Towards a Computational Framework for Distinguishing Critical and Conspiratorial Texts by Elaborating on the Context and Argumentation with LLMs

Notebook for PAN at CLEF 2024

Ariana Sahitaj[1], Premtim Sahitaj[1,2], Salar Mohtaj[1,2], Sebastian Möller[1,2] and Vera Schmitt[1,2]

[1]*Quality and Usability Lab, Technische Universität Berlin, Berlin, Germany*
[2]*German Research Center for Artificial Intelligence (DFKI), Berlin, Germany*

Abstract

The shared task of PAN 2024 addresses the need to distinguish between critical and conspiratorial texts in relation to public health measures during the COVID-19 pandemic. In early 2020, the pandemic caused a simultaneous rise in misinformation and conspiracy theories, leading to an 'infodemic' that increased societal insecurity. This notebook introduces an experimental computational framework leveraging Large Language Models (LLMs) for contextual and argumentative elaborations to enhance the classification accuracy of a reference DeBERTa base classification model. Our approach involves automatic annotations of intent and argumentation style, hypothesizing that these features aid in differentiating between conspiracy and critical texts. Experimental results, however, reveal that DeBERTa performs best without these elaborations, achieving an MCC of 0.838 and F1-macro of 0.917. The inclusion of LLM-generated feature annotations did not surpass the baseline performance. These findings suggest that while theoretically valuable, the practical application of such elaborations requires further refinement. Future work should focus on optimizing LLM outputs and exploring alternative techniques to enhance text classification without overloading models with excessive information.

Keywords

Critical Thinking, Conspiracy Theory, LLMs, Argumentation

1. Introduction

In early 2020, the World Health Organization (WHO) declared a global pandemic. Alongside the increase in new infections, the COVID-19 pandemic also sparked a concurrent *infodemic*, in which fake news and conspiracy theories spread even more rapidly than the actual virus [6, 4]. During the pandemic, Google searches related to the effect of the virus on health and society saw a significant increase as people sought information amidst the growing uncertainty. However, not only searches about the symptoms, strains, vaccines were trending, but also concatenations of the term *coronavirus* with keywords such as *"5G"*, *"laboratory"*, and *"ozone"* were extensively utilized during search. [19] The results of these searches, a mix of reliable and (very) unreliable sources and information, contributed to a rapid spread of *critical questions* as well as *conspiracies* on social media [6]. Conspiracy theories are explanations of events or situations that blame them on secret agreements between powerful groups, often relying on a lack of evidence and pattern recognition to see connections where none exist, while promoting a sense of secrecy and cover-up by hidden agendas. These theories are resistant to disconfirmation, holding on to beliefs despite contradictory evidence, and often create an "*us versus them*" mentality, dividing the world into believers and non-believers [21, 22, 27]. Critical thinking involves focused and thoughtful decision-making, guiding us on what to believe or do. It aims to meet high standards of accuracy and sound reasoning, avoiding quick judgments, unsupported claims, or biased reasoning [23]. The spread of false information and conspiracy ideologies has been amplified on platforms with a high

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France
ariana.sahitaj@tu-berlin.de (A. Sahitaj); sahitaj@tu-berlin.de (P. Sahitaj); salar.mohtaj@tu-berlin.de (S. Mohtaj); sebastian.moeller@tu-berlin.de (S. Möller); vera.schmitt@tu-berlin.de (V. Schmitt)
0009-0002-0096-9383 (A. Sahitaj); 0000-0003-3908-5681 (P. Sahitaj); 0000-0002-0032-3833 (S. Mohtaj); 0000-0003-3057-0760 (S. Möller); 0000-0002-9735-6956 (V. Schmitt)

reach and minimal content regulation, among which Telegram has emerged as a significant channel [5]. The danger of encapsulated communities such as private Telegram groups lies in their potential to radicalize users through the spread of (fabricated) conspiracy theories backed by fake news [7, 29]. Fake news refers to information that is typically not verified and inaccurate and can be categorized into *misinformation* (inaccurate information, no harmful intent), *disinformation* (inaccurate information, harmful intent) and *malinformation* (truthful information used maliciously to inflict harm) [3]. When personal beliefs about conspiracy theories are converted to real-world consequences, public safety and societal cohesion is threatened [20]. For example, the spread of disinformation during the pandemic resulted in anti-Asian racism, verbal violence, and violent hate crimes worldwide, falsely linking the origins of the virus to conspiracy theories about Asian foreign influence [8, 9]. The protection of public discourse is vulnerable when misinformation and disinformation threatens to undermine trust in scientific knowledge and institutional measures, especially when critical voices are wrongly labeled as conspiracy theories, potentially driving curious individuals to seek answers outside of official and reliable sources [6, 16, 17, 18, 34].

The necessity to distinguish between critical questioning and conspiratorial thinking has led to the development of the *Oppositional Thinking Analysis* shared task at PAN 2024 [31]. Here, teams are tasked with classifying text as either one of the categories critical and conspiracy. The first category comprises texts that critically examine public health decisions without endorsing conspiracy theories. The second category comprises texts that contribute towards the idea that the pandemic or related public health policies are the result of malicious conspiracies. Identifying and differentiating between these categories is crucial in mitigating the spread of harmful content, while preserving the integrity of public discourse on social media and in the real world.

The rest of the paper is organized as follows: Section 2 reviews recent, relevant research on the task of computational conspiracy theory detection. In Section 3 we summarize the provided dataset for the shared task that has been used to train the models. We describe the conducted experiments, as well as the expectations and hypotheses in detail in Section 4. Finally, the obtained results and the directions for future studies are presented in Sections 5 and 6, respectively.

2. Related Work

While there have been approaches for modeling computational conspiracy theory detection [16, 17, 18], there is a gap in the literature for a computational framework that distinguishes between critical and conspiratorial thinking [34]. Developing such a framework could enhance our ability to understand the computational processes that underlie the creation and spread of conspiracy theories. Thus, improving our ability to identify and address misinformation more effectively. We hypothesize, that the differentiation between critical thinking and conspiracy thinking, among other factors, involves an understanding of the context in which a claim is made and how the claim is argued for [26, 25, 24].

Van Prooijen and Douglas address how societal crises affect the spread of conspiracy theories [10]. They argue that societal crises heighten the need for meaning among individuals, which can lead to a greater susceptibility to conspiracy theories. Conspiracy theories target certain narratives with the intent to mobilize supporters and coordinate them towards certain actions [28]. Thus, the intent behind such theories often relates to providing simple explanations for complex events, offering a psychologically comforting sense of understanding and control during times of uncertainty.

Douglas et al. further explore the psychological foundation of conspiracy theories, describing how these theories serve to fulfill specific psychological needs, such as the aforementioned need for certainty, maintaining a positive self-image, and maintaining control over an increasingly complex environment. The authors highlight that conspiracy theories are speculative, resistant to falsification, and offer broad, internally coherent explanations that serve to insulate beliefs from uncertainty [11]. These theories often appear to thrive in environments where there is a lack of reliable information or when individuals experience distress due to uncertainty.

Gambini et al. conducted a comparative analysis of conspiracy theorists and random users on Twitter, revealing significant differences in their discussions, terminology, and stances on trending topics [32]. They found that conspiracy users often employ more extreme and intense language, such as *"billgates"*, *"vaccinesideeffects"*, and *"wakeup"*, compared to the more moderate language used by random users, such as *"coronavirus"*, *"covid19vaccine"*, and *"fakenews"*.
Giachanou et al. investigated the psycho-linguistic characteristics of conspiracy propagators on social media and found that conspiracy propagators tend to use more swear words and exhibit different personality traits compared to anti-conspiracy propagators [35].
To better understand the psychological basis of conspiracy theory acceptance, it is necessary to explore the relevance to one's individual analytical thinking ability. Douglas et al. suggest that conspiracy beliefs are associated with lower levels of analytical thinking and education [11]. Swami et al. investigate how analytical thinking influences the belief in conspiracy theories. They find that individuals with higher levels of analytical thinking skills are less likely to believe in conspiracy theories, due to their ability to critically analyze the argumentation and evaluate the presented evidence based on the context [12]. Their results suggest that critical thinking involves a more structured approach to processing information, as opposed to the sometimes emotionally driven and less structured argumentation found in conspiracy thinking [12].
Transitioning from analytical thinking to a broader understanding of human reasoning, Mercier and Sperber propose that the primary function of human reasoning is argumentative. This perspective suggests that reasoning is geared towards creating and assessing arguments for persuasion rather than seeking truth, which can explain why individuals are prone to believe in and defend conspiracy theories, as they seek arguments that support their preconceived conclusions [13]. More specifically, this may imply that both conspiracy thinking and critical thinking employ distinct argumentative strategies. Conspiracy thinking often relies on confirmation bias to reinforce pre-existing narrow beliefs, while critical thinking is expected to emphasize the evaluation of evidence from multiple, more open perspectives. Similarly, Ghanem et al. provide insights into fake news detection by modeling the flow of affective information, which includes elements such emotion, sentiment, imageability, and hyperbolic language [14]. Their findings suggest that fake news, often serving as the foundation for conspiracy theories [29], heavily relies on affective elements in its arguments. Thus, we hypothesize that understanding the choice of argumentation within presented claims may be a useful predictor for distinguishing between critical thinking and conspiracy theories.

3. Dataset

The dataset as provided for the PAN 2024 shared task includes a collection of oppositional texts extracted from the Telegram platform, specifically related to the COVID-19 pandemic, available in both English and Spanish. Each entry in the dataset includes a unique identifier, the content of the text, the overall category of the text as either CONSPIRACY or CRITICAL, and a list of useful annotations such as extracted *objectives* or *campaigners*. In the context of this work, we only utilize the textual claim without these annotations to investigate how much information can be extracted automatically from a presented claim.
Table 1 illustrates the dataset statistics, which reveal insights into the textual characteristics and category distribution. In terms of text distribution, the average token length varies between English and Spanish texts, with English texts having an average token length of 124.37 and Spanish texts having an average token length of 258.42. The token length ranges from a minimum of 15 tokens in English to 31 tokens in Spanish, with maximum token lengths of 1307 tokens in English and 1827 tokens in Spanish. The median token length is 87.0 for English and 189.0 for Spanish. The total token count is $497,489$ for English and $1,033,663$ for Spanish. Regarding label distribution, the majority of texts in both languages are categorized as CRITICAL, with 65.53% in English and 63.45% in Spanish. The remaining texts are categorized as CONSPIRACY, with 34.48% in English and 36.55% in Spanish.

Table 1
Distribution of text length and label information for English and Spanish

	English	Spanish
Text distribution		
Average token length	124.37	258.42
Minimum token length	15	31
Maximum token length	1307	1827
Median token length	87.0	189.0
Total token count	497,489	1,033,663
Label distribution		
CRITICAL	2621	2538
	65.53 %	63.45 %
CONSPIRACY	1379	1462
	34.47 %	36.55 %

4. Approach

Our approach towards an experimental computational framework for distinguishing between critical and conspiracy texts is based on the idea of automatically annotating specific characteristics, which we hypothesize to have an impact towards the binary classification task. We follow the findings as described in our brief review of related literature in section 2. Our first idea is based on the assumption that conspiracy theories have a certain intent that may be interpreted as harmful as opposed to critical thinking which should by definition be less destructive. We categorize intent as a specific piece of information of the underlying (hidden) context of a claim. Our second idea is based on the assumption that the tone, language, and structure utilized within conspiracies are different as compared to critical thinking. We categorize these characteristics as a specific case of the argumentation within a claim. For the automatic annotation of these two features, we employ an LLMs and our domain knowledge as gathered from the literature review for prompt design in a few-shot setting. We use an open source model to generate our feature annotations, as research should be accessible, and its components should not be hidden behind closed APIs. Specifically, we utilize Llama3 70B [1] due to its incredible performance on natural language benchmark tasks. Finally, we compare the generated annotations against the original claim by feeding the individual and combined inputs into a standard DeBERTa [15] classification model. Figure 1 illustrates our computational framework.

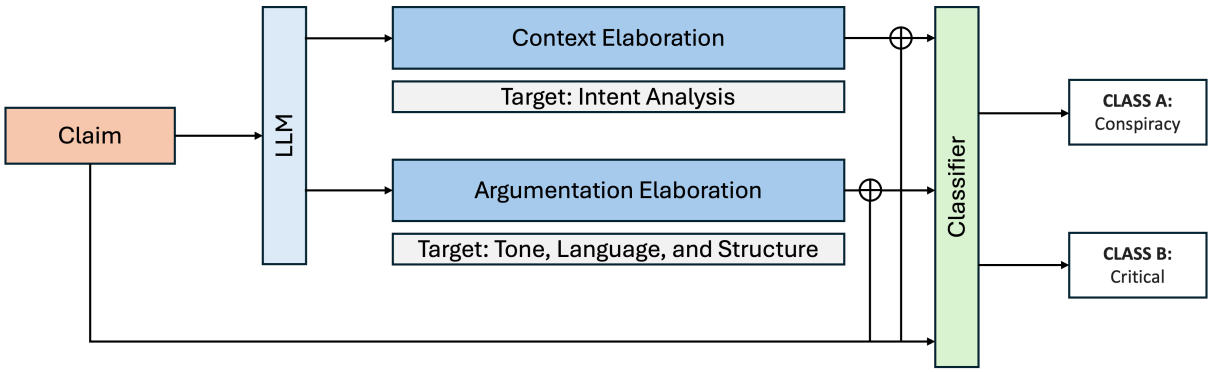

Figure 1: Framework Overview: Employing Large Language Model for Contextual Analysis and DeBERTa for Classification

[1] https://ai.meta.com/blog/meta-llama-3/

4.1. Example

To illustrate our approach, we consider the following claim from the original dataset, which we pass towards to our LLM to automatically generate the described feature annotations.

> **Claim:** "Illegal aliens exempt from covid “vaccine” mandates because they can sue drug companies (but you can’t) https://www.naturalnews.com/2021-12-22-illegal-aliens-exempt-vaccine-mandates-lawsuits-pharma.html"

Context Elaboration In our framework, context elaboration focuses on the specific case of identifying the intent behind a claim, differentiating potentially harmful conspiracies from critical thinking. For the given example claim, the elaboration mechanism identifies the intent as follows:

> **Intent:** The intent appears to be to create a sense of injustice and anger among citizens by suggesting that illegal aliens are exempt from vaccine mandates, while also implying that pharmaceutical companies are avoiding accountability.

Argumentation Elaboration Our approach to argumentation elaboration involves examining the tone, language, and structural features that differentiate conspiracy texts from critical thinking. We postulate that conspiracies utilize distinctive argumentation strategies characterized by emotional appeals, speculative assertions, and a lack of logical coherence. For the given claim, the argumentation elaboration identifies the following features:

> **Tone, Language, and Structure:** The tone is provocative and accusatory, using phrases like ’Illegal aliens’ and ’but you can’t.’ The language is inflammatory, implying unfair treatment and using scare quotes around the word ’vaccine.’ The structure is geared towards sparking outrage and misinformation, as it presents a sensationalized claim without providing balanced context or credible sources. The inclusion of a link to a non-mainstream news website (Natural News) adds to the overall conspiratorial tone of the claim.

By analyzing and annotating the intent and argumentation of a claim, our framework targets the experimental evaluation of these selected features for distinguishing between conspiracy and critical texts. The intent analysis helps in understanding the possible motives behind the claim, while the argumentation analysis reveals how the claim’s argumentation is presented. There are multiple dimensions on which a claim can be analyzed and annotated - due to page constrains, we limited our approaches to the above. In future work, we would like to thoroughly explore this concept of synthetic annotations in the domain of conspiracy theory detection.

4.2. Experimental Setup

We serve the LLM with vLLM [30] to benefit from efficient memory management and utilize two H100 GPUs for the generation of our feature annotations. Table 2 specifies the sampling parameters that we utilized with Llama3 70B.

The temperature parameter, set to 1.0, controls the randomness of the model’s outputs. A higher temperature results in more creative outputs, while a lower value makes the output more deterministic. The chosen value of 1.0 balances maximizes creativity, which can lead to issues with coherence. The top p parameter, set to 0.95, is used for nucleus sampling. It ensures that only the smallest possible set of words whose cumulative probability is greater than or equal to the specified value (0.95) are considered for the next token. This maintains the diversity and relevance of the generated text. Different combinations of the temperature and top p parameters provide a wide range of text styles, balancing

Table 2
Sampling parameters for Text Generation with vLLM

Parameter	Value
Seed	239
Max tokens	512
Temperature	1.0
Top p	0.95
Min p	0.01

between randomness and coherence to suit different applications. These parameters should be empirically evaluated and selected for a given model, task, and domain. We limit the number of tokens generated in a single pass to 512. This is mainly motivated by the maximum input token length for the classification model.

Table 3
Training parameters for Fine-Tuning Classification Model

Parameter	Value
Random state	239
Batch size	20
Batch accumulation	1
Epochs	5
Dropout	0.05
Max length	512
Learning rate	3e-5
Warmup steps	100

For the classification task, we utilize a standard training implementation based on HuggingFace Transformers and PyTorch Lightning. Table 3 specifies the training parameters. We utilize an AdamW optimizer with a learning rate of $3e-5$ and cross-entropy loss. The model is trained for 5 epochs with a dropout of 0.05 and a batch size of 20 on each mode. The model of the epoch with the highest validation metrics is utitized for testing.

4.3. Expectations and Hypotheses

Our initial hypothesis was that providing DeBERTa with richer context and well-structured argumentation would enhance its ability to classify texts accurately. We anticipated that the elaborations would clarify the differences between critical and conspiracy texts, leading to higher evaluation scores. Specifically, we expected the context elaboration to improve the model's understanding of the background and intent of the texts, while the argumentation elaboration was expected to highlight logical flows and key points, aiding in the differentiation process.

5. Experiments and Results

To evaluate our approach, we conducted a series of experiments comparing the performance of DeBERTa with and without the elaborations by LLMs. We utilized metrics such as the Matthews Correlation Coefficient (MCC) and F1 scores for our comparison.

5.1. Dataset Split

The official training data as provided for the shared task is split into three sets: 80% for training, 10% for validation, and 10% for testing. This split ensures that a model has a sufficient amount of data to learn from, while also allowing for robust testing and validation within our presented setting. As the official testing data of the shared task is held out, comparison to other approaches can be made only on the official workshop leaderboard.

5.2. Performance Comparison

We evaluated the performance of DeBERTa with two different elaboration approaches and compared it to our baseline DeBERTa model. The baseline has only been trained on the claims, while the other models have either been trained on the elaborations alone or on the concatenations of claim and respective elaborations. Here we present the performance metrics for the various models, focusing on our test set results.

Table 4
Performance metrics for DeBERTa models with and without various elaborations.

Mode	Test			Train			Validation		
	Loss	F1	MCC	Loss	F1	MCC	Loss	F1	MCC
Claim	0.250	**0.917**	**0.838**	0.050	0.985	0.971	0.236	0.917	0.836
Context w/o	0.357	0.864	0.731	0.126	0.954	0.909	0.317	0.864	0.737
Context w/	0.253	0.913	0.827	0.074	0.972	0.945	0.196	0.918	0.836
Argumentation w/o	0.267	0.885	0.772	0.137	0.952	0.905	0.265	0.895	0.795
Argumentation w/	0.217	0.910	0.821	0.077	0.975	0.951	0.215	0.904	0.809

The DeBERTa baseline performance served as our reference during development. The base DeBERTa model achieves an MCC of 0.83777 and an F1-macro of 0.91733. For the **context elaboration** of the intent, we observed that excluding the claim (Context w/o) led to a drop in performance (MCC: 0.73084, F1-macro: 0.86396). However, including the claim (Context w/) improved the results (MCC: 0.82685, F1-macro: 0.91268), but still did not surpass the performance of the claim classified by DeBERTa alone. Similarly, **argumentation elaboration** on the tone, language, and structure without the claim (Argumentation w/o) resulted in lower performance (MCC: 0.77229, F1-macro: 0.88482). Including the claim (Argumentation w/) improved the metrics (MCC: 0.82115, F1-macro: 0.9104), but again, it did not outperform the claim elaboration. Both context and argumentation elaborations showed improved results when the claim was included, yet neither surpassed the baseline performance of the claim alone. Potential reasons for lack of improvements could be due to overloading the classification models with too long elaborations that should have been controlled with a lower max token parameter during generation. While the elaboration annotations seem to provide interesting analyses of the presented claims, their impact on the results is not as expected.

6. Conclusion

Despite our efforts to enhance our reference model's performance through the use of LLM-generated context and argumentation elaborations, our findings demonstrate that DeBERTa alone, without these additional elaborations, was more effective in accurately categorizing texts. Theoretical frameworks suggesting the benefits of enhanced context and arguments did not translate into practical improvements in our experiments. This suggests that the practical application of these enhancements requires further refinement and optimization to be effective in improving text classification performance. Future work could focus on more targeted and refined elaborations, better alignment of LLM outputs with the classification task, and exploring other techniques to enhance model performance without overwhelming a model with excessive information.

References

[1] M. Fröbe, M. Wiegmann, N. Kolyada, B. Grahm, T. Elstner, F. Loebe, M. Hagen, B. Stein, M. Potthast, Continuous Integration for Reproducible Shared Tasks with TIRA.io, in: Advances in Information Retrieval. 45th European Conference on IR Research (ECIR 2023), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2023, pp. 236–241. URL: https://link.springer.com/chapter/10.1007/978-3-031-28241-6_20. doi:10.1007/978-3-031-28241-6_20.

[2] J. Bevendorff, X. B. Casals, B. Chulvi, D. Dementieva, A. Elnagar, D. Freitag, M. Fröbe, D. Korenčić, M. Mayerl, A. Mukherjee, A. Panchenko, M. Potthast, F. Rangel, P. Rosso, A. Smirnova, E. Stamatatos, B. Stein, M. Taulé, D. Ustalov, M. Wiegmann, E. Zangerle, Overview of PAN 2024: Multi-Author Writing Style Analysis, Multilingual Text Detoxification, Oppositional Thinking Analysis, and Generative AI Authorship Verification, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2024), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2024.

[3] V. Balakrishnan, W. Z. Ng, M. C. Soo, G. J. Han, C. J. Lee, Infodemic and fake news–a comprehensive overview of its global magnitude during the covid-19 pandemic in 2021: A scoping review, International Journal of Disaster Risk Reduction 78 (2022) 103144.

[4] J. Moffitt, C. King, K. M. Carley, Hunting conspiracy theories during the covid-19 pandemic, Social Media+ Society 7 (2021) 20563051211043212.

[5] R. Hohlfeld, F. Bauerfeind, I. Braglia, A. Butt, A.-L. Dietz, D. Drexel, J. Fedlmeier, L. Fischer, V. Gandl, F. Glaser, et al., Communicating covid-19 against the backdrop of conspiracy ideologies (2021).

[6] T. A. Ghebreyesus, Munich security conference, https://www.who.int/director-general/speeches/detail/munich-security-conference, 2020. Accessed: 2024-06-03.

[7] S. Phadke, M. Samory, T. Mitra, What makes people join conspiracy communities?: Role of social factors in conspiracy engagement, CoRR abs/2009.04527 (2020). URL: https://arxiv.org/abs/2009.04527. arXiv:2009.04527.

[8] A. R. Gover, S. B. Harper, L. Langton, Anti-asian hate crime during the covid-19 pandemic: Exploring the reproduction of inequality, American journal of criminal justice 45 (2020) 647–667.

[9] T. Callaghan, M. Motta, S. Sylvester, K. L. Trujillo, C. C. Blackburn, Parent psychology and the decision to delay childhood vaccination, Social science & medicine 238 (2019) 112407.

[10] J.-W. Van Prooijen, K. M. Douglas, Conspiracy theories as part of history: The role of societal crisis situations, Memory studies 10 (2017) 323–333.

[11] K. M. Douglas, R. M. Sutton, A. Cichocka, The psychology of conspiracy theories, Current directions in psychological science 26 (2017) 538–542.

[12] V. Swami, M. Voracek, S. Stieger, U. S. Tran, A. Furnham, Analytic thinking reduces belief in conspiracy theories, Cognition 133 (2014) 572–585.

[13] H. Mercier, D. Sperber, Why do humans reason? arguments for an argumentative theory, Behavioral and brain sciences 34 (2011) 57–74.

[14] B. Ghanem, S. P. Ponzetto, P. Rosso, F. Rangel, FakeFlow: Fake news detection by modeling the flow of affective information, in: P. Merlo, J. Tiedemann, R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, 2021, pp. 679–689. URL: https://aclanthology.org/2021.eacl-main.56. doi:10.18653/v1/2021.eacl-main.56.

[15] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, ArXiv abs/2006.03654 (2020). URL: https://api.semanticscholar.org/CorpusID:219531210.

[16] K. M. Douglas, R. M. Sutton, What are conspiracy theories? a definitional approach to their correlates, consequences, and communication, Annual review of psychology 74 (2023) 271–298.

[17] E. Funkhouser, A tribal mind: Beliefs that signal group identity or commitment, Mind & Language 37 (2022) 444–464.

[18] B. Franks, A. Bangerter, M. W. Bauer, M. Hall, M. C. Noort, Beyond "monologicality"? exploring conspiracist worldviews, Frontiers in psychology 8 (2017) 250235.

[19] A. Rovetta, A. S. Bhagavathula, et al., Global infodemiology of covid-19: analysis of google web searches and instagram hashtags, Journal of medical Internet research 22 (2020) e20673.
[20] B. J. Dow, A. L. Johnson, C. S. Wang, J. Whitson, T. Menon, The covid-19 pandemic and the search for structure: Social media and conspiracy theories, Social and Personality Psychology Compass 15 (2021) e12636.
[21] C. R. Sunstein, A. Vermeule, Conspiracy theories: Causes and cures, Journal of political philosophy 17 (2009) 202–227.
[22] F. Farinelli, Conspiracy theories and right-wing extremism: Insights and recommendations for p/cve, Luxembourg: Publications Office of the European Union (2021).
[23] D. Hitchcock, Critical thinking (2018).
[24] S. Clarke, Conspiracy theories and the internet: Controlled demolition and arrested development, Episteme 4 (2007) 167–180.
[25] K. M. Douglas, J. E. Uscinski, R. M. Sutton, A. Cichocka, T. Nefes, C. S. Ang, F. Deravi, Understanding conspiracy theories, Political psychology 40 (2019) 3–35.
[26] B. Schlipphak, M. Bollwerk, M. Back, Beliefs in conspiracy theories (ct): the role of country context, Political Research Exchange 3 (2021) 1949358.
[27] M. Bruder, P. Haffke, N. Neave, N. Nouripanah, R. Imhoff, Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy mentality questionnaire, Frontiers in psychology 4 (2013) 43078.
[28] A. Marie, M. B. Petersen, Political conspiracy theories as tools for mobilization and signaling, Current Opinion in Psychology 48 (2022) 101440.
[29] D. Halpern, S. Valenzuela, J. Katz, J. P. Miranda, From belief in conspiracy theories to trust in others: Which factors influence exposure, believing and sharing fake news, in: Social Computing and Social Media. Design, Human Behavior and Analytics: 11th International Conference, SCSM 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part I 21, Springer, 2019, pp. 217–232.
[30] W. Kwon, Z. Li, S. Zhuang, Y. Sheng, L. Zheng, C. H. Yu, J. E. Gonzalez, H. Zhang, I. Stoica, Efficient memory management for large language model serving with pagedattention, 2023. arXiv:2309.06180.
[31] D. Korenčić, B. Chulvi, X. Bonet Casals, M. Taulé, P. Rosso, F. Rangel, Overview of the oppositional thinking analysis pan task at clef 2024. working notes of clef 2024, in: N. Ferro, G. Faggioli, P. Galuscakova, A. G. S. D. Herrera (Eds.), Conference and Labs of the Evaluation Forum, 2024.
[32] M. Gambini, S. Tardelli, M. Tesconi, The anatomy of conspiracy theorists: Unveiling traits using a comprehensive twitter dataset, Computer Communications 217 (2024) 25–40. URL: http://dx.doi.org/10.1016/j.comcom.2024.01.027. doi:10.1016/j.comcom.2024.01.027.
[33] D. Korenčić, B. Chulvi, X. Bonet-Casals, M. Taulé, P. Rosso, F. Rangel, Overview of the oppositional thinking analysis pan task at clef 2024, in: G. Faggioli, N. Ferro, P. Galuvakova, A. G. S. de Herrera (Eds.), Working Notes of CLEF 2024 – Conference and Labs of the Evaluation Forum, 2024.
[34] A. A. Ayele, N. Babakov, J. Bevendorff, X. Bonet-Casals, B. Chulvi, D. Dementieva, A. Elnagar, D. Freitag, M. Fröbe, D. Korenčić, M. Mayerl, D. Moskovskiy, A. Mukherjee, A. Panchenko, M. Potthast, F. Rangel, N. Rizwan, P. Rosso, F. Schneider, A. Smirnova, E. Stamatatos, B. Stein, M. Taulé, D. Ustalov, X. Wang, M. Wiegmann, S. M. Yimam, E. Zangerle, Overview of pan 2024: Multi-author writing style analysis, multilingual text detoxification, oppositional thinking analysis, and generative ai authorship verification - condensed lab overview, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association CLEF-2024, 2024.
[35] A. Giachanou, B. Ghanem, P. Rosso, Detection of conspiracy propagators using psycho-linguistic characteristics, Journal of Information Science 49 (2023) 3–17.