frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 16 April 2025
DOI 10.3389/fcomm.2025.1525899

OPEN ACCESS

EDITED BY
Yi Luo,
Montclair State University, United States

REVIEWED BY
Delia Dumitrescu,
Heidelberg University, Germany
Yavuz Selim Balcıoğlu,
Doğuş University, Türkiye

*CORRESPONDENCE
Aelita Skarzauskiene
✉ aelita@mruni.eu

RECEIVED 10 November 2024
ACCEPTED 26 March 2025
PUBLISHED 16 April 2025

CITATION
Skarzauskiene A, Maciuliene M, Dirzyte A and Guleviciute G (2025) Profiling antivaccination channels in Telegram: early efforts in detecting misinformation.
Front. Commun. 10:1525899.
doi: 10.3389/fcomm.2025.1525899

Profiling antivaccination channels in Telegram: early efforts in detecting misinformation

Aelita Skarzauskiene*, Monika Maciuliene, Aiste Dirzyte and Gintare Guleviciute

Institute of Communication, Mykolas Romeris University, Vilnius, Lithuania

Introduction: Telegram's privacy-focused architecture has made it a fertile ground for the spread of misinformation, yet its closed nature poses challenges for researchers. This study addresses the methodological gap in capturing and analysing misinformation on Telegram, with a particular focus on the anti-vaccination community.

Methods: The research was conducted in three phases: (1) a structured review of literature on misinformation dissemination via Telegram, (2) development of a conceptual framework incorporating features of message creators, message content, intended targets and broader social context, and (3) application of this framework to anti-vaccination Telegram channels using latent profile analysis (LPA). A dataset comprising 7,550 messages from 151 Telegram channels was manually annotated and analysed.

Results: LPA identified distinct profiles among the channels. Malicious and non-malicious channels showed significant differences in their communication patterns, particularly in the use of crisis framing, discursive manipulation, and thematic orientation. T-tests confirmed these distinctions.

Discussion: The findings highlight Telegram's unique dynamics in misinformation spread and support the utility of the proposed framework in isolating harmful content. The study underscores the need for tailored analytical strategies for platforms with non-standard affordances and suggests that content-based profiling may assist in proactive moderation.

KEYWORDS

malicious channels, misinformation, antivaccination movement, telegram platform, latent profile analysis (LAT)

1 Introduction

Telegram is a cloud-based, cross-platform instant messaging service with over 700 million monthly active users globally (Ng et al., 2024). Beyond personal messaging, Telegram is used for news dissemination, political communication and organizing social movements, making it a fertile ground for misinformation (Sosa and Sharoff, 2022). Telegram is distinguished from other online social networks by its enhanced encryption and privacy features which appeal to users prioritizing privacy and security in their communications (Terracciano, 2023). It is known as a crucial outlet for extremist groups and the spread of politically motivated misinformation (Ruffo et al., 2022; Willaert et al., 2022). However, enhanced encryption can be an obstacle both for collecting high-quality data (Liz-López et al., 2024; Ng et al., 2024) and for detecting misinformation (Ng and Taeihagh, 2021). Despite the potential for misinformation to spread, Telegram remains under-researched in the realm of misinformation

(Urman et al., 2021; Bodrunova and Nepiyuschikh, 2022) with limited evidence that the insights from misinformation research on other platforms could apply to Telegram.

Hence, the research presented here addresses two critical questions: *What are the* methodological *challenges in collecting high-quality misinformation data on Telegram and how can a tailored conceptual analysis framework for identifying malicious channels be developed and validated?* This study aims to explore these challenges and take initial steps toward developing and testing a conceptual framework that accounts for Telegram's unique features and the nature of spreading misinformation. To answer these research questions, this study was structured into three main phases (see Figure 1).

First, a structured literature review was conducted to understand the work that has already been conducted on misinformation in Telegram. Second, we built a conceptual framework focusing on four major components: features of creators/spreaders, message content, target victims and social context. This framework was designed specifically for Telegram to address the unique challenges posed by Telegram's structure and functionality due to its encryption, private and public channel structures, and lack of content moderation, which influence how misinformation spreads. Unlike traditional social media platforms, Telegram's design allows for rapid, unchecked dissemination within closed groups and large audiences alike, requiring an approach that accounts for these unique dynamics. Third, we tested the conceptual framework by examining its utility within the context of the antivaccination community on Telegram. Data was collected from 151 anti-vaccination Telegram channels, resulting in a dataset of 7,550 messages. These messages were manually annotated according to the conceptual framework. Lastly, we employed latent profile analysis to profile misinformation channels. Profiling channels based on their track record of misinformation requires careful methodology and thorough fact-checking, and enough data is needed from each channel to make a reasonable assessment: channel metadata, textual data, social media posts, and contextual data. If done transparently, profiling channels for misinformation can be a useful tool to help audiences gauge reliability, and it can also assist platforms or researchers in understanding where and how misinformation spreads (Shen and Wu, 2024).

By focusing on the antivaccination movement, this research aims to provide valuable insights into the dynamics of malicious channels on Telegram. The World Health Organization (2019) has listed the antivaccination movement as one of the top 10 global health threats. Despite conclusive evidence that the benefits of vaccination far outweigh the risks, antivaccination misinformation continues to thrive, particularly on social media platforms where it can easily reach a broad audience (Schlette et al., 2022; Ortiz-Sánchez et al., 2020; Bode and Vraga, 2018; Chua and Banerjee, 2018). Movements against vaccination have become increasingly active online, using the COVID-19 crisis to broaden their influence (Bonnevie et al., 2021). A key strategy involves amplifying and dramatizing reports of adverse reactions to vaccines in media and public discourse (Ball, 2020; Germani and Biller-Andorno, 2021). Telegram, in particular, has emerged as a widely used platform for disseminating extreme viewpoints (Rogers, 2020). It markets itself as a fast and secure messaging service, offering strong encryption and anonymity while also enabling users to reach large audiences without content moderation or restrictions (Urman et al., 2021). Understanding malicious channels behind the antivaccination movements is crucial for developing effective strategies to combat misinformation and protect public health. Hence, this study not only addresses a significant gap in existing literature but also contributes to the broader efforts of safeguarding information integrity in the digital age.

2 Related work

To establish our study within the existing scholarly discourse, we systematically reviewed research on misinformation within Telegram, drawing from major academic databases. This review aimed to examine data collection methods, analytical approaches, available datasets, and methodological gaps in detecting misinformation on the platform. Table 1 and the subsequent section outline the systematic literature review strategy we employed.

In the Scopus database, a search query combining terms related to misinformation and data collection on Telegram yielded 29 initial articles, while the Web of Science database identified 12 articles. The initial search results were then screened by reviewing the titles, abstracts and keywords of the articles. The screening process involved evaluating the relevance of each article based on specific inclusion criteria (1) only articles published in English were included to

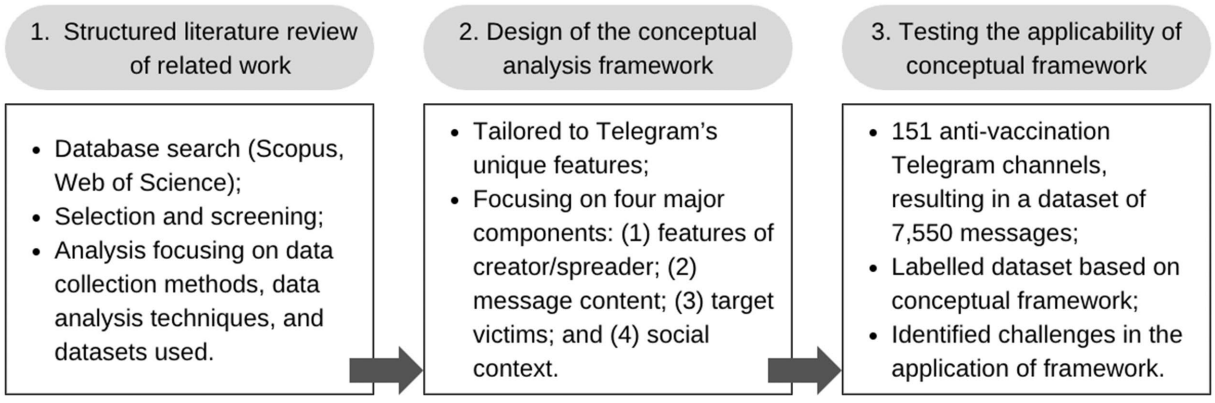

FIGURE 1
Key phases of research (source: Authors, 2024).

TABLE 1 Database search string (source: Authors, 2024).

Database	Query string/keywords	Initial	Final
Scopus	TITLE-ABS-KEY (misinformation OR disinformation OR fake news OR rumors OR rumors OR misleading) AND TITLE-ABS-KEY (dataset OR data set OR data collection OR database OR corpora) AND TITLE-ABS-KEY (telegram)	29 entries	15 entries
Web of science	TOPIC = (fake news OR misinformation OR disinformation OR rumors OR rumors OR misleading) AND (dataset OR data set OR data collection OR database OR corpora) AND (Telegram)	12 entries	

maintain accessibility; (2) studies had to involve Telegram as a primary platform for data collection or analysis; (3) peer-reviewed journal articles and conference papers were included; and (4) provide sufficient methodological detail regarding data collection, dataset creation or analytical techniques used to study misinformation on Telegram. No specific starting date was selected since research on Telegram is relatively recent. Following the screening process, 15 articles from both databases were included in the final analysis.

The literature on misinformation in Telegram research highlights several key challenges: the lack of standardized data collection methods, the fragmentation of analytical approaches, the limitations in generalizability due to linguistic and cultural constraints, and ethical concerns related to user privacy. Additionally, the evolving nature of misinformation and the platform's structural characteristics make it difficult to track and analyze false narratives. While studies employ diverse methodological techniques, their lack of integration restricts the formation of overarching conclusions. The review also identifies the need for validation and replication studies to enhance research reliability and calls for the development of cross-lingual, adaptable datasets.

A major obstacle in Telegram misinformation research is the inconsistency in data collection methods, which leads to variations in dataset quality and scope. Some researchers, such as Vanetik et al. (2023), have used web scraping and crawling techniques to compile large-scale datasets, capturing user interactions and content dissemination patterns. Others, including Claudino de Sá et al. (2023) and Ng et al. (2024), have implemented real-time monitoring systems, such as MST and BATMAN, to track misinformation as it spreads during major events. While these approaches offer valuable insights, the absence of standardized procedures makes it difficult to compare findings across studies or establish broader patterns.

The lack of methodological integration is another significant limitation. Research on Telegram misinformation employs a wide range of analytical techniques, including content analysis, sentiment analysis, natural language processing (NLP), and machine learning models. Studies such as those by Bodrunova and Nepiyuschikh (2022) and Ei and Kiat (2023) apply content and sentiment analysis to assess emotional tone and dominant narratives. Others, including Jahanbakhsh-Nagadeh et al. (2021a) and Yang et al. (2023), leverage NLP and machine learning to classify misinformation and evaluate chatbot performance. Additionally, researchers like Terracciano (2023) and Willaert et al. (2022) have introduced alternative frameworks, such as semiotics and visual network analysis, to examine misinformation dynamics. While these varied approaches provide different perspectives, their lack of integration prevents the formulation of universal conclusions.

Another challenge concerns the generalizability of findings. Most studies focus on specific linguistic or cultural contexts, limiting their applicability to broader misinformation trends. Comparative corpus analysis, such as that conducted by Maschmeyer et al. (2023) and Boumechaal and Sharoff (2024), attempts to bridge this gap by examining Telegram's anti-vaccine discourse alongside general COVID and English-language corpora. However, research remains largely fragmented, raising concerns about whether findings from one context can be applied elsewhere.

Ethical and privacy concerns add further complexity to Telegram misinformation research. The platform's encryption and privacy settings restrict access to high-quality data (Ng et al., 2024), making misinformation detection more difficult (Ng and Taeihagh, 2021). Unlike many other social media platforms, Telegram lacks content moderation, allowing misinformation to spread unchecked within echo chambers (Bodrunova and Nepiyuschikh, 2022). Additionally, forwarding mechanisms create cascading effects that amplify false narratives (Terracciano, 2023), yet the private nature of many channels makes it difficult to map misinformation flows comprehensively.

Finally, the dynamic nature of misinformation on Telegram poses ongoing challenges. Similar to other platforms, Telegram experiences constant shifts in deceptive tactics, requiring researchers to continuously adapt their methods (Ahmad et al., 2019; Aïmeur et al., 2023; Mathiesen, 2019; Panda and Levitan, 2021; Moran et al., 2023). The platform's multilingual environment further complicates misinformation detection, necessitating cross-lingual datasets and more adaptable analytical techniques. Given the relatively recent surge in Telegram misinformation research, validation and replication studies remain essential to improving reliability and ensuring methodological rigor.

Addressing these challenges requires the refinement of data collection strategies and the development of standardized, cross-lingual datasets. The growing body of research highlights the need for methodological coherence, better integration of analytical techniques, and enhanced ethical considerations to effectively study and combat misinformation on Telegram.

3 Conceptual framework for data collection

The structured literature review presented in Section 2 highlights a scarcity of conceptual frameworks explicitly designed for analyzing misinformation on Telegram. However, numerous research approaches have been documented for other social networking platforms, including Facebook (Schmidt et al., 2018), X/Twitter (Castillo et al., 2011; Horawalavithana et al., 2023) and other social media networks (Yang et al., 2012). Non-platform-specific frameworks also offer distinct perspectives for analyzing disinformation online (François, 2019; Pamment, 2020; Wardle and Derakhshan, 2017; Bontcheva et al., 2020). After thoroughly reviewing these methodologies, we have adapted Zhang and Ghorbani's (2020) approach, focusing on four

major components: (1) features of creator/spreader; (2) message content (focus on specific topics); (3) target victims (audience); and (4) social context. This multi-dimensional approach, tailored to analyze the features of datasets on Telegram, helps avoid too narrow a focus and prevents a one-dimensional interpretation of complex disinformation phenomena. By systematically categorizing and analyzing these facets of online misinformation, we aim to facilitate a deeper understanding of its spread and impact on Telegram.

3.1 Features of creators/spreaders

As part of our conceptual analysis framework, malicious channels can be defined as entities where fake news is created, published and spread (Zhou and Zafarani, 2020). They intentionally spread deceptive information to enhance their social influence, often driven by personal or financial gain (Shu et al., 2020). In examining actors on online social networks, literature frequently highlights three critical features: (1) *Creators vs. spreaders*. Creators generate and trigger the spread of disinformation, while spreaders amplify its reach. Understanding the actors behind anti-vaccine messages is crucial, as the virality of misinformation depends on these users (Karami et al., 2021). The literature consistently points out that creators are often highly motivated and capable of producing disinformation for personal or financial gain (Patel and Constantiou, 2020); (2) *Bots vs. humans*. Bots are defined as pieces of software programmed to pursue specific tasks, which can present simple and sophisticated behavior into the network, creating, sharing, and rating content and interacting with other users and bots (Moguel-Sánchez et al., 2023). In this analysis, bots were detected as they sent messages in bulk and created messages repeatedly. Both bots and humans significantly contribute to misinformation spread. Bots, or automated accounts, exploit online ecosystems to disseminate false content. Despite this, research shows that humans are still major propagators of misinformation (Schlette et al., 2022; Rogers, 2020; Vosoughi et al., 2018). Identifying bots is essential, but humans' inability to distinguish them from real accounts leads to the inadvertent spreading of misinformation (Torabi Asr and Taboada, 2019). The literature emphasizes that while bots increase the volume of misinformation, humans are crucial to its spread (Wang et al., 2019); and (3) *Individual vs. Group actors*. Misinformation is propagated by both individuals and groups. Individuals, including perceived experts and online celebrities, often blur the lines of medical authority, leveraging their influence to spread misinformation (Harris et al., 2024). Groups, such as the "disinformation dozen," play a significant role in disseminating large amounts of low-credibility content, affecting public trust, especially during pandemics (Herasimenka et al., 2023). Literature highlights that the collective actions of these groups can significantly amplify the impact of misinformation.

3.2 Target victims

In our conceptual analysis framework, the "target victims" dimension identifies the audience, individuals or groups that disinformation campaigns aim to harm (Zhang and Ghorbani, 2020). Understanding who the target victims are is essential for assessing the impact of misinformation and developing effective countermeasures. This study categorizes target victims based on patterns observed in prior research on anti-vaccine misinformation and broader disinformation strategies. The classification reflects both the frequency and strategic intent behind misinformation campaigns, with groups positioned based on their societal roles and the nature of their targeting.

1 *Activist groups* are often targeted because they advocate for social, political, or legal changes, making them a frequent focus of actors seeking to discredit or disrupt movements (Shahid et al., 2022). Their visibility and engagement in public discourse make them susceptible to misinformation designed to erode trust in their causes.
2 *Individual victims* can also be directly targeted, whether they are part of online communities or not. Smear campaigns frequently focus on single persons, aiming to damage their reputations and delegitimize their work (Lee, 2018). This includes journalists, scientists, and public figures whose influence threatens misinformation narratives.
3 *Political entities*, including political parties and politicians, are another major category of victims. Disinformation campaigns often seek to manipulate electoral outcomes or erode public trust in governance by spreading false information about political figures (Shahid et al., 2022). Given the direct impact of political misinformation on democratic processes, these actors are consistently targeted.
4 *Scientific and medical communities* face significant targeting, particularly in the context of health misinformation. During the COVID-19 pandemic, researchers and medical professionals were frequently attacked to undermine public trust in vaccines and health measures. The spread of false information about scientific institutions is often intended to delegitimize expertise and promote alternative, misleading narratives.
5 *Social identity groups*, defined by characteristics such as race, ethnicity, gender, social class, sexual orientation, or religious beliefs, are frequent targets of misinformation designed to deepen societal divisions. By exploiting existing tensions, malicious actors can manipulate public opinion and behavior, reinforcing polarization and hostility (Shahid et al., 2022; Lee, 2018).

This categorization balances specificity with comprehensiveness, ensuring that the analysis captures the primary ways misinformation operates across different societal domains. The order reflects a progression from structured organizations (activist and political groups) to individual and institutional targets, concluding with the broadest category—social identity groups—whose targeting has wide-ranging implications for societal cohesion.

3.3 Message content

As part of our conceptual analysis framework, the dimension of "message content" encompasses both linguistic and visual semiotic resources employed in disinformation campaigns to engage and mislead audiences. In this context, visual elements, such as images, typography, layout, and design choices, function as salient

communicative modes that shape how users perceive and interact with misinformation (van Leeuwen, 2005; Kress and van Leeuwen, 2006). These multimodal resources contribute to the persuasive power of disinformation by directing attention, evoking emotions, and reinforcing ideological frames.

A particularly relevant feature is attention-capturing strategies, including clickbait headlines, hashtags, and image-text juxtapositions (Lee, 2018). These elements, often strategically designed to trigger curiosity or emotional reactions, align with Dimitrova's (2011) work on media framing and visual priming, where the presentation of information significantly influences audience interpretation. For instance, the use of provocative headlines or manipulated images can frame misinformation in a way that enhances its credibility and virality.

In most cases, the actors in the anti-vaccination movement employ a multimodal content strategy that includes text, images, audio, video and interactive elements, making their messages accessible and engaging (Wawrzuta et al., 2021). Several tactics include document manipulation, which involves creating misinfographics and recontextualizing media to mislead audiences (Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy, 2023). Additionally, evidence collage compiles information from multiple sources into a single document to persuade the target audience, while distributed amplification involves campaign operators directing participants to widely disseminate materials, complicating mitigation efforts and overwhelming the information ecosystem (Krafft and Donovan, 2020).

Research indicates that the effectiveness of disinformation is less about the dissemination of technologies and more about the emotional and cognitive reactions they evoke (Martel et al., 2020; Horner et al., 2023). Cognitive biases and societal influences play significant roles in how misinformation is perceived and retained, making it essential to address these aspects in combating misinformation. Hence, the visual content includes opinions and sentiments that create polarity and influence views through trolling, memes, and viral slogans. This content leverages strong emotional appeals to make messages more sensational and memorable (Wawrzuta et al., 2021). Anti-vaccination discourse often involves narrative persuasion, where storytelling is used to engage audiences more effectively than factual arguments, reducing critical thinking and increasing susceptibility to misinformation (Covolo et al., 2017). Emotional appeals and personal stories attract attention and can lead to inaction regarding vaccination by leveraging fears and uncertainties (Guidry et al., 2015). Cultural values and personal freedom are emphasized to resonate with those skeptical of mainstream health information (Benecke and DeYoung, 2019), and anti-vaccine sites often imply false credibility or scientific authority to enhance their legitimacy (Davies et al., 2002).

Linguistic content will be analyzed through various dimensions to understand the themes and narratives used in misinformation. These dimensions include conspiracy theories, where misinformation involves elaborate conspiracy theories that undermine trust in official sources and institutions (Oliveira et al., 2022; Pierre, 2020). Political content can be politically motivated, aiming to influence public opinion or disrupt political processes (Sánchez-Castillo et al., 2023). Extremism promotes radical ideologies, while hate speech targets specific races, genders, religions, or political groups, inciting violence or discrimination (Koehler, 2023; Chen, 2024). Testimonials and captious language, such as using personal stories to elicit emotional reactions and sway opinions, often trap readers with subtly deceptive reasoning (DiResta, 2018). Emotion contagion manipulates emotions to trigger negative responses, and cloaked science uses scientific jargon to lend credibility to false claims (Herasimenka et al., 2023).

By leveraging both linguistic and visual semiotic resources, disinformation campaigns effectively manipulate audience perception and engagement. Future research should focus on cross-platform comparative analyses and the development of automated detection tools that account for the multimodal nature of disinformation. As digital media environments evolve, interdisciplinary approaches will be essential for mitigating the spread of misleading content and enhancing public resilience against information manipulation.

3.4 Social context

The social context of misinformation refers to the broader environment in which false or misleading information spreads. This encompasses the interaction between users, the technological landscape shaping information flows, and the political and societal conditions that influence how misinformation is produced, shared, and received. The social context is essential in understanding why misinformation gains traction, as it determines the speed, scale, and impact of false narratives across different communities and platforms (Castillo et al., 2011). One key aspect of social context is the interaction between users within digital communication spaces. Misinformation does not spread in isolation—it moves through networks of individuals, groups, and online communities where social relationships, trust dynamics, and engagement patterns determine its reach and persistence (Olteanu et al., 2018). Social media platforms and messaging applications structure these interactions through algorithms, content-ranking mechanisms, and platform-specific policies, all of which influence how information is shared, debated, and reinforced within different communities. Additionally, shifts in content moderation, changes in platform ownership, and evolving user norms can alter the ways misinformation circulates, making the temporal dimension of the social context a crucial factor to consider (Skafle et al., 2022).

The technological environment further shapes the social context of misinformation. The stage of technological development determines the tools available for creating, distributing, and countering misinformation. The rise of artificial intelligence, deepfake technology, and algorithmically driven content recommendation systems has transformed the landscape of misinformation, making it more sophisticated and difficult to detect (Martínez, 2023). Additionally, the increasing use of encrypted messaging apps, private forums, and decentralized platforms challenges traditional fact-checking and intervention efforts, as these environments offer reduced visibility and minimal content regulation (Cowden and Yuval-Davis, 2022).

Beyond digital infrastructure, the political and societal climate significantly influences misinformation dynamics. Certain periods, such as crises, elections, and contentious social debates, create conditions where misinformation spreads more rapidly and exerts greater influence. During crises, uncertainty and urgency drive people to seek information quickly, often before verification processes can take effect, making them more susceptible to false narratives (Clemente-Suárez et al., 2022). Election cycles amplify misinformation, as actors seeking to manipulate public opinion exploit digital platforms to shape perceptions, attack opponents, or suppress voter engagement (Seckin et al., 2024). Additionally, wedge issues—deeply divisive topics related to identity, ideology, or social policy—fuel

misinformation campaigns designed to deepen polarization and reinforce preexisting biases (Martínez, 2023).

By examining misinformation through the lens of social context, encompassing communication structures, technological development, and political conditions, researchers can better understand the factors that enable its spread and persistence. Addressing misinformation effectively requires approaches that account for how these contextual dimensions interact and evolve.

4 Testing the applicability of the conceptual framework

We assess the applicability of the conceptual framework by evaluating its utility within the context of the antivaccination community on Telegram. This section presents the methodology employed and the results obtained from this evaluation.

4.1 Methodology

4.1.1 Data collection and extraction methods

To test the applicability of our conceptual framework, we examined the antivaccination community on Telegram. We collected channels by searching Telegram with keywords such as "covid," "covid19," "vaccines," "anti-vax," "covid vaccination complications," "vaccine victims," "vaccine injuries," and "Pfizer." The data collection took place in December 2023. To ensure a comprehensive sample, we also employed a snowballing method, a well-established technique in Telegram research (Peeters and Willaert, 2022). This method assumes that if a channel forwards a message from another channel, a meaningful relationship exists between them. Additionally, Telegram channel links were retrieved from Facebook communities identified through similar keywords. Our web crawling efforts yielded an extended list of 151 channels. From these channels, we acquired all the messages, focusing on the first 50 messages from each channel. This resulted in a dataset of 7,550 messages, including original contributions in English and Lithuanian, as well as forwarded messages in English from international channels.

4.1.2 Data labeling approach

The messages were labeled according to the conceptual model discussed in Section 3, which includes the following categories: (1) *Features of spreaders/creators:* Malicious vs. Non-Malicious, Individual vs. Group, Human vs. Bot; (2) *Target victims*: activist, political, scientific/medical, minorities, undetermined; (3) *Message content*: linguistic (Conspiracy, Politics, Extremism, Hate speech, Captious language, Emotion contagious, Testimonial, Trolling, Others) and visual/multimodal strategies (Document manipulation, discourse manipulation, Evidence collage, distributed amplification, Cloaked science); and (4) *Social context*: active Crisis, Breaking News Event, Election Period, Wedge Issue. Three coders independently and manually labeled each message in the channel from March to May 2024. The subcategories were further detailed into smaller sections. The coding process was facilitated using Label Studio,[1] where annotators could tag channels and messages, and include comments and questions. Early annotation stages involved discussions among project members to refine the coding scheme. Up to 5% of messages were undefined and could not be analyzed. A codebook was maintained to document the annotations and any comments by annotators.

4.1.3 Statistical analysis

Profiling channels based on misinformation is a complex, resource-intensive process that needs not only clear definitions and reliable data but also reliable statistical analyses. To analyze the labeled messages, the statistical package JAMOVI, version 2.6.13 was applied. The sub-categories (e.g., non-malicious, activist, conspiracy, document manipulation, active crisis) were entered as variables, and the number of messages in each channel in each sub-category was entered as data. For the sub-category "malicious," a new variable was computed, summing up malicious creators, spreaders, and those that were labeled as undetermined. Based on the number of both labeled as malicious and non-malicious messages in all the channels ($n = 151$), a latent profile analysis was performed to test how many classes (profiles) can be identified in the whole sample of channels. Afterwards, differences between the identified classes of channels were analyzed concerning previously established four categories with a specific focus on subcategories.

4.1.4 Ethical considerations

All data collected were anonymized to protect the identities of individuals involved. Only public channels were accessed, with no attempts to enter private channels or chats. Messages, texts, and images shared on social media are considered part of the public domain. Data collected were publicly posted on Telegram, assuming that users expect the virtual space to be open to the public. However, channel names were replaced with codes to ensure privacy and ethical integrity.

4.2 Results

4.2.1 Preliminary analysis of data labeling approach

The preliminary analysis of the 7,550 messages from 151 anti-vaccination Telegram channels is structured around four primary dimensions: features of spreaders/creators, target victims, news content, and social context (Table A1).

Features of spreaders/creators: a significant number of messages were labeled as non-malicious (3,158), while 1,626 messages were identified as malicious. Only a few messages were classified as malicious creators (5) and none as malicious spreaders. It was challenging to differentiate between spreaders and creators based solely on message content, so the rest of the messages remained unclassified due to insufficient context. When distinguishing between individual and group actors, only six messages were identified as individual, and one as a group, with 553 remaining unclear. Similarly, identifying whether the actor was human or a bot was difficult, resulting in only four messages being labeled as human, none as bots, and 544 as unclear. The significant number of "unclear" labels highlights the difficulty in determining the nature of the spreader/creator without additional context.

Target victims were categorized as Activist, Political, Scientific/medical, Minorities, and Others. A considerable number of messages

1 https://app.heartex.com

could not be clearly labeled, indicating challenges in identifying the specific targets of disinformation campaigns. Only one message targeted activists, two targeted political entities, 14 targeted the scientific/medical community, and two fell into the "Others" category, leaving 1,611 messages undetermined. The predominance of "undetermined" labels suggests a need for more detailed content to accurately identify target victims.

Message content was divided into linguistic (text-based context) and visual/multimodal categories. Among the linguistic content, conspiracy theories (1,235), politics (105), and testimonials (141) were more frequently identified, whereas extremism (6) and hate speech (21) were less common. In the visual/multimodal category, evidence collages (1,194) and discourse manipulation (67) were prevalent, while cloaked science (17) and distributed amplification (25) were less frequently observed. While some categories like "Conspiracy" and "Evidence collage" had substantial data, others like "Extremism" and "Cloaked science" were less frequently identified, indicating variability in content types.

Social context was categorized as Active Crisis, Breaking News Event, Election Period, and Wedge Issue. Labeling in these contexts also presented challenges, with many messages lacking sufficient information to determine the context accurately. The analysis showed a higher frequency of messages related to active crises (1,117) and breaking news events (270), suggesting that disinformation campaigns often exploit these contexts. Wedge issues (218) and election periods (27) were less frequently identified but still significant.

4.2.2 Latent profile analysis

Latent Profile Analysis (LPA) was conducted to investigate unobserved heterogeneity in online content: based on the data of both labeled as malicious (1631) and non-malicious (3,158) messages in each channel, it was tested how many classes (profiles) can be identified in the whole sample of channels (n = 151). In this study, tidyLPA within JAMOVI 2.6.13 was used to explore a series of latent profile solutions. This package in R is designed to generate finite mixture models that identify unobserved subgroups (i.e., latent classes) based on continuous indicators. The models tested ranged to several classes, and it was focused on the best-fitting two-class solution according to a set of criteria. Multiple fit indices were considered, including the Akaike Information Criterion (AIC), Approximate Weight of Evidence (AWE), Bayesian Information Criterion (BIC), Classification Likelihood Criterion (CLC), Kullback Information Criterion (KIC), the sample-size-adjusted BIC (SABIC), and Integrated Completed Likelihood (ICL) to determine which of four competing two-class models offered the best balance of fit and parsimony. Table A2 displays the overall fit metrics for each model. In JAMOVI, the final selection of the model was guided by an analytic hierarchy process (AHP), a methodology that integrates multiple fit indices to recommend an optimal solution. The results indicated to selection of a model (Nr. 6) which achieves the smallest BIC (BIC = 2820.0), exhibits a comparatively high log-likelihood value (−1382.0), the lowest AIC (2786.0), the second-lowest SABIC (2785.0), and the most negative ICL (−2845.0), implying that this profiling provides a strong overall fit while retaining parsimony. Furthermore, in LPA, each participant receives a posterior probability of belonging to each latent class. For the final selected profiling model, the smallest of these average probabilities is 0.85522, and the largest is 0.93014, and these values signify that, on average, messages are assigned to their most likely class with an 85.5–93.0% probability, reflecting satisfactory classification quality. Additionally, the proportion of messages assigned to each class ranges from 0.38411 to 0.61589, providing a roughly 38–62% split in class sizes, indicating that neither class is disproportionately small, and reducing potential concerns about unstable solutions or spurious classes. Thus, this profiling model was chosen for all subsequent interpretation and reporting of parameter estimates, and this solution identified two latent classes, or profiles, each characterized by distinct estimates of the observed indicators (malicious and non-malicious messages) included in the analysis. Table 2 presents the final parameter estimates (i.e., means and variances) and associated statistics for both latent classes; in the context of LPA, the provided means are the model-estimated average values of each indicator (in this case, *Not malicious messages* and *Malicious messages*) for each of the two latent classes.

Latent Class 1 contains approximately 38.4% of the channels (specifically, 58 channels). This profile is characterized by lower scores on *Not malicious* (M = 30.40) content and higher scores on *Malicious* (M = 45.80) content. Latent Class 2 comprises roughly 61.6% of the total channels (specifically, 93 channels). In contrast to Class 1, Class 2 exhibits substantially higher *Not Malicious* content scores (M = 84.30) and lower *Malicious* scores (M = 23.40). Hence, Class 2 appears to have a much stronger inclination with *Not malicious* content and a diminished tendency toward *Malicious content* relative to the other group. Across both classes, the means of *Not malicious* and *Malicious* differ substantially in magnitude. Moreover, the standard errors (SE = 4.72–8.90) and the $p < 0.001$ suggest robust differences between the two classes. This distinction is supported by an entropy of 0.764, which falls into the range typically considered indicative of good separation in LPA, meaning that the classes are differentiated with relatively low misclassification error. Figure 2 demonstrates a line plot of latent profiles of Telegram channels based on anti-vaccination messages.

From a substantive perspective, the presence of two distinct classes suggests that the channels divide into a group that rates relatively high on *Malicious* and lower on *Not malicious* (Class 1) content, dimensions contrasted with a group showing the reverse

TABLE 2 Latent profile analysis: parameter estimates for the two-class solution for 151 channels.

Category	Parameter	Class 1		Class 2		*p*
		Mean	SE	Mean	SE	
Means	Not malicious messages	30.40	4.72	84.30	4.39	< 0.001
	Malicious messages	45.80	8.90	23.40	1.77	< 0.001
Variances	Not malicious messages	502.40	110.31	502.40	110.31	< 0.001
	Malicious messages	441.60	93.99	441.60	93.99	< 0.001

SE, Standard error.

pattern (Class 2), although LPA is an exploratory approach that does not, by itself, explain why these profiles emerge. Thus, the findings revealed the existence of two latent profiles in the data, distinguished primarily by their patterns on *Not malicious* and *Malicious* messages.

After establishing two latent classes via Latent Profile Analysis (LPA), an independent samples *T*-test was conducted to examine mean differences between Class 1 (n = 58) and Class 2 (n = 93) across a variety of variables. The *t*-test compared the actual observed values (e.g., Undetermined target, Active crisis, Breaking news events) recorded in the dataset for each channel in Class 1 vs. Class 2; the means provided below represent the average scores of messages when splitting dataset into two groups based on each channel's most likely class assignment and then computing regular descriptive statistics. Overall, several variables did yield statistically significant group differences between Class 1 and Class 2 (Table 3). Firstly, *the Undetermined target* was significantly higher in Class 1 (M = 46.33, *SD* = 28.58) exceeding Class 2 (M = 23.08, *SD* = 12.94). Another significant mean difference emerged for *Active crisis*: Class 1 (M = 39.34, *SD* = 31.64) was significantly higher than Class 2 (M = 11.51, *SD* = 11.75), also suggesting a very large effect. Although the effect size was more moderate, Class 1 (M = 3.69, *SD* = 5.87) scored lower on *Breaking news event*s than Class 2 (M = 6.41, *SD* = 7.08), (negative indicating Class 2 > Class 1). Class 1 scored lower (M = 1.22, *SD* = 2.17) than Class 2 (M = 2.63, *SD* = 3.66) on the *Type of activism: Politics*, indicating a moderate effect. However, Class 1 was substantially higher (M = 41.33, *SD* = 29.88) on *Type of activism: Conspiracy* relative to Class 2 (M = 14.05, *SD* = 11.50), this was one of the largest observed effects in the analysis. Class 1 also demonstrated higher means on *Trolling* (M = 1.64, *SD* = 4.98) than Class 2 (M = 0.16, *SD* = 0.47), as well as consistently higher scores on *Discourse manipulation* (e.g., *Cloaked science*, *Evidence collage*, *Captious language*, *Testimonial*). Notably, *Discourse manipulation: Evidence collage* had a large effect size, with Class 1 (M = 33.83, *SD* = 23.13) far exceeding Class 2 (M = 17.43, *SD* = 11.34). Class 1 also scored higher on *Identity Unclear* (M = 12.90) and *Human vs bot- Unclear* (M = 13.33) than Class 2 (M = 8.96 and 9.24, respectively). The rest of the comparisons, including those for various hate speech indicators, types of targeted groups, and other manipulation tactics, did not reach statistical significance ($ps > 0.05$) and these null results may indicate insufficient statistical power to detect smaller effects.

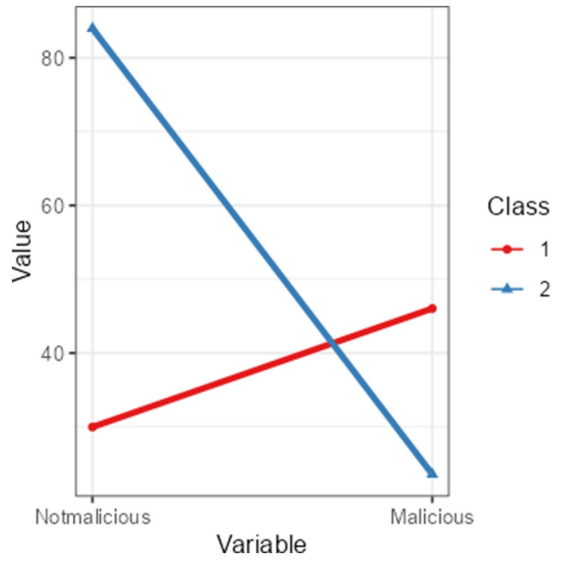

FIGURE 2
Latent profiles of Telegram channels based on anti-vaccination messages: line plot.

Overall, the *T*-test results indicate clear differences between the two latent classes in terms of misinformation characteristics and thematic emphasis. Class 1 demonstrates significantly higher scores in several key dimensions, including Undetermined target, Active crisis, Trolling, Conspiracy, Individual identity, and specific Discourse manipulation techniques such as Evidence collage, cautious language, and Testimonial strategies. These findings suggest that Class 1 is more engaged in deceptive, emotionally charged, and manipulative discourse, associated with conspiracy-driven narratives. The prominence of trolling and identity-based targeting in this class highlights a strategic use of misinformation aimed at provoking reactions, deepening divisions, or discrediting individuals and groups. In contrast, Class 2 displays lower or near-zero scores in these categories but is significantly more likely to reference Breaking news events and Political topics. This suggests that Class 2 may be more aligned with real-time information-sharing behaviors, focusing on current events rather than engaging in manipulative or deceptive tactics. However, while their references to political themes could be neutral or factual, further qualitative analysis would be necessary to determine whether this class also contributes to political misinformation or simply reacts to political discourse.

The large effect sizes observed across these variables underscore the pronounced differences between the two groups. These findings are particularly relevant because they validate the latent profile analysis (LPA)-derived classifications, confirming that the two latent profiles are not arbitrary but represent statistically meaningful distinctions in misinformation behaviors and themes. The fact that these distinctions hold across a subset of key variables suggests that each class represents a cohesive behavioral pattern, with Class 1 leaning toward manipulative, crisis-oriented, and conspiracy-driven content, while Class 2 focuses more on political and breaking news narratives.

The results of this study underscore the complexity and variability in anti-vaccination content on Telegram and highlight the need for refined analytical techniques and improved frameworks for categorizing and understanding misinformation.

5 Discussion

This study sheds light on the spread of misinformation within anti-vaccination Telegram channels, emphasizing both the utility and limitations of the applied conceptual framework. Identifying creators versus spreaders through message content alone was challenging, aligning with Leader et al. (2021) and many messages remained unclassified due to insufficient context, underscoring the need for more data that includes metadata and user behavior patterns. Similarly, identifying target victims was difficult, with many messages unclassified, as target victims are seldom explicitly mentioned (D'Ulizia et al., 2021), suggesting the necessity for additional context or integrated data sources. In terms of message content, conspiracy

TABLE 3 Independent samples *t*-test statistically significant results comparing Class 1 (n = 58) and Class 2 (n = 93).

Variable	*t* (149)	*p*	Mean diff.	SE diff.	95% CI (diff.)	Cohen's d	95% CI (d)
Features of spreaders/creators							
Individual identity	2.24	0.027	0.052	0.023	[0.006, 0.097]	0.37	[0.04, 0.70]
Identity: unclear	2.07	0.040	3.94	1.91	[0.17, 7.71]	0.35	[0.01, 0.68]
Human vs. bot - unclear	2.22	0.028	4.09	1.84	[0.45, 7.73]	0.37	[0.04, 0.70]
Number: unclear	2.23	0.027	4.15	1.86	[0.47, 7.82]	0.37	[0.04, 0.70]
Target victims							
Undetermined	6.82	< 0.001	23.25	3.41	[16.51, 29.99]	1.14	[0.79, 1.49]
Message content							
Trolling	2.84	0.005	1.48	0.52	[0.45, 2.50]	0.48	[0.14, 0.81]
Politics	−2.66	0.009	−1.41	0.53	[−2.46, −0.36]	−0.44	[−0.78, −0.11]
Conspiracy	7.92	< 0.001	27.27	3.44	[20.47, 34.08]	1.33	[0.96, 1.68]
Cloaked science	2.08	0.039	0.31	0.15	[0.02, 0.60]	0.35	[0.02, 0.68]
Evidence collage	5.81	< 0.001	16.40	2.82	[10.82, 21.97]	0.97	[0.63, 1.32]
Captious language	2.30	0.023	0.47	0.21	[0.07, 0.88]	0.38	[0.05, 0.72]
Testimonial	4.38	< 0.001	4.11	0.94	[2.26, 5.97]	0.73	[0.39, 1.07]
Discourse manipulation, overall	2.72	0.007	1.53	0.56	[0.42, 2.65]	0.45	[0.12, 0.79]
Social context							
Breaking news event	−2.45	0.016	−2.72	1.11	[−4.91, −0.52]	−0.41	[−0.74, −0.08]
Active crisis	7.69	< 0.001	27.84	3.62	[20.69, 34.99]	1.29	[0.93, 1.64]

Negative *t*-values and negative mean differences indicate higher means for Class 2 than Class 1. CI, Confidence Interval. The mean difference reflects (Class 1–Class 2). Omitted rows represent variables for which $p > 0.05$.

theories and evidence collages were prevalent, reflecting Wawrzuta et al. (2021), with the significant presence of emotional appeals and personal stories indicating the psychological dimensions of misinformation, which require targeted strategies to address. Misinformation frequently exploits active crises and breaking news events, leveraging high public interest and uncertainty to spread rapidly, making it crucial to understand these contexts for developing timely, context-specific interventions. Another key finding is the distinction between malicious and non-malicious misinformation channels, which became evident through latent profile analysis (LPA) and *t*-test results. These findings suggest that misinformation can have different forms—some actors actively manipulate narratives for ideological or disruptive purposes, while others participate in information-sharing with varying degrees of accuracy and intent.

Based on the statistically significant comparisons between the channels, several clear patterns emerge:

5.1 Features of creators/spreaders

Class 1, which can be described as a more malicious profile, reported higher levels of ambiguous or unclear identity of spreaders/creators (Individual Identity, Identity: Unclear, Human vs. bot: Unclear, Number: Unclear). Although modest in magnitude (d = 0.35–0.37), these findings suggest Class 1 misinformation may stem from (or emphasize) uncertain, difficult-to-trace sources. The literature review highlights that humans' inability to distinguish bots from real accounts leads to the inadvertent spreading of misinformation (Torabi Asr and Taboada, 2019; Schlette et al., 2022; Rogers, 2020; Vosoughi et al., 2018). The results highlight the importance of engagement metrics, timestamps, and interaction patterns in distinguishing between malicious and non-malicious misinformation actors. Analyzing content alone is insufficient for differentiating between systematic disinformation efforts and organic information-sharing.

5.2 Target victims

"Undetermined" target victims showed one of the largest gaps (d = 1.14), indicating Class 1 (more malicious profile) content is significantly more likely to remain vague about intended targets or victims. The results differ from previous research (Lee, 2018; Shahid et al., 2022), where clear strategic intents toward different audience groups were identified. Since the target audience is rarely explicitly mentioned, expanding the framework to include cross-platform data sources or indirect signals of targeting can improve accuracy. Malicious channels, in particular, exhibited higher engagement in identity-based misinformation, suggesting the need for better indicators of implicit targeting strategies.

5.3 Message content

Conspiracy content emerged as a major distinction (d = 1.33), as Class 1, a more malicious profile, is substantially more likely to include conspiratorial messages. Evidence collage (d = 0.97), Testimonial framing (d = 0.73), and Cloaked science (d = 0.35) likewise appear

more frequently in Class 1. Trolling (d = 0.48) and Captious language (d = 0.38) are also higher in Class 1, suggesting more confrontational or misleading rhetorical strategies in a more malicious profile. In contrast, Politics (d = −0.44) is higher in Class 2, implying that politically oriented messaging is more central to Class 2 (less malicious) than Class 1. The results are in alignment with other studies, which claim that anti-vaccination movements employ multimodal content strategies and document manipulation (Wawrzuta et al., 2021; Krafft and Donovan, 2020; Martel et al., 2020; Horner et al., 2023).

5.4 Social context

Active crisis (d = 1.29) is strongly elevated in Class 1, consistent with the large effect sizes seen for conspiracy-related messages (in alignment with the results of Clemente-Suárez et al., 2022; Seckin et al., 2024). However, Breaking news events (d = −0.41) are more characteristic of Class 2, indicating that Class 2 (less malicious) communications are likelier to connect to immediate, unfolding events. Less malicious Class 2 content, meanwhile, tends more toward political discussion and references to breaking news. These findings suggest two qualitatively distinct styles or "profiles" of misinformation/disinformation activity. Given that misinformation spreads differently depending on the social and political climate, the framework should incorporate real-time contextual factors, such as ongoing crises, election cycles, and wedge issues. This would allow for dynamic misinformation tracking, that accounts for how malicious actors exploit high-interest events to amplify false narratives. Lastly, since malicious channels exhibited significantly different engagement strategies compared to non-malicious ones, refining the framework by incorporating behavioral markers of coordinated activity (e.g., bot-like posting patterns, repeated message forwarding) can improve classification precision.

Our study has several limitations. Focusing solely on public Telegram channels may overlook significant misinformation activities occurring in private or semi-private groups, which often serve as key vectors for misinformation spread. Additionally, reliance on message content alone limited our ability to capture underlying intent and actor motivations, highlighting the need for metadata and engagement pattern analysis. The manual annotation process, while thorough, was also time-consuming and subject to potential bias, emphasizing the importance of automated detection tools in future research.

The dynamic nature of misinformation, especially within malicious channels, further underscores the need for continuous updates to the framework. The significant differences identified between the two classes suggest that countermeasures may need to be tailored accordingly, for example, addressing conspiracy-driven content with credibility-based interventions, while managing breaking news misinformation through real-time verification efforts. Misinformation tactics evolve rapidly, often in response to current events, fact-checking efforts, and platform policies, requiring adaptive methodologies that can detect emerging manipulation strategies in real time. Future research should compare malicious and non-malicious misinformation sources using cross-platform analyses, integrating verified medical news sources as control data to enhance reliability in identifying misinformation.

6 Conclusion

This study offers an in-depth examination of the role Telegram plays in the dissemination of misinformation, focusing on the methodological challenges and the development of a conceptual framework for profiling malicious channels. By addressing the unique features of Telegram, such as its end-to-end encryption and the diversity of its communication channels, this research highlights the complexity of tracking and analyzing misinformation on this platform. Malicious channels exhibit higher engagement with crisis-driven misinformation, conspiratorial content, trolling, vague or unidentifiable sources, and undetermined targeting. They rely on discourse manipulation techniques such as Evidence Collage, Captious Language, and Testimonial Strategies, indicating a deliberate intent to mislead. Malicious channels tend to promote deceptive, misleading, or manipulative content, often including conspiracy theories, fabricated claims, or highly emotional narratives designed to provoke strong reactions. Malicious channels often use trolling, inflammatory rhetoric, or fear-based messaging to encourage engagement.

In contrast, not-malicious channels primarily focus on Breaking News Events and Political discussions, with minimal use of deceptive framing or manipulative discourse. Not-malicious channels focus on factual reporting, discussion, or opinion-sharing without intentional distortion. Non-malicious channels have more organic dissemination patterns, with less frequent resharing of misleading content and a greater emphasis on original analysis or discussion. They are more likely to cite sources, provide context, and use measured language, even when discussing controversial topics.

The statistical findings confirm large-effect differences between the two groups, supporting the need for context-specific misinformation detection strategies. Identifying patterns in manipulation techniques, crisis exploitation, and engagement behaviors can enhance misinformation mitigation efforts, allowing for more targeted fact-checking and content moderation approaches.

Future work will focus on refining data collection methods, integrating metadata and user behavior analysis using AI, and continuously updating the framework to adapt to evolving misinformation tactics. This approach will contribute significantly to safeguarding information integrity in digital spaces.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving human data in accordance with the local legislation and institutional requirements. Written informed consent was not required, for either participation in the study or for the publication of potentially/indirectly identifying information, in accordance with the local legislation and institutional requirements. The social media data was accessed and analyzed in accordance with the platform's terms of use and all relevant institutional/national regulations.

Author contributions

AS: Writing – review & editing. MM: Writing – original draft, Writing – review & editing. AD: Writing – original draft. GG: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the CHIST-ERA grant CHIST-ERA-21-OSNEM-004 and by the Research Council of Lithuania (Grant no. S-CHIST-ERA-22-1).

Acknowledgments

We thank all the consortia partners and especially Sergio D'Antonio Maceiras from Universidad Politecnica de Madrid for helping us with the data curation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmad, A., Webb, J., Desouza, K. C., and Boorman, J. (2019). Strategically-motivated advanced persistent threat: definition, process, tactics and a disinformation model of counterattack. *Comput. Secur.* 86, 402–418. doi: 10.1016/j.cose.2019.07.001

Aïmeur, E., Amri, S., and Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Soc. Netw. Anal. Min.* 13:30. doi: 10.1007/s13278-023-01028-5

Ball, P. (2020). Anti-vaccine movement could undermine efforts to end coronavirus pandemic, researchers warn. *Nature* 581:251. doi: 10.1038/d41586-020-01423-4

Benecke, O., and DeYoung, S. E. (2019). Anti-vaccine decision-making and measles resurgence in the United States. *Global Pediatr. Health* 6:2333794X19862949. doi: 10.1177/2333794X19862949

Bode, L., and Vraga, E. K. (2018). See something, say something: correction of global health misinformation on social media. *Health Commun.* 33, 1131–1140. doi: 10.1080/10410236.2017.1331312

Bodrunova, S. S., and Nepiyuschikh, D. (2022). "Dynamics of distrust, aggression, and conspiracy thinking in the anti-vaccination discourse on Russian telegram," in *International Conference on Human-Computer Interaction* (Cham: Springer International Publishing), 468–484.

Bonnevie, E., Gallegos-Jeffrey, A., Goldbarg, J., Rosenberg, S. D., and Wartella, E. (2021). Quantifying the rise of vaccine opposition on twitter during the COVID-19 pandemic. *J. Commun. Healthc.* 14, 12–19. doi: 10.1080/17538068.2020.1858222

Bontcheva, K., Posetti, J., Teyssou, D., Meyer, T., Gregory, S., Hanot, C., et al. (2020). Balancing act: Countering digital disinformation while respecting freedom of expression. Geneva: United Nations Educational, Scientific and Cultural Organization (UNESCO).

Boumechaal, S., and Sharoff, S. (2024). Attitudes, communicative functions, and lexicogrammatical features of anti-vaccine discourse on telegram. *Applied Corpus Ling.* 4:100095. doi: 10.1016/j.acorp.2024.100095

Castillo, C., Mendoza, M., and Poblete, B. (2011). "Information credibility on twitter," in *Proceedings of the 20th International Conference on World Wide Web*, 675–684.

Chen, S. (2024). "Far-right political extremism and the radicalisation of the anti-vaccine movement in Canada" in Communicating COVID-19: Media, trust, and public engagement. eds. M. Lewis, E. Govender and K. Holland (Cham: Springer International Publishing), 303–323.

Chua, A. Y., and Banerjee, S. (2018). Intentions to trust and share online health rumours: an experiment with medical professionals. *Comput. Hum. Behav.* 87, 1–9. doi: 10.1016/j.chb.2018.05.021

Claudino de Sá, I., Galic, L., Franco, W., Gadelha, T., Monteiro, J. M., and Machado, J. (2023). BATMAN: A big data platform for misinformation monitoring. *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023)*. 1, 237–246. doi: 10.5220/0011995500003467

Clemente-Suárez, V. J., Navarro-Jiménez, E., Simón-Sanjurjo, J. A., Beltran-Velasco, A. I., Laborde-Cárdenas, C. C., Benitez-Agudelo, J. C., et al. (2022). Mis–dis information in COVID-19 health crisis: a narrative review. *Int. J. Environ. Res. Public Health* 19:5321. doi: 10.3390/ijerph19095321

Covolo, L., Ceretti, E., Passeri, C., Boletti, M., and Gelatti, U. (2017). What arguments on vaccinations run through YouTube videos in Italy? A content analysis. *Hum. Vaccin. Immunother.* 13, 1693–1699. doi: 10.1080/21645515.2017.1306159

Cowden, S., and Yuval-Davis, N. (2022). Contested narratives of the pandemic crisis: the far right, anti-vaxxers and freedom of speech. *Feminist Dissent* 6, 96–132. doi: 10.31273/fd.n6.2022.1264

Davies, P., Chapman, S., and Leask, J. (2002). Antivaccination activists on the world wide web. *Arch. Dis. Child.* 87, 22–25. doi: 10.1136/adc.87.1.22

Dimitrova, D. V. (2011). "Framing of political news in mass and online media" in Communication in U.S. elections: New agendas. ed. R. P. Hart (London: Routledge), 31–47.

DiResta, R. (2018). Of virality and viruses: the anti-vaccine movement and social media. *NAPSNet Special Reports*. Retrieved from https://nautilus.org/napsnet/napsnet-special-reports/of-virality-and-viruses-the-anti-vaccine-movement-and-social-media/

D'Ulizia, A., Caschera, M. C., Ferri, F., and Grifoni, P. (2021). Fake news detection: A survey of evaluation datasets. *PeerJ Comput. Sci.* 7:e518. doi: 10.7717/peerj-cs.518

Ei, C. H., and Kiat, C. Y. (2023). "Understanding the nature of misinformation on publicly accessible messaging platforms: the case of Ivermectin in Singapore" in Mobile communication and online falsehoods in Asia: Trends, impact and practice. ed. C. Soon (Dordrecht: Springer Netherlands), 149–172.

François, C. (2019). Actors, behaviors, content: a disinformation ABC. Cambridge, MA: Graphika and Berkman Klein Center for Internet & society at Harvard University.

Germani, F., and Biller-Andorno, N. (2021). The anti-vaccination infodemic on social media: a behavioural analysis. *PLoS One* 16:e0247642. doi: 10.1371/journal.pone.0247642

Guidry, J. P., Carlyle, K., Messner, M., and Jin, Y. (2015). On pins and needles: how vaccines are portrayed on Pinterest. *Vaccine* 33, 5051–5056. doi: 10.1016/j.vaccine.2015.08.064

Harris, M. J., Murtfeldt, R., Wang, S., Mordecai, E. A., and West, J. D. (2024). Perceived experts are prevalent and influential within an antivaccine community on Twitter. *PNAS Nexus* 3:pgae007. doi: 10.1093/pnasnexus/pgae007

Harvard Kennedy School Shorenstein Center on Media, Politics and Public Policy. (2023). Media Manipulation Casebook. Available online at: https://mediamanipulation.org/ (accessed July 28, 2024).

Herasimenka, A., Au, Y., George, A., Joynes-Burgess, K., Knuutila, A., Bright, J., et al. (2023). The political economy of digital profiteering: communication resource mobilization by anti-vaccination actors. *J. Commun.* 73, 126–137. doi: 10.1093/joc/jqac043

Horawalavithana, S., De Silva, R., Weerasekara, N., Kin Wai, N. G., Nabeel, M., Abayaratna, B., et al. (2023). Vaccination trials on hold: malicious and low credibility content on twitter during the AstraZeneca COVID-19 vaccine development. *Comput. Mathematical Org. Theor.* 29, 448–469. doi: 10.1007/s10588-022-09370-3

Horner, C. G., Galletta, D., Crawford, J., and Shirsat, A. (2023). "Emotions: the unexplored fuel of fake news on social media" in Fake news on the internet. eds. A. R. Dennis, D. F. Galletta and J. Webster (London: Routledge), 147–174.

Jahanbakhsh-Nagadeh, Z., Feizi-Derakhshi, M. R., and Sharifi, A. (2021). A semi-supervised model for Persian rumour verification based on content information. *Multimed. Tools Appl.* 80, 35267–35295. doi: 10.1007/s11042-020-10077-3

Karami, M., Nazer, T. H., and Liu, H. (2021). "Profiling fake news spreaders on social media through psychological and motivational factors," in *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, 225–230. doi: 10.1145/3465336.3475097

Koehler, D. (2023). Siren calls of anti-government extremism: far-right influences on the German anti-vax ('Querdenken') protest milieu through music. *Behav. Sci. Terror. Political Aggress.* 22, 1–22. doi: 10.1080/19434472.2023.2244571

Krafft, P. M., and Donovan, J. (2020). Disinformation by design: the use of evidence collages and platform filtering in a media manipulation campaign. *Polit. Commun.* 37, 194–214. doi: 10.1080/10584609.2019.1686094

Kress, G., and van Leeuwen, T. (2006). Reading images: The grammar of visual design. *2nd* Edn. London: Routledge.

Leader, A. E., Burke-Garcia, A., Massey, P. M., and Roark, J. B. (2021). Understanding the messages and motivation of vaccine hesitant or refusing social media influencers. *Vaccine* 39, 350–356. doi: 10.1016/j.vaccine.2020.11.058

Lee, N. M. (2018). Fake news, phishing, and fraud: a call for research on digital media literacy education beyond the classroom. *Commun. Educ.* 67, 460–466. doi: 10.1080/03634523.2018.1503313

Liz-López, H., Keita, M., Taleb-Ahmed, A., Hadid, A., Huertas-Tato, J., and Camacho, D. (2024). Generation and detection of manipulated multimodal audiovisual content: advances, trends, and open challenges. *Inf. Fusion* 103:102103. doi: 10.1016/j.inffus.2023.102103

Martel, C., Pennycook, G., and Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognit. Res. Princip. Implications* 5, 1–20. doi: 10.1186/s41235-020-00252-3

Martínez, C. R. (2023). Examining the role of wedge issues in shaping voter behavior: insights from the 2020 US presidential election. *Comillas J. Int. Relations* 27, 101–121. doi: 10.14422/cir.i27.y2023.006

Maschmeyer, L., Abrahams, A., Pomerantsev, P., and Yermolenko, V. (2023). Donetsk don't tell–'hybrid war' in Ukraine and the limits of social media influence operations. *J. Inform. Tech. Polit.* 22, 1–16. doi: 10.1080/19331681.2023.2211969

Mathiesen, K. (2019). "Fake news and the limits of freedom of speech" in Media ethics, free speech and the requirements of democracy. eds. C. Fox and J. Saunders (Abingdon-On-Thames: Routledge), 161–180.

Moguel-Sánchez, R., Martínez-Palacios, C. S., Ocharán-Hernández, J. O., Limón, X., and Sánchez-García, A. J. (2023). Bots in software development: a systematic literature review and thematic analysis. *Program Comput. Soft.* 49, 712–734. doi: 10.1134/S0361768823080145

Moran, R., Nguyễn, S., and Bui, L. (2023). Sending news back home: misinformation lost in transnational social networks. *Proc. ACM Hum. Comput. Int.* 7, 1–36. doi: 10.1145/3579521

Ng, L. H. X., Kloo, I., Clark, S., 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. Sci.* 7, 1–26. doi: 10.1007/s42001-024-00253-y

Ng, L. H., and Taeihagh, A. (2021). How does fake news spread? Understanding pathways of disinformation spread through APIs. *Policy Internet* 13, 560–585. doi: 10.1002/poi3.268

Oliveira, T., Wang, Z., and Xu, J. (2022). Scientific disinformation in times of epistemic crisis: circulation of conspiracy theories on social media platforms. *Online Media Global Commun.* 1, 164–186. doi: 10.1515/omgc-2022-0005

Olteanu, A., Kıcıman, E., and Castillo, C. (2018). "A critical review of online social data: biases, methodological pitfalls, and ethical boundaries," in *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, 785–786.

Ortiz-Sánchez, E., Velando-Soriano, A., Pradas-Hernández, L., Vargas-Román, K., Gómez-Urquiza, J. L., Cañadas-De la Fuente, G. A., et al. (2020). Analysis of the anti-vaccine movement in social networks: a systematic review. *Int. J. Environ. Res. Public Health* 17:5394. doi: 10.3390/ijerph17155394

Pamment, J. (2020). The EU's role in the fight against disinformation: Developing policy interventions for the 2020s. Washington, DC: Carnegie Endowment for International Peace.

Panda, S., and Levitan, S. I. (2021). "Detecting multilingual COVID-19 misinformation on social media via contextualized embeddings," in *Proceedings of the Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, 125–129.

Patel, S., and Constantiou, I. (2020). Human agency in the propagation of false information – a conceptual framework. In ECIS 2020 Research-in-Progress Papers. Available online at: https://web.archive.org/web/20220801210803id_/https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1021&context=ecis2020_rip (accessed December 31, 2020).

Peeters, S., and Willaert, T. (2022). Telegram and digital methods: mapping networked conspiracy theories through platform affordances. *M/C J.* 25:2878. doi: 10.5204/mcj.2878

Pierre, J. M. (2020). Mistrust and misinformation: a two-component, socio-epistemic model of belief in conspiracy theories. *J. Soc. Polit. Psychol.* 8, 617–641. doi: 10.5964/jspp.v8i2.1362

Rogers, R. (2020). Deplatforming: following extreme internet celebrities to telegram and alternative social media. *Eur. J. Commun.* 35, 213–229. doi: 10.1177/0267323120922066

Ruffo, G., Semeraro, A., Giachanou, A., and Rosso, P. (2022). Studying fake news spreading, polarisation dynamics, and manipulation by bots: a tale of networks and language. *Comput Sci Rev* 47:100531. doi: 10.1016/j.cosrev.2022.100531

Sánchez-Castillo, S., López-Olano, C., and Peris-Blanes, À. (2023). Politics, public health, and disinformation: Instagram posts by European far-right parties about COVID-19 vaccines. *Rev. Lat. Comun. Soc.* 81, 209–228.

Schlette, A., van Prooijen, J. W., and Thijs, F. (2022). The online structure and development of posting behaviour in Dutch anti-vaccination groups on telegram. *New Media Soc.* 26, 4689–4710. doi: 10.1177/14614448221128475

Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., and Quattrociocchi, W. (2018). Polarization of the vaccination debate on Facebook. *Vaccine* 36, 3606–3612. doi: 10.1016/j.vaccine.2018.05.040

Seckin, O. C., Atalay, A., Otenen, E., Duygu, U., and Varol, O. (2024). Mechanisms driving online vaccine debate during the COVID-19 pandemic. *Social Media + Soc.* 10:20563051241229657. doi: 10.1177/20563051241229657

Shahid, W., Li, Y., Staples, D., Amin, G., Hakak, S., and Ghorbani, A. (2022). Are you a cyborg, bot or human?—a survey on detecting fake news spreaders. *IEEE Access* 10, 27069–27083. doi: 10.1109/ACCESS.2022.3157724

Shen, X. L., and Wu, Y. (2024). "Multidimensional information literacy and fact-checking behavior: a person-centered approach using latent profile analysis" in Wisdom, Well-Being, Win-Win. iConference 2024. Lecture Notes in Computer Science. ed. I. Serwanga (Cham: Springer).

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., and Liu, H. (2020). Fakenewsnet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data* 8, 171–188. doi: 10.1089/big.2020.0062

Skafle, I., Nordahl-Hansen, A., Quintana, D. S., Wynn, R., and Gabarron, E. (2022). Misinformation about COVID-19 vaccines on social media: rapid review. *J. Med. Internet Res.* 24:e37367. doi: 10.2196/37367

Sosa, J., and Sharoff, S. (2022). "Multimodal pipeline for collection of misinformation data from telegram," in *Proceedings of the thirteenth language resources and evaluation conference. European Language Resources Association*. 1480–1489. Available at: https://aclanthology.org/2022.lrec-1.159/.

Terracciano, B. (2023). Accessing to a "truer truth": conspiracy and figurative reasoning from COVID-19 to the Russia-Ukraine war. *Media Commun.* 11, 64–75. doi: 10.17645/mac.v11i2.6396

Torabi Asr, F., and Taboada, M. (2019). Big Data and quality data for fake news and misinformation detection. *Big Data & Society* 6, 1–14. doi: 10.1177/2053951719843310

Urman, A., Ho, J. C., and Katz, S. (2021). Analyzing protest mobilization on telegram: the case of the 2019 anti-extradition bill movement in Hong Kong. *PLoS One* 16:e0256675. doi: 10.1371/journal.pone.0256675

van Leeuwen, T. (2005). Introducing social semiotics. London: Routledge.

Vanetik, N., Litvak, M., Reviakin, E., and Tiamanova, M. (2023). "Propaganda detection in Russian telegram posts in the scope of the Russian invasion of Ukraine," in *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, 1162–1170.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science* 359, 1146–1151. doi: 10.1126/science.aap9559

Wang, Y., McKee, M., Torbica, A., and Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Soc. Sci. Med.* 240:112552. doi: 10.1016/j.socscimed.2019.112552

Wardle, C., and Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Strasbourg: Council of Europe.

Wawrzuta, D., Jaworski, M., Gotlib, J., and Panczyk, M. (2021). Characteristics of antivaccine messages on social media: systematic review. *J. Med. Internet Res.* 23:e24564. doi: 10.2196/24564

Willaert, T., Peeters, S., Seijbel, J., and Van Raemdonck, N. (2022). Disinformation networks: a quali-quantitative investigation of antagonistic Dutch-speaking telegram channels. *First Monday*. doi: 10.5210/fm.v27i5.12533

World Health Organization (2019). Ten threats to global health in 2019. Geneva: WHO.

Yang, F., Liu, Y., Yu, X., and Yang, M. (2012). "Automatic detection of rumours on Sina Weibo," in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, 1–7.

Yang, L. W. Y., Ng, W. Y., Lei, X., Tan, S. C. Y., Wang, Z., Yan, M., et al. (2023). Development and testing of a multi-lingual natural language processing-based deep learning system in 10 languages for COVID-19 pandemic crisis: a multi-center study. *Front. Public Health* 11:1063466. doi: 10.3389/fpubh.2023.1063466

Zhang, X., and Ghorbani, A. A. (2020). An overview of online fake news: characterization, detection, and discussion. *Inf. Process. Manag.* 57:102025. doi: 10.1016/j.ipm.2019.03.004

Zhou, X., and Zafarani, R. (2020). A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Comput. Surv.* 53, 1–40. doi: 10.1145/3395046

Appendix

TABLE A1 Descriptive analysis of labeled dataset (source: Authors, 2024).

Actors		Target		Content		Context	
Non-malicious	3,158	Activist	1	Non-physical		Active crisis	1,117
Malicious	1,626	Political	2	Conspiracy	1,235	Wedge issue	218
Malicious creator	5	Scientific/medical	14	Politics	105	Breaking news event	270
Malicious spreader	0	Minorities	0	Extremism	6	Election period	27
Individual	6	Others	2	Hate speech	21		
Group	1	Undetermined	1,611	Testimonial	141		
Unclear	553			Captious language	36		
Human	4			Emotion contagious	30		
Bot	0			Others	93		
Unclear	544			*Physical*			
				Document manipulation	41		
				Discourse manipulation	67		
				Evidence collage	1,194		
				Distributed amplification	25		
				Cloaked science	17		

TABLE A2 Latent profile analysis of channels (n = 151): the overall fit metrics for each model.

Model	Classes	Logik	AIC	AWE	BIC	CAIC	CLC	KIC	SABIC	ICL	Entropy
1	2	−1425	2863	2939	2884	2891	2850	2873	2862	−2920	0.665
2	2	−1401	2819	2917	2846	2855	2803	2831	2818	−2874	0.724
3	2	−1422	2860	2947	2885	2893	2846	2871	2859	−2926	0.615
6	2	−1382	2786	2906	2820	2831	2766	2800	2785	−2845	0.764