Decoding Narratives: Towards a Classification Analysis for Stereotypical Patterns in Italian News Headlines

Matteo Berta
DAUIN
Politecnico di Torino
Turin, Italy
matteo.berta@polito.it

Salvatore Greco
DAUIN
Politecnico di Torino
Turin, Italy
salvatore_greco@polito.it

Giuseppe Tipaldo
DAUIN
Politecnico di Torino
Turin, Italy
giuseppe.tipaldo@polito.it

Tania Cerquitelli
DAUIN
Politecnico di Torino
Turin, Italy
tania.cerquitelli@polito.it

***Abstract*—Media headlines shape our initial interpretation of news, framing narratives that influence societal engagement with political and social issues. Yet, they often rely on sensationalism and bias to capture readers' attention.**

In this paper, we aim to uncover distinct patterns in Italian headline composition, examining how language and framing vary across political leanings. We analyze a dataset of daily Italian newspaper articles from two outlets with opposing political perspectives, anonymized as NEWSPAPER A and NEWSPAPER B. Our study encompasses the entire set of news and a subset of topics ($n = 8$) likely to contain stereotypes or clickbait headlines identified using a Large Language Model. Our methodology combines (1) a lexicometric analysis to identify characteristic words of each newspaper, and (2) the training of an accurate deep learning classifier ($F1 = 0.84$) to learn specific patterns for categorizing headlines into these two perspectives and leveraging explainability techniques to extract and interpret these patterns.

Our analysis reveals distinct tonal differences between the two newspapers: NEWSPAPER A generally adopts a more balanced and nuanced approach, while NEWSPAPER B often favors a more direct and sometimes provocative style, especially regarding topics like immigration and social justice. Additionally, NEWSPAPER B's headlines tend to be brief and punchy, in contrast to the longer, more detailed ones from NEWSPAPER A. Despite these tonal differences, both outlets exhibit similar stereotypical patterns in their coverage, such as consistently emphasizing nationality and group distinctions in ways that can reinforce social stereotypes. This shared tendency suggests that, although their narrative strategies differ, both outlets could contribute to a broader pattern of stereotype reinforcement.

***Disclaimer*: This research is conducted for research purposes only and does not reflect the personal political views of the authors. The objective is to analyze and understand the framing of political narratives rather than to endorse any particular political ideology.**

Code: https://github.com/MatteoBerta/Decoding-Narratives

***Index Terms*—news classification, NLP, stereotype detection**

I. INTRODUCTION

In our daily lives, as we navigate the internet, we are constantly flooded with news headlines. These short, attention-grabbing statements serve as our initial point of contact with global events, shaping our understanding and influencing our perceptions of the world. Designed to capture interest within seconds, headlines often prioritize engagement over depth, frequently highlighting sensational or emotionally charged aspects of a story, which can shape our views before we engage further [1]. Consequently, headlines significantly affect public opinion, framing discourse and influencing the stories we explore. Additionally, the communication environment influenced by digital media, social networks, and online news platforms contributes to deepening affective polarization [2].

However, headlines often rely on sensationalism and bias to capture readers' attention. As a result, these platforms often promote and amplify extreme portrayals of groups with differing beliefs, especially those that audiences may already view with suspicion or dislike. Algorithms that prioritize engagement frequently favor emotionally charged content, showcasing exaggerated differences and reinforcing divisive stereotypes [3] and, consequently, readers often encounter negative or hostile representations of "out-groups", which can lead them to perceive these groups as more extreme and antagonistic than they truly are [4]. This ongoing exposure may foster feelings of distrust and misunderstanding, contributing to divisions along ideological lines and possibly creating a sense of separation, encouraging an "us versus them" mentality that influences our social and political interactions [5–8].

Our work conducts a preliminary analysis to uncover stereotypical patterns in the news headlines of two politically opposing outlets. By leveraging Natural Language Processing (NLP) techniques, we examine how specific words and stereotypes shape narratives, reinforce particular viewpoints, and influence public perception and readers' opinions. Through this investigation, we make two main contributions:

1) *Methodology*: We employ a dual approach: a lexicometric analysis to uncover linguistic patterns and a deep learning classifier enhanced with explainability techniques to learn, extract, and interpret specific patterns for categorizing headlines into the two outlets (Section III);
2) *Analysis*: We analyzed a dataset of Italian news headlines of two opposing political outlets. Our findings reveal potential correlations between ethnicity-related terms and crime coverage across outlets (Section IV).

II. Related Works

In this section, we review research on the art of crafting news headlines, including the emergence of clickbait strategies (Section II-A), and analyze how narrative framing and word choice shape audience perceptions and reinforce biases (Section II-B). Additionally, we investigate methodologies for identifying stereotypical content in news media, utilizing NLP techniques to detect, analyze, and quantify stereotypes embedded within textual narratives (Section II-C).

A. News Headlines and Clickbaits

The study on news headlines highlights their pivotal role in shaping public opinion and influencing societal attitudes. For instance, a study conducted in collaboration among different universities in the USA and Mexico [3] investigated affective polarization, and emotional division along partisan lines, examining how political conversations contribute to this phenomenon. In this study, media play the role of an enabler of political discussions rather than a direct source of effective polarization. Still, their contribution is fundamental for starting political discussions, regardless of political alignment.

Other studies, such as [14], examine the relationship between social media news consumption and political knowledge, confidence in knowledge, and misinformation. This study shows that social media news consumption correlates with reduced factual political knowledge. Indeed, in environments like social media, news headlines play a pivotal role in how users perceive and discover what happens in the world daily, because people often slide through the headlines without clicking to read the full story.

The phenomenon described has created a demand for *catchy* headlines that effectively capture users' attention. This has prompted research into the realm of *clickbait*–content specifically designed to attract clicks and engage audiences. Various studies have examined this style of headline writing, exploring its role as a strategy for viral journalism [15], its emergence in the social media era [16], and its widespread prevalence [17]. Others have analyzed the outcomes of clickbait strategies, focusing on the engagement they generate [18, 19], revealing limited effectiveness and questioning their utility.

B. Narrative Framing and Word Selection

Studies across multiple fields confirm the importance of word choice in shaping narrative tone, perspective, and reader impact. By carefully choosing words, journalists can influence how audiences relate to the news and to the people involved.

Building on research that highlights the impact of word choice in shaping narrative tone, Roland Barthes' Cultural Code [20] offers valuable insight into how shared cultural knowledge shapes audience interpretation. Applying this code to news media it is possible to say that culturally loaded terms or historical events can reinforce certain perspectives or evoke specific emotional responses. Understanding the Cultural Code could be essential when analyzing a large corpus of text to assess how specific words or references activate cultural associations, impacting public opinion.

Academic studies show that word selection can also shape the trust and distrust of individuals [21] and that word choice is a fundamental element of narrative and could have a big impact on the perception of bias [22].

C. Stereotypes Detection in News

Various studies have attempted to quantify stereotypes in news media using NLP echniques [23–25]. These approaches aim to analyze and measure the presence of stereotypes within news articles, providing insights into how language shapes public perception and representation. These studies typically utilize word embeddings to investigate stereotypical associations in news media, focusing principally on ethnic stereotypes. Their findings reveal that groups linked to less wealthy countries, as well as those from culturally distant nations, experience stigmatization, manifesting both explicitly and implicitly [26], highlighting the role of language in perpetuating stereotypes and shaping societal perceptions.

Starting from these works, we aim to examine how two of the most representative Italian newspapers differ in their narratives based on the political biases of both writers and readers. We will specifically investigate the presence of stereotypical patterns in the formulation of news headlines, highlighting the impact of perspectives on media representation, using modern techniques like Large Language Models (LLMs) and Explainable AI (XAI) to identify the source of stereotypes.

III. Methodology

Figure 1 outlines the main data analytics steps performed to identify stereotypes in a dataset of Italian news headlines (CHANGE-IT [11]), described in Section III-A. We first utilized an LLM (Llama-3.1-8B-Instruct [9]) to categorize the topics of headlines (Section III-B). This categorization allowed us to filter the original dataset by selecting relevant topics likely to contain stereotypes. We conducted a comprehensive lexicometric analysis (Section III-C), aiming to uncover linguistic patterns and compare the explainability findings for a comprehensive view of the results. We then trained a classifier to distinguish the originating newspaper from the headline acting as a proxy to learn patterns that distinguish the newspaper's writing style (Section III-D) and we employed explainability techniques to identify significant features for both outlets (Section III-E). Finally, we exploited an LLM to categorize the identified words into categories to better understand the nature of the stereotypes (Section III-F).

A. Italian news headlines dataset

The CHANGE-IT [10, 11] dataset[1] consists of around $152,000$ article-headline pairs sourced from two Italian newspapers with contrasting political orientations. In this work, we will call them: Newspaper A and Newspaper B. Both newspapers are equally represented in terms of number of headlines ($\approx 76{,}000$ each). The dataset contains both headlines and full articles, but this work is focused on the headlines.

[1] https://huggingface.co/datasets/gsarti/change_it

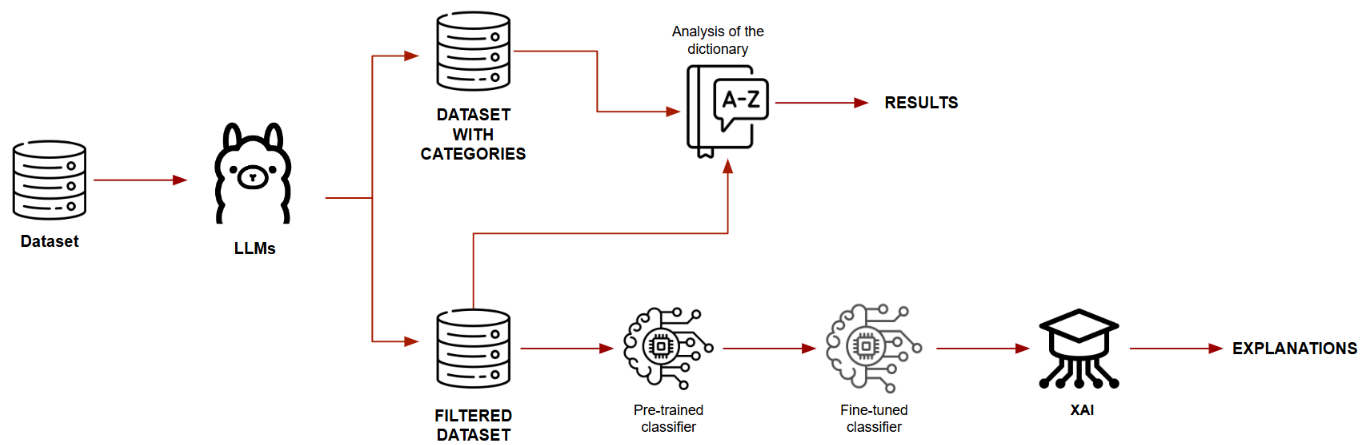
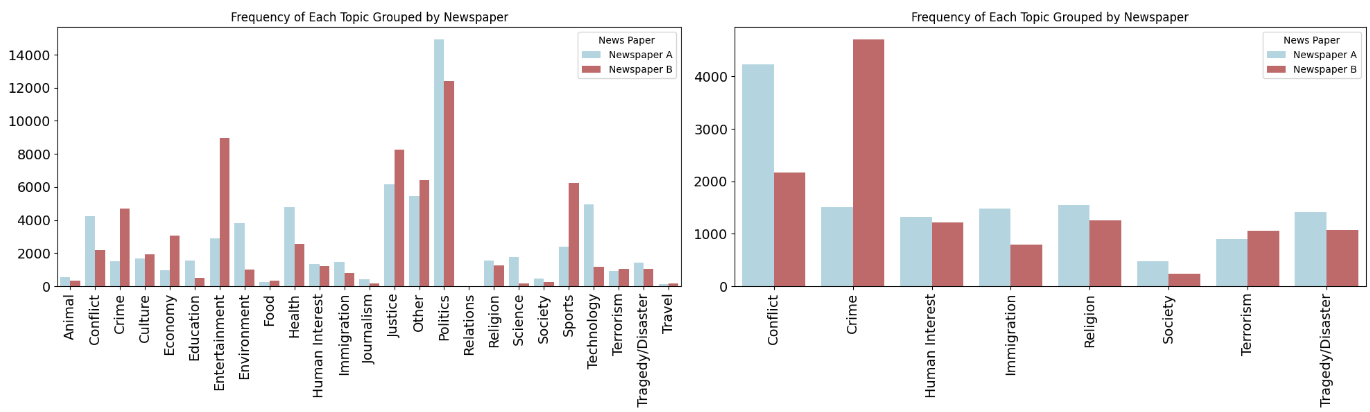


Fig. 1: **Pipeline for research workflow**. First, LLaMa 3.1 [9] was used to categorize topics in the CHANGE-IT [10, 11] dataset, enabling targeted filtering by relevant topics. Next, an Italian BERT [12] classifier identified the originating newspaper for each headline. An analysis of SHAP [13] explanations was conducted to highlight significant features across outlets, and a dictionary-based analysis of both datasets provided a comprehensive comparison of results with SHAP insights.



(a) Complete CHANGE-IT Dataset

(b) Filtered CHANGE-IT Dataset

Fig. 2: Topic Distribution for each outlet in the complete and filtered dataset

B. Topic categorization

We conducted a preliminary analysis of the dataset to identify the most frequent topics reflected in the news headlines. To achieve this, we prompted a LLaMA 3.1 [9] model with 8B parameters[2] without providing predefined topics. This process generated over 100 unique topics. These topics were then manually aggregated and refined into a final set of 25 significant topics. Finally, we prompted LLaMA again using the entire dataset, explicitly specifying the 25 selected topics in the input prompt. As a result, each headline was categorized into one and only one of these 25 topics.

The distribution of headlines per topic across the entire dataset is illustrated in Figure 2a. It reveals that NEWSPAPER B places greater emphasis on topics such as *entertainment*, *sports*, and *crime*, whereas NEWSPAPER A focuses more on themes like *conflict*, *politics*, *health*, and *technology*.

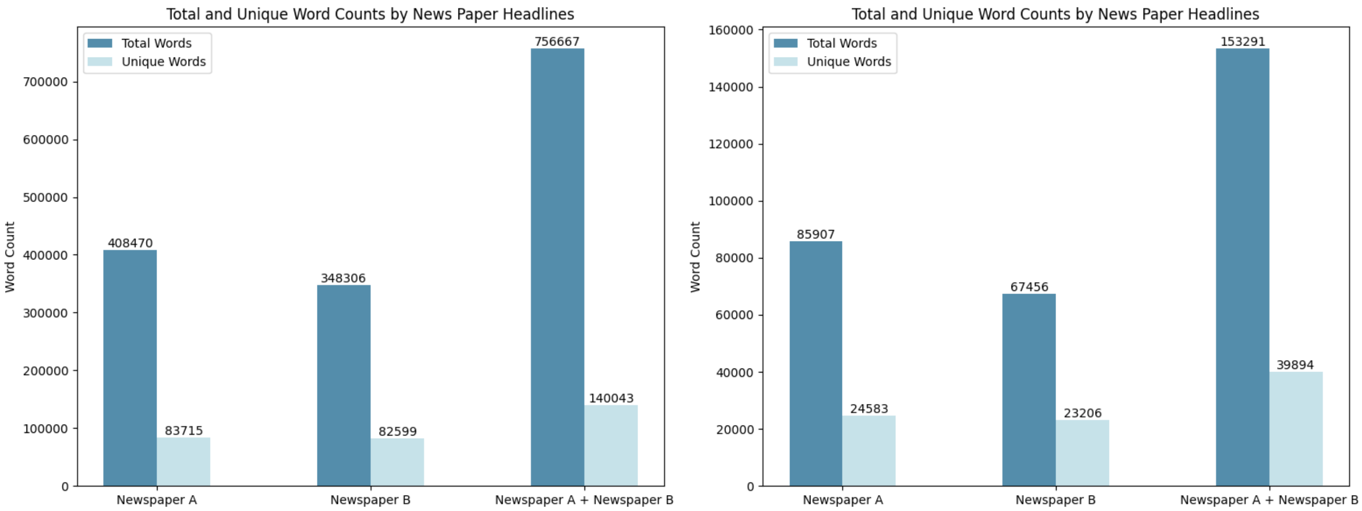
[2]Configured with $temperature = 0.2$ and $max\_tokens = 4$

Since our objective is to uncover stereotypes contained in the headlines, we selected 8 different topics that are more likely to be subjected to a stereotyped narrative related to ethnicity, religion, age, gender, and sexual orientation categories: *"conflict"*, *"crime"*, *"human interest"*, *"immigration"*, *"religion"*, *"society"*, *"terrorism"*, and *"tragedy/disaster"*.

We define this subset as the *filtered dataset*, which contains a balanced number of headlines between the two outlets, as shown in Table I. The topic distribution of the filtered dataset is shown in Figure 2b, showing that the different topic distribution is also reflected in the filtered dataset: NEWSPAPER A features more coverage of *conflict*, whereas NEWSPAPER B focuses more on *crime*. The remaining topics show a relatively balanced representation between the two outlets.

C. Lexicometric Analysis

We performed a series of analyses to explore the composition of our dataset using NLP techniques. Given that our study



(a) Complete CHANGE-IT Dataset

(b) Filtered CHANGE-IT Dataset

Fig. 3: Number of total words and unique words in each set in complete and filtered dataset

TABLE I: Number of news for each dataset version

Set	#Newspaper A	#Newspaper B	#Total News
Complete Set	66,200 (50.3%)	65,491 (49.7%)	131,691
Filtered Set	12,507 (49.3%)	12,876 (50.7%)	25,383

TABLE II: Average number of words for each headline

Set	Newspaper A	Newspaper B	Overall
Complete Set	11.37	9.60	10.48
Filtered Set	11.91	9.84	10.89

focuses on headlines, examining their word frequency provides valuable insights into the semantical differences conveyed between the newspapers. Figure 3 shows the total and distinct number of words for each newspaper separately and overall on the original (3a) and filtered (3b) datasets. Headlines from Newspaper A contain 60,000 more words than those from Newspaper B, a difference that persists in the filtered dataset. This indicates that Newspaper A tends to use longer headlines. However, the distinct word counts are similar for both newspapers and across both datasets, suggesting a comparable variety of words used in their headlines. Finally, Table II reports the average word count per headline, revealing that, on average, Newspaper A headlines are two words longer than those of Newspaper B.

After this first numerical analysis, we analyzed word frequency across both the entire dataset and a filtered subset, identifying dominant terms and themes. Next, we extended our analysis to include bigrams, which revealed relationships between word pairs and provided deeper insights into common phrases within the texts. Finally, we compiled a list of specific terms for examination and conducted an analysis within the filtered dataset. By comparing these findings to those from the complete dataset, we aimed to determine if certain words were preferentially used over others. This comparison would help us understand variations in language use across different contexts and outlets. The results of this analysis will be presented in Section IV-A.

D. Training the Newspaper Classifier

To uncover the different patterns in the two newspaper headlines, we trained a binary classifier to predict the newspaper given the headline. To this end, we fine-tuned a BERT [12] model pre-trained on the Italian language [27].[3] We fine-tuned the model over 10 epochs with a $batch\_size = 1024$, using the filtered CHANGE-IT dataset, which was obtained through LLaMa-based classification. The dataset was split into training and testing sets with an $80/20$ ratio, obtaining promising results that we will discuss later in Section IV-B.

In this way, we obtained a model capable of accurately classifying and distinguishing between headlines from Newspaper A and Newspaper B. The objective of training this classifier was not to predict the source of future headlines but rather to learn the specific word patterns to distinguish the two outlets. This approach opens up the possibility of exploring the model's decision-making process, allowing us to gain deeper insights into the distinctive patterns associated with each outlet through models of explainable AI.

E. Explainability

Explainable AI (XAI) in the context of text analysis focuses on making deep learning models transparent and interpretable [28–30]. XAI methods in text processing, such as [13, 31], aim to reveal underlying patterns, such as why certain words, phrases, or structures influence predictions. In

[3]https://huggingface.co/dbmdz/bert-base-italian-xxl-uncased

this study, we use XAI techniques to uncover and analyze stereotypical patterns associated with each outlet by identifying the patterns learned by the classifier to distinguish between the headlines produced by the two outlets. These explanations offer insights into which linguistic elements are driving the model's classification, thereby allowing us to critically evaluate the model's decision-making system, and deepening our understanding of the narrative differences that characterize each outlet, as we will discuss in Section IV-B.

We selected SHAP *(SHapley Additive exPlanations)* [13] as the XAI method to interpret the model's decisions. SHAP provides a unified measure of feature importance, allowing us to understand the contribution of individual words, sentences, or other textual features in driving model predictions. Specifically, we used SHAP to identify the words that the BERT classifier relies on to differentiate headlines between the two newspapers. By assigning Shapley values [32] to features in each prediction, SHAP allows us to determine which terms most influence the classification of headlines as originating from either NEWSPAPER A or NEWSPAPER B. We first used SHAP to produce local explanations, highlighting how much each word influenced individual predictions. Subsequently, we determined the overall importance of words by calculating the average importance of each word across predictions, by following the idea of some previous work that aggregate local explanations to provide more general insights about the model [31, 33, 34]. These aggregations provide a more general view of the most influential terms in the classification process.

F. Word categorization

To identify words more likely to be associated with stereotypes, we also categorized whether each word extracted from the filtered is related to protected attributes. Inspired by previous work that showed that LLMs can outperform humans in annotating words related to protected attributes [35], we prompted LLaMa 3.1 8B to classify each word into one of the following eight categories:

- *"ethnicity"*, *"crime"*, *"age"*, *"gender and sex"*, *"religion"*, *"migration"*, *"other"*.

We selected *ethnicity*, *age*, *gender and sex*, and *religion* as categories, as these represent four protected individual attributes. Additionally, we included categories for *crime* and *migration*, as these attributes show a strong correlation with ethnicity within our dataset. This process generates a categorized list of relevant words, allowing us to better understand the dataset's composition and enabling more effective visualization of the distribution across these categories.

IV. RESULTS

In this section, we present the main findings we obtained from the lexicometric analyses (Section IV-A) and by explaining the patterns learned by the BERT classifier (Section IV-B).

A. Lexicometric Analysis

The first analysis of the dataset focused on examining the length of the headlines across outlets. As shown in Table II, the data reveals a notable difference in headline length: headlines from the NEWSPAPER A are, on average, two words longer than those from NEWSPAPER B counterpart. This finding highlights a stylistic difference between the two outlets, reflecting varying editorial strategies between them. The NEWSPAPER A publication potentially tries to favor more detailed and nuanced headlines, while the NEWSPAPER B source opts for conciseness and a more impactful style.

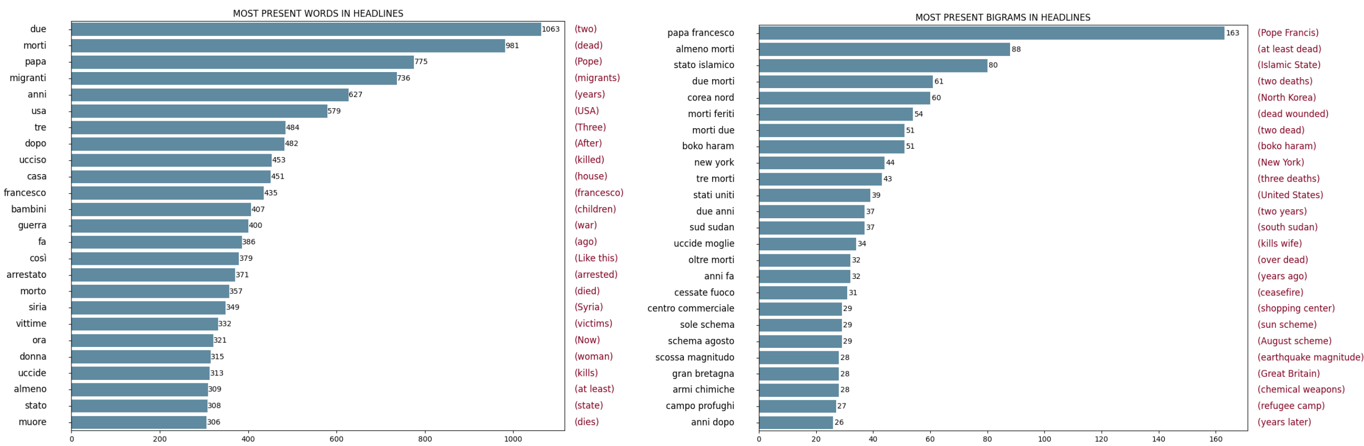
The second analysis is focused on the occurrences and frequencies of words and bigrams within the headlines of the two outlets. Figures 4a and 4b illustrate the results of this analysis on the filtered dataset. The most common words reflect the categories used to filter the dataset. For instance, the Italian word *"morti"* (EN: *"deaths"*) likely pertains to the *conflict* category, while *"Papa"* (EN: *"Pope"*) is associated with the *religion* category. However, while some words, such as *"migranti"* (EN: *"migrants"*), may be relevant, many of these frequent terms—except for a few—do not offer significant insights for detecting stereotypes in media headlines. In the following, we will focus our analysis on words related to ethnicity and migration.

Ethnicity To enhance the accuracy of our analysis, we focused on identifying words strongly associated with stereotypical patterns, particularly those relevant to migration flows of the past two decades. Like other developed nations, Italy has experienced a rise in anti-immigrant sentiment, as documented in studies such as [36] and [37]. Using data from the Italian National Institute of Statistics (ISTAT) on the most common citizenships among foreigners residing in Italy [38], we compiled a list of these groups. This approach enabled us to analyze the prevalence of these terms in media sources and to assess whether their usage differs between NEWSPAPER A and NEWSPAPER B outlets (see Fig. 5). To obtain these results, we consolidated all word forms related to a specific nationality into a single term. For example, variations such as "*nigeriano*" (masc. sing.), "*nigeriana*" (fem. sing.), "*nigeriani*" (masc. plur.), and "*nigeriane*" (fem. plur.) were all grouped under the English term "*Nigerian*".

The dataset reveals a noticeable emphasis on the nationality of individuals mentioned in the headlines, particularly through the frequent use of terms like "*Nigeriano*" (EN: "*Nigerian*"), which appears more than 40 times or "*Marocchino*" (EN: "*Moroccan*"). Although NEWSPAPER B seems to place greater emphasis on specifying the nationality of individuals involved in crimes, this pattern is also evident in the coverage by NEWSPAPER A.

For example, headlines such as (1) *"Nigeriano spintona poliziotti, era pieno di droga"* (EN: *"Nigerian pushes police officers, he was full of drugs"*) and (2) *"Un nigeriano ha ferito tre passanti a Napoli"* (EN: *"A Nigerian has injured three passers-by in Naples"*) illustrate this trend.

However, It is possible to find many headlines that respect this trend for every citizenship reported in Figure 5. Other examples are (3) *"Ladro albanese cade dalla finestra durante il furto: è grave"* (EN: *"Albanian thief falls from the window during the robbery: he is in serious condition"*), and



(a) Most common words

(b) Most common bigrams

Fig. 4: Most common words (a) and bigrams (b) in the filtered dataset

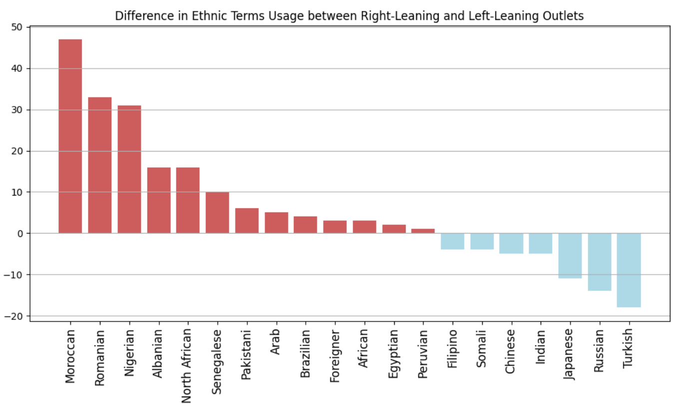


Fig. 5: **Distribution of terms referring to ethnic categories.** The plot illustrates differences in term usage between the two media outlets. Positive values represent terms more frequently used by NEWSPAPER B, while negative values indicate terms more commonly found in NEWSPAPER A.

(4) *"Firenze, marocchino prima molesta passanti poi picchia 2 poliziotti"* (EN: *"Florence, Moroccan first harasses passers-by then beats 2 police officers"*).

Using the list of words classified by LLaMa under the category *ethnicity*, it is insightful to visualize the correlation (Figure 6) between each word and the various topics in the filtered dataset. This visualization highlights the contrasting associations of these words, particularly how terms like *"Arabi"* (EN: *"Arabs"*) and *"Nigeriano"* (EN: *"Nigerian"*) are strongly correlated with the *Crime* topic across both newspapers. Additionally, it is noteworthy how NEWSPAPER A associates *Crime* with words referring to specific regions of Italy, such as *"nuorese"* (EN: *"a person from Nuoro"*), further illustrating the nuances of regional and ethnic associations in media coverage.

We also found that the correlations between *Crime* and *Conflict* are nearly opposite. This observation is significant because ethnic terms are primarily associated with these two topics in our dataset. However, it also indicates that ethnic groups involved in conflicts that gain visibility in Italy are less frequently linked to *Crime* news.

This pattern suggests a broader tendency to highlight the nationalities of individuals involved in criminal activities. Such an approach could reinforce a narrative of an "inside group" versus an "outside group," contributing to the perpetuation of stereotypes and reinforcing anti-immigrant sentiment.

Migration A similar pattern emerges when we examine terms associated with migratory flows to Italy, for instance, words like *"migrante"* (EN: *"migrant"*) are strongly correlated with headlines in the *Crime* category, in both outlets.

The most interesting observation is how the words *"migrante"* (EN: *"migrants"*) and *"immigrato"* (EN: *"immigrants"*) are distributed among the outlets (Figure 7).

The term *"migrant"* is broadly understood as someone who moves from their usual residence, either within their country or across borders, for various reasons. This term includes legally recognized categories, such as migrant workers, people involved in legally-defined movements like smuggled migrants, and those whose movement isn't defined under international law, such as international students, according to the International Organization for Migration (IOM) [39].

The term *"immigrant"* is often perceived as neutral, but an etymological analysis reveals that it may carry additional connotations. While the Treccani dictionary [40] defines it broadly as someone who relocates to another country, in Italy and much of the West, the term is frequently associated with irregular status, poverty, lack of education, and criminality.

In Italy, it is common to use the term *"clandestino"* to identify irregular migrants in a pejorative way, because it assumes the meaning of illegal. This term is not a legal one, and its use to broadly describe migrants could be associated with an unfair assumption that all migrants in Italy are irregular.

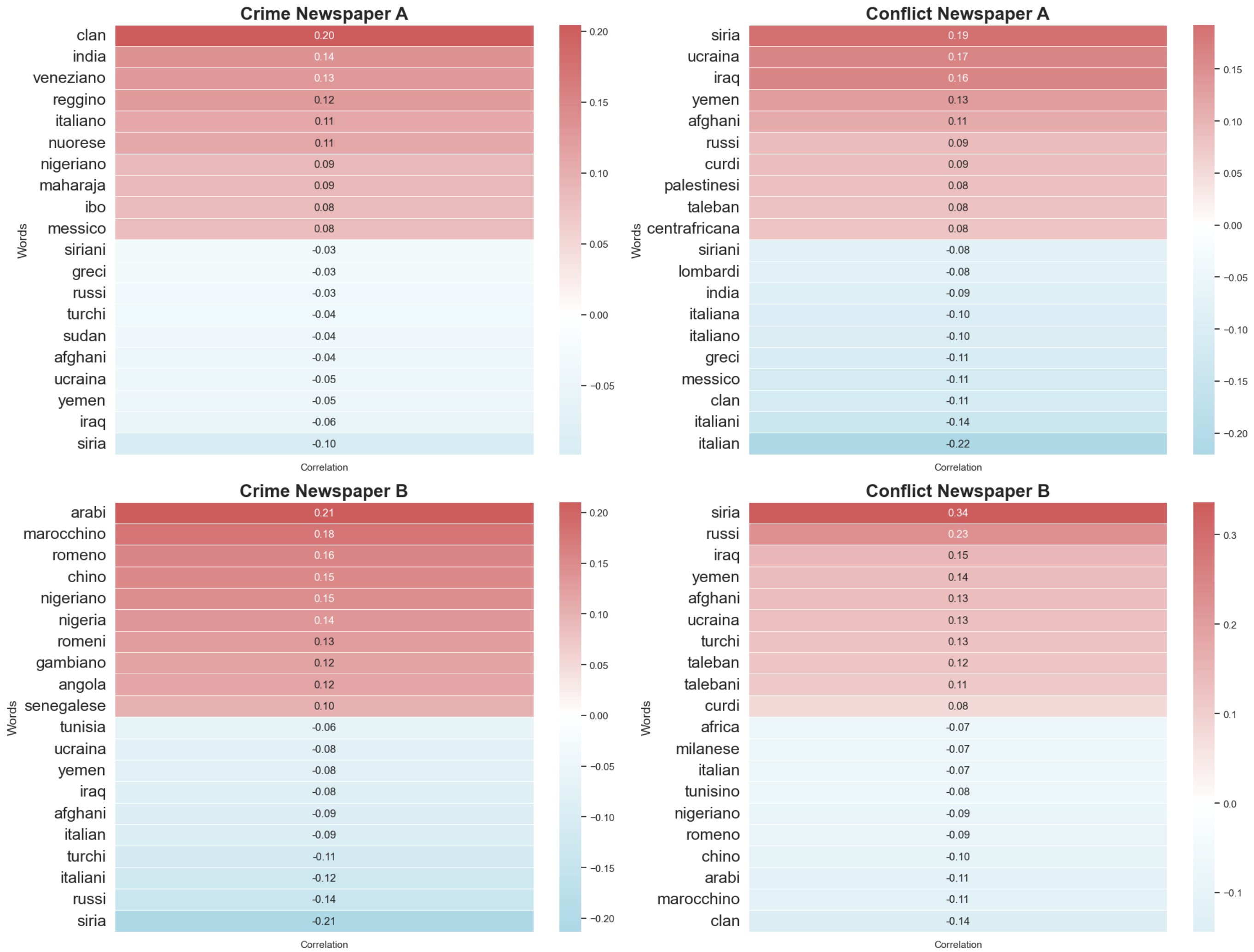


Fig. 6: **Correlations between words classified as ethnic and topic of headlines for the two outlets**. The most and least 10 correlated words for each topic and newspaper are reported. Red values represent positive correlations, indicating stronger associations between words and topics, while blue values represent negative correlations, showing weaker or inverse associations.

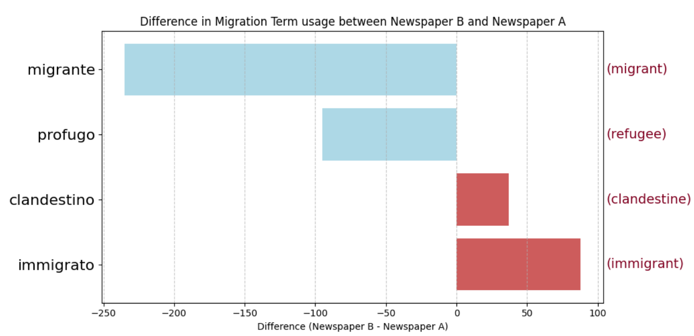


Fig. 7: **Comparison of migrant vs immigrant Usage.** The differing frequency and context in which two prominent outlets, use the terms 'migrant' and 'immigrant.' The graph highlights the varying linguistic choices.

B. SHAP explanations

The BERT classifier described in Section III-D exhibits high predictive capabilities, achieving an F1 score of 0.84 in classifying the newspaper given the headline. This performance demonstrates the model's effectiveness in learning the patterns of words that distinguish the headlines of the two outlets. As a result, we can confidently use it as a proxy for those patterns.

Leveraging SHAP [13], we can quantify the contribution of each token to the classifier's predictions for distinguishing between the two outlets. We identified the words that contributed the most in distinguishing between the two outlets for the 8 topics filtered. However, we report here for discussion the results on the topic of migration due to the significant and insightful results it yielded. Specifically, we began with the migration-related keywords identified earlier with the LLM. Figure 8 illustrates the average impact of these words on the classifier's predictions, highlighting their influence on the

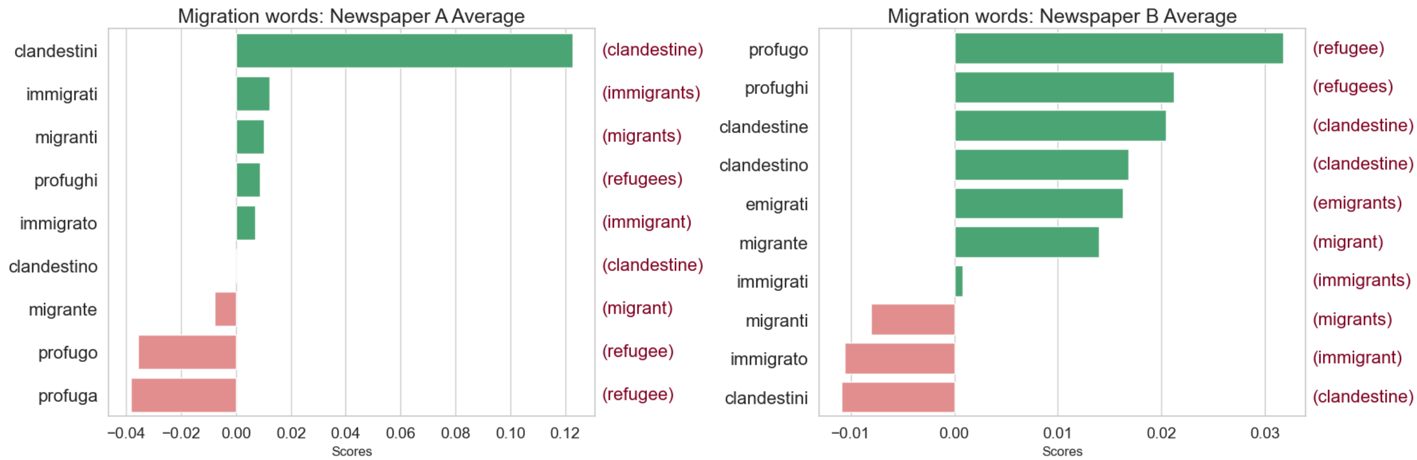


Fig. 8: **SHAP Values comparison on migration terms**. Influence scores of migration-related terms between the two outlets identified using SHAP. Positive values (green bars) indicate words that positively contribute to the classification of the outlet, while negative values (red bars) indicate words with a negative impact. English translations are reported in round brackets.

model's decision-making process.

We found that, interestingly, the word "*immigrati*" (EN: "*immigrants*") plays a stronger role in classifying articles from NEWSPAPER A than from NEWSPAPER B, contrary to initial inferences made during vocabulary analysis.

This observation supports the idea that similar patterns in word usage to describe specific groups appear consistently across the dataset, with limited distinction between the two subsets. Additionally, terms such as "*clandestino*" (EN: "*clandestine*") and "*profugo*" (EN: "*refugee*") are associated with both sets, reinforcing the lack of differentiation.

The SHAP values associated with terms LLaMa categorized as *ethnic* confirm these findings in the context of immigration-related language. Words like "*Nigeriano*" (EN: "*Nigerian*"), "*Marocchino*" (EN: "*Moroccan*"), and "*Albanese*" (EN: "*Albanian*") appear in both subsets, revealing once again that, despite some differences, both sets reflect similar patterns in the treatment of ethnic terms.

V. CONCLUSION

A. Discussion

The stereotypical patterns discussed in the previous sections appear across both NEWSPAPER A and NEWSPAPER B, highlighting that the issue of representation is not confined to a single perspective but is pervasive across the media landscape.

The variation in headline length between NEWSPAPER A and NEWSPAPER B suggests different editorial approaches. NEWSPAPER A outlets tend to adopt a more detailed style, possibly reflecting a broader focus on context and complexity, while NEWSPAPER B prioritizes brevity and directness.

A key observation in the analysis is the disproportionate frequency of ethnic terms such as "*Nigerian*" in NEWSPAPER B headlines, especially in the context of crime. This frequent association of ethnicity with criminality may contribute to the reinforcement of stereotypes. However, this trend is not absent in NEWSPAPER A, although these references are less frequent. Both outlets, despite their different editorial focuses, reflect a tendency to use ethnicity as a marker when discussing crime-related topics, reflecting a tendency to categorize individuals based on their national origin or legal status.

Our findings emphasize the need for journalism to be more responsible and aware of the significant influence that the used language can have in reinforcing biases, because this categorization can contribute to the perpetuation of stereotypes, ignoring the multifaceted nature of migration and integration, and with the risk of promoting and normalizing a potentially dangerous "us versus them" narrative. Moreover, building on prior research on the ineffectiveness of clickbait underscores the essential role of ethical journalism in delivering clear, unbiased news, our findings also emphasize the need for inclusive language to ensure fair representation [41]. We claim that AI-based applications could play a crucial role in supporting journalists to produce more inclusive and less stereotype-driven news, as demonstrated in other contexts [42]. Collectively, these insights reveal the prevalence of potentially biased reporting and highlight the pressing need for responsible journalism, which serves as a cornerstone for balanced political and social discourse in society.

B. Future Work

In future work, we would like to extend our analysis of stereotypical patterns in Italian media across a larger number of news newspapers and protected characteristics, such as age and gender. Secondly, we would like to investigate how these patterns fluctuate and change over time in response to various historical events and societal shifts. Finally, we would like to leverage the capabilities of modern LLMs to develop a comprehensive framework to assist humans in detecting and removing stereotypes in multi-modal media content, encompassing not only text but also images and videos.

Acknowledgment

This study was carried out within the project "E-MIMIC: Empowering Multilingual Inclusive Communication" (Nr. 2022WEFCFP), funded by the Ministero dell'Universitá e della Ricerca - with the PRIN 2022 (D.D. 104 - 02/02/2022) program.

References

1. Blom, J. N. & Hansen, K. R. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics* **76,** 87–100. ISSN: 0378-2166. https://www.sciencedirect.com/science/article/pii/S0378216614002410 (2015).
2. Törnberg, P. How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences* **119,** e2207159119. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.2207159119. https://www.pnas.org/doi/abs/10.1073/pnas.2207159119 (2022).
3. Suk, J., Coppini, D., Muñiz, C. & Rojas, H. The more you know, the less you like: A comparative study of how news and political conversation shape political knowledge and affective polarization. *Communication and the Public* **7,** 40–56 (2022).
4. Van Dijk, T. A. *Elite Discourse and Racism* (Sage Publications, London, 1993).
5. Hacker, K. L. Racism and The Press. *Discourse Society* **3,** 378–383. ISSN: 09579265, 14603624. http://www.jstor.org/stable/42887807 (2024) (1992).
6. Van Dijk, T. A. Principles of Critical Discourse Analysis. *Discourse & Society* **4,** 249–283 (1993).
7. Julia M. Wondolleck, B. G. & Bryan, T. Us versus Them: How Identities and Characterizations Influence Conflict. *Environmental Practice* **5,** 207–213. eprint: https://doi.org/10.1017/S1466046603035592. https://doi.org/10.1017/S1466046603035592 (2003).
8. Hayes, D. & Guardino, M. in *Influence from Abroad: Foreign Voices, the Media, and U.S. Public Opinion* 17–50 (Cambridge University Press, 2013).
9. Touvron, H. *et al. LLaMA: Open and Efficient Foundation Language Models* 2023. arXiv: 2302.13971 [cs.CL]. https://arxiv.org/abs/2302.13971.
10. European Language Grid. *Corpus 7373* Accessed: 2024-10-02. 2023. https://live.european-language-grid.eu/catalogue/corpus/7373.
11. De Mattei, L., Cafagna, M., Dell'Orletta, F., Nissim, M. & Gatt, A. *CHANGE-IT @ EVALITA 2020: Change Headlines, Adapt News, GEnerate* in *Proceedings of Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020)* (eds Basile, V., Croce, D., Di Maro, M. & Passaro, L. C.) (CEUR.org, Online, 2020).
12. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. *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)* (Association for Computational Linguistics, Minneapolis, Minnesota, June 2019), 4171–4186.
13. Lundberg, S. M. & Lee, S. A unified approach to interpreting model predictions. *CoRR* **abs/1705.07874.** arXiv: 1705.07874. http://arxiv.org/abs/1705.07874 (2017).
14. Atle Haugsgjerd, R. K. & Steen-Johnsen, K. Uninformed or Misinformed in the Digital News Environment? How Social Media News Use Affects Two Dimensions of Political Knowledge. *Political Communication* **40,** 700–718. eprint: https://doi.org/10.1080/10584609.2023.2222070. https://doi.org/10.1080/10584609.2023.2222070 (2023).
15. Bazaco, A., Redondo, M. & Sánchez-García, P. *Clickbait as a strategy of viral journalism: conceptualisation and methods* in (2019). https://api.semanticscholar.org/CorpusID:150129578.
16. Rony, M. M. U., Hassan, N. & Yousuf, M. *Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects?* in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (Association for Computing Machinery, Sydney, Australia, 2017), 232–239. ISBN: 9781450349932. https://doi.org/10.1145/3110025.3110054.
17. Chakraborty, A., Paranjape, B., Kakarla, S. & Ganguly, N. *Stop Clickbait: Detecting and preventing clickbaits in online news media* in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2016), 9–16.
18. Jung, A. K., Stieglitz, S., Kissmer, T., Mirbabaie, M. & Kroll, T. Click me...! The influence of clickbait on user engagement in social media and the role of digital nudging. *PLOS ONE* **17,** e0266743 (June 2022).
19. D. Molina, M. *et al. Does Clickbait Actually Attract More Clicks? Three Clickbait Studies You Must Read* in (Association for Computing Machinery, Yokohama, Japan, 2021). ISBN: 9781450380966. https://doi.org/10.1145/3411764.3445753.
20. Barthes, R. *S/Z* Translated by Richard Miller in 1974 (Éditions du Seuil, Paris, 1970).
21. Gefen, D., Fresneda, J. E. & Larsen, K. R. Trust and Distrust as Artifacts of Language: A Latent Semantic Approach to Studying Their Linguistic Correlates. *Frontiers in Psychology* **11.** ISSN: 1664-1078. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.00561 (2020).
22. Coats, J. & Moe, K. Choose Your Words Wisely: The Role of Language in Media Trust. *University of Florida College of Journalism and Communications.* Originally

published in the November 16 issue of InContext, a publication by Digital Content Next. https://www.jou.ufl.edu/insights/choose-your-words-wisely-the-role-of-language-in-media-trust/ (2023).
23. Sorato, D., Zavala-Rojas, D. & del Carme Colominas Ventura, M. *Using Word Embeddings to Quantify Ethnic Stereotypes in 12 years of Spanish News* in *Proceedings of the 19th Annual Workshop of the Australasian Language Technology Association* (eds Rahimi, A., Lane, W. & Zuccon, G.) (Australasian Language Technology Association, Online, Dec. 2021), 34–46. https://aclanthology.org/2021.alta-1.4.
24. Thijs, G., Trilling, D. & Kroon, A. C. *Contextualized Word Embeddings Expose Ethnic Biases in News* in *Proceedings of the 16th ACM Web Science Conference* (Association for Computing Machinery, Stuttgart, Germany, 2024), 290–295. ISBN: 9798400703348. https://doi.org/10.1145/3614419.3643994.
25. Kroon, A. C., Trilling, D. & Raats, T. Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage. *Journalism & Mass Communication Quarterly* **98,** 451–477 (2021).
26. Müller, P., Chan, C.-H., Ludwig, K., Freudenthaler, R. & Wessler, H. Differential Racism in the News: Using Semi-Supervised Machine Learning to Distinguish Explicit and Implicit Stigmatization of Ethnic and Religious Groups in Journalistic Discourse. *Political Communication* **40,** 396–414 (2023).
27. Schweter, S. *Italian BERT-Electra models* Accessed: 2024-11-03. 2020. https://github.com/stefan-it/italian-bertelectra.
28. Qian, K. *et al. XNLP: A Living Survey for XAI Research in Natural Language Processing* in *26th International Conference on Intelligent User Interfaces - Companion* (Association for Computing Machinery, College Station, TX, USA, 2021), 78–80. ISBN: 9781450380188. https://doi.org/10.1145/3397482.3450728.
29. Danilevsky, M. *et al. A Survey of the State of Explainable AI for Natural Language Processing* in (Association for Computational Linguistics, Suzhou, China, Dec. 2020), 447–459. https://aclanthology.org/2020.aacl-main.46.
30. Guidotti, R. *et al.* A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* **51,** 93:1–93:42. ISSN: 0360-0300 (Aug. 2018).
31. Ventura, F., Greco, S., Apiletti, D. & Cerquitelli, T. Trusting deep learning natural-language models via local and global explanations. *Knowl. Inf. Syst.* **64,** 1863–1907. ISSN: 0219-1377. https://doi.org/10.1007/s10115-022-01690-9 (July 2022).
32. Shapley, L. S. A value for n-person games. *Contributions to the Theory of Games* **2,** 307–317 (1953).
33. Accelerating the Global Aggregation of Local Explanations. **38,** 18807–18814. https://ojs.aaai.org/index.php/AAAI/article/view/29845 (Mar. 2024).
34. Ventura, F., Greco, S., Apiletti, D. & Cerquitelli, T. Explaining deep convolutional models by measuring the influence of interpretable features in image classification. *Data Mining and Knowledge Discovery,* 1–58 (2023).
35. Greco, S., Zhou, K., Capra, L., Cerquitelli, T. & Quercia, D. NLPGuard: A Framework for Mitigating the Use of Protected Attributes by NLP Classifiers. *Proc. ACM Hum.-Comput. Interact.* **8.** https://doi.org/10.1145/3686924 (Nov. 2024).
36. Nese, A. Migrations in Italy and Perceptions of Ethnic Threat. *International Migration & Integration* **24.** Issue Date: September 2023, 939–968. https://doi.org/10.1007/s12134-022-00985-8 (2023).
37. Papademetriou, D. G. & Banulescu-Bogdan, N. *Understanding and Addressing Public Anxiety About Immigration* Council Statement. Transatlantic Council on Migration - Council Statement, Rome, 2015 (Migration Policy Institute, Washington, DC, 2016), 1–23.
38. Italian National Institute of Statistics (ISTAT). *International Migration Data* Available online at ISTAT Data Portal. Accessed 2024. http://dati.istat.it/Index.aspx?QueryId=19675&lang=en.
39. International Organization for Migration. *About Migration* Accessed: 2024-11-05. 2024. https://www.iom.int/about-migration.
40. Treccani. *Immigrato* Accessed: 2024-11-06. n.d. https://www.treccani.it/vocabolario/immigrato/.
41. Attanasio, G. *et al. E-MIMIC: Empowering Multilingual Inclusive Communication* in *2021 IEEE International Conference on Big Data (Big Data)* (2021), 4227–4234.
42. La Quatra, M., Greco, S., Cagliero, L. & Cerquitelli, T. *Inclusively: An AI-Based Assistant for Inclusive Writing* in *Machine Learning and Knowledge Discovery in Databases: Applied Data Science and Demo Track* (Springer Nature Switzerland, Cham, 2023), 361–365. ISBN: 978-3-031-43430-3.