The role of gender in conflict prediction

Sylvie Cerise
University of Valle d'Aosta
Department of Economic and Political Sciences
Aosta, Italy
s.cerise4@univda.it

Consuelo Rubina Nava
University of Valle d'Aosta
Department of Economic and Political Sciences
Aosta, Italy
c.nava@univda.it

Paola Pisano
University of Turin
Department of Economics and Statistics
Torino, Italy
paola.pisano@unito.it

Stefano Tedeschi
University of Valle d'Aosta
Department of Economic and Political Sciences
Aosta, Italy
s.tedeschi@univda.it

***Abstract*—This paper examines gendered patterns of emotional expression and participation in the legislative discourse of the Regional Council of Aosta Valley (Italy), with a particular focus on identifying gender bias and potential indicators of gender-based conflict or violence in political communication. We construct a multilingual dataset that integrates parliamentary transcripts with speaker-level and contextual metadata, and apply a pre-trained transformer-based model (BERT) to classify the sentiment and identify violent or harmful language in each intervention.**

Building on these outputs, we employ a suite of supervised machine learning algorithms – including Naive Bayes, Logistic Regression, Decision Trees, Random Forest, and XGBoost – to predict sentiment polarity and detect subtle markers of verbal violence and bias, leveraging structural, contextual, and socio-demographic variables. Our findings reveal that female councilors intervene less frequently but are more likely to express negative sentiment, often linked to contentious or emotionally charged issues. Furthermore, structural features – such as intervention length and agenda position – emerge as stronger predictors of sentiment and violent expressions than gender alone. Notably, tree-based models deliver superior performance and offer interpretable insights into the complex institutional dynamics surrounding emotional and potentially violent tones. This study highlights the potential of AI-driven approaches to uncover latent forms of gender-based bias and violence in public discourse, fostering more inclusive and conscious institutional analysis.

***Index Terms*—Gender-based violence, gender bias detection, recognition of texts with violent content, deep learning for violence recognition, sentiment metrology, BERT.**

I. Introduction

Violence in various forms, including verbal and digital aggression, remains a pervasive social issue with significant consequences. While much research focuses on physical violence, non-violent conflict in institutional settings can also foster environments of exclusion and bias. Institutional conflict, although often perceived as a non-violent phenomenon, can exert a form of structural and psychological violence within an organization [1], [2]. Drawing from Galtung's concept of structural violence, institutional conflicts can manifest through exclusionary practices, power asymmetries, and implicit biases that limit access to opportunities, resources, or fair treatment [3]. These conflicts, whether arising from bureaucratic inefficiencies, discriminatory policies, or interpersonal power struggles, can systematically disadvantage certain groups, reinforcing social inequities and psychological distress.

Public administration, as a central arena for policy-making and governance, presents unique challenges in detecting and managing these forms of conflict. Gender plays a critical but often under-explored role in shaping these dynamics, influencing not only participation and tone but also the escalation and resolution of institutional disputes [4], [5]. Institutional discourse is, thus, not only a site of policy deliberation but also a stage where power relations and social norms – including those related to gender – are performed and reproduced.

At the intersection of computer science, statistics, and gender studies, this study explores how gendered dynamics unfold in institutional communication. While qualitative approaches have long examined these issues, computational tools for large-scale, multilingual, and context-sensitive analysis remain limited. The growing availability of digital legislative records, combined with advances in artificial intelligence (AI) and natural language processing (NLP), now enables researchers to examine these patterns at scale systematically.

This paper presents a novel, AI-driven approach to analyzing gendered conflict within legislative discourse. Focusing on the Regional Council of Aosta Valley (Italy), we construct an original dataset that integrates parliamentary transcripts with socio-demographic information of elected officials. Leveraging machine learning techniques – including transformer-based sentiment classifiers such as BERT [6] – we investigate how speaker gender, political affiliation, and discourse structure correlate with variations in emotional tone and sentiment polarity, quantifying discourse dynamics through both sentiment

Funded by the European Union – NextGenerationEU, Mission 4 Component 1.5 – ECS00000036 – CUP B63B22000010001 (C.R. Nava and S. Tedeschi) and the project "Gender Inclusion and Artificial Intelligence for Conflict Prediction" as part of the "2024 Ordinary Grants Call" – first session issued by Fondazione CRT – CUP B67G24000380009 (S. Cerise and C. R. Nava).

analysis and structural features.

Specifically, our objectives are threefold:

1) Detect patterns of participation and sentiment expression across gender;
2) Quantify the role of contextual and structural variables in shaping sentiment;
3) Assess the effectiveness of supervised learning methods for classifying sentiment in institutional contexts.

In doing so, we contribute to ongoing debates on the role of AI in political text analysis, particularly regarding fairness, representation, and the detection of implicit conflict. We also highlight how computational methods can support more inclusive and reflexive governance practices by producing interpretable indicators of discourse quality and emotional dynamics.

The structure of the paper is as follows. Section II presents related work. Section III introduces the dataset, describes the feature engineering process, and outlines the sentiment analysis and classification methodology. Section IV presents the results while Section V discusses implications and limitations. Finally, Section VI concludes with directions for future research.

II. Related work

While a vast body of research has focused on violent conflict, the study of non-violent forms of conflict – particularly within institutional and administrative settings – remains relatively underexplored. This gap is especially notable when considering the role of gender, which, despite its recognized influence on conflict dynamics, receives limited attention in computational analyses of public discourse.

Existing literature suggests that gender influences both the emergence and management of conflict [7], [8]. However, in institutional discourse, few works have systematically analyzed how gendered dynamics manifest, escalate, or shape the tone and affective style of public deliberation. For example, female politicians are often subject to stricter tone policing and are more likely to adopt cooperative or conciliatory speech patterns [5], [9]. Yet, these dynamics are rarely captured or quantified in large-scale datasets.

In this work, we treat sentiment polarity as a proxy for institutional conflict, assuming that affective language – particularly negative or emotionally charged tone – reflects underlying tensions, disagreements, or strategic confrontation. Sentiment analysis offers a scalable method for capturing such dimensions across thousands of institutional interventions.

Recent studies have employed sentiment analysis in parliamentary speeches to examine affective trends, often with gender as a key factor. For instance, [10] analyzed over 50,000 speeches in the Austrian Parliament and found that women tend to use less negative language than men, and that higher female participation correlates with a less hostile tone overall. Similar computational methods have been used in political communication to detect polarization, emotional appeal, and strategic tone shifts on social media [11].

Technically, sentiment analysis has evolved from rule-based systems to modern transformer-based architectures, such as BERT [6], which offer strong performance on nuanced and context-dependent tasks. These models have been applied in multiple political domains, including hate speech detection, campaign analysis, and policy framing [12]. However, applications in institutional discourse – particularly at the local government level and with attention to speaker demographics – remain limited.

Our work contributes to filling this gap by applying multilingual sentiment classification to a novel dataset of council interventions, enriched with speaker metadata and gender information. We aim to explore whether gender correlates with differences in sentiment, both in absolute terms and relative to contextual factors such as political role, agenda type, or session dynamics.

III. Methodology

This study employs a multi-step computational approach to investigate emotional tone and gendered communication in institutional debates. The pipeline consists of four main stages, as shown in Figure 1: (i) data acquisition and preprocessing, (ii) feature engineering, (iii) sentiment classification using BERT, and (iv) predictive modeling with supervised machine learning models.

A. Data acquisition and preprocessing

The analysis rests on a novel dataset constructed from the official transcripts of the Regional Council of Aosta Valley (Italy). Specifically, it covers proceedings from the last two legislative terms (2018–2024), totaling over 5,400 documents and approximately 30,000 individual interventions. The raw data were sourced from publicly accessible institutional archives and consist of full transcripts of plenary sessions[1].

For each document, we extracted speaker metadata (name, gender, political affiliation, and institutional role), session-level metadata (date and agenda item), and the verbatim text of each intervention. Non-substantive speech acts – such as procedural remarks or routine moderation by the chairperson – were filtered out to retain only content that was emotionally or argumentatively relevant.

Additional socio-demographic attributes of council members, including gender, year of birth, and seniority (measured as the number of legislatures served), were manually integrated from public records and official biographies. Text preprocessing involved tokenizing and segmenting interventions using regular expressions and rule-based parsing, allowing for precise alignment between textual content and speaker/session metadata.

An overview of the raw intervention dataset is provided in Table I, which illustrates representative entries before feature extraction.

[1]https://www.consiglio.vda.it/app/oggettidelconsiglio

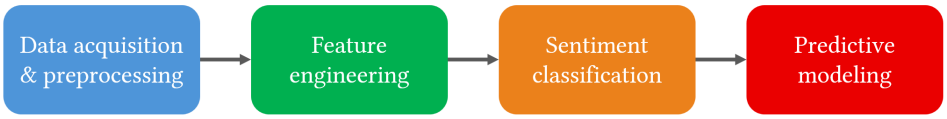

Fig. 1. Methodology pipeline.

	ID_file	leg	date	class	ID_cons	chunk
0	41041	XV	09/01/2019	CONSIGLIO REGIONALE, Attività consiliare d'aula	211	Non pensavo di intervenire, però quanto è ...
1	43345	XVI	24/06/2021	ASSISTENZA SOCIALE, Disabili	185	Bene, sono felice di apprendere dalle sue parole ...
2	48096	XVI	19/12/2024	ATTIVITÀ CULTURALI	178	Vede, Assessore, me l'aspettavo sicuramente ...

TABLE I
EXAMPLE ENTRIES FROM THE RAW INTERVENTION DATASET.

B. Feature engineering

To model sentiment dynamics and their potential predictors, we engineered features at three main levels of granularity:

- Speaker-level features: gender (binary), year of birth, political group affiliation, and seniority (number of legislative terms served);
- Intervention-level features: character length of the speech act and position within the debate;
- Discussion-level features: legislature, agenda position, item category (e.g., legislation, questioning, budget), and month of the year.

Feature extraction was performed by processing both transcript content and associated metadata. For example, intervention length was calculated as the number of characters in each speech act, while debate position was derived from the sequential order of interventions within each agenda item. Speaker attributes were matched from institutional records, and discussion-level variables such as agenda category and session month were derived from session metadata.

These structured features serve both descriptive and inferential purposes. On the one hand, they enable a rich characterization of institutional communication practices; on the other hand, they are employed as predictors in supervised machine learning models for sentiment classification.

By integrating linguistic, structural, and socio-demographic dimensions, our feature engineering strategy facilitates a nuanced exploration of institutional discourse. It allows us to investigate how emotional tone varies across individuals and contexts, and to identify potential gender-based or institutional biases in communication style, participation, and framing.

Representative examples of the structured feature set used for modeling are shown in Table II.

C. Sentiment classification with BERT

To measure emotional tone, we employ a transformer-based architecture, specifically BERT (Bidirectional Encoder Representations from Transformers) [6], which is known for its ability to capture deep contextual relationships in language. We specifically use the *bert-base-multilingual-uncased-sentiment* model, a variant fine-tuned for multilingual sentiment analysis across six languages. This model is particularly suited for political discourse in bilingual contexts such as the Aosta Valley, where interventions may switch between Italian and French.

The model architecture comprises 12 transformer layers, 768 hidden units, and 12 self-attention heads, which are helpful in capturing nuanced emotional tones and conflict trends more effectively than traditional methods [13]. Its uncased vocabulary ensures consistent tokenization across all capitalizations, which helps normalize institutional speech patterns. For each intervention, the model outputs a probability distribution over five sentiment classes: very negative, negative, neutral, positive, and very positive – offering a nuanced and multidimensional view of emotional dynamics.

Since BERT performs its tokenization and normalization, no additional text preprocessing was required beyond the exclusion of procedural utterances (typically by the presiding officer), which were assumed to be structurally neutral and thus not informative for sentiment modeling.

This fine-grained sentiment output enables us to track tonal shifts over time and across topics, and to investigate correlations between emotional expression and speaker characteristics, with a particular focus on gender-based patterns.

D. Predictive modeling and model selection

To examine which factors most influence sentiment polarity, we cast the task as a binary classification problem: predicting whether an intervention is emotionally positive or negative. Neutral interventions were excluded from this step. Moreover, we focused specifically on interventions with marked polarity, i.e., those where the model assigned an above-average score to either the very positive or very negative class. To determine the final label, we summed the very positive and positive scores to compute a composite positive score, and likewise summed the very negative and negative scores to obtain a composite negative score. An intervention was labeled as positive if its positive score exceeded its negative score and its very positive score was above the dataset mean; conversely, it was labeled as negative if its negative score exceeded its positive score and its very negative score was above the mean. This selective filtering enhances the signal-to-noise ratio by concentrating on emotionally salient interventions that are more likely to reflect strategic or expressive choices by speakers.

	leg	month	class	obj_pos	year_birth	gender	group	seniority	position	length
0	0	1	3	1	1960	1 (F)	0 (ALPE)	3	3	2663
1	1	6	5	10	1983	0 (M)	2 (LEGA VDA)	2	4	406
2	1	12	1	24	1968	0 (M)	1 (UV)	3	3	3409

TABLE II
STRUCTURED FEATURE SET USED FOR MODELING.

We tested five standard machine learning algorithms: Naive Bayes, Logistic Regression, Decision Tree, Random Forest, and XGBoost. Each model was trained using an 80/20 stratified train-test split, with hyperparameter tuning performed via grid search and 5-fold cross-validation.

Among the models, XGBoost achieved the best trade-off between predictive performance and interpretability. It demonstrated high accuracy and robustness across folds while offering clear feature importance metrics. These metrics were used to assess the relative influence of gender, political group, speech structure, and session characteristics on emotional tone.

Evaluation metrics include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). These allow us to quantify the predictive power of our features and assess the extent to which gender acts as a statistically significant determinant of emotional polarity in discourse.

IV. Preliminary Results

This section presents the empirical findings based on the 2018-2024 subset of the legislative transcript dataset. Our analysis focuses on (i) gender-based differences in discourse patterns and sentiment, (ii) the predictive performance of the tested models, and (iii) the interpretability of feature contributions to sentiment variation[2].

A. Descriptive statistics

Preliminary analysis revealed notable differences in discourse engagement between male and female councilors. On average:

- Male councilors intervened 4.02 times per session, with an average length of 3,390 characters;
- Female councilors intervened 2.46 times per session, with a slightly longer average of 3,296 characters.

Although the length difference is marginal, the frequency gap is substantial, indicating lower participation by women in formal debates.

Sentiment scores across interventions were aggregated and compared by gender. Table III shows the mean distribution of sentiment classes and corresponding p-values (from two-sample t-tests).

Statistically significant differences emerged across all sentiment classes ($p < 0.01$) except for the "neutral" score. Female councilors exhibited higher negative sentiment, while male councilors were more often associated with positive tones. These findings support the hypothesis that gender influences the emotional framing of institutional discourse.

[2]The source code of the analysis is available at: https://colab.research.google.com/drive/14Anr0uTOaD3bXi-7GT5ZRh9M9BNI9S17

	mean very_negative	mean negative	mean neutral	mean positive	mean very_positive
Female	0.171	0.265	0.255	0.204	0.105
Male	0.165	0.251	0.254	0.218	0.112
p-value	6.6e-03	1.9e-17	3.5e-01	1.6e-15	4.0e-6

TABLE III
COMPARISON OF THE SENTIMENT ANALYSIS BETWEEN GENDERS.

	accuracy	precision	recall	f1-score
NB	0.53	0.53	0.53	0.53
LR	0.55	0.56	0.55	0.55
DT	0.64	0.64	0.64	0.64
RF	0.66	0.66	0.66	0.66
XGB	0.67	0.67	0.67	0.67

TABLE IV
COMPARISON OF THE PERFORMANCE OF THE DIFFERENT MODELS.

B. Predictive models performance

The predictive models were trained to classify interventions as having either a positive or negative sentiment. Table IV and Figure 2 summarize the performance of the tested models.

Although overall performance remained moderate, the results indicate that structural and speaker-level features hold explanatory value in modeling emotional tone. As shown in Table IV, the best-performing models were Random Forest and XGBoost, both achieving an AUC of 0.72 and an F1-score around 0.67. Simpler classifiers, such as Naive Bayes and Logistic Regression, underperformed, with AUC scores of 0.52 and 0.58, respectively. These findings highlight the added value of tree-based ensemble methods in capturing complex patterns in institutional discourse data.

To better understand what drives model predictions, Figures 3 and 4 present the features ranked by importance according to the Random Forest and XGBoost models, re-

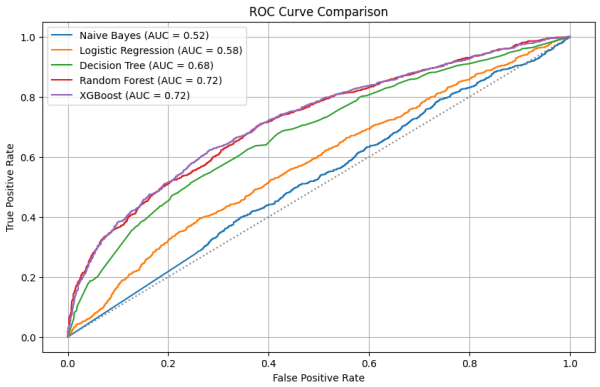

Fig. 2. ROC curve comparison across models.

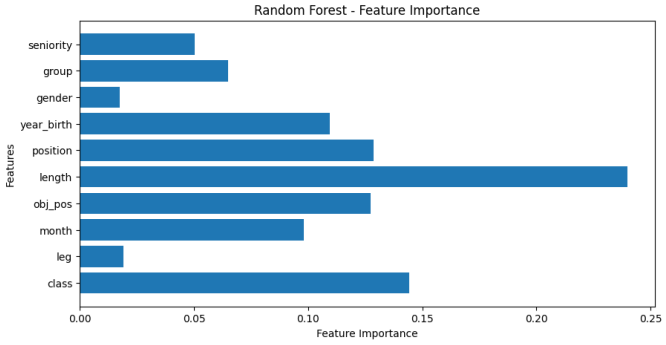

Fig. 3. Feature importance for Random Forest model.

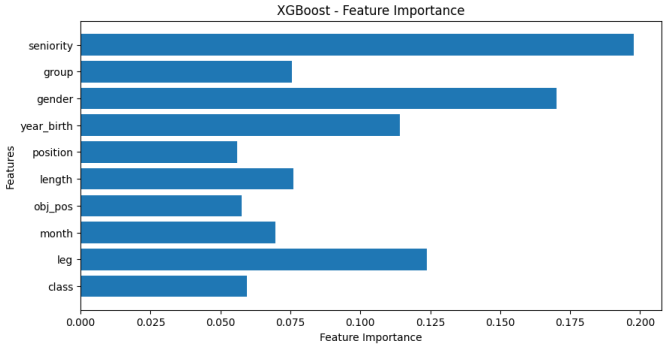

Fig. 4. Feature importance for XGBoost model.

spectively. In the Random Forest classifier, intervention length emerged as the most influential predictor of sentiment polarity, followed by discussion class and position in the agenda and discussion. Variables such as birth year, discussion month, and seniority showed moderate relevance, while gender and legislature played a minimal role.

By contrast, the XGBoost model assigned greater importance to seniority and gender, followed by legislative term, year of birth, and type of discussion. This shift suggests that different ensemble methods may emphasize distinct aspects of the data structure or interactions between features. Notably, while gender appeared marginal in the Random Forest model, it ranked among the top predictors in XGBoost – pointing to possible model-specific sensitivities or latent patterns. These discrepancies highlight the importance of comparing multiple models not only in terms of performance but also in their interpretability and emphasis on different explanatory variables.

V. Discussion

The results presented in the previous section demonstrate that institutional sentiment can be effectively modeled using supervised machine learning techniques, particularly ensemble-based classifiers. Among the tested models, Random Forest provided the highest performance across all metrics, confirming its robustness and interpretability in this context.

The successful application of a multilingual BERT-based sentiment classifier further supports the feasibility of applying large pre-trained language models to bilingual institutional corpora. The use of probability distributions across five sentiment classes allowed a more granular assessment of emotional tone, which was then filtered and reduced to a binary task to improve classification reliability. This selective labeling procedure enhanced the signal quality by focusing on strongly polarized interventions, which are more likely to carry meaningful emotional or strategic content.

Feature importance analysis revealed that structural and contextual variables – such as intervention length, discussion type, and agenda position, were the most influential in predicting sentiment polarity. By contrast, speaker-level features such as gender and age contributed less to predictive performance, suggesting that gendered dynamics may operate indirectly through institutional structures that constrain participation and framing rather than through sentiment itself. However, the descriptive statistics reveal consistent and statistically significant differences in sentiment expression across genders, particularly in the higher prevalence of negative tones among female councilors. This discrepancy underscores the importance of interpreting feature importance within machine learning models carefully, as predictive power does not necessarily align with sociopolitical relevance.

Overall model performance remains moderate (F1-score = 0.67), indicating both the promise and current limitations of this approach. While the existing feature set captures key metadata and structural aspects of the discourse, further improvements may be achieved by integrating semantic or interactional features, such as dialogue flow, interjections, or speaker replies. These could help capture more complex patterns of emotional alignment, escalation, or dissent within institutional communication.

From a methodological standpoint, the workflow ensures internal validity through cross-validation, stratified sampling, and a clear model selection process. The interpretability of Random Forest enables the identification of dominant predictors, supporting the broader aim of producing transparent and explainable indicators of institutional behavior. These aspects align with current research priorities in computational metrology, particularly in the areas of trust, reproducibility, and fairness in AI-assisted policy analysis.

Beyond its immediate focus on gendered dynamics and sentiment polarity, the proposed predictive system may have broader implications for the early detection of emotionally charged or potentially conflictual discourse. In this respect, the ensemble of machine learning models – particularly XGBoost and Random Forest – could be repurposed as a monitoring tool for verbal tension in institutional settings. By identifying interventions with extreme sentiment profiles in real time, such a system could inform preventive strategies aimed at mitigating verbal escalation or fostering more constructive deliberation.

In summary, the proposed framework demonstrates that AI-based analysis can provide scalable and interpretable tools for measuring emotional dynamics in legislative discourse, laying the foundation for future developments in monitoring and assessing institutional quality.

VI. Conclusion

This study proposed a novel pipeline for analyzing emotional tone in institutional discourse, integrating multilingual sentiment classification with supervised machine learning on structured parliamentary data. Focusing on the Regional Council of Aosta Valley, the analysis revealed measurable gender-based differences in participation and sentiment expression, as well as structural patterns linked to agenda dynamics and intervention length.

The combination of BERT-based sentiment analysis and Random Forest classification proved effective in identifying emotionally polarized content and modeling its predictors. While speaker-level features such as gender showed limited predictive power in isolation, their role remains relevant in descriptive terms, warranting further investigation into indirect forms of institutional bias.

Methodologically, the approach emphasizes interpretability, scalability, and cross-linguistic adaptability – key criteria for deploying AI in sensitive institutional contexts. These findings contribute to the emerging field of computational metrology for institutional quality, where emotional tone can serve as a proxy for conflict, exclusion, or procedural imbalance.

While this study does not directly address verbal violence, the methodological framework it introduces – particularly the combination of sentiment detection and supervised modeling – may serve as a foundational layer for systems designed to monitor and prevent escalatory discourse. The ability to flag emotionally extreme interventions based on predictive features opens the possibility of real-time alert mechanisms to support moderation, de-escalation, or mediation protocols in deliberative settings. As such, the pipeline may contribute not only to measuring emotional tone but also to the broader aim of promoting safer and more inclusive institutional communication.

Future work will aim to expand the dataset, incorporate semantic and dialogic features, and refine classification models to enhance predictive accuracy. Moreover, the framework can be extended to other institutional contexts to support comparative analysis and contribute to the development of more inclusive and responsive governance systems.

References

[1] L. Bosi, S. Malthaner *et al.*, "Political violence," *The Oxford handbook of social movements*, pp. 440–451, 2015.

[2] E. Chenoweth and A. Lawrence, *Rethinking violence: States and non-state actors in conflict*. MIT press, 2010.

[3] A. Dilts, Y. Winter, T. Biebricher, E. V. Johnson, A. Y. Vázquez-Arroyo, and J. Cocks, "Revisiting johan galtung's concept of structural violence," *New political science*, vol. 34, no. 2, pp. e191–e227, 2012.

[4] J. Birchall, "Gender as a causal factor in conflict," The Institute of Development Studies and Partner Organisations, Tech. Rep., 2019.

[5] J. Gray, *Men are from Mars, women are from Venus*. HarperCollins, 1993.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[7] M. Caprioli, "Primed for violence: The role of gender inequality in predicting internal conflict," *International studies quarterly*, vol. 49, no. 2, pp. 161–178, 2005.

[8] E. Melander, "Gender equality and intrastate armed conflict," *International Studies Quarterly*, vol. 49, no. 4, pp. 695–714, 2005.

[9] O. N. Bordean, D. S. Rácz, S. I. Ceptureanu, E. G. Ceptureanu, and Z. C. Pop, "Gender diversity and the choice of conflict management styles in small and medium-sized enterprises," *Sustainability*, vol. 12, no. 17, p. 7136, 2020.

[10] M. Haselmayer, S. C. Dingler, and M. Jenny, "How women shape negativity in parliamentary speeches—a sentiment analysis of debates in the austrian parliament," *Parliamentary Affairs*, vol. 75, no. 4, pp. 867–886, 2022.

[11] K. Karpouzis, S. Kaperonis, and Y. Skarpelos, "Identification of common trends in political speech in social media using sentiment analysis," *arXiv preprint arXiv:2210.07600*, 2022.

[12] E. del Valle and L. de la Fuente, "Sentiment analysis methods for politics and hate speech contents in spanish language: a systematic review," *IEEE Latin America Transactions*, vol. 21, no. 3, pp. 408–418, 2023.

[13] A. Sahoo, R. Chanda, N. Das, and B. Sadhukhan, "Comparative analysis of bert models for sentiment analysis on twitter data," in *2023 9th International Conference on Smart Computing and Communications (ICSCC)*. IEEE, 2023, pp. 658–663.