

More than Memes: A Multimodal Topic Modeling Approach to Conspiracy Theories on Telegram

Elisabeth Steffen

TU Berlin
HTW Berlin
elisabeth.steffen@htw-berlin.de

Abstract

To address the increasing prevalence of (audio)visual data on social media, and to capture the evolving and dynamic nature of this communication, researchers have begun to explore the potential of unsupervised approaches for analyzing multimodal online content. However, existing research often neglects visual content beyond memes and, in addition, lacks methods to compare topic models across modalities. Our study addresses these gaps by applying multimodal topic modeling for analyzing conspiracy theories in German-language Telegram channels. We use BERTopic with CLIP for the analysis of textual and visual data in a corpus of ~40,000 Telegram messages posted in October 2023 in 571 German-language Telegram channels known for disseminating conspiracy theories. Through this data set, we provide insight into unimodal and multimodal topic models by analyzing the symmetry and intersections of topics across modalities. We demonstrate the variety of textual and visual content shared in the channels discovered through the topic modeling, and propose a conceptual framework for the analysis of textual and visual discursive strategies in the communication of conspiracy theories. We apply the framework in a case study of the topic group *Israel Gaza*.

Introduction

The rise of social media has enabled the dissemination of harmful online content, including conspiracy theories (CTs), at an unprecedented speed and scale, and has promoted the formation of communities around them (Bessi et al. 2016; Cinelli et al. 2021). Despite the increasingly multimodal nature of social media, research on CTs has predominantly focused on textual data. According to Mahl, Schäfer, and Zeng (2022), 78.5% of the studies in this area analyze text, while less than 10% examine (audio)visual data. Unsupervised methods offer an opportunity to develop flexible approaches to analyze evolving online content. They can be leveraged to explore large amounts of unstructured data, without the need for human-labeled training corpora. A small number of recent studies explore the potential of such approaches for clustering and analyzing multimodal online communication (Zannettou et al. 2018, 2020; Qu et al. 2023; Theisen and Scheirer 2024b). However, existing research often neglects

visual content beyond memes and, in addition, lacks methods to compare topic models across modalities. Therefore, many questions remain: What information does each modality convey? How do different modalities interact? What are the specific contributions of each modality? We contribute to this field of research by exploring multimodal topic modeling of text-based, image-based, and text-image-based communication on German-language conspiracist Telegram channels.¹ We apply BERTopic (Grootendorst 2022) with the vision language model CLIP (Radford et al. 2021) to text, image, and text-image data. We analyze data from October 2023, a period marked by the Hamas attack on Israel and the Israel-Gaza war, which represent crisis-like events that often spur the proliferation of CTs (van Prooijen and Douglas 2017). Using this data set, we offer insights into both unimodal and multimodal topic models, highlighting the diverse textual and visual content shared in the channels as revealed through the topic modeling process. Further, we aim to understand whether there is correspondence or overlap of similar topics across modalities.

Our work is guided by the following research questions:

- RQ 1: What kinds of content (topic types) and visuals (image types) can be identified in the data set through text-based, image-based, and multimodal topic models?
- RQ 2: How do similar topics from different topic models correspond or intersect across modalities?
- RQ 3: What kinds of textual and visual discursive strategies of conspiracy theories does a qualitative analysis of the results reveal?

For RQ1, our analysis reveals that most identified topics exhibit a clear thematic focus, while others fulfill functional roles, such as advertising or serving as captions. We further observe a variety of image types, with photos and screenshots as the predominant image types of topics, whereas memes play a relatively minor role. This underscores the need for a broader focus on diverse forms of visual online content, beyond memes. Regarding RQ2, our analysis yields one main topic, *Israel Gaza*, consistent across all modalities, followed by topics related to protests and the war in Ukraine. Although similar topics appear across modalities, they are

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹We use the term ‘multimodal’ to refer to text-image-data, although it can encompass a variety of modalities.

represented by different sets of documents in each modality, with text, image, and multimodal topics frequently showing limited to no overlap. Our examination of topic relations indicates less correspondence than anticipated for similar topics across modalities. Different input modalities yield divergent document sets for the same topics. While some topics share a core set of common documents across modalities, others lack such overlap. Our findings suggest that combining modality-specific models could enhance topic modeling by producing richer and more comprehensive topic representations compared to unimodal models. However, further research is needed to better understand the mechanisms and factors underlying the differing patterns of topics across modalities. Extending our analysis to data from a later time period, November 2024, shows that the discovered characteristics regarding symmetry and varying intersection patterns are quite consistent, even under topic shifts. We also demonstrate the applicability of a generative Large Language Model (LLM) for automated topic annotation to improve efficiency.

Drawing on prior research and insights from our data set, we propose a conceptual framework for analyzing textual and visual discursive strategies in CTs (RQ 3). We apply this framework in a case study, demonstrating that integrating textual and visual discursive strategies provides a more holistic understanding of how conspiracy theories are communicated through a qualitative perspective. The case study reveals how key strategies in the dissemination of conspiracy theories can be conveyed through text and images, or both; including providing evidence, authenticating claims, discrediting mainstream narratives, exposing connections, constructing enemy images, framing ingroups and outgroups, issuing directive statements, and fostering emotional responses. To the best of our knowledge, no such framework exists yet for joint text and image analysis.

Related Work

Visuals play an important role for the communication of CTs: images appear to provide immediate ‘knowledge’, and are powerful means in presenting complex narratives (Caumanns and Önnerrfors 2020). A small number of studies examine the relevance of visual content in CTs: Tuters and Willaert (2022) analyze over 470,000 CT-related Instagram posts, guided by the concept of narrative convergence, showing that different conspiracy narratives (e.g., anti-vax, QAnon, and anti-5G), converge over time, with figures like Donald Trump and Bill Gates playing central roles in connecting them. Holzer (2021) finds that visual communication triggers more activity than text, and is strategically used to circumvent detection mechanisms, or, as Al-Rawi (2024) shows, to tactically disguise hateful content as humor using emojis or memes. Qualitative research of CT-related memes has shown that they serve to build ingroup community and criticize outgroups like governments and elites (Gür-Şeker, Boonen, and Wentker 2022). However, other research on far-right visual communication on social media reveals that memes are less prevalent compared to other image types such as photographs and promotional materials (McSwiney et al. 2021), or screenshots of news and social media posts

(Sosa and Sharoff 2022). This indicates an important gap in existing research with its focus on memes.

Clustering and topic modeling (and combinations thereof) are widely used approaches for partitioning data into meaningful groups. Depending on the underlying mathematical approach, they not only assign documents to one or more groups, but might also allow to discover latent themes within large corpora, each representing a “human-interpretable semantic concept” (Abdelrazek et al. 2023). Respective methods involve technical steps such as feature extraction, typically via pre-trained models, and the application of clustering algorithms, but also human-guided steps such as validation and interpretation of results (Peng, Lock, and Ali Salah 2023). While these approaches allow for the discovery of textual, visual or multimodal ‘topics’, their application to social and political phenomena comes with the challenge of connecting automatically generated clusters with theoretically meaningful concepts post-hoc, a process which requires background knowledge and cultural understanding (Peng, Lock, and Ali Salah 2023). The choice of modality is relevant to the technical steps (i.e. how the data is processed, what features are extracted), but is also likely to influence the human interpretation of the topics.

Text-based topic modeling is a common approach for analyzing CTs and other deceptive content on various social media platforms such as Twitter, Reddit, or Telegram. Researchers use it to analyze German far-right communities on Telegram (Wich et al. 2021; Zehring and Domahidi 2023), the QAnon movement (Hoseini et al. 2021), CTs about COVID-19 (Vergani et al. 2022; Kant et al. 2022; Ng, Kloo, and Carley 2024), and Russian disinformation about the Ukraine war (Hanley and Durumeric 2024).

Image-Based Approaches

Unsupervised approaches are also leveraged to analyze image and text-image data on social media. Much of the existing research is on memes, and thus a specific type of visual content. The work of Zannettou et al. (2018) uses image hashing and clustering to analyze the dissemination of memes on the fringe sites Gab and 4chan, and the mainstream platforms Twitter and Reddit. Their findings indicate that politics-related memes are prevalent across both types of platforms, highlighting their potential role in supporting or attacking political figures. Zannettou et al. (2020) use the same pipeline to analyze antisemitic memes on Gab and 4chan. Theisen et al. (2021) develop a pipeline for analyzing meme genres through feature extraction, indexing, graph construction, and clustering. Their method reveals five ‘super-genres’ of meme clusters, including ‘Political’ (27%), and ‘Tweet or Text’ (18%), indicating that screenshots and other images containing text represent a relevant portion of visual online communication.

Multimodal Approaches

Multimodal approaches extend image-based approaches by including textual data, often in the form of image captions (Peng, Lock, and Ali Salah 2023). The work of Li et al. (2023) jointly models text and image data with multiple labels, using a bag of visual words for images and word

embeddings for text. The data set is sourced from an e-commerce platform, and therefore contains distinct visual categories, and rather short and descriptive texts. However, social media data are noisier, have different image and text types, and various text-image relations, i.e., the text is not always a description of the image content. The work of Qu et al. (2023) adapts a CLIP model to this domain by fine-tuning it on a 4chan data set. Text, image, and text-image embeddings obtained from the fine-tuned model are then clustered, finding that dominant clusters are captured in all three settings, however with a high noise level of $\sim 50\text{-}60\%$. Unimodal embeddings form clusters by textual or visual similarity, while text-image embeddings cluster by common semantics. Our work extends these studies by examining a broader range of textual and visual content. Additionally, we apply the approach in the novel context of Telegram, a hitherto unexplored platform in this regard, and the German-language conspiracist movement.

Data and Methods

Telegram was selected as the data source due to its important role in spreading deceptive political content, including CTs. The platform’s lax content moderation, encrypted communication, unlimited message forwarding, and capacity for large broadcast channels make it appealing to political protest movements, but also extremists and violent actors (Baumgartner et al. 2020; Urman and Katz 2022). Previous research highlights Telegram’s importance for the German far-right (Fielitz and Schwarz 2020; Marcks 2023) and for disseminating CTs (Hoseini et al. 2021; Vergani et al. 2022; Imperati et al. 2023; Alvisi, Tardelli, and Tesconi 2024).

Data Collection

We collected text and image data from 683 German-language Telegram channels, previously identified as disseminators of CTs and related content by a research project.² We relied on six CT-related categories (see Table 1), including channels from the German *Querdenken* and *Reichsbürger* movements.³ Data was collected using *Telethon*⁴ between November 4 and 6, 2023, covering all of October 2023. The initial data set contained 244,569 messages, with 15.5% being text-only, and 38.5% including images, highlighting the importance of analysis beyond text.

To obtain meaningful and comparable topic models across modalities, we removed all service messages (which log channel events, such as users joining), kept only messages that contain both text and image, and applied basic text cleaning (removing URLs and user mentions, parsing hashtags) yielding 62,239 messages in total. For each image, we

²<https://dashboard.bag-gegen-hass.net/>, accessed: 2023/17/08

³‘Querdenken’ emerged from the COVID-19 pandemic protests against government measures, uniting far-right, conspiracist, and esoteric groups. The antisemitic ‘Reichsbürger’ movement questions Germany’s sovereignty. It was also part of the COVID-19 protests, visible through the display of the German Reich flag (Rathje 2021).

⁴<https://docs.telethon.dev/en/stable/>

Channel Category	Messages	Channels
Conspiracy Ideology	12,177 (30%)	186
QAnon	11,894 (30%)	73
Esotericism	5,623 (14%)	89
Querdenken	4,508 (11%)	125
COVID-19 Disinfo	3,538 (9%)	72
Reichsbürger	2,370 (6%)	26
Total	40,110	571

Table 1: Channel category distribution in the base corpus (percentages rounded).

calculated a dhash (Krawetz 2013) using the *imagehash* library⁵. We deduplicated messages based on the cleaned text and dhash pair, yielding a base corpus of 40,110 messages, most of which were from the categories *Conspiracy ideology* and *QAnon* (see Table 1).

The base corpus still contained many texts and images occurring multiple times, leading to topics dominated by these duplicates. We applied stricter, modality-specific deduplication, resulting in a text-based (36,717 messages), image-based (34,850 messages), and text-image-based (31,982 messages) subcorpus. Topic modeling was conducted on both the base and the deduplicated subcorpora.⁶

Topic Modeling

We used BERTopic, which applies UMAP to embeddings for dimensionality reduction, and then clusters them with HDBSCAN. A recently added multimodal CLIP-based backend also allows to process image and text-image input. We used the BERTopic standard multilingual model *paraphrase-multilingual-MiniLM-L12-v2* to generate embeddings for text and the *clip-ViT-B-32* model for visual and multimodal data. Dimensionality reduction and clustering were performed using default hyperparameters. Note that the image clusters are based on the image embeddings, while the topic representations are based on image captions automatically generated by the *nlpconnect/vit-gpt2-image-captioning* model, as recommended for BERTopic. Such automated captions allow to handle image-only data sets or corpora with very short texts, which are often poorly suited for topic modeling, but quite common in social media. The multimodal topics are based on averaged text-image embeddings, while topic representation is derived from text only.

HDBSCAN is a soft clustering method that labels noisy data as outliers, rather than forcing them into clusters. In BERTopic, this often results in many documents left unanalyzed, a common issue with this approach (de Groot, Alianajadi, and Haas 2022). We mostly used standard recommended settings for manageability, leaving alternative clustering algorithms and hyperparameter variations for future

⁵<https://pypi.org/project/ImageHash/>

⁶The data sets will be made available upon request for academic use through Zenodo following FAIR principles. Due to potential legal and ethical concerns, image data will not be shared directly, but reconstruction details will be provided.

Modality	Messages	Topics	Outliers
Text	36,717	167	15,189 (41.4%)
Image	34,850	157	12,302 (35.3%)
Multimodal	31,982	146	11,429 (35.7%)

Table 2: Topic modeling results for modality-specific deduplication and `min_topic_size = 20`.

research. We varied the minimum number of documents assigned to a cluster (`min_topic_size`, values 10, 20, 50) and examined the effect of modality-specific deduplication using the base corpus and the subcorpora. Computation was performed on a server with two Nvidia A30 GPUs, an Intel Xeon Gold 6346 CPU, and 251 GB RAM. Execution times were ~ 1 minute for text, ~ 15 -35 minutes for images, and ~ 15 minutes for text-image data, faster than image-only due to skipping of image captioning.⁷

Evaluation Our quantitative and qualitative evaluation of different hyperparameter configurations focused on the number and size of topics, the number of outliers, and a manual inspection of topic quality based on their textual and/or visual representations. A topic was considered high quality if its focus was clearly conveyed through its name and its representative documents. Conversely, vague or thematically mixed names and documents indicated low quality. Topic quality was similar for `min_topic_size` of 10 or 20, but 10 led to more outliers and more topics, increasing the manual annotation effort. A `min_topic_size` of 50 reduced outliers but decreased topic quality. Modality-specific deduplication improved topic quality. The optimal configuration across modalities was thus `min_topic_size = 20`, with modality-specific deduplication (see Table 2).

Annotation and Grouping The validation and interpretation of topic modeling results can be challenging, and available quantitative measures do not necessarily correspond with human judgment. Manual inspection of topics and the assignment of names (topic assignment) is a common validation approach (Zhang and Peng 2022; Ying, Montgomery, and Stewart 2022). In the context of our study, topic annotations were conducted by the author, who has field knowledge and practical experience in social media data annotation. In addition to the assignment of a topic name, the topic type and image type of each topic were annotated. The corresponding label sets were developed inductively after an initial exploration of the results.

The category **topic type** characterizes the general type of a topic and helps determine its relevance for further analysis. A topic was labeled as *content*, if its main thematic focus could be identified. In contrast, the type labels *advertising* and *caption* were assigned to functional topics deemed irrelevant for content analysis. The topic type *content unclear*, was used for topics with ambiguous or indeterminate thematic focus. Each topic was assigned exactly one topic type label. **Image type** was assigned to image

⁷The code is available under: <https://github.com/eli-quereli/multimodal-tm/>

Category	Text		Image		Multimodal	
	κ	F1	κ	F1	κ	F1
Topic type	0.91	0.68	0.35	0.62	0.86	0.46
Image type	–	–	0.85	0.74	0.70	0.51

Table 3: Intra-annotator agreement (Cohen’s κ) and LLM-based annotation performance (macro average F1) for topic types and image types.

and text-image topics to distinguish different visual types. The initial label set included *photo*, *photo with text*, *photo comic*, *photo black-white*, *screenshot news*, *screenshot social media*, *meme*, *quote image*, *graphic*, *infographic*, *book cover*, *poster*, *document*, *template*, and *map*. Each topic received exactly one label. Due to the limited instances in some categories, the label set was later revised to a subset (*photo*, *screenshot news*, *screenshot social media*, *meme*, *infographic*, and *map*), ensuring a manageable number of well-represented categories.

Text topics were labeled based on topic keywords and representative documents. Image topics were based on nine representative images selected by the model. The textual representation of image topics, based on the generated captions, was generally not useful for annotation. Multimodal topics were labeled using text and image representations. For each *content* topic, a topic name was then assigned by the annotator.

Finally, we manually grouped the topics into broader topic groups by compiling them into a single spreadsheet and analyzing thematic similarities across modalities. For example, topics related to Israel or Gaza were grouped under *Israel Gaza*. Topics that did not clearly relate to another topic were grouped under *Misc*.

Intra-Annotator Agreement For validation, the author re-annotated the topic type and the image type assignments on a random sample of 50 topics per modality after three days. The intra-annotator agreement, measured as Cohen’s κ (see Table 3), was moderate to high, except for the topic type in the image-only setting. Error analysis revealed that inconsistencies in annotating image topics stemmed from relying on stylistic similarities rather than semantic coherence (Zhang and Peng 2022, cf.). We resolved label conflicts to obtain the final topic and image type labels.

LLM-Based Topic Annotation To improve efficiency, we evaluated the use of a generative Large Language Model (LLM), *gemini-1.5-flash*, for annotating and naming topics. The model was prompted to generate concise topic names and assign topic and image types (using a simplified label set for image types) and short definitions for each label.⁸ The model’s performance was evaluated against human annotations as the gold standard, with macro average F1 scores showing acceptable to good agreement (see Table 3). For topic name evaluation, we computed cosine similarity between manual and LLM-generated names using

⁸For prompt details please refer to the accompanying code repository.

the paraphrase-multilingual-MiniLM-L12-v2 model. The median similarity scores indicate moderate to high similarity for all modalities (text: 0.69, image: 0.62, multimodal: 0.65). The annotation process was reduced from about 4 hours (manual) to 15 minutes per modality, while topic grouping (around 30 minutes in total), remained manual to preserve human assessment.

Temporal Transfer To assess the generalizability of our insights, we analyzed data from the same Telegram channels in November 2024, covering events like the U.S. presidential elections (Nov 6th) and the breakup of Germany’s ‘Ampel’ coalition (Nov 7th). We collected 219,076 messages, 43% with images and 13% text-only. Using the same hyperparameters (`min_topic_size = 20`, modality-specific deduplication) for topic modeling, and LLM-based topic annotation, we identified 212 topics in the text setting, and 160 topics each in the image and in the multimodal settings.

Results

Topic and Image Types

Examining and comparing topic types and images types across modalities allows us to answer RQ 1. As Table 4 shows, the **topic type content** is predominant across modality settings, with the highest proportion found in the image-based topic model. Interpreting image-only topics appears to be easier than text-based or multimodal topics. The higher proportion of *advertising* topics in November compared to October, along with the lower proportion of *caption* topics, may result from overlap in the content or characteristics of these two types. The higher proportion of *content unclear* in the multimodal setting could be due to the difficulty for human annotators to interpret dual-modality topics, which requires to consider both textual and visual aspects of a topic.

The predominant **image type** category in October 2023 is *photo*, followed by *screenshots*, which in total account for around 20% of the topics. The prominence of screenshots increases in November 2024, reaching around 50%, while the photo proportion decreases but remains relevant. *Memes* range between 1% and 5% across modalities and time periods. *Maps* drop from 6-7% in October 2023 to (less than) 1% in November 2024, likely due to reduced relevance of geopolitical topics like *Israel Gaza*. The low proportions of *infographics* and maps also reflect differences in category granularity, with broader categories like photo and screenshots dominating. However, despite also being a broad category, memes remain minor, clearly indicating they are not key genres.

Dominant Topic Groups

The manually assigned topic groups offer a broad view of thematic differences and similarities across modalities. Table 5 shows the five largest topic groups per modality (excluding the *Misc* group due to its lack of thematic expressiveness) in October 2023. Across all modalities, *Israel Gaza* dominates, reflecting the data set’s focus on October 2023, including the Hamas attack on Israel (Oct 7th), and the ensuing war. Its prevalence varies by modality, with

36.2% in text, 12.9% in images, and 16.7% in the multimodal data. In the text-based setting, it forms a single large topic, while in the image and multimodal setting, it is divided into 18 and 11 smaller topics, respectively. Text captures more messages with a clear main focus, but lacks the finer granularity provided by image and multimodal settings, which offer more detailed perspectives at the cost of comprehensiveness. The image and multimodal setting share the topic group *Protests*, while *Ukraine War* appears in text and image, though with differences in rank and size. The topic group *Self-Love*, present in text and multimodal settings, likely originates from channels in the Esotericism channel category. Among these top-ranked topic groups, we will focus on *Israel Gaza*, *Protests*, and *Ukraine war* in the following, as they exhibit a greater potential for CTs compared to self-love topics.

In November 2024, we find three topic groups ranking in the top five across all modalities: *German Politics*, *Ukraine War*, and *U.S. Politics*. *COVID-19* appears in text and multimodal, while *Finance, Military & War*, and *Conspiracy (specific)* each rank in the top five of one modality. This marks a clear thematic shift, with *Israel Gaza* losing its dominant role.

Topic Relations Across Modality Settings

To better understand topic relations across modality settings (RQ 2), we conducted a series of analyses examining how topics are assigned and overlap across different models.

Topic Symmetry For each *source topic*, we identified the *target topic* in another modality setting with the largest intersection in terms of documents with the source topic. We then checked if the target topic maps back to the source topic, indicating symmetric alignment. The symmetry ratio, calculated as the proportion of symmetric pairs, is detailed in Table 6. Text-image and image-text pairs have the lowest symmetry, while image-multimodal and multimodal-image have the highest, indicating stronger alignment between visual and multimodal data. The results for the November 2024 data are notably consistent.

Messages Over Time Figure 1 shows the message frequency over time for the dominant topic groups. *Israel Gaza* shows a sharp increase on October 7th (Hamas attack on Israel) and a slight decline over time. Text dominates with about 50% of messages, while image and multimodal make up roughly 25% each. We found consistent modality-based proportions over time for the other (smaller) topic groups as well, with differing modality distributions. E.g., *Ukraine war* is mostly represented through multimodal topics (two-thirds) and text (one-third), with marginal relevance of image topics. In contrast, *Protests* is dominated by image (two-thirds), and multimodal topics (one-third), with minimal text presence.

Topic Intersections The topic intersections across models differ notably between the dominant topic groups in October 2023 (see Figure 2). *Israel Gaza* has an overlap of all three settings, but also distinct documents unique to each. Around 60% of documents in the text group are distinct, while image

Topic Type	Text		Image		Multimodal	
	Oct'23	Nov'24	Oct'23	Nov'24	Oct'23	Nov'24
Content	53%	53%	80%	89%	60%	61%
Content unclear	16%	17%	17%	8%	34%	24%
Advertising	10%	21%	3%	4%	4%	9%
Caption	22%	9%	–	–	2%	6%
Image Type						
Photo	–	–	35%	28%	43%	23%
Screenshot news	–	–	11%	20%	10%	24%
Screenshot social media	–	–	11%	23%	7%	30%
Infographic	–	–	7%	5%	8%	6%
Map	–	–	6%	<1%	7%	1%
Meme	–	–	2%	5%	1%	5%

Table 4: Topic and image type distribution per modality for October 2023 and November 2024.

Topic Group	Text Rank	Img Rank	Mm Rank	Text Size	Img Size	Mm Size
Israel Gaza	I	I	I	4,700 (36.2%)	1,900 (12.9%)	2,071 (16.6%)
<i>Self-Love</i>	II		V	894 (6.9%)		873 (7.0%)
<i>Protests</i>		II	III		1,666 (11.3%)	1,205 (9.7%)
<i>Ukraine War</i>	V		II	560 (4.3%)		1,491 (12.0%)
COVID-19	III			817 (6.3%)		
German Gov		III			1,251 (8.5%)	
Conspiracy (general)	IV			725 (5.6%)		
Esotericism		IV			838 (5.7%)	
Trump			IV			1,073 (8.6%)
Far-Right Statements		V			829 (5.6%)	

Table 5: October 2023 data set: Top five topic groups by message count, per modality. *Israel Gaza* ranks first in all three settings, while three other groups (in italics), appear in the top five of two modalities. Percentages reflect modality-based proportions.

Modality Pairs	SR Oct'23	SR Nov'24
Text-Img / Img-Text	0.15 / 0.16	0.17 / 0.23
Text-Mm / Mm-Text	0.24 / 0.28	0.28 / 0.37
Img-Mm / Mm-Img	0.58 / 0.63	0.57 / 0.57

Table 6: Symmetry ratio (SR) for modality pairs in October 2023 and November 2024 topics.

and multimodal show larger overlap. In contrast, the *Protests* topic group has a very small core, but a large overlap of text and multimodal. The *Ukraine War* topic group completely misses a shared core, and generally shows little overlap. Almost 90% of the documents from the multimodal setting are not identified by the text nor the image model. While *Israel Gaza* is predominantly text-based, *Protests* is dominated by the image setting. For *Ukraine War*, multimodal messages prevail, and the image setting is marginal. We found similar variations of overlap in the November 2024 data. These insights show that topic representation varies not only by input modality but also by topic. This requires further investigation to understand what drives modality predominance per topic, or how core documents differ from those found in only one modality setting of a topic. After this quantitative investigating of topics, we now turn to qualitative insights to explore CT-related discursive strategies in greater depth.

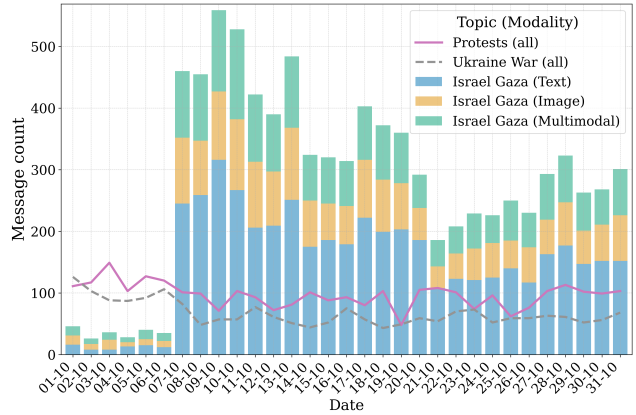


Figure 1: Message frequency over time for topic groups *Israel Gaza*, *Protests*, and *Ukraine War*.

Textual and Visual Discursive Strategies in Conspiracy Theories

Conspiracy theories are “efforts to explain an event or practice by referencing the machinations of powerful people” (Sunstein and Vermeule 2009). As such, they are dynamic meaning-making processes (Wood 2018) in which conspiracy theorists position themselves as narrators or whistle-

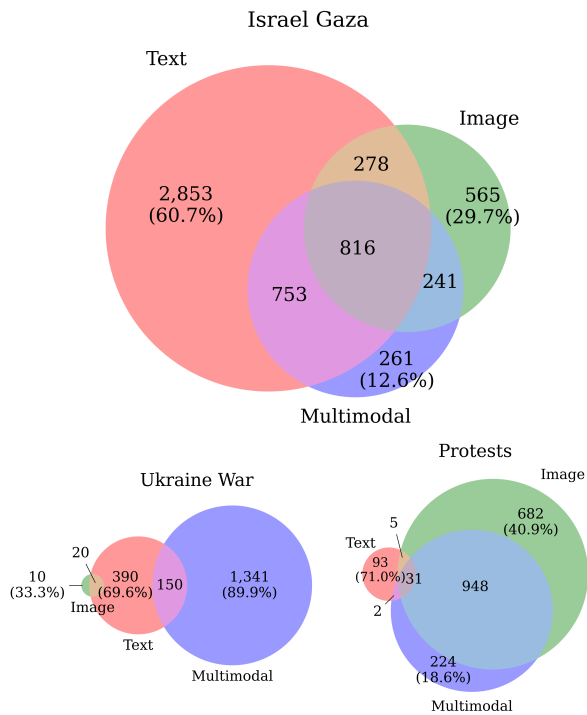


Figure 2: Intersections of topic groups *Israel Gaza*, *Ukraine War* and *Protests* across models. Percentages show document proportions within each modality setting.

blowers, challenging established views and attempting to persuade audiences (Caumanns and Önnersfors 2020), and employing various *discursive strategies* (Kou et al. 2017).

While much research examines textual strategies (Kou et al. 2017; Wood 2018), the role of visuals remains underexplored, and no unified framework exists for analyzing both modalities. To address this gap, we propose a conceptual framework for analyzing textual and visual strategies in CTs, informed by existing research and insights from our data set. We apply this framework in a case study of the *Israel Gaza* topic group to address RQ 3. Developing this framework presents two challenges: First, it requires categories which are applicable to both textual and visual strategies. Second, the role of visuals as discursive tools can be debated due to their often ambiguous nature. However, following Richardson and Wodak (2009), we argue that visuals, like text, can serve as meaning-making tools, with both requiring context-sensitive interpretation.

Conceptual Framework

To develop a framework that captures CTs as active meaning-making processes involving both textual and visual strategies, we draw on prior research from discourse analysis in political communication, social psychology on rumors, and CTs in online discussions (Bordia and Difonzo 2004; Kou et al. 2017; Wood 2018), narratological approaches to CT images (Caumanns and Önnersfors 2020), and discourse analysis of CT memes (Gür-Şeker, Boonen, and Wentker 2022). We define *strategy* based on Richardson and Wodak

(2009) as:

a (more or less accurate and more or less intentional) plan of practices, including discursive practices, adopted to achieve a particular social, political, psychological, or linguistic goal.

In the context of CTs, the goal is to persuade audiences of a CT’s veracity or to delegitimize official accounts (Caumanns and Önnersfors 2020). We thus define *CT-related discursive strategies* as textual or visual techniques employed to convince audiences of a conspiracy theory’s validity or to challenge and undermine official accounts. Both CT believers, aiming to ‘enlighten’ others, and opportunistic actors, such as those using CTs in information warfare (Madisson and Ventsel 2020), may employ these strategies. This implies that the intentionality of these strategies may vary (Richardson and Wodak 2009), and that their use can be analyzed independently of the poster’s beliefs or intentions.

Previous studies provide important insights into textual CT-related strategies. The Rumor Interaction Analysis System (RIAS), was originally developed by Bordia and Difonzo (2004) for analyzing rumor discussions in newsgroups, categorizing strategies like authenticating claims (e.g., citing media or personal expertise), providing information, directives, and using rhetorical or sarcastic statements. Their study shows how people attempt to persuade each other, often citing personal experiences or media sources to confirm or contradict rumors. The five most common types of statements are interrogatory statements, sense-making (attempting to falsify or verify a rumor), providing information, disbelief, and digressive (unrelated) statements. The authors conceptualize rumors as efforts to reduce uncertainty in crisis-like, often ambiguous situations, typically involving events with significant impact – characteristics shared with CTs. These commonalities motivated Wood (2018) to adapt the RIAS schema for analyzing CTs related to the Zika virus on Twitter in 2015-2016, focusing on six categories: reference, belief, disbelief, authenticating, directives, and rhetorical questions. They find that CT-supporting tweets more often cite authorities and use rhetorical questions compared to CT-opposing posts. We incorporate authenticating, directives, and rhetorical questions into our framework, as the other categories do not represent strategies, but refer to stance, which is outside the scope of our analysis. While the work of Wood (2018) offers some suitable categories, we find the range of provided strategies too limited for analyzing our data. To extend the set of strategies, the work of Kou et al. (2017) offers a valuable source. Their work draws on a different strand of rumor research (a study on AIDS denialism by Meylakhs et al. (2014)), which they adapt to analyze CT-related posts about the Zika virus on Reddit. They identify a number of strategies in CT-supporting posts including selective citation of authoritative sources, cultural references (e.g., movies), narrative elaboration (adding unverifiable details), ‘connecting the dots’ (linking unrelated information), ideological arguments (e.g., system critique), casting doubt, and promoting skepticism. While selective citation and connecting the dots align with authentication in Wood (2018), other strategies such as system critique, casting doubt, and

promoting skepticism demonstrate how conspiracy theorists undermine official narratives while constructing alternative explanations. We consider this strategy of discrediting mainstream accounts as important aspect of CTs, and therefore include it in our framework.

While the studies presented so far are restricted to text data, the work of Caumanns and Önnersfors (2020) addresses the role of visuals in CT meaning-making. Applying a narratological approach to historical CT imagery, they identify the following key strategies: pattern recognition (e.g., network imagery), agency detection, coalition mapping (e.g., collages), enemy imagery, and visualizing secrecy. They also highlight the interplay of text and visuals in creating complex causal links or pseudo-evidence, e.g. in pseudo-scientific infographics or collages. While developed for images, we find their schema useful for our joint framework, because the presented strategies can be employed both through text and images. Finally, Gür-Şeker, Boonen, and Wentker (2022) highlight the role of multimodal CT posts in community building and expressing skepticism toward elites, governments, and the media. While their analysis extends beyond the post itself to include user interactions (e.g., comments, likes, profiles), which is beyond the scope of our study, their work offers valuable insights for our study: the expression of skepticism aligns with the strategy of discrediting mainstream accounts, as noted by Kou et al. (2017). Additionally, the focus on community building introduces the crucial aspect of constructing an ingroup vs. outgroup dynamic in CTs.

Building on these approaches and insights from our data set, we select categories that are applicable to both textual and visual dimensions, and suitable for our data. The resulting framework, presented in Table 7, includes eight discursive strategies. While tailored to this study, the categories are designed for broader use with CT-related data across contexts and time periods.

Case Study: The Topic Group *Israel Gaza*

We used the proposed framework to analyze the predominant topic group in the October 2023 data set, *Israel Gaza*. For each modality, we explore the thematic aspects of the topics and analyze CT-related textual and visual discursive strategies.⁹

Due to the large volume of 4,700 messages in the **text-based topic**, we performed a second round of topic modeling to obtain more granular topics. Using similar hyperparameters for consistency (`min_topic_size = 15`), we found 51 topics. These were manually grouped by thematic similarity, following the same process as with the initial topic modeling.

The identified topics mention political and military actors (individuals and groups) in the region, and cover different aspects of military actions in Gaza, with the attack on a hospital in Gaza standing out as a distinct event. They also refer to protests against Israel's actions, and discuss the role of the U.S. administration and the German government in the war.

⁹Note that all text examples are presented as English translations of the posts.

Other topics address the history of the Israeli-Palestinian conflict, or claim to reveal the reputed 'truth' about the Hamas attack on Israel. More general topics include Zionism, media propaganda, and the purported relation between Islam, immigration, and antisemitism.

Applying the conceptual framework reveals a number of discursive strategies: Topics discussing the history of the Israeli-Palestinian conflict as well as claims to 'reveal the truth' about the Hamas attack can be interpreted as providing evidence. A closely related strategy are acts of disclosure and revealing connections: For instance, the topic *Truth Hamas attack* suggests that the attack was actually orchestrated by Israel, the USA, 'Zionists', and Prime Minister Netanyahu in collaboration with a 'globalist elite'. This claim explicitly connects multiple actors and aligns them with a shared, malicious goal, reinforcing a conspiratorial worldview. One particularly explicit post claims that the incident was "*an 'inside job' by the globalist elite*" in cooperation with the USA and Israel, "*as part of the grand master plan for World War III.*" The strategy of authentication occurs in posts which claim to refer to insider knowledge (without naming them explicitly). For example, a user refers to unnamed "*insiders*" and "*officials*", which creates an impression of credibility. Irony and sarcasm are employed to discredit mainstream accounts: The following message sarcastically criticizes the framing of oppositional views in mainstream discourse: "*Corona deniers, anti-vaccinationists and Putin haters are out! Now there are ISRAEL HATERS.*" The statement also reflects the ingroup vs. outgroup strategy, portraying the ingroup as marginalized rebels unjustly dismissed by the mainstream. Another post adapts Reichsbürger rhetoric to a new context by claiming that Israel is not a nation-state but a corporate entity. This aligns with the strategy of constructing enemy images by using delegitimizing language. An example of emotionalization is this post, which uses capslock and exaggerations while also serving as an act of disclosure, claiming the true goal behind the Israel-Gaza war is:

THAT ISRAEL AND THE ZIONISTS THEN OWN THE GAZA STRIP'S TENS OF TRILLIONS OF DOLLARS WORTH OF GAS RESERVES FOR THE NEXT HUNDRED YEARS OR MORE.

or by asking the audience if they want to "*understand who is behind the Israel-Palestine war and what goals are being pursued?*", prompting them to click on link for answers, attempting to involve readers further.

The **image-based topics** feature diverse visuals, including photos of explosions and debris, portraits of key figures (e.g., Benjamin Netanyahu), screenshots, memes, graphic representations like maps, cartoons, and symbolic images such as flags. Like in the text setting, topics cover political and military actors, aspects of the Israel Gaza war, and the history of the conflict. Photos of explosions and debris (Figure 3, row 1), highlighting the war's devastation and evoking fear, exemplify emotionalization. Symbols like flags (e.g., of Ukraine and Israel, Figure 3, row 2) visually reveal connections, linking nations or ideologies to construct associations. Memes and cartoons (Figure 3, row 2) support the strategy of

Discursive strategy	Textual manifestation	Visual manifestation
Providing evidence	Presenting information as factual claims or evidence	Images conveying factual authority (e.g., maps, infographics, documentary-style photos)
Authenticating	Referring to perceived experts or insiders to build credibility	Quote-based sharepics or screenshots showing statements by recognized figures
Discrediting mainstream accounts	Employing irony, sarcasm, and derogatory language; highlighting perceived contradictions or inconsistencies; rhetorical questions	Political memes or cartoons; annotated screenshots emphasizing perceived contradictions
Acts of disclosure and revealing connections	Identifying actors and alleging hidden connections, often tied to a shared malevolent agenda	Depicting actors jointly to illustrate alleged connections
Constructing enemy images	Employing derogatory or dehumanizing language	Visual representations such as portraits, caricatures, or memes
Ingroup vs. outgroup	Framing the ingroups as enlightened, rebels, or persecuted victims	Images of supportive crowds or prominent leaders or heroes (e.g., Trump, Musk)
Directives	Encouraging audience engagement through directives (e.g., click a link, explore further)	Sharepics or posters designed to mobilize action (e.g., attend a demonstrations, listen to a radio show)
Emotionalization	Leveraging emotionally charged language to evoke fear, anger, or other strong reactions	Employing emotionally evocative imagery to provoke fear, anger, or similar responses

Table 7: Conceptual framework for the analysis of textual and visual CT strategies

opposing or discrediting mainstream accounts. These visuals often employ irony or sarcasm and create links between unrelated events such as the Israel-Gaza war, the Ukraine war, and COVID-19. In doing so, they reinforce oppositional narratives through humor and simplification. Textual elements play a central role in many images, with screenshots being the most text-heavy examples. The technique of ‘quoting’ via screenshots, where text from other sources is directly captured and presented, aligns with strategies of providing evidence and authenticating. Highlighting details in screenshots serves to emphasize key information, expose inconsistencies, or delegitimize specific actors (Figure 3, row 3). Maps and infographics often serve the purpose of providing evidence but can also reflect the ideological framing. For example, some maps label the Hamas attack as “*Palestinian uprising*”, implicitly framing Israel as a suppressor and supporting the construction of enemy images (Figure 3, row 4).

The **multimodal topics** offers fewer aspects than the image setting, but the inclusion of textual information enhances interpretation and yields visually more diverse topics. We found one topic focused on actors, while others address the Israel Gaza war, protests, and geopolitical issues, including references to historical events like the topic *1948 Nakba*. Figure 4 (row 1) shows images of Israel’s Prime Minister Benjamin Netanyahu. Accompanying texts portray him as a conspiratorial actor, often in antisemitic terms, illustrating the construction of enemy images and revealing connections strategies. One post labels him a “*Zionist prostitute [...] with the Zionist master lord Jacob Rothschild*”, while another accuses Netanyahu and Zelensky of being “*Jewish N...s*” (omission in original) opposed to peace, and a third represents the strategy of providing evidence, referring to “*internal sources*”, claiming that Netanyahu supports using force against Iran’s nuclear program. Figure 4 (row 2) shows examples of posts employing providing evidence and acts of

disclosure: the first image, a black-and-white map purportedly from the radical Israeli group Irgun, is presented as evidence of Israel’s alleged colonial agenda, with accompanying text claiming that it “*is not an antisemitic invention that Israel is striving for a ‘Greater Israel’*”. The second image features a long and cryptic QAnon-style text with phrases like “*1st Map Palestine 1905*” and “*Balfour Declaration*”. Notably, the post was assigned as outlier in the image setting, but assigned to the *Israel Gaza* topic in both the text and multimodal setting. Despite no visual similarity to the other two images in the row, apparently it’s textual dimension establishes a thematic link. This post exemplifies the authenticating strategy, presenting a screenshot of a post from the platform X by a conspiracy theorist with over 400K followers, an apparently influential figure. The third image is an Al Jazeera infographic, accompanied by text suggesting Gaza’s size could have tripled by 2020, possibly referencing then former US-American President Donald Trump’s alleged plan for a larger Palestinian state on the Sinai Peninsula. The use of an infographic conveys information in a factual manner, aimed at providing evidence.

Cartoons in the context of the topic are employed for acts of disclosure, but also serve to create enemy images (Figure 4, row 3). The first cartoon shows a US-American puppet master with a Jewish ally, identified by national, religious, and political symbols. The text claims Hamas and IS were founded by the USA and Jews, who are further demonized through Nazi associations. The middle cartoon depicts Gaza as a prison under constant threat of bombs, while a Jewish soldier, marked by a Star of David on his helmet, portrays Israel, represented by its army, as victim. The cartoon claims to reveal Israel’s alleged evil and hypocrisy in falsely presenting itself as a victim. The accompanying text, “*I love it*”, reflects the poster’s support for the cartoon’s message. The cartoon on the right conveys a similar idea, highlighting double standards in how the war is perceived: it suggests

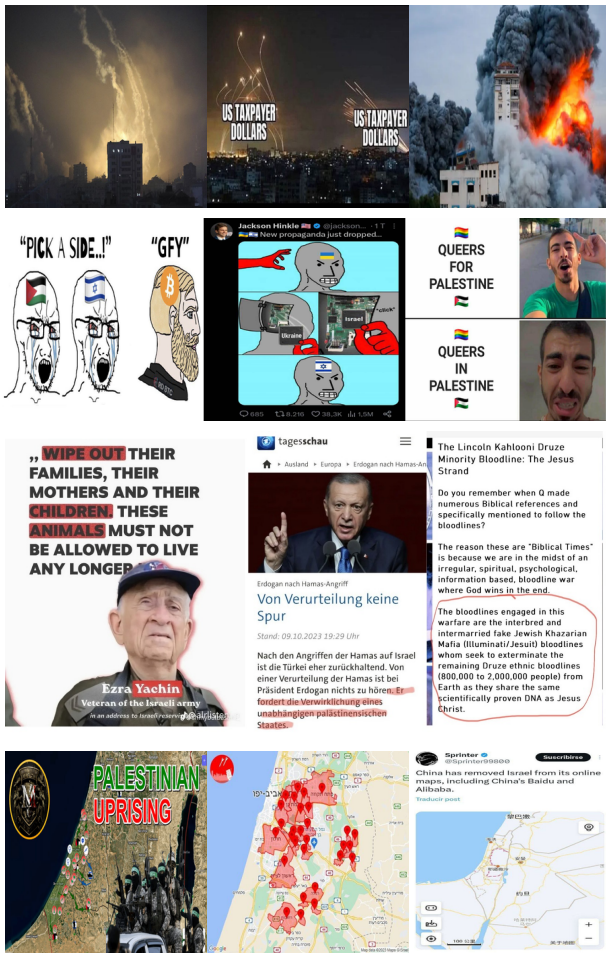


Figure 3: Image-based model: Examples from the *Israel Gaza* topic group.

that Palestinian civilian deaths are disregarded, while Israeli deaths receive more attention. The post’s text reinforces this, stating that “*In the Anglo-Saxonian media it’s simple: in Israel there are many murders, and in Gaza there are many deaths*”. This implies that ‘the media’ denies the deliberate killing of Palestinians while emphasizing Israeli casualties. The cartoon thus serves the strategies of discrediting mainstream accounts, accompanied by emotionalization.

Conclusion

Our study applies multimodal topic modeling to German-language conspiracist Telegram data from October 2023, with additional analysis of a November 2024 data set to evaluate temporal transferability. We demonstrate the utility of BERTopic for analyzing German-language text, image, and text-image Telegram posts, and integrate LLM-based topic annotation to increase efficiency and practical utility of the approach. Regarding predominant topics, we find that in October 2023, the *Israel Gaza* topic dominates across all topic models, while November 2024 focuses on *German Politics*, *Ukraine War*, and *U.S. Politics*. Our analysis further reveals



Figure 4: Multimodal model: Examples from the *Israel Gaza* topic group.

various image types, with photos and screenshots dominating, while memes play a minor role. The discovered visual types illustrate the diversity of visual content on Telegram, and also reflect varying text-image ratios: photos prioritize visuals, screenshots emphasize text, and memes represent an intermediary type. Our analysis of topic characteristics in different models reveals that topics are communicated differently: For instance, the largest part of the *Israel Gaza* topic group is present in text, while *Protests* rely heavily on images. Further, each model captures a substantial number of documents not captured by the other models, resulting in often small topic intersections. Topic symmetry is lower between text and image, and higher between image and multimodal models. Temporal transfer overall confirms our findings regarding the distribution of topic types, image types, symmetry and intersections of similar topics across models. Further, building on prior research on CT-related textual and visual meaning-making, we propose a novel framework for analyzing CT-related discursive strategies, comprising eight categories suited for the joint analysis of text and image data. Applying it to the *Israel Gaza* topic, we identify strategies such as providing evidence, authenticating, discrediting mainstream accounts, revealing connections, constructing enemy images, ingroup-outgroup framing, directives, and emotionalization, in text and image data.

Discussion

Our study is subject to the following limitations: first, while we selected October 2023 to capture a period marked by heightened CT activity due to major crisis events such as the Hamas attack on Israel and the Israel-Gaza war, this choice may have biased the data set toward these events. To address this, we analyzed data from a later period, confirming findings such as screenshot predominance and low symmetry,

even with thematic shifts. Second, our embedding model, pre-trained on descriptive image captions, is not optimized for noisy, user-generated texts with complex text-image relationships (e.g., commentative text-image pairs) (Theisen and Scheirer 2024a, cf.). Fine-tuning CLIP on Telegram data and refining multimodal embedding strategies, such as adjusting text embedding weights, could improve topic quality. Third, the absence of standardized evaluation metrics for BERTopic models introduces subjectivity and limits comparability. Expanding the annotator pool in future studies could enhance validation reliability, and provide an even more solid gold standard for the evaluation of LLM-based performance.

Despite its limitations, our work makes important contributions to the advancement of multimodal topic modeling. It demonstrates the potential of combining text-based, image-based, and multimodal topic models: each model identifies unique sets of documents for shared topics, and their integration can provide a more comprehensive understanding of the data set while reducing the prevalence of outlier documents. The suggested approach is particularly valuable for researchers and practitioners in social media monitoring, as it uncovers emerging themes, including image-driven topics, in dynamic contexts. Our integration of LLM-based topic annotation enhances the practical utility of our approach by increasing its efficiency. By incorporating images, our study reveals topics that are primarily expressed visually, and therefore would be missed by text-based topic models. We contribute to multimodal research by highlighting the variety of visual types in Telegram channels, particularly the overlooked role of screenshots. The identified image clusters can also be useful in strategic sampling of training data for supervised classification models. Our conceptual framework enhances the integrated analysis of textual and visual data, and our case study demonstrates how these strategies empirically manifest in CT-related online content.

Future work could explore the reasons behind modality-specific differences in topic representation. While our study highlights notable variations in topic symmetry and intersections, a deeper investigation into the underlying causes is needed. Investigating the temporal dynamics of CT content across different periods, especially during less eventful times, could provide a broader understanding of how topics evolve over time. A systematic comparison of different platforms and content types could also offer valuable insights into the cross-platform dynamics of multimodal discourse. Insights into the relevance of screenshots open new avenues for studying sharing practices, information propagation and content dissemination, as well as the role of (manipulated) screenshots in disinformation on social media. Future work could explore these practices further in order to complement existing research by including visual content. Our conceptual framework offers a foundation for both qualitative studies, such as multimodal discourse analysis, and quantitative studies, including annotation schema design for content analysis. It could be applied to diverse corpora, enhancing analyses of how CTs are communicated across platforms through text, images, or their combination.

References

- Abdelrazek, A.; Eid, Y.; Gawish, E.; Medhat, W.; and Hassan, A. 2023. Topic modeling algorithms and applications: A survey. *Information Systems*, 112.
- Al-Rawi, A. 2024. The Far-Right Online Communities and the Dissemination of Hate. In Al-Rawi, A., ed., *Online Hate on Social Media*, 23–57. Springer, Cham.
- Alvisi, L.; Tardelli, S.; and Tesconi, M. 2024. Unraveling the Italian and English Telegram Conspiracy Spheres Through Message Forwarding. In *Proceedings of ASONAM 2024*. Berlin, Heidelberg.
- Baumgartner, J.; Zannettou, S.; Squire, M.; and Blackburn, J. 2020. The Pushshift Telegram Dataset. In *Proceedings of ICWSM 2020*, 840–847. AAAI.
- Bessi, A.; Zollo, F.; Vicario, M. D.; Puliga, M.; Scala, A.; Caldarelli, G.; Uzzi, B.; and Quattrociocchi, W. 2016. Users Polarization on Facebook and Youtube. *PLOS ONE*, 11(8).
- Bordia, P.; and Difonzo, N. 2004. Problem Solving in Social Interactions on the Internet: Rumor As Social Cognition. *Social Psychology Quarterly*, 67(1): 33–49.
- Caumanns, U.; and Önnersfors, A. 2020. Conspiracy Theories and Visual Culture. In *Routledge Handbook of Conspiracy Theories*. Routledge.
- Cinelli, M.; De Francisci Morales, G.; Galeazzi, A.; Quattrociocchi, W.; and Starnini, M. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9).
- de Groot, M.; Aliannejadi, M.; and Haas, M. R. 2022. Experiments on Generalizability of BERTopic on Multi-Domain Short Text. arXiv:2212.08459.
- Fielitz, M.; and Schwarz, K. 2020. Hate not Found?! Das Deplatforming der extremen Rechten und seine Folgen. https://www.idz-jena.de/fileadmin/user_upload/Hate_not_found/WEB_IDZ_FB_Hate_not_Found.pdf. Accessed: 2024-07-30.
- Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794.
- Gür-Şeker, D.; Boonen, U. K.; and Wentker, M. 2022. #conspiracymemes: A framework-based analysis of conspiracy memes as digital multimodal units and ensuing user reactions on Instagram. In *Discourse Approaches to Politics, Society and Culture*, volume 98, 267–294. John Benjamins Publishing Company.
- Hanley, H. W. A.; and Durumeric, Z. 2024. Partial Mobilization: Tracking Multilingual Information Flows Amongst Russian Media Outlets and Telegram. arXiv:2301.10856.
- Holzer, B. 2021. Zwischen Protest und Parodie: Strukturen der "Querdenken"-Kommunikation auf Telegram (und anderswo). In Reichardt, S., ed., *Die Misstrauensgemeinschaft der "Querdenker": Die Corona-Proteste aus kultur- und sozialwissenschaftlicher Perspektive*, 125–157. Frankfurt: Campus Verlag.
- Hoseini, M.; Melo, P.; Benevenuto, F.; Feldmann, A.; and Zannettou, S. 2021. On the Globalization of the QAnon Conspiracy Theory Through Telegram. In *Proceedings of WebSci 2023*, 75–85.
- Imperati, V.; La Morgia, M.; Mei, A.; Mongardini, A. M.; and Sassi, F. 2023. The Conspiracy Money Machine: Uncovering Telegram's Conspiracy Channels and their Profit Model. arXiv:2310.15977.
- Kant, G.; Wiebelt, L.; Weisser, C.; Kis-Katos, K.; Luber, M.; and Säfken, B. 2022. An iterative topic model filtering framework for short and noisy user-generated data: analyzing conspiracy theories on twitter. *Int J Data Sci Anal*, 6: 1–21.

- Kou, Y.; Gui, X.; Chen, Y.; and Pine, K. 2017. Conspiracy Talk on Social Media: Collective Sensemaking during a Public Health Crisis. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW): 1–21.
- Krawetz, N. 2013. Kind of Like That. <https://www.hackerfactor.com/blog/index.php?/archives/529-Kind-of-Like-That.html>. Accessed: 2023-10-03.
- Li, H.; Qian, Y.; Jiang, Y.; Liu, Y.; and Zhou, F. 2023. A novel label-based multimodal topic model for social media analysis. *Decision Support Systems*, 164.
- Madisson, M.-L.; and Ventsel, A. 2020. *Strategic Conspiracy Narratives: A Semiotic Approach*. London: Routledge.
- Mahl, D.; Schäfer, M. S.; and Zeng, J. 2022. Conspiracy theories in online environments: An interdisciplinary literature review and agenda for future research. *New Media & Society*, 25(7): 1781–1801.
- Marcks, H. 2023. Die Kanalisation des Protests: Demokratiefreundliche Mobilisierung via Telegram. <https://machine-vs-rage.bag-gegen-hass.net/die-kanalisation-des-protests/>. Accessed: 2024-07-30.
- McSwiney, J.; Vaughan, M.; Heft, A.; and Hoffmann, M. 2021. Sharing the hate? Memes and transnationality in the far right’s digital visual culture. *Information, Communication & Society*, 24(16): 2502–2521.
- Meylakhs, P.; Rykov, Y.; Koltsova, O.; and Koltsov, S. 2014. An AIDS-Denialist Online Community on a Russian Social Networking Service: Patterns of Interactions With Newcomers and Rhetorical Strategies of Persuasion. *Journal of Medical Internet Research*, 16(11).
- Ng, L. H. X.; Kloo, I.; and Carley, K. M. 2024. An Exploratory Analysis of COVID Bot vs Human Disinformation Dissemination stemming from the Disinformation Dozen on Telegram. *J Comput Soc Sc*, 7: 695–720.
- Peng, Y.; Lock, I.; and Ali Salah, A. 2023. Automated Visual Analysis for the Study of Social Media Effects: Opportunities, Approaches, and Challenges. *Communication Methods and Measures*, 1–23.
- Qu, Y.; He, X.; Pierson, S.; Backes, M.; Zhang, Y.; and Zannettou, S. 2023. On the Evolution of (Hateful) Memes by Means of Multimodal Contrastive Learning. arXiv:2212.06573.
- Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020.
- Rathje, J. 2021. “Reichsbürger” und Souveränismus. <https://www.bpb.de/shop/zeitschriften/apuz/verschwoerungstheorien-2021/339286/reichsbuerger-und-souveraenismus/>. Accessed: 2024-12-05.
- Richardson, J. E.; and Wodak, R. 2009. The Impact of Visual Racism: Visual Arguments in Political Leaflets of Austrian and British Far-right Parties. *Controversia*, 6(2): 45–77.
- Sosa, J.; and Sharoff, S. 2022. Multimodal Pipeline for Collection of Misinformation Data from Telegram. arXiv:2204.12690.
- Sunstein, C. R.; and Vermeule, A. 2009. Conspiracy Theories: Causes and Cures. *Journal of Political Philosophy*, 17(2): 202–227.
- Theisen, W.; Brogan, J.; Thomas, P. B.; Moreira, D.; Phoa, P.; Weninger, T.; and Scheirer, W. 2021. Automatic Discovery of Political Meme Genres with Diverse Appearances. In *Proceedings of ICWSM 2021*, 714–726.
- Theisen, W.; and Scheirer, W. 2024a. C-CLIP: Contrastive Image-Text Encoders to Close the Descriptive-Commentative Gap. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 7241–7250.
- Theisen, W.; and Scheirer, W. 2024b. N-Modal Contrastive Losses with Applications to Social Media Data in Trimodal Space. arXiv:2403.12747.
- Tuters, M.; and Willaert, T. 2022. Deep state phobia: Narrative convergence in coronavirus conspiracism on Instagram. *Convergence*, 28(4): 1214–1238.
- Urman, A.; and Katz, S. 2022. What they do in the shadows: examining the far-right networks on Telegram. *Information, Communication & Society*, 25(7): 904–923.
- van Prooijen, J.-W.; and Douglas, K. M. 2017. Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies*, 10(3): 323–333.
- Vergani, M.; Martinez Arranz, A.; Scrivens, R.; and Orellana, L. 2022. Hate Speech in a Telegram Conspiracy Channel During the First Year of the COVID-19 Pandemic. *Social Media + Society*, 8(4).
- Wich, M.; Gorniak, A.; Eder, T.; Bartmann, D.; Çakici, B. E.; and Groh, G. 2021. Introducing an Abusive Language Classification Framework for Telegram to Investigate the German Hater Community. In *Proceedings of ICWSM 2022*, 1133–1144.
- Wood, M. J. 2018. Propagating and Debunking Conspiracy Theories on Twitter During the 2015–2016 Zika Virus Outbreak. *Cyberpsychology, Behavior, and Social Networking*, 21(8): 485–490.
- Ying, L.; Montgomery, J. M.; and Stewart, B. M. 2022. Topics, Concepts, and Measurement: A Crowdsourced Procedure for Validating Topics as Measures. *Political Analysis*, 30(4): 570–589.
- Zannettou, S.; Caulfield, T.; Blackburn, J.; De Cristofaro, E.; Sirivianos, M.; Stringhini, G.; and Suarez-Tangil, G. 2018. On the Origins of Memes by Means of Fringe Web Communities. In *Proceedings of IMC 2018*, 188 – 202.
- Zannettou, S.; Finkelstein, J.; Bradlyn, B.; and Blackburn, J. 2020. A Quantitative Approach to Understanding Online Antisemitism. In *Proceedings of ICWSM 2020*, 786–797.
- Zehring, M.; and Domahidi, E. 2023. German Corona Protest Mobilizers on Telegram and Their Relations to the Far Right: A Network and Topic Analysis. *Social Media + Society*, 9(1).
- Zhang, H.; and Peng, Y. 2022. Image Clustering: An Unsupervised Approach to Categorize Visual Data in Social Science Research. *Sociological Methods & Research*, 0(0): 1–54.

Paper Checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? [Yes](#)
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes](#)
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? [Yes](#)
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? [Yes, see Discussion.](#)
 - (e) Did you describe the limitations of your work? [Yes](#)
 - (f) Did you discuss any potential negative societal impacts of your work? [Yes](#)
 - (g) Did you discuss any potential misuse of your work? [Yes](#)
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? [Yes](#)
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes](#)
 2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? [NA](#)
 - (b) Have you provided justifications for all theoretical results? [NA](#)
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? [NA](#)
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? [NA](#)
 - (e) Did you address potential biases or limitations in your theoretical framework? [NA](#)
 - (f) Have you related your theoretical results to the existing literature in social science? [NA](#)
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? [NA](#)
 3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? [NA](#)
 - (b) Did you include complete proofs of all theoretical results? [NA](#)
 4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes, the code is available in a publicly available repository. The dataset will be made available under restricted access on Zenodo.](#)
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes](#)
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [NA](#)
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes](#)
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? [Yes](#)
 - (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? [NA](#)
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity...**
 - (a) If your work uses existing assets, did you cite the creators? [Yes, the channel list used for data collection, as well as all used models and libraries.](#)
 - (b) Did you mention the license of the assets? [NA](#)
 - (c) Did you include any new assets in the supplemental material or as a URL? [NA](#)
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes, see Ethical Statement.](#)
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes](#)
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? [Yes, see Data and Methods.](#)
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? [Yes, it will be provided along with the dataset following the anonymous review process.](#)
 6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity...**
 - (a) Did you include the full text of instructions given to participants and screenshots? [NA](#)
 - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? [NA](#)
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [NA](#)
 - (d) Did you discuss how data is stored, shared, and de-identified? [NA](#)

Ethical Statement

Our research uses publicly available Telegram data, which may include hateful speech. While obtaining informed consent from all users is not feasible, we adhere to ethical guide-

lines (Rivers and Lewis 2014) by presenting results in aggregate form, and removing personally identifiable information. The data set will be released under restricted access for academic purposes only. Potential negative effects include the misuse of data for spreading and automatically generating harmful content, and affecting individuals targeted by hateful speech. Our goal is to strengthen the potential positive impact of our work, namely improving our understanding of harmful online communication to aid in developing effective monitoring and countermeasures.

References

- FORCE11. 2020. The FAIR Data principles. <https://force11.org/info/the-fair-data-principles/>.
- Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.
- Rivers, C. M.; and Lewis, B. L. 2014. Ethical research standards in a world of big data. Technical Report 3:38, F1000Research.