsiveness, and Returns to Extremism. *The Journal of Politics*, 88(2):681–697.

Mikhaylov, S., Laver, M., and Benoit, K. (2012). Coder Reliability and Misclassification in the Human Coding of Party Manifestos. *Political Analysis*, 20(1):78–91.

Mimno, D. and McCallum, A. (2008). Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, pages 411–418.

Morris, C. N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*, 78(381):47–55.

OpenAI (2024). Structured Outputs. https://platform.openai.com/docs/guides/structured-outputs. OpenAI API documentation.

OpenAI (2025). GPT-4.1. https://platform.openai.com/docs/models/gpt-4.1. Large language model, API snapshot `gpt-4.1`.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong.

Rheault, L. and Cochrane, C. (2020). Word embeddings for the analysis of ideological dimensions in political texts. *Political Analysis*, 28(1):112–133.

Roberts, M. E., Stewart, B. M., and Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2):1–40.

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

Slapin, J. B. and Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722.

Speer, R. (2019). ftfy: Fixes text for you (version 6). https://github.com/rspeer/python-ftfy.

Tribunal Superior Eleitoral (2020). Candidatos: Eleições municipais 2020 (candidate microdata). Repositório de Dados Eleitorais, https://dadosabertos.tse.jus.br/.

Varin, C., Reid, N., and Firth, D. (2011). An Overview of Composite Likelihood Methods. *Statistica Sinica*, 21(1):5–42.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika*, 61(3):439–447.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25.