



Research Note

A playbook for mapping adolescent interactions with misinformation to perceptions of online harm

Digital misinformation is rampant, and understanding how exposure to misinformation affects the perceptions and decision-making processes of adolescents is crucial. In a four-part qualitative study with 25 college students 18–19 years old, we found that participants first assess the severity of harms (e.g., emotion, trust) that misinformation can cause, and then think about the possibilities for reputation harm, discrimination harm, or safety harm for certain kinds of misinformation. Qualities of misinformation including mis-contextualization, deceptive imagery, and impersonation factor into adolescent assessments. From these qualities, we developed a playbook for understanding adolescents' perceptions of the harms caused by digital misinformation. This playbook can be used by researchers and technologists working to enhance and develop online governance standards by creating digital navigation practices to mitigate misinformation-related harm towards adolescents.

Authors: Gowri S. Swamy (1), Morgan G. Ames (1), Niloufar Salehi (1)

Affiliations: (1) School of Information, University of California, Berkeley, USA

How to cite: Swamy, G. S., Ames, M. G., & Salehi, N. (2024). A playbook for mapping adolescent interactions with misinformation to perceptions of online harm. *Harvard Kennedy School (HKS) Misinformation Review*, 5(5).

Received: April 17th, 2024. Accepted: October 4th, 2024. Published: October 30th, 2024.

Research questions

- How do adolescents navigate, experience, and negotiate trust when exposed to misinformation on social media?
- What are adolescents' perceptions of the harms propagated by misinformation?
- How do qualities of misinformation change the perception of harm and the impact it may have on adolescent information sensibility practices?

Research note summary

- We conducted a study with 25 participants aged 18–19 who regularly interact with information online and interface with a wide range of media sources to assess their information sensibility practices when interfacing with misinformation.
- Each 60-minute session involved four parts. First, we conducted brief semi-structured interviews to understand existing perceptions of trust, mistrust, and harms of misinformation. Then, we led participants through a guided artifact retrieval to understand their navigation patterns and

¹ A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.

previous interactions with misinformation. Third, participants walked through a pre-structured misinformation newsfeed (which included different types of misinformation: satire or parody, false connection, imposter content, fabricated content, false context, and manipulated content). In this section, they both assessed the level of trust they had in the information they were seeing and determined the level of harm they hypothesized the information would cause. Finally, participants engaged in a situational mapping exercise where they viewed different types of misinformation and assessed their perceptions of who the potential recipients of harm were and which stakeholders may be held accountable for the dissemination and moderation of online content.

- Our findings demonstrate that participants use a variety of cues within examples of misinformation to predict potential harm. Using a taxonomy of types of harm—emotion, trust, reputation, discrimination, safety—we created a playbook to map participants’ understandings of misinformation to these types of harm.
- This exploration between types of misinformation, severity of harmful online content, and adolescent information sensibility practices allowed us to explore participant information sensemaking processes. The framework helps understand how misinformation qualities (e.g., mis-contextualization, deceptive imagery, impersonation) may increase perceptions of harm.

Implications

We build on a growing body of related work from human-computer interaction (HCI) research which has explored adolescent interactions with misinformation and harms caused by digital information systems. This paper’s specific goal is to contribute novel information on how adolescent interactions with misinformation and their perceptions of harm can be used to further enhance online governance standards. Past research has focused on online harms, misinformation, or adolescent information sensibility practices, but not on mapping how they interact with one another in detail. Past work around young adults’ interactions with misinformation (Borah et al., 2022) and misinformation harms on social media (Tran et al., 2020) have demonstrated a growing need to consider how these experiences weave together and have social consequences. The literature has also demonstrated the need for research which considers diverse attributes of information and employs a framework-based approach to evaluating perceptions around misinformation and its implications (Scheurman et al., 2021). In a world of increasing misinformation-related media and complex adolescent digital interactions, this study aspires to inform misinformation intervention practices by designing a harm-centric playbook. The playbook considers a variety of perceived harms and maps elements of misinformation and adolescent navigation patterns in order to critically examine their digital information landscape.

Prior research has established the harmful effects that misinformation can have on audiences (Galvão, 2021; Gisondi et al., 2022). Research in the area has often focused on distrust, which may appear in the media due to politically driven misinformation disseminated through social media platforms (Jerit & Zhao, 2020; Rashkin et al., 2017). Research by Schoenebeck et al. (2021) demonstrates the ways in which social media platforms may exacerbate unique forms of trauma and advocates for a trauma-informed approach to acknowledge the agency and experiences of users. Our research questions and design emphasize the importance of trust and belief online while attempting to navigate changes in perceptions of harm based on misinformation type and qualities.

A growing body of research has also focused on better understanding the nature of harms from digital interactions, which are often situated around interpersonal harm within specific user groups (e.g., gaming communities, those with disordered eating, teenage girls; see respectively Xiao et al., 2023; Gak et al., 2022; George et al., 2019) and through specific interpersonal actions (e.g., hate speech and harassment;

Gelber & McNamara, 2016; Im et al., 2022). A number of prior studies have addressed the types of harms (e.g., Scheuerman et al., 2021) that may appear during digital navigation processes and some have also defined frameworks addressing the relationships between different forms of harm. Moreover, prior literature has shown that there are significant harms associated with misinformation as seen through research in health advice during a pandemic (Rosenberg et al., 2020) and climate change (Treen et al., 2020).

Adolescents are the fastest-growing population on social media platforms (O’Keeffe et al., 2011), and understanding the ways in which this population navigates and perceives social media interactions is crucial to assessing information sensibility practices and the creation of healthy information ecosystems. Our study engaged with this user group in order to increase understanding of adolescent information sensibility practices (Hassoun et al., 2023), trust on social media platforms (Winstone et al., 2021), and understanding of misinformation (Paciello et al., 2023).

Hassoun et al. (2023) assessed the trustworthiness of information online among members of the Gen Z population (ages 15–26 as of 2024) to understand how they seek out, assess, and interface with information online. They found that Gen Z’s dialogic and exploratory information journey informs how they handle and treat misinformation. Moreover, Gen Zers’ information and social needs are entangled: Their information journeys do not begin with a truth-seeking query, and their use of information to orient themselves socially helps inform the framework of identifying the harms of misinformation. The study by Hassoun et al. informs our work, as it builds context around how present-day adolescents navigate online environments, adds definition around the ways to address the harms posed by misinformation, and provides guidance for creating frameworks for online governance standards.

Prior work on the varied expectations of justice within younger communities (Masucci et al., 2020) highlights the importance of working with adolescents directly to understand their perceptions of harm in order to create education and moderation practices that support their growing digital needs. Research that has demonstrated Gen Zers’ sensitivity and sympathy towards injustice on social media (Popat & Tarrant, 2023) and the negative impacts of social media and other online information systems on adolescent populations further highlights the importance of understanding this group’s information sensibility and online harm navigation process. Understanding their navigation of online harms will also further inform and assist in ongoing policy and legal efforts around information younger populations interface with online (Montgomery, 2000; Palfrey, 2010).

Evidence

In this section, we present a playbook for understanding our participants’ perceptions of misinformation and the harm it produces. The playbook, summarized in Table 1, works as a tool for making sense of the complex perceptions of the harms that misinformation spreads, the qualities within misinformation that lead to these perceptions, and how severity guidelines play a role in adolescent understanding of digital content. Overall, we learned that adolescents navigated misinformation on social media by creating a mental model of the information they see and systematically considering trust online. They used their knowledge about the media platform, their past experiences, the social and political context, and visual cues in this process. We also found that their perceptions of types of harm generated by misinformation—emotion harm, trust harm, discrimination harm, reputation harm, and safety harm—are connected to misinformation qualities such as mis-contextualization, deceptive imagery, and impersonation.

Table 1. *Playbook for mapping adolescent interactions with misinformation to perceptions of harm.*

Harm Type	Adolescent Interaction	Misinformation	Question
Emotion Harm	Unfair treatment or actions, specifically towards minority populations	Content intended to fool and/or mislead	<i>How does information trigger emotion to either impact decision making or incite psychological harm?</i>
Trust Harm	Lack of reliability and credibility of information sources Mistrust which significantly influences lifestyle or advocacy choices	Headlines, captions, visuals, content not in alignment Genuine content with false context, manipulated imagery, and/or framing of individual/issue	<i>How and to what degree does information urge mistrust towards a government, institution, and/or surrounding community?</i>
Reputation Harm	Information that adversely alters perceptions around a person or other entity	Impersonation of genuine sources New content designed to deceive, manipulate	<i>Does information significantly alter perceptions of a person?</i>
Discrimination Harm	Information that adversely alters perceptions of vulnerable and/or minority communities	Impersonation of genuine sources Content intended to fool and/or mislead	<i>Does information target or intentionally mislead a minority group?</i>
Safety Harm	Risk of threat to self or identity Lack of control around privacy and digital security	Genuine content with false context, manipulated imagery, and/or framing of individual/issue New content designed to deceive, manipulate	<i>Does information pose danger to viewers' physical or other forms of wellbeing which may deter them from their present lifestyle?</i>

1) Misleading misinformation may lead to emotion-based harm

Adolescents largely regarded misinformation that contains content intended to fool and/or mislead with an increased perception of emotion-based harms. When they encountered potential misinformation, they would often start by noting whether it seemed to intentionally elicit strong emotional reactions. They saw misinformation as geared towards evoking strong and often distressing feelings, which would in turn motivate them to decide whether to disseminate content further or take an action based on the sentiments it elicited. Participants discussed how their emotional reaction caused them to skeptically take

a closer look at language, citations, and other contextual qualities to determine whether something was misinformation. Most paid close attention to the emotions that a piece of information sparked in them as a first cue to whether or not the information might be trying to manipulate their beliefs in some way. To do so, they considered what political, social, or economic motivations it had and whether it was intentionally misleading.

2) Trust changes with misaligned content and manipulations

Adolescent participants cited headlines, captions, visuals, and content that are not in alignment with one another as indicative of decreasing perceived trust in misinformation. They also stated that genuine content with false context, manipulated imagery, or framing of an individual or issue impacts their perception of trust in misinformation. We found that evaluations of trust were the cornerstone of experiences with all types of misinformation. During the curated newsfeed task, participants interacted with misinformation that was interspersed with fact-checked information and were asked to rank their level of trust. Participants were unaware of the type of misinformation they were seeing, but trust varied considerably. Interestingly, satirical misinformation was nearly identical to authentic information in how much participants were willing to trust it: 47.1% of participants stated that they were very likely or somewhat likely to interpret satirical content as untrustworthy and 46.8% for authentic content (see Figure 1). On the other hand, participants were especially non-trusting of imposter content (76.8% were likely or somewhat likely to interpret it as untrustworthy), false context (71.9%), and false connection (61.6%). Both false connection and false context misinformation involve falsifying information of a genuine source.

Misinformation and Likelihood of Distrust

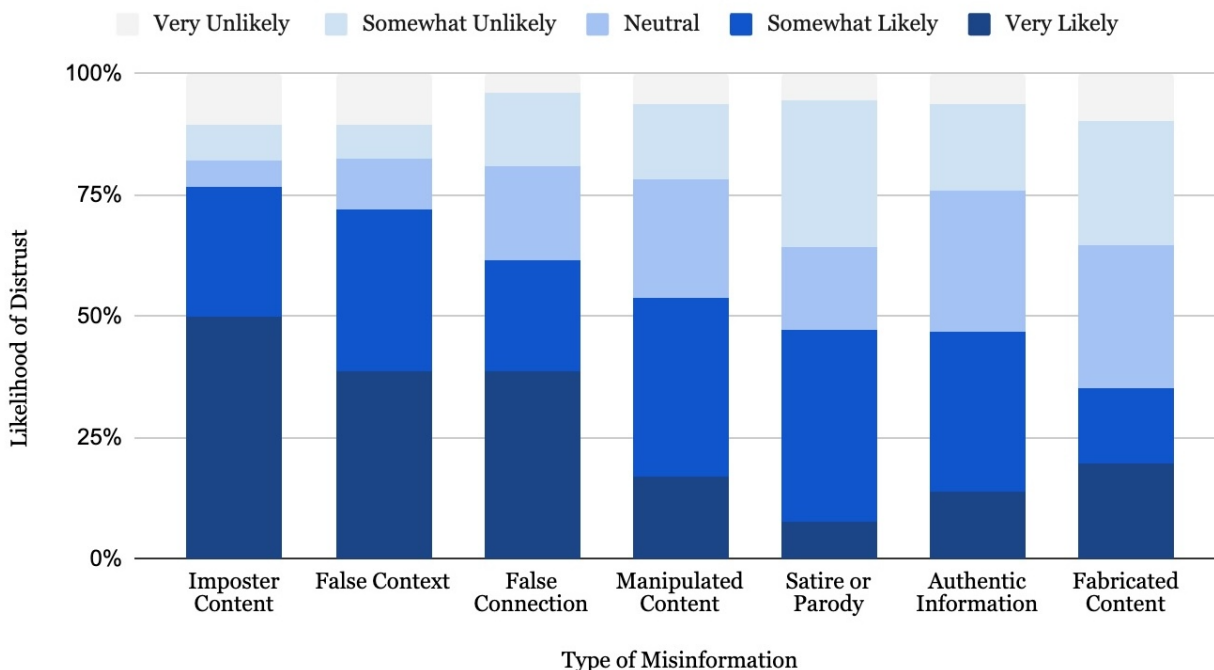


Figure 1. The reported likelihood that participants would distrust each type of misinformation.

Participants told us that mismatching headlines, captions, or visualizations, as well as incorrect contextual information were especially pernicious in how they could deceive by subverting participants' abilities to discern the trustworthiness in the content they were seeing. As the highest rate of mistrust, imposter

content impersonates genuine sources. During both qualitative interviews and newsfeed interactions, participants repeatedly emphasized the dangers of impersonating information and inaccurately crediting it to a well-known or highly reputed source. They cited the media platform where they view information, labeling and other marked associations with a source, and the inclusion of citations as the cues that they use to establish trust; thus, we have categorized violations of this trust as “trust harm.”

Misinformation categorized as manipulated content—particularly manipulated imagery—could also create trust harm: 53.7% of participants stated that they were very likely or somewhat likely to interpret it as untrustworthy. Participants tended to quickly detect this kind of misinformation’s potential to deceive; this was not as pronounced as sources that were more pernicious in their deception.

On the other end of the spectrum, only 35.3% of participants thought fabricated content was very likely or somewhat likely to be untrustworthy. The lack of an association to credible or well-reputed sources, along with the form of media that the information tends to appear on, meant that even when unreliable, this information elicited feelings of mistrust among fewer participants than even the authentic information did.

Participants’ practices and comments provided additional context for how they established trust in information. As mentioned above, assessing the source was most participants’ first step. They began by considering where the information may have come from (social media, newspaper, magazine, etc.), and many emphasized the importance of the scale of the media source, their pre-existing associations of trust, and their understanding of content moderation policies and user agency on the platforms. Participant 4 verbalized this assessment process when looking at a misinformation source in a tweet:

The information is coming from an individual Twitter user, not a credible source. The photo used in the tweet is also very ‘meme-like’ and unserious. The user is making a joke and presenting their own personal stance and interpretation of an existing policy.

Once gaining an understanding of the source, participants attempted to map the information they were looking at to past experiences. Participants discussed the credibility or “seriousness” of content as compared to other information they had recently seen. One participant reported that considering levels of bias based on prior understandings of the media channel assists them in determining their level of trust. They cited that even if information comes from CNN, which they described as a “large company and news source,” it is possible that the article is biased as CNN is “sometimes known to be.” Using past experiences and information on the platform was also a critical factor in assessment of creation, dissemination, scale, and moderation.

Participants then used a nuanced approach of their experience and understanding to test initial assessments. For instance, participant 21, had pre-existing notions about Fox News and used contextual cues (e.g., viewers, sponsorship) to add nuance to their approach of forming an assessment of reliability and trust. Overall, after thinking about emotion, participants tended to think about the trustworthiness of information in determining whether something was misinformation.

3) Deception may suggest increased risk for reputation harm

Participants stated that misinformation which involves impersonation of genuine sources and is designed to deceive or manipulate increases risk for reputation-based harms. Adolescents noted that some kinds of misinformation, such as the impersonation of real sources, could cause reputation harm. In our study, reputation harm included damage to the public opinion of a person, institution, or government; damage to credibility; and financial losses. Participants discussed how misinformation tarnishing someone’s or something’s image could even come to overshadow their achievements, expertise, or positive

contributions, leading to a lower reputation. For example, participant 19 discussed how retweets and comments on Twitter assist them in creating a more confident judgment on overall believability:

The inflammatory nature of this statement leads me to believe that the speaker is exaggerating his words for dramatic flair. Furthermore, the imbalanced number of retweets to likes makes me believe that the people in the comments either heavily disagree or they are pointing out misinformation.

In determining the severity of possible reputation harm, participants assessed the reach or scalability of the social media platform and the potential of misinformation sharing. Many also noted that the potential for information to spread broadly and rapidly also makes participants increasingly unable to alter or correct narratives. Participants perceived the severity and type of reputation harm based on the contextual information around a social media post. Specifically, participants spoke about the comments, thumbnails, likes, dislikes, shares, and other features on social media platforms that can help them better assess how information can impugn someone's reputation.

The reputation of the information source itself was also at stake. Participants stated that if they encountered and flagged enough misinformation over time on a particular platform, they were more likely to perceive the platform as untrustworthy—and as a platform's credibility sank in their eyes, so too did their beliefs in the expertise and authority of its information. Misinformation can thus increase reputational harm not just of individuals, but of platforms, institutions, and other information sources. Participants also discussed how this reputation harm can have financial impacts. In particular, they recognized that reputation harm could cause a loss of audience, especially when the audience is seeking content that they can trust.

Overall, we found that misinformation that caused reputation harm constituted an important subgenre of misinformation—one that most of our participants had encountered. Participants worried about the viral nature of reputation-related misinformation on social media in particular and the difficulty in controlling a narrative on social media. They also noted that the reputation of the platform was at stake when it propagated misinformation without adequate recourse.

4) Impersonation may lead to discrimination-based harm

Adolescents demonstrated that impersonations of genuine sources that mislead their beliefs may increase perceptions of discrimination-based harms when interfacing with misinformation. While reputation harm is individualistic, targeting one person or institution, discrimination harm targets a group. Factors such as the reinforcement of pre-existing stereotypes, exaggeration of biased perceptions, emphasis of divisions, and attempts to marginalize particular groups were important in identifying discrimination harm. Participants discussed their understandings of a number of systemic inequalities such as race, gender, socioeconomic status, and more that defined the boundaries of what they considered to be discrimination harm.

When looking at factors that may influence the severity of discrimination harm, participants' perceptions of discrimination harm were in part dependent on who spread the information and how frequently they saw it. As was also the case with reputation harm, the scalability of misinformation led participants to think of the harm as more severe and hold those involved in spreading it more accountable. More unique to discrimination harm was the role that current events often played in making this kind of misinformation spread.

Our participants discussed how their understanding of social and political contexts influenced their assessments of perceived discrimination harm, as well as what motivations others might have in disseminating it. This kind of determination of the sensitivity of content and what incentives actors may

have in spreading it helped participants develop a more nuanced understanding of trust and harm. Participants identified some commonalities between the two—particularly the importance of the platform policies and scalability—as well as key differences, especially the targeting of groups rather than individuals.

5) Combinations of genuine and manipulated or false content may also increase safety harm

Participants stated that misinformation that contains genuine content with false context, manipulation of imagery, and content designed to deceive may pose increased safety-based harms. The last type of harm that we found to be a significant factor in our participants' experiences of misinformation is safety harm. Participants noted that misinformation about physical or mental health (e.g., vaccinations, drug consumption, mental health disorders), responses to public crises or world events (e.g., COVID-19 pandemic, elections, natural disasters), and cybersecurity risks (e.g., phishing attacks, online scams) were especially common instances that triggered feelings of safety harm.

With the COVID-19 pandemic a cornerstone of adolescent experiences as of our writing (our participants had been in high school during pandemic-related lockdowns), participants were especially attuned to the safety harms of vaccine-related misinformation. Participants recognized that false claims around vaccine efficacy and potential side effects, which they had all seen throughout the COVID-19 pandemic, were prevalent and had significant public health consequences. Related work on pandemic misinformation demonstrates similar notions of social media's role in spreading COVID-19 misinformation, sometimes termed the "infodemic," which refers to the perils of misinformation during the management of disease outbreaks (Cinelli et al., 2020; Pennycook et al., 2020). The chance for misinformation to influence the way people think about how a viral infection can get transmitted, potential measures for prevention, and how the disease can get treated increased the proliferation of perceptions of safety harm.

Participants also indicated that their perceptions of safety harm increased when misinformation had the potential to impact their emotional well-being, indicating a relationship between safety harm and emotion harm. For instance, participants stated that pandemic-related misinformation that perpetuated safety harms also increased their sense of confusion and panic. Social media platforms that contained misinformation about false remedies and treatments led them to worry about how their safety may be compromised depending on the actions they took. When it came to health-related misinformation, participants were inclined to take actions that minimized harm to their safety and well-being. For instance, participant 23, after seeing a piece of misinformation regarding a mistake that Dr. Fauci (then the U.S. Chief Medical Advisor on COVID-19) made, indicated that a public health crisis such as COVID-19 implies a need for increased caution:

I would probably read more into the article to see the claims backing up this headline and then decide for myself whether or not I believe the information being provided. However, with the amount of unknowns when it came to COVID, I always felt it was better to be safe than sorry.

Another common topic of misinformation that propagated safety harms amongst adolescents was election misinformation. Participants worried that this kind of misinformation could diminish the legitimacy of election processes and promote voter discrimination and suppression, impacting public safety by challenging democracy. Participants also noted that cybersecurity threats such as phishing and online scams were common sources of safety harm.

Misinformation that propagated safety harm thus represented an important subgroup of information in participants' experiences. In particular, participants described encountering a lot of threats to public health related to the pandemic and public safety related to elections, as well as the perennial threat of

phishing and scams. This highlights the importance of taking public events into consideration when designing systems to combat safety harm.

Methods

Study design

We gathered data in four interrelated explorations with 25 participants: 1) a semi-structured background interview, 2) a guided retrieval of examples of misinformation, 3) a guided interaction with a researcher-created newsfeed of misinformation (Table 2 demonstrates the types of misinformation used), and 4) a situational mapping exercise, totaling 60 minutes per participant.

Table 2. *Types of misinformation used in the misinformation newsfeed portion of study.*

Misinformation Type	Definition
Satire or Parody	No intention to cause harm but has potential to fool
False Connection	When headlines, visuals or captions don't support content
Imposter Content	When genuine sources are impersonated
Fabricated Content	Content is false and designed to deceive and do harm
False Context	When headlines, visuals or captions don't support content

Participants

We recruited through a campus service that maintains a pool of students for research studies. These students were 18 (76%) or 19 (24%) years old and were enrolled full-time as undergraduate students. Eighteen participants identified as women, six identified as men, and one identified as nonbinary; 11 were Democrats, three were Independent, and 14 did not have or did not specify a political party affiliation.

Data analysis

For our data analysis, we primarily employed a constructivist grounded approach (Clarke et al., 2017). We began by transcribing recordings with a speech-to-text application (Otter.ai) and then used MAXQDA, a qualitative data analysis software, to iteratively code data from all 25 transcripts. We reached a point of saturation wherein we began to note the repeated presence of certain themes and patterns in the information. Further information regarding steps within the research methods, recruitment and participant demographics, data analysis, and limitations can be found in the Appendix.

Bibliography

- Borah, P., Irom, B., & Hsu, Y. C. (2022). 'It infuriates me': Examining young adults' reactions to and recommendations to fight misinformation about COVID-19. *Journal of Youth Studies*, 25(10), 1411–1431. <https://doi.org/10.1080/13676261.2021.1965108>
- Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-73510-5>
- Clarke, A. E., Friese, C., & Washburn, R. S. (2017). *Situational analysis: Grounded theory after the interpretive turn*. Sage Publications.
- Gak, L., Olojo, S., & Salehi, N. (2022). The distressing ads that persist: Uncovering the harms of targeted weight-loss ads among users with histories of disordered eating. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2). <https://doi.org/10.1145/3555102>
- Galvão, J. (2021). COVID-19: The deadly threat of misinformation. *The Lancet Infectious Diseases*, 21(5), e114. [https://doi.org/10.1016/S1473-3099\(20\)30721-0](https://doi.org/10.1016/S1473-3099(20)30721-0)
- Gelber, K., & McNamara, L. (2016). Evidencing the harms of hate speech. *Social Identities*, 22(3), 324–341. <https://doi.org/10.1080/13504630.2015.1128810>
- George, M. (2019). The importance of social media content for teens' risks for self-harm. *Journal of Adolescent Health*, 65(1), 9–10. <https://doi.org/10.1016/j.jadohealth.2019.04.022>
- Gisondi, M. A., Barber, R., Faust, J. S., Raja, A., Strehlow, M. C., Westafer, L. M., & Gottlieb, M. (2022). A deadly infodemic: Social media and the power of COVID-19 misinformation. *Journal of Medical Internet Research*, 24(2), e35552.
- Hassoun, A., Beacock, I., Consolvo, S., Goldberg, B., Kelley, P. G., & Russell, D. M. (2023). Practicing information sensibility: How Gen Z engages with online information. In A. Schmidt, K. Väänänen, & T. Goyal, T. (Eds.), *CHI'23: Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–17). Association of Computing Machinery. <https://dl.acm.org/doi/10.1145/3544548.3581328>
- Hawes, T., Zimmer-Gembeck, M. J., & Campbell, S. M. (2020). Unique associations of social media use and online appearance preoccupation with depression, anxiety, and appearance rejection sensitivity. *Body Image*, 33, 66–76. <https://doi.org/10.1016/j.bodyim.2020.02.010>
- Im, J., Schoenebeck, S., Iriarte, M., Grill, G., Wilkinson, D., Batool, A., Alharbi, R., Funwie, A., Gankhuu, T., Gilbert, E., & Naseem, M. (2022). Women's perspectives on harm and justice after online harassment. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2). <https://doi.org/10.1145/3555775>
- Jerit, J., & Zhao, Y. (2020). Political misinformation. *Annual Review of Political Science*, 23(1), 77–94. <https://doi.org/10.1146/annurev-polisci-050718-032814>
- Laplante, A. (2012). Who influence the music tastes of adolescents? A study on interpersonal influence in social networks. In *MIRUM'12: Proceedings of the second international ACM workshop on music information retrieval with user-centered and multimodal strategies* (pp. 37–42). Association of Computing Machinery. <https://doi.org/10.1145/2390848.2390857>
- Masucci, M., Pearsall, H., & Wiig, A. (2021). The smart city conundrum for social justice: Youth perspectives on digital technologies and urban transformations. *Annals of the American Association of Geographers*, 110(2), 476–484. <https://doi.org/10.1080/24694452.2019.1617101>
- Montgomery, K. (2000). Youth and digital media: A policy research agenda. *Journal of Adolescent Health*, 27(2), 61–68. [https://doi.org/10.1016/S1054-139X\(00\)00130-0](https://doi.org/10.1016/S1054-139X(00)00130-0)

- Musgrave, T., Cummings, A., & Schoenebeck, S. (2022). Experiences of harm, healing, and joy among black women and femmes on social media. In S. Barbosa, C. Lampe, & C. Appert (Eds.), *CHI'22: Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–17). Association of Computing Machinery. <https://doi.org/10.1145/3491102.3517608>
- O'Keeffe, G. S., Clarke-Pearson, K., & Council on Communications and Media (2011). The impact of social media on children, adolescents, and families. *Pediatrics*, *127*(4), 800–804. <https://doi.org/10.1542/peds.2011-0054>
- Paciello, M., Corbelli, G., & D'Errico, F. (2023). The role of self-efficacy beliefs in dealing with misinformation among adolescents. *Frontiers in Psychology*, *14*. <https://doi.org/10.3389/fpsyg.2023.1155280>
- Palfrey, J. (2010). The challenges of developing effective public policy on the use of social media by youth. *Federal Communications Law Journal*, *63*(1), 5–18. <https://www.repository.law.indiana.edu/fclj/vol63/iss1/3>
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, *31*(7), 770–780. <https://doi.org/10.1177/0956797620939054>
- Popat, A., & Tarrant, C. (2023). Exploring adolescents' perspectives on social media and mental health and well-being: A qualitative literature review. *Clinical Child Psychology and Psychiatry*, *28*(1), 323–337. <https://doi.org/10.1177/13591045221092884>
- Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In M. Palmer, R. Hwa, & S. Riedel (Eds.), *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2931–2937). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D17-1317>
- Rosenberg, H., Syed, S., & Rezaie, S. (2020). The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian Journal of Emergency Medicine*, *22*(4), 418–421. <https://doi.org/10.1017/cem.2020.361>
- Salac, J., Oleson, A., Armstrong, L., Le Meur, A., & Ko, A. J. (2023). Funds of knowledge used by adolescents of color in scaffolded sensemaking around algorithmic fairness. In K. Fisler, P. Denny, D. Franklin, & M. Hamilton (Eds.), *ICER'23: Proceedings of the 2023 ACM conference on international computing education research* (Vol. 1, pp. 191–205). Association for Computing Machinery. <https://doi.org/10.1145/3568813.3600110>
- Scheuerman, M. K., Jiang, J. A., Fiesler, C., & Brubaker, J. R. (2021). A framework of severity for harmful content online. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2). <https://doi.org/10.1145/3479512>
- Schoenebeck, S., Scott, C. F., Hurley, E. G., Chang, T., & Selkie, E. (2021). Youth trust in social media companies and expectations of justice: Accountability and repair after online harassment. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1). <https://doi.org/10.1145/3449076>
- Tran, T., Valecha, R., Rad, P., & Rao, H. R. (2020). An investigation of misinformation harms related to social media during two humanitarian crises. *Information Systems Frontiers*, *23*(4), 931–939. <https://doi.org/10.1007/s10796-020-10088-3>
- Treen, K. M. d'I., Williams, H. T., & O'Neill, S. J. (2020). Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *11*(5), e665. <http://dx.doi.org/10.1002/wcc.665>
- Winstone, L., Mars, B., Haworth, C. M. A., & Kidger, J. (2021). Social media use and social connectedness among adolescents in the United Kingdom: A qualitative exploration of displacement and stimulation. *BMC Public Health*, *21*(1). <https://doi.org/10.1186/s12889-021-11802-9>

- Xiao, S., Cheshire, C., & Salehi, N. (2022). Sensemaking, support, safety, retribution, transformation: A restorative justice approach to understanding adolescents' needs for addressing online harm. In S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, & K. Yatani (Eds.), *CHI'22: Proceedings of the 2022 CHI conference on human factors in computing systems* (pp. 1–15). Association of Computing Machinery. <https://doi.org/10.1145/3491102.3517614>
- Xiao, S., Jhaver, S., & Salehi, N. (2023). Addressing interpersonal harm in online gaming communities: The opportunities and challenges for a restorative justice approach. *ACM Transactions on Computer-Human Interaction*, 30(6), 1–36. <https://doi.org/10.1145/3603625>

Acknowledgements

We thank the Experimental Social Science Laboratory (Xlab) at the University of California, Berkeley for their support towards this research, the Berkeley Center for New Media (BCNM) and Isabel Li for assistance with initial data analysis, and the anonymous reviewers and editorial group of the *Harvard Kennedy School (HKS) Misinformation Review* for their feedback.

Funding

This research was supported by the Experimental Social Science Laboratory (Xlab) at the University of California, Berkeley.

Competing interests

The authors declare no competing interests.

Ethics

This study was approved by the University of California, Berkeley's Institutional Review Board. Informed consent was received from all subjects.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data availability

All materials needed to replicate this study are available via the Harvard Dataverse: <https://doi.org/10.7910/DVN/JQWE0U>

Appendix: Methodology details, participant recruitment, and data analysis

Semi-structured qualitative interviews

We first conducted 15-minute semi-structured qualitative interviews with participants. Our questions covered two themes: The first focused on how participants established trust or mistrust in their digital interactions and the second investigated their thoughts and perceptions of misinformation, perceptions of harm, and accountability. Questions included how participants identify misinformation in their online interactions, how their lack of trust arises, who they perceive may encounter harm and/or may be accountable, what moderation practices they've encountered, and how their exposure to different media sources influences their perceptions.

Guided artifact retrieval

Next, we asked participants to navigate through their search history (with the option to select from their mobile phone, tablet, or laptop) to display pieces of misinformation they encountered in the last 30 days. After giving participants 2–3 minutes to navigate through search history, we asked them to explain why the found source was identified as misinformation, how they navigated through the content, and accountability/content moderation practices they engaged with when they encountered the information. We then asked participants to further explain their interpretation of the media policies on the service that influenced their perceptions of trust and harm. This exploration took 15–20 minutes in total.

Misinformation newsfeed interaction

In our third exploration, which took 10–15 minutes, we presented participants with a pre-generated newsfeed of information that included different types of misinformation interspersed with truthful sources created in Qualtrics. The examples of misinformation we used had been recently trending on various online platforms and spanned multiple types of misinformation. All pieces of misinformation references were assessed based on prevalence on social media platforms and inclusion of certain characteristics when mapping to a specific type of misinformation. This part of the study explored participants' information consumption practices and navigation strategies. We showed participants two to three pieces of each type of misinformation described in Table 2 and Table 3 and two pieces of authentic content for a total of 15 newsfeed interactions (see Figure 1). After each, we asked participants to rank how likely it is that the media they are seeing is not truthfully representing information (on a 5-point scale from 1 = very unlikely to 5 = very likely) and to write down their reasoning in a text field. We also asked them to verbalize their thought processes using "think-aloud" prompting. At the end of this exploration, participants were asked to rank the media they had encountered by the level of harm it caused.

Situational mapping

In the final part of the study, participants were informed that they had interacted with pieces of misinformation (i.e., content that has now been verified as factually incorrect) and they then engaged in a situational mapping activity where they were asked to identify those who were harmed and those who were or should be accountable for misinformation on a collaborative workspace called Mural. First, we presented participants with a chart naming and describing the six different types of misinformation and a

thumbnail example. We then guided participants through a large interactive table, and we continued to prompt participants to think aloud as they filled it out. For each type of misinformation, we asked participants to describe:

- 1) whether or not they thought that kind of misinformation caused harm (yes, no, or maybe);
- 2) who they thought the recipients of harm were (from a list of kids/pre-teens, teenagers, politics/political parties, educators, government, differently abled, parents, and other);
- 3) who they thought was accountable for that harm or for preventing future harm (from a list of policymakers, government [with the option to list a specific branch or agency], media other than journalists, journalists, educators, public, technology companies, or other);
- 4) what level of accountability they felt that each stakeholder they listed had (neutral, somewhat accountable, or very accountable); and
- 5) what could be done to mitigate or decrease the risk of harm, based on their interactions with digital media as well as the information they gained through these activities.

Recruiting

Participants were compensated hourly, per the policies set by the campus service and our IRB protocol. Our recruiting message stated that we were looking for participants who use a variety of media channels on a regular basis and are willing to engage in an activity that explores their recent interactions with misinformation.

Data analysis

In the first round of coding, we identified patterns within each participant's interview. In the second round, we identified broader themes across participant interviews and wrote memos on connections to related literature. We then generated descriptive statistics on the data from newsfeed interactions and created affinity mapping diagrams to discover patterns in the participant assessments for reasons behind their trust assignments. Finally, we explored participants' assignments of levels of accountability and harm from the situational mapping exercise, as well as their reasoning, which helped us develop ideas to identify and mitigate harms spread by misinformation.

Limitations

There are a few limitations in our study which are important to note. First, for the purpose of this study, while we worked with participants of ages 18 and 19, we broadly categorized this group as adolescents. Precedence established by prior studies (Laplante et al., 2021; Salac et al., 2023; Xiao et al., 2022) has demonstrated the use of data from similar age groups, and this study aspires to understand information sensibility practices of adolescents by interfacing with 18- and 19-year-old users. However, working with a limited age group poses risks of not being able to authoritatively map to all adolescent users. We also recruited our sample from one university and note that this may limit the overall generalizability of our findings.

Second, while the design of the study intended to best use participant time and thought processes to understand their past experiences with misinformation and notions of content online, it may be the case that the ordering of the tasks may have led participants to think differently or more critically about misinformation than they otherwise would. Sessions were also conducted in a lab setting and, therefore, participant perceptions may change as they were not in their natural environment. In future work, we hope to structure our sessions with participants in such a way that we are able to limit potential biases.

Finally, while the proposed frameworks are based on findings from our study and existing literature, the scope of the paper did not involve the evaluation and testing of the ideas presented. Testing of the playbook will allow for an increasingly detailed and nuanced presentation of the potential uses of the frameworks suggested.