Linking Stance and Stereotypes About Migrants in Italian Fake News

Alessandra Teresa Cignarella[1,2], Simona Frenda[1,2], Tom Bourgeade[1], Cristina Bosco[1] and Francesca D'Errico[3]

[1]Dipartimento di Informatica, Università di Torino, Turin, Italy
[2]aequa-tech, Turin, Italy
[3]Dipartimento di Formazione, Psicologia, Comunicazione, Università di Bari "Aldo Moro", Italy

Abstract
This paper investigates stance and stereotypes within a dataset of Twitter conversational threads in Italian. The starting point of these conversations are tweets containing misinformation, in the form of racial hoaxes targeted at migrants, identified as untrustworthy by fake news debunking websites. The conversational structure of the dataset gives us the opportunity to observe and collect evidence about some linguistic and social phenomena at play in the propagation of stereotypes and the interactions between users which stem from them. We propose a theoretical background, as well as quantitative and qualitative analyses of our annotated data, at different levels of granularity, which can provide insights into the dynamics of Italian online discourses on the topic of migration.

Keywords
Stance Detection, Stereotypes, Rumors, Fake News, Misinformation, Italian

Warning: *This work contains words and expressions that could be considered vulgar or offensive to varying degrees. We emphasize that all authors of this paper are deeply involved in activities to counter the spread of online hatred and do not condone the use of such expressions in any way.*

1. Introduction and Motivation

In the era of information overload and widespread digital communication, the terms "disinformation", "fake news", "hoaxes", "misinformation", and "rumors" have become buzzwords that dominate public discourse [1]. While these terms are often used interchangeably, it is important to recognize the nuanced differences between them and establish some order in the terminology. "Fake news" refers to fabricated or misleading information presented as legitimate news, often with the intention to deceive or manipulate public opinion. "Disinformation" encompasses a broader range of intentionally false or misleading information disseminated with the aim of influencing beliefs or actions. On the other hand, "rumor" refers to unverified or unsubstantiated pieces of information that circulate widely within communities [2]. Finally, a particular type of rumor that disseminates information including threat claims to the health or safety of a person or group based on their race, ethnicity or religion, has been recently defined as "racial hoax" by Cerase and Santoro [3], reshaping a previous definition from Russell [4]. In this paper, we aim at stressing a connection with the area of research that investigates the above-mentioned phenomena and the research conducted so far in the field of Stance Detection (SD). Indeed, as it can also be seen in Figure 1, in a recent survey paper, Küçük and Can [5] illustrate the relationships occurring among the different tasks and subtasks in the field of Sentiment Analysis, putting at the center of attention SD. In the boxes highlighted in blue in Figure 1 we see the tasks of *Rumor Stance Classification* and *Fake News Stance Classification* being strictly related to SD. In this work, we connect the dimensions of stance and stereotypes, based on a re-annotation of our previous corpus, in particular studying the conversational structure of the data.

The paper is organized as follows. In Section 2 we briefly survey the related work on Fake News and Stance Detection on one side, and Racial Hoaxes and Stereotypes on the other side. In Section 3 we describe the corpus collection, the annotation process concerning especially the dimension of stance. In Section 4 we provide a corpus-based analysis taking into consideration also the dimension of stereotype, and we show some examples from the corpus. Finally, in Section 5 we discuss some insights gained by the observation of the annotated phenomena, and we conclude the paper with final remarks and possible ideas for future research.

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 — Dec 02, 2023, Venice, Italy
alessandrateresa.cignarella@unito.it (A. T. Cignarella); simona.frenda@unito.it (S. Frenda); tom.bourgeade@unito.it (T. Bourgeade); cristina.bosco@unito.it (C. Bosco); francesca.derrico@uniba.it (F. D'Errico)

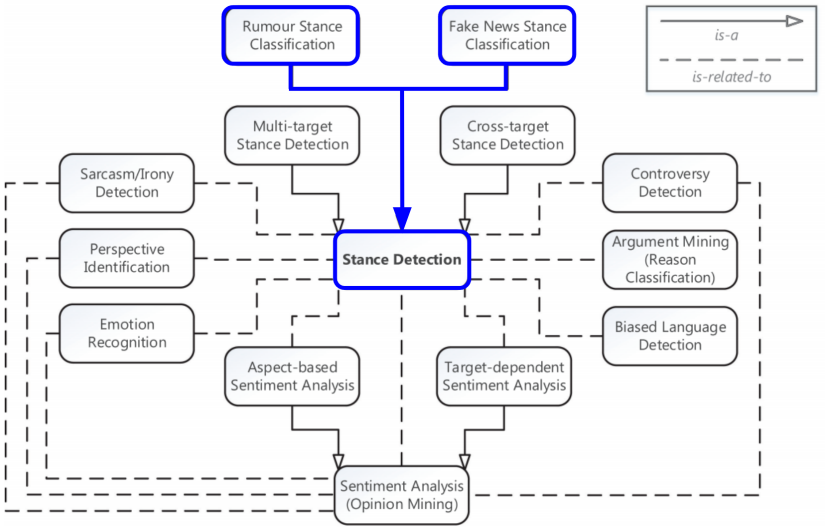

Figure 1: NLP detection tasks and subtasks related to Stance Detection and Sentiment Analysis. Adapted from Küçük and Can [5].

2. Related Work

2.1. Stance Detection, Fake News and Rumors

In SemEval 2016 [6] introduced the first shared task in the domain of stance detection establishing for the first time a formal framework for target-specific stance classification, with admissible labels: *Against*, *Neutral* and *Favor*. The task of stance detection has also proven valuable in distinguishing misinformation from genuine stories. Within the Fake News Challenge participants had to classify the stance towards a claim made in a news headline [7]. By categorizing headlines and news bodies as *Agrees*, *Disagrees*, *Discusses* (a given topic), or *Unrelated*, researchers aimed to identify and combat the spread of fake news more effectively.

In a more general context, researchers started to study the development of stance in online conversational contexts and have employed a slightly different annotation scheme, to classify attitudes toward rumors or broader topics: *Support*, *Deny*, *Query*, *Comment*, often represented as SDQC. This categorization has provided a versatile approach especially classifying tweets belonging to the same conversational thread.

Aker et al. [8] proposed the four labels described above, for the first time at a SemEval shared task: *RumorEval 2017*, which provided a standardized framework for evaluating rumor detection techniques and assessing their effectiveness. The same setting was also proposed in a second edition [9] by introducing additional exercises, such as stance prediction and veracity prediction. This allowed for a more comprehensive evaluation of rumor detection systems, focusing on not only identifying rumors, but also understanding their stance.

Finally, in the context of the Italian language, the *SardiStance* shared task was introduced in EVALITA 2020[1], offering a pioneering challenge for Italian stance detection [10]. As far as the conversational dimension is concerned, Stranisci et al. [11] presented the MoralConvITA corpus, in which moral values and conversational relations linking the components of pairs of messages are annotated with similar categories: *Attack*, *Support* or *Same topic*.

2.2. Racial Hoaxes and Stereotypes

In Bosco et al. [12] we conducted an insightful investigation into the presence of Italian stereotypes on Facebook using a combination of psychology and natural language processing frameworks. We delved into the dynamics of racial stereotyping by extracting replies and comments written below a controversial post written by the famous Italian singer Gianni Morandi, where he compared nowadays migrants in the Mediterranean Sea to Italians immigrating to the USA in the 1920s. We explored how these stereotypes manifest and spread, providing valuable insights into the prevalence and impact of Italian stereotypes in online spaces.

Similarly, D'Errico et al. [13] examines stereotypes and prejudices that arise from racial hoaxes using a psycholinguistic analysis approach. The study investigates in-

[1] http://www.di.unito.it/~tutreeb/sardistance-evalita2020/index.html

stances where false information or hoaxes related to immigrants are spread, leading to the reinforcement or creation of negative stereotypes and prejudices.

Building upon this line of research, our team presented our latest paper at the EACL 2023 conference. In the paper titled *"A Multilingual Dataset of Racial Stereotypes in Social Media Conversational Threads"* we introduced a novel multilingual dataset called MULTI-STEREOHOAX [14]. Our study aimed at studying racial hoaxes and stereotypes in three different languages: Italian, Spanish and French. The dataset is labeled with a complex annotation scheme, based on the Stereotype Content Model (SCM) proposed by Fiske et al. [15]. It is a theoretical framework that provides a psychological understanding of how stereotypes are formed, maintained, and applied in social contexts.

These studies provide crucial insights into the origins, dissemination, and potential consequences of stereotypes, paving the way for future efforts to mitigate their harmful effects and promote a more inclusive online environment. In this study, we attempt to bridge the gap between these two areas of research. Specifically we extracted the Italian portion of the dataset (STEREOHOAX-IT), which was created for the study of racial hoaxes and stereotypes, and we further annotated it with stance information, enabling a more comprehensive analysis of the propagation and impact of racial stereotypes in Italian online discourse.

3. Describing the Corpus

STEREOHOAX-IT [14] is the Italian subset of a corpus of conversations collected on Twitter originated from hoaxes targeting migrants. We started from an initial list of hoaxes deemed *racial* as they tend to explicitly or implicitly attack immigrants, inciting to adopt a contestant stance to the phenomenon of immigration. This initial list was created by consulting debunking websites (`bufale.net`[2] and BUTAC[3]).

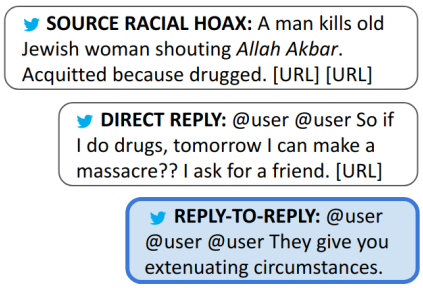

Figure 2: Example of conversational thread.

We were able to collect 273 conversations that discuss these racial hoaxes. The dataset is composed of a total of 2,850 tweets of which 597 are direct replies to the tweets that mention the racial hoax, and 2,253 are replies-to-replies, that is, replies to direct replies. Therefore, the corpus preserves the conversational structure of the Twitter threads, allowing a better analysis of the relations between these conversations' participants. An example of a conversational thread is reported in Figure 2.

In this work, we are interested in studying the stance expressed in the messages of the conversations towards the veracity of the hoax. Considering the purpose of our analysis, we chose to adopt the SDQC schema of annotation adding a label called "*Head*" to identify the texts that spread the hoax or start the conversational thread (identified in Figure 2 as "Source Racial Hoax"). Inspired by Aker et al. [8], we conceived the schema as follows:

- **H** (*Head*): the tweet contains the racial hoax at the root of the conversation;
- **S** (*Support*): the author of the message supports the veracity of the hoax;
- **D** (*Deny*): the author of the message denies the veracity of the hoax;
- **Q** (*Query*): the author of the message asks for additional evidence in relation to the veracity of the hoax;
- **C** (*Comment*): the author of the message makes their own comment without a clear contribution to assess the veracity of the hoax.

As an example, in Figure 3 a source racial hoax, i.e., the *Head* of a Twitter conversation, and four replies (one per SDQC label):

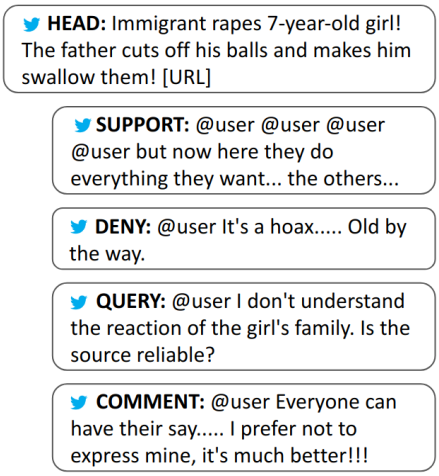

Figure 3: Example of Head with four replies showing different labels.

[2] https://www.bufale.net/
[3] https://www.butac.it/

Differently from the standard schema used by Mohammad et al. [16] where annotators determine if the author of the message is in *favor*/*against*/*neutral* towards a specific phenomenon, the SDQC+H gives us the possibility to identify, more precisely, the attitude of the author with respect to the hoax that targets immigrants. This annotation was applied to direct replies and replies-to-replies toward the hoax declared in the *Head* (or "Source Racial Hoax" in Figure 2 as defined in Bourgeade et al. [14]).

3.1. Enriching the Corpus with Stance Labels

Two different annotators, a male and a female Italian native between 25 and 35 years old (one master student in Linguistics and a PhD student in Digital Humanities) have participated in the annotation campaign. They both annotated all the tweets contained in StereoHoax-It and later additionally annotated it for the dimension of stance as described above. The annotation was performed using *Label Studio*[4] – an open-source annotation platform. Annotator 1 (A1) and Annotator 2 (A2), were both assigned 5,255 tweets in total, and they were asked to label them accordingly to the scheme presented in the previous section (i.e., SDQC+H).

Due to the complexity of the task, and to the fact that annotators could *skip* annotating a tweet in case of uncertainty, in this phase, we were able to collect only 3,123 complete annotations. Once the first round of labeling was completed, we performed an inter-annotator agreement test by calculating Cohen's kappa coefficient, which resulted in $\kappa = 0.3318$ (*fair agreement*). The cases in which A1 and A2 provided two different labels were solved by a third experienced female annotator (A3), an Italian native, 25-35 years old post-doc researcher in NLP.

Thanks to this, some tweets with disagreements were adjudicated, thus increasing the size of the gold-labeled data. However, despite this effort, some disagreements remained for some instances, and we refer to them as "*complex cases*". In Table 1 we report the numbers that are the outcome of the annotations and some more details regarding their nature.

n# tweets	details
2,132	skipped tweets / off-topic
449	incomplete annotation from either A1 or A2
202	*complex cases*
2,472	agreement between A1, A2 + A3 (*gold*)
5,255	**total**

Table 1
Number of tweets annotated for stance.

[4] http://labelstud.io/

In the remainder of the paper we provide analyses only focusing on the 2,472 tweets that present agreement between annotators, and leave the study of "complex cases" and incomplete annotations for future versions of the corpus.

4. Analyzing the Corpus

4.1. Annotation Analysis

In this section, we provide quantitative and qualitative analyses regarding annotation of stance. In Figure 4, the bar chart shows the distribution of labels annotated by A1 and A2. We can observe how both annotators had similar judgements when handling users' stance towards migrants. From the same figure, it can also be seen that for both annotators, *Comment* is the predominant label (blue), followed by *Support* (green), *Query* (yellow) and *Deny* (red). The same percentages are respected in the final label distribution of stance calculated over the gold-labeled portion of the dataset, i.e. 2,472 tweets (see Figure 5).

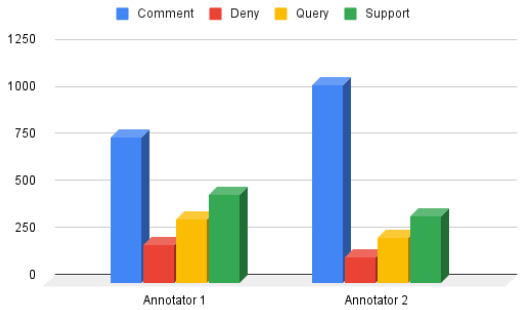

Figure 4: Annotations of A1 and A2.

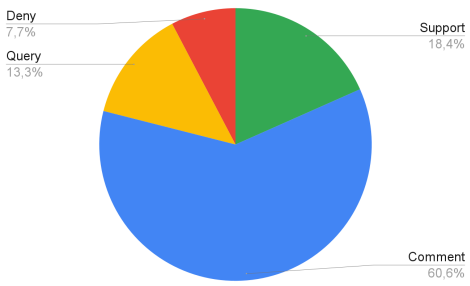

Figure 5: Stance distribution over 2,472 tweets.

Finally, in Table 2 we show a confusion matrix which intersects the newly annotated dimension of stance with the pre-existing annotation of stereotypes. The numbers

reported in this table do not add up to 100% because we removed the percentages relative to the tweets annotated with the label *Head*, since they are not relevant for the analyses.

	stance			
stereo	Comment	Deny	Query	Support
no	63.60%	8.16%	14.42%	12.07%
yes	12.77%	1.28%	1.06%	34.68%
Total	**53.91%**	**6.85%**	**11.88%**	**16.38%**

Table 2
Confusion matrix showing the percentages (%) of Support, Deny, Query and Comment labels with respect to the dimension of Stereotype.

The results do not seem to show a particularly significant co-occurrence of one phenomenon with the other. Although, as expected, the majority of tweets annotated as *Support*, also contain racial stereotypes (34.68%), and the majority of tweets annotated as *Comment* do not contain forms of stereotyping towards migrants (63.60%).

We could have expected a significant portion of the *Deny* label to co-occur with the absence of stereotypes, but the tweets annotated with that label are very sparse (they are only 6.58% in total), therefore it is not sufficient for drawing meaningful conclusions.

4.2. Analysis on the Conversational Structure

In order to evaluate the influence of conversational structure on the distribution of stance labels within our dataset, we measured the "conversation depth" of each individual tweet. Specifically, each *Head* tweet was assigned a depth of 0 (however, these *Head* tweets were not considered for the rest of this analysis). The conversation depth of each subsequent tweet was then determined by calculating the length of the reply-chain leading back to the original *Head* of its conversation. Unfortunately, due to the nature of the phenomenon we are investigating here, numerous tweets (1,947) presented gaps in their respective reply-chains, due to the deletion of content (either by their authors, or by moderation of the microblogging platform). In these cases, we assigned the minimum potential depth value of 2, given that all *Heads* are accounted for in our dataset. Table 3 thus depicts the distribution of varying stance labels according to depth within the dataset.

Although the label distribution across depths largely mirrors the overall dataset distribution, we can observe a higher proportion of *Support* messages in direct replies (depth 1). This might suggest that users who aim to challenge the veracity of a racial hoax might be more inclined to express themselves as replies to replies, rather than directly under the initial posts. To estimate the correlation

	stance			
depth	Comment	Deny	Query	Support
1	88	16	26	103
***2**	1158	134	247	269
3	74	15	18	27
4	9	3	2	3
5	1	1	0	1
Total	**1330**	**169**	**293**	**403**

(a) In number of labels

	stance			
depth	Comment	Deny	Query	Support
1	6.62%	9.47%	8.87%	25.56%
2*	87.07%	79.29%	84.30%	66.75%
3	5.56%	8.88%	6.14%	6.70%
4	0.68%	1.78%	0.68%	0.74%
5	0.08%	0.59%	0.00%	0.25%
Total	**100%**	**100%**	**100%**	**100%**

(b) In percentages of labels

Table 3
Confusion matrices showing the distribution of *Support*, *Deny*, *Query* and *Comment* labels with respect to the depth of each tweet in its conversation. *the minimum depth for conversations with missing links is 2.

of stance labels with depth, we perform a Chi-squared test and compute Cramér's V: we find a Chi^2 value of 130.762, with a p-value of $4.34 \times 10^{-}22$, as well as a V of 0.134, which thus only indicates a small association [17]. To investigate differences among specific conversations, we compute two measures of "controversiality":

1. *Support-Deny Balance* (SD-B) is simply derived from the proportion of *Support* minus *Deny* messages, as a positive or negative percentage of their sum:

$$\frac{\text{count}(Support) - \text{count}(Deny)}{\text{count}(Support) + \text{count}(Deny)}$$

2. `max P-index` implements the measure proposed by Akhtar et al. [18], the *Polarization Index*, where we consider each conversation as an instance with its *Support* and *Deny* replies as annotations. We then iterate over all possible $k = 2$ partitions of these annotations (*proponents* and *opponents*) for each conversation, and find the maximum *P-index* which we report here.

Table 4 presents these measures for the 10 largest conversations (in number of tweets) in the dataset, as well as their percentage of messages containing stereotypes. We only display the top-10 both for space reasons and because further conversations are too small to compute meaningful metrics (starting from the 27th largest conversation the number of messages is 4 or less, and many

Conv #	SD-B	max P-idx	Stereo %	Size
1	14.29%	97.96%	3.45%	898
2	50.48%	74.52%	5.18%	657
3	100.00%	0.00%	39.06%	64
4	64.44%	58.47%	20.00%	60
5	-71.43%	48.98%	7.69%	52
6	55.56%	69.14%	5.00%	40
7	100.00%	0.00%	11.43%	35
8	-33.33%	88.89%	20.59%	34
9	100.00%	0.00%	21.21%	33
10	100.00%	0.00%	20.00%	30

Table 4
Balance (SD-B) of *Support* (100%) vs. *Deny* (-100%), maximum *P-index* [18], and percentage of stereotypes in messages, for the 10 largest conversations in the dataset (in number of tweets).

have no *Support* and *Deny* replies). Conversation #1 can be considered the most "controversial" in this dataset (SD-Balance closest to 0%, largest max P-index), and it also happens to be the largest. The *Head* of this conversation is the following tweet (adapted into English): *"Now Matteo Salvini is in court in Catania for defending the borders, please also tweet #IstandWithSalvini, let's make him feel our affection!"*. As this is a call for support for a controversial figure in Italian politics, this explains the relative balance of *Support* and *Deny* replies, though the number of messages presenting stereotypes remains relatively low, possibly due to supporters' intent not to have their messages moderated by the platform. Examples of more polarized conversations are Conversations #3 and #5: the former does not have a single *Deny* response, with the *Head* being a tweet criticizing the verdict for the 2017 Kobili Traoré murder trial in France, which attracted a significant number of replies containing stereotypes against immigrants and Muslims; whereas the latter provoked a larger proportion of *Deny* responses compared to *Support*, with its *Head* propagating a racial hoax about the Italian government supposedly secretly bringing in illegal immigrants by plane during the COVID-19 pandemic. Interestingly, Conversation #8 (also displayed in Figure 2) concerns the same subject as Conversation #3, but displays a greater proportion of *Deny* responses than the former, indicating that the same subject may be received wildly differently, depending on the context it is introduced in.

4.3. Lexical analysis

To investigate the vocabulary employed by users supporting and denying the hoaxes expressed in the heads of conversational threads, we present: the most relevant n-grams (unigrams, bigrams, and trigrams) of the messages annotated with the presence of stereotypes. The n-grams are weighted using the TF-IDF measure on normalized texts after a specific phase of preprocessing that involves: the deletion of all user mentions, stop-words, punctuation and URLs, leaving only words that were lexically significant. For the tokenization and lemmatization, we employed the small model for the Italian language available in the *SpaCy*[5] library.

By looking at the resulting lists of words and expressions, we noticed that texts labeled as *Support* are explicitly offensive towards immigrants. On the contrary, the ones that are labeled as *Deny* tend to stress the condition of need and poverty of immigrants, and are more empathetic. In this second list, we also noticed some offensive words but towards political parties leaning far-right. Some of the most relevant n-grams of both lists with their TF-IDF values are reported in Table 5.

Support	TF-IDF	Deny	TF-IDF
casa	5.357	nave	0.348
governo	3.889	disperato	0.346
entrare	3.750	pelle povero	0.346
clandestino	3.323	propaganda	0.346
bastardo	2.901	difeso	0.281
cinese	2.677	minacciare	0.281
potere	2.599	povero cristo	0.266
dare	2.521	poveraccio	0.242
merde	1.822	andare cagare leghista	0.175
schifoso	1.605	affamato malato	0.167
invasione	1.574	indifese	0.167
risorsa	1.338	raccogliere pomodoro	0.157
peso	1.314	approdo coraggio	0.121

Table 5
The most relevant n-grams extracted from messages that support and deny the hoax.

In Table 5 we can also notice how the TF-IDF scores greatly vary between *Support* and *Deny*. This is due to the different number of tweets labeled with one or the other category (see Table 4). The top ranking term for the *Support* category is the word *"casa"* (lit. house), probably coming from derogatory expressions like *"rimandiamoli a casa loro"* (lit. let's send them back to their house).

5. Conclusions and Future Work

In this article, we explored the expression of stance and stereotypes as occurring in a dataset of Twitter conversational threads in Italian, focused on the topic of migration. The dataset consists of dialogues originating from tweets containing misinformation marked as untrustworthy by experts.

The analysis of the dataset shed light on the distribution of stance labels and their relationship with stereotypes. The majority of tweets were annotated as *Comment*, followed by *Support*, *Query*, and finally *Deny*. While

[5] https://spacy.io/

there was no significant co-occurrence between stance and stereotypes, tweets annotated as *Support* were more likely to contain racial stereotypes. On the other hand, tweets annotated as *Comment* were less likely to exhibit forms of stereotyping.

The corpus analysis provided insights into how the structure and nature of conversations, and lexical choices in messages, affect the perceived stance of users towards racial hoaxes.

In conclusion, this work paves the way for further investigations about topics closely related to the social phenomenon of misinformation that should be countered to stimulate accurate information dissemination and create a more inclusive online environment. In future research, we may increase the size of the dataset, improve the annotation guidelines and consider the feedback provided by the annotators. We may moreover further investigate the relationship between stance and stereotypes, as well as explore interventions to mitigate the harmful effects of stereotypes in online conversations.

Limitations

In line with the recent trend of the main NLP conferences, we add a brief section addressing the limitations of our work. In this work, we enrich our corpus previously introduced in Bourgeade et al. [14], using a similar annotation framework, and therefore the same limitations brought forward in this work still apply here: more specifically, regarding the practical reliability of the theoretical social-psychological framework used to derive the annotation guidelines. In addition, the Italian subset of the multilingual StereoHoax corpus has a very limited size and presents many unbalanced dimensions and high data sparsity. If in the future it will be used for computational tasks, as it is intended, it should be made more balanced and more inclusive in terms of data sources.

Acknowledgments

This work was partially funded by the International project *STERHEOTYPES - Studying European Racial Hoaxes and sterEOTYPES*, funded by the Compagnia di San Paolo and VolksWagen Stiftung under the 'Challenges for Europe' Call for Projects (CUP: B99C20000640007). The work of T. Bourgeade is funded by the project StereotypHate, funded by the Compagnia di San Paolo for the call 'Progetti di Ateneo - Compagnia di San Paolo 2019/2021 - Mission 1.1 - Finanziamento ex-post'.

References

[1] G. Ruffo, A. Semeraro, A. Giachanou, P. Rosso, Studying fake news spreading, polarisation dynamics, and manipulation by bots: A tale of networks and language, Computer Science Review 47 (2023) 100531.

[2] E. Aïmeur, S. Amri, G. Brassard, Fake news, disinformation and misinformation in social media: A review, Social Network Analysis and Mining 13 (2023) 30.

[3] A. Cerase, C. Santoro, From racial hoaxes to media hypes: Fake news' real consequences., Amsterdam University Press, 2018, pp. 333–354. URL: http://www.jstor.org/stable/j.ctt21215m0.20.

[4] K. K. Russell, The color of crime: Racial hoaxes, white fear, black protectionism, police harassment, and other macroaggressions, New York University Press New York, 1998.

[5] D. Küçük, F. Can, Stance detection: A survey, ACM Computing Surveys (CSUR) 53 (2020) 1–37.

[6] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, C. Cherry, SemEval-2016 Task 6: Detecting Stance in Tweets, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), ACL, 2016.

[7] FNC, Fake News Challenge Stage-1 (FNC-1): Stance Detection, http://www.fakenewschallenge.org/, 2017.

[8] A. Aker, L. Derczynski, K. Bontcheva, Simple Open Stance Classification for Rumour Analysis, in: Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2017), INCOMA Ltd., 2017.

[9] G. Gorrell, E. Kochkina, M. Liakata, A. Aker, A. Zubiaga, K. Bontcheva, L. Derczynski, SemEval-2019 Task 7: RumourEval 2019: Determining Rumour Veracity and Support for Rumours, in: Proceedings of the 13th International Workshop on Semantic Evaluation: NAACL HLT 2019, ACL, 2019, pp. 845–854.

[10] A. T. Cignarella, M. Lai, C. Bosco, V. Patti, P. Rosso, SardiStance@EVALITA2020: Overview of the Stance Detection in Italian Tweets, in: Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020), 2020.

[11] M. Stranisci, M. De Leonardis, C. Bosco, V. Patti, The expression of moral values in the twitter debate: a corpus of conversations, Italian Journal of Computational Linguistics 7 Special Issue: Computational Dialogue Modelling: The Role of Pragmatics and Common Ground in Interaction (2021).

[12] C. Bosco, V. Patti, S. Frenda, A. T. Cignarella, M. Paciello, F. D'Errico, Detecting racial stereotypes: An

Italian social media corpus where psychology meets NLP, Information Processing & Management 60 (2023) 103118.
[13] F. D'Errico, C. Papapicco, M. T. Delor, 'Immigrants, hell on board': Stereotypes and prejudice emerging from racial hoaxes through a psycho-linguistic analysis., Journal of Language & Discrimination 6 (2022).
[14] T. Bourgeade, A. T. Cignarella, S. Frenda, M. Laurent, W. Schmeisser-Nieto, F. Benamara, C. Bosco, V. Moriceau, V. Patti, M. Taulé, A Multilingual Dataset of Racial Stereotypes in Social Media Conversational Threads, in: Findings of the Association for Computational Linguistics: EACL 2023, 2023, pp. 674–684.
[15] S. T. Fiske, A. J. Cuddy, P. Glick, Universal dimensions of social cognition: Warmth and competence, Trends in cognitive sciences 11 (2007) 77–83.
[16] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, C. Cherry, A dataset for detecting stance in tweets, in: N. C. C. Chair), K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), ELRA, Paris, France, 2016.
[17] C. M. B. McLaughlin, The association between two categorical variables: A review of cramér's v and alternative statistics, The Journal of the Royal Statistical Society. Series D (The Statistician) 26 (1977) 187–195.
[18] S. Akhtar, V. Basile, V. Patti, A new measure of polarization in the annotation of hate speech, in: AI*IA 2019 – Advances in Artificial Intelligence, volume 11946, Springer International Publishing, Cham, 2019, pp. 588–603. URL: http://link.springer.com/10.1007/978-3-030-35166-3_41. doi:10.1007/978-3-030-35166-3_41.