Unveiling Disinformation Narratives with AI

Collaborative Insights from Fact-Checkers and Computer Scientists' Work in Analyzing Climate Misinformation Narratives

Irene Larraz*

Universidad de Navarra, ilarraz@alumni.unav.es

Ramón Salaverría

Universidad de Navarra, rsalaver@unav.es

Javier Serrano-Puche

Universidad de Navarra, jserrano@unav.es

Fact-checkers are transitioning from debunking falsehoods to analyzing disinformation narratives, aiming to uncover common themes and underlying messages within these pieces of misinformation. This shift seeks to piece together the puzzle of the disinformation ecosystem, providing a comprehensive view to better understand how false ideas propagate and their ultimate objectives. Artificial intelligence (AI) plays a pivotal role, facilitating the analysis of a vast array of misinformation messages and synthesizing key insights. This paper explores the collaborative efforts between fact-checkers and computer scientists through a case study focusing on the analysis of climate misinformation narratives following farmers' protests in Europe within the Climate Facts Europe project, led by the EFCSN. The findings underscore the role of AI in assisting journalists to extract primary narratives and assess their impact over time.

CCS CONCEPTS • Fact-checking • Disinformation narratives • Artificial intelligence • Computer Sciences • Journalism • Climate disinformation

1 Introduction

Fact-checkers are leaping from debunking hoaxes to dismantling the whole narratives that support them, which transcend individual pieces to provide the big picture of disinformation. Understanding the narratives allows for addressing the root of the problem and better tackling the causes and interests behind disinformation. Additionally, it enables geographic and temporal comparisons between the messages circulating in each country, how they travel from one to another, and their evolution over time.

Behind this shift lie questions such as which narratives are being promoted by each actor, who is behind their dissemination, and how these ideas translate into specific pieces in each context. In this task, artificial intelligence has become a key tool. New large language models (LLMs) allow for the analysis of dozens of false claims already fact-checked by independent fact-checkers to cluster and infer the main ideas of these messages. Thus, automated analysis is a fundamental step to facilitate the work of journalists and better decipher the set of false messages that have been spread on a specific topic or at a particular time.

In this context, this article aims to identify how this joint work between fact-checkers and engineers takes place in a newsroom to understand the intersection between computer science and journalism. To this end, the article is based on a case study of the development of a narrative report[1] between the fact-checking organizations Newtral (Spain) and ScienceFeedback (France) as part of the Climate Facts Europe project,[2] led by the European Fact-Checking Standards Network (EFCSN), with the support of the European Climate Foundation. Three journalists from Newtral and Science-Feedback worked alongside two engineers from both

[1] https://climatefacts.efcsn.com/reports/3760d048-299e-4ee9-9b0b-bedb524584d8

[2] https://climatefacts.efcsn.com/

organizations to automatically extract the main narratives. The work started with the collection of a preliminary database of verifications related to agricultural protests from the Elections24Check repository,[3] which compiles and categorizes verified information for the 2024 European elections in collaboration with over 40 European fact-checking organizations and the EFCSN, with the support of the Google News Initiative. The team performed a manual analysis of disinformation narratives parallel to the automated analysis by the model to reveal the main narratives of climate disinformation that emerged from these protests.

This research examines the collaborative efforts between journalists and engineers from both teams, highlighting their joint strategies and assessing their potential impact on newsroom workflows and the ongoing battle against disinformation.

2 Related Work

The case study specifically focuses on narratives, understood as a set of messages, statements, or arguments constructed and disseminated to promote a distorted, misleading, or false view of facts to influence public perception and behavior. In the context of disinformation, these narratives can consist of multiple pieces of disinformation that reinforce a central theme or idea, designed to be persuasive and emotionally impactful [1, 2, 3, 4].

Numerous studies highlight the importance of this leap in disinformation analysis made by some fact-checkers to identify related content reinforcing particular narratives. Some of these works focus on detecting the main disinformation narratives as opposed to isolated content, to test the reach and impact of disinformation narratives as well as dissemination patterns [5]. Previous studies have also explored how organized disinformation campaigns exacerbate social polarization and distrust through strategic narratives [2]. Both content analysis and narrative analysis have proven effective in assessing the reach and potential impact of specific disinformation [6, 7].

However, few studies examine how artificial intelligence can assist fact-checkers in this process of analyzing disinformation narratives, and even fewer examine how the collaborative work between journalism and computing professionals is conducted to draw conclusions. Despite the exponential growth in research on models capable of detecting hoaxes and disinformation patterns to fact-check misinformation [8], there are not many studies that reflect the analysis of sets of fact-checks to identify the underlying common narratives. Some authors have attempted to characterize the use of artificial intelligence to detect similar narratives and offer proposals for better-automated tracking [9]. In another study, the author projects strategic narratives disseminated through disinformation campaigns using a natural language processing algorithm [10]. Following a process similar to that of the case study, the author details that this automation allowed for the extraction of information and the categorization of recurring themes in individual news pieces. To achieve this, the algorithm was trained to determine the frequency of mentions of recurring themes and their relational structure based on co-mentions.

3 Results

The initial step in the collaborative effort between journalists and engineers involved constructing a database as a foundation for their work. To this end, they developed a comprehensive search strategy to retrieve pertinent verifications from the Elections24Check repository. This strategy employed keywords, Boolean operators, and search filters to refine the search results effectively. Following the retrieval of this preliminary database, a data cleaning process was undertaken to eliminate verifications not directly related to agricultural protests. With this curated dataset, the analytical work commenced.

[3] https://elections24.efcsn.com/

3.1 **Manual and Automated Narrative Extraction**

The objective of the narrative analysis was to categorize the verifications based on key themes and messages to identify patterns, trends, and similarities among disinformation pieces, thereby mapping the disinformation ecosystem related to the topic. Two parallel processes were followed to analyze the articles. First, a manual review of the fact-checks was conducted to extract initial conclusions about the most common messages from the disinformation claims, which were then iteratively categorized into a refined list. Additionally, this manually annotated list of narratives was reviewed with experts from the European Climate Foundation to compare findings from the protests. Concurrently, the computing teams from both organizations performed an automated analysis of the verified claims using GPT to group them and measure their volume (Table 1)."

Table 1: Comparison Between Narratives Identified Manually and by the Algorithm

	Manual Narrative Analysis		Automated Narrative Analysis
1	Climate change measures are an excuse to control what you eat, how you make a living, etc., which are not really necessary.	1	Controversies surrounding geoengineering and climate change narratives in the European Union.
2	Farmers and ranchers are going to lose their livelihoods with these measures to protect the environment.	2	French farmers protest by pelting the Ukrainian embassy in Paris with manure.
3	Food products from countries outside of the EU are contaminated.	3	Spain faces water scarcity and environmental controversies due to the destruction of water infrastructures.
4	Institutions (WHO, EU) are banning to cultivate food at home.	4	Lab-grown meat from animal cells sold in Germany.
5	Measures are being taken to reduce climate change, but it actually increases drought.	5	Police charges and arrests at farmers' protests in Don Benito and Logroño.
6	The government is controlling the weather.	6	Controversial statements suggest replacing farmers with artificial intelligence robots.
7	The farmers' protests are very aggressive.	7	Lab-grown meat company Redefine Meat supplies German restaurants with 3D-printed fillets.
8	Police brutality against farmers.	8	Moroccan strawberries under scrutiny for contamination and germs.
9	Politicians such as Macron are ignoring the protests.	9	French President Macron faces backlash from farmers as protests escalate.
10	Farmers are protesting against the Green Deal.	10	EU parliament passes laws to protect land and sea areas, promote nature restoration, and mandate property renovations.
11	Russian propaganda: false data on how agricultural imports from Ukraine are affecting us.	11	European Union plans to ban cultivation of subsistence fruit and vegetables in private gardens as part of its Green Deal strategy.

[a]Source: own elaboration.

Once both automated lists were obtained, a comparative analysis was conducted with the results of the manually extracted narratives to refine the categories and specify the groups more precisely. As shown in Table 1, there are categories that overlap, such as those in positions 4 and 7 in the automated analysis.

3.2 **Development of an Integrated Proposal**

The final phase of the process involved creating a definitive list of narratives based on the unified categories (Table 2). Using this list, a final manual review of the database was conducted to ensure it encompassed the majority of the fact-checked claims. Subsequently, the fact-checks were annotated according to these

categories. Additionally, a comparison table was created to assess the accuracy of the automation by comparing the categories generated by the algorithm with the definitive ones, identifying potential improvements.

Table 2: Integrated Proposal of Narratives

	Narrative analysis manually done
1	Climate change measures are used as a pretext to exert control over individuals' dietary choices and livelihoods.
2	Environmental protection initiatives threaten the livelihoods of farmers.
3	Allegations suggest the government is secretly manipulating weather patterns.
4	Farmers are protesting against the Green Deal
5	Unfair competition promoted by the EU is detrimental to farmers, exacerbating their challenges.
6	The farmers' protests are aggressive
7	Law enforcement's response to protesters is deemed excessively aggressive.
8	Support for Ukraine is resulting in the wasteful disposal of our food, adversely impacting local farmers.

[a]Source: own elaboration.

The categories of false texts do not necessarily indicate that what they say is false in general, but rather that there have been specific hoaxes claiming so that turned out to be false. For instance, the analysis does not assert the absence of farmers protesting against the Green Deal, or that there have been no aggressive protests or police repression. What these statements indicate is that among the claims verified by fact-checking organizations, there were some falsehoods reinforcing these ideas. It's essential to note that fact-checks pertain to specific cases, while narrative analysis aims to distill the overarching messages encapsulated within them.

4 Discussion

The results from the collaborative efforts of the journalistic and computational teams underscore the significance of automated narrative analysis as an initial step. Journalists particularly emphasized the speed and scalability of obtaining results, given the algorithm's capacity to handle large volumes of data. Moreover, they highlighted the specificity of certain model responses, noting the opportunity for cross-referencing and complementing their findings. In line with previous studies utilizing AI for narrative detection [10], automated narrative analysis facilitated the matching of insights and identification of trends and outliers within the dataset. Nonetheless, limitations were also identified:

- Abstraction Capability: Automated analysis can group false claims but is not able to abstract underlying ideas or draw conclusions from those messages. For instance, the algorithm detected the recurrence of fact-checks regarding "Moroccan-origin strawberries being scrutinized for contamination and germs." However, manual analysis extracted the idea that "food products from countries outside the EU are contaminated."
- Specificity of Narratives: Some of the narratives identified by the algorithm were too specific to a location or actor, failing to establish common features with similar disinformation from other contexts, essential for this type of analysis. For example, the algorithm yielded the category "Police charges and arrests at farmers' protests in Don Benito and Logroño" because there were several verifications on this topic. However, it did not establish a broad category about violence or repression of protests, which could have included other verified claims such as "The army was brought out to intimidate the farmers and truckers protesting in Baia Mare" or "A farmer subdued by the police when trying to enter Madrid with his tractor."
- Orphaned Claims: As a result of the above, automated detection left 50 claims orphaned under the classification of "Protests across Europe highlight farmers' grievances against EU agricultural

policies" and "Protests by European farmers against agricultural policies and government leaders in various countries." In the case of manual review, many of these claims fit into the final categories, while others were excluded because they referred to the representation of protests themselves, for example, using videos from past demonstrations or with images generated with artificial intelligence.

5 Conclusions

Artificial intelligence can play a relevant role in the automation of narrative analysis to better understand and connect the dots in disinformation patterns and get the big picture. By integrating algorithmic models into the analysis process, they can facilitate journalists' work clustering and synthesizing disinformation claims with speed and scalability opportunities. This is especially relevant with large volumes of data in a short space of time. Comparing automated results has proven effective in complementing manual analyses and undercover trends. This way, collaboration between journalists and computer scientists allowed for the synthesis of diverse insights, enhancing the depth and accuracy of narrative analysis.

However, AI-driven analyses have also shown limitations in abstracting underlying ideas and identifying broader patterns. In some instances, the provided categories lacked contextual understanding, resulting in overly specific classifications and orphaned claims. Nevertheless, workflows involving the interaction between journalists and computer scientists can be improved to better detect broader narratives and capture a greater number of claims effectively.

The case study highlights the evolving landscape of collaborative research on how AI can improve fact-checkers' work and how they tackle disinformation. Future research could explore enhanced processes to consolidate a hybrid methodology in the narrative analysis. Overall, the research underscores the importance of interdisciplinary approaches in combating disinformation.

REFERENCES

[1] Nodes. (2023). Are you resisting the green dystopia or preparing for the climate apocalypse? Polarising narratives are unsettling the climate change debate. Nodes. Retrieved from https://nodes.eu/are-you-resisting-the-green-dystopia-or-preparing-for-the-climate-apocalypse-polarising-narratives-are-unsettling-the-climate-change-debate

[2] Benkler, Yochai; Faris, Robert; Roberts, Hal. 2018. "Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics." Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001

[3] VoxCheck Team. (2023). "Same Narratives, Shifting Fakes: VoxCheck's Investigation into Russian Falsehoods in Europe." Vox Ukraine. Retrieved from https://voxukraine.org/en/same-narratives-shifting-fakes-voxchecks-investigation-into-russian-falsehoods-in-europe

[4] Tucker, Joshua A.; Guess, Andrew; Barbera, Pablo; Vaccari, Cristian; Siegel, Alexandra; Sanovich, Sergey; Stukal, Denis; Nyhan, Brendan. 2018. "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature." Political Science & Politics. https://doi.org/10.2139/ssrn.3144139

[5] Suau, J., & Puertas-Graell, D. (2023). Narrativas de desinformación en España: alcance, impacto y patrones de difusión. Profesional de la Informacion, 32(5), Article e320508. https://doi.org/10.3145/epi.2023.sep.08

[6] Strand, C., & Svensson, J. (2022). Foreign Norm Entrepreneurs' Misinformation and Disinformation Narratives on LGBT+ Rights in Europe. Medijska Istraživanja, 28(2), 109–132. https://doi.org/10.22572/mi.28.2.5

[7] Herman, L., and Vervaeck, B. (2019). Handbook of Narrative Analysis. Second edition. Lincoln: University of Nebraska Press. https://doi.org/10.2307/j.ctvr43mhw

[8] Al-Asadi, M.A., Tasdemir, S. (2022). Using Artificial Intelligence Against the Phenomenon of Fake News: A Systematic Literature Review. In: Lahby, M., Pathan, AS.K., Maleh, Y., Yafooz, W.M.S. (eds) Combating Fake News with Computational Intelligence Techniques. Studies in Computational Intelligence, vol 1001. Springer, Cham. https://doi.org/10.1007/978-3-030-90087-8_2

[9] Skumanich, A., & Kim, H. K. (2024). Modes of Analyzing Disinformation Narratives With AI/ML/Text Mining to Assist in Mitigating the Weaponization of Social Media. *arXiv:2405.15987 [cs.CY]*. https://doi.org/10.48550/arXiv.2405.15987

[10] Bánkuty-Balogh, L. S. (2021). Novel technologies and Geopolitical Strategies: Disinformation Narratives in the Countries of the

Visegrád Group. Politics in Central Europe, 17(2). https://doi.org/10.2478/pce-2021-0008