International Journal of Communication 19(2025), 3647–3666 1932–8036/20250005

Unveiling Disinformation Narratives With AI: Collaborative Insights from Fact-Checkers and Computer Scientists' Work in Analyzing Climate Misinformation Narratives

IRENE LARRAZ
RAMÓN SALAVERRÍA
JAVIER SERRANO-PUCHE
Universidad de Navarra, Spain

Fact-checkers are shifting from debunking isolated falsehoods to analyzing broader 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 study 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 European Fact-Checking Standards Network (EFCSN). The findings underscore AI's role in helping journalists extract primary narratives and assess their impact over time.

Keywords: fact-checking, disinformation narratives, artificial intelligence, computer sciences, journalism, climate disinformation

Fact-checkers are leaping from debunking isolated hoaxes to dismantling entire narratives that sustain them, which transcend individual pieces to provide the big picture of disinformation. Understanding the narratives allows them to address the root of the problem and better tackle the causes and interests behind disinformation. In addition, 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. Although fact-checkers have traditionally focused on individual claims, artificial intelligence (AI) has become a key player capable of abstracting narrative patterns from vast amounts of disinformation data. New large language

Irene Larraz: ilarraz@alumni.unav.es
Ramón Salaverría: rsalaver@unav.es
Javier Serrano-Puche: jserrano@unav.es
Date submitted: 2025-02-20

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 content that has been spread on a specific topic or at a particular time.

Despite these advancements, research has largely focused on the detection and verification of isolated claims, with limited exploration of scalable AI-driven tools for narrative-level analysis. In addition, there remains a gap in interdisciplinary frameworks that integrate the expertise of journalists and engineers to systematically track and interpret disinformation narratives, which is vital for newsrooms to detect, analyze, and counter disinformation effectively.

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 (Newtral & Science Feedback, 2024) between the fact-checking organizations Newtral (Spain) and ScienceFeedback (France) as part of the Climate Facts Europe project,[1] 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 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,[2] 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. It also contributes to filling the identified gaps by demonstrating the application of scalable AI tools for narrative abstraction and exploring how interdisciplinary approaches can enhance the battle against disinformation.

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 (Benkler, Faris, & Roberts, 2018; Nodes, 2023; Tucker et al., 2018; VoxCheck Team, 2023).

[1] https://web.archive.org/web/20250322064311/https://climatefacts.efcsn.com

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

These narratives are part of broader strategic efforts to reinforce specific ideas or messages through disinformation. According to Fisher's (1984) Narrative Paradigm, humans interpret the world through storytelling grounded in narrative coherence (logical consistency) and narrative fidelity (perceived truthfulness). In the context of disinformation, narratives leverage these storytelling principles by disseminating multiple, interrelated pieces of misleading information that collectively construct a distorted understanding of reality. These narratives rely on cognitive schemas such as confirmation bias, where individuals favor information that aligns with their existing beliefs, or the illusory truth effect, where repeated exposure to information increases its perceived truthfulness. In addition, strategic narrative theory expands on how such narratives are deliberately crafted and circulated to serve political or ideological purposes, including how the polarization of public opinion or the erosion of trust in institutions (Bennett & Edelman, 1985) provides a framework to understand how disinformation is constructed and disseminated with specific strategic goals, such as polarizing public opinion or discrediting institutions. These theoretical perspectives underscore the importance of analyzing disinformation pieces not in isolation, but through a broader and interconnected perspective to highlight disinformation campaigns.

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 (Suau & Puertas-Graell, 2023). Previous studies have also explored how organized disinformation campaigns exacerbate socyial polarization and distrust through strategic narratives (Benkler et al., 2018). Both content analysis and narrative analysis have proven effective in assessing the reach and potential impact of specific disinformation (Herman & Vervaeck, 2019; Strand & Svensson, 2022).

AI can assist fact-checkers in this process of analyzing disinformation narratives and drawing conclusions (Piper, So, & Bamman, 2021; Santos, 2023). Computational methods such as data classification, clustering analysis, and decision tree algorithms have proven effective for categorizing and grouping content based on the underlying messages they support (Akhtar et al., 2023; Tianda et al., 2024). These methods rely on text representation techniques, with word embeddings being among the most used to capture the frequency of terms and their semantic relationships (Al-Tarawneh, Al-irr, Al-Maaitah, Kanj, & Aly, 2024). Recent models, such as BERT, RoBERTa, and Sentence Transformers, have demonstrated high performance in understanding context and uncovering common structures in disinformation narratives, leveraging deep learning to map complex linguistic patterns (Angelov & Inkpen, 2024; Levi, Mor, Sheafer, & Shenhav, 2022).

When scaling up to narrative-level analysis, clustering and topic-modeling techniques are often employed to identify latent themes in disinformation content. Traditional approaches like K-Means and DBSCAN are well-regarded for their simplicity and interpretability, whereas BERTopic has gained attention for its ability to extract semantically coherent topics from complex datasets (Koterwa & Świtała, 2025). Complementary to these techniques, models such as FrameNet and Narrative Structure Analysis help map thematic roles and rhetorical strategies in disinformation, providing deeper insights into the narrative construction. In a further step, Stance Detection Models assess the text's orientation toward specific narratives, offering a structural understanding of propaganda techniques (Puczyńska & Djenouri, 2024).

Generative AI and LLMs have added a new layer for narrative analysis, allowing researchers to generate summaries of misinformation clusters, capturing the essence of the narratives detected.

Several researchers have explored the application of AI for detecting common narratives and have proposed improvements for automated tracking (Al-Asadi & Tasdemir, 2022). For instance, Skumanich and Kim (2024) examined the detection of thematic patterns for enhanced monitoring of campaigns, whereas Padalko (2025) focused on deriving insights useful for policy assessment. In another study, Bánkuty-Balogh (2021) mapped strategic narratives disseminated through disinformation campaigns using a natural language processing (NLP) algorithm. This method enabled the automated extraction of information and categorization of recurring themes in individual news pieces. The algorithm was trained to identify the frequency of mentions and their relational structures based on cooccurrences, highlighting how narratives are framed and distributed. Further research has extended these methodologies to extract narratives not only from text but also from propagation patterns and visual features. For example, Deepthi and Shastry (2024) used machine learning to incorporate visual analysis into misinformation detection, identifying recurring images and symbols that support textual disinformation.

Despite significant advances in leveraging LLMs and other AI techniques to detect narrative patterns across disinformation messages, there remains a gap in the literature concerning the analysis of fact-check data sets to uncover underlying common narratives. Addressing this gap would contribute to understanding not only isolated claims but the overarching strategies employed in disinformation campaigns.

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 data set, the analytical work commenced.

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. In addition, this manually annotated list of narratives was reviewed with experts from the European Climate Foundation to compare findings from the protests.

The original database collected relevant fact-checks using keywords and search filters, resulting in 978 preliminary entries. Following a manual review, 130 articles were identified as meeting the criteria for

fact-checks focused on the farmer protests. This repository was curated by 40 fact-checking organizations across Europe, based on a standardized inclusion criterion defined by the project coordinators. This collective approach minimized individual researcher bias during claim selection. The narrative selection process emerged organically from the analysis of these claims, guided by thematic patterns rather than subjective interpretation. Engineers contributed their technical expertise to structure the data analysis, while journalists applied their contextual understanding to identify recurring themes. Efforts were made to actively reflect on disciplinary perspectives, recognizing that journalistic experience might prioritize communicative clarity, while engineering expertise emphasized data structure and scalability. Regular cross-disciplinary discussions were held to align technical outputs with narrative interpretation to ensure balanced contributions, reducing potential biases and reinforcing methodological consistency.

The manual analysis followed a coding method that combined deductive and inductive approaches. Fact-checkers first reviewed the data set to extract initial insights on common types of misinformation using thematic analysis, which allowed for identifying predefined categories and emerging new themes during the process. The categorization criteria were established based on the content of the claims, focusing on similar messages and recurring underlying ideas. Categories were iteratively refined to enhance coherence and representation; if a category contained only one or two claims, it was merged with others following a clear set of guidelines to maintain thematic consistency. Discrepancies among the journalists annotating it were resolved through discussion.

For example, claims grouped under a common theme were like the following: "A factory in Holland prints 500 tons of steaks per month using 3D printers" (Zwin, 2024); "Around 110 German restaurants already buy 'meat' from Redefine Meat" (Zwin, 2024); "Company Redefine Meat sells 3D-printed meat cultivated from animal stem cells in Germany" (Mortkowitz, 2024); "A Dutch company supplies German restaurants with steaks made in the laboratory from animal cells" (Zwin, 2024); "Cultivated meat is produced and sold in the European Union" (Kirkova, 2024); and "The new agricultural rules and the opening of markets to Ukraine have put European agriculture in an impossible situation. Instead of healthy, locally produced food, we are being fed artificial meat and GMO junk" (Totth, 2024). The initial category was labeled as "Artificial meat is being turned to in order to mitigate the significant pollution caused by livestock." After iterative refinement, it was summarized as "The European Union is promoting lab-grown meat." From this categorization, the underlying narrative was extracted: "Climate change measures are an excuse to control what you eat, how you make a living, and more—controls that are not truly necessary."

Another example was the narrative "Unfair competition promoted by the European Union is detrimental to farmers, exacerbating their challenges." To reach this narrative, journalists summarize several claims as "Food products from countries outside of the European Union are contaminated or do not meet the standards." This included false claims such as "Moroccan strawberries contaminated with norovirus were marketed in Spain" (Domínguez, 2024); "Beans from Morocco with traces of some treatment visible under ultraviolet light" (Maldita.es, 2024); "Birds die because of Ukrainian wheat and we will go after them, in Poland they are promoting a video of dozens of birds dying because of Ukrainian wheat" (Facta, 2024); and "Agricultural products imported from third countries outside the European Union do not pass phytosanitary controls" (Ocaña, 2024).

Concurrently, the computing teams from both organizations conducted an automated analysis of the verified claims using GPT-4o-mini to group them and measure their volume. To do so, they followed three main stages: embedding generation and dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP); clustering using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN); and narrative extraction using an LLM, in this case, GPT-4o-mini.

Embedding Transformation

The initial step consisted of transforming textual data into numerical representations known as embeddings. This process allows for semantic information within the text to be captured in a structured, machine-readable format. Specifically, each claim was converted into an embedding vector, representing its linguistic and contextual attributes.

Given that embeddings are often high-dimensional, in our case of 1024 dimensions, a dimensionality reduction step was applied to streamline the complexity of the data while preserving essential features. This transformation enables more efficient clustering and visualization, mitigating the risk of overfitting and enhancing interpretability. This was done using UMAP, which is a dimension reduction technique.

Clustering

The reduced embeddings were then subjected to a clustering algorithm to group similar claims. Although K-Means is a standard choice for this task, this study opted for HDBSCAN because of its ability to identify clusters of varying densities and manage noise more effectively. Clustering allowed for the detection of narrative patterns within the data set, categorizing claims based on semantic similarity.

To complete it, a decision tree algorithm was employed to better classify them into categories. This enabled identifying branching paths of thematic similarity and grouping them based on the narrative strategies used by evaluating specific features of disinformation content. Table 1 presents the hierarchical branching paths of disinformation narratives identified through the decision tree algorithm.

Table 1. Classification of Narratives Using a Decision Tree Algorithm

Level 0—Main Theme	Level 1—Subcategory	Level 2—Subcategory	Level 3—Subcategory	Level 4—Subcategory	Level 5—Subcategory
French farmers carry out protests involving specific demands	Protests by European farmers against agricultural policies	European farmers protest agricultural regulations	European agriculture faces challenges related to market access	Impact of agricultural and environmental policies on production	
Controversies and misconceptions surrounding climate policies	European Union environmental policies and controversies	Protests by European farmers against agricultural policies	European farmers protest agricultural regulations	European agriculture faces challenges related to market access	Impact of agricultural and environmental policies on production
Contamination concerns rise because of Moroccan products	European agriculture faces challenges related to trade	Impact of agricultural and environmental policies on production			
Lab-grown 3D-printed meat in Germany	Redefine Meat introduces 3D-printed meat from animal cells	Impact of agricultural and environmental policies on production			
Challenges of European Union membership impacting industrial agriculture	European Union environmental policies and controversies	Protests by European farmers against agricultural policies	European farmers protest agricultural regulations	European agriculture faces challenges related to market access	Impact of agricultural and environmental policies on production

Narrative Extraction with GPT

Once clusters were formed, a narrative extraction process was initiated using GPT, a state-of-the-art LLM. For each cluster, a subset of representative claims, typically three to five, depending on the cluster size, was selected and processed through GPT to generate a coherent narrative that encapsulates the underlying theme of the grouped claims. This step enabled the abstraction of high-level narratives from the data, moving beyond surface-level textual similarities. The decision to employ GPT instead of a BERT model was primarily driven by considerations of time efficiency and data volume. Fine-tuning a BERT model for classification tasks traditionally requires substantial amounts of annotated data, which, in turn, demands considerable time and resources. Moreover, LLMs like GPT have demonstrated superior accuracy in classification tasks across various contexts, further justifying their selection for this purpose.

The prompt used at Newtral is a few-shot example, which provides samples of how the input and output of the request should appear. This approach helped to generate narratives that align more closely with the original concept (see Figure 1).

```
You are given a list of text collections on various topics. Summarize each collection with a
short, descriptive title.
Don't use words like various, multiple, diverse, or controversial... be specific on the
description, it must be a
narrative not a list of things. Avoid including any extra details beyond the topic name. The
texts inside each
collection talk about similar themes, your task is to extract the common narrative they are
talking about in general
terms without losing detail.

Here is an example of how the input/output should look like:
``` Sample texts from this topic:
- La Comisión Europea ha impuesto una multa de 1.800
millones de euros a Apple por prácticas competitivas desleales relacionadas con su servicio
de música en streaming. La
multa se produce después de una denuncia presentada por Spotify.
- The European Union has imposed a record-breaking
€1.8 billion antitrust fine on Apple in a dispute with Spotify.
- Apple was accused of favoring its own music streaming service, Apple Music,
over competitors like Spotify in the App Store. This decision highlights the EU's efforts to
ensure fair competition in
 the digital market and prevent dominant companies from abusing their power.

Topic name: The European Union imposes a record-breaking €1.8 billion antitrust fine on
Apple for favoring its music
streaming service over competitors like Spotify. ```

Sample texts from this topic:
[DOCUMENTS]
Topic name:
```

Figure 1. Prompt used in the automated phase.

The team's problem was that there were too general clusters around the protests since they heavily depended on the parameters used throughout the process. Since there are many steps in the clustering process, tweaking some parameters significantly affects the final result. They created specific clusters for smaller narratives. From there, they iterated to group them further to capture clusters, even if there were just two or three claims that corresponded to them.

Finally, an iterative optimization and validation process was conducted to enhance performance. The clustering process was refined through successive adjustments of model parameters to improve both accuracy and granularity. The clusters were compared with the manual annotations to assess alignment and identify discrepancies.
```

For instance, in the previous example, both the automated analysis and the manual review grouped the claims under the same category. However, the summary from the automated extraction was "Lab-grown meat from animal cells sold in Germany." In a second iteration, the automated analysis generated the summary "Lab-grown meat company Redefine Meat supplies German restaurants with 3D-printed fillets," which still omitted certain nuances and failed to capture the underlying narrative.

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 2, some categories overlap, such as those in positions 4 and 7 in the automated analysis.

Table 2. 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 protested by pelting the Ukrainian embassy in Paris with manure.
3	Food products from countries outside the European Union are contaminated.	3	Spain faces water scarcity and environmental controversies because of the destruction of water infrastructure.
4	Institutions (World Health Organization, European Union) are banning the cultivation of food at home.	4	Lab-grown meat from animal cells is 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 AI 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 are 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	The European 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	The European Union plans to ban the cultivation of subsistence fruit and vegetables in private gardens as part of its Green Deal strategy.

Development of an Integrated Proposal

The final phase of the process involved creating a definitive list of narratives based on the unified categories (Table 3). Using this list, a final manual review of the database was conducted to ensure it encompassed most fact-checked claims. Subsequently, the fact-checks were annotated according to these

categories. In addition, 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 3. Integrated Proposal of Narratives

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

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 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, some falsehoods were reinforcing these ideas. It is essential to note that fact-checks pertain to specific cases, whereas narrative analysis aims to distill the overarching messages encapsulated within them.

Validation and Editorial Assessment

The validation of the AI-assisted methodology was conducted through an editorial assessment between journalistic teams from both organizations to ensure the consistency and accuracy of the identified narratives. Journalists and domain experts reviewed the AI-generated narratives alongside the human-extracted narratives, contrasting them with the original claims from the fact-checks. The comparative analysis allowed the team to evaluate the consistency across both approaches and to ensure that they cover the main narratives without leaving any nuances or subtle messages.

Narratives deemed too broad or too narrow were either refined or discarded. For instance, although the AI provided a cluster titled "Police charges and arrests at farmers' protests in Don Benito and Logroño," human reviewers recognized the need for a more generalized category encompassing similar events across different regions. This led to the formulation of a broader narrative as "Violence or repression of protests," which encapsulated various related incidents.

From the computational perspective, the evaluation focused on assessing the alignment between AI-generated clusters and manually annotated categories. The primary metric involved calculating the

proportion of claims within each AI-generated cluster that matched the corresponding human-annotated category. Clusters where over 70% of the claims corresponded to a single manual category were considered well-aligned, indicating effective clustering. Conversely, clusters with lower alignment percentages were flagged for further review, as they potentially represented noise or misclassification.

Discussion

The results from 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 using AI for narrative detection (Bánkuty-Balogh, 2021), automated narrative analysis facilitated the matching of insights and identification of trends and outliers within the data set.

When comparing discrepancies between manual and automated analyzes, the evaluation showed that 21 out of 25 narratives were accurately generated, with approximately 60% of the records correctly clustered within the same grouping. This corresponds to an accuracy rate of 84% in narrative generation, reflecting the LLM's effectiveness in producing coherent and relevant narratives. Furthermore, the clustering percentage serves as an indicator of the precision achieved by the HDBSCAN algorithm.

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 example, while the algorithm identified recurring fact-checks related to "Moroccan-origin strawberries being scrutinized for contamination and germs," it did not infer the subtle narrative that "food products from countries outside the European Union are contaminated." This limitation reflects broader challenges in natural language processing, where models often fail to capture implied meanings and underlying framings beyond surface-level text. Addressing this gap would require more advanced architectures capable of not just summarizing the common textual elements but also capturing the intended messages behind disinformation pieces.

However, as noted by the engineering team, there is a deliberate focus on literal interpretation during the automated process. While increasing abstraction could enhance narrative detection, it also risks introducing noise or misrepresentation by deviating from the precise language used in the original claims. Research has shown that traditional topic modeling methods, such as Latent Dirichlet Allocation (LDA) and even advanced models like BerTopic, frequently overfit to repetitive language patterns while overlooking deeper contextual signals (Grootendorst, 2022).

- For the current analysis, literal consistency was prioritized to ensure alignment with fact-checkers' findings, even if it sacrificed some abstraction capability. Future studies could explore the development of agents designed for higher-level abstraction, capable of capturing overarching

themes without compromising factual precision. Until such advancements are achieved, maintaining a human-in-the-loop approach remains essential for identifying the broader implications of grouped claims.

- 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 that are 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: The automated analysis employed a decision-tree model to group claims at varying levels of granularity. This hierarchical structure enabled the categorization of claims based on broad themes at the top levels, with more specific subcategories emerging in deeper layers. However, 50 claims remained at the first level, identified as "orphaned claims" because they did not fit into any of the more refined subcategories. This outcome was not because of model error but rather a reflection of their thematic uniqueness or ambiguity, which made them difficult to associate with specific narratives, a challenge also noted by other studies (Starbird, 2019). These claims often lacked clear categorization during the manual analysis. Although many fit into the final categories, others were excluded because they referred to the representation of protests themselves, for example, using videos from past demonstrations or AI-generated images. Although such themes were observable, they were not relevant to the study's focus on climate change narratives.

These limitations might highlight automation bias in individuals to overrely on automated systems. This could help explain why certain narratives were misclassified or why some claims remained ungrouped. Moreover, the automated decision tree model's rigid structure may have contributed to these gaps, as it lacks the flexibility to adapt to ambiguous or contextually layered disinformation claims. To address these limitations, future work could explore hierarchical clustering or dynamic threshold adjustments. Unlike static decision trees, they adaptively merge or split clusters based on evolving narrative cues, potentially capturing multilayered disinformation strategies (Tianda et al., 2024).

Finally, the findings suggest that although automated clustering is a promising step toward scalable disinformation monitoring, it is not yet capable of fully abstracting the strategic ambiguity and geopolitical subtleties present in certain claims. As Saxena, Moon, Chaurasia, Guan, and Guha (2023) concluded, the results emphasize the importance of hybrid approaches, combining computational methods with human expertise in analyzing complex narratives to achieve a more effective extraction.

Ethical Considerations

Narrative analysis introduces an additional layer of complexity to the traditional fact-checking process. By shifting the focus from individual claims to a broader message, it often requires a degree of

abstraction and interpretation to identify the underlying intent of the disinformation piece. Although this abstraction enables the identification of sophisticated narrative patterns, it also raises important ethical considerations. Unlike the fact-checking of explicit claims, narrative detection may involve interpreting the implied meanings or intended effects of disinformation. This interpretative step, although grounded in contextual analysis, lacks a standardized methodology and risks introducing subjective bias if not rigorously monitored.

These ethical considerations are notably underexplored in AI benchmarks and academic literature. As automated narrative analysis becomes more integrated into journalistic and fact-checking practices, establishing ethical guidelines for abstraction, interpretation, and data handling becomes essential. This includes ensuring that human-in-the-loop processes validate algorithmic outputs and that interpretative steps are consistently checked against transparent criteria to minimize bias and uphold factual integrity.

Conclusion

AI 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 by 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 analyzes and uncovering 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. Future improvements should consider the implementation of more consistent workflows involving iterative human-AI refinement loops, where journalists and analysts periodically validate and adjust model outputs to ensure contextual accuracy and narrative cohesion. This would help mitigate errors and increase alignment with human understanding of complex disinformation narratives. In addition, the integration of adversarial testing, a method for evaluating models' behavior when provided with harmful inputs, could further enhance the models' robustness by exposing current failures that are evident to humans, but not to machines.

To integrate AI-driven narrative analysis into their existing fact-checking workflows, newsrooms can establish a new workflow to enable journalists to rapidly cluster disinformation claims, identify emerging narratives, and reinforce fact-checking efforts based on repeated disinformation. The creation of repositories of fact-checks published by organizations, such as those developed by the European Digital Media Observatory (EDMO, n.d.), a network of regional hubs bringing together fact-checkers and academics, alongside open-source annotated data sets curated with input from fact-checkers, would significantly enhance model training and improve contextual accuracy when addressing multilingual and cross-regional disinformation. Collaborative initiatives, such as crowdsourced verification platforms and journalist-AI co-monitoring systems, could further democratize access to these technologies. Finally, building interfaces that

facilitate real-time human-AI interaction would empower journalists to validate insights from machine-generated clusters.

There are two main points that deserve special attention:

- **Understanding Targeted Disinformation Through Narrative Analysis.** Narrative analysis serves as a foundational approach to unravel how disinformation campaigns strategically disseminate distinct pieces of misleading information to diverse audiences while maintaining a cohesive subliminal message. This method allows researchers to trace the fragmented yet interconnected nature of disinformation, highlighting how bad actors leverage platform algorithms to microtarget specific demographic or ideological groups.

Investigating the mechanisms by which algorithms amplify tailored disinformation narratives is crucial for dismantling coordinated influence operations and understanding the propagation of falsehoods. Future research might focus on the intersection of targeted disinformation and narrative construction to expose the pathways through which disinformation permeates different social contexts.

- **Policy Implications.** These findings carry significant implications for policy development, particularly concerning the regulatory oversight of very large online platforms (VLOPs) under the Digital Services Act (DSA). The ability to map narrative dissemination patterns provides a data-driven foundation for reinforcing VLOPs' obligations to counter disinformation through algorithmic transparency and enhanced content moderation practices. Furthermore, policy can benefit from narrative analysis by complementing established frameworks such as the ABC model (focused on actors, behaviors, and content), offering insights into how disinformation is operationalized across diverse audiences. It also supports strategies to dismantle foreign information manipulation and interference (FIMI) campaigns, offering a granular understanding of how false narratives evolve and spread across countries and contexts. This analytical depth is instrumental for policymakers aiming to design effective interventions that address both the spread mechanisms and the algorithmic enablers of disinformation.

Overall, 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

Akhtar, P., Ghouri, A. M., Khan, H. U. R., Haq, M. A. U., Awan, U., Zahoor, N., . . . & Ashraf, A. (2023). Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions. *Annals of Operations Research, 327*, 633–657. doi:10.1007/s10479-022-05015-5

Al-Asadi, M. A., & Tasdemir, S. (2022). Using artificial intelligence against the phenomenon of fake news: A systematic literature review. In M. Lahby, A. S. K. Pathan, Y. Maleh, & W. M. S. Yafooz (Eds.), *Combating fake news with computational intelligence techniques* (pp. 39–54). Cham, Switzerland: Springer. doi:10.1007/978-3-030-90087-8_2

Al-Tarawneh, M. A. B., Al-irr, O., Al-Maaitah, K. S., Kanj, H., & Aly, W. H. F. (2024). Enhancing fake news detection with word embedding: A machine learning and deep learning approach. *Computers, 13*(9), 239. doi:10.3390/computers13090239

Angelov, D., & Inkpen, D. (2024). Topic modeling: Contextual token embeddings are all you need. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Findings of the Association for Computational Linguistics*: *EMNLP 2024* (pp. 13528–13539). Miami, FL: Association for Computational Linguistics. doi:10.18653/v1/2024.findings-emnlp.790

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), 247–264. doi:10.2478/pce-2021-0008

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics.* New York, NY: Oxford University Press. doi:10.1093/oso/9780190923624.001.0001

Bennett, W. L., & Edelman, M. (1985). Toward a new political narrative. *Journal of Communication, 35*(4), 156–171. doi:10.1111/j.1460-2466.1985.tb02979.x

Deepthi, K., & Shastry, A. K. (2024). A deep learning methodology-based unsupervised misinformation detection (UMD) technique for identifying fabricated narratives. In *Proceedings of the Second International Conference on Networks, Multimedia and Information Technology (NMITCON 2024)* (pp. 1–6). Pune, India: IEEE. doi:10.1109/NMITCON62075.2024.10699051

Domínguez, G. (2024, February 23). *Las fresas marroquíes con norovirus no se comercializaron en España* [Moroccan strawberries with norovirus were not sold in Spain]. Verifica EFE. Retrieved from https://verifica.efe.com/las-fresas-marroquies-con-norovirus-no-se-comercializaron-en-espana/

European Digital Media Observatory. (n.d.). *About us—EDMOeu*. Retrieved from https://edmo.eu/about-us/edmoeu/

Facta. (2024, March 12). *Non ci sono prove che questi uccelli in Polonia siano morti per aver mangiato grano ucraino* [There is no evidence that these birds in Poland died from eating Ukrainian grain]. Facta. Retrieved from https://facta.news/antibufale/2024/03/12/uccelli-morti-polonia-grano-ucraina/

Fisher, W. R. (1984). Narration as a human communication paradigm: The case of public moral argument. *Journal of Communication, 51*(1), 1–22. doi:10.1080/03637758409390180

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. arXiv preprint arXiv:2203.05794. Retrieved from https://arxiv.org/abs/2203.05794

Herman, L., & Vervaeck, B. (2019). *Handbook of narrative analysis* (2nd ed.). Lincoln: University of Nebraska Press. doi:10.2307/j.ctvr43mhw

Kirkova, M. (2024, March 12). *Подвеждаща публикация твърди, че в Европейския съюз се произвежда и продава култивирано месо* [Misleading publication claims that cultivated meat is produced and sold in the European Union]. Factcheck.bg. Retrieved from https://factcheck.bg/podvezhdashta-publikaciya-tvardi-che-v-evropejskiya-sajuz-se-proizvezhda-i-prodava-kultivirano-meso/

Koterwa, D., & Świtała, M. (2025). *Enhancing BERTopic with intermediate layer representations*. Arxiv. doi:10.48550/arXiv.2505.06696

Levi, E., Mor, G., Sheafer, T., & Shenhav, S. R. (2022). Detecting narrative elements in informational text. *Findings of the Association for Computational Linguistics: NAACL 2022*, 1755–1765. doi:10.18653/v1/2022.findings-naacl.133

Maldita.es. (2024, March 6). *No, estas judías verdes de Marruecos no están "contaminadas" aunque tengan manchas visibles bajo luz ultravioleta* [No, these green beans from Morocco are not "contaminated" even if they show marks under ultraviolet light]. Maldita.es. Retrieved from https://maldita.es/alimentacion/20240306/judias-marruecos-manchas-luz-ultravioleta/

Mortkowitz, L. (2024, February 22). *Maso vypěstované v laboratoři z kmenových buněk zvířat není v současné době v EU povoleno* [Meat grown in a lab from animal stem cells is currently not permitted in the EU]. AFP Austria. Retrieved from https://napravoumiru.afp.com/doc.afp.com.34JV6WM

Newtral & Science Feedback. (2024). *Fertile ground for disinformation: From spreading climate change misinformation to undermining climate action: How the farmers' protests were used to influence audiences* [Report]. European Fact-Checking Standards Network. Retrieved August 1, 2025, from https://efcsn.s3.eu-central-1.amazonaws.com/files/04031834c7a3f21e0446f6ac3d026e271c4c81b5c03eff08dd42540885eaff53d6f491faaf32f84eb9c4ad2d9120eed1ca89141213f9f767b8265a3459176444.pdf

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

Ocaña, J. (2024, February 14). *Es falso que las importaciones agrícolas de fuera de la UE no pasen controles fitosanitarios* [It is false that agricultural imports from outside the EU are not subject to phytosanitary controls]. EFE Verifica. Retrieved from https://verifica.efe.com/productos-agricolas-controles-fitosanitarios-terceros-paises-ue/

Padalko, H. (2025). *Russian disinformation about the US election: AI analysis of narratives*. Centre for International Governance Innovation. Retrieved from https://www.cigionline.org/static/documents/DPH-paper-Padalko.pdf

Piper, A., So, R. J., & Bamman, D. (2021). Narrative theory for computational narrative understanding. In M.-F. Moens, X. Huang, L. Specia, & S. W.-t. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 298–311). Punta Cana, Dominican Republic: Association for Computational Linguistics. doi:10.18653/v1/2021.emnlp-main.26

Puczyńska, J., & Djenouri, Y. (2024). AI in disinformation detection. *Applied Cybersecurity & Internet Governance, 3*(2), 211–232. doi:10.60097/ACIG/200200

Santos, F. C. C. (2023). Artificial intelligence in automated detection of disinformation: A thematic analysis. *Journalism and Media, 4*(2), 679–687. doi:10.3390/journalmedia4020043

Saxena, D., Moon, E. S.-Y., Chaurasia, A., Guan, Y., & Guha, S. (2023). Rethinking "risk" in algorithmic systems through a computational narrative analysis of casenotes in child-welfare. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–19). Association for Computing Machinery, Hamburg, Germany. doi:10.1145/3544548.3581308

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].* doi:10.48550/arXiv.2405.15987

Starbird, K. (2019). Disinformation's spread: Bots, trolls and all of us. *Nature, 571*(7766), 449. doi:10.1038/d41586-019-02235-x

Strand, C., & Svensson, J. (2022). Foreign norm entrepreneurs' misinformation and disinformation narratives on LGBT+ rights in Europe. *Medijska Istraživanja, 28*(2), 109–132. doi:10.22572/mi.28.2.5

Suau, J., & Puertas-Graell, D. (2023). Narrativas de desinformación en España: Alcance, impacto y patrones de difusión [Disinformation narratives in Spain: Reach, impact, and diffusion patterns]. *Profesional de la Información, 32*(5), Article e320508. doi:10.3145/epi.2023.sep.08

Tianda, I. M., Ubadah, M. N., Mardianto, M. F. F., Munawwarah, S. A. A., Ishak, N., Amelia, D., & Ana, E. (2024). Clustering fake news with k-means and agglomerative clustering based on Word2Vec. *International Journal Of Mathematics And Computer Research, 12*(2), 3999–4007. doi:10.47191/ijmcr/v12i2.01

Totth, B. (2024, March 1). *Jön Brüsszel és „ránk sózza" a műhúst?* [Brussels is coming to "force lab-grown meat on us"?]. Lakmusz. Retrieved from https://www.lakmusz.hu/jon-brusszel-es-rank-sozza-a-muhust/

Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., . . . & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *SSRN Electronic Journal.* doi:10.2139/ssrn.3144139

VoxCheck Team. (2023, November 16). *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

Zwin, K. (2024, February 23). *European Union has not yet approved sale of 3D-printed cultivated meat*. AFP Austria. Retrieved from https://faktencheck.afp.com/doc.afp.com.34JA8R9