FOCUS PAPER NO. 8

June 2025

TACKLING ONLINE DISINFORMATION AT THE INSTITUTIONAL AND SOCIETAL LEVEL

Elizaveta Gaufman
University of Groningen

Martijn Schoonvelde
University of Groningen

REGROUP
REBUILDING GOVERNANCE AND RESILIENCE OUT OF THE PANDEMIC

FUNDED BY THE EUROPEAN UNION
HORIZON EUROPE - RESEARCH AND INNOVATION ACTIONS
GRANT NO. 101060825

REGROUP

REBUILDING GOVERNANCE AND
RESILIENCE OUT OF THE PANDEMIC

Culminating more than a decade of crisis in Europe, the Covid-19 pandemic has opened an important window of opportunity for institutional and policy change, not only at the "reactive" level of emergency responses, but also to tackle more broadly the many socio-political challenges caused or exacerbated by Covid-19. Building on this premise, the Horizon Europe project REGROUP (*Rebuilding governance and resilience out of the pandemic*) aims to: 1) provide the European Union with a body of actionable advice on how to rebuild post-pandemic governance and public policies in an effective and democratic way; anchored to 2) a map of the socio-political dynamics and consequences of Covid-19; and 3) an empirically-informed normative evaluation of the pandemic.

www.regroup-horizon.eu

info@regroup-horizon.eu

@regroup-horizon.bsky.social

@regroup-horizon

@regroup_horizon

FUNDED BY THE EUROPEAN UNION
HORIZON EUROPE - RESEARCH AND INNOVATION ACTIONS
GRANT NO. 101060825

Executive summary

Disinformation has become an increasing concern for European policymakers and the broader public, raising pressing questions about safeguarding political discourse and institutional trust in an evolving digital landscape. This focus paper begins by mapping the current landscape of disinformation policies within the European Union, including public and private regulatory approaches. While significant efforts have been made—particularly in targeting foreign information manipulation—we argue that existing policy frameworks overlook important aspects of how disinformation works. Above all, there is a lack of attention to the issue of credibility: how individuals determine whether information is trustworthy and what role identity cues play in their assessments.

To address this gap, we conducted a survey experiment to understand how young people evaluate the credibility of online content. The experiment involved 152 university students at the undergraduate and postgraduate levels. Respondents were sorted into different groups based on their self-identified gender and country of origin at the beginning of the survey. They were then asked to assess the credibility of a series of social media posts covering a range of topics. Finally, respondents were invited to reflect in writing on the reasons behind their evaluations, allowing us to compare quantitative trust ratings with qualitative reasoning.

Our findings confirm that content quality remains the strongest predictor of perceived credibility. However, they also reveal that subtle identity cues—such as a shared gender or national background between the respondent and the post's author—exert a small but consistent influence on trust evaluations. These effects often occur below the level of conscious awareness, suggesting that social proximity and identity alignment can shape how people perceive truth, even when they believe they are evaluating content objectively. Notably, the results also challenge some common assumptions: posts from authors with Anglo-American names, often associated with epistemic authority, were not rated as more credible, and male-presenting sources were not favored over female-presenting ones.

Taken together, our findings underline that perceived credibility of information is not a fixed quality but shaped by content, identity, and situational cues. Disinformation policies may therefore fall short if they do not account for these more subjective dynamics of trust. In terms of policy recommendations, we suggest the following:

- Countering disinformation will require adaptive communication strategies that reflect changing digital habits—especially among younger audiences moving toward newer platforms.
- Effective responses must go beyond enforcement and fact-checking. They must include pre-bunking and media literacy initiatives sensitive to identity dynamics, and outreach that respects audience diversity rather than assuming universal standards of credibility.
- A one-size-fits-all approach to trust will not suffice. Instead, the EU must pursue a more flexible, inclusive, and evidence-based strategy—one that addresses not only the what of disinformation but also the why of belief.

Key words: disinformation, credibility, source cues, experiment.

Introduction

Numerous policymakers have emphasized the challenge of the global spread of misinformation (Pasquetto et al. 2020; Mont'Alverne et al. 2024). Online disinformation presents significant societal and political challenges, undermining trust in democratic processes and public health initiatives. Among the most pressing concerns is its impact on electoral integrity. False narratives about election fraud, voter suppression tactics, and manipulated information campaigns can erode public trust in democratic institutions, leading to decreased voter participation and increased skepticism about election outcomes. Similarly, disinformation has fueled vaccine hesitancy by spreading misleading claims about vaccine safety and effectiveness. This has contributed to lower vaccination rates, prolonged public health crises, and heightened strain on healthcare systems (see also Moutselos 2023, Neely et al 2022).

Beyond these immediate effects, disinformation deepens political polarization and weakens trust in key institutions such as the media and political parties (Volk et al 2023). By amplifying divisive narratives, disinformation campaigns create an environment where consensus becomes nearly impossible, making democratic governance more difficult. The erosion of trust in traditional sources of information leaves the public more susceptible to conspiracy theories and alternative narratives that serve political or financial interests. In short, disinformation has short- and long-term consequences, not all of which are entirely visible (see also Böck and Kettemann, 2024).

So far, the main way to target misinformation is to fact-check it (Walter et al. 2020, Pretus et al. 2023) and label it as such (Aslett et al. 2022, Dias, Pennycook, and Rand 2020, Kreps and Kriner 2022). However, meta-analyses show that "the ability to correct political misinformation with fact-checking is substantially attenuated by participants' preexisting beliefs, ideology, and knowledge" (Walter et al. 2020). Thus, a number of scholars focused specifically on inoculation and pre-bunking (Lewandowsky Van Der Linden 2021), but it is also important to analyze what makes misinformation believable in the first place (Strömbäck et al. 2020). Many such studies test the content of posts: how sensational or emotional it is supposed to be that would indicate potentially a malicious actor (Mourão and Robertson 2019, Leng et al. 2021, Hamby, Kim, and Spezzano 2024, Staender et al. 2022), even though experimental studies show a weak role of sensationalism for average users (Staender et al. 2022). At the same time, several studies have pointed out that flows of rage (Ganesh 2020, Ganesh and Faggiani 2024, Young and Young 2020, Lee et al. 2023) can have a significant effect on the spread and circulation of misinformation, underlining the fact that affect is often more significant than fact for social media users (Weeks 2023), not to mention the (ethnoracial) identity of the user themself (Crowder-Meyer & Ferrín 2021).

Addressing online disinformation requires not just fact-checking but also broader efforts to strengthen media literacy, improve platform accountability, and rebuild institutional trust in an era of digital manipulation. Effectively countering disinformation requires a shift in focus from simply identifying and removing false or misleading information (FIMI) to domestic information manipulation and interference (DIMI). This means addressing the structural factors that enable disinformation to thrive in local information environments rather than just reacting to individual instances of false content. A key component of this approach is stricter platform regulation and rigorous enforcement of the Digital Services Act (DSA), ensuring that major digital platforms are held accountable for spreading harmful content. Additionally, raising barriers for network entry—such as stricter verification processes for new accounts and reducing the ability to automate disinformation campaigns—can help curb the rapid dissemination of false narratives.

At a societal level, countering disinformation requires adapting to shifting digital habits, particularly among younger audiences who are moving away from platforms like *X* and engaging more with decentralized or emerging social networks. Public-private cooperation will be essential in this effort, particularly through partnerships with influencers and content creators who can effectively communicate credible information to diverse audiences. Multilevel communication strategies that integrate government messaging with grassroots digital engagement can help rebuild trust and improve resilience against disinformation. By combining stronger institutional enforcement with dynamic, audience-specific outreach, the EU can create a more comprehensive and adaptive strategy to counter the evolving threat of disinformation.

Overview of existing public and private approaches to regulating information in the EU

According to the European Union policy, Foreign Information Manipulation and Interference (FIMI), including disinformation, poses a growing security and foreign policy threat to the EU. The European External Action Service (EEAS) has taken a leading role in addressing this challenge by significantly enhancing its capabilities to identify, analyze, and respond to FIMI since 2015. The EEAS's efforts focus on delivering targeted, impactful, and coordinated responses to protect the EU's Common Foreign and Security Policy, democratic processes, citizens, and global partners. This includes monitoring and addressing FIMI activities from actors such as Russia and China, facilitating the Rapid Alert System (RAS) for coordination among EU Member States and Institutions, building resilience by empowering civil society and supporting independent media, and developing strong legislation like the Digital Services Act (DSA).

Additionally, the EEAS has developed a standardized analytical model to detect FIMI activities, integrating narrative analysis with examining the tactics, techniques, and procedures used by threat actors. This approach aims to uncover how FIMI campaigns are structured, executed, and adapted over time. Through these comprehensive measures, the EEAS promotes information integrity and fundamental freedoms, contributing to long-term stability, institutional robustness, and democratic security in the face of evolving FIMI threats. EEAS also emphasizes the role of strategic communications, public awareness, fact-checking networks, cross-sector collaboration for and with EU institutions, EU Member States and partners. These measures can be effective only within the context of a strong civil society, where transparency, accountability, and democratic resilience are central, as digital/information literacy is crucial to resist FIMI.

As mentioned above, apart from FIMI detection, the EU's approach to combating disinformation has largely focused on fact-checking and debunking false narratives (Rodríguez-Pérez et al. , 2025, Graves and Cherubini, 2016). While these are valuable tools, they do not fully address disinformation's complex and evolving nature (cf. Hinds 2019; Ecker et al. 2022). A significant gap in EU policy is the lack of integration of academic insights, particularly from media studies, psychology, and political science (Nenadic 2019). Research has consistently shown that disinformation is not merely a cognitive processing problem but is deeply tied to emotional manipulation, social identity, and digital platform dynamics (Birks 2021). This misperception has led to an overemphasis on correcting false information rather than addressing the mechanisms through which disinformation spreads and influences public opinion. Moreover, misinformation often has a significant gendered component (Banet-Weiser 2021; Marwick et al 2021), where misogyny is normalized and is not even considered as misinformation even though it often plays a significant role in polarization dynamics (Hedling 2024).

Moreover, the digital landscape constantly shifts, presenting new challenges in tackling disinformation, as reflected in the European Union's (EU) 2030 Digital Decade Policy Programme (DDPP). Differences in digital architecture and network features, as well as levels of digitalization, play a crucial role in how disinformation is disseminated and consumed (Torchio 2025). The phenomenon of "enshittification" highlights the way digital platforms increasingly serve their own profit-driven interests at the expense of users, making it difficult for individuals to migrate to alternative networks (cf. Hagen 2023). This entrenchment fosters an environment where harmful content can thrive, exacerbating radicalization dynamics (Kruglanski et al 2019). To develop a more effective disinformation policy, the EU must go beyond fact-checking and incorporate a broader understanding of these systemic issues, ensuring that policy measures address not only falsehoods but also the digital structures that enable their spread.

Experiment findings

While much of the experimental literature returns to message characteristics as the main object of analysis (Schaewitz et al., 2020), and occasionally engages with the notion of source credibility (Marecos et al., 2024), our work focuses on a less discussed but no less insidious tactic in the arsenal of digital deception: the infiltration of local bubbles. Misinformation, after all, does not merely announce itself; it adopts familiar signifiers to pass as members of the target community (Yin et al., 2018).

Rather than dissecting the source in the usual terms—such as trustworthiness, competence, or transparency (Luo, Yang, and Kang, 2022)—we ask a different question: what happens when the source feels close? Can a name that sounds like it could belong to your neighbor increase the perceived credibility of content that might otherwise raise eyebrows? In cases where the message itself is ambiguous, proximity—whether national or gender-based—might act as a heuristic shortcut, a cognitive nudge toward trusting. In other words, closeness becomes credibility.

To explore this, we constructed an experiment in which participants were shown repackaged posts from VKontakte from known disinformation actors. We chose posts that walked the line, containing a certain amount of truthful information. After some light linguistic editing to avoid tripping our respondents, the posts were re-made as *X* (formerly Twitter) content, complete with fake names and AI-generated faces courtesy of whichfaceisreal.com.

Each name was selected to resonate with a particular national identity that we had in our cohort of students at our home institution—Dutch, French, German, and so on—thus signaling a proximity that, we expected, could influence credibility assessments. Our respondents consisted of a student cohort of 152 (undergraduate and postgraduate) participants, and the survey (administered through Qualtrics software) was fully anonymized. Respondents were funneled into demographic-based flows tailored to their national and gender identities. These flows included German and Austrian, Dutch, Romanian, Italian, Polish and Czech, and a generic Anglo-American cluster. Then, the respondents were asked to rank the credibility of 6 posts on a 5-likert scale ranging from 1 ('not all credible') to 5 ('completely credible'). The posts themselves were on 6 different topics: war in Ukraine, war in Gaza, Telegram, stabbing in a German town, stocks, and drug mafia in the Netherlands.[1] This experimental set-up allowed us to examine if respondents rated posts as more credible if it came from a source that they shared their self-identified gender and country of origin with.

After the survey, participants were invited to reflect in their own words why they judged certain posts to be more or less credible. This qualitative reflection offered a window into the reasoning—or, perhaps, the rationalizations—behind the ratings, allowing for a

1. We are not quoting the posts themselves to protect the identity of the poster(s).

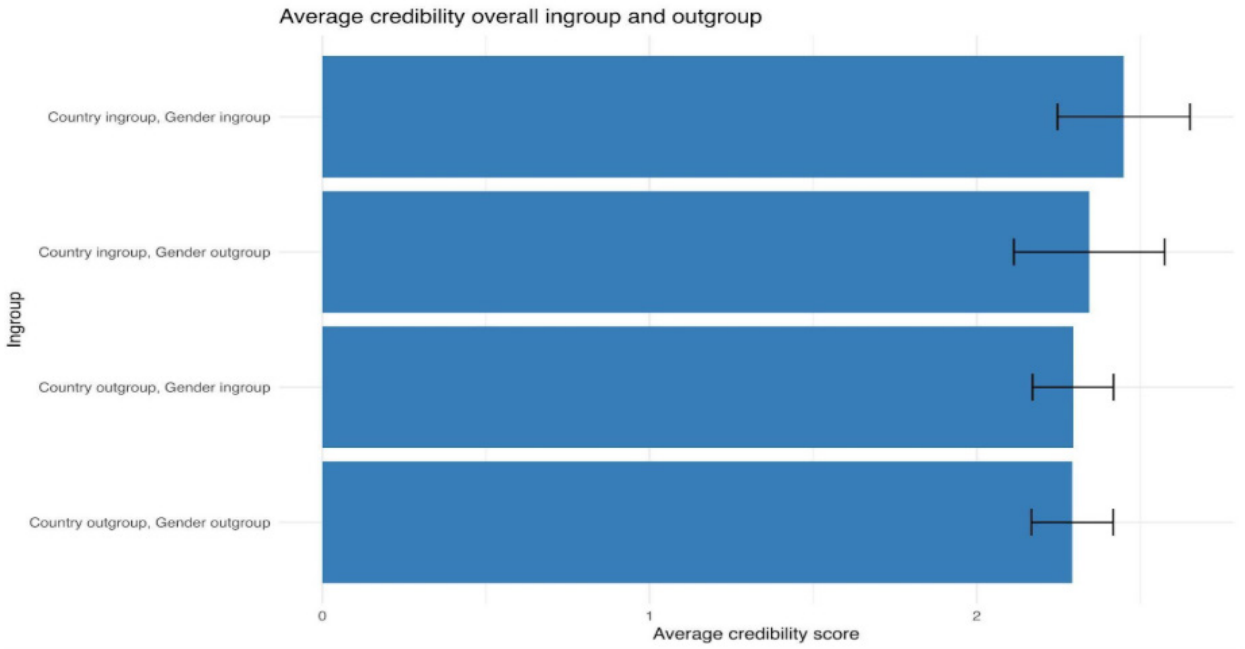
comparison between perceived credibility as a number and as a narrative. Importantly, posts were randomly shuffled so that each respondent encountered a set of fictional posters: one that bore national resemblance (same country of origin, both male and female variants), one that hailed from elsewhere in Europe (again, male and female), and one from a more "global" Anglo-American source—our control cosmopolitan. Respondents who selected "other" for country of origin, those who defied neat national categorization, were selected into the control group and exposed exclusively to the Anglo-American condition.

We included an explicit warning in the consent form regarding the use of deception, and a thorough debriefing followed the survey's completion. The design was reviewed and approved by our university's ethics review board.

Results indicated that the credibility of *X* is in severe decline. Several respondents noted in their open-ended answers that the mere fact that the posts they reviewed originated from that platform negatively influenced their credibility assessments - regardless of content. As one of the respondents stated "I don't think Twitter is a good source to get your information from". Given the average age of the respondents, we argue that the efforts on combating misinformation among the younger generation should not focus on *X*. Instead, attention should shift toward the platforms they actively use. Further research is needed to study the impact of digital affordances on credibility beyond *X* (Tripodi et al 2024, Masip et al 2021). One respondent pointed out specifically that the "blue check mark" was meaningless because anyone can buy it.

In the qualitative responses, most participants commented more on the appearance of the posts than the identity of the poster. Mistake-free language and professional-looking visuals were often cited as indicators of credibility. Interestingly, while participants rarely reflected explicitly on the source's identity, our quantitative data shows that identity—particularly shared nationality and gender—had a small but consistent effect on credibility ratings (see Figure 1). This suggests that identity cues may operate subconsciously in the credibility assessment process. Future research should explore how platforms that foreground identity more prominently—such as Instagram or TikTok—affect these dynamics.

Figure 1: Average credibility ratings (higher scores denote more credible) for in- and outgroup posters



Engagement metrics, such as the number of likes, retweets, or replies, had surprisingly little influence on credibility assessments. We purposefully randomized the engagement numbers (answer, retweet and likes) to make the posts appear popular and to mimic the IRA troll strategy of "engagement farming" that is supposed to boost credibility (Saeed et al 2022, DiResta 2022). Yet, only one respondent mentioned engagement numbers as affecting their perception of credibility. This suggests that, at least among this cohort, superficial popularity signals may not play as central a role as often assumed.

We also concluded that the students possessed several media literacy skills that have been identified as critical in combating misinformation (Friesem 2019): several students questioned the credibility of some posts because they did not contain a source for the information. Additionally, most of the students identified the tone of the posts as "opinion" rather than "fact" which also showcases the presence of critical skills necessary to combat misinformation (Art 2018).

Finally, our findings support the widely publicized strategy of Kremlin-linked disinformation campaigns that seek to embed themselves within local communities. Respondents, often unconsciously, placed more trust in sources that felt familiar—those who shared their gender or national background. This aligns with the notion that people tend to trust "the devil they know." At the same time, EU disinformation policy focuses primarily on foreign information manipulation. Yet, our data suggest that domestic actors who share linguistic, cultural, or demographic ties with a target audience may pose a more credible—and thus more insidious—threat.

Policy recommendations

Proximity-based credibility can intensify the impact of disinformation when it is spread within tight-knit regional or local online communities, even if the source is ultimately manipulative or deceptive. Platforms may unintentionally reinforce this bias through algorithmic amplification, by prioritizing locally relevant content or interactions with nearby users. This dynamic significantly increases the systemic risk of disinformation undermining democratic processes, public trust, or crisis response (e.g., in pandemics or elections). Because such risks transcend national boundaries and are tied to platform design and moderation practices, Article 34 of the DSA empowers the European Commission to supervise and enforce Very Large Online Platforms' compliance with systemic risk mitigation obligations.

Under Article 34, the Commission can audit platforms, require access to internal data, and impose corrective measures, ensuring a harmonized and effective response to systemic threats—including those magnified by geographic proximity bias. Based on the results of our survey experiment, we recommend the following 4 policy priorities to counter disinformation.

1. Focus disinformation interventions on content, not just source detection

Our study finds that content is the strongest predictor of perceived credibility, more so than source characteristics. Under the DSA it should be explicitly required to evaluate how local trust dynamics (e.g., geographic proximity bias) exacerbate systemic risks during their mandated risk assessments (Article 34(1)). The European Commission can issue guidance to ensure such biases are considered when identifying vectors of disinformation.

Policy recommendation:

- Invest in tools and campaigns that identify manipulative or misleading content patterns (e.g. emotional framing, logical fallacies) rather than solely focusing on labeling "foreign" or "untrusted" sources.
- Train fact-checkers, moderators, and AI systems to prioritize content-based risk signals.
- Localized content moderation teams and fact-checking partnerships. These should reflect regional linguistic, cultural, and political contexts to detect manipulative content more effectively. The Commission can encourage this by integrating it into best practice recommendations.

2. Design pre-bunking and media literacy efforts with awareness of subtle identity cues

While content matters most, respondents in our survey experiment still show slight, unconscious bias toward sources that match their gender and nationality, trusting information that comes from these 'ingroup members' slightly more. Platforms often use location data to enhance engagement, but this same data should be leveraged to identify and preempt the viral spread of disinformation within regional clusters. The Commission should encourage platforms to use proximity signals for containment, such as friction mechanisms (e.g., warnings, sharing limits) when localized virality is detected.

Policy recommendation:

- Include in media literacy and pre-bunking campaigns awareness of how subtle identity cues (like name or profile picture) can affect our trust in information.
- Create educational materials that show how in-group cues (e.g., a person "like me") can be exploited—even by malicious actors.
- Be alert to the tactical use of local messengers: Disinformation campaigns may co-opt visual and demographic 'localness' (e.g., female-presenting or minority messengers) to increase credibility and evade suspicion, especially in contexts where audiences expect inclusivity. Watch for content that exploits diversity without transparency or verifiable affiliations.
- Flag performative localness as a potential red flag: Inauthentic or strategic deployment of 'ingroup' spokespersons—especially in high-stakes or polarizing contexts—may indicate source manipulation. Disinformation efforts can use "trusted messenger" aesthetics as camouflage for misleading content.

3. Avoid overreliance on "Western epistemic authority" in counter-disinformation messaging

The results from our study challenge the assumption that Anglo-American sources are automatically seen as more credible compared to non-Anglo sources. Cooperation with the European Digital Media Observatory (EDMO) and EU vs DisInfo on localized narratives could be helpful. This would help coordinate enforcement actions under Article 34 and support data-driven policymaking.

Policy recommendation:

- Ensure that counter-disinformation initiatives represent a broad spectrum of linguistic, cultural, and geographic perspectives. This includes spotlighting knowledge producers from non-Western contexts.

- Be cautious of disinformation disguised in localized packaging: Campaigns may use cultural proximity by using familiar or local-sounding messengers—even if the content originates from external actors. Fact-checkers and analysts should consider the possibility that familiarity is being used deceptively, especially when messages reinforce existing grievances or polarization.
- Evaluate credibility based on verifiable networks, not cultural cues: Since source credibility is not automatically tied to Anglo-American presentation, identifying disinformation requires careful tracing of institutional affiliations, funding, and dissemination patterns, rather than relying on perceived cultural legitimacy.

4. Treat credibility as context-dependent—avoid one-size-fits-all responses

Our study shows that trust is shaped by subtle and context-specific cues, and is not universally driven by any single factor. The Commission should strengthen mechanisms for coordination between national Digital Services Coordinators and bolster its own central enforcement powers under Article 34(2).

Policy recommendation:

- Develop audience-specific communication strategies that are tested and refined across different demographic and cultural groups (such as younger users).
- Use experimental testing and behavioural insights to identify which content and messengers are most effective in different contexts.

References

Art, S. (2018). Media literacy and critical thinking. *International Journal of Media and Information Literacy*, 3(2), 66-71.

Aslett, Kevin, Andrew M Guess, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2022. "News credibility labels have limited average effects on news diet quality and fail to reduce misperceptions." *Science Advances* 8 (18):eabl3844.

Banet-Weiser, S. (2021). Misogyny and the politics of misinformation. In *The Routledge companion to media disinformation and populism* (pp. 211-220). Routledge.

Belkina, YU. A., & Kutsenko, Ye. V. (2014). Mem kak chast' internet-diskursa. *Azimut nauchnykh issledovaniy: pedagogika i psikhologiya*, (4 (9)), 77-79.

Birks, J. (2021). Fact-checking false claims and propaganda in the age of post-truth politics: The Brexit referendum. In *Research handbook on political propaganda* (pp. 390-404). Edward Elgar Publishing.

Böck, C., & Kettemann, M. C. (2024). Mapping the future of technological innovations. *Zenodo*. https://doi.org/10.5281/zenodo.11276069

Crowder-Meyer, M., & Ferrín, M. (2021). The effects of ideological and ethnoracial identity on political (mis) information. *Public Opinion Quarterly*, 85(3), 753-779.

Dias, Nicholas, Gordon Pennycook, and David G Rand. 2020. "Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media."

DiResta, R., Grossman, S., & Siegel, A. (2022). In-house vs. outsourced trolls: How digital mercenaries shape state influence strategies. *Political Communication*, 39(2), 222-253.

Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13-29.

Friesem, Y. (2019). Teaching truth, lies, and accuracy in the digital age: Media literacy as project-based learning. *Journalism & Mass Communication Educator*, 74(2), 185-198.

Ganesh, Bharath. 2020. "Weaponizing white thymos: Flows of rage in the online audiences of the alt-right." *Cultural Studies* 34 (6):892-924.

Ganesh, Bharath, and Nicolò Faggiani. 2024. "The flood, the traitors, and the protectors: affect and white identity in the Internet Research Agency's Islamophobic propaganda on Twitter." *Ethnic and Racial Studies* 47 (5):982-1008.

Graves, L., & Cherubini, F. (2016). The rise of fact-checking sites in Europe. *Digital News Project Report*.

Hagen, S. (2023). No space for Reddit spacing: mapping the reflexive relationship between groups on 4chan and Reddit. *Social Media+ Society*, 9(4), 20563051231216960.

Hamby, Anne, Hongmin Kim, and Francesca Spezzano. 2024. "Sensational stories: The role of narrative characteristics in distinguishing real and fake news and predicting their spread." *Journal of Business Research* 170:114289.

Hedling, E. (2024). Gendered Disinformation. In *Feminist Foreign Policy Analysis* (pp. 137-153). Bristol University Press.

Hinds, S. (2019). *The European Union approach to disinformation and misinformation: the case of the 2019 European Parliament elections* (Doctoral dissertation).

Kreps, S. E., & Kriner, D. L. (2022). The COVID-19 infodemic and the efficacy of interventions intended to reduce misinformation. *Public Opinion Quarterly*, 86(1), 162-175.

Kruglanski, A. W., Bélanger, J. J., & Gunaratna, R. (2019). *The three pillars of radicalization: Needs, narratives, and networks*. Oxford University Press.

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, 82(13), 1-26.

Lee, Jiyoung, Callie Kalny, Stefanie Demetriades, and Nathan Walter. 2023. "Angry Content for Angry People: How Anger Appeals Facilitate Health Misinformation Recall on Social Media." *Media Psychology*:1-27.

Leng, Yan, Yujia Zhai, Shaojing Sun, Yifei Wu, Jordan Selzer, Sharon Strover, Hezhao Zhang, Anfan Chen, and Ying Ding. 2021. "Misinformation during the COVID-19 outbreak in China: cultural, social and political entanglements." *IEEE Transactions on Big Data* 7 (1):69-80.

Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European review of social psychology*, 32(2), 348-384.

Lin, X., & Spence, P. R. (2018). Identity on social networks as a cue: Identity, retweets, and credibility. *Communication Studies*, 69(5), 461-482.

Luo, Yi Fang, Shu Ching Yang, and Seokmin Kang. 2022. "New media literacy and news trustworthiness: An application of importance–performance analysis." *Computers & Education* 185:104529.

Marecos, Joao, Duarte Tude Graça, Francisco Goiana-da-Silva, Hutan Ashrafian, and Ara Darzi. 2024. "Source Credibility Labels and Other Nudging Interventions in the Context of Online Health Misinformation: A Systematic Literature Review." *Journalism and Media* 5 (2):702-717.

Masip, P., Suau, J., Ruiz-Caballero, C., Capilla, P., & Zilles, K. (2021). News engagement on closed platforms. Human factors and technological affordances influencing exposure to news on WhatsApp. *Digital Journalism*, 9(8), 1062-1084.

Marwick, A. E., Losh, E., Schlüter, M., Markham, A., & Phipps, E. B. (2022). Feminist Approaches to Disinformation Studies. AoIR Selected Papers of Internet Research.

Mathisen, R. B. (2025, January 6). Growing Apart: Ideological Polarization between Teenage Boys and Girls. Working Paper. https://doi.org/10.31219/osf.io/7z2va

Molina, M. D. (2025). Do people believe in misleading information disseminated via memes? The role of identity and anger. *New Media & Society*, 27(2), 847-870.

Mont'Alverne, C., Arguedas, A. R., Banerjee, S., Toff, B., Fletcher, R., & Nielsen, R. K. (2024). The electoral misinformation nexus: how news consumption, platform use, and trust in news influence belief in electoral misinformation. *Public Opinion Quarterly*, 88(SI), 681-707.

Mourão, Rachel R, and Craig T Robertson. 2019. "Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information." *Journalism studies 20* (14):2077-2095.

Moutselos, M. (2025). Geographic variation in COVID-19 protests in the EU-27: Actors, grievances, and protest frames during a public health crisis. *Zenodo*. https://doi.org/10.5281/zenodo.14971670

Nedelcu, D. (2021). The role of source credibility and message credibility in fake news engagement. Perspectives from an experimental study. *Journal of Media Research*, 14(3), 42-62.

Neely, S. R., Eldredge, C., Ersing, R., & Remington, C. (2022). Vaccine hesitancy and exposure to misinformation: a survey analysis. *Journal of General Internal Medicine*, 1-9.

Nenadić, I. (2019). Unpacking the" European approach" to tackling challenges of disinformation and political manipulation. *Internet policy review*, 8(4), 1-22.

Pasquetto, Irene V, Briony Swire-Thompson, Michelle A Amazeen, Fabrício Benevenuto, Nadia M Brashier, Robert M Bond, Lia C Bozarth, Ceren Budak, Ullrich KH Ecker, and Lisa K Fazio. 2020. "Tackling misinformation: What researchers could do with social media data." *The Harvard Kennedy School Misinformation Review*.

Pretus, Clara, Camila Servin-Barthet, Elizabeth A Harris, William J Brady, Oscar Vilarroya, and Jay J Van Bavel. 2023. "The role of political devotion in sharing partisan misinformation and resistance to fact-checking." *Journal of Experimental Psychology: General*.

Saeed, M. H., Ali, S., Blackburn, J., De Cristofaro, E., Zannettou, S., & Stringhini, G. (2022, May). Trollmagnifier: Detecting state-sponsored troll accounts on Reddit. In *2022 IEEE symposium on security and privacy (SP)* (pp. 2161-2175). IEEE.

Rodríguez-Pérez, C., Sánchez-del-Vas, R., & Tuñón-Navarro, J. (2025). From Fact-Checking to Debunking: The Case of Elections24Check During the 2024 European Elections. *Media and Communication*, 13.

Staender, Anna, Edda Humprecht, Frank Esser, Sophie Morosoli, and Peter Van Aelst. 2022. "Is sensationalist disinformation more effective? Three facilitating factors at the national, individual, and situational level." *Digital Journalism* 10 (6):976-996.

Tripodi, F. B., Garcia, L. C., & Marwick, A. E. (2024). 'Do your own research': affordance activation and disinformation spread. *Information, Communication & Society*, 27(6), 1212-1228.

Strömbäck, J., Tsfati, Y., Boomgaarden, H., Damstra, A., Lindgren, E., Vliegenthart, R., & Lindholm, T. (2020). News media trust and its impact on media use: Toward a framework for future research. *Annals of the International Communication Association*, 44(2), 139-156.

Torchio, G. (2025). Europe's digital transition: Assessing the effectiveness of the 2030 Digital Decade Policy Programme. *Zenodo*. https://doi.org/10.5281/zenodo.14871590

Volk, S., De Jonge, L., & Rensmann, L. (2023). Populism and the pandemic: The politicization of COVID-19 and cleavage agency among populist radical right parties. *Zenodo*. https://doi.org/10.5281/zenodo.10259440

Walter, Nathan, Jonathan Cohen, R Lance Holbert, and Yasmin Morag. 2020. "Fact-checking: A meta-analysis of what works and for whom." *Political communication* 37 (3):350-375.

Weeks, Brian E. 2023. "Emotion, Digital Media, and Misinformation." *Emotions in the Digital World: Exploring Affective Experience and Expression in Online Interactions*:422.

Yin, Leon, Franziska Roscher, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2018. Your friendly neighborhood troll: The Internet Research Agency's use of

local and fake news in the 2016 US presidential campaign. SMaPP Data Report, Social Media and Political Participation Lab, New York ….

Young, Dannagal Goldthwaite, and Dannagal G Young. 2020. *Irony and outrage: The polarized landscape of rage, fear, and laughter in the United States*: Oxford University Press, USA.