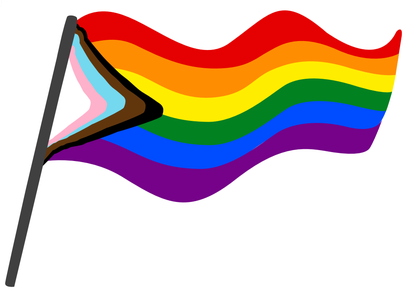
Alessandra Teresa Cignarella [1,2]
Manuela Sanguinetti [3]
Simona Frenda [1,2]
Andrea Marra [1]
Cristina Bosco [1]
Valerio Basile [1]

1 • Computer Science Department
University of Turin, Italy

2 • aequa-tech
Turin, Italy

3 • Department of Mathematics and Computer Science
University of Cagliari, Italy

QUEEREOTYPES: A MULTI-SOURCE ITALIAN CORPUS OF STEREOTYPES TOWARDS LGBTQIA+ COMMUNITY MEMBERS



INTRODUCTION

The QUEEREOTYPES corpus includes social media texts regarding **LGBTQIA+** individuals, behaviors, ideology, and events. The texts were collected from Facebook and Twitter in 2018 and were annotated for the presence of stereotypes, and orthogonal dimensions (such as **hate speech, aggressiveness, offensiveness, and irony** in one sub-corpus, and **stance** in the other). The resource was developed by Natural Language Processing researchers together with activists from an Italian LGBTQIA+ not-for-profit organization. The creation of the dataset allows the NLP community to **study stereotypes against marginalized groups**, individuals and, ultimately, to develop proper tools and measures to reduce the online spread of such stereotypes.

DOWNLOAD THE DATASET

DATASET CONSTRUCTION

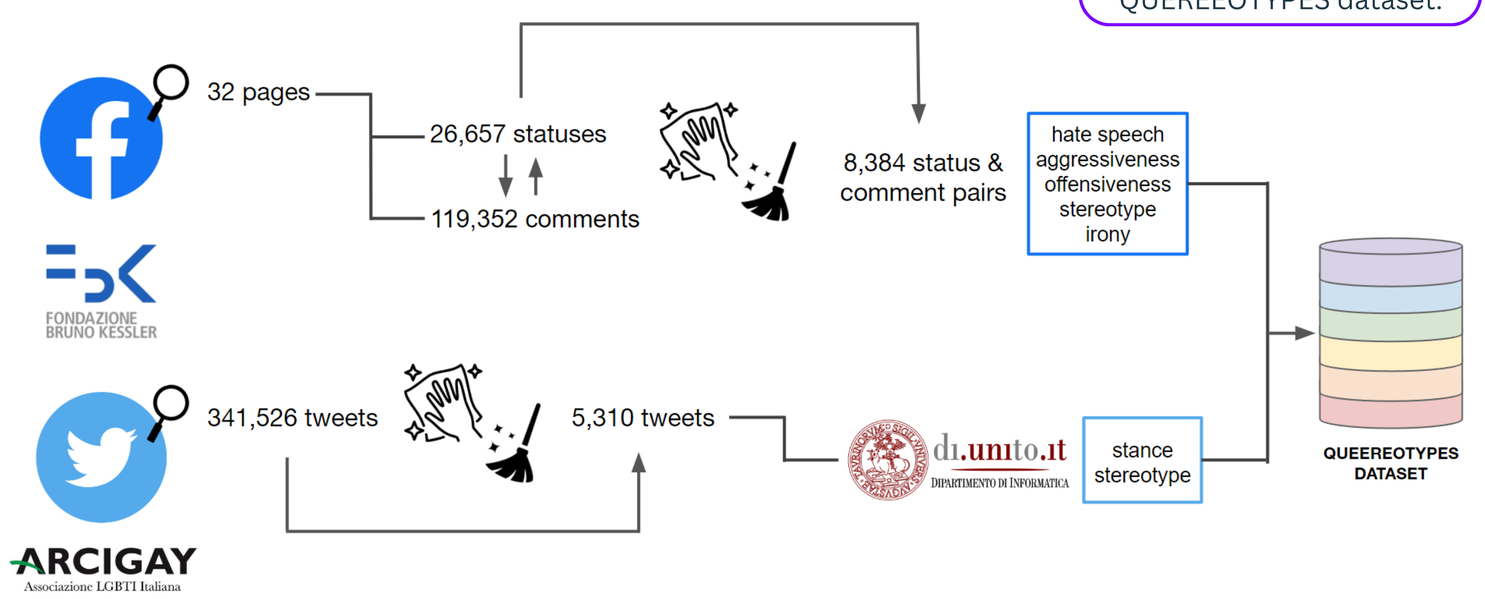


Figure 1: Pipeline of the creation of the multi-source QUEREEOTYPES dataset.

With this contribution, we aim at joining the research efforts of two different projects that are devoted to a similar goal. First, by merging two datasets and refining them in order to be used for **Natural Language Processing tasks,** such as the study of **hateful language and stereotypes/negative biases** towards LGBTQIA+ individuals.

In particular, we believe useful insights could emerge from the comparison between the two sub-corpora, which are of a same textual genre, but from different sources (**Facebook** and **Twitter/X**), thus encompassing multiple views and perspectives from different kind of users. This motivated the creation of this multi-source dataset, that we henceforth call QUEEREOTYPES.

QUEEREOTYPES SIZE AND SOME EXAMPLES

	"Non- aggregated"	"Gold Standard"
Facebook	8,384	2,888
Twitter / X	5,310	3,427
Total	**13,694**	**6,215**

Table 1: Total number of texts with 'non-aggregated' labels and of texts of the 'gold standard'.

STEREOTYPE CLASSIFICATION

	HaSpeeDe2 Setting			*Expanded Setting*		
	P	**R**	**F1**	**P**	**R**	**F1**
mBERT	.740	.719	.698	.739	.740	.735
AlBERTo	751	.729	.716	.746	.744	**.744**

Table 2: Results of textual classification experiments on the Stereotype dimension with BERT and AlbERTo.

BASELINES

- Majority class baseline MCB (F1 = 0.355)
- Random Baseline RB (F1 = 0.504)
- HaSpeeDe2 Baseline SVC (F1 = 0.715)*
- HaSpeeDe2 Baseline MFC (F1 = 0.355)*

*) Note that in HaSpeeDe2 stereotypes are towards a different target (i.e., migrants)

CONCLUSIONS AND TAKE-HOME MESSAGES

1) In the Expanded Setting, the results show that **models fine-tuned on a broader training set improve their performance**, with Recall and F1-score being higher than in the HaSpeeDe2 Setting.

2) Overall, the obtained results highlight **the importance of dataset diversity** and extension in training models to enhance performance

3) Results also seem to point out that **stereotypes towards different targets share common traits**, therefore, the phenomenon of 'stereotyping' could be more generalizable, and the same models might be employed also for detection of stereotypes towards other vulnerable groups (women, elderly, disabled bodies, non-white people, ethnic minorities, homeless, etc.)

ACKNOWLEDGEMENTS

- "STERHEOTYPES - Studying European Racial Hoaxes and sterEOTYPES" funded by the Compagnia di San Paolo and Volkswagen Stiftung under the 'Challenges for Europe' call for Project (CUP: B99C20000640007).
- Compagnia di San Paolo - Bando ex-post 2020 "StereotypHate" and "Bando ex-post 2020 "Toxic Language Understanding in Online Communication - BREAKhateDOWN".
- 'Multilingual Perspective-Aware NLU' project in partnership with Amazon Alexa.
- DEMON "Detect and Evaluate Manipulation of ONline information" funded by MIUR under the PRIN 2022 grant BAXSPY (CUP F53D23004270006, NextGenerationEU).