Instruction for Participants:

MultiPRIDE @ EVALITA2026

Lia Draetta
Department of Computer Science
University of Turin
lia.draetta@unito.it

Chiara Ferrando
Department of Computer Science
University of Turin
chiara.ferrando@unito.it

Marco Madeddu
Department of Computer Science
University of Turin
marco.madeddu@unito.it

Mae Sosto
Human Center Data Analytics Group
Centrum Wiskunde & Informatica
mae.sosto@cwi.nl

Viviana Patti
Department of Computer Science
University of Turin
viviana.patti@unito.it

Paolo Rosso
Departamento de Sistemas Informáticos y Computación
Universitat Politècnica de València
prosso@dsic.upv.es

Cristina Bosco
Department of Computer Science
University of Turin
cristina.bosco@unito.it

Jacinto Mata
Departamento de Tecnologias de la Informacion
Universidad de Huelva
mata@uhu.es

Estrella Gualda
Departamento de Sociologia, Trabajo Social y Salud Publica
Universidad de Huelva
estrella@uhu.es

1 Task description

With this task, we invite participants to explore features and issues related to reclaimed language in the LGBTQ+ community. In particular, we suggest participants to focus on both the textual content of the inputs (such as arguments, slurs, denigratory words, self-labeling, figures of speech), and the contextual information that can be inferred from the users' profile (when available), such as their being part of the LGBTQ+ community or the political orientation. We propose a **binary classification task**, in which systems must classify whether a term related to LGBTQ+ context in a sentence is used with a reclamatory intent or not. Overall, we propose two different tasks:

- **Task A - Textual Content:**
  Participants are provided only with the textual content of the message.
  You can approach the task in *constrained* or *unconstrained* manner: in the first case, participants must use only the given training data, other resources, such as lexicons are allowed; however, it is not allowed to use additional training data in the form of tweets or sentences. In the second case (unconstrained), participants can use additional data for training, e.g., additional dataset annotated for reclaimed language. If participants chose to approach the task in unconstrained manner it must be specified within the submission of the runs.

  The task is composed of three subtasks depending on the language. Participants could choose to work indifferently on one, two or three languages but **multilingual experiments are welcomed** :
  – **Subtask A1.** ***Italian:*** participants are provided with Italian texts;
  – **Subtask A2.** ***Spanish:*** participants are provided with Spanish texts;
  – **Subtask A3.** ***English:*** participants are provided with English texts.

- **Task B - Contextual Content:**
  In addition to the texts, participants can leverage contextual information related to the author's profile, such as their biography (when available). This task is composed of two subtasks depending on the language:
  – **Subtask B1.** ***Italian:*** participants are provided with Italian texts;
  – **Subtask B2.** ***Spanish:*** participants are provided with Spanish texts;

Biographical information of the users is available only for Spanish and Italian. As for the Task A participants may choose to work on one or more languages. Although not mandatory, participants are encouraged to foster cross-linguistic analysis.

Examples The following table shows some examples of tweets annotated for the task.

Table 1: Reclamated and not-reclamated data examples from the multilingual dataset

Lang	User	Bio	Tweet	Reclamation
it	@user	Certo le circostanze non sono favorevoli. 🏳️✊🌈	In quanto disabile e frocia questi sono i miei PrideMonths. Ma vorrei anche dire che il giorno in cui nel manifesto di un evento lgbtqiatransfemminista verrà citato anche l'antiabilismo oltre ad anti sessismo/obtfobia/razzismo/specismo offro da bere	yes
it	@user	I veri partigiani furono i primi sovranisti! w la patria!	Ecco, adesso pensate all'iter o in affitto ed al male che fate al bambino branco di finocchi arcobaleno	no
es	@user	Me llaman feminista, roja y bollera.✊🌈💜	Buenas tardes a rojos, feministas, republicanos, maricones, bolleras y demás LGTBI 🌈 #LGTBI #pride	yes
es	@user	I live for that energy!	Hace rato pasó una caravana de movimiento LGT…etc y algunos me vieron parado observando y me gritaron que yo también era marica. Órale, bien «respetuosas» estas personas que exigen respeto. #PrideMonth #Pride2022 #LGTBQ	no
es	N/A	N/A	I use the word tranny all the time...but that's only in reference to working on my cars. Transgendered, transvestite, and drag queen folk are too fabulous to have their descriptions abbreviated.	yes
es	N/A	N/A	Actually that's what s faggot is. Fag is just something that needs to be burnt.	no

2 Training and Test Data

For the challenge purpose we collected a multilingual dataset by combining existing resources in three languages: Italian, Spanish, and English. All data were collected from social networks, blog post and TV series covering a period between 2020 and 2022.

As data sources, we used the TWITA collection for Italian [1], a dataset containing LGBTQ+ related messages for Spanish [3], and a dataset focused on examples of reappropriated language within the LGBTQ+ community [4] for English. The first two datasets comprehend only inputs from twitter, the English one aggregates data from different sources such as Twitter, Reddit and TV-series. To filter the initial datasets, we applied the same methodology across the three languages. First, we performed keyword-based filtering using terms related to homosexuality drawn from the Hurtlex lexicon [2] (e.g., fag, bitch, gay). Second, we refined the selection by targeting sentences with a higher likelihood of reappropriative language, identified through positive terms typically expressing pride and community belonging (e.g., pride, LGBT, queer, proud, rainbow). Some example of data are presented in table 1.

For each language we finally collected 1811 inputs for Italian, 1711 for English and 1461 for the Spanish language. The distribution of the label are presented in table **??**. Three different datasets, one for language, will released. Data will be splitted in three sets for each language:

- **Training Set (60%)**: fully labeled which can be used for training the model.
- **Test Set (40%)**: will be published after the assessment time frame.

The training set will be made available to the task participants under a Creative Commons 4.0 license[1], which defines the terms governing data use and citation. The full dataset will be accessible upon request after the challenge.

2.1 Format and Distribution

A single training set will be provided for both the Tasks A and B. The training sets are released in three different csv file, one for each language: `train_it.csv` for Italian, `train_en.csv` for english and `train_es.csv` for spanish. The training data will be released in the following format:

```
id  text  label  bio  lang
```

where `id` is the unique identifier of the message, `text` is the message content, `label` indicates whether the text contains reappropriated terms, `bio` contains the profile biography of the user who authored the message, when available, and `lang` specifies the primary language of the text.

[1] `https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode`

3 How to submit your runs

Once you have run your system over the test data, please submit the results in accordance with the following recommendations:

1. Choose a team name and name the files containing your runs in the following way: `multipride2025_TeamName_RunNumber`
2. Send all relevant files to the following address: `multipride.evalita@gmail.com` using the subject "multipride2025 – TeamName".
3. In the body of the email, **please specify all resources you used in the unconstrained run** (if any).
4. In addition, all teams will have to compile a report describing the methodology used and the results obtained in detail. Further information, along with the template and format, will be available soon on the Evalita website[2].

4 Submission format

Results for all tasks should be submitted in a plain text file with tab-separated fields (tsv). The format of the run files submitted by participants is the same for both tasks:

`id  label  lang`

Each submitted run must contain one text per line, including the `text_id` and the predicted label.

The teams can use additional data for training (e.g., additional data annotated), if so, details must be specified within the submission of the runs. Participants are invited to submit a maximum of two runs to experiment with different models and architectures, but discouraged from submitting slight variations of the same model.

5 Evaluation

We will provide a separate official ranking for Task A and Task B and each language. Participants will have to report if their runs are constrained or unconstrained[3], but there will be no separate rankings based on this distinction. Systems will be evaluated using macro F1-score computed over the *Reappropriation* binary label. We allow for a maximum of two runs.

Considering that the size of models may impact their performance, we will ask participants to provide the total number of parameters of their models. We will **not** make separate rankings for different ranges of parameters, but we will include the information in the paper to allow clearer interpretation of the results.

[2] `https://www.evalita.it/campaigns/evalita-2026/`

[3] unconstrained runs allow for the use of any material in addition to the provided training dataset

Baselines For all sub-tasks, we will compute a baseline by finetuning a language specific BERT model using weighed cross-entropy as the loss function (weights are based on the label distribution).

6 Updates and contact info

For additional details or updates, please visit the MultiPRIDE website: `https://multipride-evalita.github.io`. For any questions or issues, you can contact us at: `multipride.evalita@gmail.com`.

References

[1] Basile, V., Lai, M., Sanguinetti, M., et al. Long-term social media data collection at the university of turin. In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)* (2018), CEUR-WS, pp. 1–6.

[2] Bassignana, E., Basile, V., Patti, V., et al. Hurtlex: A multilingual lexicon of words to hurt. In *CEUR Workshop proceedings* (2018), vol. 2253, CEUR-WS, pp. 1–6.

[3] Mata, J., and Gualda, E. A dataset of spanish tweets on people and communities lgbtqi+ during the covid-19 pandemic 2020-2022 [lgbtqi+ dataset 2020-2022_es]. *A dataset of Spanish tweets on people and communities LGBTQI+ during the COVID-19 pandemic 2020-2022 [LGBTQI+ Dataset 2020-2022_es]* (2025).

[4] Zsisku, E., Zubiaga, A., and Dubossarsky, H. Hate speech detection and reclaimed language: Mitigating false positives and compounded discrimination. In *Proceedings of the 16th ACM Web Science Conference* (2024), pp. 241–249.