




MALICIOUS CREATIVITY IN *TELEGRAM'S* (SOFTWARE) ANTI-VACCINATION ECOSYSTEM: PROFILING ACTORS AND EARLY MISINFORMATION DYNAMICS

Aelita SKARŽAUSKIENĖ¹, Monika MAČIULIENĖ^{2,3}, Gintarė GULEVIČIŪTĖ^{2,3}, Asta ZELENKAUSKAITĖ¹, Aistė DIRŽYTĖ⁴, Sergio D'ANTONIO MACEIRAS⁵

¹*Department of Entertainment Industries, Faculty of Creative Industries, Vilnius Gediminas Technical University, Trakų str. 1, LT-01132 Vilnius, Lithuania*

²*Institute of Communication, Faculty of Human and Social Studies, Mykolas Romeris University, Ateities str. 20, LT-08303 Vilnius, Lithuania*

³*Life Quality Research Laboratory, Social Innovations Laboratory Network, Mykolas Romeris University, Ateities str. 20, LT-08303 Vilnius, Lithuania*

⁴*Institute of Management and Psychology, Institute of Psychology, Mykolas Romeris University, Ateities str. 20, LT-08303 Vilnius, Lithuania*

⁵*Research Group of Natural Language Processing and Deep Learning, Department of Computer Systems, Technical University of Madrid, Ramiro de Maeztu 7, 28040 Madrid, Spain*

Article History:

- received 18 September 2024
- accepted 9 April 2025

Abstract. This study examines methodological challenges in collecting and analysing misinformation on *Telegram* (software) and develops a platform-sensitive conceptual framework for identifying malicious actors. Addressing gaps in existing research, the framework accounts for *Telegram's* distinctive features, including limited moderation, privacy affordances, and channel-based dissemination. The study combines a structured literature review with the development and empirical testing of a four-dimensional framework encompassing creators, message content, target victims, and social context. The framework is applied to the anti-vaccination ecosystem on *Telegram* using a dataset of 7550 messages collected from 151 public channels and manually annotated. The results demonstrate both the analytical value of structured content-based approaches and their limitations in attributing malicious activity without behavioural and network-level data.

Keywords: anti-vaccination movement, data collection approached, malevolent creativity, misinformation, *Telegram* (software) platform.

 Corresponding author. E-mail: aelita.skarzauskiene@vilniustech.lt

1. Introduction

Misinformation, commonly understood as false or misleading information shared intentionally or unintentionally, poses a growing challenge in contemporary society. Existing definitions emphasize different aspects of the phenomenon. Some authors (Fei Luo et al., 2024) describe misinformation as false information spread recklessly or unknowingly, often causing societal harm, while others (Guess & Lyons, 2020) define it as information resistant to factual verification. Osman (2024) further highlights its ambiguous position between fact and opinion, frequently lacking sufficient evidentiary support. Despite these differences, scholars converge on the view that misinformation distorts collective knowledge and generates serious consequences, particularly in public health and political contexts. *Telegram* provides a particularly conducive environment for misinformation due to its highly active communities,

limited content moderation, and channel-based communication structure. Its growing user base and increasing role in shaping public discourse make the platform especially relevant for misinformation research, as demonstrated during the COVID-19 pandemic. Understanding how misinformation circulates on *Telegram* is therefore essential for developing effective mitigation strategies. This study addresses the research question of how a targeted framework can be designed and tested to identify actors spreading misinformation on *Telegram*. Given the platform's distinct architecture and weak regulatory mechanisms, specialized analytical approaches are required. The research follows a three-phase design. First, a structured literature review examines existing studies on *Telegram*-based misinformation. Second, a conceptual framework is developed focusing on message creators, content characteristics, target audiences, and social context. Third, the framework is empirically tested using data from 151 anti-vaccination *Telegram* channels, comprising 7550 messages. The annotated dataset enables assessment of misinformation patterns and framework limitations, informing future methodological refinement.

2. Conceptual analysis framework

The literature review highlights the lack of *Telegram*-specific conceptual frameworks for misinformation analysis. While research on misinformation is well-established for platforms like *Facebook* (Schmidt et al., 2018), *X* (formerly known as *Twitter*) (Castillo et al., 2011; Horawalavithana et al., 2023), and other social networks (Wei Yun Yang et al., 2023), *Telegram*'s unique structure and communication model remain understudied. Existing general frameworks for disinformation analysis (François, 2019; Pamment, 2020; Wardle & Derakhshan, 2017; Bontcheva & Posetti, 2020) lack the specificity needed to address *Telegram*'s decentralized, encrypted, and networked nature. To fill this gap, we adapted Zhang and Ghorbani's (2020) framework, which focuses on four interconnected dimensions: actors (creators and spreaders), examining who produces and amplifies misinformation (e.g., malicious *versus* non-malicious sources, individual *versus* group, human *versus* bot); audience (target victims), identifying who is affected by misinformation (e.g., scientific communities, political figures, activists, minorities); content (message characteristics), analyzing what misinformation looks like, including themes such as conspiracy theories, manipulated evidence, and emotional rhetoric; social context, investigating when and why misinformation spreads, particularly during crises, elections, and breaking news events. This framework is the most suitable for *Telegram* misinformation analysis because it captures *Telegram*'s distinct communication features, which differ from mainstream social media.

2.1. Features of creators/spreaders

As part of our conceptual analysis framework, malicious actors are defined as entities that create, publish, and spread false information with the intent to deceive and manipulate public perception (Zhou & Zafarani, 2020). These actors often engage in misinformation campaigns to enhance their social influence or achieve personal, ideological, or financial gains (Shu et al., 2019). *Telegram*'s minimal content moderation policies enable these actors to operate freely,

amplifying misinformation without restrictions present on mainstream platforms (Curley et al., 2022). Malicious actors frequently employ malevolent creativity, which involves the use of innovation for deceptive and harmful purposes (D. Cropley & A. Cropley, 2008; Perchtold-Stefan et al., 2021). While creativity is typically associated with progress and positive societal contributions, it can also be strategically applied to manipulate public perception by crafting misinformation in novel and persuasive ways (Mitchell & Reiter-Palmon, 2023). The novelty of misinformation narratives makes them particularly difficult to detect and counteract, contributing to their rapid spread and persistent influence (Shi et al., 2023). Malevolent creativity is defined by its combination of general creative characteristics – such as originality and practicality – alongside its uniquely harmful intent (Harris & Reiter-Palmon, 2015). This form of deception adapts to changing circumstances, exploiting evolving crises, public anxieties, and political instability to enhance its credibility and impact. Misinformation framed in an innovative or unexpected way is harder for individuals and fact-checkers to identify quickly, allowing falsehoods to gain momentum before they can be debunked (Shi et al., 2023).

2.2. Target victims

In our conceptual analysis framework, the target victims dimension refers to individuals or groups that disinformation campaigns seek to harm (Zhang & Ghorbani, 2020). On *Telegram*, vaccine misinformation primarily targets individuals with specific psychological and ideological tendencies, such as distrust of official authorities, belief in conspiracy theories, and preference for alternative healthcare perspectives. Research indicates that susceptibility to misinformation is often linked to low scientific literacy, alongside strong pseudoscientific and conspiratorial beliefs (Bossert & Jongman-Sereno, 2024). Existing literature on anti-vaccine misinformation identifies several categories of target victims. Activist groups campaign for social, political, or legal changes, making them frequent targets for disinformation aimed at undermining their movements. Next, misinformation campaigns often target individual victims, particularly through smear campaigns designed to discredit their work or reputation (Lee, 2018). Political entities such as parties, politicians, and electoral systems are common targets, with disinformation used to influence public opinion, sway elections, or erode trust in political institutions. Important to note, that scientific and medical communities frequently face targeted attacks, especially in the realm of health misinformation, as seen during the COVID-19 pandemic. Finally, social identity groups, defined by race, ethnicity, gender, social class, sexual orientation, or religious beliefs are often targeted to fuel division and manipulate public perceptions (Lee, 2018). By identifying these distinct categories, researchers can better understand how disinformation campaigns operate and develop strategies to mitigate their harmful effects.

2.3. Message content

Misinformation related to vaccines found on *Telegram* aligns primarily with three key groups (conspiracy theories, safety doubts, and false information about vaccine effectiveness) that comprise conspiracy theories and safety doubts along with false information about vaccine effectiveness (Sing Bik Ngai et al., 2022). The combination of text, images and links

on the *Telegram* platform supports efficient anti-vaccine rhetoric distribution. People find misinformation more relatable because it uses emotional language and conversational tones that makes content sharable to *Telegram* users (Herasimenka et al., 2023). Messaging platforms distribute content which asserts health fears and suggests governmental manipulation while reporting stories about vaccine-associated injuries to generate distrust among users (D'Almonte et al., 2023; Ginossar et al., 2022; Fumiyo Tokojima Machado et al., 2020). Anti-vaccine content pretends to be credible by adopting scientific terminology and news presentation methods (Sing Bik Ngai et al., 2022). Such presentation methods appear trustworthy towards those who do not understand scientific language. The dissemination of personal stories and memes throughout *Telegram* misinformation campaigns proves effective for these groups because they foster emotional bonds between readers (Baker & Walsh, 2024). As part of our conceptual analysis framework, the dimension of message content was split into physical and non-physical elements used in disinformation campaigns to engage and mislead audiences. In this regard, physical content refers to tangible elements such as headings, images, and visual features designed to attract user attention. Clickbait and hashtags are common examples of physical content that initially capture viewers' interest (Lee, 2018). Non-physical content consists of opinions and sentiments that shape perceptions by creating polarity and influencing public views. This type of content relies on strong emotional appeals to make messages more sensational and memorable (Wawrzuta et al., 2021).

2.4. Social context

The social context dimension of our conceptual analysis framework refers to the broader environment in which misinformation spreads. Politics plus ideology together form significant conditions which drive misinformation spread particularly during movements like anti-vaccination activism. The COVID-19 pandemic witnessed vaccine misinformation spread aggressively because of combined health apprehensions and mounting political dispute (Burns, 2023, pp. 59–89). *Telegram* emerged as a vital platform because of its many global users to facilitate and distribute anti-vaccine content. Through its platform *Telegram* enabled messages to grow through connections with overall political agendas. Specifically, during the COVID-19 pandemic political disagreements caused widespread distrust toward government institutions alongside public health protocols. Through *Telegram* different political organizations at both national and international levels spread anti-vaccine messages which targeted specific ideological groups (Flesher Fominaya, 2022). Vaccine hesitancy became linked to government limitations and individual rights through these movements which developed into a political conflict zone where conspiracy theories also emerged (Bertuzzi et al., 2022). Through *Telegram* users received additional opportunities to discuss topics which resulted in expanding the breadth of misinformation within an expanding campaign of political polarization (Lenti et al., 2023). The combination of health-related misinformation with political ideology allowed vaccine misinformation to find deeper ground within groups who originally distrusted government policies.

3. Testing the applicability of the conceptual framework

3.1. Methodology

To assess the applicability of our conceptual framework, we conducted a structured data collection process targeting the anti-vaccination community on *Telegram*. Data collection took place in December, 2023 and involved identifying relevant *Telegram* channels using keyword searches such as *covid*, *covid19*, *vaccines*, *anti-vax*, *covid vaccination complications*, *vaccine victims*, *vaccine injuries*, and *Pfizer*. To ensure a more comprehensive dataset, we employed a snowballing method, a recognized technique in *Telegram* research. This approach assumes that when a channel forwards a message from another, there is a meaningful link between them, allowing us to trace interconnected communities. Additionally, *Telegram* channel links were extracted from *Facebook* groups discussing vaccine-related topics using similar keywords. The web crawling process resulted in an expanded dataset of 151 *Telegram* channels. From each channel, we extracted the first 50 messages, yielding a dataset of 7550 messages. This dataset included original messages in English and Lithuanian, as well as forwarded messages from international *Telegram* channels, reflecting the global nature of anti-vaccination discussions. The collected messages were categorized using the conceptual framework outlined in section 2. The framework classified messages across four dimensions: 1) features of spreaders/creators: malicious *versus* non-malicious, individual *versus* group, human *versus* bot; 2) target victims: activist, political, scientific/medical, minorities, others; 3) message content: non-physical (conspiracy, politics, extremism, hate speech, captious language, emotional contagion, testimonial, others) and physical (document manipulation, discourse manipulation, evidence collage, distributed amplification, cloaked science); 4) social context: active crisis, breaking news event, election period, wedge issue. Three independent coders manually annotated the messages between March and May, 2024. The coding process was conducted using *Label Studio* (app.heartex.com, 2026), a tool that facilitated tagging, commenting, and refining annotation criteria. During early annotation stages, discussions were held among project members to align coding interpretations and refine subcategories. Approximately 5% of messages remained undefined due to ambiguity or lack of context, making them unsuitable for classification. A codebook was maintained to document coding rules, annotator decisions, and emerging challenges, ensuring consistency and transparency throughout the process.

3.2. Ethical considerations

This study followed strict ethical protocols to protect user privacy and ensure responsible data handling. All collected data were fully anonymized, with no identifying information retained. Only publicly available *Telegram* channels were analyzed, and no attempts were made to access private channels or direct messages. Given that messages, texts, and images shared on public social media platforms are generally considered part of the public domain, data collection focused solely on content users had openly posted. However, to further safeguard privacy, channel names were replaced with coded identifiers, preventing the direct association of collected content with specific groups or individuals. These measures align with ethical research standards while maintaining the integrity of the study.

3.3. Descriptive analysis

The descriptive analysis of the 7550 messages from 151 anti-vaccination *Telegram* channels is structured around four primary dimensions: 1) features of spreaders/creators; 2) target victims; 3) news content; 4) and social context. The results of the coding, adjusted by averaging the numbers across the three annotators, are presented in Table 1 below.

Table 1. Descriptive analysis of labelled dataset (source: created by authors)

Actors		Target		Content		Context	
Non-malicious	3158	Activist	1	Non-physical		Active crisis	1117
Malicious	1626	Political	2	Conspiracy	1235	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	1611	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	1194		
				Distributed amplification	25		
				Cloaked science	17		

3.3.1. Features of spreaders/creators

The majority of messages were classified as non-malicious ($n = 3158$), while malicious messages accounted for 1626 instances. However, only a small subset was attributed to malicious creators ($n = 5$), and no messages were classified as malicious spreaders. The inability to distinguish between creators and spreaders based solely on message content indicates a limitation in attribution without additional metadata. When differentiating between individual and group actors, only six messages were identified as originating from individuals, one from a group, while 553 remained unclear. Similarly, classifying the actor as human or bot proved difficult, with only four messages labeled as human, none as bots, and 544 remaining unclear. The large number of messages labeled as unclear underscores the difficulty of determining the nature of misinformation actors without contextual or behavioural metadata.

3.3.2. Target victims

Messages were categorized based on their intended targets: activists, political entities, scientific/medical community, minorities, and others. However, a significant number of messages lacked explicit targeting cues, making classification challenging. Scientific/medical community was the most frequently identified target ($n = 14$). Political entities ($n = 2$) and activists ($n = 1$) were rarely targeted. Minorities were not explicitly identifiable as a distinct target category. Others ($n = 2$) captured miscellaneous targeting instances. A substantial proportion ($n = 1611$) remained undetermined, suggesting that disinformation messages often do not specify explicit targets or that additional context is required to infer targeting patterns.

3.3.3. Message content

Misinformation content was classified into non-physical and physical categories. In non-physical misinformation the conspiracy theories ($n = 1235$) were the most common, followed by testimonials ($n = 141$) and political content ($n = 105$). Extremism ($n = 6$) and hate speech ($n = 21$) were less frequently observed, indicating that explicit hostility was less prevalent in the dataset. Evidence collages ($n = 1194$) were the most commonly identified physical misinformation format, followed by discourse manipulation ($n = 67$). Cloaked science ($n = 17$) and distributed amplification ($n = 25$) appeared less frequently, suggesting that certain misleading strategies, such as selectively manipulating scientific discourse, were less prominent in the dataset. While conspiracy-related content was highly prevalent, other categories, particularly extremism, cloaked science, and distributed amplification, were less frequently detected, indicating variability in misinformation strategies.

3.3.4. Social context

Messages were categorized based on their association with active crises, breaking news events, election periods, and wedge issues. Active crises were the most frequently observed context ($n = 1117$), indicating that misinformation thrives in times of uncertainty and societal distress. Breaking news events ($n = 270$) were also a common context for misinformation propagation, aligning with the tendency of false information to spread rapidly during evolving events. Wedge issues ($n = 218$) were less frequent but notable, suggesting that misinformation is sometimes used to deepen polarization. Election periods ($n = 27$) were the least frequent, though still relevant, emphasizing that *Telegram* misinformation is not exclusively political but extends across various domains. The high proportion of messages classified under active crises and breaking news events suggests that misinformation campaigns often exploit moments of heightened public attention, where fact-checking mechanisms may be weaker and misinformation spreads rapidly. These findings underscore the challenges of misinformation classification on *Telegram* and point to the need for context-aware methods, such as behavioural analysis and metadata tracking, to improve detection accuracy.

4. Discussion and conclusions

This study analyses how anti-vaccination misinformation circulates on *Telegram* through the lens of malicious creativity, drawing on a large hand labelled dataset and a four-dimensional

analytical framework. The findings show that *Telegram* constitutes a distinct and methodologically challenging environment where assumptions derived from research on other platforms cannot be applied without adjustment. First, the analysis confirms that identifying misinformation actors on *Telegram* is difficult in practice. Most messages were coded as non-malicious and only a small share could be confidently attributed to specific creators or spreaders. Forwarding chains, missing metadata and fragmented message histories obscure initiation and amplification roles, leaving a substantial proportion of actors unclassified. This indicates that content-based approaches alone are insufficient and that future research must integrate behavioural and network level signals to distinguish creators, spreaders and passive participants more reliably. Second, target classification suggests that anti-vaccination content on *Telegram* is rarely addressed to clearly defined groups. Explicit references to scientific, medical, or political actors are uncommon, while most messages remain undetermined in terms of targets. Rather than textual microtargeting, repeated exposure appears to result from channel interconnections, meaning audience concentration emerges structurally rather than rhetorically. Third, content analysis shows a strong dominance of conspiracy narratives and evidence collages, supported by testimonials and emotionally charged storytelling. More overt forms such as hate speech or extremist rhetoric are present but marginal. The prevalence of collage like formats highlights how fragmented visuals and selective evidence facilitate circulation and encourage forwarding with minimal commentary. Fourth, misinformation activity is closely tied to periods of acute crisis and breaking news. Health related uncertainty generates high information demand, creating conditions in which malicious creativity thrives. Electoral contexts play a comparatively minor role, distinguishing *Telegram* dynamics from those observed on platforms such as *Twitter* or *Facebook*. Despite limitations related to data scope, reliance on manual annotation and the exclusion of private channels, the study demonstrates the value and limits of structured content analysis. It provides a foundation for future work combining annotation, network analysis, and machine learning to track evolving forms of malicious creativity in messaging environments.

Note

This work was supported by the CHIST-ERA grant CHIST-ERA-21-OSNEM-004 and by Research Council of Lithuania (Grant no. S-CHIST-ERA-22-1).

References

- app.heartex.com. (2026). *Human signal: Creators of label studio*. <https://app.heartex.com/user/login/>
- Baker, S. A., & Walsh, M. J. (2024). "Memes save lives": Stigma and the production of antivaccination memes during the COVID-19 pandemic. *Social Media + Society*, 10(1). <https://doi.org/10.1177/20563051231224729>
- Bertuzzi, N., Legalisse, E., Lello, E., Gobo, G., & Sena, B. (2022). Politics during and after COVID-19: Science, health and social protest. *Partecipazione e Conflitto*, 15(3), 507–529.
- Bontcheva, K., & Posetti, J. (Eds.). (2020). *Balancing act: Countering digital disinformation while respecting freedom of expression: Broadband Commission research report on "freedom of expression and addressing disinformation on the internet"*. International Telecommunication Union/United Nations Educational, Scientific and Cultural Organization.

- Bossert, A., & Jongman-Sereno, K. (2024). Misinformation beliefs, intellectual humility, and vaccine attitudes and status. *Psi Chi: Journal of Psychological Research*, 29(2), 158–167. <https://doi.org/10.24839/2325-7342.JN29.2.158>
- Burns, R. B. (2023). *The human impact of the COVID-19 pandemic: A review of international research*. Springer Nature Singapore Pte Ltd. <https://doi.org/10.1007/978-981-99-1710-5>
- Castillo, C., Mendoza, M., & Poblete, B. (2011, 28 March–1 April). Information credibility on Twitter. In S. Sadagopan & K. Ramamritham (General Chairs.), *WWW'11: Proceedings of the 20th International Conference on World Wide Web*, Hyderabad, India (pp. 675–684). Association for Computing Machinery. <https://doi.org/10.1145/1963405.1963500>
- Cropley, D., & Cropley, A. (2008). Elements of a universal aesthetic of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 2(3), 155–161. <https://doi.org/10.1037/1931-3896.2.3.155>
- Curley, C., Siapera, E., & Carthy, J. (2022). COVID-19 protesters and the far right on Telegram: Co-conspirators or accidental bedfellows? *Social Media + Society*, 8(4). <https://doi.org/10.1177/20563051221129187>
- D'Almonte, E. F., Lima Siqueira, E., & Araújo e Silva, de G. (2023). Vaccines and disinformation: A content analysis of fake news verified by debunking platforms on social networks. *Reciis – Revista Eletrônica de Comunicação, Informação y Inovação em Saúde*, 17(3), 593–615. <https://doi.org/10.29397/reciis.v17i3.3821>
- Fei Luo, Ch., Shayanfar, R., Bhambhoria, R., Dahan, S., & Zhu, X. (2024, 12–16 November). Misinformation with Legal Consequences (MisLC): A new task towards harnessing societal harm of misinformation. In *Findings of the Association for Computational Linguistics: EMNLP 2024. The 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, Florida, United States (pp. 15749–15768). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.findings-emnlp.924>
- Flesher Fominaya, C. (2022). Mobilizing during the COVID-19 pandemic: From democratic innovation to the political weaponization of disinformation. *American Behavioral Scientist*, 68(6), 829–852. <https://doi.org/10.1177/00027642221132178>
- François, C. (2019). *The transatlantic working group papers series. Actors, behaviors, content: A disinformation ABC: Highlighting three vectors of viral deception to guide industry and regulatory responses*. https://www.annenbergpublicpolicycenter.org/wp-content/uploads/ABC_Framework_TWG_Francois_Sept_2019.pdf
- Fumiyo Tokojima Machado, D., Fioravante de Siqueira, A., & Gitahy, L. (2020). Natural stings: Selling distrust about vaccines on Brazilian YouTube. *Frontiers in Communication*, 5. <https://doi.org/10.3389/fcomm.2020.577941>
- Ginossar, T., Cruickshank, I. J., Zheleva, E., Sulskis, J., & Berger-Wolf, T. (2022). Cross-platform spread: Vaccine-related content, sources, and conspiracy theories in YouTube videos shared in early Twitter COVID-19 conversations. *Human Vaccines and Immunotherapeutics*, 18(1), 1–13. <https://doi.org/10.1080/21645515.2021.2003647>
- Guess, A. M., & Lyons, B. A. (2020). Misinformation, disinformation, and online propaganda. In N. Persily & J. A. Turcker (Eds.), *SSRC Anxieties of Democracy. Social media and democracy: the state of the field, and prospects for reform* (pp. 10–33). Cambridge University Press. <https://doi.org/10.1017/9781108890960.003>
- Harris, D. J., & Reiter-Palmon, R. (2015). Fast and furious: The influence of implicit aggression, premeditation, and provoking situations on malevolent creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 9(1), 54–64. <https://doi.org/10.1037/a0038499>
- Herasimenka, A., Au, Y., George, A., Joynes-Burgess, K., Knuutila, A., Bright, J., & Howard, Ph. N. (2023). The political economy of digital profiteering: Communication resource mobilization by anti-vaccination actors. *Journal of Communication*, 73(2), 126–137. <https://doi.org/10.1093/joc/jqac043>
- Horawalavithana, S., Silva, de R., Weerasekera, N., Wai, N. G. K., Nabeel, M., Abayaratna, B., Elvitigala, Ch., Wijesekera, P., & Iamnitich, A. (2023). Vaccination trials on hold: Malicious and low credibility content on Twitter during the AstraZeneca COVID-19 vaccine development. *Computational and Mathematical Organization Theory*, 29, 448–469. <https://doi.org/10.1007/s10588-022-09370-3>
- Lee, N. M. (2018). Fake news, phishing, and fraud: A call for research on digital media literacy education beyond the classroom. *Communication Education*, 67(4), 460–466. <https://doi.org/10.1080/03634523.2018.1503313>

- Lenti, J., Mejova, Y., Kalimeri, K., Panisson, A., Paolotti, D., Tizzani, M., & Starnini, M. (2023). Global misinformation spillovers in the vaccination debate before and during the COVID-19 pandemic: Multilingual Twitter study. *JMIR Infodemiology*, 3. <https://doi.org/10.2196/44714>
- Mitchell, K. S., & Reiter-Palmon, R. (2023). Malevolent creativity: Personality, process, and the larger creativity field. In H. Kapoor & J. C. Kaufman (Eds.), *Explorations in creativity research. Creativity and morality* (pp. 47–68). Elsevier Inc. <https://doi.org/10.1016/B978-0-323-85667-6.00004-9>
- Osman, M. (2024). Public evaluations of misinformation and motives for sharing it. *Journalism and Media*, 5, 766–786. <https://doi.org/10.3390/journalmedia5020050>
- Pamment, J. (2020). *Carnegie endowment for international peace. Vol. 3: Future threats, future solutions. The EU's role in fighting disinformation: Developing policy interventions for the 2020s*. https://assets.production.carnegie.fusionary.io/static/files/Pamment_-_Regulation.pdf
- Perchtold-Stefan, C. M., Fink, A., Rominger, Ch., & Papousek, I. (2021). Creative, antagonistic, and angry? Exploring the roots of malevolent creativity with a real-world idea generation task. *Journal of Creative Behavior*, 55(3), 710–722. <https://doi.org/10.1002/jocb.484>
- Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., & Quattrocioni, W. (2018). Polarization of the vaccination debate on Facebook. *Vaccine*, 36(25), 3606–3612. <https://doi.org/10.1016/j.vaccine.2018.05.040>
- Shi, Z., Zhou, Zh., Tian, L., Zhu, Y., Liu, Ch., & Xu, L. (2023). What causes malevolent creative people to engage in malevolent behaviors? Mediating role of moral disengagement and moderating effects of conscience. *Thinking Skills and Creativity*, 49. <https://doi.org/10.1016/j.tsc.2023.101329>
- Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2019). *FakeNewsNet: A data repository with news content, social context and spatiotemporal information for studying fake news on social media*. *arXiv*. <https://arxiv.org/pdf/1809.01286>
- Sing Bik Ngai, C., Gill Singh, R., & Yao, L. (2022). Impact of COVID-19 vaccine misinformation on social media virality: Content analysis of message themes and writing strategies. *Journal of Medical Internet Research*, 24(7). <https://doi.org/10.2196/37806>
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>
- Wawrzuta, D., Jaworski, M., Gotlib, J., & Panczyk, M. (2021). Characteristics of antivaccine messages on social media: Systematic review. *Journal of Medical Internet Research*, 23(6). <https://doi.org/10.2196/24564>
- Wei Yun Yang, L., Yan Ng, W., Lei, X., Yuan Tan, Sh. Ch., Wang, Zh., Yan, M., Pargi, M. K., Zhang, X., Sujuan Lim, J., Gunasekeran, D. V., Chee Ping Tan, F., Ee Lee, Ch., Yeo, K. K., Khoon Tan, H., Sun Sien Ho, H., Wee Bor Tan, B., Wong, T. Y., Yung Chiang Kwek, K., Siow Mong Goh, R., Liu, Y., & Shu Wei Ting, D. (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. *Frontiers in Public Health*, 11. <https://doi.org/10.3389/fpubh.2023.1063466>
- Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *arXiv*. <https://doi.org/10.1145/3395046>
- Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing and Management*, 57(2). <https://doi.org/10.1016/j.ipm.2019.03.004>