

Zootopi at HOPE2023@IberLEF: Is Zero-Shot ChatGPT the Future of Hope Speech Detection?

Anh Ngo^{1,†}, Hanh Thi Hong Tran^{2,3,4,*,†}

¹Wroclaw University of Science and Technology, Poland

²Jožef Stefan International Postgraduate School, Slovenia

³Jožef Stefan Institute, Slovenia

⁴University of La Rochelle, France

Abstract

Hope Speech Detection is a Natural Language Processing (NLP) task where we aim to detect any message or text that can relax a hostile environment and inspire people in a time of suffering illness, stress, loneliness or depression with optimism. In this paper, we proposed two different mechanisms to identify whether the text contains Hope Speech or not: (1) We consider the problem as a sequence classification task and examine whether the language model improves the performance when additional training data from another language is added and its capacity to apply knowledge in a language to another unseen one with XLMR; (2) We evaluate ChatGPT's performance and examine how significant the language and prompting model's prediction affected by certain sensitive words, which could potentially lead to data poisoning. The experiments are conducted on two datasets from different social networks with informal languages proposed by the IberLEF 2023 Task [1], including Spanish tweets and English Youtube comments, and the performance is evaluated using Precision, Recall, and F1-score. Overall, ChatGPT outperforms other approaches when applying to the less-resourced Spanish tweets while monolingual XLMR surpassed other performances in rich-resourced English comments.

Keywords

Hope Speech, Transformers, Multilingual, Cross-lingual, ChatGPT

1. Introduction

Discriminatory messages, such as Hate Speech, are prevalent on social media platforms and often target individuals because of their race, ethnicity, gender, sexual orientation, nationality, or religion. Research [2] has been conducted on the significance of social media in the lives of vulnerable groups (e.g., the LGBT community, people with disabilities, and racial minorities) and the findings indicate that a vulnerable individual's social media engagement is crucial in shaping their personality and their perspectives on society [3, 4]. To counteract the negative impact of Hate Speech, our study focuses on Hope Speech, which comprises encouraging, supportive, and inclusive messages that can relax a hostile environment and motivate several people effectively when they are in times of suffering from illness, stress, loneliness, or depression [2].

IberLEF 2023, September 2023, Jaén, Spain

*Corresponding author.

†These authors contributed equally.

✉ anh@zootopi.dev (A. Ngo); hanh@zootopi.dev (H. T. H. Tran)

🌐 <https://github.com/haanh764> (A. Ngo); <https://github.com/honghanhh> (H. T. H. Tran)

🆔 0000-0002-5993-1630 (H. T. H. Tran)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Unlike the well-established domain of Hate Speech Detection, the detection of Hope Speech is a relatively nascent area of research. Inspired by the second workshop on Language Technology for Equality, Diversity, and Inclusion (LT-EDI-2022) for five languages (Tamil, Malayalam, Kannada, English, and Spanish), *IberLEF 2023 Task - HOPE. Multilingual Hope Speech Detection* [5], in pursuit of equality, diversity, and inclusion, focuses on Hope Speech Detection in two languages, Spanish and English, with expansion and improvement in both corpora. Given the Spanish tweets [6], and English Youtube comments [2] of the task, we aim to identify whether it contains Hope Speech (HS) or not (NHS).

The novelties of our contribution are threefold: (1) Examining whether the language model’s performance improves when additional training data from another language is added; (2) Examining the language model’s capacity to apply knowledge of Hope Speech learned in one language to another unseen language in both directions (from rich-resourced corpus to less-resourced one and vice versa); and (3) Investigating the extent to which sensitive words can cause data poisoning by influencing the language and prompting model’s predictions. Our work brings valuable insights for Hope Speech Detection in the less-resourced languages niche.

2. Related work

The emergence and widespread use of social media platforms have significantly transformed how people communicate and, consequently, led to a significant effort of research dedicated to detecting harmful content, such as offensive language, toxic comments, and Hate Speech Detection. However, on the reverse side, Hope Speech Detection is a relatively novel field of study and currently has limited available literature reviews.

In an initial work by Palakodety et al. [7], the identification of hopeful language in YouTube comments was explored, specifically in the context of the Pakistan and India war. Until now, only a small number of datasets have been developed for the Hope Speech Detection task, such as the HopeEDI dataset [2], which contains YouTube comments in English, Tamil, and Malayalam; the KanHope dataset [8], which includes YouTube comments in code-mixed Kannada-English language; and the SpanishHopeEDI dataset [6], which covers LGBT-related tweets in Spanish.

Hossain et al. [9], Aggarwal et al. [10], S et al. [11] employed various machine learning and deep learning models (e.g., Support Vector Machine, Naive Bayes, Ensemble, Convolutional Neural Network, Bi-LSTM, and Recurrent Neural Network) in combination with word embeddings (e.g., TF-IDF, Word2vec, and fastText) to handle the HopeEDI dataset. Additionally, Hossain et al. [9], Mahajan et al. [12], Aggarwal et al. [10] used Transformer-based techniques for fine-tuning multiple language models, including mBERT, XLMR, Indic-BERT, and XLNet. Furthermore, Puranik et al. [13] employed the fine-tuned ULMFit model, while Ghanghor et al. [14] used customized mBERT and customized XLMR. Moreover, Hande et al. [8] fine-tuned the Dual-channel BERT4Hope model for the code-mixed language in the KanHope dataset.

Regarding incorporating ChatGPT into NLP tasks, recent research has primarily focused on evaluating ChatGPT’s performance in text generation [15, 16, 17, 18] and question-answering [19, 20, 21] tasks. In addition, Kocoń et al. [22]’s study also evaluated ChatGPT’s performance in various NLP tasks, including offensiveness detection, while in Huang et al. [23], the authors assessed its ability to generate explanations for implicitly hateful speech.

In the realm of considering the task as a sequence-classification problem, our work adopts a similar evaluation of multilingual and cross-lingual learning that has been applied to other languages in other downstream tasks (e.g., sentiment analysis [24], keyword or term extraction [25], to mention a few). Meanwhile, we evaluated the performance of ChatGPT to our specific tasks under the inspiration of Huang et al. [23] for hateful speech. None of our two directions had been previously explored in Hope Speech Detection.

3. Dataset

In the context of IberLEF 2023 Task - HOPE. Multilingual Hope Speech Detection [5], we focus on two proposed corpora with informal texts from two different social platforms, including training, development, and testing sets of Spanish tweets [6] from Twitter and English comments [2] from Youtube. For both corpora, the data distribution is presented in Table 1 while the sequence length distribution for each class in the training and development sets is presented in Figure 1.

Table 1
Data distribution by class and set.

Labels	English		Spanish	
	Train - Dev.	Test	Train - Dev.	Test
Hope Speech (HS)	2,229	-	791	-
Non Hope Speech (NHS)	23,221	-	821	-
Total	25,450	4,805	1,612	450

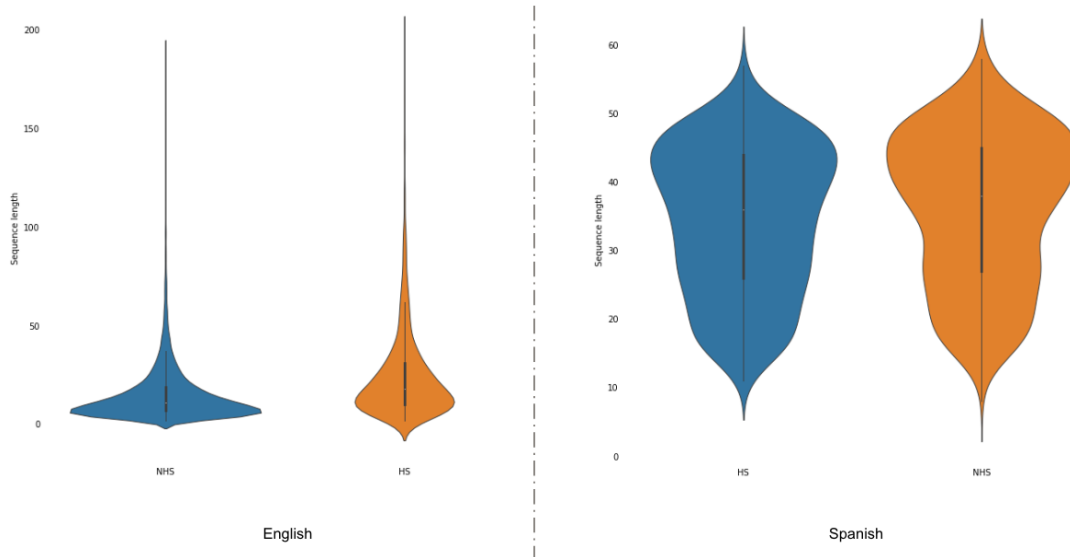


Figure 1: The distribution of sequence length for each class in the training and development sets.

Table 1 demonstrates the two different characteristics of the two datasets: (1) English comments can be considered as a rich-resourced corpus with 25,450 examples collected for training and development and 4,805 examples for testing. In contrast, Spanish tweets belong to a less-resourced corpus where the amount of training and development examples is approximately 16 times smaller than the number of English comments; (2) The label distribution in the training and development phase is highly imbalanced in the English set where the amount of HS examples is 10 times larger than the amount of NHS samples. Meanwhile, the Spanish set contains a relatively balanced amount of HS and NHS examples.

Besides, Figure 1 provides another insight on the distribution of sequence length, that is, the number of words within a sequence. Although English comments have a positively skewed distribution, most of the English comments and Spanish tweets are short, which leads to a lack of context for the language model training phase. Additionally, both use informal language, which includes misspellings (e.g., *Hi, teachet Hadar. Youre my favorite teachet. L liked your explanation very much...*), emojis, and onomatopoeias (e.g., *Hmmm...*). Figure 2 and 3 depict data samples from the corpus used in this task.

Figure 2: Data samples from English corpus.

Text: *I feel so base for that guy! They treated him as if he wasn't a human just because of who he loved!*

Category: *HS*

Text: *So did I. She's clearly grooming herself and behaving in a more masculine manner. Lol when she said she didn't agree I was like "...you don't?"*

Category: *NHS*

Figure 3: Data samples from Spanish corpus.

Text: *Cada 28 de Junio me levanto escuchando Born This Way. Sigue siendo un himno. Porque cada persona del colectivo #LGTB+ nació así. Porque nadie escogería a propósito una vida de opresión. Sigamos luchando por nuestra dignidad. Feliz #Orgullo #Orgullo2021*

Category: *HS*

Text: *Ánimo, donde están los defensores del mal llamado movimiento #LGTBI ? Los mismos q defienden la Inmigración sin barrerasDONDE?*

Category: *NHS*

4. Methodology

4.1. Sequence Classification

We consider our task as an experiment on sequence classification problems with XLMR [26], a Transformer-based model pretrained on 2.5TB of filtered CommonCrawl data containing 100 languages. Three experimental setups are proposed to verify our first two hypotheses, including monolingual, multilingual, and cross-lingual learning.

- **Monolingual setup:** We fine-tune two monolingual XLMR classifiers for English and Spanish, respectively, and test each model in the same language. In this scenario, we evaluate how well the model performs when there is a match between the language of the training set and the test set.
- **Cross-lingual setup:** We fine-tune XLMR classifiers in a language and test it on another language not appearing in the training set. In this scenario, we examine the model’s performance to apply knowledge of Hope Speech learned in a language to another unseen language. Note that we consider both directions: transferring knowledge from a rich-resourced language (e.g., English) to a lesser-resourced language (e.g., Spanish) and vice versa.
- **Multi-lingual setup:** We fine-tune our classifier using the training set from both languages and then apply the model to the test sets of all languages, respectively. In this scenario, we examine whether the language model’s performance improves when additional training data from another language is added.

We divide the dataset into training, validation, and testing sets, where the training-development set is split into two subsets of training and validation set, separately with a ratio of 3:1. The training split is used for fine-tuning the models while the validation split is used to prevent over-fitting during the fine-tuning phase. Finally, the test split is used for evaluation and is excluded during model training. More specifically, the model is fine-tuned on the training set to predict the probability for each sequence whether it is Hope Speech (HS) or not (NHS).

4.2. Prompting

We formulate two different prompting scenarios for the English and Spanish datasets. The Spanish dataset focuses solely on the LGBT+ community, and therefore, the prompt will furnish both sentence context and criteria for determining whether a text is classified as HS or NHS. Figure 4 illustrates the prompting sample for the Spanish dataset while Figure 6 describes the prompting scenario for the English dataset.

Figure 4: Spanish dataset prompt scenario

Describe the sentiment of the given text using one of these two attributes: “Hope Speech”, “Non Hope Speech”. Knowing that a text is considered as “Hope Speech” if the text: (1) explicitly supports the social integration of minorities; (2) is a positive inspiration for the LGTBI community; (3) explicitly encourages LGTBI people who might find themselves in a situation; or (4) unconditionally promotes tolerance. And knowing that a text is considered as “Non Hope Speech” if it: (1) expresses negative sentiment towards the LGTBI community; (2) explicitly seeks violence; or (3) uses gender-based insults. Do not give explanation, give only an answer as either “Hope Speech” or “Non Hope Speech”. Text: *Hilo- Hoy en día del #OrgulloLGTBI hoy más que nunca los derechos civiles , deben ser sobre el individuo sin excluir a nadie por sus gusto, orientaciones o religión . Eso ya es de libertad del individuo. Pero tampoco caigamos en leyes especiales para personas o grupos*

ChatGPT answer: *Hope Speech.*

Figure 5: English dataset prompt first scenario

Describe the sentiment of the given text using one of these two attributes: ‘Hope Speech’, ‘Non Hope Speech’. Knowing that the domains are relevant to social topics, such as Equality, Diversity and Inclusion, including LGBTIQ issues, COVID-19, women in STEM, or Black Lives Matter. Do not give explanation. Select one of these two labels: ‘Hope Speech’, ‘Non Hope Speech’. Text: *So pulling a statue out is going to re-write the past is it? Mindless fools. It would be different if there still was slavery in Bristol. But the present has moved on from the past and these moronic individuals need to also.*

ChatGPT answer: *Non Hope Speech.*

Figure 6: English dataset prompt second scenario

Describe the sentiment of the given text using one of these two attributes: “Hope Speech”, “Non Hope Speech”. Do not give explanation. Select one of these two labels: “Hope Speech”, “Non Hope Speech”. Text: *So pulling a statue out is going to re-write the past is it? Mindless fools. It would be different if there still was slavery in Bristol. But the present has moved on from the past and these moronic individuals need to also.*

ChatGPT answer: *Non Hope Speech.*

Furthermore, regarding the English dataset, as the context relates to different fields (e.g., equality, diversity, and inclusion, with topics like LGBTIQ issues, COVID-19, women in STEM, and Black Lives Matter), we created the second scenario represented in Figure 5, that provides some domain-specific information to the chatbot. To establish the experiment, we utilized PyGPT, an unofficial API client for ChatGPT.

4.3. Evaluation metrics

We measure the performance of our proposed classifiers using Precision, Recall, and F1-score (F1) per category and averaged using the macro-average method. This is also the evaluation metrics suggested by the *IberLEF 2023 Task - HOPE. Multilingual Hope Speech Detection*, which makes our work more comparable with the related work.

5. Results and Analysis

5.1. Results

Table 2 presents the performance of our approaches in both sequence classification and prompting regarding Precision, Recall, F1-score per class, and macro-average F1-score where the macro-average F1-score will be used for the final ranking of the model’s performance. Regarding the sequence-classification mechanism, the results demonstrate that rich-resourced cross-lingual learning (i.e., using knowledge from English to classify Spanish texts) can significantly enhance the model’s performance. Specifically, the cross-lingual setup in Spanish outperforms the monolingual setup by up to 7 percentage points (pp) in F1-score and up to 5 pp compared to the multilingual setup. However, for rich-resourced languages (e.g., English), the monolingual setup performs better without the need for additional knowledge from other less-resourced languages.

On the other hand, the best prompting with ChatGPT surpasses the performance of other sequence-classification approaches we applied with a large margin for all three evaluation metrics regarding the clean and less-resourced language, Spanish, but failed to capture the information from English contexts to provide a good classification.

In addition, the stability of the sequence-classification approach was evident as it achieved the sixth position on the official leaderboards for both the English and Spanish datasets. Conversely, the prompting method utilizing ChatGPT displayed potential inconsistency, outperforming other groups with first place for the Spanish dataset but finishing last for the English dataset. This result highlights the advantage of the ChatGPT prompting approach, which benefitted from the precise information provided by designated prompts. For instance, in the case of the Spanish corpus, the prompts provided specific scenarios, contexts, and domains where hope speech or non-hope speech occurred, leading to significantly improved accuracy. In contrast, for the English corpus, the information regarding the context and situation in which hope speech and non-hope speech happened was broader and less specific. As a result, we observed a notable decrease in performance.

5.2. Error analysis

5.2.1. Data ambiguity

Regarding the English datasets, there exist multiple instances of ambiguity which directly affect the performance of both proposed mechanisms, especially the prompting. For example, these sentences do not express any attitude towards Hope Speech or Non-Hope Speech, such as

Table 2
Performance comparison.

Model	Training language	Avg. Macro F1	HS			NHS		
			Precision	Recall	F1	Precision	Recall	F1
Spanish								
XLMR	English	0.7034	0.7765	0.4400	0.5617	0.7699	0.9367	0.8451
	Spanish	0.6349	0.8036	0.3000	0.4369	0.7335	0.9633	0.8329
	English + Spanish	0.6198	0.8511	0.2667	0.4061	0.7271	0.9767	0.8336
ChatGPT	-	0.9161	0.8671	0.9133	0.8896	0.9555	0.9300	0.9426
English								
XLMR	English	0.4862	0.0128	0.2857	0.0246	0.9965	0.9036	0.9478
	Spanish	0.4304	0.0064	0.3810	0.0125	0.9963	0.7385	0.8483
	English + Spanish	0.4794	0.0109	0.2857	0.0209	0.9965	0.8859	0.9379
ChatGPT	-	0.4429	0.0065	0.3333	0.0128	0.9962	0.7770	0.8730

“*Its for the birds*”, “*Peter Peter wake up*” or “*When I was little*”, or the sentence is too short and composed of only one or two words, such as “*Lol*”, “*Elo*”, “*Hahaa*”, or “*The sweetest*”.

Figure 7: ChatGPT response for a fail case in the first scenario.

<p>Describe the sentiment of the given text using one of these two attributes: ‘Hope Speech’, ‘Non Hope Speech’. Knowing that the domains are relevant to social topics, such as Equality, Diversity and Inclusion, including LGBTIQ issues, COVID-19, women in STEM, or Black Lives Matter. Do not give explanation. Select one of these two labels: ‘Hope Speech’, ‘Non Hope Speech’. Text: <i>The sweetest</i></p> <hr/> <p>ChatGPT answer: <i>I’m sorry, but the given text “The sweetest” does not contain enough information to determine its sentiment or whether it belongs to the category of “Hope Speech” or “Non Hope Speech”. Could you please provide me with a more substantial text to work with?</i></p>

Regarding the ChatGPT’s initial prompt scenario of the English dataset, where we provided the chatbot with context about the subject matter of comments, there were multiple instances where it failed to answer due to the natural ambiguity of the input text. Although we forced the chatbot not to provide any explanation for its response, in these situations, ChatGPT still gave a reason for being unable to perform the task, as described in Figure 7. To address the problem, we carried out the second scenario prompt without providing the context of the comments. This approach forced ChatGPT, and as a result, it was generally successful in assigning labels for those ambiguous instances.

In contrast, the Spanish dataset contains very rare ambiguous instances, which partially help ChatGPT perform significantly better than other methods and also than it does on the English dataset.

5.2.2. Model Bias

We noticed some biased responses from ChatGPT, particularly when it relates to sensitive words, such as the word “*Trump*”, which usually refers to President Donald Trump. As indicated in Figure 8, we evaluated ChatGPT’s performance using a subset of the training and development set, where we discovered that for several sentences that pertain to President Donald Trump or merely include the word “*Trump*” or “*Donald*”, ChatGPT categorized them as NHS, despite the expected correct label being HS. Other sentences also exhibit similar patterns of behavior, such as “*Even Israelis here don’t think that Trump’s stargate is racist*”, and “*There is a God who is benevolent. Intelligence and physical attractiveness are more important than anything else. She is probably one of the most intelligent CEOs and the most attractive speaker on TED Talks ever.*”

Figure 8: An example of a biased response from ChatGPT by the word “*Trump*”. The comment’s author stated that they support former President Donald Trump, but they are against Islam and feminism. However, they express support for the LGBT community and women’s rights, which made this comment classified as HS as providing ground truth from the training set.

Describe the sentiment of the given text using one of these two attributes: ‘Hope Speech’, ‘Non Hope Speech’. Knowing that the domains are relevant to social topics, such as Equality, Diversity and Inclusion, including LGBTIQ issues, COVID-19, women in STEM, or Black Lives Matter. Do not give explanation. Select one of these two labels: ‘Hope Speech’, ‘Non Hope Speech’. Text: *I’m a trump supporter. I’m anti Islam and anti feminism BUT I’m pro LGBT pro women’s rights (just not for all the femenzi BS.*

ChatGPT answer: *Non Hope Speech.*

Ground truth: *Hope Speech.*

To understand the cause of this phenomenon, we allow ChatGPT to explain its answer. The results in Figure 9 demonstrate that ChatGPT can associate the word “*trump*” with the former US President Donald Trump. However, due to the sensitive and controversial nature of the topic, ChatGPT categorized all related responses as NHS to avoid potential false positive alerts. It also explains the relatively lower performance of ChatGPT with the English dataset compared to other approaches.

On the contrary, most of the instances in the Spanish dataset clearly convey an attitude or sentiment towards the LGBT+ community. This also explains why ChatGPT performs significantly better than other methods as well as than it does on the English dataset.

6. Conclusion

In conclusion, we presented two different mechanisms to identify whether the text contains Hope Speech or not and verified our hypothesis regarding whether the language model improves the performance when additional training data from another language and the model’s capacity to apply knowledge in a language to another unseen one regarding sequence classification

Figure 9: An example to convey the reason behind the biased behavior of ChatGPT toward the word “Trump”

Describe the sentiment of the given text using one of these two attributes: ‘Hope Speech’, ‘Non Hope Speech’. Select one of these two labels: ‘Hope Speech’, ‘Non Hope Speech’. Text: *go trump*

ChatGPT answer: *The sentiment of the given text “go trump” is a “Non Hope Speech”. This is because the statement is an expression of support for former US President Donald Trump, and it does not convey a hopeful or positive message related to any of the mentioned social topics. The word “Trump” is often associated with controversial and divisive political opinions and actions, and expressing support for a particular politician does not necessarily convey a positive sentiment towards any of the social issues mentioned.*

task such as Hope Speech Detection. We also evaluated the impact of prompting scenarios on ChatGPT’s predictions and examined the ways in which ChatGPT can be influenced and prejudiced by sensitive words. These words have the potential to be manipulated in a harmful manner through data poisoning, thereby directing ChatGPT’s responses. The experiments were conducted on two datasets from different social networks with informal languages proposed by IberLEF 2023 Task, including Spanish tweets and English Youtube comments. Overall, prompting with ChatGPT outperforms other approaches when applying to the less-resourced Spanish tweets while monolingual XLMR surpassed other performances in rich-resourced English comments.

Acknowledgments

The authors express their gratitude for the 10-year friendship that they have shared, which has been a continuous source of motivation and inspiration for this paper. This friendship has not only made our research work enjoyable but also enriched and created HOPE in our personal lives. We are thankful to each other for the unwavering support and collaborative efforts that have culminated in this work. The author also would like to express their gratitude to their beloved kitty, Đô Đô, for being an absolute rockstar throughout the writing of this paper - even if it meant using the laptop as a comfy throne. Purrfect job, Đô Đô!

References

- [1] S. M. Jiménez-Zafra, F. Rangel, M. Montes-y Gómez, Overview of IberLEF 2023: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS.org, 2023.
- [2] B. R. Chakravarthi, Hopeedi: A multilingual hope speech detection dataset for equality, di-

- versity, and inclusion, in: Proceedings of the Third Workshop on Computational Modeling of People’s Opinions, Personality, and Emotion’s in Social Media, 2020, pp. 41–53.
- [3] P. Burnap, W. Colombo, J. Scourfield, Machine classification and analysis of suicide-related communication on twitter, in: Proceedings of the 26th ACM conference on hypertext & social media, 2015, pp. 75–84.
- [4] D. N. Milne, G. Pink, B. Hachey, R. A. Calvo, Clpsych 2016 shared task: Triaging content in online peer-support forums, in: Proceedings of the third workshop on computational linguistics and clinical psychology, 2016, pp. 118–127.
- [5] S. M. Jiménez-Zafra, M. Á. García-Cumbreras, D. García-Baena, J. A. García-Díaz, B. R. Chakravarthi, R. Valencia-García, L. A. Ureña-López, Overview of HOPE at IberLEF 2023: Multilingual Hope Speech Detection, *Procesamiento del Lenguaje Natural* 71 (2023).
- [6] D. García-Baena, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-García, Hope speech detection in spanish: The lgbt case, *Language Resources and Evaluation* (2023) 1–28.
- [7] S. Palakodety, A. R. KhudaBukhsh, J. G. Carbonell, Hope speech detection: A computational analysis of the voice of peace, 2020. [arXiv:1909.12940](https://arxiv.org/abs/1909.12940).
- [8] A. Hande, R. Priyadharshini, A. Sampath, K. P. Thamburaj, P. Chandran, B. R. Chakravarthi, Hope speech detection in under-resourced kannada language, *ArXiv abs/2108.04616* (2021).
- [9] E. Hossain, O. Sharif, M. M. Hoque, NLP-CUET@LT-EDI-EACL2021: Multilingual code-mixed hope speech detection using cross-lingual representation learner, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Kyiv, 2021, pp. 168–174. URL: <https://aclanthology.org/2021.ltedi-1.25>.
- [10] P. Aggarwal, P. Chandana, J. Nemade, S. Sharma, S. Saumya, S. Biradar, Hope speech detection on social media platforms, 2022. [arXiv:2212.07424](https://arxiv.org/abs/2212.07424).
- [11] T. S, R. t. Tasubilli, K. Sai rahul, Amrita@LT-EDI-EACL2021: Hope speech detection on multilingual text, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Kyiv, 2021, pp. 149–156. URL: <https://aclanthology.org/2021.ltedi-1.22>.
- [12] K. Mahajan, E. Al-Hossami, S. Shaikh, TeamUNCC@LT-EDI-EACL2021: Hope speech detection using transfer learning with transformers, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Kyiv, 2021, pp. 136–142. URL: <https://aclanthology.org/2021.ltedi-1.20>.
- [13] K. Puranik, A. Hande, R. Priyadharshini, S. Thavareesan, B. R. Chakravarthi, IIIT@LT-EDI-EACL2021-hope speech detection: There is always hope in transformers, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Kyiv, 2021, pp. 98–106. URL: <https://aclanthology.org/2021.ltedi-1.13>.
- [14] N. Ghanghor, R. Ponnusamy, P. K. Kumaresan, R. Priyadharshini, S. Thavareesan, B. R. Chakravarthi, IIITK@LT-EDI-EACL2021: Hope speech detection for equality, diversity, and inclusion in Tamil , Malayalam and English, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion, Association for Computational Linguistics, Kyiv, 2021, pp. 197–203. URL: <https://aclanthology.org/2021.ltedi-1.30>.
- [15] M. Song, H. Jiang, S. Shi, S. Yao, S. Lu, Y. Feng, H. Liu, L. Jing, Is chatgpt a good keyphrase

- generator? a preliminary study, 2023. [arXiv:2303.13001](#).
- [16] H. Wu, W. Wang, Y. Wan, W. Jiao, M. Lyu, Chatgpt or grammarly? evaluating chatgpt on grammatical error correction benchmark, 2023. [arXiv:2303.13648](#).
 - [17] B. Guo, X. Zhang, Z. Wang, M. Jiang, J. Nie, Y. Ding, J. Yue, Y. Wu, How close is chatgpt to human experts? comparison corpus, evaluation, and detection, 2023. [arXiv:2301.07597](#).
 - [18] O. Aydın, E. Karaarslan, Openai chatgpt generated literature review: Digital twin in healthcare, *SSRN Electronic Journal* (2022). doi:10.2139/ssrn.4308687.
 - [19] Y. Tan, D. Min, Y. Li, W. Li, N. Hu, Y. Chen, G. Qi, Evaluation of chatgpt as a question answering system for answering complex questions, 2023. [arXiv:2303.07992](#).
 - [20] R. Omar, O. Mangukiya, P. Kalnis, E. Mansour, Chatgpt versus traditional question answering for knowledge graphs: Current status and future directions towards knowledge graph chatbots, 2023. [arXiv:2302.06466](#).
 - [21] N. Bian, X. Han, L. Sun, H. Lin, Y. Lu, B. He, Chatgpt is a knowledgeable but inexperienced solver: An investigation of commonsense problem in large language models, 2023. [arXiv:2303.16421](#).
 - [22] J. Kocoń, I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, J. Bielaniewicz, M. Gruza, A. Janz, K. Kanclerz, A. Kocoń, B. Koptyra, W. Mieleśczenko-Kowszewicz, P. Miłkowski, M. Oleksy, M. Piasecki, Łukasz Radliński, K. Wojtasik, S. Woźniak, P. Kazienko, Chatgpt: Jack of all trades, master of none, 2023. [arXiv:2302.10724](#).
 - [23] F. Huang, H. Kwak, J. An, Is chatgpt better than human annotators? potential and limitations of chatgpt in explaining implicit hate speech, *ArXiv abs/2302.07736* (2023).
 - [24] A. Pelicon, M. Pranjić, D. Miljković, B. Škrlić, S. Pollak, Zero-shot learning for cross-lingual news sentiment classification, *Applied Sciences* 10 (2020) 5993.
 - [25] H. T. H. Tran, M. Martinc, A. Doucet, S. Pollak, Can cross-domain term extraction benefit from cross-lingual transfer?, in: *Discovery Science: 25th International Conference, DS 2022, Montpellier, France, October 10–12, 2022, Proceedings*, Springer, 2022, pp. 363–378.
 - [26] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, *arXiv preprint arXiv:1911.02116* (2019).