McRock at SemEval-2024 Task 4: Mistral 7B for Multilingual Detection of Persuasion Techniques in Memes

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

Abstract

One of the most widely used content types in internet misinformation campaigns is memes. Since they can readily reach a big number of users on social media sites, they are most successful there. Memes used in a disinformation campaign include a variety of rhetorical and psychological strategies, including smearing, name-calling, and causal oversimplification, to achieve their goal of influencing users. The shared task's objective is to develop models for recognizing these strategies solely in a meme's textual content (Subtask 1) and in a multimodal context where both the textual and visual material must be analysed simultaneously (Subtasks two and three). In this paper, we discuss the application of a Mistral 7B model to address the Subtask one in English about finding the persuasive strategy that a meme employs from a hierarchy of twenty based just on its textual content. Only a portion of the reward is awarded if the technique's ancestor node is chosen. This classification issue is multilabel hierarchical. Our approach based on the use of a Mistral 7B model obtains a Hierarchical F1 of 0.42 a Hierarchical Precision of 0.30 and a Hierarchical Recall of 0.71. Our selected approach is able to outperform the baseline provided for the competition.

1 Introduction

When information is intentionally crafted to serve a predetermined agenda, we often classify it as propaganda (Geissler et al., 2023). Propaganda employs various psychological and rhetorical strategies to achieve its goals (Çakmak, 2023). These methods encompass the utilization of logical fallacies and the manipulation of audience emotions (Soares et al., 2023). Logical fallacies can be particularly deceptive as they may initially appear sound and impartial, yet upon closer examination, it becomes evident that the conclusion cannot be logically derived from the premises. Another tactic involves employing emotionally charged language to sway the audience's opinion, bypassing rational analysis in favour of an emotional connection. Memes, typically comprising images overlaid with text, serve as a platform for propagandistic dissemination. Within deceptive memes, images either reinforce or complement textual techniques, or they themselves convey persuasive strategies.

To address these objectives, there is an ongoing demand for automated tools capable of extracting and categorizing data from online sources, facilitating the response to both established and emerging societal concerns. Recent advancements in machine and deep learning architectures have spurred heightened interest in Natural Language Processing (NLP). Substantial endeavours have been directed towards devising techniques for the automated identification and categorization of textual content accessible on the internet today. In the literature, to perform text classification tasks, several strategies have already been proposed. In the last fifteen years, some of the most successful strategies have been based on SVM (Colas and Brazdil, 2006; Croce et al., 2022), on Convolutional Neural Network (CNN) (Kim, 2014; Siino et al., 2021), on Graph Neural Network (GNN) (Lomonaco et al., 2022), on ensemble models (Miri et al., 2022; Siino et al., 2022) and, recently, on Transformers (Vaswani et al., 2017; Siino et al., 2022b).

The surge in the adoption of Transformer-based architectures within academic research has been further propelled by diverse methodologies showcased at SemEval 2024. These methodologies address a range of tasks and yield notable outcomes. For instance, in Task 2 (Jullien et al., 2024), T5 is utilized to confront the challenge of identifying the inference relation between plain language statements and Clinical Trial Reports (Siino, 2024b). In Task 10, a Mistral 7B model is employed to perform emotion Recognition in Conversation (ERC) within Hindi-English code-mixed conversations

Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 53–59

(Siino, 2024c). Additionally, in Task 8 (Wang et al., 2024), a DistilBERT model is leveraged to identify machine-generated text (Siino, 2024a).

Finally, for the Task 4 at SemEval 2024 (Dimitrov et al., 2024) – Multilingual Detection of Persuasion Techniques in Memes – three Subtasks were proposed. As already stated, in a disinformation campaign, memes effectively manipulate users through various rhetorical and psychological strategies, including causal oversimplification, name-calling, and smear tactics. The objective of this shared task is to develop models capable of detecting these techniques within the textual content of memes alone (Subtask 1), as well as within a multimodal framework where both textual and visual elements are analysed jointly (Subtasks 2 and 3). To face with the first Subtask in English, 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, we provided the definitions of the 20 techniques to the model to identify, given each sample, all the techniques detected. We opted for Mistral 7B because the comparative analysis between Mistral 7B and other leading models, namely Llama 2 and Llama 1, reveals noteworthy advancements in common NLP tasks. Across multiple benchmark evaluations, Mistral 7B consistently exhibits superior performance in comparison to Llama 2, a prominent open 13B model. Moreover, its efficacy extends beyond mere parity with, but rather exceeds, the achievements of Llama 1, a state-of-the-art 34B model, particularly in tasks pertaining to reasoning, mathematics, and code generation. These findings underscore Mistral 7B's substantive contributions to the advancement of NLP, suggesting its potential as a benchmark model in the field.

The rest of the paper is made as follows. In Section 2 we provide some background on the Task 4 hosted at SemEval 2024. In Section 3 we provide a description of the models 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].

[1] https://github.com/marco-siino/SemEval2024/

2 Background

This section furnishes background information regarding Task 4 (Subtask 1), held at SemEval 2024. The task entails identifying, based solely on the textual content of a meme, which of the 20 persuasion techniques, organized hierarchically, are employed. The selection of an ancestor node of a technique warrants only partial reward. The task thus presents a hierarchical, multilabel classification challenge. The hierarchical structure is illustrated in the official task's page, with 22 techniques depicted, although "Transfer" and "Appeal to Strong Emotion" are excluded from Subtask 1. For comprehensive details, please refer to provided resources.

For all Subtasks, the annotations from the PTC corpus, comprising over 20,000 sentences, were utilized where feasible. Although the corpus pertains to news articles, annotations adhere to identical guidelines, albeit with fewer techniques considered. As highlighted by the task coordinators, certain meme content may be deemed offensive or excessively potent by certain audiences. A similar multilingual corpus was also accessible during SemEval 2023 (Piskorski et al., 2023). Here again, the corpus revolves around news articles across nine languages, yet the number of techniques and annotation guidelines differ marginally. A training set for local system development was additionally provided. Furthermore, the organizers furnished a development set and a public leaderboard for real-time result sharing among task participants. Ultimately, the organizers supplied a test set devoid of annotations and an online submission platform to evaluate the system performance.

Subtask 1 relies on the textual content extracted from memes as input data. Training, development, and test sets for all Subtasks are disseminated as JSON files, with each Subtask having its own individual file. For Subtasks 2a and 2b, in addition to the meme's textual content, input data includes the meme's image. In the Figure 1, is reported a sample from the official competition website[2].

Given the Figure 1:

- *ID* consists of a unique identifier of the example across all the Subtasks;
- *text* represents the textual expression within the meme, formatted as a singular UTF-8

[2] https://propaganda.math.unipd.it/semeval2024task4/

Subtask 1

The entry for that example in the json file for subtask 1 is


```
{
        "id": "125",
        "text": "I HATE TRUMP\n\nMOST TERRORIST DO",
        "labels": [
                "Loaded Language",
                "Name calling/Labeling"
],
"link": "https://..."
},
```


Figure 1: An example from the dataset. In this case, two labels are assigned to the sample's text.

string. Initially, this text is automatically extracted from the meme, subsequently undergoing manual post-processing to rectify errors and arrange it such that each sentence occupies a distinct row. Furthermore, segments of text originating from distinct regions within the image are demarcated by blank rows. Notably, Task 1 qualifies as an NLP endeavour, given that image input is absent;

- *labels* denotes a compilation of permissible technique names identified within the text. These labels serve as the gold standard and will solely be furnished for the training set. In this particular instance, two techniques were identified: "Loaded Language" and "Name calling/Labelling."

3 System Overview

Even if it has already been proved that the Transformers are not necessarily the best option for any text classification task (Siino et al., 2022a), depending on the goal some strategies like domain-specific fine-tuning (Sun et al., 2019; Van Thin et al., 2023), or data augmentation (Lomonaco et al., 2023; Mangione et al., 2022) can be beneficial in several applications.

Our approach is a few-shot one (Littenberg-Tobias et al., 2022) and make use of the above-mentioned Mistral 7B. Mistral 7B, a language model equipped with 7 billion parameters, is designed to excel in both performance and efficiency. Compared to the leading open 13B model (Llama 2), Mistral 7B demonstrates superior performance across all evaluated benchmarks. Moreover, it outperforms the top released 34B model (Llama 1) in tasks related to reasoning, mathematics, and code generation. The model leverages grouped-query attention (GQA) to expedite inference, along with sliding window attention (SWA) to efficiently process sequences of varying lengths while minimizing inference costs. Additionally, a fine-tuned variant, Mistral 7B – Instruct, tailored for adhering to instructions, surpasses the 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. Notably, this variant lacks any moderation mechanisms. The Mistral 7B Instruct variant requires a specific input format, as stated below:

> <s>[INST] Instruction [/INST] Model answer</s>[INST] Follow-up instruction [/INST]

Instruction, along with the following *Model answer*, can be a single sample with the related label or a set of sample/label pairs (realizing, in this case, a few-shot use of the model). Then, *Follow-up instruction* is the current sample for which the prediction has to be provided by the model. More specifically, given the 20 persuasion techniques in memes, we have prepared a text string containing the techniques and their definitions to provide context in the template ready. The definitions of the 20 techniques are provided by the task organizers[3]. At this point, the full text containing the twenty definitions plus the sample to be classified were provided as prompt to Mistral.

Then the question provided as prompt to mistral was: *"Given the above Definitions of the Persuasion Techniques, Identify the Persuasion Techniques used in the Sentence. Answer using ONLY one or more numbers in the range 1-20 separated by commas. No text nor other options are allowed."*.

To this request, the model replied with one or more techniques detected in the corresponding sample. So, as an example, to the sentence: "Happy April Fools Day - - Ooop I mean: March Fools day" the model replied to the prompt with the numbers 3 and 5. These two numbers correspond to the technique 3 (i.e., *Whataboutism*) and to the technique 5 (i.e., *Obfuscation, Intentional vagueness, Confusion*). It is important to mention that we also tried to use the model in a zero-shot configuration. In this case, we just asked the model to pick one or more categories given a meme. Unfortunately, the model did not report one or more correct categories, while developing discussions as answers.

Finally, we collected all the predictions provided

[3] https://propaganda.math.unipd.it/semeval2024task4/definitions.html

on the test set to into a JSON file with the required format to submit our predictions.

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 performing 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. The low impact of the best preprocessing techniques - or combinations of techniques - using Transformers, as reported in the study, is due to several factors like preserving the quantity and the quality of the original information available.

4 Experimental Setup

We implemented our model on Google Colab. The library we used comes from Hugging Face and as already mentioned is Mistral 7B[4]. We employed the v0.2 iteration of Mistral 7B, which represents an enhanced version of the Mistral-7B-Instruct-v0.1 model. To harness the capabilities of instruction fine-tuning, prompts must be enclosed within [INST] and [/INST] tokens. Additionally, the initial instruction should commence with a sentence identifier. The next instructions should not. The assistant generation will be ended by the end-of-sentence token ID. We also imported the Llama library (Touvron et al., 2023) from *llama_cpp*. The library is fully described on GitHub[5]. The dataset provided for all the phases are available on the Official Competition page. We did not perform any additional fine-tuning on the model. To run the experiment, a T4 GPU from Google has been used. After the generation of predictions, we exported the results on the format required by the organizers. As already mentioned, all of our code is available on GitHub.

5 Results

The evaluation was done by submitting to the leaderboard the predictions provided by the model. Subtask 1 and 2a are reliant on a hierarchical structure. The gold label consistently corresponds to a leaf node within the Directed Acyclic Graph (DAG). However, any node within the DAG can serve as a predicted label:

- If the prediction does not correspond to a leaf node and is an ancestor of the correct gold label, a partial reward is issued, with the reward magnitude contingent upon the distance between the two nodes. For instance, if the gold label is "Red Herring" and the predicted label is "Distraction" or "Appeal to Logic."

- If the prediction does not align with any ancestor node of the correct label, no reward is granted. For instance, if the gold label is "Red Herring" and the predicted label is "Black and White Fallacy" or "Appeal to Emotions." A graphical representation illustrating this concept is provided.

However, it's worth noting that the hierarchical structure can be disregarded by confining predictions solely to technique names. This approach renders the task analogous to SemEval 2023 Task 3 (Piskorski et al., 2023).

An illustrative example of the evaluation function can be accessed online[6]. In this case, the Subtask consists of a hierarchical multilabel classification task. Drawing from the aforementioned figure depicting the hierarchy, any node within the DAG can be designated as a predicted label. The gold label consistently corresponds to a leaf node within the DAG. Hierarchical-F1, detailed in (Kiritchenko et al., 2006), is employed as the official evaluation metric.

In the Table 1, the results obtained by the first three teams and by the last one, as showed on the official page[7], are reported. Compared to the best performing models, our simple approach exhibits some room for improvements, although it is able to outperform the baseline. 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.

6 Conclusion

This paper presents the application of Mistral 7B-model for addressing the Task 4 at SemEval 2024.

[4] https://huggingface.co/TheBloke/Mistral-7B-Instruct-v0.2-GGUF

[5] https://github.com/ggerganov/llama.cpp

[6] https://propaganda.math.unipd.it/semeval2024task4/data/hierarchy_evaluation.html

[7] https://propaganda.math.unipd.it/semeval2024task4/SemEval2024task4_test.html

	H-F1	H-Prec	H-Recall
914isthebest (1)	0.752	0.684	0.836
BCAmirs (2)	0.698	0.668	0.732
OtterlyObsessedWithSemantics (3)	0.697	0.648	0.755
Mistral 7B (30)	0.42	0.30	0.71
BASELINE (31)	0.369	0.477	0.300
IIMAS1UTM1LaSalle (33)	0.199	0.755	0.115

Table 1: Comparing performance on the test set for Subtask 1 in English. In the table are shown the results obtained by the first three users and by the last one. Furthermore, is included the result of the baseline considered and of our approach making use of Mistral 7B. In parentheses is reported the position in the official final ranking.

For our submission, we decided to follow a few-shot learning approach, employing as-is, an in-domain pre-trained Transformer. After several experiments, we found beneficial to build a prompt containing the definitions of the techniques in memes. Then we provide, as a prompt, the definitions together with a sample. The model was asked to select all the techniques detected in the sentence. The task is challenging, and there is still opportunity for improvement, as can be noted looking at the final ranking. Possible alternative approaches include utilizing the zero-shot capabilities of other models like GPT and T5, increasing the size of the training set by using further data, or directly integrating ontology-based domain knowledge differently than what has been proposed in our work. Further improvements could be obtained with a fine-tuning and modelling the problem as a different 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 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 extend our gratitude to the anonymous reviewers for their insightful comments and valuable suggestions, which have significantly enhanced the clarity and presentation 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.

Dominique Geissler, Dominik Bär, Nicolas Pröllochs, and Stefan Feuerriegel. 2023. Russian propaganda on social media during the 2022 invasion of ukraine. *EPJ Data Science*, 12(1).

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.

Svetlana Kiritchenko, Stan Matwin, Richard Nock, and A. Fazel Famili. 2006. Learning and evaluation in the presence of class hierarchies: Application to text categorization. In *Canadian AI*, volume 4013 of *Lecture Notes in Computer Science*, pages 395–406. Springer.

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.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multilingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

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. 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. 2024c. Transmistral at semeval-2024 task 10: Using mistral 7b for emotion discovery and reasoning its flip in conversation. 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.

Felipe Bonow Soares, Anatoliy Gruzd, and Philip Mai. 2023. Falling for russian propaganda: Understanding the factors that contribute to belief in pro-kremlin disinformation on social media. *Social Media and Society*, 9(4).

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.

Veysel Çakmak. 2023. *Social media use and propaganda techniques: An evaluation of the Ukraine-Russia war*. IGI Global.