SM+S
social media + society

Social Media + Society
July-September 2023: 1–6
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20563051231196878
journals.sagepub.com/home/sms
Sage

Trust and Safety on Social Media: Understanding the Impact of Anti-Social Behavior and Misinformation on Content Moderation and Platform Governance

Anatoliy Gruzd[1], Felipe Bonow Soares[2], and Philip Mai[1]

Abstract

The Special Issue on Trust and Safety on Social Media delves into two pressing and interlinked concerns: the growing prevalence of anti-social behavior and the widespread presence of misinformation within and across various social media platforms. The collection of articles featured in the issue collectively examines factors that contribute to these concerns and proposes potential strategies to mitigate their negative impact on social media users and society. The articles included in the issue are extended versions of research first presented at the 2022 International Conference on Social Media & Society (#SMSociety), organized by the Social Media Lab at Toronto Metropolitan University.

Keywords

social media, anti-social behavior, misinformation, disinformation, platform governance, content moderation, interventions

Introduction

The Special Issue on Trust and Safety on Social Media is a collection of peer-reviewed articles that examine the rise of anti-social behavior, misinformation, and other forms of problematic content within and across various social media settings, contexts, and user groups. The special issue emerged from the presentations and deliberations by interdisciplinary scholars at the 2022 International Conference on Social Media & Society, organized by the Social Media Lab at Toronto Metropolitan University. The issue explores two dangerous and interconnecting trends: the rise of anti-social behavior and the spread of misinformation online. Its aim is to help the public, policymakers, and platform operators better understand the factors contributing to the rise of these minacious trends on social media.

In recent years, anti-social behaviors, such as trolling, harassment, and bullying have surged online. For instance, the percentage of adults (above 18 years of age) in the United States who have reported being harassed online has sharply increased from 23% in 2022 to 33% in 2023 (ADL Center for Technology & Society, 2023). According to the same survey, the increase is even more pronounced among teens (13–17 years old), going from 36% to 51%. This rise in online anti-social behavior has impacted various groups of people, including journalists (e.g., Holton et al., 2023; Lewis et al., 2020), politicians (e.g., Anne et al., 2023; Tenove et al., 2023), academics (e.g., Gosse et al., 2023; Houlden et al., 2022), women (e.g., Esposito & Breeze, 2022; Kumar et al., 2021), LGBTQ+ (e.g., Mkhize et al., 2020; Strand & Svensson, 2022), ethnic minorities (e.g., Chaudhry & Gruzd, 2020; Li & Nicholson Jr., 2021), religious minorities (e.g., Ferguson et al., 2021; Wahlström et al., 2021), and other vulnerable groups. The motives for engaging in online anti-social behavior are varied, as studies have indicated. Social media users may participate in such acts for various reasons, including seeking revenge, seeking amusement, seeking social approval, displaying sadistic tendencies, or simply due to a lack of empathy (e.g., Santre, 2023; Soares et al., 2023; Vismara et al., 2022; Volkmer et al., 2023).

The diverse targets of and motives for engaging in anti-social behavior make content moderation a challenge.

[1] Toronto Metropolitan University, Canada
[2] University of the Arts London, UK

Corresponding Author:
Anatoliy Gruzd, Social Media Lab, Ted Rogers School of Management, Toronto Metropolitan University, 350 Victoria Street Toronto, ON M5B 2K3, Canada.
Email: gruzd@torontomu.ca

Platforms have been testing and building both automated and manual approaches to mitigate problems on the social web (e.g., Gibson, 2023; Horta Ribeiro et al., 2023; Papaevangelou & Smyrnaios, 2022). For example, Twitter tested (with some success) a feature that prompted users to reconsider their messages if they were potentially harmful (Katsaros et al., 2022). Other platforms, such as Reddit, rely more on community-led moderation and use jargon-free community rules to promote "healthier" conversations by enhancing civility and other pro-social behaviors (e.g., Del Valle et al., 2020; Trujillo & Cresci, 2022).

The situation surrounding the spread of misinformation online is not much different. Even though measuring the extent of misinformation on social media can be challenging due to its various forms and often ephemeral nature, researchers have extensively documented its impact, spread, and prevalence across social media platforms (e.g., Chen, Xiao & Kumar, 2023; van der Linden, 2022). This phenomenon is particularly evident in discussions of politically polarized topics, such as gun control (Williams, 2022), climate change (Falkenberg et al., 2022), abortion (Pagoto et al., 2023), vaccination (Gruzd et al., 2023), refugees (Zhen et al., 2023), and more recently, the COVID-19 pandemic (Gruzd et al., 2021). In its most dangerous form, misinformation turns into disinformation when it is deliberately used to deceive, polarize, or radicalize the population. Disinformation has been weaponized by malicious actors during events, such as the 2020 US election (e.g., Chang et al., 2021; Lee & Jones-Jang, 2022) and, more recently, Russia's full-scale invasion of Ukraine (Gruzd et al., 2022; Milmo, 2022). Previous efforts to understand the reasons for the spread of misinformation and propose mitigation strategies include identifying and flagging false claims, fake accounts, coordinated sharing of misleading links, and manipulated media (e.g., Calleja et al., 2021; Carnahan & Bergan, 2022; Gruzd et al., 2022). In addition, there have been some attempts to implement intervention strategies, such as accuracy prompts, debunking, friction, inoculation, media literacy, and self-reflection tools (e.g., Koch et al., 2023; Scharrer et al., 2022; Singh & Banga, 2022; Traberg et al., 2022; Vivion et al., 2022). Further research is necessary to fully grasp the phenomenon of misinformation and develop effective countering strategies, given the constantly evolving techniques of misinformation spreaders and changing nature of social media platforms and social norms.

Building on the previous research in these areas, the special issue proposes an integrated and multidisciplinary approach toward addressing the challenges posed by the growth of anti-social behavior and the proliferation of misinformation on social media. The articles in the issue demonstrate the complex and intertwined nature of these matters; for example, in cases where online harassment is employed as part of strategic disinformation campaigns, and vice versa. The issue incorporates a diverse range of works, including both conceptual and empirical case studies, and examines both human and algorithmic aspects of content moderation and platform governance. It also takes a critical view of user practices, community engagement, and the network effects of anti-social behavior. Below is a brief overview of the works included in the issue.

Platform Governance

The issue starts with three papers on platform governance (Haythornthwaite, 2023; Nagappa, 2023; Zuckerman & Rajendra-Nicolucci, 2023). By examining the role of platform and community governance, these studies offer potential strategies for improving online moderation practices and fostering healthier online spaces.

In "Moderation, Networks, and Anti-Social Behavior Online," Haythornthwaite discusses the challenges of moderating extreme content on major social media platforms. The author examines the questions of how to define and identify anti-social behavior online and how to effectively use automation and human review to manage offending content. The paper suggests a framework of three layers (environment, community, and crowd) and emphasizes the importance of understanding the network impact of anti-social behavior.

In "From Community Governance to Moderation and Back Again: Re-examining Pre-Web Models of Online Governance to Address Trust and Safety's Crisis of Legitimacy," Zuckerman and Rajendra-Nicolucci argue that the issues of content moderation on social media are central to democratic participation. They review early models of social media content moderation, considering whether the "free speech" and "public health" approaches to moderation might have obscured an earlier model of community-led content moderation. The authors advocate for a community moderation approach to social media, which they argue could address persistent challenges of social media moderation and provide valuable training in democratic participation.

In "Narratives of Change to Platform Governance on DTube, an Emerging Blockchain-based Video-sharing Platform," Nagappa discusses the emergence of blockchain-based social media platforms as alternatives to mainstream social media platforms. The paper focuses on the changes to one such platform called DTube and its governance structure. It highlights how user practices, rather than technology, steer platform functions.

Computational Techniques for Investigating Anti-Social Behavior and Disinformation and Misinformation

The next set of three articles are focused on methodology and present advanced computational techniques for investigating anti-social behavior and the dissemination of disinformation and misinformation on social media (Angus et al., 2023; Giglietto et al., 2023; Steinfeld, 2023).

In their paper, "Computational Communication Methods for Examining Problematic News-Sharing Practices on Facebook at Scale," Angus et al. present a novel approach for analyzing the spread of "fake news" and other problematic information on Facebook. By analyzing networks of content sharing between public pages, groups, and external sources, the proposed method can pinpoint the most prominent sources of problematic content. The result is a comprehensive understanding of the impact that misinformation and disinformation can have on societies that are interconnected through social media.

In "A Workflow to Detect, Monitor and Update Lists of Coordinated Social Media Accounts Across Time: The Case of 2022 Italian Election," Giglietto et al. propose a methodology for detecting coordinated social media accounts in the context of political elections. The authors applied their technique to uncover various instances of potentially coordinated accounts engaged in discussions about the 2022 Italian election, including politically motivated, click-driven, and religiously motivated operations.

In "How Do Users Examine Online Messages to Determine If They Are Credible? An Eye-Tracking Study of Digital Literacy, Visual Attention to Metadata and Success in Misinformation Identification," Steinfeld investigates how users assess the credibility of online information and the association between user attention to metadata, digital literacy, and the identification of misinformation. The author found that users with advanced digital literacy tend to focus more on information metadata and are better equipped at spotting online misinformation.

Complexities of Managing Online Content and Behavior across Various Platforms

The issue concludes with five case studies that offer insights into the complexities of managing online content and behavior across various platforms, contextual backgrounds, and user demographics. These case studies have been authored by B. Chen, Lukito, and Koo (2023), Salles et al. (2023), Morales (2023), Hodson and O'Meara (2023), and Musiyiwa and Jacobson (2023).

In their research article, "Comparing the #StopTheSteal Movement across Multi-platform: Differentiating Discourse on Facebook, Twitter, and Parler," Chen, Lukito, and Koo analyze the discourse around the #StopTheSteal movement on Facebook, Twitter, and Parler in the aftermath of the 2020 US Presidential election and leading up to the Capitol Riot. The authors employ Snow and Benford's Social Movement Frames typology and specifically explore the presence of violence cues. Their findings indicate that Parler, an alternate platform, was more inclined toward inciting violence through the use of aggressive language, thereby exacerbating the call to action, as opposed to Facebook or Twitter.

In "The Far-Right Smokescreen: Environmental Conspiracy and Culture Wars on Brazilian YouTube" by Salles et al., the authors examine how the conservative YouTube channel "Brasil Paralelo" used platform affordances and political alignment to gain social relevance on environmental conspiracies. The study, which used topic modeling and network analysis, discovered that far-right rhetoric opposing environmentalism is employed as a contemporary culture war weapon. This narrative often contains unfounded allegations concerning politics, gender, religion, and other ideological themes.

In the article on "Ecologies of Violence on Social Media: An Exploration of Practices, Contexts, and Grammars of Online Harm," Morales argues for the need to understand the ways that violence is performed and communicated on social media. Using a case study of young adults in Colombia, the author demonstrates the complexity of violence on social media platforms, which is often multifaceted, overlapping, and interconnected. Morales proposes a framework for understanding and addressing harmful practices as ecologies of violence, which includes examining the practices, contexts, and grammars that contribute to online harm. The author stresses the importance of recognizing and addressing this complexity to build online and offline "cultures of peace."

Hodson and O'Meara's study, "Curating Hope: The Aspirational Self and Social Engagement in Early-Onset Cancer Communities on Social Media," examines online communities and discourse among young cancer patients and caregivers on social media. The authors note that these communities offer more than just information and emotional support, they also inspire hope and confront the traditional division between online authenticity and the aspirational self often present on social media.

Finally, in their work titled "Sponsorship Disclosure in Social Media Influencer Marketing: The Algorithmic and Non-Algorithmic Barriers," Musiyiwa and Jacobson explore the obstacles that hinder compliance with sponsorship disclosure on social media. The study reveals that both algorithmic and non-algorithmic challenges exist, making it difficult to ensure proper disclosure. Such challenges include the deprioritization of disclosure by algorithms and time-consuming disclosure procedures. The researchers suggest various strategies that influencers can employ to effectively disclose their sponsored content upfront and in a prominent manner.

Closing Thoughts

Overall, the diverse perspectives and innovative approaches presented in all 11 papers contribute to a deeper understanding of the challenges and opportunities in managing social media platforms, fostering healthier online spaces, and guiding future research in this rapidly evolving field. While the papers tackle various issues, their collective message emphasizes the importance of a comprehensive strategy that incorporates automated and manual review, community governance, and user literacy to effectively combat emerging challenges in digital spaces.

As the editors of the special issue, our hope is that it will reach a broad audience of social media stakeholders, including policymakers, developers, and platform owners and users. Researchers from a range of fields, such as communication, information technology, media studies, sociology, psychology, computer science, and other fields, will benefit from the theoretical frameworks, methodologies, and findings provided in these papers. Social media platform developers and managers can use the insights gained from the issue to inform design decisions related to content moderation, community governance, and user engagement. In addition, policymakers and regulators, educators, digital literacy advocates, and the general public may use this research to develop a deeper understanding of issues related to online behavior, moderate content, misinformation, and user safety.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by Government of Canada, Social Sciences and Humanities Research Council of Canada (PIs: Gruzd, Jacobson, Hodson).

ORCID iDs

Anatoliy Gruzd https://orcid.org/0000-0003-2366-5163
Felipe Bonow Soares https://orcid.org/0000-0003-4850-9255

References

ADL Center for Technology & Society. (2023). *Online hate and harassment: The American experience 2023*. https://www.adl.org/resources/report/online-hate-and-harassment-american-experience-2023

Angus, D., Bruns, A., Hurcombe, E., Harrington, S., & Tan, X. Y. (2023). Computational communication methods for examining problematic news-sharing practices on Facebook at scale. *Social Media + Society*.

Anne, S., Gronthos, C., & Crouch, C. (2023). The harassment of parliamentarians and judicial officers: A South Australian perspective. *Psychiatry, Psychology and Law*. Advance online publication. https://doi.org/10.1080/13218719.2023.2222379

Calleja, N., AbdAllah, A., Abad, N., Ahmed, N., Albarracin, D., Altieri, E., Anoko, J. N., Arcos, R., Azlan, A. A., Bayer, J., Bechmann, A., Bezbaruah, S., Briand, S. C., Brooks, I., Bucci, L. M., Burzo, S., Czerniak, C., Domenico, M. D., Dunn, A. G., . . .Purnat, T. D. (2021). A public health research agenda for managing infodemics: Methods and results of the first WHO infodemiology conference. *JMIR Infodemiology*, *1*(1), Article e30979. https://doi.org/10.2196/30979

Carnahan, D., & Bergan, D. E. (2022). Correcting the misinformed: The effectiveness of fact-checking messages in changing false beliefs. *Political Communication*, *39*(2), 166–183. https://doi.org/10.1080/10584609.2021.1963358

Chang, H.-C. H. C., Chen, E., Zhang, M., Muric, G., & Ferrara, E. (2021). Social bots and social media manipulation in 2020: The year in review. In U. Engel, A. Quan-Haase, S. X. Liu, & L. E. Lyberg (Eds.), *Handbook of computational social science* (Vol. 1, pp. 304–323). Routledge.

Chaudhry, I., & Gruzd, A. (2020). Expressing and challenging racist discourse on Facebook: How social media weaken the "spiral of silence" theory. *Policy & Internet*, *12*(1), 88–108. https://doi.org/10.1002/poi3.197

Chen, B., Lukito, J., & Koo, G. H. (2023). Comparing the #StopTheSteal movement across multi-platform: Differentiating discourse on Facebook, Twitter, and Parler. *Social Media + Society*.

Chen, S., Xiao, L., & Kumar, A. (2023). Spread of misinformation on social media: What contributes to it and how to combat it. *Computers in Human Behavior*, *141*, Article 107643. https://doi.org/10.1016/j.chb.2022.107643

Del Valle, M. E., Gruzd, A., Kumar, P., & Gilbert, S. (2020). Learning in the Wild: Understanding networked ties in Reddit. In N. B. Dohn, P. Jandrić, T. Ryberg, & M. de Laat (Eds.), *Mobility, data and learner agency in networked learning* (pp. 51–68). Springer. https://doi.org/10.1007/978-3-030-36911-8_4

Esposito, E., & Breeze, R. (2022). Gender and politics in a digitalised world: Investigating online hostility against UK female MPs. *Discourse & Society*, *33*(3), 303–323. https://doi.org/10.1177/09579265221076608

Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, N., Larosa, F., Sas, M., Mekacher, A., Pearce, W., Zollo, F., Quattrociocchi, W., & Baronchelli, A. (2022). Growing polarization around climate change on social media. *Nature Climate Change*, *12*(12), Article 12. https://doi.org/10.1038/s41558-022-01527-x

Ferguson, J., Ecklund, E. H., & Rothschild, C. (2021). Navigating religion online: Jewish and muslim responses to social media. *Religions*, *12*(4), Article 4. https://doi.org/10.3390/rel12040258

Gibson, A. D. (2023). What teams do: Exploring volunteer content moderation team labor on Facebook. *Social Media + Society*, *9*(3), 20563051231186108. https://doi.org/10.1177/20563051231186109

Giglietto, F., Marino, G., Mincigrucci, R., & Stanziano, A. (2023). A workflow to detect, monitor and update lists of coordinated social media accounts across time: The case of 2022 Italian election. *Social Media + Society*.

Gosse, C., O'Meara, V., Hodson, J., & Veletsianos, G. (2023). Too rigid, too big, and too slow: Institutional readiness to protect and support faculty from technology facilitated violence and abuse. *Higher Education*. Advance online publication. https://doi.org/10.1007/s10734-023-01043-7

Gruzd, A., Abul-Fottouh, D., Song, M. Y., & Saiphoo, A. (2023). From Facebook to YouTube: The potential exposure to COVID-19 anti-vaccine videos on social media. *Social Media + Society*, *9*(1), 20563051221150404. https://doi.org/10.1177/20563051221150403

Gruzd, A., De Domenico, M., Sacco, P. L., & Briand, S. (2021). Studying the COVID-19 infodemic at scale. *Big Data & Society*, *8*(1), 20539517211021116. https://doi.org/10.1177/20539517211021115

Gruzd, A., Mai, P., & Soares, F. B. (2022). How coordinated link sharing behavior and partisans' narrative framing fan the spread of COVID-19 misinformation and conspiracy theories. *Social Network Analysis and Mining*, *12*(1), Article 118. https://doi.org/10.1007/s13278-022-00948-y

Haythornthwaite, C. (2023). Moderation, networks, and anti-social behavior online. *Social Media + Society*.

Hodson, J., & O'Meara, V. (2023). Curating hope: The aspirational self and social engagement in early-onset cancer communities on social media. *Social Media + Society*.

Holton, A. E., Bélair-Gagnon, V., Bossio, D., & Molyneux, L. (2023). "Not their fault, but their problem": Organizational responses to the online harassment of journalists. *Journalism Practice*, *17*(4), 859–874. https://doi.org/10.1080/17512786.2021.1946417

Horta Ribeiro, M., Cheng, J., & West, R. (2023, April). Automated content moderation increases adherence to community guidelines. In *Proceedings of the ACM web conference 2023* (pp. 2666–2676). Association for Computing Machinery. https://doi.org/10.1145/3543507.3583275

Houlden, S., Hodson, J., Veletsianos, G., Gosse, C., Lowenthal, P., Dousay, T., & Hall, N. C. (2022). Support for scholars coping with online harassment: An ecological framework. *Feminist Media Studies*, *22*(5), 1120–1138. https://doi.org/10.1080/14680777.2021.1883086

Katsaros, M., Yang, K., & Fratamico, L. (2022). Reconsidering Tweets: Intervening during Tweet creation decreases offensive content. Proceedings of the International AAAI Conference on Web and Social Media, 16(1), 477–487.

Koch, T. K., Frischlich, L., & Lermer, E. (2023). Effects of fact-checking warning labels and social endorsement cues on climate change fake news credibility and engagement on social media. *Journal of Applied Social Psychology*, *53*(6), 495–507. https://doi.org/10.1111/jasp.12959

Kumar, P., Gruzd, A., & Mai, P. (2021). Mapping out violence against women of influence on Twitter using the cyber–lifestyle routine activity theory. *American Behavioral Scientist*, *65*(5), 689–711. https://doi.org/10.1177/0002764221989777

Lee, S., & Jones-Jang, S. M. (2022). Cynical nonpartisans: The role of misinformation in political cynicism during the 2020 U.S. presidential election. *New Media & Society*. Advance online publication. https://doi.org/10.1177/14614448221116036

Lewis, S. C., Zamith, R., & Coddington, M. (2020). Online harassment and its implications for the journalist–audience relationship. *Digital Journalism*, *8*(8), 1047–1067. https://doi.org/10.1080/21670811.2020.1811743

Li, Y., & Nicholson Jr, H. L. (2021). When "model minorities" become "yellow peril"—Othering and the racialization of Asian Americans in the COVID-19 pandemic. *Sociology Compass*, *15*(2), Article e12849. https://doi.org/10.1111/soc4.12849

Mkhize, S., Nunlall, R., & Gopal, N. (2020). An examination of social media as a platform for cyber-violence against the LGBT+ population. *Agenda*, *34*(1), 23–33. https://doi.org/10.1080/10130950.2019.1704485

Morales, E. (2023). Ecologies of violence on social media: An exploration of practices, contexts, and grammars of online harm. *Social Media + Society*.

Musiyiwa, R., & Jacobson, J. (2023). Sponsorship disclosure in social media influencer marketing: The algorithmic and non-algorithmic barriers. *Social Media + Society*.

Nagappa, A. (2023). Narratives of change to platform governance on DTube, an emerging blockchain-based video-sharing platform. *Social Media + Society*.

Pagoto, S. L., Palmer, L., & Horwitz-Willis, N. (2023). The next infodemic: Abortion misinformation. *Journal of Medical Internet Research*, *25*(1), Article e42582. https://doi.org/10.2196/42582

Papaevangelou, C., & Smyrnaios, N. (2022). The case of a Facebook content moderation debacle in Greece. In S. Iordanidou, N. Jebril, & E. Takas (Eds.), *Journalism and digital content in emerging media markets* (pp. 9–26). Springer. https://doi.org/10.1007/978-3-031-04552-3_2

Salles, D., Medeiros, P., Santini, R. M., & Barros, C. E. (2023). The far-right smokescreen: Environmental conspiracy and culture wars on Brazilian YouTube. *Social Media + Society*.

Santre, S. (2023). Cyberbullying in adolescents: A literature review. *International Journal of Adolescent Medicine and Health*, *35*(1), 1–7. https://doi.org/10.1515/ijamh-2021-0133

Scharrer, L., Pape, V., & Stadtler, M. (2022). Watch out: Fake! How warning labels affect laypeople's evaluation of simplified scientific misinformation. *Discourse Processes*, *59*(8), 575–590. https://doi.org/10.1080/0163853X.2022.2096364

Singh, N., & Banga, G. (2022). Media and information literacy for developing resistance to 'infodemic': Lessons to be learnt from the binge of misinformation during COVID-19 pandemic. *Media, Culture & Society*, *44*(1), 161–171. https://doi.org/10.1177/01634437211060201

Soares, F. B., Gruzd, A., Jacobson, J., & Hodson, J. (2023). To troll or not to troll: Young adults' anti-social behaviour on social media. *PLOS ONE*, *18*(5), Article e0284374. https://doi.org/10.1371/journal.pone.0284374

Steinfeld, N. (2023). How do users examine online messages to determine if they are credible? An eye-tracking study of digital literacy, visual attention to metadata and success in misinformation identification. *Social Media + Society*.

Strand, C., & Svensson, J. (2022). Towards a situated understanding of vulnerability—An analysis of Ugandan LGBT+ exposure to hate crimes in digital spaces. *Journal of Homosexuality*. Advance online publication. https://doi.org/10.1080/00918369.2022.2077679

Tenove, C., Tworek, H., Lore, G., Buffie, J., & Deley, T. (2023). Damage control: How campaign teams interpret and respond to online incivility. *Political Communication*, *40*(3), 283–303. https://doi.org/10.1080/10584609.2022.2137743

Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science*, *700*(1), 136–151. https://doi.org/10.1177/00027162221087936

Trujillo, A., & Cresci, S. (2022). Make Reddit great again: Assessing community effects of moderation interventions on r/The_Donald. *Proceedings of the ACM on Human-computer Interaction*, 6(CSCW2), 5261–526:28. https://doi.org/10.1145/3555639

van der Linden, S. (2022). Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine*, *28*(3), Article 3. https://doi.org/10.1038/s41591-022-01713-6

Vismara, M., Girone, N., Conti, D., Nicolini, G., & Dell'Osso, B. (2022). The current status of Cyberbullying research: A short review of the literature. *Current Opinion in Behavioral Sciences*, *46*, Article 101152. https://doi.org/10.1016/j.cobeha.2022.101152

Vivion, M., Anassour Laouan Sidi, E., Betsch, C., Dionne, M., Dubé, E., Driedger, S. M., Gagnon, D., Graham, J., Greyson, D., Hamel, D., Lewandowsky, S., MacDonald, N., Malo, B., Meyer, S. B., Schmid, P., Steenbeek, A., van der Linden, S., Verger, P., Witteman, H. O., & Yesilada, M. (2022). Prebunking messaging to inoculate against COVID-19 vaccine misinformation: An effective strategy for public health. *Journal of Communication in Healthcare*, *15*(3), 232–242. https://doi.org/10.1080/17538068.2022.2044606

Volkmer, S. A., Gaube, S., Raue, M., & Lermer, E. (2023). Troll story: The dark tetrad and online trolling revisited with a glance at humor. *PLOS ONE*, *18*(3), Article e0280271. https://doi.org/10.1371/journal.pone.0280271

Wahlström, M., Törnberg, A., & Ekbrand, H. (2021). Dynamics of violent and dehumanizing rhetoric in far-right social media. *New Media & Society*, *23*(11), 3290–3311. https://doi.org/10.1177/1461444820952795

Williams, M. F. (2022). Gun control and gun rights: A conceptual framework for analyzing public policy issues in technical and professional communication. *Technical Communication Quarterly*, *31*(1), 33–43. https://doi.org/10.1080/10572252.2021.1963487

Zhen, L., Yan, B., Tang, J. L., Nan, Y., & Yang, A. (2023). Social network dynamics, bots, and community-based online misinformation spread: Lessons from anti-refugee and COVID-19 misinformation cases. *The Information Society*, *39*(1), 17–34. https://doi.org/10.1080/01972243.2022.2139031

Zuckerman, E., & Rajendra-Nicolucci, C. (2023). From community governance to moderation and back again: Re-examining pre-web models of online governance to address trust and safety's crisis of legitimacy. *Social Media + Society*.

Author Biographies

Anatoliy Gruzd (PhD, University of Illinois at Urbana-Champaign) is a Canada Research Chair, Professor, and Director of Research at the Social Media Lab at Toronto Metropolitan University. Situated at the intersection of social media research, information management, and communication, Gruzd's multidisciplinary program explores how social media and the growing availability of user data are changing the ways in which people and organizations communicate, collaborate, and disseminate information and how these changes impact the social, economic, and political norms and structures of modern society.

Felipe Bonow Soares (PhD, Federal University of Rio Grande do Sul) is a Senior Lecturer in Communications and Media—Social Analytics at the University of the Arts London (UAL) in the United Kingdom. He was formerly a Postdoctoral Fellow at the Social Media Lab at Toronto Metropolitan University in Canada where he has worked on multi-year initiatives in the area of anti-social behavior and misinformation on social media, funded by the Canadian Tri-Council agencies. Soares research interests include online discourse, political communication, social media, and disinformation.

Philip Mai, (MA, JD, Syracuse University) is a Co-Director and a Senior Researcher of the Social Media Lab at Toronto Metropolitan University in Canada. He is also a Co-Founder of the International Conference on Social Media & Society (#SMSociety). In his work at the Social Media Lab, Mai works on technology policy and data-related issues, knowledge mobilization, information diffusion, business and research partnerships, and practical application of social media analytics. His research interest is mainly in the areas of dis/misinformation campaigns, online toxicity, hate, conspiracy theories, and extremism.