FZZG at WILDRE-7: Fine-tuning Pre-trained Models for Code-mixed, Less-resourced Sentiment Analysis

Gaurish Thakkar, Marko Tadić, Nives Mikelić Preradović
Faculty of Humanities and Social Sciences, University of Zagreb
{gthakkar, marko.tadic, nmikelic}@ffzg.unizg.hr

Abstract

This paper describes our system used for a shared task on code-mixed, less-resourced sentiment analysis for Indo-Aryan languages. We are using the large language models (LLMs) since they have demonstrated excellent performance on classification tasks. In our participation in all tracks, we use *unsloth/mistral-7b-bnb-4bit* LLM for the task of code-mixed sentiment analysis. For track 1, we used a simple fine-tuning strategy on PLMs by combining data from multiple phases. Our trained systems secured first place in four phases out of five. In addition, we present the results achieved using several PLMs for each language.

Keywords: sentiment analysis, code-mixed, LLM, Indo-Aryan

1. Introduction

Expression of sentiment-bearing information is natural to humans. The information expressed can span a spectrum of positive, negative, neutral, and mixed connotations. Sentiment (Turney, 2002) plays a major role in the interaction of people through social media as a tool of expression. Social media has evolved as an effective tool for people to express their views and ideas on a wide range of issues (Alodat et al., 2023; Kapoor et al., 2018). The interaction of social media users around the world has led to numerous phenomena (Nasir Ansari and Khan, 2020). One of them is code-mixing, also called intra-sentential code switching or intra-sentential code alternation and it occurs when speakers use two or more languages below clause level within one social situation (Mónica et al., 2009). For instance, the phrase "**Superhit bahut Achcha**" translates to "*superhit very good*" in English. The phrase is written in Roman characters instead of Devanagari and incorporates terms from both English and Hindi. This example does not necessarily follow standard writing rules (Das and Gambäck, 2014), but it effectively demonstrates its amalgamating nature, it poses a significant problem to process this text as it contains language constructs borrowed from multiple languages. Therefore, it is important to develop systems that can handle these phenomena to better understand sentiment. The code-mixed dataset can be understood by individuals who understand both languages; hence, developing the system for modelling can be challenging.

The WILDRE-7 shared task was organised for language pairs and triplets of less-resourced closely related languages: Magahi-Hindi-English (Rani et al., 2024a), Maithili-Hindi (Rani et al., 2024b), Bangla-English-Hindi (Raihan et al., 2023), and Hindi-English. Each code-mixed comment or sentence in Magahi-Hindi-English and Hindi-English had been annotated with four sentiment labels (positive, negative, neutral or mixed). However, the Bangla-English-Hindi is labelled with only three sentiment labels (positive, negative, or neutral).

Our approach to the code-mixed sentiment classification is to use the entire data in a multilingual training setup to aid transfer-learning between languages. The multilingual training helps low-resourced languages owing to the sharing of features between instances of different languages (Schmidt et al., 2022; Alves et al., 2023; Thakkar et al., 2021). We explore three large language models with fine-tuning setups. We combine all the data from different phases into a single dataset and fine-tune two XLM-RoBERTa-based models (Conneau et al., 2020; Barbieri et al., 2022) and one quantized version of the Mistral-7b model (Jiang et al., 2023).

Our final submission for all the phases used supervised fine-tuning on the "unsloth/mistral-7b-bnb-4bit" model[1]. Our proposed model performed well in the Bangla-English code-mixed and combined code-mixed phases. In other phases, despite achieving the best scores compared to other participants, the performance for the relevant languages in the test set was below 0.50 F1.

2. Related Work

An initial investigation into the code switching phenomenon was conducted by Warschauer et al. (2002). They investigated the use of English and Arabic by a group of youthful professionals in email correspondence. It was discovered that English

[1] https://huggingface.co/unsloth/mistral-7b-bnb-4bit

WILDRE-7 Workshop, pages 59–65

was used more frequently in both formal (business-related) email exchanges and Internet searches.

Chittaranjan et al. (2014) employed word-level language identification in code-mixed texts, in which various characteristics were utilised to identify the language of a given word. Contextual features, capitalization features, special character features, and lexicon features were all implemented by the system. Annotated data is then utilised to train the CRF model. The authors attained results with high precision for the majority of language pairs. The accuracy was compromised when the distribution of languages in the test data differed from that of the training data.

Veríssimo dos Santos Neto et al. (2020) proposed, for the Semeval 2020 submission (shared task 9), a combination of four models predicated on the application of transfer learning and language models. The task required conducting sentiment analysis on code-mixed languages that combine English and Hindi. Ma et al. (2020) presented a novel approach in SemEval-2020 for sentiment analysis problem by utilising weighted loss of several multilingual models, with a specific emphasis on the difficulty of code-mixing phrases. The authors employed XLM models in conjunction with machine translation as a form of data augmentation.

3. System Overview

In this section, we describe the task, the different LLMs used, along with preprocessing steps and training configurations.

3.1. Task description

The task had two different evaluation tracks. Track 1 dealt with the classification of the polarity (positive, negative, neutral or mixed) of the comment in the code-mixed setting for the following phases.

1. Hindi-English
2. Magahi-Hindi-English
3. Bangla-English
4. Combined all the language pairs/triplets (1+2+3)

In Track 2, the task was to use the given unlabeled test data for the code-mixed Maithili language (Maithili-Hindi-English) and leverage any or all of the available training datasets in Track 1 to determine the sentiment of a comment in the target language. The dataset was divided into the train, validation and test sets with a ratio of 70:15:15. However, for the fourth part of Track 1 (combining all the language pairs), we combined the provided training and validation datasets of each code-mixed language to train the model.

3.2. Approach

We experimented with two approaches: supervised fine-tuning (Severyn and Moschitti, 2015) and instruction tuning (Efrat and Levy, 2020). Instruction tuning involves providing the model with a collection of instructions or prompts and subsequently modifying the model's parameters to enhance its performance on the tasks specified by these instructions. One way to do this is through the use of techniques such as reinforcement learning (Bai et al., 2022), in which the model receives rewards for behaviours that result in favourable outcomes, or gradient descent (Chen et al., 2022), in which the model's parameters are continuously modified to minimise a loss function.

The following insights served as the foundation for our instruction tuning strategy. For several benchmark datasets, the models (Touvron et al., 2023; Jiang et al., 2023) that were trained using instruction tuning were at the top of the Open LLM Leaderboard[2]. Given that the training cases in the competition were annotated at the sentence level, we concentrated on representing the problem as a single task classification problem without exploring other sub-tasks such as language identification and classification. Since the non-quantized version of Mistral requires extensive processing capabilities, we used the quantized version that can be effortlessly trained on a single GPU with 24 GB of memory.

3.3. Dataset

The organisers provided a dataset (Rani et al., 2024a) containing Magahi-Hindi-English and Hindi-English, which was collected from various YouTube channels and annotated with the help of native speakers of the language. For Bangla-English code-mixed data set 1, we are using the SentMix-3L dataset (Raihan et al., 2023). Table 1 shows the statistics of the provided dataset. In addition, we used SAIL 2017 (Patra et al., 2018), a Hindi code-mixed shared task dataset. In Table 2, the number of instances from the SAIL 2017 dataset is presented.

3.4. Pretrained language models (PLMs)

3.4.1. XLM-RoBERTa-base

XLM-RoBERTa (Conneau et al., 2020) is pre-trained on a vast text and code dataset, which includes BooksCorpus, Wikipedia, and the Pile. This

[2]https://tinyurl.com/3s3zfsu8

Phase	Pos	Neg	Neu	Mix
Ben-Eng	293	247	163	
Hin-Eng	1989	419	77	113
Mag-Hin	615	194	26	30
Total	2806	860	266	143

Table 1: Distribution of the dataset released by the organisers.

Split	Pos	Neg	Neu
train	3190	2312	4577
test	399	290	573

Table 2: Additional dataset used for training - SAIL 2017 (Patra et al., 2018)

pre-training technique combines language modelling with natural language task-specific cues, resulting in increased performance on a wide range of activities. It builds on RoBERTa's (Zhuang et al., 2021) great performance by offering new architectural advancements, such as larger model sizes and additional training data. This leads to improved accuracy and efficiency on many NLP tasks.

3.4.2. cardiffnlp/twitter-roberta-base-sentiment

Twitter-RoBERTa-base-sentiment[3] (Camacho-Collados et al., 2022; Loureiro et al., 2022) is a RoBERTa (Zhuang et al., 2021) model trained on ≈124M tweets from January 2018 to December 2021, and fine-tuned for sentiment analysis with the TweetEval benchmark (Barbieri et al., 2020).

3.4.3. unsloth/mistral-7b-bnb-4bit

The Mixtral-8x7B Large Language Model (LLM) is a pre-trained generative Sparse Mixture of Experts. The unsloth/mistral-7b-bnb-4bit model is quantized model of Mixtral-8x7B that has been saved as a LoRA (Hu et al., 2022) adapter through the Unsloth library[4]. The LoRA weights can be retrained during the fine-tuning phase. The model supports a maximum sequence length of 2048, which works optimally with larger contexts.

3.5. Data preparation

In order to generate the training set, we combine all of the code-mixed training sets. We also merge the validation sets of all the datasets provided as part of the competition to create a single validation set. In addition, we incorporate the SAIL 2017 (Patra et al., 2018) dataset as an additional resource into the training to increase the training data size for training the Hindi-English code-mixed model.

3.5.1. XLM-RoBERTa and Twitter-RoBERTa

No special format is required for fine-tuning the model other than tokenizing the dataset with the respective pre-trained tokenizer.

3.5.2. Mistral-7b model

The fine-tuning of the dataset is performed in the form of Instructions. We followed the Alpaca (Taori et al., 2023) dataset format and converted the dataset into the following format:

```
Instruction: Classify the given
article as either positive or
negative or neutral or mix sentiment.

alpaca_prompt = """Below is an
instruction that describes a
task, paired with an input that
provides further context. Write a
response that appropriately
completes the request.

### Instruction:
{Classify the given article as either
positive or negative or neutral
or mix sentiment}

### Input:
{Ekdam sahi bat bolalahi bhaiya}

### Response:
{positive}"""
```

The sentence "**Ekdam sahi bat bolalahi bhaiya**" (hi-en) can be translated to "*You said the right thing brother*" (en). The expected input to the LLM is a single tuple consisting of a prompt, instruction, input, and response. The prompt was the description of the task, the instruction was set to the classification of the text, the input was defined as the code-mixed text, and the response was the expected sentiment label.

4. Experimental Setup

4.1. Fine-tuning

For fine-tuning XLMR models, we used a learning-rate of $5e^{-5}$ with a batch size of 16 and a maximum sequence length of 512. We trained for a maximum of 16 epochs with early stopping and a patience of 3 on the validation set. We used the weighted cross-entropy loss to handle the class imbalance.

[3]cardiffnlp/twitter-roberta-base-sentiment
[4]https://github.com/unslothai/unsloth

4.2. Instruction tuning

For instruction tuning (Efrat and Levy, 2020; Mishra et al., 2022), we used a batch size of 8 and a gradient accumulation of 2. The learning rate was set to $2e^{-5}$ after a few trials. We used the maximum sequence length of 2048. An early stopping mechanism based on a validation set was used to prevent model overfitting.

5. Results

Table 3 presents the initial experiments conducted with the XLM-RoBERTa models. We found that the XLM-RoBERTa performed better than Twitter-RoBERTa, even though Twitter-RoBERTa is trained with Twitter data. The evaluation scores on the target language validation set when using unsloth/mistral-7b-bnb-4bit were better compared to XLM-RoBERTa models.

Model	Eval-F1
XLM-RoBERTa	0.60
Twitter-RoBERTa	0.54

Table 3: Evaluation F-1 scores.

Tr	Phase	F1	P	R
1	Ben-Eng (all)	0.97	0.97	0.97
1	Hin-Eng (all)	0.43	0.50	0.44
	Hin-Eng (Hi+SAIL)	0.44	0.48	0.43
	Hin-Eng (Hi)	0.54	0.54	0.56
1	Mag-Hin-Eng (all)	0.45	0.44	0.57
1	Combined (all)	0.60	0.64	0.57
2	Mai (Hi+SAIL)	0.49	0.45	0.59

Table 4: Final scores reported by the submission system. The scores are reported using predictions obtained using 'unsloth/mistral-7b-bnb-4bit'. The first column (Tr) denotes the track's task number. 'all': A combined training set from the shared task was used for training.

In Table 4, we present the results for the instruction tuning experiments. The model, trained using a combined training dataset, demonstrated strong performance on the test set for Bangla-English, Magahi-Hindi-English, and in combination code-mixed setting. The model achieved higher scores in the Hindi-English test case when exclusively trained on Hindi-English cases. We also attempted alternative combinations, but none of them yielded superior results compared to only using the data instances given as part of the shared task. The findings align with prior research (Thakkar et al., 2021, 2023) indicating that including data from comparable languages with a larger number of training instances improves performance in the case of lower-resourced languages. However, when data instances from lower-resourced languages are combined with higher-resourced languages, there is a decrease in performance for the latter. The combination of the SAIL dataset with Hindi-English training examples was found to be effective combination for training the model to be tested on Hindi-English and Maithili test set.

6. Conclusion

This paper describes the proposed model used for a shared task on code-mixed, less-resourced sentiment analysis for Indo-Aryan languages. We experimented with PLM-based XLM-Roberta and a customised version of Mistral-7b to model the task of code-mixed sentiment. Our analysis shows that code-mixed, less-resourced sentiment analysis for Indo-Aryan languages is a difficult task for the PLMs, and there is scope for further improvements that we will take up in future works. For future work, we would like to use other available code-mixed datasets to improve the performance of sentiment analysis systems in code-mixed settings.

7. Acknowledgements

This work was partially funded by the European Union under the grant agreements No. LC-01641480 – 101018166 (ELE) and No. LC-01884166 – 101075356 (ELE 2). This work was partially funded from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement No 101070631 and from the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (Grant No 10039436).

8. Bibliographical References

Abdelsalam M. Alodat, Lamis F. Al-Qora'n, and Muwafaq Abu Hamoud. 2023. Social Media Platforms and Political Participation: A Study of Jordanian Youth Engagement. *Social Sciences*, 12(7):1–19.

Diego Alves, Božo Bekavac, Daniel Zeman, and Marko Tadić. 2023. Corpus-based syntactic typological methods for dependency parsing improvement. In *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 76–88, Dubrovnik, Croatia. Association for Computational Linguistics.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn

Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, Nicholas Joseph, Saurav Kadavath, Jackson Kernion, Tom Conerly, Sheer El Showk, Nelson Elhage, Zac Hatfield-Dodds, Danny Hernandez, Tristan Hume, Scott Johnston, Shauna Kravec, Liane Lovitt, Neel Nanda, Catherine Olsson, Dario Amodei, Tom B. Brown, Jack Clark, Sam McCandlish, Chris Olah, Benjamin Mann, and Jared Kaplan. 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback. *CoRR*, abs/2204.05862.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. 2020. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online. Association for Computational Linguistics.

Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. 2022. XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 258–266, Marseille, France. European Language Resources Association.

Jose Camacho-Collados, Kiamehr Rezaee, Talayeh Riahi, Asahi Ushio, Daniel Loureiro, Dimosthenis Antypas, Joanne Boisson, Luis Espinosa Anke, Fangyu Liu, Eugenio Martínez Cámara, et al. 2022. TweetNLP: Cutting-edge natural language processing for social media. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–49, Abu Dhabi, UAE. Association for Computational Linguistics.

Yanda Chen, Ruiqi Zhong, Sheng Zha, George Karypis, and He He. 2022. Meta-learning via language model in-context tuning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 719–730, Dublin, Ireland. Association for Computational Linguistics.

Gokul Chittaranjan, Yogarshi Vyas, Kalika Bali, and Monojit Choudhury. 2014. Word-level language identification using CRF: Code-switching shared task report of MSR India system. In *Proceedings of the First Workshop on Computational Approaches to Code Switching*, pages 73–79, Doha, Qatar. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Amitava Das and Björn Gambäck. 2014. Identifying languages at the word level in code-mixed Indian social media text. In *Proceedings of the 11th International Conference on Natural Language Processing*, pages 378–387, Goa, India. NLP Association of India.

Avia Efrat and Omer Levy. 2020. The turking test: Can language models understand instructions? *CoRR*, abs/2010.11982.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *CoRR*, abs/2310.06825.

Kawal Kapoor, Kuttimani Tamilmani, Nripendra Rana, Pushp Patil, Yogesh Dwivedi, and Sridhar Nerur. 2018. Advances in social media research: Past, present and future. *Information Systems Frontiers*, 20:531–558.

Daniel Loureiro, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke, and Jose Camacho-collados. 2022. TimeLMs: Diachronic language models from Twitter. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 251–260, Dublin, Ireland. Association for Computational Linguistics.

Yili Ma, Liang Zhao, and Jie Hao. 2020. XLP at SemEval-2020 task 9: Cross-lingual models with focal loss for sentiment analysis of code-mixing language. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 975–980, Barcelona (online). International Committee for Computational Linguistics.

Swaroop Mishra, Daniel Khashabi, Chitta Baral, and Hannaneh Hajishirzi. 2022. Cross-task generalization via natural language crowdsourcing instructions. In *Proceedings of the 60th Annual*

Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 3470–3487, Dublin, Ireland. Association for Computational Linguistics.

Stella Mónica, Mónica Cárdenas-Claros, and Neny Isharyanti. 2009. Code switching and code mixing in internet chating: betwen "yes", "ya", and "si" a case study. *The jaltcall Journal*, Vol 5:67–78.

Jamal Nasir Ansari and Nawab Khan. 2020. Exploring the role of social media in collaborative learning the new domain of learning. *Smart Learning Environments*, 7(1):9.

Braja Gopal Patra, Dipankar Das, and Amitava Das. 2018. Sentiment analysis of code-mixed indian languages: An overview of sail_code-mixed shared task @icon-2017.

Md Nishat Raihan, Dhiman Goswami, Antara Mahmud, Antonios Anastasopoulos, and Marcos Zampieri. 2023. SentMix-3L: A novel code-mixed test dataset in Bangla-English-Hindi for sentiment analysis. In *Proceedings of the First Workshop in South East Asian Language Processing*, pages 79–84, Nusa Dua, Bali, Indonesia. Association for Computational Linguistics.

Priya Rani, Gaurav Negi, Theodorus Fransen, and John P. McCrae. 2024a. Macms: Magahi code-mixed dataset for sentiment analysis.

Priya Rani, Gaurav Negi, Saroj Jha, Shardul Suryawanshi, Atul Kr. Ojha, Paul Buitelaar, and John P. McCrae. 2024b. Findings of the wildre shared task on code-mixed less-resourced sentiment analysis for indo-aryan languages. In *Proceedings of the 7th Workshop on Indian Language Data: Resources and Evaluation @LREC-COLING-2024 (WILDRE-7)*, Turin, Italy. ELRA Language Resources Association (ELRA) and the International Committee on Computational Linguistics (ICCL).

Fabian David Schmidt, Ivan Vulić, and Goran Glavaš. 2022. Don't stop fine-tuning: On training regimes for few-shot cross-lingual transfer with multilingual language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10725–10742, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Aliaksei Severyn and Alessandro Moschitti. 2015. UNITN: Training deep convolutional neural network for Twitter sentiment classification. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 464–469, Denver, Colorado. Association for Computational Linguistics.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford alpaca: An instruction-following llama model. `https://github.com/tatsu-lab/stanford_alpaca`.

Gaurish Thakkar, Nives Mikelić Preradović, and Marko Tadić. 2021. Multi-task learning for cross-lingual sentiment analysis. In *Proceedings of the 2nd International Workshop on Cross-lingual Event-centric Open Analytics co-located with the 30th The Web Conference (WWW 2021), Ljubljana, Slovenia, April 12, 2021*, volume 2829 of *CEUR Workshop Proceedings*, pages 76–84. CEUR-WS.org.

Gaurish Thakkar, Nives Mikelić Preradović, and Marko Tadić. 2023. CroSentiNews 2.0: A Sentence-Level news sentiment corpus. In *Human Language Technologies as a Challenge for Computer Science and Linguistics - 2023*, pages 294–299, Poznan. Adam Mickiewicz University Press.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models.

Peter Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Manoel Veríssimo dos Santos Neto, Ayrton Amaral, Nádia Silva, and Anderson da Silva Soares.

2020. Deep learning Brasil - NLP at SemEval-2020 task 9: Sentiment analysis of code-mixed tweets using ensemble of language models. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1233–1238, Barcelona (online). International Committee for Computational Linguistics.

Mark Warschauer, Ghada R El Said, and Ayman G Zohry. 2002. Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication*, 7(4):744.

Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. 2021. A robustly optimized BERT pre-training approach with post-training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.