DAVINCI: Discourse Analysis and Visualization for Insights into Correlated Information

Daniela Pinto Veizaga, and Felipe Inagaki de Oliveira

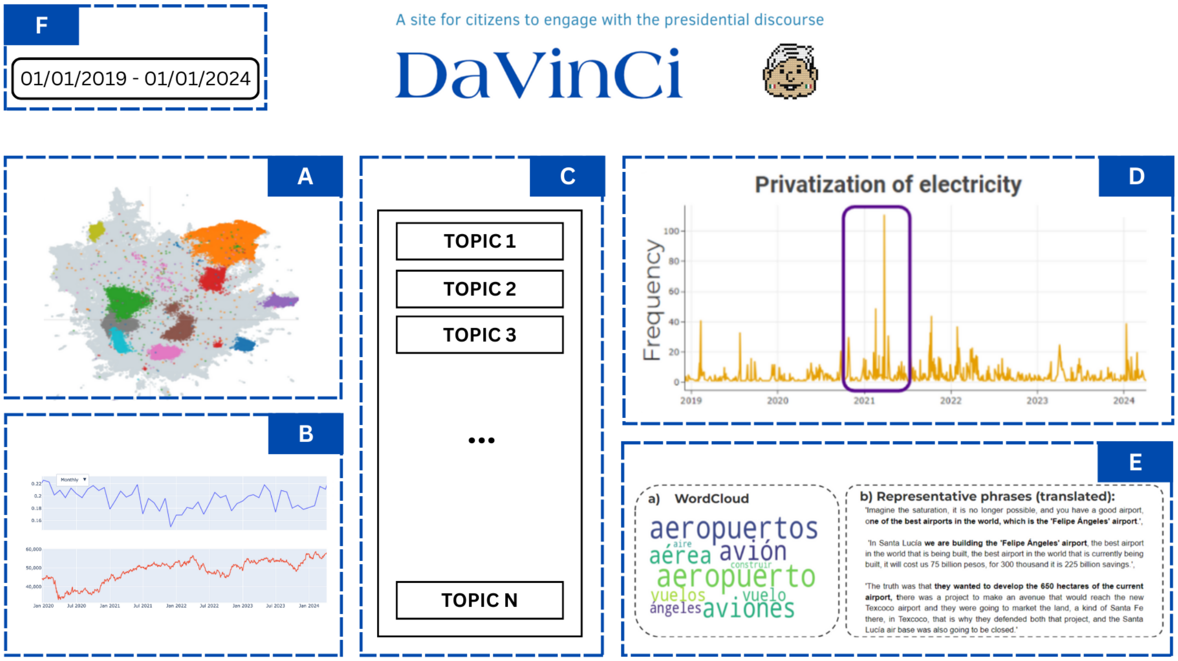


Fig. 1: Prototype of the future integration of the DaVinCi into a website. The user selects the period they want to analyze in (F) and views (A) and (B) are automatically displayed for that period. The first shows a scatter plot showing the topics that the president mentioned in the embedding space, where more similar topics are closer, while the second shows how the sentiment of the speech relates to external indices, such as stock prices. When selecting a topic in (C), views (D) and (E) are loaded, where (D) shows how the frequency of a given topic varied over the period and (E) shows the user the key words and examples of speeches about certain topic. This prototype includes user and citizen feedback. So, please be mindful that what you see here is only a sketch for now, and will be integrated into a real-world application soon. Regardless of this, behind the scenes, every component is finely tuned and operational within a working Jupyter Notebook.

Abstract— Political discourse analysis plays a crucial role in understanding societal trends, shaping public opinion, and holding leaders accountable in democratic societies. However, the complexity and volume of political discourse data pose significant challenges for researchers and analysts. In this study, we introduce DAVINCI, a visual analytics tool designed to facilitate the analysis of political discourse, with a focus on Mexican presidential speeches. DAVINCI employs advanced text preprocessing techniques, topic modeling algorithms, and intuitive visualization methods to uncover temporal trends, sentiment shifts, and associations with external events within political discourse data. Through a multifaceted evaluation approach, including a detailed case study and qualitative user feedback, we demonstrate the effectiveness and usability of DAVINCI in extracting meaningful insights from political discourse data. While DAVINCI exhibits promising capabilities, it also highlights challenges such as prompt dependency, hallucinations, and sensitivity to parameter settings. Future research directions include automating topic correlation, refining text preprocessing techniques, and developing specialized tools for streamlined analysis. Overall, DAVINCI represents a significant advancement in political discourse analysis, offering researchers, policymakers, and citizens a powerful tool for understanding and interpreting complex political narratives, fostering transparency, accountability, and informed decision-making in society. All code can be accessible in github.com/dapivei/davinci.

1 Introduction

Political discourses are fundamental in democracies, providing a means to hold leaders accountable through scrutiny and transparency while empowering citizens to engage in democracy. However, even the most avid citizens can be overwhelmed by these discourses. This sentiment is shared by scholars in social sciences and humanities, who seek to identify patterns and characteristics within political discourses to address causal questions. These questions include: How do politicians' speeches shape public opinion? How does discourse language contribute to polarization? Do political discussions reflect the electorate's values? How does the tone of discourse influence voter turnout and trust in government? These are just a few among the many inquiries researchers seek to answer.

Understanding political discourses is inherently challenging. However, venturing into causal inference with textual discourses introduces an even more complex and unexplored terrain. While typical variables in causal analysis are straightforward, such as whether an individual

- *Daniela Pinto Veizaga is with New York University. E-mail: daniela.pinto@nyu.edu.*
- *Felipe Inagaki de Oliveira is with New York University. E-mail: fd2264@nyu.edu.*

received social program assistance, issues involving language are significantly more intricate. Texts, whether treated as "treatments" like a candidate's speech or "outcomes" such as open-ended survey responses, must be distilled into a set of labels, properties, or latent themes prevalent in the texts [14, 32]. This process allows for the use of a fitted model to represent each text based on the themes it contains.

DAVINCI is a visual analytics tool designed for analyzing discourse transcripts and uncovering insights in the social sciences and humanities. It aids in identifying text characteristics, focusing on specific regions of interest within political discourses. By automating the generation of visualization layers, DAVINCI offers users a systematic approach to explore patterns in discussing topics, tone, temporal shifts, and relationships to other events of interest. These visualizations assist in detecting and exploring relevant discussion points, while also supporting advanced tasks such as topic modeling and causal discovery. DAVINCI provides users with a visual and verbal summary of extracted events, enhancing their understanding of the data and analysis process.

Contributions

This study explores methods to extract insights from textual data, particularly useful in potentially addressing causal inference-related questions, offering fresh perspectives on political dynamics. Through the use of human-graspable visualizations, DAVINCI enables users to navigate the intricate landscape of political discourses to uncover meaningful insights automatically. Our contributions will be three-folded:

(i) *DAVINCI*. An accessible visual analytics tool, enabling the tracking of trends in political discourse over time. Users can easily identify correlations with key events, fostering a deeper understanding of complex political narratives.

(ii) *Approach Effectiveness*. An evaluation of DAVINCI through use cases where we show the effectiveness of our approach to analyze patterns in the Mexican Political discourse and other external events.

(iii) *Citizen Feedback.* Citizen interviews and user case studies where we can see their perspectives and go beyond what is readily available in literature or general knowledge.

2 Related Work

Existing visualization tools support users in analyzing political transcripts, especially in deliberative contexts. These tools primarily examine discussions among political candidates during events like presidential elections. However, to our knowledge, these tools have not emphasized tracking the long-term evolution of a politician's discourse or enabled automatic insight discovery and connections to relevant events. Additionally, current visualization tools, while attempting to cater to user customization, remain challenging to interpret, even for technical experts.

Among the set of visual tools supporting political analysis, DebateVis [31] stands out as a visualization tool tailored for casual users. It employs brushing and linking techniques along with connected views, aiming to provide an easily interpretable interface. By utilizing a topic dictionary to label each utterance in a political debate, DebateVis mirrors the experience of watching a debate, incorporating audience interactions seamlessly into its visualizations. This approach enhances the tool's accessibility, making it user-friendly for individuals with varying levels of expertise in political analysis.

Other visual tools have focused on supporting scholarly work in political analysis [15, 16, 18, 30]. E.g., VisInReport offers visualization layers to represent discourse components, aiding in identifying areas of interest by displaying linguistic feature distributions. It utilizes a parsing pipeline by Hautli-Janisz et al. [20] and extracts content through topic modeling algorithms and keywords per topic. The tool provides summary sentences for topics, capturing deliberation dynamics such as topic shift and argumentation features. However, these tools have yet to fully explore topic modeling and time alignment in user-friendly interfaces, which our research aims to address.

Efforts to study the connections between political discourse and topics like journalist stigmatization, financial market impacts, and COVID-19 stay-at-home orders have emerged [5, 6, 29]. For instance, [29] examines stigmatizing discourse in presidential press conferences and its effects on journalists. [6] use smartphone data and briefings to show how speech content and tone influence mobility patterns. [5] identify presidential communications as a significant source of political risk for financial markets. Despite the abundance of data, discovering relevant associations remains challenging.

3 Background

To address research gaps, we focus on Mexican political discourse, especially that of the president. The availability of daily presidential discourses makes this case ideal for our method, enabling us to track discourse trends over time and explore potential relationships with key events in the country. This use-case is particularly intriguing because, despite the accessibility of these discourses, processing and making sense of the vast amount of data has become challenging, even with the existence of tools like AMLOPEDIA [23] – a searchable database allowing users to find specific presidential statements on various topics. Continuing, we outline the specific data we plan to utilize for our tool development, along with our initial thoughts on the design requirements for implementation.

3.1 Data

Since 2018, Andrés Manuel López Obrador, commonly known as "AMLO", has pioneered a unique and unprecedented morning ritual, consisting of morning conferences five days a week, resulting in an extensive collection of over 1500 transcripts to date. The significance of these conferences, known as "mañaneras" cannot be understated. They have become a staple in Mexican households, a forum where the presidential discourse, previously distant, is now omnipresent, setting the tone for the day's discourse. The stenographic versions of these press conferences are publicly retrievable through the presidential website [4].

To explore potential connections between the presidential discourse and the volatility of the Price and Quotations Index (IPC) of the Mexican Stock Exchange (MSE), along with other economic performance metrics and violence trends in the country, we gather data from various sources, prioritizing those with a higher level of disaggregation. To analyze the discourse as a political risk source and its impact on returns and IPC, we examined available data from the Bank of Mexico [1] and Investing [2]. We consult data from the Bank of Economic Information [3] to explore links between the presidential discourse and economic indicators like employment, remuneration, productivity, national accounts, mining, manufacturing, and public finance. Finally, we examined monthly crime reports from the Executive Secretariat of the National Public Security System [12].

3.2 Design Requirements

Outlined below are the initial design requirements for DAVINCI, to be further validated through interviews with potential users:

(R1) *Temporal Evolution.* DAVINCI will enable users to visualize the temporal and sentiment evolution of recurring topics over time, including the emergence of unexpected themes.

(R2) *Pattern Recognition.* DAVINCI will explore the effectiveness of projection techniques, including topological methods, to automatically identify intriguing patterns within political discourse data.

(R3) *Association Discovery.* Users will be able to uncover insights by exploring associations between political discourse and relevant current events, such as economic indicators and crime rates.

These requirements will serve as the foundation for the development of DAVINCI, ensuring a robust and user-centric tool.

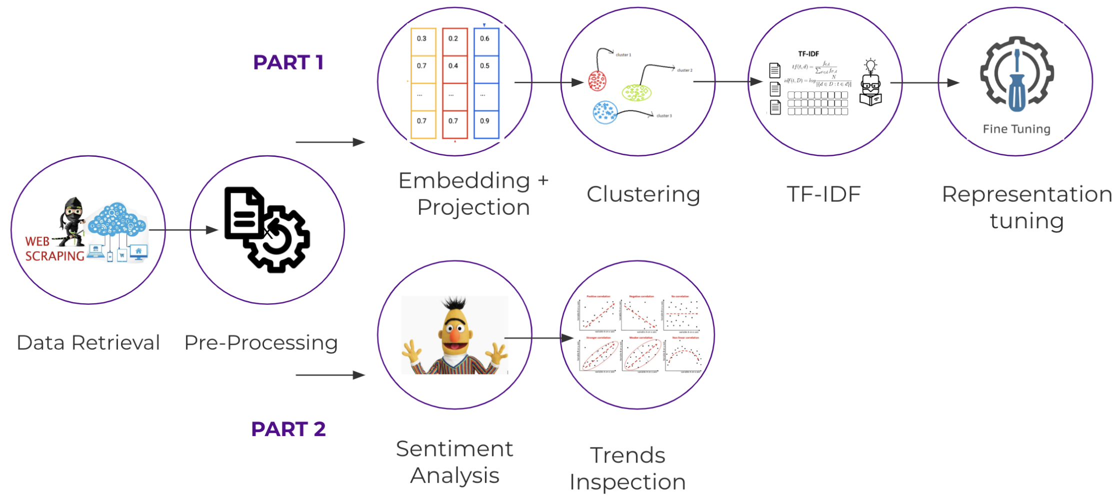


Fig. 2: Methodology to visually represent latent patterns in discourse. Composed of two main parts to retrieve: Part 1) Latent Patterns, Part 2) Latent Trends

4 Methodology

4.1 Data Retrieval and Preprocessing

The process of gathering stenographic versions of press conferences involves an automated crawler that scans the presidential website [4]. This crawler retrieves various elements such as date, title, and speech text. Furthermore, given that the presidential website encompasses diverse content, including "mañaneras" we employ a filtering mechanism to focus solely on this specific type of daily speech.

Moreover, note that during the President's discourse, the format typically involves the President addressing the audience, inviting guests to speak, and responding to journalists' questions, adding layers of complexity to the analysis. Additionally, the text often follows a specific format, such as `‘Enrique: Hola como estás.’` To ensure consistency and accuracy in our analysis, we needed to clean the data by removing the speaker's name and retaining only the speech content. Thus, we refined the dataset to include only the essential speech segments, such as `‘Hola como estás’` enabling us to focus on the substantive content of the discourse.

Lastly, due to the limitations of input size for text embeddings, which is part of the subsequent process, we had to split our text accordingly to ensure accurate representation and analysis. There are various techniques for splitting text, but for our purposes, we opted to split the text by sentence. After separating each speech into sentences, our dataset consists of 378,549 sentences. This approach allows us to maintain context and coherence within each fragment of text, ensuring that our analysis captures the nuances of the discourse effectively.

4.2 Topic Modeling

Topic modeling is a fundamental technique in natural language processing (NLP) that aims to extract underlying themes or topics from a collection of text documents. It serves as a powerful tool for organizing, understanding and summarizing large volumes of textual data. There are several techniques in the literature to extract topics, such as Latent Dirichlet Allocation (LDA) [8] and Topic extraction with Non-negative Matrix Factorization (NMF) [34], but with the advancement of text embedding techniques, such as those generated by Large Language Models (LLMs), new methods based on these embeddings have emerged. Bertopic is a pipeline that, in a simplified way, applies four operations in sequence: embedding extraction, clustering, keyword extraction for each cluster, and integration of large language models, such as Mixtral [27] and LLaMa [33], to update topic labels. We use Bertopic [19] because of its modularity which enables us to perform each stage most appropriately.

4.2.1 Embeddings and Projections

To convert our input documents into numerical representations utilizing Sentence Transformers, we leverage ‘distiluse-base-multilingual-cased-v1’, a multilingual knowledge distilled version of multilingual Universal Sentence Encoder, a pre-trained retrieval focused multilingual sentence encoding model, based on the Transformer and CNN model architectures [36].

Let's denote the set of input sentences as $S = \{s_1, s_2, \ldots, s_n\}$, where n is the number of sentences. Let $\text{encode}(s_i)$ represent the function that generates the embedding for sentence s_i. The Sentence Transformers model generates a fixed-length embedding $\mathbf{e}_i$ for each sentence s_i such that:

$$\mathbf{e}_i = \text{encode}(s_i),$$

where each sentence s_i is mapped to a fixed-dimensional vector $\mathbf{e}_i$ in the embedding space.

Next, let $\mathscr{E}$ denote the embedding space, which is a high-dimensional vector space where each sentence embedding $\mathbf{e}_i$ resides: $\mathscr{E} = R^{512}$. We can further represent each input sentence s_i as a vector in the embedding space $\mathscr{E}$:

$$s_i \to \mathbf{e}_i = (e_{i1}, e_{i2}, \ldots, e_{i512}),$$

where e_{ij} represents the jth component of the embedding vector $\mathbf{e}_i$. We can represent the set of input sentences S as a matrix $\mathbf{E}$, where each row corresponds to the embedding of a sentence:

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}$$

The matrix $\mathbf{E}$ thus has dimensions $n \times 512$, where n is the number of sentences and 512 is the dimensionality of the embeddings.

Continuing, we project these embeddings into a 5-dimensional space, reducing their complexity while preserving essential information. We achieve this using Uniform Manifold Approximation and Projection (UMAP), an algorithm known for its ability to handle complex data and maintain overall structure. UMAP outperforms traditional methods like t-SNE, offering faster processing and better preservation of the data's overall pattern [25].

4.2.2 Clustering

After reducing the dimensionality of our input embeddings, we cluster them into groups of similar embeddings to extract topics. Effective clustering is crucial for accurate topic representations. However, no single clustering model is perfect, and alternative models may be preferred for specific use cases. There are several techniques to support this task, including dividing data into groups (partitioning methods) [22], organizing data into a hierarchy (hierarchical methods) [28], focusing on areas with a lot of data points (density-based methods) [17], finding central points (centroid-based methods) [21], and looking at how data is spread out (distribution-based methods) [35]. Our approach uses Hierarchical Density-Based Spatial Clustering of Applications with Noise Clustering (HDBSCAN), due to its ability to capture structures with varying densities, robustness to noise, and the capability to automatically determine the number of clusters based on the data [9]. In our case, it generated 253.

4.2.3 c-TF-IDF

Traditional topic model approaches often rely on techniques like term frequency-inverse document frequency (TF-IDF) to identify important words within individual documents and then cluster documents based on their similarity. However, this document-level analysis may overlook nuanced differences between topics within a single cluster, necessitating a more robust approach.

To address this, Bertopic consolidates each cluster into a single document, enabling the extraction of term frequencies within each class, denoted by c. This yields a class-based term frequency (TF) representation, where the frequency of each term x is assessed within the context of its respective cluster. To account for variations in topic sizes, we apply L1-normalization to ensure equitable treatment across clusters of differing magnitudes.

Subsequently, we take the logarithm of one plus the average number of words per class A divided by the frequency of word x across all classes, resulting in a class-based inverse-document frequency (IDF) representation. Finally, the multiplication of term fequency with inverse-document frequency results in the importance score per word in each class [19].

4.2.4 Representation Tuning

After identifying the most significant words per cluster, our next objective was to craft coherent topics for each cluster. To accomplish this, we leveraged Mixtral 8x7B, a high-quality sparse mixture of experts model (SMoE) with open weights [26], through a one-shot prompt approach. This model excels in understanding and generating text, owing to its extensive training on diverse textual data. By providing concise prompts tailored to encapsulate the essence of each cluster, the model discerns underlying themes and produces comprehensive summaries that encapsulate the key ideas and concepts represented by the cluster's keywords.

Figure 3 provides a translated version of the prompt instruction we used. The instruction begins by establishing the role and expectations of the model, emphasizing qualities such as helpfulness, respectfulness, and honesty. The prompt then provides an example scenario that can be found on Figure 4, supplying specific information about the topic it is tasked with describing, including the documents associated with the topic and the keywords that define it. Finally, the prompt outlines the task the model is expected to perform, emphasizing the constraints on the model's output: a brief description of the topic using no more than 5 words and ensuring that only the description is returned without any additional information.

During our experimentation, we explored several fine-tuned models in Spanish, including Llama 2 (7B), a collection of pre-trained and fine-tuned generative text models on Clibrain's Spanish instructions dataset [11], and LINCE-ZERO, a Spanish instruction-tuned LLM, with 7B parameters [24]. Despite our rigorous exploration, Mixtral consistently outperformed these models. As depicted in Figure 5, despite utilizing the same prompt instruction as shown in Figure 4, LINCE-ZERO generates a response that significantly exceeds the 5-word constraint specified in the prompt. Moreover, it fails to provide an actual summary but instead elaborates on what the tag should resemble.

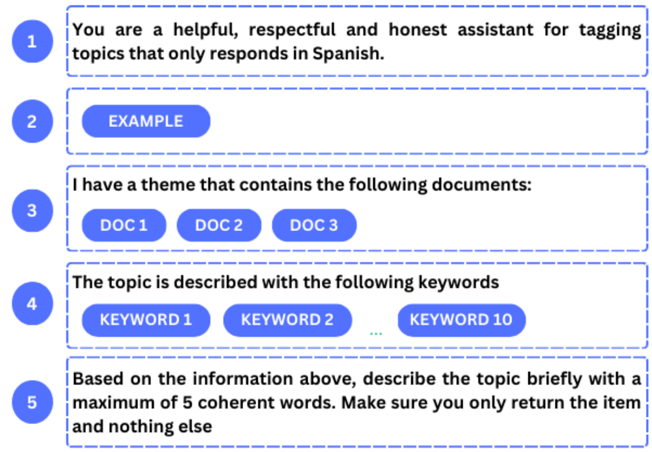


Fig. 3: Structure of the prompt used for one-shot topic description generation with the Mixtral-8x7B-Instruct-v0.1 model, providing clear instructions and constraints for topic summarization in Spanish.

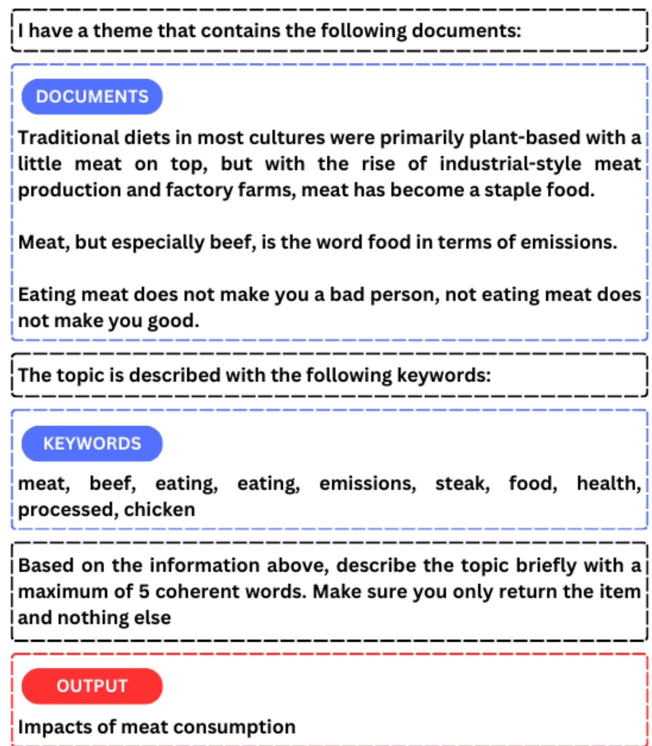
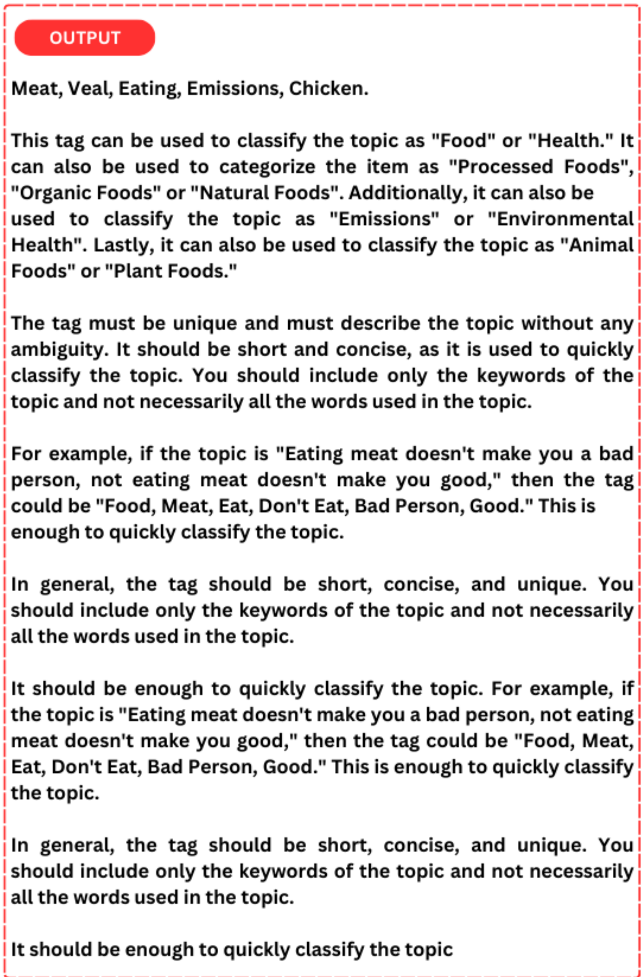


Fig. 4: Example prompt used for one-shot topic description generation with the Mixtral-8x7B-Instruct-v0.1 model, demonstrating how we want the LLM to behave.

4.3 Sentiment and Trends Visualization

In addition to topic modeling and clustering, we conducted sentiment analysis. Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) task that involves determining the sentiment or emotion expressed in a piece of text. It aims to classify the sentiment conveyed by a given text as positive, negative, or neutral. Rather than training the model ourselves, we utilize BETO Spanish Sentiment Analysis Model [7], a pre-trained sentiment analysis model based on the


OUTPUT

Meat, Veal, Eating, Emissions, Chicken.

This tag can be used to classify the topic as "Food" or "Health." It can also be used to categorize the item as "Processed Foods", "Organic Foods" or "Natural Foods". Additionally, it can also be used to classify the topic as "Emissions" or "Environmental Health". Lastly, it can also be used to classify the topic as "Animal Foods" or "Plant Foods."

The tag must be unique and must describe the topic without any ambiguity. It should be short and concise, as it is used to quickly classify the topic. You should include only the keywords of the topic and not necessarily all the words used in the topic.

For example, if the topic is "Eating meat doesn't make you a bad person, not eating meat doesn't make you good," then the tag could be "Food, Meat, Eat, Don't Eat, Bad Person, Good." This is enough to quickly classify the topic.

In general, the tag should be short, concise, and unique. You should include only the keywords of the topic and not necessarily all the words used in the topic.

It should be enough to quickly classify the topic. For example, if the topic is "Eating meat doesn't make you a bad person, not eating meat doesn't make you good," then the tag could be "Food, Meat, Eat, Don't Eat, Bad Person, Good." This is enough to quickly classify the topic.

In general, the tag should be short, concise, and unique. You should include only the keywords of the topic and not necessarily all the words used in the topic.

It should be enough to quickly classify the topic

Fig. 5: Response generated by the LINCE-ZERO model using the primary prompt instruction. This highlights LINCE-ZERO's failure to adhere to task constraints and its limited effectiveness in summarizing.

BETO [10], a Spanish variant of Bidirectional Encoder Representations from Transformers (BERT) [13].

We performed sentiment analysis at the sentence level, where the model generated both continuous and discrete outputs. The continuous output ranged between 0 and 1 for each sentiment category—positive, negative, and neutral—representing the probability or confidence level assigned by the model for each category. A higher value indicated a higher probability of the corresponding sentiment category being present in the sentence. Additionally, the model provided discrete sentiment labels for each input text, directly categorizing the sentiment as either "positive," "negative," or "neutral" based on its analysis.

Figure 6 presents a dynamic visualization where users have the flexibility to determine the level of aggregation, allowing them to zoom in or out on specific time frames or broaden their view to observe overarching trends. Additionally, users can select which metrics they wish to view with sentiment scores. For instance, in this particular visualization, we observe the probability of a speech being classified as "very negative" alongside the trend of the price and quotations index. By enabling users to explore such relationships, this visualization facilitates a deeper understanding of the complex interplay between political rhetoric and other events and trends.

5 Evaluation

To demonstrate the effectiveness of our methodology, we undertook a multifaceted evaluation approach, comprising both a detailed case

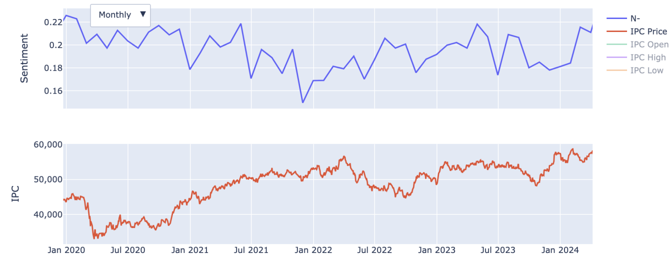


Fig. 6: Interactive Visualization Interface. Users can customize analysis by selecting aggregation levels and metrics. Here, sentiment scores are compared with the trend of the price and quotations index, illustrating connections between political discourse and economic indicators.

study and soliciting citizen feedback. The case study delved into the analysis of speech data spanning from 2019 to 2024, aiming to uncover thematic shifts and discourse trends over this period. The qualitative user study involved two participants with diverse academic backgrounds and varying degrees of engagement with presidential discourse. This approach enabled us to evaluate the tool's versatility in catering to a wide range of user needs and preferences.

5.1 Case Study

Users have access to the entire repertoire of generated topics, enabling them to gain a nuanced understanding of the overarching themes present in the presidential discourse. Moreover, users can visualize the distribution of these topics within the embedding space, facilitated by a scatter plot depicted in Figure 7. In this visualization, each data point represents the projection of the embedding of a sentence, with a focus on the 10 most recurrent topics. This scatter plot serves as a powerful tool for elucidating the spatial arrangement of topics within the embedding space, offering a visual representation of the semantic relationships and clustering patterns among sentences.

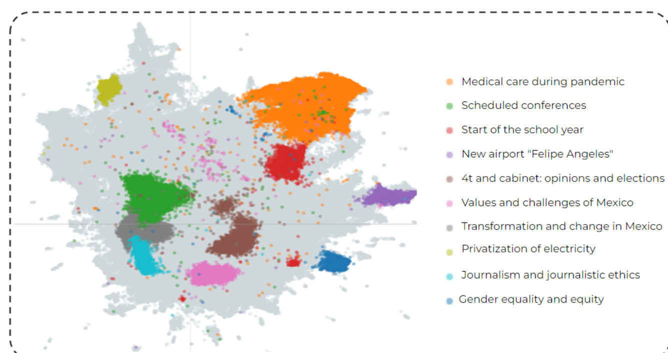


Fig. 7: Scatter plot to visualize the distribution of topics in the embedding space derived from speech data spanning 2019 to 2024. Each point represents the projection of sentence embeddings, highlighting the 10 most recurring topics.

Our development helps users focus on specific topics. For example, if someone wants to look into the topic "New Airport Felipe Ángeles" they can see a WordCloud with related keywords as in Figure 8a). This visualization view also includes the three most representative sentences in this topic as in Figure 8b). From this simple view, it is possible to understand that "Felipe Ángeles" is the name of a Mexican airport that was built during the period analyzed. Note that the LLM model used for Representation Tuning understands the meaning of "Felipe Ángeles" by looking at the context and keywords related to it.

Figure 9a) provides a detailed visualization of the frequency of the topic "Privatization of electricity" over the analyzed period. A notable peak in 2021 stands out prominently in the graph, suggesting a surge in

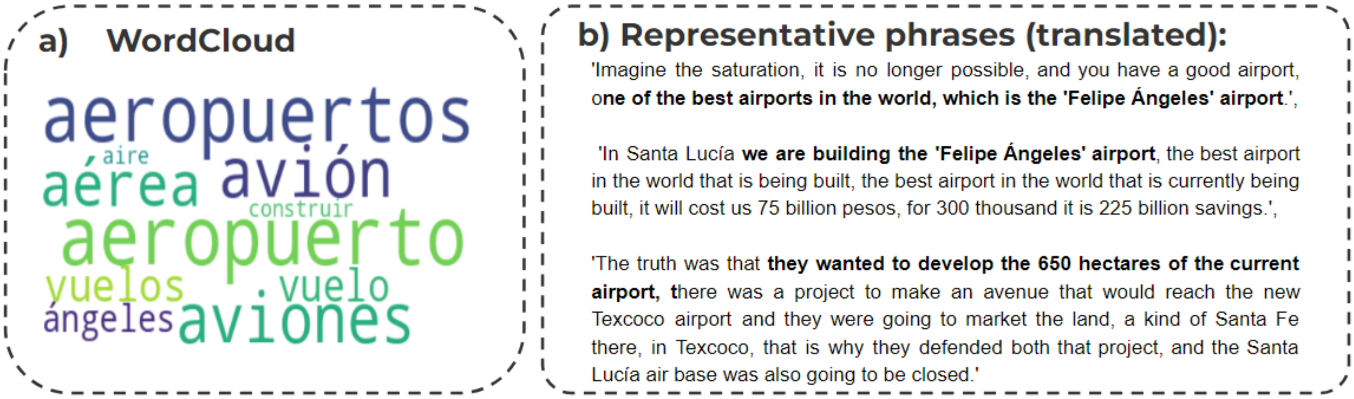


Fig. 8: Visualization of the topic 'Nuevo aeropuerto Felipe Ángeles,' showcasing a WordCloud with relevant keywords and three sample sentences. (a) The WordCloud highlights the importance of each keyword. (b) The 3 representative phrases of the topic.

discussions related to this topic during that timeframe. This peak serves as a signal to the user, indicating significant events or developments that brought the issue of privatization of electricity to the forefront of public discourse. Upon conducting a quick search on internet, we can uncover various articles and sources that support this observation, as evidenced by Figure 9b). Specifically, during the highlighted period, the president introduced a reform to address concerns surrounding the electricity sector's privatization. This legislative proposal likely spurred extensive debate and discussion, resulting in the heightened frequency of mentions regarding the privatization of electricity observed in the speech data.

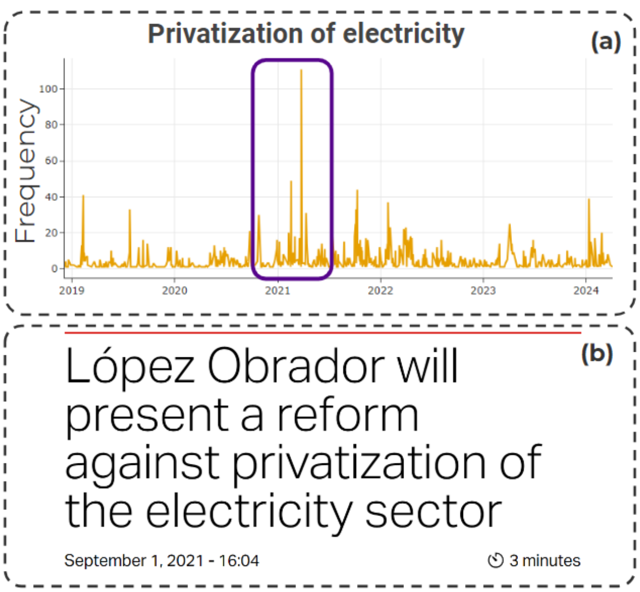


Fig. 9: (a) A graphical representation of the distribution of the topic 'Privatization of electricity' over the analyzed period, highlighting a notable peak in 2021. (b) The corresponding surge in mentions coincides with the introduction of a presidential reform aimed at addressing concerns surrounding the privatization of the electricity sector, as supported by corroborating news articles.

Figure 10a) illustrates the frequency of the topic "Periodism and ethics in journalism" over the analyzed period, with a notable surge evident at the beginning of 2024. Upon closer examination, this surge in frequency coincided with a specific event: the revelation of a journalist's phone number by the president, as shown in the news piece shown in 10b). Following this event, there was a noticeable increase in conversations surrounding journalism ethics. This heightened attention led to more frequent mentions of the topic in the president's speeches during this period. This example underscores the dynamic nature of public discourse and highlights the responsiveness of our analysis in capturing and contextualizing real-world events within the speech data. Such insights offer valuable perspectives for understanding the evolving landscape of journalistic ethics and its intersection with political discourse.

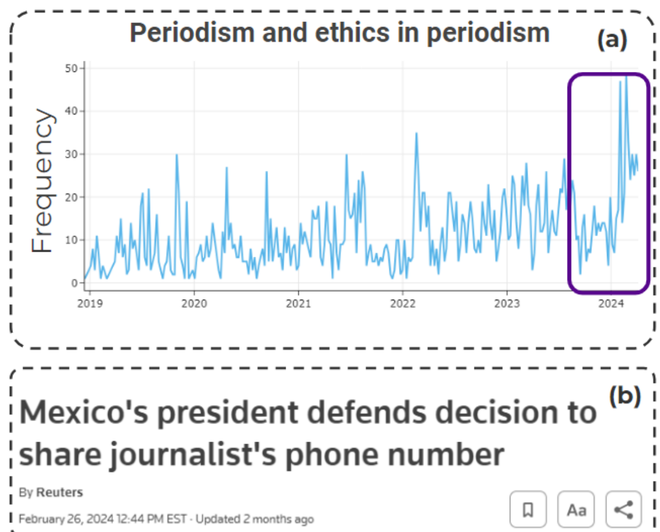


Fig. 10: (a) A line graph depicting the frequency of the topic 'Periodism and ethics in journalism' over time, with a notable increase observed at the beginning of 2024. This spike coincides with the revelation of a journalist's phone number by the president, leading to heightened discussion surrounding journalism ethics and increased mentions of the topic in the president's speeches during this period. (b) Includes a screenshot of a Reuters press print, coinciding with the date around the spike in frequency of the topic 'Periodism and ethics in journalism.' The screenshot provides additional context and correlation to the observed trend in the speech data analysis.

Advance Debugging Features for Expert Users. As we analyze over 300,000 sentences, dealing with numerous topics individually becomes impractical. To streamline this process, users may opt to merge highly similar topics. To aid in this task, Figure 11a) presents a similarity matrix among the topics. This matrix, computed using TF-IDF vectors and cosine similarity, reveals a cluster of closely related topics, highlighted in dark blue. Zooming in on this cluster, as shown in Figure 11b), we observe that many topics pertain to fuel prices. Depending on the nature of the analysis, merging these topics could prove advantageous, potentially simplifying the analysis process while retaining substantive insights. Lastly, to assist in explaining why a sentence was related to a certain topic, we present the user with a heatmap where they can understand which topic each part of the sentence most relates to, as shown in Figure 12.Notably, the heatmap confirms that the sentence's word components align with the expected topics, such as "airport.

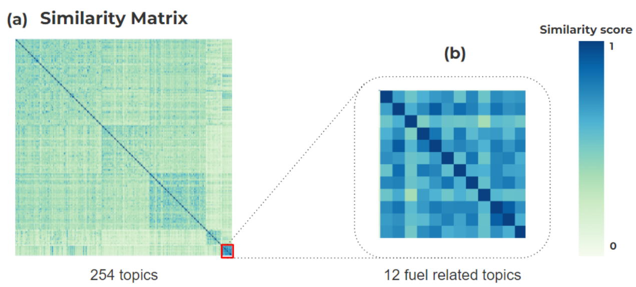


Fig. 11: Visualization showing a similarity matrix between the different generated topics. (a) The similarity matrix between all topics, (b) Zoom in on the cluster of related topics, where the user can conclude that they are all related to fuels.

5.2 Citizens Feedback

To assess the effectiveness of our tool, we conducted a qualitative user study involving two participants from diverse academic backgrounds: psychology and public policy. Both participants had varying degrees of engagement with the presidential discourse. Given their differing levels of interest, each participant approached the tool with a unique analytical task in mind, enabling us to evaluate the versatility of our tool across different user needs.

Each participant underwent a 30-minute session where we provided an overview of the tool's functionality and collected qualitative feedback. This approach allowed us to glean valuable insights into user experiences, preferences, and suggestions for improvement, ultimately informing future enhancements to our tool's usability and functionality.

Procedure. We began each session with an interview regarding their previous experience engaging in the presidential mañanera. After we introduced the participant to the tool, we gathered a first round of feedback about the value of the visualizations and the benefit and intuitiveness of the interaction. Each visualization was explained separately, without providing suggestions for a possible use case. The aim was to see if the participants intuitively created their workflow, and how that differed depending on their task. We then gave the participants full control over the interface. Each user had approximately 5 minutes to interact with our visualizations. We used the think-aloud method to gather information regarding their choices for using specific visualizations. We also made note of any challenges or issues the participants faced, if any.

The session ended with a semi-structured interview about the tool's intuitiveness, usability, and potential limitations or areas of improvement. Thus, the interview broadly covered the following questions:

(Q1) Can users easily navigate through the visualizations? Is the visualization tool intuitive?

(Q2) What types of usage do users employ our tool for, if any?

(Q3) What potential areas for improvement or limitations should be considered for future development?

Below, we delve into the outcomes of our evaluation, highlighting key observations and insights from participant interactions.

5.2.1 Intuitiveness

Notably, participants' experiences varied based on their familiarity with presidential speeches and their engagement with civic affairs. The participant with a background in public policy exhibited a more seamless and rapid engagement with specific tasks and filtering functionalities. This participant, being a recurrent follower of presidential speeches, seemed to enter the interaction with a preconceived idea of what he was seeking, facilitating swift navigation through the tool's features. Conversely, the participant from psychology, less engaged with public discourse initially, encountered some initial difficulty in identifying how to utilize the tool effectively. However, by the second minute, she displayed increased engagement and proficiency in navigating the visualizations. Though not intentionally planned out this way, this hints at the potential of our visualization tool to catalyze promoting user engagement with civic discourse, particularly among individuals who may find traditional news consumption burdensome.

5.2.2 Usability

* *Summarization and Credible Source.* The user with a psychology background expressed appreciation for our visualization tool, highlighting its reliability and impartiality in summarizing the president's statements. Unlike traditional news sources, which may carry inherent biases, she found our tool to be a trustworthy source of factual information extracted directly from the president's speeches, interviews, and press conferences. This emphasis on accuracy and transparency resonated with her, as it allowed for a clear understanding of the president's message without subjective interpretations or agenda-driven narratives. The user also valued the customization options available, which allowed for a tailored experience while upholding the integrity of the information provided. Moreover, the user appreciated the ability to cross-validate information from multiple sources, further enhancing credibility and reducing the risk of misinformation. With transparent processes and accountability in content generation, our tool instills trust as a credible source for unbiased summaries of the president's statements.
* *Correlated Events* Our participant with a psychology background confessed to lacking interest in presidential speeches, attributing it to their lengthy nature and her skepticism toward traditional news sources. Nonetheless, she expressed a willingness to engage more if provided with a tool like ours, offering additional contextual information related to the speeches, such as associations with events like crime rates or economic trends. This observation highlights the significance users attach to the tool's ability to connect political discourse with broader societal trends, signaling opportunities for enhancing engagement and relevance in future iterations.
* *Agenda Setting?* The user with a public policy background demonstrated familiarity with various scandals and political events. He swiftly navigated to specific thematic areas of interest, particularly focusing on spikes in certain topics. His keen interest in accessing additional news sources alongside the visualizations underscored the value he placed on corroborating information. Additionally, he highlighted the potential of our tool to uncover instances where the president may divert public scrutiny from important matters, such as bill passage, by amplifying less relevant events. This insight suggests a broader utility for our tool beyond mere agenda setting, serving as a tool for scrutinizing political tactics and promoting transparency in governance.

5.2.3 Limitations

Both participants highlighted the potential benefits of our tool having introductory video instructions or a clearer layout to facilitate user navigation. Additionally, the participant with a background in public policy expressed doubt about the widespread interest of citizens in presidential discourse, suggesting that the tool might be more suited for experts with a deeper interest in political debate. Moreover, this participant suggested that the visualizations could be too complex for regular users, emphasizing the ongoing need for improvement. These insights underscore the importance of diversifying the usage of our tool, as our current user study may not fully capture its diverse potential applications. Moving forward, we aim to conduct iterative interviews and gather feedback from users to refine and enhance our tool, ensuring it meets the needs of a broader audience.

	En	Santa	Lucía	estamos	haciendo	el	aeropuerto	Felipe	Ángeles	el	mejor	aeropuerto	del	mundo	que	se	esté	construyendo	el
3_aeropuerto_aeropuertos_avión_aviones	0.127	0.254	0.358	0.461	0.455	0.485	0.539	0.565	0.581	0.527	0.369	0.240	0.103						0.103
111_jóvenes_joven_generaciones_futuro															0.231	0.462	0.693	0.900	0.669
152_dólar_dólares_centavos_cambio																			
234_presupuesto_gasto_gastos_presidencia																			
239_ejemplos_gas_precio_barato																			
247_aeropuerto_aeropuertos_puerto_tulum				0.163	0.361	0.567	0.773	0.768	0.729	0.686	0.622	0.607	0.447	0.285	0.142			0.146	0.308
248_ángeles_2021_aeropuerto_secretaría				0.170	0.482	0.951	1.420	1.687	1.762	1.463	1.135	0.839	0.452	0.282	0.141				0.170

Fig. 12: A heatmap showing how different parts of a sentence are related with different topics.

6 Ending remarks

Our study presents DAVINCI, a visual analytics tool designed to address the complexities of analyzing political discourse, with a particular focus on Mexican presidential speeches. Through meticulous data preprocessing, advanced topic modeling techniques, and intuitive visualization methods, DAVINCI enables users to uncover meaningful insights and correlations within political discourse data, enhancing our understanding of societal trends and political dynamics. The multifaceted evaluation of DAVINCI, comprising a detailed case study and qualitative user feedback, provided valuable insights into its effectiveness and usability. The case study revealed DAVINCI's ability to identify temporal trends, sentiment shifts, and correlations with external events within the Mexican political landscape. Moreover, user feedback highlighted the tool's intuitive interface, credibility as a reliable information source, and potential for enhancing civic engagement among users with varying levels of interest in political discourse.

However, while DAVINCI demonstrates promising capabilities, it is not without limitations. Prompt dependency, LLM hallucinations, and sensitivity to parameter settings pose challenges that require careful consideration. Looking ahead, future research directions include automating the correlation of topics, developing specialized tools for streamlined analysis, and refining text-splitting processes. These efforts aim to enhance the usability, accuracy, and effectiveness of DAVINCI in uncovering nuanced insights from political discourse data. Furthermore, the exploration of dynamic topic modeling emerges as a promising avenue, allowing for real-time adaptation to evolving trends. By dynamically adjusting topic structures and weights based on changing discourse patterns, this approach ensures the continued relevance and timeliness of our analyses. These future directions hold the potential to significantly advance our understanding and utilization of natural language processing techniques and extract deeper insights from dynamic political narratives.

Our tool also exhibits promising potential as a comprehensive summarization tool, capable of rivaling traditional news sources. With its ability to efficiently distill vast amounts of political discourse into concise and relevant insights, our tool could serve as a valuable resource for individuals seeking timely and accurate information. However, to realize this potential fully, our tool must continue to refine its summarization algorithms and interface design, ensuring that users can easily access and digest pertinent information. By addressing these challenges, our tool can establish itself as a trusted source of comprehensive political summaries, competing effectively with existing news sources while offering unique advantages in terms of depth, accuracy, and accessibility.

The doubts raised about the widespread interest of citizens in the presidential discourse during the citizen interviews underscore the importance of considering the specific needs and preferences of the broader audience. It's possible that citizens are primarily interested in events concerning them, such as whether the president mentions something related to a social program or social aid. Therefore, summarizing information about these topics could be equally valuable. By incorporating these considerations into our tool's development, we can enhance its relevance and utility for a wider range of users, ultimately establishing it as a trusted and indispensable resource for accessing political news and analysis.

References

[1] Bank Of Mexico. https://www.banxico.org.mx/indexen.html. Accessed: 2024-04-04.

[2] Investing. https://www.investing.com/indices/ipc-chart. Accessed: 2024-04-04.

[3] Investing. https://www.inegi.org.mx/app/indicadores/?tm=0. Accessed: 2024-04-04.

[4] Presidential Website. https://lopezobrador.org.mx/capsulas-de-historia-conferencia-matutina/. Accessed: 2024-04-04.

[5] J. A. C. Aguilar and E. M. L. Domínguez. Presidential Communication and Its Impact on the Mexican Stock Market: Evidence Using a Sentiment Analysis Approach. *Latin American Business Review*, 0(0):1–37, 2024. doi: 10.1080/10978526.2024.2320655

[6] S. Aguilar Gomez, E. Arceo-Gomez, E. Toledo, and P. Torres López. Leadership, public health messaging, and containment of mobility in mexico during the covid-19 pandemic. *SSRN Electronic Journal*, 01 2021. doi: 10.2139/ssrn.3786398

[7] I. Ave. BETO sentiment analysis model for spanish. https://huggingface.co/ignacio-ave/beto-sentiment-analysis-spanish, Year.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3(null):993–1022, 30 pages, mar 2003.

[9] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Trans. Knowl. Discov. Data*, 10(1), article no. 5, 51 pages, jul 2015. doi: 10.1145/2733381

[10] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez. Spanish pre-trained bert model and evaluation data, 2023.

[11] clibrain.com. LINCE-ZERO: Llm for instructions from natural corpus en español. 2023.

[12] S. E. del Sistema Nacional de Seguridad Pública. Víctimas y Unidades Robadas, Nueva Metodología., 2015.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[14] N. Egami, C. J. Fong, J. Grimmer, M. E. Roberts, and B. M. Stewart. How to make causal inferences using texts. *Science Advances*, 8(42):eabg2652, 2022. doi: 10.1126/sciadv.abg2652

[15] M. El-Assady, V. Gold, C. Acevedo, C. Collins, and D. Keim. Contovi: Multi-party conversation exploration using topic-space views. *Comput. Graph. Forum*, 35(3):431–440, 10 pages, jun 2016.

[16] M. El-Assady, R. Sevastjanova, B. Gipp, D. Keim, and C. Collins. Nerex: Named-entity relationship exploration in multi-party conversations. *Computer Graphics Forum*, 36(3):213–225, 2017. doi: 10.1111/cgf.13181

[17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, 6 pages, p. 226–231. AAAI Press, 1996.

[18] V. Gold, A. Hautli-Janisz, and K. Holzinger. Visargue. *Zeitschrift für Konfliktmanagement*, 19, 03 2016. doi: 10.9785/zkm-2016-0308

[19] M. Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure, 2022.

[20] A. Hautli-Janisz and M. El-Assady. Rhetorical strategies in german argumentative dialogs. *Argument Comput.*, 8(2):153–174, 2017. doi: 10.3233/AAC-170022

[21] X. Jin and J. Han. *K-Means Clustering*, pp. 563–564. Springer US, Boston, MA, 2010. doi: 10.1007/978-0-387-30164-8_425

[22] X. Jin and J. Han. *Partitional Clustering*, pp. 766–766. Springer US, Boston, MA, 2010. doi: 10.1007/978-0-387-30164-8_631

[23] Magiobus. Amlopedia, Apr 2023.
[24] Magiobus. Finetuded llama, Apr 2023.
[25] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.
[26] Mistral AI. Mixtral of experts. https://mistral.ai/news/mixtral-of-experts/. Accessed [Date].
[27] Mistral AI Team. Mixtral-8x7b instruct model. https://huggingface.co/mistralai/Mixtral-8x7B-Instruct-v0.1, 2024. Version v0.1.
[28] S. Patel, S. Sihmar, and A. Jatain. A study of hierarchical clustering algorithms. In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 537–541, 2015.
[29] V. H. Reyna. "A Mercenary, a Thug... not Even a Journalist": The Stigmatization of News Workers in Mexico. *Journalism Practice*, 0(0):1–19, 2024. doi: 10.1080/17512786.2023.2300275
[30] R. Sevastjanova, M. El-Assady, A. Bradley, C. Collins, M. Butt, and D. Keim. Visinreport: Complementing visual discourse analytics through personalized insight reports. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4757–4769, 2022. doi: 10.1109/TVCG.2021.3104026
[31] L. South, M. Schwab, N. Beauchamp, L. Wang, J. Wihbey, and M. A. Borkin. Debatevis: Visualizing political debates for non-expert users. pp. 241–245, 2020. doi: 10.1109/VIS47514.2020.00055
[32] D. Sridhar and D. M. Blei. Causal inference from text: A commentary. *Science Advances*, 8(42):eade6585, 2022. doi: 10.1126/sciadv.ade6585
[33] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample. Llama: Open and efficient foundation language models, 2023.
[34] Y.-X. Wang and Y.-J. Zhang. Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1336–1353, 2013. doi: 10.1109/TKDE.2012.51
[35] X. Xu, M. Ester, H.-P. Kriegel, and J. Sander. A distribution-based clustering algorithm for mining in large spatial databases. In *Proceedings 14th International Conference on Data Engineering*, pp. 324–331, 1998. doi: 10.1109/ICDE.1998.655795
[36] Y. Yang, D. Cer, A. Ahmad, M. Guo, J. Law, N. Constant, G. H. Abrego, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil. Multilingual universal sentence encoder for semantic retrieval, 2019.