TransMistral at SemEval-2024 Task 10: Using Mistral 7B for Emotion Discovery and Reasoning its Flip in Conversation

Marco Siino
Department of Electrical, Electronic
and Computer Engineering
University of Catania
Italy
marco.siino@unipa.it

Abstract

The EDiReF shared task at SemEval 2024 comprises three subtasks: Emotion Recognition in Conversation (ERC) in Hindi-English code-mixed conversations, Emotion Flip Reasoning (EFR) in Hindi-English code-mixed conversations, and EFR in English conversations. The objectives for the ERC and EFR tasks are defined as follows: 1) Emotion Recognition in Conversation (ERC): In this task, participants are tasked with assigning an emotion to each utterance within a dialogue from a predefined set of possible emotions. The goal is to accurately recognize and label the emotions expressed in the conversation; 2) Emotion Flip Reasoning (EFR): This task involves identifying the trigger utterance(s) for an emotion-flip within a multi-party conversation dialogue. Participants are required to pinpoint the specific utterance(s) that serve as catalysts for a change in emotion during the conversation. In this paper we only address the first subtask (ERC) making use of an online translation strategy followed by the application of a Mistral 7B model together with a few-shot prompt strategy. Our approach obtains an F1 of 0.36, eventually exhibiting further room for improvements.

1 Introduction

Affective computing has experienced a resurgence, largely propelled by recent advancements in artificial intelligence. Emotion Recognition in Conversations (ERC) has emerged as a prominent task within affective computing, garnering increasing attention (Poria et al., 2019; Kumar et al., 2023). Its objective is to discern the emotion conveyed in each utterance during conversations, with implications for various applications including the development of effective dialogue systems, facilitating social viewpoint mining, and creating intelligent medical systems. Current research in ERC primarily focuses on capturing the emotional state of speakers through contextual analysis and establishing distinct contexts for different speakers, often leveraging multimodal data to support this endeavour. Despite recent strides, two major challenges persist: (1) Ensuring emotional consistency and (2) Generating contextual information. Current research efforts broadly fall into two categories: the first involves obtaining contextual representations of utterances using temporal neural networks, while the second entails capturing long-distance information through graph networks. However, these approaches overlook a crucial aspect: changes in utterance order can alter the meaning of utterances, potentially leading to varying emotional expressions. A shift in utterance order impacts the underlying meaning of the utterance, consequently influencing the speakers' emotions.

The increasing demand for automated tools capable of extracting and categorizing data from online sources underscores the need to address both established and emerging societal concerns efficiently. Recent strides in machine and deep learning architectures have sparked significant interest in Natural Language Processing (NLP). Efforts have been intensified towards developing techniques for automating the identification and categorization of textual content prevalent on the internet today. In the literature, various strategies have been proposed for performing text classification tasks. Over the past fifteen years, some of the most successful strategies have included Support Vector Machines (SVM) (Colas and Brazdil, 2006; Croce et al., 2022), Convolutional Neural Networks (CNN) (Kim, 2014; Siino et al., 2021), Graph Neural Networks (GNN) (Lomonaco et al., 2022), ensemble models (Miri et al., 2022; Siino et al., 2022), and more recently, Transformers (Vaswani et al., 2017; Siino et al., 2022b).

At SemEval-2024 Task 10 (Kumar et al., 2024a) – Emotion Discovery and Reasoning its Flip in Conversation (EDiReF) – three Subtasks were proposed. All the three subtasks are presented and

Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 298–304
June 20-21, 2024 ©2024 Association for Computational Linguistics

discussed in the Section 2.

To face with the first subtask (ERC), we proposed a Transformer-based approach which made use of Mistral 7B (Jiang et al., 2023). We used the model in a particular few-shot way described in the rest of this paper. Specifically, after translating all the code-mixed samples in English, we provided the samples from the labelled train and dev set to the model, asking while prompting to predict the emotion for the current utterance (i.e., the current sample from the test set).

The rest of the paper is made as follows. In Section 2 we provide some background on the Task 10 hosted at SemEval-2024. In Section 3 we provide a description of the approach presented. In Section 4 we provide details about the experimental setup to replicate our work. In Section 5, the results of the official task and some discussions are provided. In section 6 we present our conclusion and proposals for future works.

We make all the code publicly available and reusable on GitHub[1].

2 Background

This section furnishes background information regarding Task 10 (Subtask 1), held at SemEval-2024.

The EDiReF shared task at SemEval 2024 (Kumar et al., 2024a) encompasses three distinct subtasks, namely: Emotion Recognition in Conversation (ERC) within Hindi-English code-mixed conversations, Emotion Flip Reasoning (EFR) within Hindi-English code-mixed conversations, and EFR within English conversations (Kumar et al., 2022, 2024b). The ERC task involves the assignment of emotions to each utterance within a dialogue, drawn from a predefined set of potential emotions. Conversely, the EFR task focuses on identifying the trigger utterance(s) responsible for inducing an emotion-flip within a multi-party conversation dialogue.

In the Figure 1, is reported a sample from the official competition website[2] and specifically related to the Subtask 1 (i.e., ERC).

For Subtask 1, a submission entails a singular JSON file with each emotion in a new line. Every emotion correspond to each utterance in the official provided test set.

The second and the third subtasks differ on the

Speaker	Utterance
Sp_1	Aaj to bhot awful day tha! *(I had an awful day today!)*
Sp_2	Oh no! Kya hua? *(Oh no! What happened?)*
Sp_1	Kisi ne mera sandwich kha liya! *(Somebody ate my sandwich!)*
Sp_2	Me abhi tumhare liye new bana deti hun! *(I can make you a new one right now!)*
Sp_1	Wo great hoga! Thanks! *(That would be great! Thanks!)*

ERC aims to assign emotions to each utterance

Speaker	Utterance	Emotion
Sp_1	Aaj to bhot awful day tha! *(I had an awful day today!)*	Sad
Sp_2	Oh no! Kya hua? *(Oh no! What happened?)*	Sad
Sp_1	Kisi ne mera sandwich kha liya! *(Somebody ate my sandwich!)*	Sad
Sp_2	Me abhi tumhare liye new bana deti hun! *(I can make you a new one right now!)*	Joy
Sp_1	Wo great hoga! Thanks! *(That would be great! Thanks!)*	Joy

The example dialogue has two emotion flips. Sp_1's emotion changed from Sad to Joy while Sp_2's emotion also shifted from Sad to joy. EFR aims to justify such emotion flips using triggers.

Figure 1: Example of some samples from the dataset. In this case, we report samples related to the first ERC task in which we took part.

Speaker	Utterance	Emotion	Trigger
Sp_1	Aaj to bhot awful day tha! *(I had an awful day today!)*	Sad	0
Sp_2	Oh no! Kya hua? *(Oh no! What happened?)*	Sad	0
Sp_1	Kisi ne mera sandwich kha liya! *(Somebody ate my sandwich!)*	Sad	0
Sp_2	Me abhi tumhare liye new bana deti hun! *(I can make you a new one right now!)*	Joy	1
Sp_1	Wo great hoga! Thanks! *(That would be great! Thanks!)*	Joy	0

Figure 2: Example of some samples from the dataset. In this case, we report samples related to the EFR task.

language used. In one case, the language is code-mixed Hindi-English and in another case only in English. In both cases, the participants were asked to propose automatic detection systems able to detect a trigger utterance that determined a changed in the emotion. Also in this case, an example from the official task webpage is reported in the Figure 2. The EFR sample contains a trigger (i.e., 1) in proximity of the fourth sentence contained in the dialogue.

3 System Overview

While it has been established that Transformers may not always be the optimal choice for text classification tasks (Siino et al., 2022a), the efficacy of various strategies, such as domain-specific fine-tuning (Sun et al., 2019; Van Thin et al., 2023) and data augmentation (Lomonaco et al., 2023; Mangione et al., 2022), depends on the specific objectives.

The increasing adoption of Transformer-based architectures in academic research has also been bolstered by various methodologies showcased at SemEval 2024. These methodologies tackle di-

[1] https://github.com/marco-siino/SemEval2024/
[2] https://lcs2.in/SemEval2024-EDiReF/

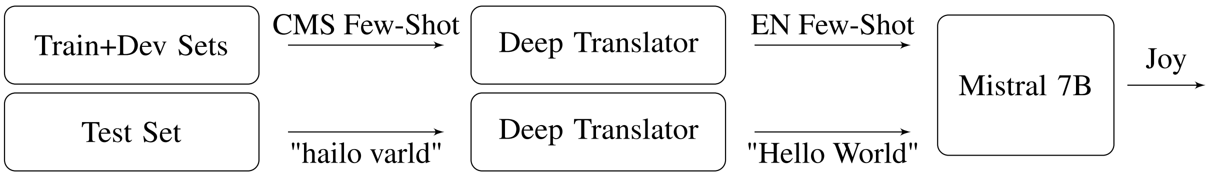

Figure 3: The system overview of our proposed approach. Given a set of Code Mixed Samples (i.e., CMS) from the train and dev sets, they are translated to English using Deep Translator (i.e., ENglish samples). Then they are all provided as input - a few-shot samples from the training set, together with the emotion definitions - to Mistral 7B. Following these few shot samples and the definitions as input, there is one utterance from the test set for which the emotion has to be predicted.

verse tasks and yield noteworthy findings. For instance, at the Task 2 (Jullien et al., 2024), where to address the challenge of identifying the inference relation between a plain language statement and Clinical Trial Reports is used T5 (Siino, 2024c); Task 4 (Dimitrov et al., 2024) where is employed a Mistral 7B model to detect persuasion techniques in memes (Siino, 2024b); and Task 8 (Wang et al., 2024), that utilizes a DistilBERT model to identify machine-generated text (Siino, 2024a).

Our approach is few-shot (Littenberg-Tobias et al., 2022) and make use of Mistral 7B. Mistral 7B, a language model boasting 7 billion parameters, is engineered to excel in both performance and efficiency. In comparison to the leading open 13B model (Llama 2), Mistral 7B demonstrates superior performance across all assessed benchmarks. Furthermore, it surpasses the top released 34B model (Llama 1) in tasks related to reasoning, mathematics, and code generation. The model capitalizes on grouped-query attention (GQA) to expedite inference, complemented by sliding window attention (SWA) to effectively process sequences of varying lengths while minimizing inference costs. Additionally, a fine-tuned variant, Mistral 7B – Instruct, is tailored for adhering to instructions, and it outperforms Llama 2 13B – chat model across both human and automated benchmarks. The introduction of Mistral 7B Instruct underscores the ease with which the base model can be fine-tuned to achieve notable performance enhancements.

For our task, before prompting the model with the current sample from the test set, we made an online and real-time use of *Google Translator* from the *deep_translator*[3] library. Then we randomly selected eighty samples from the provided labelled training set and other eighty from the provided labelled dev set. Then we formatted the samples in each set in the following way:

[3]https://pypi.org/project/deep-translator/

EMOTIONS
1. *disgust*
2. *joy*
3. *neutral*
4. *anger*
5. *sadness*
6. *contempt*
7. *surprise*
8. *fear*

Table 1: The list of all the motions available for the task.

speaker1 - utterance1 - emotion1
speaker2 - utterance2 - emotion2
...
speakerX - utterance80 - emotionY

After merging the formatted samples from both the training and the dev set, we fed the model, appending to the few-shot samples the current unlabelled sample from the official test set. At this point, the full text containing the few-shot samples plus the sample to be classified were provided as prompts to Mistral.

Then the question provided as prompt to the model was: *" Use the CONTEXT to complete the SENTENCE using ONLY one emotion among: disgust, joy, neutral, anger, sadness, contempt, surprise, fear. Do not explain!"*. Where the *CONTEXT* were the few-shot samples provided. For all the samples from the test set the model correctly predicted one of the available emotions from the list provided and shown in the Table 1.

As noted in the recent study by (Siino et al., 2024b), the contribution of preprocessing for text classification tasks is generally not impactful when using Transformers. More specifically, the best combination of preprocessing strategies is not very different from doing no preprocessing at all in the case of Transformers. For these reasons, and to keep our system fast and computationally light, we have not performed any preprocessing on the text.

4 Experimental Setup

Our model implementation was executed on Google Colab, utilizing the Mistral 7B library from Hugging Face, specifically the Mistral-7B-Instruct-v0.2-GGUF [4] version. Additionally, we utilized the *deep_translator* package with Google Translator [5] for the translation task. The Mistral 7B version employed represents an enhanced iteration of the Mistral-7B-Instruct-v0.1 model, geared towards instruction fine-tuning. Instructions for instruction fine-tuning should be enclosed within [INST] and [/INST] tokens, with the initial instruction beginning with a sentence identifier, and subsequent instructions omitting this identifier. The generation process is terminated by the end-of-sentence token ID. Furthermore, we imported the Llama library (Touvron et al., 2023) from *llama_cpp*, with comprehensive details available on GitHub [6].

All datasets required for the various phases of the experiment are accessible on the Official Competition page. No additional fine-tuning was conducted on the model. The experiment was executed using a T4 GPU provided by Google. Upon generating the predictions, the results were exported in the format specified by the organizers. As previously mentioned, our complete codebase is accessible on GitHub.

5 Results

As described on the official task page[7], the evaluation criteria for the three tasks are delineated as follows:

- Task 1 (ERC for code-mixed): Weighted F1 score computed across all emotions.
- Task 2 (EFR for code-mixed): F1 score computed specifically for triggers (label '1.0').
- Task 3 (EFR for English): F1 score computed for triggers (label '1.0') in English.

To generate the prediction file:

	T1-F1	T2-F1	T3-F1
TransMistral 7B	0.36	0.10	0.22

Table 2: The method's performance on the test set. Even if we did not participate in the tasks 2 and 3, in the *answer* file we included a list of 0s to complete all the lines as suggested by the task organizers.

1. For Task 1: Each line should depict an emotion associated with an utterance, with no additional lines separating dialogues. Each emotion should be in lowercase, devoid of extra spaces.

2. For Tasks 2 and 3: Each line should represent either 0.0 or 1.0, reflecting the label of triggers assigned to an utterance. The formatting should adhere to a string format with a floating precision of 1 (e.g., 0.0 or 1.0, rather than 0 or 1).

Then, it was asked to aggregate the outputs from all tasks into a single file named 'answer.txt', structured as follows:

- Lines 1–1580: Predictions for Task 1.
- Lines 1581–9270: Predictions for Task 2.
- Lines 9271–17912: Predictions for Task 3.

Upon compilation, the organizers asked to create a zip file encompassing *'answer.txt'* and proceed with submission as per guidelines.

In the Table 2 we report the result obtained by the proposed approach on the official test set. Thanks to our application of an online translation followed by a Mistral 7B we have been able to reach the twenty-second position on the final ranking for the evaluation phase.

The Table 3 presents the performance outcomes achieved by the top three teams and the last-ranked team, as delineated on the official task page. While our simplistic approach showcases potential for enhancement in comparison to the leading models, it is noteworthy that our method necessitated no additional pre-training. Moreover, the computational resources utilized to tackle the task remained feasible, courtesy of the complimentary resources provided by Google Colab.

[4] https://huggingface.co/TheBloke/Mistral-7B-Instruct-v0.2-GGUF
[5] https://pypi.org/project/deep-translator/
[6] https://github.com/ggerganov/llama.cpp
[7] https://codalab.lisn.upsaclay.fr/competitions/16769

TEAM NAME	T1-F1	T2-F1	T3-F1
MasonTigers	0.78 (1)	0.79 (2)	0.79 (1)
Knowdee	0.73 (2)	0.66 (4)	0.61 (9)
IASBS	0.70 (3)	0.12 (7)	0.25 (12)
GAVx	0.08 (34)	0.79 (2)	0.76(2)

Table 3: Comparing performance on the test set. In the table are shown the results obtained by the first three users and by the last one ordered considering the first task. In parentheses is reported the position for each task in the official final ranking.

6 Conclusion

This paper presents the application of Mistral 7B-model for addressing the Task 10 at SemEval-2024. For our submission, we decided to follow few-shot learning approach, employing as-is, an in-domain pre-trained Transformer. After several experiments, we found it beneficial to build a prompt containing some samples from the training and from the dev set. Then we provide as a prompt the current translated sample together with the few-shot samples. The model was asked to select one of the emotion among the ones available. The task presents inherent challenges, with evident scope for refinement, as underscored by the final ranking. Potential alternative methodologies encompass leveraging the zero-shot capabilities inherent in other models such as GPT and T5, expanding the training set size through the incorporation of additional data, or adopting alternative strategies for integrating ontology-based domain knowledge beyond the approaches delineated in our study. Furthermore, refinement opportunities exist through fine-tuning and recontextualizing the problem as a text classification task.

Furthermore, given the interesting results recently provided on a plethora of tasks, also other few-shot learning (Wang et al., 2023; Maia et al., 2024; Siino et al., 2023; Meng et al., 2024) or data augmentation strategies (Muftie and Haris, 2023; Siino et al., 2024a; Tapia-Téllez and Escalante, 2020; Siino and Tinnirello, 2023) could be employed to improve the results. Looking at the final ranking, our simple approach exhibits some room for improvements. However, it is worth notice that it required no further pre-training and the computational cost to address the task is manageable with the free online resources offered by Google Colab. Also, thanks to the proposed approach, we have been able to outperform the baseline provided by the task organizers.

Acknowledgments

We express our sincere appreciation to the anonymous reviewers for their constructive feedback and invaluable suggestions. Their insightful comments have greatly contributed to the refinement and clarity of this paper.

References

Fabrice Colas and Pavel Brazdil. 2006. Comparison of svm and some older classification algorithms in text classification tasks. In *IFIP International Conference on Artificial Intelligence in Theory and Practice*, pages 169–178. Springer.

Daniele Croce, Domenico Garlisi, and Marco Siino. 2022. An SVM ensemble approach to detect irony and stereotype spreaders on twitter. In *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022*, volume 3180 of *CEUR Workshop Proceedings*, pages 2426–2432. CEUR-WS.org.

Dimitar Dimitrov, Firoj Alam, Maram Hasanain, Abul Hasnat, Fabrizio Silvestri, Preslav Nakov, and Giovanni Da San Martino. 2024. Semeval-2024 task 4: Multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation*, SemEval 2024, Mexico City, Mexico.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Maël Jullien, Marco Valentino, and André Freitas. 2024. SemEval-2024 task 2: Safe biomedical natural language inference for clinical trials. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Association for Computational Linguistics.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29,* *2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1746–1751. ACL.

Shivani Kumar, Md Shad Akhtar, Erik Cambria, and Tanmoy Chakraborty. 2024a. Semeval 2024 – task 10: Emotion discovery and reasoning its flip in conversation (ediref). In *Proceedings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics.

Shivani Kumar, Shubham Dudeja, Md Shad Akhtar, and Tanmoy Chakraborty. 2024b. Emotion flip reasoning in multiparty conversations. *IEEE Transactions on Artificial Intelligence*, 5(3):1339–1348.

Shivani Kumar, Ramaneswaran S, Md Akhtar, and Tanmoy Chakraborty. 2023. From multilingual complexity to emotional clarity: Leveraging commonsense to unveil emotions in code-mixed dialogues. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9638–9652, Singapore. Association for Computational Linguistics.

Shivani Kumar, Anubhav Shrimal, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. Discovering emotion and reasoning its flip in multi-party conversations using masked memory network and transformer. *Knowledge-Based Systems*, 240:108112.

Joshua Littenberg-Tobias, G. R. Marvez, Garron Hillaire, and Justin Reich. 2022. Comparing few-shot learning with GPT-3 to traditional machine learning approaches for classifying teacher simulation responses. In *AIED (2)*, volume 13356 of *Lecture Notes in Computer Science*, pages 471–474. Springer.

Francesco Lomonaco, Gregor Donabauer, and Marco Siino. 2022. COURAGE at checkthat!-2022: Harmful tweet detection using graph neural networks and ELECTRA. In *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022*, volume 3180 of *CEUR Workshop Proceedings*, pages 573–583. CEUR-WS.org.

Francesco Lomonaco, Marco Siino, and Maurizio Tesconi. 2023. Text enrichment with japanese language to profile cryptocurrency influencers. In *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), Thessaloniki, Greece, September 18th to 21st, 2023*, volume 3497 of *CEUR Workshop Proceedings*, pages 2708–2716. CEUR-WS.org.

Beatriz Matias Santana Maia, Maria Clara Falcão Ribeiro de Assis, Leandro Muniz de Lima, Matheus Becali Rocha, Humberto Giuri Calente, Maria Luiza Armini Correa, Danielle Resende Camisasca, and Renato Antonio Krohling. 2024. Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer. *Expert Systems with Applications*, 241:122418.

Stefano Mangione, Marco Siino, and Giovanni Garbo. 2022. Improving irony and stereotype spreaders detection using data augmentation and convolutional neural network. In *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022*, volume 3180 of *CEUR Workshop Proceedings*, pages 2585–2593. CEUR-WS.org.

Zong Meng, Zhaohui Zhang, Yang Guan, Jimeng Li, Lixiao Cao, Meng Zhu, Jingjing Fan, and Fengjie Fan. 2024. A hierarchical transformer-based adaptive metric and joint-learning network for few-shot rolling bearing fault diagnosis. *Measurement Science and Technology*, 35(3).

Mohsen Miri, Mohammad Bagher Dowlatshahi, Amin Hashemi, Marjan Kuchaki Rafsanjani, Brij B Gupta, and W Alhalabi. 2022. Ensemble feature selection for multi-label text classification: An intelligent order statistics approach. *International Journal of Intelligent Systems*, 37(12):11319–11341.

Fuad Muftie and Muhammad Haris. 2023. Indobert based data augmentation for indonesian text classification. In *2023 International Conference on Information Technology Research and Innovation, ICITRI 2023*, page 128 – 132.

Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE access*, 7:100943–100953.

Marco Siino. 2024a. Badrock at semeval-2024 task 8: Distilbert to detect multigenerator, multidomain and multilingual black-box machine-generated text. In *Proceedings of the 18th International Workshop on Semantic Evaluation*, SemEval 2024, Mexico City, Mexico.

Marco Siino. 2024b. Mcrock at semeval-2024 task 4: Mistral 7b for multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation*, SemEval 2024, Mexico City, Mexico.

Marco Siino. 2024c. T5-medical at semeval-2024 task 2: Using t5 medical embeddings for natural language inference on clinical trial data. In *Proceedings of the 18th International Workshop on Semantic Evaluation*, SemEval 2024, Mexico City, Mexico.

Marco Siino, Elisa Di Nuovo, Ilenia Tinnirello, and Marco La Cascia. 2021. Detection of hate speech spreaders using convolutional neural networks. In *Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st - to - 24th, 2021*, volume 2936 of *CEUR Workshop Proceedings*, pages 2126–2136. CEUR-WS.org.

Marco Siino, Elisa Di Nuovo, Ilenia Tinnirello, and Marco La Cascia. 2022a. Fake news spreaders detection: Sometimes attention is not all you need. *Information*, 13(9):426.

Marco Siino, Marco La Cascia, and Ilenia Tinnirello. 2022b. Mcrock at semeval-2022 task 4: Patronizing and condescending language detection using multi-channel cnn, hybrid lstm, distilbert and xlnet. In *Proceedings of the 16th International Workshop on Semantic Evaluation, SemEval@NAACL 2022, Seattle, Washington, United States, July 14-15, 2022*, pages 409–417. Association for Computational Linguistics.

Marco Siino, Francesco Lomonaco, and Paolo Rosso. 2024a. Backtranslate what you are saying and i will tell who you are. *Expert Systems*, n/a(n/a):e13568.

Marco Siino, Maurizio Tesconi, and Ilenia Tinnirello. 2023. Profiling cryptocurrency influencers with few-shot learning using data augmentation and ELECTRA. In *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), Thessaloniki, Greece, September 18th to 21st, 2023*, volume 3497 of *CEUR Workshop Proceedings*, pages 2772–2781. CEUR-WS.org.

Marco Siino and Ilenia Tinnirello. 2023. Xlnet with data augmentation to profile cryptocurrency influencers. In *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), Thessaloniki, Greece, September 18th to 21st, 2023*, volume 3497 of *CEUR Workshop Proceedings*, pages 2763–2771. CEUR-WS.org.

Marco Siino, Ilenia Tinnirello, and Marco La Cascia. 2022. T100: A modern classic ensemble to profile irony and stereotype spreaders. In *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022*, volume 3180 of *CEUR Workshop Proceedings*, pages 2666–2674. CEUR-WS.org.

Marco Siino, Ilenia Tinnirello, and Marco La Cascia. 2024b. Is text preprocessing still worth the time? a comparative survey on the influence of popular preprocessing methods on transformers and traditional classifiers. *Information Systems*, 121:102342.

Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to fine-tune bert for text classification? In *Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18*, pages 194–206. Springer.

José Medardo Tapia-Téllez and Hugo Jair Escalante. 2020. Data augmentation with transformers for text classification. In *Advances in Computational Intelligence*, pages 247–259, Cham. Springer International Publishing.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Dang Van Thin, Duong Ngoc Hao, and Ngan Luu-Thuy Nguyen. 2023. Vietnamese sentiment analysis: An overview and comparative study of fine-tuning pretrained language models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(6):1–27.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Xixi Wang, Xiao Wang, Bo Jiang, and Bin Luo. 2023. Few-shot learning meets transformer: Unified query-support transformers for few-shot classification. *IEEE Trans. Circuits Syst. Video Technol.*, 33(12):7789–7802.

Yuxia Wang, Jonibek Mansurov, Petar Ivanov, Jinyan Su, Artem Shelmanov, Akim Tsvigun, Chenxi Whitehouse, Osama Mohammed Afzal, Tarek Mahmoud, Giovanni Puccetti, Thomas Arnold, Alham Fikri Aji, Nizar Habash, Iryna Gurevych, and Preslav Nakov. 2024. Semeval-2024 task 8: Multigenerator, multidomain, and multilingual black-box machine-generated text detection. In *Proceedings of the 18th International Workshop on Semantic Evaluation*, SemEval 2024, Mexico, Mexico.