How fake is climate change? A comparison of fake news detection approaches.

Braunegg, Florian[1] and Second Author[2]

[1]Graz University of Technology, Graz, Austria
[2]University of Graz, Graz, Austria

July 8, 2025

Abstract

Your abstract goes here. It should be a single paragraph of 150–250 words summarizing the report's context, aims, methods, results, and conclusions. The entire paper should not exceed 6 pages (excluding limitations, ethical considerations, references, appendix). Please do not forget to fill out the contributions for each team member in the table.

1 Introduction

Fake news have become a major concern in contemporary political and social discourse. They contribute to the polarization and fragmentation of society (Au et al., 2022) and, in extreme cases, may even pose a threat to public safety. This is exemplified by incidents such as Pizzagate, in which an armed man stormed a pizzeria in Washington, D.C., attempting to rescue nonexistent children based on fabricated information (Lipton, 2016). Due to their wide-ranging impact, addressing fake news effectively requires an approach that integrates both social and technical perspectives. While disciplines like psychology and sociology ofte explore how such misinformation spreads and shapes public opinion, computational fields, especially Natural Language Processing (NLP), provide tools to systematically analyze and detect such content. NLP enables the identification of linguistic patterns and contextual signals that are typical of deceptive or false information. Consequently, the reliable detection and classification of fake news remains a core challenge that NLP is uniquely equipped to address.

As part of the "Natural Language Processing" course at Graz University of Technology, we analyzed and classified fake news in three consecutive stages, each illustrating increasingly sophisticated NLP techniques for fake news detection.

2 Related Work

Fake news can be broadly understood as information presented in the form of news that is intentionally false and designed to mislead its audience. While the definition is not universally agreed upon, broader interpretations such as the one adopted in this paper treat fake news as an umbrella term that includes both misinformation (false information shared without intent to deceive) and disinformation (false information shared deliberately). In this broader perspective, any misleading or incorrect content presented as news, regardless of intent, may be classified as fake news (de Oliveira et al. 2021; for a discussion of the terminology see Tandoc et al. 2017). Although instances of fake news can be traced as far back as 2100–1200 BC (Roozenbeek and van der Linden, 2024), scholarly interest in the phenomenon has only gained significant momentum in recent years (Tătaru et al., 2024). Despite definitional differences, fake news have far-reaching consequences for social cohesion, public opinion, institutional trust, and political development. Prominent examples of these consequences include the election of U.S. President Donald Trump (Allcott and Gentzkow, 2017), the Brexit referendum (Orlando, 2023), and the COVID-19 crisis (Ferreira et al., 2022), all of which are closely linked to the digitization of society.

Interest in fake news detection within the field of NLP began to grow significantly in the second decade of the 21st century. Early research on fake news detection primarily relied on traditional machine learning algorithms and, in some cases, rule-based systems. These systems operate by defining sets of rules or linguistic heuristics to classify articles. Rule-based systems generally exhibit lower accuracy compared to more advanced approaches, primarily due to their limited ability to capture contextual dependencies and adapt to dynamic linguistic variations. Consequently, they are prone to generating a high rate of false positives, especially when applied to sophisticated fake narratives or texts that deviate from their predefined rule configurations (Polu, 2024; Repede and Brad, 2023). Applications of such rule-based approaches can be found, for example, in Alotaibi and Alhammad (2022), who examined the spread of Arabic fake news during the COVID-19 pandemic, and in Yuliani et al. (2019), who developed a rule-based framework for hoax detection. When applied to larger or more complex text corpora, traditional machine learning algorithms such as Logistic Regression or Random Forest often outperform rule-based systems. Traditional machine learning algorithms rely on statistical modeling techniques and are capable of learning patterns directly from data. Although traditional machine learning algorithms are more flexible than rule-based systems, they still lack the ability to capture contextual information within text and depend heavily on manually engineered features and expert domain knowledge. Moreover, large amounts of data are often required to effectively train and fine tune these algorithms for specific tasks. Consequently, the performance and applicability of such models are highly dependent on how the problem is defined and on the quality of the manually engineered features they utilize (Pittman, 2025; Polu, 2024). Despite these limitations, comparative studies such as that by Sudhakar and Kaliyamurthie (2022) demonstrate the considerable potential of traditional machine learning methods. The advent of deep learning marked a significant advancement in computational classification tasks, with models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs) capable of learning hierarchical and sequential patterns directly from data. However, the introduction of the transformer architecture and the subsequent development of models such as BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), and GPT (Radford et al., 2018) represented a major breakthrough in natural language processing. Unlike earlier architectures, transformers leverage contextual embeddings and attention mechanisms to capture subtle dependencies and nuanced patterns in text that traditional machine learning models cannot handle. However, deep learning approaches still require substantial amounts of data, annotated datasets, and significant computational resources in order to achieve strong performance. Another major limitation of deep learning models is their lack of interpretability, as they often function as black box systems whose internal decision-making processes are difficult to trace or explain (Pittman, 2025; Polu, 2024). In response to the limitations of individual approaches, recent research has proposed hybrid models for fake news detection that aim to combine the strengths of different techniques. Nasir et al. (2021) for example proposed a CNN-RNN hybrid model and Albahar (2021) an SVM-RNN-BI-GT hybrid model for fake news detection.

3 Materials and Methods

Our dataset consists of 150 articles of both fake and real news, collected and provided by students and faculty at Graz University of Technology. For stages two and three, we implemented all code using Python. In the first stage, we manually evaluated all articles to create a labeled dataset for the subsequent stages and to gain an initial understanding of how fake news is written. In doing so, we followed the definition of fake news outlined earlier in the related work section.

In the second stage, we used pandas (McKinney, 2010) to import the dataset and the Random Forest classifier from scikit-learn (Pedregosa et al., 2011) to classify 150 articles as fake or real. We specifically chose features that are straightforward and interpretable, aiming to make our results easily understandable and transparent. These include part-of-speech tags (PoS-tags), fake claim matches, emotion scores, readability and difficulty measures, as well as grammatical and spelling errors, all extracted separately for headlines and body text.

PoS-tags were extracted using spaCy (Honnibal et al., 2020), which also handled the tokenization. We restricted the PoS-tags to the 17 universal PoS categories defined in the Universal Dependencies framework (universaldependencies.org, n.d.) to extract easily interpretable logical units. The frequency of each tag was calculated relative to the total number of tokens in the article.

Our fake claim detection process was based on a knowledge base consisting of two components: entities (with aliases) and concise fake claims. To reduce bias toward our articles and due to time constraints, all entries were extracted using a one-shot prompt from GPT-4o mini (OpenAI, 2025). Both elements were tokenized and lowercased using spaCy to facilitate matching; claims were further lemmatized and filtered with spaCy's stopword list. Entity matching in the articles was performed using spaCy's PhraseMatcher. For each matched entity, all sentences from the article containing it were collected, lowercased, lemmatized, and cleaned of non-alphanumeric tokens and stopwords. Each entity in multi-entity sentences was checked separately. Since exact wording was not guaranteed, synonyms were retrieved via WordNet from NLTK (Bird et al., 2009), then also lemmatized and lowercased. To enhance fuzzy matching, we used Python's SequenceMatcher with a similarity threshold of 0.9 for entities and 0.7 for claims. Each entity, claim pair was counted only once per article, regardless of the number of occurrences. The final feature was calculated as the number of unique matched pairs divided by the total number of sentences in the article.

To detect emotions in the text, we compared all tokens extracted with spaCy against an extended version of the NRC Word–Emotion Association Lexicon (Mohammad and Turney, 2013), as provided by Bailey (n.d.). For each article, we calculated the relative frequency of each emotion in proportion to all emotion-bearing terms. Based on insights from our manual analysis in Stage 1, we focused on four prevalent emotions: anger, anticipation, fear, and sadness.

To assess article readability, we used the Flesch Reading Ease (Flesch, 1948) and the Automated Readability Index (ARI) (Senter and Smith, 1967), both calculated with textstat (Bansal, n.d.). The Flesch score ranges from 0 (very difficult) to 100 (very easy), while the ARI reflects U.S. grade levels. Additionally, we measured word difficulty as the proportion of difficult words, based on the list from Chall and Dale (1995).

Grammatical and spelling errors were detected using LanguageTool for Python (Morris, n.d.). We calculated the number of detected issues relative to the total word count as an additional feature.

To determine the optimal feature set and model parameters, we used an 80/20 holdout split. The models hyperparameters were tuned using GridSearchCV from scikit-learn with accuracy as the evaluation metric, and overfitting was controlled with RepeatedStratifiedKFold (five folds, three repeats). Feature selection was refined using the Boruta algorithm (Homola, n.d.; Kursa and Rudnicki, 2010), and feature importance was assessed by the mean decrease in impurity from the Random Forest implementation. We decided to retain all features above the 90th percentile in importance. The final model performance was evaluated using repeated holdout validation across 1000 random splits, reporting accuracy, precision, recall, F1-score, feature importances, and mean feature values.

4 Results

Using the Boruta algorithm described in the last section, we identified 13 features as statistically significant predictors for distinguishing fake from real news with our Random Forest classifier. Table 1 presents the evaluation metrics of the model in comparison to a mean baseline.

Table 1: Evaluation metrics (mean and 95% CI) over 1000 random splits.

Random Forest Classifier	
Accuracy	0.842 (0.837-0.846)
Precision	0.848 (0.844-0.852)
Recall	0.841 (0.836-0.845)
F1-Score	0.840 (0.835-0.844)
Mean Baseline	
Accuracy	0.506 (0.500-0.512)
Precision	0.505 (0.499-0.511)
Recall	0.504 (0.498-0.510)
F1-Score	0.500 (0.494-0.506)

The Random Forest classifier achieved a mean accuracy of 0.839 (0.834–0.843), with similarly high precision, recall, and F1-scores. In contrast, the baseline only reached around 0.5 on these metrics, highlighting the added value of our feature set.

To assess the contribution of each selected feature to the model's predictive performance, Table 2 presents their relative importance and mean values. The headline-based PoS-tag ratios and the ARI together account for nearly 60% of the overall feature importance. Among emotion features, only anticipation contributes meaningfully, and it is more prevalent in real news. Grammatical function words, error counts, and the knowledge-base feature have negligible impact. Feature means further show that real news tend to have higher noun and verb ratios and more readable headlines, whereas fake news rely more on proper nouns, pronouns, adjectives, determiners and punctuation.

Table 2: Top 13 features ranked by mean decrease in impurity (MDI) and their mean values for fake and real news over 1000 random splits.

Feature	**MDI**	**Fake**	**Real**
Proper noun ratio in headline	0.167	0.316	0.125
Noun ratio in headline	0.128	0.142	0.292
Anticipation word ratio	0.078	0.082	0.107
Verb ratio in headline	0.077	0.085	0.141
Subordinating conjunction ratio	0.074	0.021	0.018
Punctuation ratio in headline	0.073	0.134	0.073
Adjective ratio	0.071	0.083	0.072
Verb ratio	0.068	0.101	0.111
ARI in headline	0.066	10.87	8.97
Noun ratio	0.060	0.212	0.225
Proper noun ratio	0.053	0.061	0.068
Pronoun ratio	0.050	0.044	0.035
Determiner ratio in headline	0.037	0.050	0.020

MDI: Mean Decrease in Impurity.

5 Discussion

Our qualitative analysis in Stage 1 revealed that fake news articles are characterized by sensational and sometimes very emotionally charged narratives, frequent misinterpretations of sources, and the advancement of one sided or even conspiratorial viewpoints. This is often mirrored in the headlines of the articles, which serve as eye catching hooks, featuring named individuals, organizations, or groups to lend authority or appeal to emotions, or employ mockery and ridicule to delegitimize opposing viewpoint.

Rather than relying primarily on accepted facts, fake news articles achieve their effect through provocative, informal rhetorical devices that shape their style. In contrast, real news tends to exhibit a more formal, informative, and reportative style, as reflected in the different PoS-tag patterns and the ARI captured by our Random Forest classifier. Notably, in the context of climate change reporting, real news articles also place greater emphasis on anticipation, signaling a forward looking and constructive framing of information. These stylistic contrasts, while most apparent in the headlines, are not limited to them but extend consistently throughout the body of the text.

Recognizing these recurring stylistic and narrative patterns proved crucial for feature engineering and for the effective use of our traditional machine learning method. Since such approaches depend on explicitly defined features, our qualitative analysis provided valuable grounding in what to look for beyond surface level content. The resulting set of descriptive features enabled us to capture the relevant stylistic and rhetorical patterns with only twelve features, demonstrating that a compact, well-chosen feature set can provide robust performance for fake news detection in our domain.

While we included a rule-based knowledge base for explicit factual errors, this approach quickly reached its limits. The changing and creative ways misinformation is expressed made it difficult for our system to keep pace, especially regarding subtle or stylistic strategies. Our attempts to make the system more flexible ultimately resulted in an increase of false positives, underscoring the need for adaptable, data-driven models that better reflect the complexity of real world texts.

6 Conclusion

Summarize the key findings and implications (design your report that one get the main insights from reading abstract/introduction/conclusions and glancing at the illustrations). Suggest future research (very briefly).

Limitations

Since our dataset comprises only 150 articles, the models are highly tailored to this small sample. Additionally, the data was collected over a short period, focusing primarily on recent fake news related to climate change. To improve especially the external validity and generalizability of our results, a larger dataset covering a broader range of topics and a longer collection timeframe would have been necessary.

As outlined in the methodology, we also relied on GPT-4o mini to generate information for our knowledge base in Stage Two. While GPT-4o mini is a powerful tool for producing general statements about fake news, it lacks the depth and nuance that manual data collection can offer. As a result, the knowledge base used for our rule-based algorithm, whose outputs served as features for the Random Forest classifier, remained quite generic and therefore less robust.

Contributions

Table 3: Contribution

Name	Contribution
Florian Braunegg	Stage0, Stage1, Stage2, Introduction, Related Work, Methods & Materials (Materials, Stage1, Stage2), Results (Stage1, Stage2), Discussion (Stage1, Stage2), Limitations, Acknowledgments
Another Name	Did not even care to show up.

Acknowledgments

This paper was written as part of the "Natural Language Processing" course held during the summer semester 2025 at Graz University of Technology. The materials used were collected by the course participants and individuals affiliated with the university. No monetary resources were involved in the preparation of this paper.

References

Marwan Albahar. A hybrid model for fake news detection: Leveraging news content and user comments in fake news. *IET Information Security*, 15(2):169–177, 2021. doi: 10.1049/ise2.12021.

Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236, 2017. doi: 10.1257/jep.31.2.211.

Fatimah L. Alotaibi and Muna M. Alhammad. Using a rule-based model to detect arabic fake news propagation during covid-19. *International Journal of Advanced Computer Science and Applications*, 13(1), 2022. doi: 10.14569/IJACSA.2022.0130114.

Cheuk Hang Au, Kevin K. W. Ho, and Dickson K. W. Chiu. The role of online misinformation and fake news in ideological polarization: Barriers, catalysts, and implications. *Information Systems Frontiers*, 24(4):1331–1354, 2022. doi: 10.1007/s10796-021-10133-9.

Mark Bailey. Nrc lexicon, n.d. https://github.com/DemetersSon83/NRCLex/blob/master/nrc_en.json. Accessed: 2025-06-17.

Shivam Bansal. textstat, n.d. https://github.com/textstat/textstat. Accessed: 2025-06-17.

Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, Inc., 1st edition, 2009. ISBN 0596516495.

Jeanne S. Chall and Edgar Dale. *Readability Revisited: The New Dale–Chall Readability For-*

mula. Brookline Books, Cambridge, MA, 1995. ISBN 1571290087.

Nicollas R. de Oliveira, Pedro S. Pisa, Martin Andreoni Lopez, Dianne Scherly V. de Medeiros, and Diogo M. F. Mattos. Identifying fake news on social networks based on natural language processing: Trends and challenges. *Information*, 12(1), 2021. doi: 10.3390/info12010038.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *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)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423.

Carlos Ferreira, Vivian Ferreira, Rebeca Bandeira, and Luisa Ferreira. The impact of misinformation on the covid-19 pandemic. *AIMS Public Health*, 9(2):262–277, 2022. doi: 10.3934/publichealth.2022018.

Rudolf Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221–233, 1948. doi: 10.1037/h0057532.

Daniel Homola. Borutapy: An all relevant feature selection method for python, n.d. https://github.com/scikit-learn-contrib/boruta_py. Accessed: 2024-06-17.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spacy: Industrial-strength natural language processing in python, 2020.

Miron B. Kursa and Witold R. Rudnicki. Feature selection with the boruta package. *Journal of Statistical Software*, 36(11):1–13, 2010. doi: 10.18637/jss.v036.i11.

Eric Lipton. Man motivated by 'pizzagate' conspiracy theory arrested in washington gunfire. *The New York Times*, December 2016. https://www.nytimes.com/2016/12/05/us/pizzagate-comet-ping-pong-edgar-maddison-welch.html?_r=0. Accessed: 2025-06-16.

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, 2019. https://arxiv.org/abs/1907.11692 Accessed: 2024-06-23.

Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56–61, 2010. doi: 10.25080/Majora-92bf1922-00a.

Saif M. Mohammad and Peter D. Turney. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3): 436–465, 2013. doi: 10.1111/j.1467-8640.2012.00460.x.

Jack Morris. language-tool-python, n.d. https://github.com/jxmorris12/language_tool_python. Accessed: 2025-06-17.

Jamal Abdul Nasir, Osama Subhani Khan, and Iraklis Varlamis. Fake news detection: A hybrid cnn-rnn based deep learning approach. *International Journal of Information Management Data Insights*, 1(1):100007, 2021. doi: 10.1016/j.jjimei.2020.100007.

OpenAI. Gpt-4o mini (large language model), 2025. https://chat.openai.com/ Accessed: 2024-06-17.

Vittorio Orlando. *Post-Truth Politics, Brexit, and European Disintegration*, pages 103–127. Springer International Publishing, Cham, 2023. doi: 10.1007/978-3-031-13694-8_6.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Jason M. Pittman. Truth in text: A meta-analysis of ml-based cyber information influence detection approaches, 2025. https://arxiv.org/abs/2503.22686. Accessed: 2025-06-17.

Omkar Reddy Polu. Ai-based fake news detection using nlp. *International Journal of Artificial Intelligence & Machine Learning*, 3(2): 231–239, 2024. doi: 10.34218/IJAIML_03_02_019.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. Technical report, OpenAI, 2018. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf. Accessed: 2024-06-23.

Ştefan Emil Repede and Remus Brad. A comparison of artificial intelligence models used for fake news detection. *Bulletin of "Carol I" National Defence University*, 12(1):114–131, 2023. doi: 10.53477/2284-9378-23-10.

Jon Roozenbeek and Sander van der Linden. *The Psychology of Misinformation*. Contemporary Social Issues Series. Cambridge University Press, 2024.

Ross J. Senter and Edward A. Smith. Automated readability index. Technical Report AMRL-TR-6620, Aerospace Medical Research Laboratories, U.S. Air Force, Wright-Patterson Air Force Base, OH, November 1967.

M. Sudhakar and K.P. Kaliyamurthie. Effective prediction of fake news using two machine learning algorithms. *Measurement: Sensors*, 24:100495, 2022. doi: 10.1016/j.measen.2022.100495.

Edson Tandoc, Zheng Lim, and Rich Ling. Defining "fake news": A typology of scholarly definitions. *Digital Journalism*, 6:1–17, 08 2017. doi: 10.1080/21670811.2017.1360143.

George-Cristian Tătaru, Adrian Domenteanu, Camelia Delcea, Margareta Stela Florescu, Mihai Orzan, and Liviu-Adrian Cotfas. Navigating the disinformation maze: A bibliometric analysis of scholarly efforts. *Information*, 15(12), 2024. doi: 10.3390/info15120742.

universaldependencies.org. Universal pos tags, n.d. https://universaldependencies.org/u/pos/. Accessed: 2025-06-16.

S.Y. Yuliani, Mohd Faizal Bin Abdollah, Shahrin Sahib, and Yunus Supriadi Wijaya. A framework for hoax news detection and analyzer used rule-based methods. *International Journal of Advanced Computer Science and Applications*, 10(10), 2019. doi: 10.14569/IJACSA.2019.0101055.

A Appendix - GPT-4o mini prompt (Stage 2)

Prompt: Write fake news claims about climate change using the exact same structure as the example given. Do not alter the format at any time. Example: { ”entity”: ”Paris Agreement”, ”aliases”: [”PA”, ”Paris Accord”], ”claims”: [”aims to limit global warming to 4°C”, ”was signed by China alone”, ”has no legal force”] }

B Appendix - Usage of AI

Beside the mentioned use of GPT-4o mini in the text above, ”AI” was only used for grammatical and structural corrections. All contents and ideas presented in this paper, where not explicitly marked as the intellectual property of other authors, are solely the result of the authors' own work.

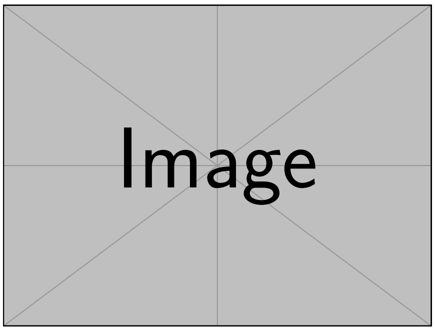

Figure 1: Example figure caption. Please consider to explain to the reader, what is depicted