A Literature Review of NLP Approaches to Fake News Detection and Their Applicability to Romanian-Language News Analysis

Costin BUSIOC, Stefan RUSETI, Mihai DASCALU

Universitatea Politehnica din București
Politehnica University of Bucharest
Personal e-mail: mihai.dascalu@me.com

A Literature Review of NLP Approaches to Fake News Detection and Their Applicability to Romanian-Language News Analysis

Fighting fake news is a difficult and challenging task. With an increasing impact on the social and political environment, fake news exert an unprecedently dramatic influence on people's lives. In response to this phenomenon, initiatives addressing automated fake news detection have gained popularity, generating widespread research interest. However, most approaches targeting English and low-resource languages experience problems when devising such solutions. This study focuses on the progress of such investigations, while highlighting existing solutions, challenges, and observations shared by various research groups. In addition, given the limited amount of automated analyses performed on Romanian fake news, we inspect the applicability of the available approaches in the Romanian context, while identifying future research paths.

Keywords: fake news identification, Natural Language Processing Techniques, Romanian language

Introduction and Fake News Characterization

Against the changes the world is currently facing, an increasing number of daily tasks are moving online, including reading the news and being informed about relevant topics. The increase in the volume of information made available led, in turn, to fake news dissemination becoming a trending topic on the Internet. Recent events such as the COVID-19 pandemic have shown that fake news exerts a significant negative influence on societies by highlighting stories that do not necessarily report on facts.[1] Moreover, fake news serves to misinform people and manipulate their opinions for several reasons.[2]

Research communities expressed concern about this flood of misinformation and introduced automated fake news identification solutions based on Natural Language Processing (NLP) techniques. In this study, we reference existing datasets and related work in the analysis of English fake news, discuss potential detection techniques for Romanian fake news, as well as establish future work plans for this research initiative.

Literature Review

In this section, we discuss frequently employed methods for compiling fake news corpora, alongside the machine learning solutions developed to analyze data and build classifiers capable of distinguishing between fake and true news.

Datasets

A challenging and crucial step in fake news identification consists of building a relevant corpus containing labeled articles. However, the effort required to compile a clear and objective set of articles, especially one containing thousands of entries, is quite high, as is evident from the analysis performed by Hassan et al.[3] In order to reduce the time and energy invested in such activities, certain research groups empower that the content generated by specific entities is objective and describes facts – for example, CNN[4] and PolitiFact[5].

The Fake News Challenge

One of the most important datasets that is intensively used for testing machine learning models is The Fake News Challenge (FNC-1).[6] The challenge tackles a slightly different, but related task, namely stance detection, whereby the system is presented with the headline and the content of the article, and programmed to decide if the content agrees, disagrees, discusses, or is completely unrelated to the headline. The decision to tackle stance detection instead of the general fake news detection problem was influenced by the difficulty of achieving the latter, even for humans. This specific task is more effectively approached by machine learning models; in this manner, the challenge could attract more attention from researchers and, in turn, solving this challenge would bring the research community a step closer to identifying fake news. Human fact checkers can use such a tool to gather different opinions for a specific headline, which considerably reduces the time needed to research a given subject. The challenge is based on the Emergent dataset,[7] which contains 300 claims and 2,595 associated news articles. New headlines were added and paired with the news articles, resulting in almost 50,000 training examples with an associated stance type.

Satires

Although serving a different purpose, satires are stories which can still be considered fake as their nature is to present events by extrapolating facts in a humorous way. The concept of satires is deeply discussed by Rubin et al.[8] because satires are definitely encouraging the fake news trend, without having this intention. Satires are written in such a way that people with certain interests and a specific sense of humor can understand the joke; however, such text may easily become viral, with other groups of people interpreting the story as completely true. In addition, it is highly possible for such texts to get mixed up with serious news articles and be intensively shared on social media platforms as they usually focus on general interest topics, such as public figures or politics. Since social media platforms only highlight the title of articles, it is not that difficult for people to limit their attention to one or two sentences, and not read the entire story in order to get a sense of the underlying humor. Rubin et al. pinpoint out that, at the time of writing, several of the most read sources are also the least trustworthy, indicating the ease of an unreliable source to become a trusted source of information.

Drawing on the link between fakes news and satire, the corpus Rubin et al. compiled for their experiment comprises of 360 articles, which were added via a 2-step process. First, they collected 120 satires from 2 sources and 120 valid articles from 2 sources. The subjects were split across 12 different categories such as science, business and civics, and the groups of articles were complementary, so as to include similar topics in both a satirical and an objective manner. Second, 120 more articles were collected to add more diversity on the source set.

While analyzing the semantic differences between satires and legitimate articles, a series of interesting facts emerged. For example, satires usually quote obsessively to build up to the story's punchline, whereas real articles merely report on facts. There is also common practice for satires to start the story by reiterating the title and to wrap it up with a humorous conclusion. Based on these observations, characteristics like humor, absurdity, grammar, negative affect, and punctuation were defined for the analyzed texts, with a view to outlining a classification model.

ClaimBuster

ClaimBuster[9] is a tool which detects check-worthy sentences in political discourses and debates, highlighting pieces from candidates' speeches that may require additional attention. Such a solution reduces the time journalists invest in interpreting complex discourses, by spotting top priority phrases in need of a check-worthy analysis before becoming viral and misinforming citizens through controversial information. Roughly 20,000 sentences from the past 15 US Presidential Elections were selected to be analyzed and categorized by volunteers according to one of the following three categories: "Non-Factual Sentence (NFS)" – sentences that contain subjective information or words that do not contain fact related data, "Unimportant Factual Sentence (UFS)" – sentences reporting on well-known information or essential data, "Check-worthy Factual Sentence (CFS)" – sentences that actually need to be identified and feature potential misleading information.

The volunteers interacted with the labeling process through a web page where random sentences were selected to be tagged accordingly. The process provided more contextual details on the sentences to identify the category as effectively as possible and avoid producing entries the compilers felt unsure of. All volunteers were trained before actually using the platform, and most of them were students, professors, and journalists. In addition, payments were offered to stimulate the volunteers' attention, with rewards being awarded in accordance with the number of sentences labeled and their length. Once a specific phrase was similarly classified by multiple users, the label was considered final and was not shown again to future users.

Additional measures were taken to ensure the volunteers' attention and the quality of data they produced. A subset of statements was pre-annotated by experts, thus providing clear baselines; these statements were randomly sent for classification to volunteers, and a quality user index was created for each participant. The index influenced both the level of trust in selecting the statement labels and the formula whereby the financial rewards of the volunteers were calculated. Once the data were completely annotated, the next steps consisted of applying a binary supervised learning classification algorithm, which clusters check-worthy sentences under one category, and the other two types of sentences under another.

Politifact

Politifact[10] is also referenced by Wang,[11] where 12,800 short-statements were uniformly selected based on 6 categories defined by Politifact: pants-fire, false, barely-true, half-true, mostly-true, and true. "The problem of fake news detection is more challenging than detecting deceptive reviews" is highlighted in the paper, as a side note to the 2016 US Presidential Elections. The LIAR dataset described by the paper introduces the potential of applying Machine Learning algorithms to analyze such statements and classify politics-related information, by defining a structured, annotated, and large enough dataset for such experiments.

FEND

More recently, Zhang et al.[12] emphasized how a continuously updated source of truth corpus can alleviate the limitations of applying an algorithm that does not have a sufficient degree of generality to new, unseen articles. By developing the algorithm for such a database, deceptive information would constantly be highlighted and marked for double checking.

FEND (FakE News Detection), the architecture proposed by the authors, is centered on the idea of clustering news with similar subjects into groups, in order to easily compare fake articles with real ones addressing the same topic. The underlying assumption was that CNN and New York Times are legitimate news providers. The authors defined the concept of an event, consisting of a (subject, verb, object) tuple and a topic, represented by a subject and object pair. The initial step consisted of identifying events and topics for every news article, while building clusters that share a similar subset of topics in that group, whereas different clusters do not share common topics in their representative subset.

Two popular algorithms were used to obtain the clusters, Affinity Propagation and K-Means. One large difference between them is that K-Means requires the number of clusters to be defined beforehand, whereas Affinity Propagation does not depend on such preliminary data. First, the Affinity Propagation algorithm was applied to the articles compiled from CNN and the New York Times. The number of clusters obtained were used as input for the K-Means algorithm, which generated a second set of clusters. The authors discovered that the two algorithms returned identical results for the given dataset.

Labelling a new article as true or false is effected through two simple processes: if the clustering algorithm does not redirect the article to an already true built cluster, the article is considered false. If it can be integrated into one of the clusters, the topics of the article are compared to topic of the cluster, and a specific value is attributed based on the threshold value.

Four datasets from different sources known to post tendentious information were created to test fake news detection as follows: advocate.com, naturalnews.com, greenvillegazette.com and politicot.com. The process of identifying fake news consisted of a two-step filtering. First, the clustering is used to detect fake news based on topics known to be fake. Secondly, a credibility score is given based on the differences between the events in a cluster and the events from the analyzed articles. By experimenting with different threshold values for the credibility score, accurate detection results were ultimately obtained, averaging a success rate of 97% with a threshold value of 0.7. In addition, a second experiment was performed with a folding test approach: 75% of the legitimate CNN and New York Times articles being used in the train clustering step, and several random articles from the remaining 25% were mixed with fake news from one of the selected 4 fake sources. The accuracy ranged between 89.55% and 93.77% for the four experiments (one for each fake news article).

Fake News Identification Methods

As most datasets are not large enough to train complex neural networks, classical algorithms were used to tackle the problem of fake news identification. For example, Rubin et al.[13] used a SVM model on their corpus. The results obtained after deploying a ten-fold cross validation process to juxtapose several mixes of characteristics with Tf-Idf vector representations are as follows: 85% accuracy with a 89% recall when using absurdity (i.e., a heuristic accounting for the unexpected introduction of newly named entities in the final sentence of satirical news); 93% accuracy with 82% recall, in the cases were either grammar or punctuation were assessed, and 90% accuracy with 84% recall, when all the features were taken into consideration.

In the case of ClaimBuster, the classification step consisted of several algorithms and experiments based on supervised learning methods, including Multinomial Naive Bayes Classifier (MNB; Kibriya et al.[14]), Support Vector Machine (SVM; Crammer and Singer[15]), and Random Forest Classifier (RFC; Hassan et al.[16]). The models included features such as sentence sentiment and word or part-of-speech counts. Since the size of the dataset grew, the SVM model was the only framework to preserve a high level of accuracy, averaging a precision rate of 72%, a recall of 67%, and an accuracy of 96% when compared to the top 100 phrases.

During the 2016 US Presidential Elections, ClaimBuster covered the debates alongside two trusted entities, CNN and Politifact, in an attempt to identify the check-worthy sentences in the candidates' speeches; 40% of the selected items were commonly singled out, whereas 25% of the sentences were only marked by ClaimBuster, therefore highlighting its potential for success in such scenarios.

The classifiers experimented on the LIAR dataset[17] consisted of a majority baseline, a regularized logistic regression classifier (LR), an SVM, a bi-directional long short-term memory networks model (Bi-LSTMs)[18], and a convolutional neural network model (CNNs)[19]. Among the solutions, the CNN approach surpassed the other ones, achieving a .270 accuracy on the 6-classes classification task.

Several models were tested on FNC-1, which is large enough to test more complex NLP models. Bhatt et al.[20] experimented with the usual methods used in NLP for encoding and then comparing two fragments of text, like CNN, BiLSTM, BiLSTM with attention, or CNN + BiLSTM. All these models did not fare well, performing worse than a simple TF-IDF baseline. Instead, Bhatt et al. combined a pretrained skip-thought network to generate embeddings for the headline, and the text with TF-IDF vectors, together with hand-crafted features counting the overlap of words and character n-grams from the two texts. Until Transformer-based models[21] started being used in NLP tasks, this was the best performing solution on the FNC-1 dataset with a weighted accuracy of 83%. Slovikovskaya[22] tested the more popular Transformer architectures, like Bert[23], XLNet[24], and RoBERTa[25] on the FNC-1 dataset, and improved the best result significantly, with RoBERTa achieving a weighted accuracy of more than 89%.

Potential Automated Detection Tools for Romanian-Language Fake News

Particular attention must be paid to Romania and the Republic of Moldova, as in the case of low-resource languages. Fake news has not yet been the topic of major public debates, neither in the mass media nor in academic national journals. Yet the 2018 Flash Eurobarometer showed that Romania is the European country with the highest incidence of consumers to use and trust unfiltered online content. Numerous government and independent reports have signaled that, in this country, the dissemination of fake news is not only a consequence of an "orchestrated misinformation" campaign[26] but also "a threat to national security."[27] Therefore, the apparent lack of impact of fake news on Romanian society is an indication not of the relative absence of the phenomenon but of insufficient observation and understanding. An in-depth analysis of fake news in the Romanian language is not only a priority but also a matter of urgency.[28]

As with any other task, the construction of a fake news classifier must start with a dataset of labeled examples. As seen in other studies, building such a dataset is a difficult and laborious task, which involves multiple annotators in order to reduce the level of subjectivity in the dataset. In addition, it is of paramount importance to produce a clear definition, which prevents the phenomenon from being mistaken for other similar types of discourse. To this end, we envision fake news as a narrative microgenre of the journalistic discourse whereby false information or inferences are willfully disseminated with a view to producing an immediate practical effect in the target audience. The corpus could be developed on a taxonomy of 6 surveyed types of news: (i) true news, (ii) plausible news, (iii) propaganda news, (iv) fabricated news, (v) fictional news, and (vi) satirical news. Of them, only categories (iii) and (iv), in which facts themselves are fabricated, are fake news in every sense of the concept, with the other categories serving to single out and delineate the former two.

An approach similar to Zhang et al.'s[29] can be adopted here, in that a relatively small number of highly credible sources may be used as reference points. This way, the dataset construction could be substantially simplified, yet in turn, the variability of subjects and writing styles would be significantly reduced. This might limit the capability of the model to generalize for new uncharted texts. Our goal is to build a large and diversified dataset with articles from multiple sources that cover as many subjects as possible. This is the main setback encountered when constructing an accurate classifier, as pretrained Natural Language Processing models like RoBERT[30] are already available for Romanian. From this point of view, previous analyzes, such as Dragomir et al.[31] and Terian et al.[32] can serve as possible models.

Conclusions and Future Work

In this paper, we highlighted the impact of fake news at a social level and examined the challenges faced when combating this phenomenon, which can be successfully addressed by devising automated detection methods. We assessed the available solutions, potential obstacles, the results obtained by using various NLP techniques, and the data recent experiments relied on. Drawing on the existing international literature, and in places on our own related research findings, we also outlined potential approaches to the automated detection of Romanian fake news articles and the key points at issue.

Acknowledgement:

This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-1794, within PNCDI III.

Notes:

1. Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson J. Lu, and David G. Rand. "Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-nudge Intervention." *Psychological Science* 31, no. 7 (20202. Hunt Allcott, Matthew Gentzkow, and Chuan Yu. "Trends in the Diffusion of Misinformation on Social Media." *Research and Politics* 6, no. 2 (2019): 1-8.
3. Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne, "Toward Automated Fact-checking: Detecting Check-worthy Factual Claims by Claimbuster," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Part F1296 (New York: ACM, 2017), 1803–1812.
4. CNN. Retrieved November 2, 2020, from https://edition.cnn.com/
5. PolitiFact. Retrieved November 2, 2020, from https://www.politifact.com/
6. See http://www.fakenewschallenge.org
7. William Ferreira and Andreas Vlachos, "Emergent: A Novel Data-set for Stance Classification," *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016*, *1* (San Diego: ACL, 2016), 1163–1168.
8. Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell, "Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News," *Proceedings of the Second Workshop on Computational Approaches to Deception Detection* (San Diego, ACL, 2016), 7–17.
9. Hassan, Arslan, Li, and Tremayne, "Toward Automated Fact-checking."
10. PolitiFact. Retrieved November 2, 2020, from https://www.politifact.com/
11. William Yang Wang, "'*Liar, liar pants on fire*': A New Benchmark Dataset for Fake News Detection." *ACL 2017: 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2 (Vancouver: ACL, 2017), 422–426.
12. Chao Zhang, Ashish Gupta, Christian Kauten, Amit V. Deokar, and Xiao Qin, "Detecting Fake News for Reducing Misinformation Risks Using Analytics Approaches," *European Journal of Operational Research* 279, no. 3 (2019): 1036–1052.
13. Rubin, Conroy, Chen, and Cornwell, "Fake News or Truth?"
14. Ashraf M. Kibriya, Eibe Frank, Bernhard Pfahringer, and Geoffrey Holmes, "Multinomial Naive Bayes for Text Categorization Revisited," *Australasian Joint Conference on Artificial Intelligence* (Berlin-Heidelberg: Springer, 2004), 488–499.
15. Crammer, Koby and Yoram Singer, "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines," *Journal of Machine Learning Research* 2 (2001): 265–292.
16. Hassan, Arslan, Li, and Tremayne, "Toward Automated Fact-checking."
17. Wang, "'*Liar, liar pants on fire*': A New Benchmark Dataset for Fake News Detection."
18. Sepp Hochreiter and Jürgen Schmidhuber, "Long Short-term Memory," *Neural Computation* 9, no. 8 (1997): 1735-1780; Graves, Alex, and Jürgen Schmidhuber, "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures," *Neural Networks* 18, no. 5-6 (2005): 602–610.
19. Yoon Kim, "Convolutional Neural Networks for Sentence Classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Doha, Qatar: ACL, 2014), 1746–1751.
20. Gaurav Bhatt, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal, "Combining Neural, Statistical and External Features for Fake News Stance Identification," *WWW '18: Companion Proceedings of the The Web Conference* (International World Wide Web Conferences Steering Committee Republic and Canton of Geneva, Switzerland, 2018), 1353–1357.
21. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems* (New York: Curran Associates, 2017), 5998–6008.
22. Valeriya Slovikovskaya and Giuseppe Attardi, "Transfer Learning from Transformers to Fake News Challenge Stance Detection (FNC-1) Task," *Proceedings of the 12th Language Resources and Evaluation Conference* (Marseille: European Languages Resources Association, 2020), 1211–1218.
23. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT 2019* (Minneapolis, Minnesota: ACL, 2019), 4171–4186.
24. Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," *Advances in Neural Information Processing Systems 32 (*New York: Curran Associates, 2019*)*, 5753–5763.
25. Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
25. Sorin Ioniță and Septimiu Pârvu, *Fake news în România și Moldova. Dezinformarea, proasta guvernare și combaterea lor în Europa de Est* (Bucharest: EFOR, 2019).
27. Iuliana Grigore and Andrada Halgaș, "'Fabricate' pentru România. Interese ascunse în știrile false", *Intelligence*, 9 Jan. 2019. https://intelligence.sri.ro/fabricate-pentru-romania-interese-ascunse-stirile-false/
28. Anca-Simina Martin and Simina-Maria Terian, "Știrile false: limite și perspective ale analizei lingvistice," *Transilvania* 49, no.

10 (2020): 72-77.

29. Chao Zhang, Ashish Gupta, Christian Kauten, Amit V. Deokar, and Xiao Qin, "Detecting Fake News for Reducing Misinformation Risks Using Analytics Approaches," *European Journal of Operational Research* 279, no. 3 (2019): 1036–1052.
30. Mihai Masala, Stefan Ruseti, and Mihai Dascalu, "RoBERT – A Romanian BERT Mode," *Proceedings of COLING 2020* (Barcelona: ACL, 2020). Online.
31. Isabela-Anda Dragomir, Maria-Dorinela Sirbu, Mihai Dascalu, Simina-Maria Terian, and Stefan Trausan-Matu, "Exploring Differences in NATO Discourses Using the ReaderBench Framework," *University Politehnica of Bucharest Scientific Bulletin – Series C* 82, no. 1 (2020): 3–18.
32. Simina-Maria Terian, Teodor-Mihai Cotet, Maria-Dorinela Sirbu, Mihai Dascalu, and Stefan Trausan-Matu. "Discourses of Economic Crisis in Romanian Media: An Automated Analysis Using the ReaderBench Framework." *Transylvanian Review* 28, Suppl. 1 (2020): 245–260.

Bibliography:

Allcott, Hunt, Matthew Gentzkow, and Chuan Yu. "Trends in the Diffusion of Misinformation on Social Media." *Research and Politics* 6, no. 2 (2019): 1-8.

Bhatt, Gaurav, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal. "Combining Neural, Statistical and External Features for Fake News Stance Identification." In *WWW '18: Companion Proceedings of the The Web Conference*, 1353–1357. International World Wide Web Conferences Steering Committee Republic and Canton of Geneva, Switzerland, 2018.

CNN. Retrieved November 2, 2020, from https://edition.cnn.com/

Crammer, Koby and Yoram Singer. "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines." *Journal of Machine Learning Research* 2 (2001): 265–292.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of NAACL-HLT 2019*, 4171–4186. Minneapolis, Minnesota: ACL, 2019.

Dragomir, Isabela-Anda, Maria-Dorinela Sirbu, Mihai Dascalu, Simina-Maria Terian, and Stefan Trausan-Matu. "Exploring Differences in NATO Discourses Using the ReaderBench Framework." *University Politehnica of Bucharest Scientific Bulletin – Series C* 82, no. 1 (2020): 3–18.

Ferreira, William, and Andreas Vlachos. "Emergent: A Novel Data-set for Stance Classification." In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016*, *1*, 1163–1168. San Diego: ACL, 2016.

Graves, Alex, and Jürgen Schmidhuber. "Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures." *Neural Networks* 18, no. 5-6 (2005): 602–610.

Grigore, Iuliana, and Andrada Halgaș. "'Fabricate' pentru România. Interese ascunse în ştirile false ['Made' for Romania: Hidden Interests in Fake News]." *Intelligence*, 9 Jan. 2019. https://intelligence.sri.ro/fabricate-pentru-romania-interese-ascunse-stirile-false/

Hassan, Naeemul, Fatma Arslan, Chengkai Li, and Mark Tremayne. "Toward Automated Fact-checking: Detecting Check-worthy Factual Claims by Claimbuster." In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Part F1296, 1803–1812. New York: ACM, 2017.

Hochreiter, Sepp, and Jürgen Schmidhuber. "Long Short-term Memory." *Neural Computation* 9, no. 8 (1997): 1735-1780.

Ioniță, Sorin, and Septimiu Pârvu. *Fake news în România și Moldova. Dezinformarea, proasta guvernare și combaterea lor în Europa de Est* [Fake News in Romania and Moldova: Misinformation, Bad Governance, and the Fight Against Them in Eastern Europe]. Bucharest: EFOR, 2019.

Kibriya, Ashraf M., Eibe Frank, Bernhard Pfahringer, and Geoffrey Holmes. "Multinomial Naive Bayes for Text Categorization Revisited." In *Australasian Joint Conference on Artificial Intelligence*, 488–499. Berlin-Heidelberg: Springer, 2004.

Kim, Yoon. "Convolutional Neural Networks for Sentence Classification." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. Doha, Qatar: ACL, 2014.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." arXiv preprint arXiv:1907.11692, 2019.

Martin, Anca-Simina, and Simina-Maria Terian. "Știrile false: limite și perspective ale analizei lingvistice [Fake News: Limits and Perspectives of Linguistic Analysis]." *Transilvania* 49, no. 10 (2020): 72-77.

Masala, Mihai, Stefan Ruseti, and Mihai Dascalu. "RoBERT – A Romanian BERT Mode." In *Proceedings of COLING 2020*. Online. Barcelona: ACL, 2020.

Pennycook, Gordon, Jonathon McPhetres, Yunhao Zhang, Jackson J. Lu, and David G. Rand. "Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-nudge Intervention." *Psychological Science* 31, no. 7 (2020): 770–780.

PolitiFact. Retrieved November 2, 2020, from https://www.politifact.com/

Rubin, Victoria, Niall Conroy, Yimin Chen, and Sarah Cornwell. "Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News." In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, 7–17. San Diego, ACL, 2016.

Slovikovskaya, Valeriya, and Giuseppe Attardi. "Transfer Learning from Transformers to Fake News Challenge Stance Detection (FNC-1) Task." In *Proceedings of the 12th Language Resources and Evaluation Conference*, 1211–1218. Marseille: European Languages Resources Association, 2020.

Terian, Simina-Maria, Teodor-Mihai Cotet, Maria-Dorinela Sirbu, Mihai Dascalu, and Stefan Trausan-Matu. "Discourses of Economic Crisis in Romanian Media: An Automated Analysis Using the ReaderBench Framework." *Transylvanian Review* 28, Suppl. 1 (2020): 245–260.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, 5998–6008. New York: Curran Associates, 2017.

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V. Le. "XLNet: Generalized Autoregressive Pretraining for Language Understanding." In *Advances in Neural Information Processing Systems 32*, 5753–5763. New York: Curran Associates, 2019.

Wang, William Yang. "'*Liar, liar pants on fire*': A New Benchmark Dataset for Fake News Detection." In *ACL 2017: 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2, 422–426. Vancouver: ACL, 2017.

Zhang, Chao, Ashish Gupta, Christian Kauten, Amit V. Deokar, and Xiao Qin. "Detecting Fake News for Reducing Misinformation Risks Using Analytics Approaches." *European Journal of Operational Research* 279, no. 3 (2019): 1036–1052.