Overview of Third Shared Task on Homophobia and Transphobia Detection in Social Media Comments

Bharathi Raja Chakravarthi[1], **Prasanna Kumar Kumaresan**[2], **Ruba Priyadharshini**[3]
Paul Buitelaar[2], **Asha Hegde**[4], **Hosahalli Lakshmaiah Shashirekha**[4], **Saranya Rajiakodi**[5],
Miguel Ángel García-Cumbreras[6], **Salud María Jiménez-Zafra**[6],
José Antonio García-Díaz[7], **Rafael Valencia-García**[7], **Kishore Kumar Ponnusamy**[8],
Poorvi Shetty[9], **Daniel García-Baena**[6]
[1] School of Computer Science, University of Galway, Ireland.
[2] Data Science Institute, University of Galway, Ireland.
[3] Gandhigram Rural Institute-Deemed to be University, India.
[4] Mangalore University, Mangalore, India.
[5] Central University of Tamil Nadu, India. [6] SINAI, Universidad de Jaén, Spain.
[7] UMUTeam, Universidad de Murcia, Spain.
[8] Digital University of Kerala, India. [9] JSS College, Mysore.
bharathiraja.akr@gmail.com

Abstract

This paper provides a comprehensive summary of the "Homophobia and Transphobia Detection in Social Media Comments" shared task, which was held at the LT-EDI@EACL 2024. The objective of this task was to develop systems capable of identifying instances of homophobia and transphobia within social media comments. This challenge was extended across ten languages: English, Tamil, Malayalam, Telugu, Kannada, Gujarati, Hindi, Marathi, Spanish, and Tulu. Each comment in the dataset was annotated into three categories. The shared task attracted significant interest, with over 60 teams participating through the CodaLab platform. The submission of prediction from the participants was evaluated with the macro F1 score.

1 Introduction

The growth of the internet has given rise to the widespread use of social media, and numerous other online spaces (Chakravarthi, 2023). The use of social media in particular has seen a significant increase in communication across various languages around the world (Al-Hassan and Al-Dossari, 2022). These platforms enable users to post, share content and freely express their opinions on any subject at any time (Chakravarthi et al., 2022a)(Kumar et al., 2018). However, the liberty of expression found on the internet also comes with downsides. It allows individuals, who might otherwise feel powerless, to impact and even harm others' lives (Ponnusamy et al., 2023a). This is often facilitated by the anonymity and emotional detachment that online interactions provide (Kumaresan et al., 2023b).

The rapid increase in online content has raised significant concerns within digital communities (Kumaresan et al., 2022). This issue is particularly acute for individuals identifying as lesbian, gay, bisexual, transgender, and other LGBTQ+ identities, who often face heightened vulnerability (Díaz-Torres et al., 2020). Members of the LGBTQ+ community are frequently targets of harassment, discrimination, violence, and in extreme cases, even death, due to their appearance, who they love, or their gender identity (Kumaresan et al., 2023a). Sexual orientation and gender identity are fundamental aspects of personal identity and should be respected rather than used as grounds for discrimination (Thurlow, 2001). In several regions, being identified as LGBTQ+ can be life-threatening. Consequently, many seek support and connection through social media, hoping to find others with similar experiences and form supportive communities (Chakravarthi et al., 2022c)(Ponnusamy et al., 2023b).

The task at hand involves utilizing a newly established gold standard dataset designed for identifying instances of homophobia and transphobia. This shared task uses a new gold standard dataset in Dravidian and Indo-Arian languages Tamil, Malayalam, Telugu, Kannada, Gujarati, Tamil-English (code-mixed), Tulu, Hindi, Spanish, and English languages.

In this overview, we conducted a shared task on homophobia and transphobia at LT-EDI[1] in ten lan-

[1] https://sites.google.com/view/lt-edi-2024/

Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 124–132

guages which were annotated in 3 labels. In the upcoming section, we will describe the task description, dataset statistics, and participant-provided experiment analysis to investigate homophobia and transphobia detection from the YouTube comments on Dravidian languages.

2 Related Work

In the realm of natural language processing and computational linguistics, recent studies have made significant strides in understanding and analyzing the nuances of language as it pertains to social issues. A notable example is the work of Zhang and Luo (2019), who compiled a corpus to examine the linguistic behaviors of homosexual individuals in China, shedding light on cultural and linguistic patterns. Similarly, Chakravarthi et al. (2022a) developed a fine-grained taxonomy specifically for homophobia and transphobia in English and Tamil languages, providing a structured framework for analyzing such content (Ponnusamy et al., 2023a).

Expanding on this, Chakravarthi et al. (2022c) spearheaded a shared task focused on the identification of homophobia, transphobia, and non-anti-LGBT+ content in Tamil, English, and Tamil-English (code-mixed) languages (Lande et al., 2023). This initiative was crucial in understanding the subtleties and variations of discriminatory language across different linguistic contexts. Complementing this, Chinnaudayar Navaneethakrishnan et al. (2022) conducted a study on sentiment analysis and homophobia detection in code-mixed Dravidian language YouTube comments, covering Tamil, Malayalam, and English. This research was pivotal in exploring the intersection of sentiment analysis and social bias detection in multilingual online spaces (Shanmugavadivel et al., 2022)(Subramanian et al., 2022).

In a related vein, Manikandan et al. (2022) employed transformer-based model methodologies like BERT and XLMRoBERTa to identify transphobic and homophobic insults in social media comments. Their work highlighted the efficacy of advanced computational models in detecting subtle and explicit forms of hate speech. Further, the growing prevalence of social media and its impact on communication and relationship building has been explored in depth by researchers like Chakravarthi et al. (2022c) and Chakravarthi et al. (2022b). Their studies delved into the dynamics of social networking sites like YouTube, where user interactions through comments, likes, and shares can significantly influence public discourse and perception.

However, this increased interaction on social platforms also brings challenges, as highlighted by Diefendorf and Bridges (2020), who explored the prevalence of antisocial behaviors like misogyny, sexism, homophobia, and transphobia. Larimore et al. (2021) further contributed to this discussion by examining the occurrence of racism and other forms of bias in online spaces. These studies underscore the importance of developing robust computational methods to detect and analyze such harmful content. The field has seen a surge in research focusing on text-based algorithms for identifying abusive language (Pannerselvam et al., 2023) and hate speech, as demonstrated by the work on YouTube comment mining and the analysis of social media data for detecting discriminatory language.

Building on this foundation, a notable study, conducted in 2021, delved into Homophobia and Transphobia identification, providing valuable insights and methodologies for future research in this area. This body of work collectively emphasizes the crucial role of computational linguistics in addressing social issues and fostering more inclusive and respectful online environments.

3 Task Description

This task marks the third year we have conducted a shared task focused on homophobia and transphobia detection[2]. We present a diverse dataset sourced from YouTube comments and posts in ten different languages: English, Tamil, Malayalam, Telugu, Kannada, Gujarati, Hindi, Marathi, Spanish, and Tulu. This dataset is thoughtfully annotated with three distinct labels: homophobia, transphobia, and non-anti-LGBT+ content (a category designated for content that does not exhibit either of these prejudiced behaviors). Participants in this task are provided with extensive training, development, and testing datasets. The primary objective for participants is to devise robust algorithms capable of accurately categorizing these comments and posts. Their systems must discern whether the text under scrutiny contains instances of homophobia, transphobia, or falls into the non-anti-LGBT+ category. This challenge not only addresses the pressing issue of online hate speech but also contributes to

[2] https://codalab.lisn.upsaclay.fr/competitions/16056

inclusive language detection in a global context, promoting safer online spaces for all.

4 Dataset

Social media platforms like Twitter, Facebook, and YouTube significantly influence public opinion through user-generated content, impacting reputations. Recognizing this, there's an increasing need for tools to extract emotions and identify irrelevant content online, especially on platforms like YouTube, where user comments are rapidly growing. This is particularly relevant for the LGBTQ+ community, who engage with such platforms and share their thoughts on various topics. Focusing on YouTube, we collected comments from videos related to LGBTQ+ themes. We avoided personal stories from LGBTQ+ individuals to maintain privacy. Using the YouTube Comment Scraper tool[3], we gathered comments and manually annotated them with three labels: 'Homophobic', 'Transphobic', and 'Non-anti-LGBT+ content'. Our dataset expanded to include ten languages: English, Tamil, Malayalam, Telugu, Kannada, Gujarati, Hindi, Marathi, Spanish, and Tulu. This diverse dataset was compiled following the annotation guidelines provided in the dataset research paper (Kumaresan et al., 2023b). Table 1 shows the dataset statistics for all languages with all three labels.

Languages	Set	H	T	N
English	Train	179	7	2,978
	Dev	42	2	748
	Test	55	4	931
Tamil	Train	453	145	2,064
	Dev	118	41	507
	Test	152	47	634
Malayalam	Train	476	170	2,468
	Dev	197	79	937
	Test	140	52	674
Telugu	Train	2,907	2,647	3,496
	Dev	588	605	747
	Test	624	571	744
Kannada	Train	2,765	2,835	4,463
	Dev	585	617	955
	Test	599	606	951
Gujarati	Train	2,267	2,004	3,848
	Dev	498	454	788
	Test	510	436	794
Hindi	Train	45	92	2,423
	Dev	2	13	305
	Test	3	10	308
Marathi	Train	551	377	2,572
	Dev	129	80	541
	Test	112	69	569
Spanish	Train	250	250	700
	Dev	93	93	200
	Test	150	150	300
Tulu	Train	188		542
	Test	67		312

Table 1: Dataset statistics for all languages (H-Homophobia, T-Transphobia, and N-Non-anti-LGBT+ content)

5 Participants Methodology

In our shared task, we had a total of 61 participants registered, 12 teams who submitted results in various languages. Various teams employed innovative methodologies to tackle the challenge of detecting homophobia and transphobia in social media comments. The "dkit_nlp" (Yadav et al., 2024) team utilized a BERT (bert-base-uncased) (Devlin et al., 2018) model, combining training and development sets and fine-tuning it with specific hyper-parameters for optimal performance. "MUCS" approached the task with voting classifiers, employing techniques like Syllable tf-idf, oversampling, and transformer-based BERT models, alongside mvlearn. "SCaLAR_sys1" utilized AdaBoost, integrating multiple classification models and focusing on hyper-parameter tuning to enhance the performance of their ensemble model. The "Hypnotize" team analyzed deep learning and transformer-based models across eight languages, focusing on data preprocessing and hyper-parameter tuning to address imbalances in certain languages.

"catnlp" adopted a transformer-based approach, retraining XLM-RoBERTa (Conneau et al., 2019) with script-switched Wikipedia[4] abstracts and customizing language profiles for multi-class classification. They evaluated their model across various pre-trained language models without significant improvement from additional social media data. "Quartet" (Allan H et al., 2024) implemented a thorough dataset analysis and preprocessing, followed by the use of traditional machine learning models and BERT models, selecting the best-performing model for the final evaluation. "MEnTr" (Arora et al., 2024) em-

[3] https://pypi.org/project/youtube-comment-scraper-python/

[4] https://en.wikipedia.org/wiki/ScriptSwitch

Team name	Run	M_F1-score	Rank
dkit (Yadav et al., 2024)	Run1	0.496	1
MUCS	Run2	0.493	2
KEC_AIDS	-	0.466	3
CUTN_CS_HOMO	BERT	0.457	4
SCaLAR	Run3	0.438	5
MEnTr (Arora et al., 2024)	-	0.407	6
Hypnotize	-	0.384	7
KEC_AI_NLP (Shanmugavadivel et al., 2024)	Run1	0.369	8
quartet (Allan H et al., 2024)	-	0.347	9
cantnlp	Run1	0.323	10
MasonTigers (Goswami et al., 2024)	-	0.323	10

Table 2: Rank list for English dataset

Team name	Run	M_F1-score	Rank
Hypnotize	-	0.880	1
MUCS	Run3	0.860	2
bytellm (Manukonda and Kodali, 2024)	-	0.801	3
MEnTr (Arora et al., 2024)	-	0.746	4
MasonTigers (Goswami et al., 2024)	-	0.512	5
quartet (Allan H et al., 2024)	-	0.483	6
KEC_AI_NLP (Shanmugavadivel et al., 2024)	Run1	0.315	7

Table 3: Rank list for Tamil dataset

ployed an ensemble model integrating three transformer models—Multilingual BERT (Devlin et al., 2018), XLM-RoBERTa(Conneau et al., 2019), and MuRIL (Khanuja et al., 2021) with dataset augmentation to enhance generalization across languages. "KEC_AI_NLP" (Shanmugavadivel et al., 2024) used a combination of machine learning and deep learning techniques, with a focus on preprocessing and SMOTE oversampling, finding that the random forest model yielded the highest accuracy. "MasonTiger" (Goswami et al., 2024) used XLM-R for nine languages and few-shot prompting for Tulu, addressing the challenge posed by imbalanced datasets. Finally, "bytesizedllm" (Manukonda and Kodali, 2024) utilized custom-built subword tokenizers and embeddings from AI4Bharat's data, employing a Bidirectional Long Short-Term Memory (Bi-LSTM) classifier for their classification tasks. The "CUTN_CS_HOMO" team approached the shared task with Malayalam and English datasets, addressing class imbalances with RandomOverSampler for oversampling. They utilized mBERT and MuRIL (Khanuja et al., 2021) for Malayalam and BERT and RoBERTa (Conneau et al., 2019) for English, training with a learning rate of 2e-5 over four epochs. Their models yielded high accuracy, achieving 94% with both BERT and RoBERTa for English and up to 96% with MuRIL for Malayalam, ranking them 1st in Malayalam and 4th in English. Each team's unique approach contributed to the advancement of understanding in the field of online hate speech detection on homophobia and transphobia.

6 Results

There was a total of 61 participants from the 12 teams submitted their results. For English 11 teams, Tamil 7 teams, Spanish 4 teams, Hindi 7 teams, Gujarati 6 teams, Telugu 8 teams, Kannada 8 teams, Malayalam 9 teams, Marathi 6 teams, and Tulu 4 teams submitted the final results of all languages. Table 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 shows the final rank list of all languages. We used the average macro F1 score to rank the teams as it identifies the F1 score in each label and calculates their unweighted average. Macro F1 scores arrange the runs in descending order. The "dkit" (Yadav et al., 2024) achieved Rank 1 in English, owing to their effective combination of training and development sets, a strategic cap on sequence length at 128, and meticulous hyperparameter tuning on the BERT

Team name	Run	M_F1-score	Rank
MEnTr (Arora et al., 2024)	-	0.582	1
MUCS	Run3	0.532	2
MasonTigers (Goswami et al., 2024)	-	0.499	3
KEC_AI_NLP (Shanmugavadivel et al., 2024)	Run1	0.369	4

Table 4: Rank list for Spanish dataset

Team name	Run	M_F1-score	Rank
MUCS	Run2	0.458	1
SCaLAR	Run1	0.410	2
Hypnotize	-	0.403	3
cantnlp	Run1	0.326	4
quartet (Allan H et al., 2024)	-	0.326	4
MasonTigers (Goswami et al., 2024)	-	0.326	4
MEnTr (Arora et al., 2024)	-	0.325	5

Table 5: Rank list for Hindi dataset

Team name	Run	M_F1-score	Rank
Hypnotize	-	0.968	1
cantnlp	Run1	0.962	2
MEnTr (Arora et al., 2024)	-	0.960	3
MUCS	Run2	0.958	4
MasonTigers (Goswami et al., 2024)	-	0.935	5
quartet (Allan H et al., 2024)	-	0.893	6

Table 6: Rank list for Gujarati dataset

Team name	Run	M_F1-score	Rank
Hypnotize	-	0.971	1
MasonTigers (Goswami et al., 2024)	-	0.971	1
MEnTr (Arora et al., 2024)	-	0.960	2
byteLLM (Manukonda and Kodali, 2024)	-	0.959	3
MUCS	Run1	0.958	4
SCaLAR	Run1	0.911	5
quartet (Allan H et al., 2024)	-	0.891	6
KEC_AI_NLP (Shanmugavadivel et al., 2024)	Run1	0.369	7

Table 7: Rank list for Telugu dataset

(bert-base-uncased) model. The “Hypnotize" team showed versatility across languages, securing Rank 1 in Tamil, Gujarati, Telugu, and Marathi, while also obtaining Rank 2 in Malayalam and Kannada and Rank 3 in Hindi. Their success was due to their comprehensive approach that included deep learning and transformer-based models, rigorous data preprocessing, and hyperparameter adjustments.

“MUCS", demonstrating their prowess, achieved Rank 1 in Hindi and Kannada, and Rank 2 in English, Tamil, Spanish, Marathi, and Tulu. Their methodology centered around voting classifiers trained with Syllable tfidf, augmented by oversampling and TL bert models, along with mvlearn. “MEnTr" (Arora et al., 2024), with their ensemble model integrating mBERT, XLM-RoBERTa, and MURIL, and complemented by strategic dataset augmentation, earned Rank 1 in Tulu and Spanish, and Rank 3 in Marathi, Telugu, and Gujarati. “CUTN_CS_HOMO", although specific details of their methodology were not provided, achieved Rank 1 in Malayalam, showcasing their expertise

Team name	Run	M_F1-score	Rank
MUCS	Run2	0.948	1
Hypnotize	-	0.946	2
MasonTigers (Goswami et al., 2024)	-	0.945	3
cantnlp	Run1	0.943	4
MEnTr (Arora et al., 2024)	-	0.935	5
bytellm (Manukonda and Kodali, 2024)	-	0.922	6
SCaLAR	Run1	0.903	7
quartet (Allan H et al., 2024)	-	0.887	8

Table 8: Rank list for Kannada dataset

Team name	Run	M_F1-score	Rank
CUTN_CS_HOMO	MuRIL	0.942	1
Hypnotize	-	0.909	2
bytellm (Manukonda and Kodali, 2024)	-	0.891	3
KEC_AI_NLP (Shanmugavadivel et al., 2024)	Run1	0.883	4
quartet (Allan H et al., 2024)	-	0.877	5
MUCS	Run3	0.870	6
cantnlp	Run2	0.775	7
MEnTr (Arora et al., 2024)	-	0.744	8
MasonTigers (Goswami et al., 2024)	-	0.505	9

Table 9: Rank list for Malayalam dataset

Team name	Run	M_F1-score	Rank
Hypnotize	-	0.626	1
MUCS	Run2	0.537	2
MEnTr (Arora et al., 2024)	-	0.488	3
MasonTigers (Goswami et al., 2024)	-	0.438	4
cantnlp	Run1	0.433	5
quartet (Allan H et al., 2024)	-	0.391	6

Table 10: Rank list for Marathi dataset

Team name	Run	M_F1-score	Rank
MEnTr (Arora et al., 2024)	Run1	0.707	1
MUCS	Run2	0.620	2
MasonTigers (Goswami et al., 2024)	Run1	0.452	3
cantnlp	Run1	0.452	3

Table 11: Rank list for Tulu dataset

in the field. These results underscore the diverse and innovative computational strategies employed by the teams in addressing the challenging task of detecting homophobia and transphobia across different languages on social media platforms.

7 Conclusion

We presented the third shared task overview on homophobia and transphobia detection in social media comments on ten different language datasets. We expect this task to have a long-term impact on the NLP domain because we received a variety of submissions with various methodologies. the most successful system was achieved by synthesizing advanced machine learning techniques, custom data preprocessing, and strategic model fine-tuning, effectively addressing the complex challenge of detecting homophobia and transphobia in

multilingual social media content. The prediction evaluation was evaluated with a macro F1 score. The increased number of participants and improved system performance indicate a growing interest in Dravidian NLP.

Acknowledgements

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223, supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2(Insight_2). It has also been partially supported by Project CONSENSO (PID2021-122263OB-C21), Project MODERATES (TED2021-130145B-I00) and Project SocialTox (PDC2022-133146-C21) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR, Project PRECOM (SUBV-00016) funded by the Ministry of Consumer Affairs of the Spanish Government, Project FedDAP (PID2020-116118GA-I00) and Project TrustReDaS (PID2020-119478GB-I00) supported by MICINN/AEI/10.13039/501100011033, and WeLee project (1380939, FEDER Andalucía 2014-2020) funded by the Andalusian Regional Government. Salud María Jiménez-Zafra has been partially supported by a grant from Fondo Social Europeo and the Administration of the Junta de Andalucía (DOC_01073). This work is also part of the research project LaTe4PoliticES (PID2022-138099OB-I00) funded by MCIN/AEI/10.13039/501100011033 and the European Fund for Regional Development (FEDER)-a way to make Europe, and the research project LT-SWM (TED2021-131167B-I00) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

References

Areej Al-Hassan and Hmood Al-Dossari. 2022. Detection of hate speech in Arabic tweets using deep learning. *Multimedia systems*, 28(6):1963–1974.

Shaun Allan H, Samyuktaa Sivakumar, Rohan R, Nikilesh Jayaguptha, and Thenmozhi Durairaj. 2024. Quartet@LT-EDI 2024: Support Vector Machine based Approach for Homophobia/Transphobia Detection in Social Media Comments. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Adwita Arora, Aaryan Mattoo, Divya Chaudhary, Ian Gorton, and Bijendra Kumar. 2024. MEnTr@LT-EDI-2024: Multilingual Ensemble of Transformer Models for Homophobia/Transphobia Detection. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Bharathi Raja Chakravarthi. 2023. Detection of homophobia and transphobia in YouTube comments. *International Journal of Data Science and Analytics*, pages 1–20.

Bharathi Raja Chakravarthi, Adeep Hande, Rahul Ponnusamy, Prasanna Kumar Kumaresan, and Ruba Priyadharshini. 2022a. How can we detect Homophobia and Transphobia? Experiments in a multilingual code-mixed setting for social media governance. *International Journal of Information Management Data Insights*, 2(2):100119.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Cn, S Sangeetha, Malliga Subramanian, Kogilavani Shanmugavadivel, Parameswari Krishnamurthy, Adeep Hande, Siddhanth U Hegde, Roshan Nayak, et al. 2022b. Findings of the Shared Task on Multitask Learning in Dravidian Languages. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 286–291.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Philip McCrae, Paul Buitelaar, Prasanna Kumaresan, and Rahul Ponnusamy. 2022c. Overview of the shared task on homophobia and transphobia detection in social media comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 369–377.

Subalalitha Chinnaudayar Navaneethakrishnan, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Malliga Subramanian, Prasanna Kumar Kumaresan, Bharathi, Lavanya Sambath Kumar, and Rahul Ponnusamy. 2022. Findings of shared task on Sentiment Analysis and Homophobia Detection of YouTube Comments in Code-Mixed Dravidian Languages. In *Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation*, pages 18–21.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised Cross-lingual Representation Learning at Scale. *CoRR*, abs/1911.02116.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of

Deep Bidirectional Transformers for Language Understanding. *CoRR*, abs/1810.04805.

María José Díaz-Torres, Paulina Alejandra Morán-Méndez, Luis Villasenor-Pineda, Manuel Montes, Juan Aguilera, and Luis Meneses-Lerín. 2020. Automatic detection of offensive language in social media: Defining linguistic criteria to build a Mexican Spanish dataset. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 132–136.

Sarah Diefendorf and Tristan Bridges. 2020. On the enduring relationship between masculinity and homophobia. *Sexualities*, 23(7):1264–1284.

Dhiman Goswami, Sadiya Sayara Chowdhury Puspo, Md Nishat Raihan, and Al Nahian Bin Emran. 2024. MasonTigers@LT-EDI-2024: An Ensemble Approach towards Detecting Homophobia and Transphobia in Social Media Comments. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, and Partha Talukdar. 2021. MuRIL: Multilingual Representations for Indian Languages.

Ritesh Kumar, Atul Kr Ojha, Shervin Malmasi, and Marcos Zampieri. 2018. Benchmarking aggression identification in social media. In *Proceedings of the first workshop on trolling, aggression and cyberbullying (TRAC-2018)*, pages 1–11.

Prasanna Kumar Kumaresan, Kishore Kumar Ponnusamy, Kogilavani Shanmugavadivel, Subalalitha Chinnaudayar Navaneethakrishnan, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2023a. VEL@ LT-EDI-2023: Detecting Homophobia and Transphobia in Code-Mixed Spanish Social Media Comments. *LTEDI 2023*, page 233.

Prasanna Kumar Kumaresan, Rahul Ponnusamy, Ruba Priyadharshini, Paul Buitelaar, and Bharathi Raja Chakravarthi. 2023b. Homophobia and transphobia detection for low-resourced languages in social media comments. *Natural Language Processing Journal*, 5:100041.

Prasanna Kumar Kumaresan, Rahul Ponnusamy, Elizabeth Sherly, Sangeetha Sivanesan, and Bharathi Raja Chakravarthi. 2022. Transformer Based Hope Speech Comment Classification in Code-Mixed Text. In *International Conference on Speech and Language Technologies for Low-resource Languages*, pages 120–137. Springer.

Kaustubh Lande, Rahul Ponnusamy, Prasanna Kumar Kumaresan, and Bharathi Raja Chakravarthi. 2023. KaustubhSharedTask@LT-EDI 2023: Homophobia-transphobia detection in social media comments with NLPAUG-driven data augmentation. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 71–77, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Savannah Larimore, Ian Kennedy, Breon Haskett, and Alina Arseniev-Koehler. 2021. Reconsidering annotator disagreement about racist language: Noise or signal? In *Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media*, pages 81–90.

Deepalakshmi Manikandan, Malliga Subramanian, and Kogilavani Shanmugavadivel. 2022. A System For Detecting Abusive Contents Against LGBT Community Using Deep Learning Based Transformer Models. In *Working Notes of FIRE 2022-Forum for Information Retrieval Evaluation (Hybrid). CEUR*.

Durga Prasad Manukonda and Rohith Gowtham Kodali. 2024. byteLLM@LT-EDI-2024: Homophobia/Transphobia Detection in Social Media Comments - Custom Subword Tokenization with Subword2Vec and BiLSTM. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Kathiravan Pannerselvam, Saranya Rajiakodi, Rahul Ponnusamy, and Sajeetha Thavareesan. 2023. CSS-CUTN@DravidianLangTech:Abusive comments Detection in Tamil and Telugu. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 306–312, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kishore Kumar Ponnusamy, Charmathi Rajkumar, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2023a. Team_Tamil at HODI: Few-Shot Learning for Detecting Homotransphobia in Italian Language. In *Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR. org, Parma, Italy*.

Rahul Ponnusamy, Malliga S, Sajeetha Thavareesan, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2023b. VEL@LT-EDI-2023: Automatic detection of hope speech in Bulgarian language using embedding techniques. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 179–184, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Kogilavani Shanmugavadivel, Sai Haritha Sampath, Pramod Nandhakumar, Prasath Mahalingam, Malliga Subramanian, Prasanna Kumar Kumaresan, and Ruba Priyadharshini. 2022. An analysis of machine learning models for sentiment analysis of Tamil code-mixed data. *Computer Speech Language*, 76:101407.

Kogilavani Shanmugavadivel, Malliga Subramanian, Shri Durga Ramanathan, Samyuktha Kathirvel, Srigha S, and Nithika Kannan. 2024. KEC-AI-NLP@LT-EDI-2024: Homophobia and Transphobia Detection in Social Media Comments using Machine Learning. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Malliga Subramanian, Ramya Chinnasamy, Prasanna Kumar Kumaresan, Vasanth Palanikumar, Madhoora Mohan, and Kogilavani Shanmugavadivel. 2022. Development of multi-lingual models for detecting hope speech texts from social media comments. In *International Conference on Speech and Language Technologies for Low-resource Languages*, pages 209–219. Springer.

Crispin Thurlow. 2001. Naming the "outsider within": Homophobic pejoratives and the verbal abuse of lesbian, gay and bisexual high-school pupils. *Journal of adolescence*, 24(1):25–38.

Sargam Yadav, Abhishek Kaushik, and Kevin McDaid. 2024. dkit@LT-EDI-2024: Detecting Homophobia and Transphobia in English Social Media Comments. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Ziqi Zhang and Lei Luo. 2019. Hate speech detection: A solved problem? the challenging case of long tail on Twitter. *Semantic Web*, 10(5):925–945.