AIT_FHSTP at CheckThat! 2022: Cross-Lingual Fake News Detection with a Large Pre-Trained Transformer

Mina Schütz[1], Jaqueline Böck[2], Medina Andresel[1], Armin Kirchknopf[2], Daria Liakhovets[1], Djordje Slijepčević[2] and Alexander Schindler[1]

[1]*Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria*
[2]*St.Pölten University of Applied Sciences, 3100 St. Pölten, Austria*

Abstract

The increase of fake news in today's society, partially due to the accelerating digital transformation, is a major problem in today's world. This year's CheckThat! Lab 2022 challenge addresses this problem as a Natural Language Processing (NLP) task aiming to detect fake news in English and German texts. Within this paper, we present our methodology and results for both, the monolingual (English) and cross-lingual (German) tasks of the CheckThat! challenge in 2022. We applied the multilingual transformer model XLM-RoBERTa to solve these tasks by pre-training the models on additional datasets and fine-tuning them on the original data as well as its translations for the cross-lingual task. Our final model achieves a macro F1-score of 15,48% and scores the 22th rank in the benchmark. Regarding the second task, i.e., the cross-lingual German classification, our final model achieves an F1-score of 19.46% and reaches the 4th rank in the benchmark.

Keywords

Fake News Detection, Pre-Training, Transformer, Cross-Lingual

1. Introduction

Due to the information overload on the web and the rapid spread of content on social media platforms, fake news articles circulate faster and are difficult to distinguish from journalistic articles [1]. The term *fake news* is commonly used since the US presidential election in 2016 [2] and can include multiple aspects of incorrect information propagation, such as propaganda, pure fabrications, hoaxes, click-bait and rumors [2, 3, 4].

In this year's shared task at CLEF2022 CheckThat! Lab [5] the third task is fake news detection with four classes [6, 7, 8]: *false, partly false, true,* and *other*. We decided to take part in both fake news detection sub-tasks: a) English and b) German. The latter was proposed as a cross-lingual task without training data in German language. We propose a large pre-trained XLM-RoBERTa model [9], which we additionally pre-trained on a non-publicly available dataset with roughly 200,000 news articles from journalistic as well as citizen sources, such as blogs. After pre-training the model, we fine-tuned it with the given English training data as well as their translations into German to increase its generalization ability.

Our paper is structured as follows: The first Section 1.1 presents the current state-of-the-art and related work. Section 1.2 describes our methodological approach, including the employed datasets and models. Our experimental setup is explained in Section 2, followed by a documentation of the results (Section 3) and a discussion with final conclusions (Section 4).

1.1. Related Work

Detecting fake news is a task that is becoming increasingly important due to digitalization and the rapid spread of information. Current approaches to detect fake news can be categorized into: *feature-based approaches* which describe the task of learning different writing styles and *knowledge-based approaches*, where the model learns latent information about the text and its domain [10]. Gasparetto et al. [11] provided a structured and comprehensive overview of the existing methods for text classification. In the past, supervised machine learning (ML) models such as stochastic gradient descent (SGD), support vector machines (SVM), linear support vector machines (LSVM), k-nearest neighbour (KNN) and decision trees (DT) have been used for solving that task [12]. These methods have been overtaken soon by deep learning (DL) models like long short-term memory (LSTM), convolutional neural networks (CNN) and attention-based bidirectional long short-term memory (BiLSTM with Attention) models. However, transformer models including BERT, ALBERT and XLNET outperformed recent ML and DL algorithms [13]. These are becoming increasingly popular as pre-trained models trained on large corpora of data [14] are made publicly available (e.g., https://huggingface.co/). Another popular transfomer is T5, which was used by Sabry et al. [15]. The authors trained an English T5 transformer (t5-base) for an English hate speech classification task and compared the results to several other state-of-the-art classification models; The authors stated that their T5 model outperformed the RoBERTa model in all tasks. This results show that the use of sequence-to-sequence models like T5 can be beneficial when it comes to text classification.

Fine-tuning a pre-trained ML models to the target data is used in many deep learning applications, especially for small datasets. Previous studies have shown that models which got pre-trained and fine-tuned on similar data as the task-specific one leads to improvements in performance of the models [16, 17]. This has also been demonstrated in our work, where we have used additional datasets (external datasets and translations) for pre-training and fine-tuning [18].

1.2. Methodological Approach

In this paper we propose a feature-based approach for fake news detection. We define pre-training as unsupervised re-training of a transformer model and fine-tuning as supervised training on the specific classification task. For training our models we used additional as well as translated data.

- **Pre-Training Strategy:** Transformer models are usually already pre-trained on a large set of generic text data [14]. However, to adapt these models to a specific classification task, we experiment with further pre-training on domain-related data (which might be relevant to the classification task).

- **Fine-Tuning Strategy:** Training pre-trained models for a given downstream task on the given training data for classification is also called fine-tuning. This can be performed either on the upper layers of the model or on all layers.

1.3. CheckThat! 2022 Data (CT)

This year's training data consisted of 900 news articles and an additional development set containing 364 instances - both only in English [19]. The test set in English (sub-task a) contained 612 articles, the one in German 586 (sub-task b). The dataset contained in total four different classes with the following split for the training set [20, 21]: partially false (217), false (465), true (142), and other (76); and the development set: *partially false* (141), *false* (113), *true* (69), and *other* (41). Since the provided datasets do not include German data, we translated the original English CheckThat data to German using Google Translator.

1.4. External Data:

In this section, we describe briefly the external data we used for pre-training our models. The abbreviations for the datasets *AD* and *FN* are later used to describe which datasets we used for pre-training. *FN* is used during the following sections for the combination of both presented fake news datasets.

Article Dataset (AD)   This dataset was collected over a period of 1,5 years as part of a nationally funded Austrian research project, i.e., Defalsif-AI https://science.apa.at/project/defalsifai/, and therefore is not publicly available. It contains 194,332 gathered news articles from different sources. The articles are multilingual, however the majority are either in English or German. The articles are not annotated in terms of whether they are Fake News or not, and are only used for pre-training the transformer models.

Fake News Dataset German (FN)   The Fake News Dataset German [22] contains approximately 63,000 fake and non-fake news articles from the fields of economics and sports. As some of the text include HTML and JavaScript snippets we removed lines which contained such snippets.

Fake and real news dataset (FN)   This dataset can be found on Kaggle.com [23] and consists of text data from news, political and other articles. This dataset comprises approximately 19,000 fake and 21,000 non-fake texts. More details on the described data can be found in the papers [24, 25].

2. Experimental Setup

We employed and experimented with two different transformer models: XLM-R [9] and T5 [26]. The experimental setup is depicted in Figure 1. We evaluated each experiment with the same training (90%) and validation split (10%).

- **XLM-R** is a multilingual model, trained on 100 languages, which is designed for standard NLP tasks. The underlying architecture is a combination of RoBERTa [27] and XLM [28] which leads to very good performance, outperforming other state-of-the-art models like the multilingual version of BERT (mBERT) [29] without the use of *Next Sentence Prediction* (relying only on *Masked Language Modeling*) as pre-training strategy.
- **T5** is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks and for which each task is converted into a text-to-text format. The small variant of the English T5 model is pre-trained on the English C4 [26] dataset as well as the Wiki-DPR [30] data. This publicly available model has been fine-tuned in the past for several downstream tasks using different datasets. In our work, this small version of the T5 got further fine-tuned for detecting fake news.

2.1. Unsupervised Pre-Training

For pre-training our models, which we obtained from HuggingFace [1] we experimented with two different additional datasets (AD, FN).

- **T5-PRET:** The smaller version of the T5 model (T5-small) that can be found on Huggingface [26] was trained on:
    1. the original CheckThat data: **T5-PRET-CT**
    2. the original + translated CheckThat data: **T5-PRET-CT-TL**
    3. as well as on a combination of the original CheckThat and the additional fake news dataset (*FN*): **T5-PRET-CT-FN**

  Since the additional fake news datasets (*FN*) are relatively large, experiments were conducted smaller subsets. In this paper, only the results for one of these splits are mentioned, since similar results were obtained for all splits. The additional data did not show better results than the model only re-trained on the original CheckThat data. The mentioned T5-PRET-CT-FN model was re-trained on a split of the fake news datasets (*FN*) which had a length of about 10 million characters. The distribution of English and German texts is about 50%. All models got re-trained with a batch size of 8 and a learning rate of $1e^{-4}$. Each model was trained for about 8 to 15 epochs.
- **XLM-R-PRET-AD:** We trained the available XLM-R model provided by HuggingFace with the *AD* dataset. It was trained for 5 epochs, with a batch size of 16 and a learning rate of $2e^{-5}$. The probability for masked language modeling was 15% as used in the original BERT paper [14], where this type of pre-training was introduced. We only trained it for such a low amount of epochs because the training time took roughly 55 hours on one GPU for all articles.
- **XLM-R-PRET-FN:** The second pre-trained XLM-R transformer we trained was with the other fake news datasets (*FN*). We used a similar strategy by also using the 15% probability for the Masked Language Modeling and a character length of 40 million. Since the available fake news datasets (*FN*) are less than the additional dataset (*AD*), the training time on GPU was only around 13 minutes. We also trained it with a learning rate of $2e^{-5}$ and for 5 epochs and a smaller batch size of 8.
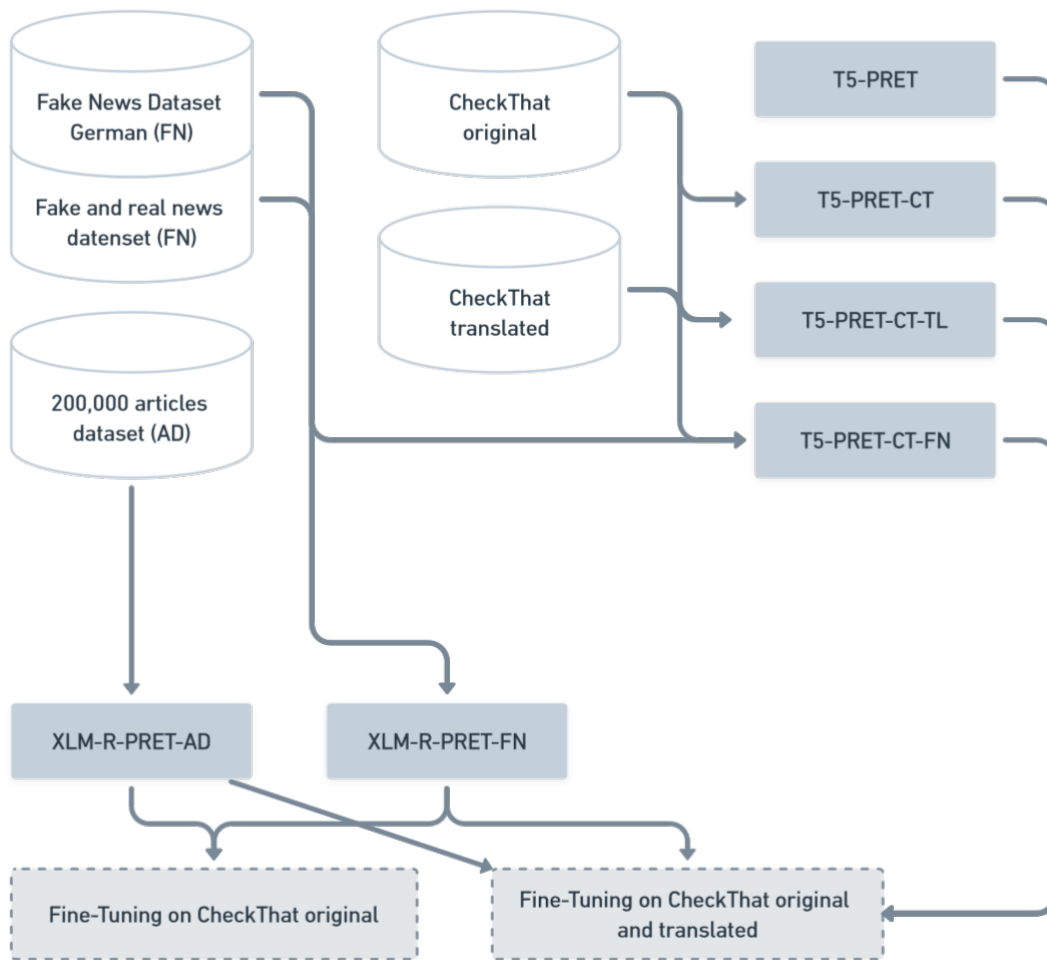
---

[1]https://huggingface.co/

Figure 1: Overview of the experimental setup for training the two transformer architectures, including both training strategies, i.e., unsupervised pre-training and supervised fine-tuning.

2.2. Supervised Fine-Tuning

For fine-tuning our pre-trained models we experimented with different hyperparameters and data combinations. For the fine-tuning of the XLM-R we used the titles as well as the article content. If a title was available it was added to the front of the content, due to the maximum sequence length of models. Using transformer models, usually the content after the maximum sequence length gets padded and not cropped off. This has shown to push the performance on fake news classification tasks in other settings [31] and was also used for the official baseline of this year's shared task [32]. This approach was not used for the T5 model, for which only the article content was used for training.

Our experiments show that the T5 was not performing well due the low amount of data, even though it was pre-trained as well. Hence, after some experiments with the re-trained T5 models

Table 1
Experiment results (in %). The models were all evaluated on a merge of the original English development set and its German translations. The evaluated performance metrics are accuracy, macro-averaged precision, recall, and F1-score.

No.	Model	Dataset	Accuracy	Precision	Recall	F1
1	XLM-R-PRET-AD	CheckThat	53.80	56.77	51.68	51.68
2	XLM-R-PRET-AD	CheckThat	42.93	39.94	36.29	34.00
3	XLM-R-PRET-AD	CheckThat + Translated	53.85	54.31	52.04	50.65
4	**XLM-R-PRET-AD**	**CheckThat + Translated**	**54.81**	**52.56**	**51.59**	**50.29**
5	XLM-R-PRET-FN	CheckThat + Translated	53.30	53.45	50.67	50.16
6	XLM-R-PRET-FN	CheckThat + Translated	43.96	29.03	35.32	29.09
7	T5-PRET	CheckThat	48.35	48.53	44.70	44.24
8	T5-PRET-CT	CheckThat	54.40	49.70	50.89	49.76
9	T5-PRET-CT-TL	CheckThat	49.18	45.59	46.14	45.26
10	T5-PRET-CT-FN	CheckThat	53.02	55.06	48.48	48.98

fine-tuned on the original CheckThat data, we did not continue with further experiments and focused on the better performing XLM-R models. Since we only had a small dataset for the English language we found the best amount of epochs for fine-tuning the XLM-R models is 30, even though it has been shown that usually 3-5 epochs are enough or sometimes even 5-10 epochs - depending on the available dataset size [31]. We additionally experimented with only using the training data as well as the German translations. Table 1 shows the final results of our experiments, which we used to determine our best model for the submission. Experiment 4 (XLM-PRET-AD) shows the best performance in terms of accuracy.

For a more detailed overview we documented all investigated hyperparameters in Table 2. For each experiment we evaluated different learning rates, i.e., $1e^{-5}$ and $2e^{-5}$. Our assumption was that a pre-trained model needs a lower learning rate during fine-tuning, because of the additional data it was pre-trained on. As shown in Table 2 and Table 1 the higher learning rate of $2e^{-5}$ did result in significantly worse predictions on the development set in two setups. For the second pre-trained model XLM-R-PRET-FN, we only performed experiments with the translated data, as preliminary results showed that using the translated data yielded more stable results with both learning rates.

3. Results

Our submitted models for subtask 3a and subtask 3b are both pre-trained on the large additional dataset (AD) and only fine-tuned on the CheckThat data as well as its translations into German (see Table 3). For subtask 3a we rank 22th out of 25 and for the cross-lingual task we rank 4th of 8. The best performing models from other teams for subtask 3a achieved an F1-score of 33.91% and for subtask 3b 29.98%. These results indicate that none of the models achieves a performance that is suitable for real-world applications. An interesting observation is that our model performed better in the cross-language task, even without using training data in German.

Table 2
Investigated hyperparameters.

No.	Epochs	Batch Size	Learning Rate	Max. Seq. Length
1	30	8	$1e^{-5}$	512
2	30	8	$2e^{-5}$	512
3	30	8	$1e^{-5}$	512
4	30	8	$2e^{-5}$	512
5	30	8	$1e^{-5}$	512
6	30	8	$2e^{-5}$	512
7	7	4	$1e^{-4}$	512
8	10	4	$1e^{-4}$	512
9	4	4	$1e^{-4}$	512
10	9	4	$1e^{-4}$	512

Table 3
Results, model, rank. The F1-score is macro-averaged and shown in percent (%).

Model	Dataset	Task	F1	Accuracy	Rank
XLM-R-PRET-AD	CheckThat + Translated	Subtask 3a	15.48	19.93	22th
XLM-R-PRET-AD	CheckThat + Translated	Subtask 3b	19.46	25.42	4th

Table 4
Results for subtask 3a. Values are macro-averaged and shown in percent (%).

Class	Precision	Recall	F1
False	35.80	9.20	14.65
Other	2.81	6.45	3.92
Partially False	11.50	23.21	15.38
True	22.47	37.14	28.07

3.1. Subtask 3a: English

Table 4 shows the overall results per class for subtask 3a (English data). The proposed model performs best for the class *True* (F1: 28.07%). The *False* (F1: 14.65%) and *Partially False* (F1: 15.38%) classes are classified considerably worse. However, XLM-R fails to model the class *Other* at all (F1: 3.92).

The low results for some classes are probably due to the unbalanced class distributions. In general, studies have shown that even humans have difficulty distinguishing between the different fake-news related categories, even for binary classification tasks [33].

3.2. Subtask 3b: Cross-Lingual

To train one model for both subtasks, we translated the original English data into German to train and fine-tune our multilingual XLM-R model on the specific classes. In comparison to our

Table 5
Results for subtask 3b. Values are macro-averaged and shown in percent (%).

Class	Precision	Recall	F1
False	22.03	13.61	16.82
Other	9.09	7.27	8.08
Partially False	13.28	17.52	15.11
True	34.45	41.97	37.84

approach, the baseline approach of the organizers was a standard BERT model, trained on the English CheckThat dataset. For the cross-lingual task, they translated the German test data to English and then evaluated the performance [32]. Even though there was no training data available, our model performs slightly better for each class compared to our results for subtask 3a. The results are shown in Table 5. The results for each class are similar to the results for subtask 3a, where the model performed best for the *True* class (F1: 37.84%) and worst for the *Other* class (F1: 8.08%). We assume that this behavior is due to the fact that the class *Other* class was significantly underrepresented in the training set.

4. Discussion & Conclusion

In this paper, we provide the details on our submission to the CheckThat! 2022 Lab for *Task 3: Fake News Detection*, which consists of two tasks on the classification of fake content. Our experiments show that the unsupervised pre-training strategy of the XLM-R model with additional generic (not task-specific) data with more data instances is the more promising strategy compared to using domain-specific data with fewer training instances. Our model - XLM-R-PRET-AD achieves an F1-score of 15.48% in subtask 3a and 19.46% in subtask 3b. However, the model shows great signs of overfitting, especially on the class *Other*. We conclude that using translations of the original data and using similar content for fine-tuning increases the performance of these models, rather than just fine-tuning them on the provided training data. In future work, we want to compare the influence of pre-training models with more domain-specific data than general content data.

Acknowledgments

References

[1] Z. I. Mahid, S. Manickam, S. Karuppayah, Fake news on social media: Brief review on detection techniques, in: 2018 Fourth International Conference on Advances in Computing,

Communication Automation (ICACCA), 2018, pp. 1–5.

[2] S. A. Khan, M. H. Alkawaz, H. M. Zangana, The use and abuse of social media for spreading fake news, in: 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), 2019, pp. 145–148.

[3] E. Tandoc, Z. W. Lim, R. Ling, Defining "fake news": A typology of scholarly definitions, Digital Journalism 6 (2018) 137–153. URL: https://doi.org/10.1080/21670811.2017.1360143. doi:10.1080/21670811.2017.1360143. arXiv:https://doi.org/10.1080/21670811.2017.1360143.

[4] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, Y. Liu, Combating fake news: A survey on identification and mitigation techniques, ACM Trans. Intell. Syst. Technol. 10 (2019). URL: https://doi.org/10.1145/3305260. doi:10.1145/3305260.

[5] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, The clef-2022 checkthat! lab on fighting the covid-19 infodemic and fake news detection, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2022, pp. 416–428.

[6] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, M. Wiegand, M. Siegel, J. Köhler, Overview of the CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection, in: Proceedings of the 13th International Conference of the CLEF Association: Information Access Evaluation meets Multilinguality, Multimodality, and Visualization, CLEF '2022, Bologna, Italy, 2022.

[7] J. Köhler, G. K. Shahi, J. M. Struß, M. Wiegand, M. Siegel, T. Mandl, Overview of the CLEF-2022 CheckThat! lab task 3 on fake news detection, in: Working Notes of CLEF 2022—Conference and Labs of the Evaluation Forum, CLEF '2022, Bologna, Italy, 2022.

[8] G. K. Shahi, D. Nandini, FakeCovid – a multilingual cross-domain fact check news dataset for covid-19, in: Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media, 2020. URL: http://workshop-proceedings.icwsm.org/pdf/2020_14.pdf.

[9] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[10] Z. Khanam, B. N. Alwasel, H. Sirafi, M. Rashid, Fake news detection using machine learning approaches, IOP Conf. Ser. Mater. Sci. Eng. 1099 (2021) 012040. doi:0.1088/1757-899X/1099/1/012040.

[11] J. Dižo, M. Blatnický, R. Melnik, O. K. https://orcid.org/0000 0003-4677-2535), A mathematical model of operation of a semi-trailer tractor powertrain, Komunikácie (2022).

[12] R. Malhotra, A. Mahur, Achint, Covid-19 fake news detection system, in: 2022 12th International Conference on Cloud Computing, Data Science Engineering (Confluence), 2022, pp. 428–433. doi:10.1109/Confluence52989.2022.9734144.

[13] S. Gundapu, R. Mamidi, Transformer based automatic covid-19 fake news detection system, 2021. URL: https://arxiv.org/abs/2101.00180. doi:10.48550/ARXIV.2101.00180.

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional trans-formers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota, publisher = Association for Computational Linguistics, url = https://www.aclweb.org/anthology/N19-1423, doi = 10.18653/v1/N19-1423, 2019, pp. 4171–4186.

[15] S. S. Sabry, T. Adewumi, N. Abid, G. Kovacs, F. Liwicki, M. Liwicki, Hat5: Hate language identification using text-to-text transfer transformer, 2022. URL: https://arxiv.org/abs/2202.05690. doi:10.48550/ARXIV.2202.05690.

[16] Z. Liu, Y. Xu, Y. Xu, Q. Qian, H. Li, A. B. Chan, R. Jin, Improved fine-tuning by leveraging pre-training data: Theory and practice, 2022. URL: https://openreview.net/forum?id=kQns9y_JH6.

[17] M. E. Peters, S. Ruder, N. A. Smith, To tune or not to tune? adapting pretrained represen-tations to diverse tasks, in: Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019), Association for Computational Linguistics, Florence, Italy, 2019, pp. 7–14. URL: https://aclanthology.org/W19-4302. doi:10.18653/v1/W19-4302.

[18] M. Schütz, J. Boeck, D. Liakhovets, D. Slijepcevic, A. Kirchknopf, M. Hecht, J. Bogensperger, S. Schlarb, A. Schindler, M. Zeppelzauer, Automatic sexism detection with multilingual transformer models ait fhstp@exist2021, in: IberLEF@SEPLN, 2021.

[19] G. K. Shahi, J. M. Struß, T. Mandl, Overview of the clef-2021 checkthat! lab task 3 on fake news detection, Working Notes of CLEF (2021).

[20] G. K. Shahi, A. Dirkson, T. A. Majchrzak, An exploratory study of covid-19 misinformation on twitter, Online Social Networks and Media 22 (2021) 100104.

[21] G. K. Shahi, Amused: An annotation framework of multi-modal social media data, arXiv preprint arXiv:2010.00502 (2020).

[22] A. Ströckl, Fake news dataset german, 2020. URL: https://www.kaggle.com/datasets/astoeckl/fake-news-dataset-german.

[23] C. BISAILLON, Fake and real news dataset, 2020. URL: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset?resource=download.

[24] H. Ahmed, I. Traore, S. Saad, Detecting opinion spams and fake news using text classifica-tion 1 (2018).

[25] H. Ahmed, I. Traore, S. Saad, Detection of online fake news using n-gram analysis and machine learning techniques, in: Lecture Notes in Computer Science, Lecture notes in computer science, Springer International Publishing, Cham, 2017, pp. 127–138.

[26] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, arXiv e-prints (2019). arXiv:1910.10683.

[27] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoy-anov, Roberta: A robustly optimized bert pretraining approach, 2019. arXiv:1907.11692.

[28] A. Conneau, G. Lample, Cross-lingual language model pretraining, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 32, Curran Associates, Inc., 2019. URL: https://proceedings.neurips.cc/paper/2019/file/c04c19c2c2474dbf5f7ac4372c5b9af1-Paper.pdf.

[29] I. Turc, M.-W. Chang, K. Lee, K. Toutanova, Well-read students learn better: On the

importance of pre-training compact models, arXiv preprint arXiv:1908.08962 (2019).

[30] V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, W. tau Yih, Dense passage retrieval for open-domain question answering, 2020. `arXiv:2004.04906`.

[31] M. Schütz, A. Schindler, M. Siegel, K. Nazemi, Automatic fake news detection with pre-trained transformer models, in: D. Bimbo, et al (Eds.), Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021. Lecture Notes in Computer Sciences, volume 12667, Springer, Cham, 2021. doi:`10.1007/978-3-030-68787-8\_45`.

[32] M. Schütz, M. Siegel, Baseline for clef2022 - checkthat! lab task 3, 2022. URL: https://doi.org/10.5281/zenodo.6362498. doi:`10.5281/zenodo.6362498`.

[33] X. Zhou, R. Zafarani, Fake news: A survey of research, detection methods, and opportunities, CoRR abs/1812.00315 (2018). URL: http://arxiv.org/abs/1812.00315. `arXiv:1812.00315`.