THE MISINFORMATION MACHINE

Improving Accountability and Transparency to Limit Polarization on Social Media

Isabel Schimpff and Kiley Sherlock

ELON UNIVERSITY | Phoenix Policy Institute

ABSTRACT

As social media becomes a heavily relied on source for news, entertainment, and information, the algorithmic system paired with the use of artificial intelligence exacerbates the spread of misinformation and polarization between users. These issues of misinformation and polarization are left to grow with little to no enforced policies currently in place. To combat this, this memo encourages the implementation of a two-part policy which pairs behavioral friction with mandated transparency reports. These seemingly small steps will go far to bring this “mysterious” problem to light and turn the complex innerworkings of these digital systems into digestible topics.

TABLE OF CONTENTS

01 INTRODUCTION

02 PROBLEM DEFINITION

03 HOW MISINFORMATION SPREADS

04 RATIONALE FOR GOVERNMENT INTERVENTION

05 POLICY ALTERNATIVES

06 RECOMMENDATIONS AND IMPLEMENTATION

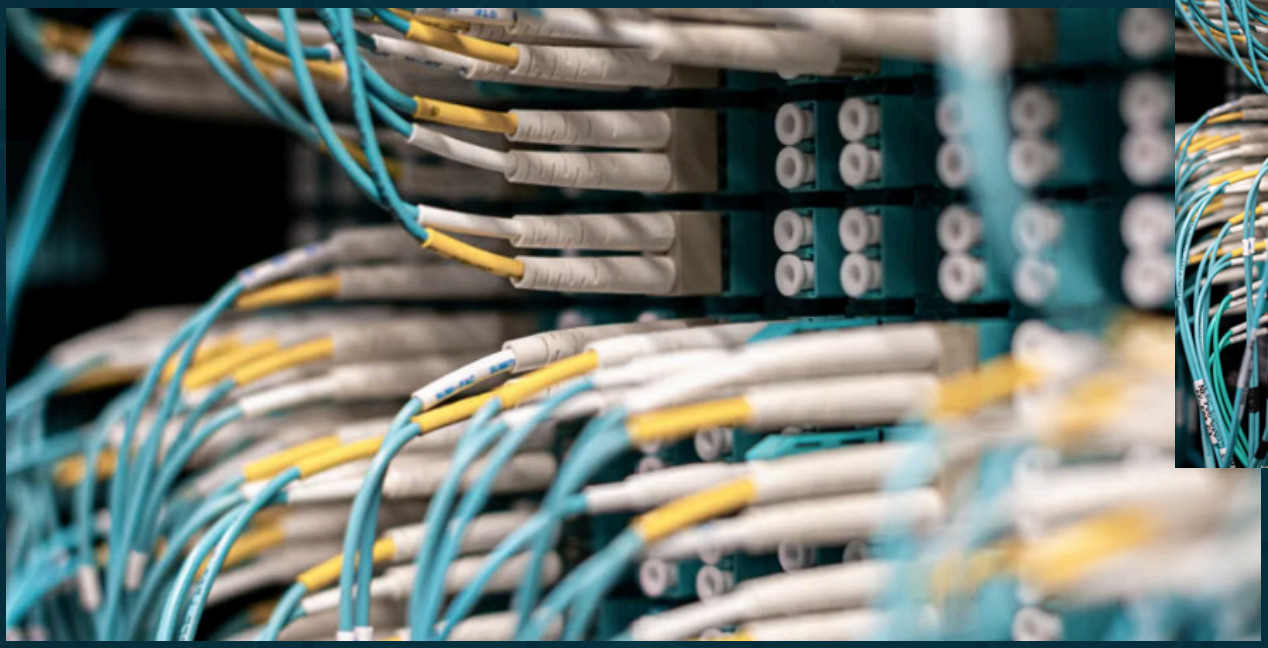
07 CALL TO ACTION

08 REFERENCES

01 Introduction

The rise of citizen journalism and social media as a source of news has transformed how Americans access and share information. The promotion of apps like Instagram, TikTok, and Facebook, has welcomed the rise of mistrust and polarization between social media users, which is then exacerbated by the unregulated use of artificial intelligence imbedded in these platforms. We recommend the adoption of a federal policy mandating transparency reports by social media companies, as well as the incorporation of an extra-click measure to encourage users to review the content before sharing.





02 Problem Definition

Social media has become a foundational news source as technology continues to advance. According to the Reuters Institute Digital News Report, 54% of respondents from a sample of 48 countries identified social media as the most common way to access online news, found across all age groups (Egan, 2026). Relying on social media for news has been prevalent in the United States for quite some time and is only continuing to grow, particularly through short-form video content. From 2020 to 2025, the percentage of TikTok users who regularly received news updates on the platform increased from 22% to 55%, and 35% of U.S. adults identified YouTube as a common news source (Liedke & Aubin, 2025). These high rates of social media-driven news consumption have transformed the news media landscape, ushering in drastic changes in media literacy and news consumption habits.



02 Problem Definition

The rise of social media as a primary form of news consumption has also coincided with an increase in misinformation. Social media platforms are not required to comply with the regulations and journalistic values of traditional news media organizations, resulting in higher likelihood of misinformation being published and spread by algorithms (APA, 2024).



Adding another layer to the problem of misinformation is the increase of content generated by artificial intelligence (AI). The use of AI on social media platforms has expedited the mistrust that is already growing on these platforms, further polarizing users. AI spreads misinformation by not only creating highly believable but fabricated images (Rutgers, 2026), but also in the fact that social media platforms use AI to tailor algorithms and messages to keep people scrolling on their platforms for longer (Virginia Tech, 2024).

03 How Misinformation Spreads

A driving force of misinformation in news published on social media is the incentivization of misinformation by algorithms. A study examining the sharing behavior of habitual Facebook users found that habitual users were more likely to share misinformation and were less likely to correctly discern whether a given headline was true. In this study, 15% of habitual users contributed to 39% of shared misinformation across all experimental conditions, and additional findings proposed that the increased amounts of likes and shares associated with misinformation incentivizes habitual users to share politically polarizing headlines, even if the topic does not align with their own political beliefs (Ceylan et al., 2023).



03 How Misinformation Spreads

Rewarding the spread of misinformation on social media by increasing opportunities for engagement has a negative impact on media literacy, encouraging the spread of politically polarizing content. This model prevents social media executives from considering the implementation of regulations and safeguards, as the current model of misinformation boosted by algorithms draws more users to their platform.



The removal of factcheckers from Facebook and Instagram by Meta in 2025 confirms this, with CEO Mark Zuckerberg acknowledging that “more harmful content will appear on the platform as a result of the content moderation changes” (Duffy, 2025, para. 4)

04 Rationale for Government Intervention

The rampant spread of misinformation on social media requires government intervention to ensure a policy solution that is consistent and enforceable across all social media platforms. While conversations regarding the regulation of content on social media have occurred for years in the United States, the rapid emergence of AI has exacerbated these discussions, exacerbating the problem of misinformation. As the public, politicians, and AI leaders alike call for increased safeguards to regulate the speed of AI development, there seems to be an opportunity for government intervention (Rosen, 2026; Isaac, 2026).

04 Rationale for Government Intervention (Cont.)

The difficulty then lies in formulating an effective policy to limit misinformation that does not threaten free speech. The debate surrounding social media regulation and free speech is largely fueled by competing perspectives on what constitutes hateful or violent speech. For a government policy designed to curb misinformation on social media to succeed, it must seek to balance these perspectives by prioritizing transparency and implementing less restrictive regulations (Hwang, 2025).



05 Policy Alternatives

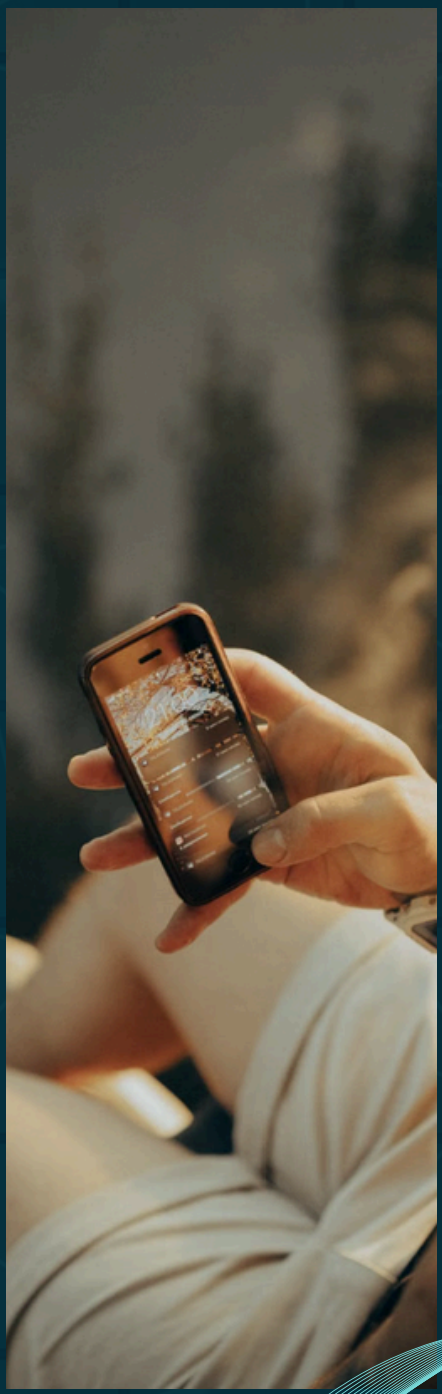
A proposed policy solution to mitigate the spread of misinformation on social media is behavioral friction, which refers to in-app protections that make it more difficult to share social media posts through deterrents such as extra clicks before posting. This type of deterrent seeks to address the spread of misinformation by encouraging users to slow down and consider the content they are sharing before posting. When users go to post, an additional message pops up on the screen, warning the user of potential misinformation and requiring an additional click to publish the post. Apps including Twitter and Nextdoor have implemented this strategy in the past to reduce the spread of misinformation, relying on a shift in user behavior (Garland et al., 2026). While behavioral friction does not prevent users from posting misinformation, it offers an additional safeguard to encourage critical thinking on what is being shared. A University of Copenhagen study on the use of friction and quality control to mitigate misinformation found an increase in post quality when small amounts of friction measures are in place, but a negative impact on post quality when "excessive" amounts of friction are implemented (Jahn et al., 2025). This shows that while the approach may be successful in some aspects of influencing behavior, it may not be enough on its own to mitigate the spread of misinformation.

05 Policy Alternatives

An additional proposed policy solution to help rectify the issue of AI leading to the spread of misinformation on social media is the implementation of unified and outside transparency reports. These reports will require that these platforms publicly disclose information quarterly on certain pertinent metrics. Legal requests of this nature like Government Requests for User Data have already been implemented by companies like Meta (Meta, 2026). These proposed reports would focus on AI usage and include statistics about how many posts include AI generated or manipulated content among other statistics. Similar ideas are gaining traction with institutions like Harvard implementing systems like their Transparency Hub (Harvard, 2026). This policy alternative would narrow in on the issue of AI in social media and hold companies to a more public reporting system. In the past, transparency reporting on social media has been criticized for not being truly transparent (TechPolicy, 2024). To combat this, this policy will include an outside body to help facilitate these reports from the companies publishing them. This outside body would exist in the form of the federal government, specifically the Office of Information Policy.

06 Recommendations and Implementation

We recommend the adoption of a policy that combines both transparency reports and behavioral friction. This policy should be implemented at the federal level, in the Office of Information Policy in the U.S. Department of Justice, to ensure consistency and enforceability (Office of Information Policy, 2024). By coupling transparency reports with an extra-click behavioral friction tool, social media users will be given access to data regarding the spread of misinformation and will be prompted to consider misinformation each time before posting or reposting content. The data collected from quarterly transparency reports will be made public on the Office of Information Policy Website, as well as the website of each social media platform. Additionally, users will be notified on the app each time the report is published to increase awareness of the resource.



06 Recommendations and Implementation

The critique of regulating misinformation on social media due to free-speech concern is remediated by the transparency report policy alternative, as social media companies will be responsible for reporting misinformation rather than the consumer.This method does not directly prevent the spread of misinformation, but it provides information on the prevalence of misinformation on a given app to encourage users' awareness of the issue.

This also applies to the extra-click tool, which does not prevent users from posting, but encourages them to review the content they are sharing to mitigate the spread of misinformation. By providing information, people are given the opportunity to shift their behavior to be cautious of reposting misinformation.



07 Call to Action

It has become apparent that social media as a platform for the spread of news and entertainment for the masses is here to stay. This inevitably means that policies must be seriously considered in order to combat the growing issues of misinformation accessibility on social media. These issues of increased polarization and misinformation do not yet have a formidable opponent, and they are only increased by the usage of AI. Narrowing down this large and complex problem makes it appear more manageable, inspiring the development of more regulations on the spread of misinformation. At the current moment, strong regulations lack support for national implementation.

Developing policies that increase information accessibility and encourage media literacy addresses the problem of the spread of misinformation while remaining non-invasive, avoiding critiques of free-speech violations. Transparency reports and extra-click tools are important steps to take in the development of effective social media policy. While these policy alternatives may seem like a small step, in the age of misinformation, we believe that access to information is the most important tool to combat the spread of misinformation on social media.

References

Ai and the spread of fake news sites: Experts explain how to counteract them (2024, February 22). Virginia Tech News | Virginia Tech. (2024, February 22). https://news.vt.edu/articles/2024/02/AI-generated-fake-news-experts.html%E2%80%AF

American Psychological Association. (2024). How And Why Does Misinformation Spread? American Psychological Association. https://www.apa.org/topics/journalism-facts/how-why-misinformation-spreads

Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. Data & Policy, 3, e32. doi:10.1017/dap.2021.20

Ceylan, G., Anderson, I. A., & Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. Proceedings of the National Academy of Sciences, 120(4), e2216614120.

Egan, J. (2026, June 16). Overview and key findings of the 2026 Digital News Report | Reuters Institute for the Study of Journalism. Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2026/dnr-executive-summary

Garland, J., Bak-Coleman, J., Benesch, S., DeDeo, S., DiResta, R., Eissfeldt, J., ... & Semenova, V. (2026). The case against efficiency: friction in social media. npj Complexity, 3(1), 5.

Government transparency. Office of Information Policy. (2024, July 2). https://www.justice.gov/oip/government-transparency

How ai-generated images are fueling misinformation. (2026, March 24). Rutgers University Libraries. https://www.libraries.rutgers.edu/news/how-ai-generated-images-are-fueling-misinformation

How AI shapes misinformation and disinformation: Key differences explained. Southampton Online. (n.d.). https://www.online.southampton.ac.uk/blog/how-ai-shapes-misinformation-and-disinformation-key-differences-explained

References

Hwang, L. (2025). Regulation of Misinformation in the Digital Age: First Amendment Rights and Government's Role in Social Media. Critical Debates in Humanities, Science and Global Justice, 5(2).

Isaac, M. (2026). Top A.I. Leaders Call for Slowing Down A.I. Development. The New York Times. https://www.nytimes.com/2026/09/12/technology/anthropic-dario-amodei-ai-slowdown.html

Jahn, L., Rendsvig, R. K., Flammini, A., Menczer, F., & Hendricks, V. F. (2025). A perspective on friction interventions to curb the spread of misinformation. npj Complexity, 2(1), 31.

James, L. (2026). Social Media Terms of service reporting. Social Media Terms of Services . https://ag.ny.gov/resources/organizations/social-media-tos-reporting

Kraemer, J. (2024, February 15). On social media, transparency reporting is anything but transparent. Tech Policy Press. https://www.techpolicy.press/on-social-media-transparency-reporting-is-anything-but-transparent/

Liedke, J., & St. Aubin, C. (2025). Social Media and News Fact Sheet. Pew Research Center. https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/

Office of Information Policy (2024). Government transparency. U.S. Department of Justice. Www.Justice.Gov. https://www.justice.gov/oip/government-transparency

Rosen, J. (2026). Americans strongly support regulations on AI. HUB, John Hopkins University. https://hub.jhu.edu/2026/06/15/americans-strongly-support-regulations-on-ai/

Shin, jieun. (2026, February 23). Ai in the age of fake (imagined) content. STIMSON. https://www.stimson.org/2026/ai-in-the-age-of-fake-imagined-content/

Transparency Center. (n.d.). https://transparency.meta.com/reports/

Transparency Hub. (2026, June 29) Applied Social Media Lab at Harvard University. https://asml.cyber.harvard.edu/transparency-hub/