SemanticCuetSync at CheckThat! 2024: Finding Subjectivity in News Articles using Llama

Notebook for the CheckThat! Lab at CLEF 2024

Ashraful Islam Paran[1,†], Md. Sajjad Hossain[1,†], Symom Hossain Shohan[1,†], Jawad Hossain[1], Shawly Ahsan[1] and Mohammed Moshiul Hoque[1,*]

[1]Chittagong University of Engineering and Technology, Chattogram-4349, Bangladesh

Abstract

This study introduces an LLM-based technique for detecting subjectivity and objectivity in English and Arabic news articles. Although several transformers, deep learning (DL), and machine learning (ML)- based techniques were exploited for the task, the LLM (Llama-3-8b) outperformed other models, obtaining the highest F1-scores of 72.6% (Arabic) and 50.36% (English). The suggested LLM-based solution provides a rank of 4th (Arabic) and 12th (English) in the task competition. The research emphasizes the potential of advanced LLMs like Llama-3-8b in achieving high subjectivity and objectivity detection accuracy, which is essential for applications in media analysis, sentiment analysis, and automated content moderation. This study contributes to developing robust multilingual text classification systems, paving the way for more sophisticated and accurate linguistic analysis tools.

Keywords

Natural Language Processing, Subjectivity, Large Language Model (LLM), Llama, Objectivity

1. Introduction

In the era of technology, the internet is a constant source of textual information, including news articles, social media posts, blogs, and reviews. These texts offer a diverse range of information, opinions, and narratives. The ability to distinguish between subjective and objective content, especially in news articles, is crucial. Subjective text often includes personal opinions, emotions, and biases, significantly influencing the reader's perception. Objective text, on the other hand, presents factual and impartial observations. The automatic classification of text sequences into subjective or objective categories has wide-ranging applications, including media analysis, sentiment analysis, and information retrieval. This capability can significantly enhance the quality of information processing and extraction across various domains, leading to more accurate and reliable results.

How news editorials address political issues may influence individuals with differing ideological beliefs [1]. Researchers have proposed a variety of methods to categorize subjective and objective news articles [2], [3], [4], [5], [6]. The majority of these approaches focus on high-resource languages. The primary challenge in classifying subjectivity and objectivity is addressing language's intricate and context-dependent nature. Most of the time, subjective texts incorporate subtle linguistic markers to convey personal viewpoints. The critical contributions of this work are:

- Introduced an LLM-based technique for classifying text into subjective and objective categories in Arabic and English.

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

*Corresponding author.

†These authors contributed equally.

u1904029@student.cuet.ac.bd (A. I. Paran); u1904031@student.cuet.ac.bd (Md. S. Hossain); u1904048@student.cuet.ac.bd (S. H. Shohan); u1704039@student.cuet.ac.bd (J. Hossain); u1704057@student.cuet.ac.bd (S. Ahsan); moshiul_240@cuet.ac.bd (M. M. Hoque)

0009-0001-4795-3816 (A. I. Paran); 0009-0008-8670-8857 (Md. S. Hossain); 0009-0004-0834-2037 (S. H. Shohan); 0009-0006-6051-8989 (J. Hossain); 0009-0003-9940-9681 (S. Ahsan); 0000-0001-8806-708X (M. M. Hoque)

- Investigated the task performance leveraging various ML, DL, transformer, and LLM models to discover a reasonable solution for classifying Arabic and English news into subjective and objective categories.

2. Related Work

The rise of yellow journalism has made it more crucial than ever to determine an article's objectivity or subjectivity. NLP can be crucial in identifying subjectivity and objectivity in news articles. Annotation rules to identify if an article is subjective or objective were provided by Antici et al. [7]. It is possible to use these ideas in other languages. An Arabic dataset containing several news item types for subjectivity and sentiment analysis was presented by Abdul-Mageed et al. [8]. Dey et al. [9] proposed a transformer-based approach (XLM-RoBERTa large) to identify subjectivity in news articles. Their model recorded an F1 score of 0.81 in multilingual datasets. Pachov et al. [2] provided an ensemble technique for detecting subjectivity, which recorded an F1 score of 0.77. AI-generated news from ChatGPT was used by Shushkevich et al. [10] to balance the dataset, which improved the F1 score by 3% in Italian and 9% in English by using mBERT. A back translation method in conjunction with a transformer-based solution (RoBERTa, BERT) was suggested by Tran et al. [11], which achieved an F1 score of 0.69 in English.

Frick et al. [12] proposed to use ChatGPT to detect subjectivity. They used GPT-3.5, and on the English test dataset, they obtained an F1 score of 0.73. Using GPT-3.5, they obtained F1 values of 0.68 on the German and 0.73 on the English datasets. Furthermore, ChatGPT can be applied in a few-shot and zero-shot manner to detect subjectivity in news articles [13]. This work leverages the LLMs for classifying texts into objective and subjective.

3. Dataset and Task Description

The dataset used in this work includes two classes (SUBJ and OBJ) and features sentences in English and Arabic. Table 1 illustrates the distribution of train, dev, dev-test, and test sets. We trained all models using the training set and evaluated the model's performance based on the test set.

Table 1
Dataset statistics for Task-2, where TW stands for total words and UW stands for unique words.

Language	Train	Dev	Dev-Test	Test	Total	TW	UW
English	830	219	243	484	1776	30821	5785
Arabic	1185	297	445	748	2675	53041	17477
Total	**2015**	**516**	**688**	**1232**	**4451**	**83862**	**23262**

CLEF 2024 - CheckThat! Lab [14, 15, 16] consists of six tasks [17, 18, 19, 20, 21]. We participated in task-2 of this shared task. Task-2 [18] focused on distinguishing whether a sentence from a news article expresses the subjective view of the author behind it or presents an objective view on the covered topic instead. Table 2 depicts an example of training data for the different languages.

4. System Overview

The ML techniques employed include linear regression (LR), support vector machine (SVM), multinomial naive Bayes (MNB), k-nearest neighbors (KNN), and random forest (RF). The DL techniques involved CNN, CNN+LSTM, and CNN+BiLSTM. Finally, two LLMs are fine-tuned for each language to address the given task. Figure 1 illustrates the schematic process of subjectivity detection.

Textual Feature Extraction: Textual feature extraction is a crucial step in natural language processing, involving converting raw text data into numerical formats. This numerical format helps models

Table 2
Task-2 sample with the text and corresponding label

Text	Label
Gone are the days when they led the world in recession-busting	SUBJ
The trend is expected to reverse as soon as next month.	OBJ
كلها تتحدث عن إسرائيل وموقعها العالمي والاقتصادي والتنموي والعسكري. (They all talk about Israel and its global, economic, developmental and military position.)	SUBJ
غسل اليدين بصفة متكررة بالماء والصابون أو بمطهر يحتوي على كحول. (Wash hands frequently with soap and water or an alcohol-based sanitizer.)	OBJ

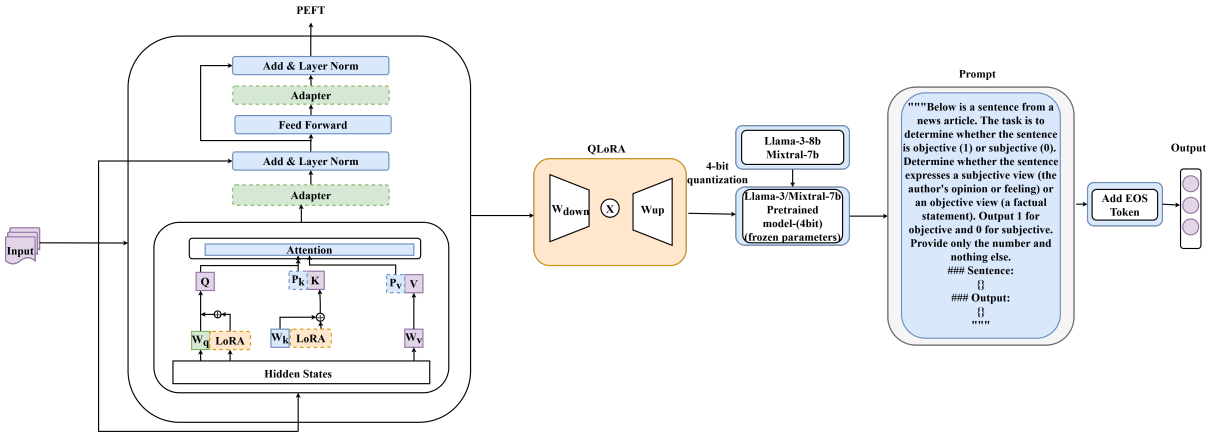

Figure 1: Schematic process of subjectivity detection in the best performing model.

interpret and process textual information. A Count Vectorizer is employed in the ML models explored in this study. It is a popular method for textual feature extraction that converts text data into a matrix of token counts. In DL models, tokenization and padding transform raw texts into structured numerical data. These numerical formats are fed through an embedding layer, which captures more sophisticated features such as semantic relationships.

ML Models: This study explores several ML models, including LR, SVM, MNB, KNN, and RF. The hyperparameter configurations for these models are detailed in Table 3.

Table 3
Parameters of the employed ML models.

Classifier	Parameters	Value
LR	solver	lbfgs
	max_iter	20000
MNB	alpha	1.0
	fit-prior	False
SVM	kernel	linear
	gamma	auto

CNN: This study utilizes a CNN model, starting with an embedding layer with an output dimension of 200. The architecture includes two Conv1D layers, containing 64 and 128 filters, respectively, employing a kernel size of 2 and ReLU activation. For downsampling, a GlobalMaxPooling1D layer is used. Following this, a dense layer with 128 units and ReLU activation is added, along with a dropout layer at a rate of 0.5 to mitigate overfitting. The final output layer consists of a single unit with sigmoid activation. The model is trained using the '*binary_crossentropy*' loss function and the '*Nadam*' optimizer, with a batch size of 32 over three epochs.

CNN+LSTM: The CNN-LSTM model implemented in this study shares a similar architecture with

the CNN model but includes an LSTM layer with 64 units and a 0.2 dropout rate for sequence modeling. Additionally, the dense layer in this design has 64 units and uses the ReLU activation function. The other hyperparameter settings remain the same as those used in the CNN model.

CNN+BiLSTM: This model has an architecture similar to the CNN+LSTM model but replaces LSTM with BiLSTM.

Transformer-based models: In this study, three transformer-based models were fine-tuned using the English dataset, while another three were fine-tuned using the Arabic dataset. The models employed in English were MdeBERTav3 [22], BERT-base-uncased [23], and RoBERTa [24]. Several text preprocessing measures were implemented to reduce noise in the dataset and concentrate on meaningful words, including lowercasing, emoji removal, stop word removal, stemming, contraction expansion, simple spelling correction using Unicode, and HTML tag elimination. For stopword removal, the NLTK stopwords list was employed.

MdeBERTav3 is a multilingual BERT-based architecture specifically designed for multilingual tasks. This model demonstrates superior linguistic comprehension across various languages and performs excellently in this study. The BERT-base-uncased model, another pre-trained transformer architecture, has previously shown exceptional performance in various natural language processing (NLP) tasks and delivered satisfactory results. Finally, RoBERTa, another optimized version of BERT, is used here for the specified task and obtained comparative results.

In Arabic tasks, AraBERTv2 [25] is additionally used besides MdeBERTav3 and RoBERTa. AraBERTv2 leverages its prior training on an Arabic dataset and has demonstrated its usefulness in various natural language processing tasks, such as sentiment analysis, named entity recognition, and question answering. This model has achieved notable results in the task at hand.

Table 4 shows the learning rate (LR), weight decay (WD), warmup steps (WS), and epochs (EP) used for training the large language models.

Table 4
Hyperparameters for the transformers.

Language	Models	LR	WD	WS	EP
English	MdeBERTav3	$5e^{-5}$	0	200	10
	BERT-base-uncased	$5e^{-5}$	0.01	200	10
	RoBERTa	$5e^{-5}$	0.01	200	10
Arabic	MdeBERTav3	$3e^{-5}$	0.30	500	4
	RoBERTa	$3e^{-5}$	0.30	500	4
	AraBERTv2	$3e^{-5}$	0.30	500	4

Mixtral-7b: In this study, the Mixtral-7b [26] was fine-tuned for subjectivity and objectivity detection in news articles in both English and Arabic. Mixtral-7b was chosen due to its advanced capabilities in handling multilingual data. The ability to effectively comprehend and process English and Arabic texts helped this model perform better in the given task. To achieve the desired results, the model was trained on labeled datasets in both English and Arabic, which enabled it to discern between subjective and objective content.

Llama-3-8b: Llama-3-8b [27], another multilingual large language model, is used in this task. Llama-3-8b is a versatile and practical model for performing multilingual NLP tasks. It demonstrated its potential to handle complex subjectivity and objectivity detection in English and Arabic.

Table 5 shows the learning rate (LR), weight decay (WD), warmup steps (WS), max-length, Lora-alpha (LA), gradient accumulation steps (GAS), and epochs (EP) used for training the large language models.

5. Results and Analysis

Table 6 illustrates an in-depth analysis of the performance of machine learning (ML), deep learning (DL), transformer-based models, and large language models across English and Arabic languages on the test set.

Table 5
Hyperparameters for the LLMs.

Language	Models	LR	WD	WS	Max_len	LA	GAS	EP
English	Mixtral-7b	$5e^{-5}$	$1e^{-3}$	5	50	16	4	12
	Llama-3-8b	$5e^{-5}$	$1e^{-3}$	5	50	16	4	12
Arabic	Mixtral-7b	$6e^{-5}$	$1e^{-3}$	10	50	16	4	10
	Llama-3-8b	$5e^{-5}$	$1e^{-3}$	10	50	16	4	10

Table 6
Performance of the employed models on the test set. The bold rows denote the performance of the best-performing model

Language	Method	Classifier	Pr(%)	Re(%)	Ac(%)	Macro-F1(%)	SUBJ-F1(%)
English	ML Models	LR	60.82	59.34	71.69	59.83	38.01
		SVM	**56.11**	**56.43**	**66.12**	**56.23**	**35.43**
		MNB	56.52	57.63	64.26	56.63	38.43
	DL Models	CNN+LSTM	37.40	50.00	74.79	42.79	0.00
		CNN+BiLSTM	**52.22**	**51.56**	**68.18**	**51.11**	**22.22**
	Transformers	MdeBERTav3	70.45	65.11	77.89	66.65	86.01
		BERT-base-uncased	69.68	66.47	77.48	67.63	85.49
		RoBERTa	**70.01**	**66.61**	**77.69**	**67.82**	**85.64**
	LLMs	Mixtral-7b	37.40	50.00	74.79	42.79	0.00
		Llama-3-8b	**76.96**	**70.32**	**81.61**	**72.46**	**56.59**
Arabic	ML Models	LR	51.83	50.53	56.28	42.42	14.17
		SVM	**49.73**	**49.87**	**54.81**	**44.53**	**20.66**
		MNB	53.75	51.26	56.82	44.08	17.39
	DL Models	CNN+LSTM	28.41	50.00	56.82	36.23	0.00
		CNN+BiLSTM	**52.41**	**50.54**	**56.55**	**41.33**	**11.44**
	Transformers	**AraBERTv2**	**51.28**	**51.06**	**54.01**	**49.91**	**64.24**
		RoBERTa	28.41	50.00	56.82	36.23	72.46
		MdeBERTav3	49.99	49.99	53.48	48.04	31.23
	LLMs	Mixtral-7b	49.17	49.20	50.80	49.07	39.67
		Llama-3-8b	**51.40**	**51.20**	**53.88**	**50.36**	**37.16**

On English test data, among ML models, SVM emerged as the top-performing model with a precision of 56.11%, recall of 56.43%, accuracy of 66.12%, macro-F1 score of 56.23%, and SUBJ-F1 score of 35.43%. Among the DL models, CNN+BiLSTM demonstrated the best performance with a precision of 52.22%, recall of 51.56%, accuracy of 68.18%, macro-F1 score of 51.11%, and SUBJ-F1 score of 22.22%. In the transformer category, RoBERTa outperformed others with a precision of 70.01%, recall of 66.61%, an accuracy of 77.69%, macro-F1 score of 67.82%, and SUBJ-F1 score of 85.64%.

In the Arabic test data, Llama-3-8b demonstrated the best performance with a precision of 76.96%, recall of 70.32%, accuracy of 81.61%, macro-F1 score of 72.46%, and SUBJ-F1 score of 56.59%. The large language models dominated other ML, DL, and transformer-based models in English and Arabic. Llama-3-8b is the best-performing LLM in both languages.

5.1. Error Analysis

A comprehensive quantitative and qualitative error analysis is conducted to provide detailed insights into the proposed model's performance.

Quantitative Analysis

Figure 2 illustrates the confusion matrix of Llama-3-8b for English and Arabic. Out of 484 English test cases, the model successfully detects the positive class, with 337 True Positives and 25 False Positives. This reflects high precision, meaning the model is accurate when predicting 'SUBJ.' Moreover,

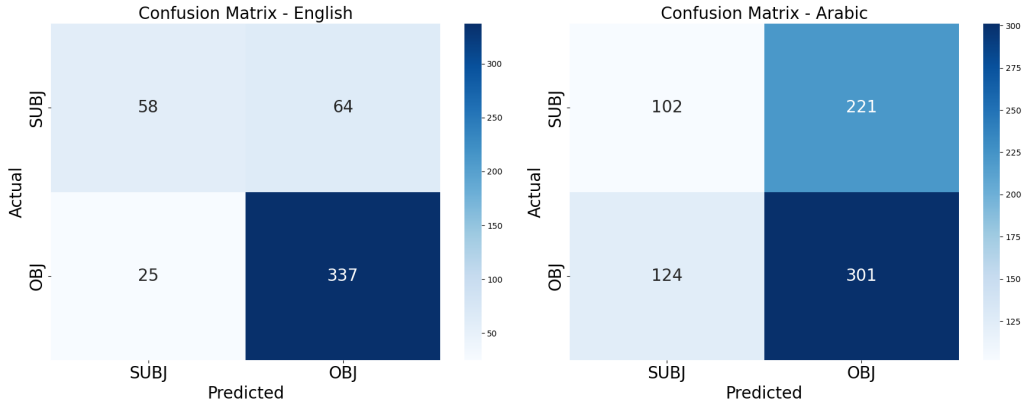

Figure 2: Confusion matrix of Llama-3-8b for English and Arabic.

it correctly identifies many negative instances with 58 True Negatives. However, 64 False Negatives indicate some positive instances are missed. Overall, the model displays a balanced approach with significant proficiency in minimizing incorrect positive predictions, leading to a high F1 score for positive samples.

In contrast, for 748 Arabic test cases, the model exhibits a different pattern in its confusion matrix. It accurately identifies instances from both classes with 301 True Positives and 102 True Negatives. However, many False Positives (221) and False Negatives (124) exist. This indicates that while the model can detect positive instances, it tends to misclassify many negative instances as positive, resulting in lower precision. Additionally, the high number of False Negatives suggests better recall, emphasizing the need for improved distinction between the two classes.

Qualitative Analysis

Table 7 presents some actual labels (AL) and predicted labels (PL) of the developed models.

Table 7
Few predictions with actual and predicted labels.

Text	AL	PL
A sip can really hit the spot after a long bike ride or a walk.	SUBJ	SUBJ
I just believe in being the best version of myself that I can possibly be, it makes me feel good.	OBJ	OBJ
House Democrats and the remaining pro-Ukraine House Republicans are casting about behind the scenes ...	SUBJ	OBJ
القرن الأخير، شنّت الدوائر الغربية والإعلام الغربي المسخّر لها حملات مدروسة ومركّزة (In the last quarter century, Western circles and the Western media under their control have launched deliberate and focused campaigns to confuse...)	SUBJ	SUBJ
موقع ذا هيل: مناقشات مسؤولي الإدارة بشأن تفعيل التعديل 25 من الدستور لعزل... (The Hill website: Administration officials' discussions regarding activating the 25th Amendment to the Constitution to remove the president seem limited and not...)	SUBJ	OBJ

It is evident that the models accurately predicted the labels for samples 1, 2, and 4, but made errors with samples 3 and 5. For the third sample, the sentence's intent is ambiguous, leading to an incorrect prediction by the model. In the case of the fifth sample, although the sentence is subjective, the model mislabels it due to the large language models being trained on insufficient Arabic data, and also, the provided training dataset is small.

6. Conclusion

This study evaluated several techniques to detect subjectivity in news articles, including ML, DL, transformer, and LLMs. Transformer-based solutions score better than ML and DL-based models. However, Llama-3-8b performed better than all these models, obtaining the highest F1 scores of 72.46% and 50.36% in Arabic and English, respectively. This study demonstrates how effective LLMs are in identifying subjectivity in articles. Even with limited resources, such as Arabic, LLMs outperform other models regarding results. Future improvements can be made using GPT-4, Llama-3-70b, or other LLMs with significant parameters.

References

[1] R. El Baff, H. Wachsmuth, K. Al Khatib, B. Stein, Analyzing the persuasive effect of style in news editorial argumentation, Association for Computational Linguistics, 2020.

[2] G. Pachov, D. Dimitrov, I. Koychev, P. Nakov, Gpachov at checkthat! 2023: a diverse multi-approach ensemble for subjectivity detection in news articles, arXiv preprint arXiv:2309.06844 (2023).

[3] E. V. Tunyan, T. Cao, C. Y. Ock, Improving subjective bias detection using bidirectional encoder representations from transformers and bidirectional long short-term memory, International Journal of Cognitive and Language Sciences 15 (2021) 329–333.

[4] H. Huo, M. Iwaihara, Utilizing bert pretrained models with various fine-tune methods for subjectivity detection, in: Web and Big Data: 4th International Joint Conference, APWeb-WAIM 2020, Tianjin, China, September 18-20, 2020, Proceedings, Part II 4, Springer, 2020, pp. 270–284.

[5] I. B. Schlicht, L. Khellaf, D. Altiok, Dwreco at checkthat! 2023: enhancing subjectivity detection through style-based data sampling, arXiv preprint arXiv:2307.03550 (2023).

[6] H. T. Sadouk, F. Sebbak, H. E. Zekiri, Es-vrai at checkthat! 2023: Enhancing model performance for subjectivity detection through multilingual data aggregation (2023).

[7] F. Antici, A. Galassi, F. Ruggeri, K. Korre, A. Muti, A. Bardi, A. Fedotova, A. Barrón-Cedeño, A corpus for sentence-level subjectivity detection on english news articles, arXiv preprint arXiv:2305.18034 (2023).

[8] M. Abdul-Mageed, M. Diab, Subjectivity and sentiment annotation of modern standard arabic newswire, in: Proceedings of the 5th linguistic annotation workshop, 2011, pp. 110–118.

[9] K. Dey, P. Tarannum, M. A. Hasan, S. R. H. Noori, Nn at checkthat!-2023: Subjectivity in news articles classification with transformer based models., in: CLEF (Working Notes), 2023, pp. 318–328.

[10] E. Shushkevich, J. Cardiff, Tudublin at checkthat! 2023: Chatgpt for data augmentation, Working Notes of CLEF (2023).

[11] S. Tran, P. Rodrigues, B. Strauss, E. Williams, Accenture at checkthat! 2023: Impacts of back-translation on subjectivity detection, Working Notes of CLEF (2023).

[12] R. A. Frick, Fraunhofer sit at checkthat! 2023: can llms be used for data augmentation & few-shot classification? detecting subjectivity in text using chatgpt, Working Notes of CLEF (2023).

[13] M. D. Türkmen, G. Coşgun, M. Kutlu, Tobb etu at checkthat! 2023: Utilizing chatgpt to detect subjective statements and political bias (2023).

[14] A. Barrón-Cedeño, F. Alam, J. M. Struß, P. Nakov, T. Chakraborty, T. Elsayed, P. Przybyła, T. Caselli, G. Da San Martino, F. Haouari, C. Li, J. Piskorski, F. Ruggeri, X. Song, R. Suwaileh, Overview of the CLEF-2024 CheckThat! Lab: Check-worthiness, subjectivity, persuasion, roles, authorities and adversarial robustness, in: L. Goeuriot, P. Mulhem, G. Quénot, D. Schwab, L. Soulier, G. M. Di Nunzio, P. Galuščáková, A. García Seco de Herrera, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024), 2024.

[15] A. Barrón-Cedeño, F. Alam, T. Chakraborty, T. Elsayed, P. Nakov, P. Przybyła, J. M. Struß, F. Haouari, M. Hasanain, F. Ruggeri, X. Song, R. Suwaileh, The CLEF-2024 CheckThat! Lab: Check-worthiness, subjectivity, persuasion, roles, authorities, and adversarial robustness, in: N. Goharian, N. Tonel-

lotto, Y. He, A. Lipani, G. McDonald, C. Macdonald, I. Ounis (Eds.), Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2024, pp. 449–458.
[16] G. Faggioli, N. Ferro, P. Galuščáková, A. García Seco de Herrera (Eds.), Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, CLEF 2024, Grenoble, France, 2024.
[17] M. Hasanain, R. Suwaileh, S. Weering, C. Li, T. Caselli, W. Zaghouani, A. Barrón-Cedeño, P. Nakov, F. Alam, Overview of the CLEF-2024 CheckThat! lab task 1 on check-worthiness estimation of multigenre content, in: [16], 2024.
[18] J. M. Struß, F. Ruggeri, A. Barrón-Cedeño, F. Alam, D. Dimitrov, A. Galassi, G. Pachov, I. Koychev, P. Nakov, M. Siegel, M. Wiegand, M. Hasanain, R. Suwaileh, W. Zaghouani, Overview of the CLEF-2024 CheckThat! lab task 2 on subjectivity in news articles, in: [16], 2024.
[19] J. Piskorski, N. Stefanovitch, F. Alam, R. Campos, D. Dimitrov, A. Jorge, S. Pollak, N. Ribin, Z. Fijavž, M. Hasanain, N. Guimarães, A. F. Pacheco, E. Sartori, P. Silvano, A. V. Zwitter, I. Koychev, N. Yu, P. Nakov, G. Da San Martino, Overview of the CLEF-2024 CheckThat! lab task 3 on persuasion techniques, in: [16], 2024.
[20] F. Haouari, T. Elsayed, R. Suwaileh, Overview of the CLEF-2024 CheckThat! Lab Task 5 on Rumor Verification using Evidence from Authorities, in: [16], 2024.
[21] P. Przybyła, B. Wu, A. Shvets, Y. Mu, K. C. Sheang, X. Song, H. Saggion, Overview of the CLEF-2024 CheckThat! lab task 6 on robustness of credibility assessment with adversarial examples (incrediblae), in: [16], 2024.
[22] P. He, J. Gao, W. Chen, Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, 2021. arXiv:2111.09543.
[23] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018). URL: http://arxiv.org/abs/1810.04805. arXiv:1810.04805.
[24] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.
[25] W. Antoun, F. Baly, H. Hajj, Arabert: Transformer-based model for arabic language understanding, arXiv preprint arXiv:2003.00104 (2020).
[26] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. L. Scao, T. Lavril, T. Wang, T. Lacroix, W. E. Sayed, Mistral 7b, 2023. arXiv:2310.06825.
[27] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., Llama: Open and efficient foundation language models, arXiv preprint arXiv:2302.13971 (2023).