International Journal of Communication 19(2025), 3559–3576 1932–8036/20250005

(Generative) AI and Disinformation

Introduction

AQSA FAROOQ
CLAES DE VREESE
University of Amsterdam, The Netherlands

Generative artificial intelligence (GenAI) represents a pivotal development in the contemporary information ecosystem. Large Language and Image Models now enable rapid and scalable creation of (hyper)realistic yet synthetic content. As these models become more accessible and sophisticated, so too do their capacities to distort public discourse, manipulate perceptions, and undermine trust in democratic institutions. At the same time, these technologies offer promising tools for detection, resilience building, and possibly countering falsehoods. As such, there is rapid global and cross-disciplinary interest in understanding how AI-driven tools have added another dimension to the existing challenge of disinformation. This Special Section brings together timely and original scholarship on this challenge, with the aim to explore the multifaceted role of AI, including how it can both contribute to, as well as potentially provide solutions to, the problem of disinformation.

Keywords: (generative) artificial intelligence, disinformation, journalism, review, regulation, citizens

In a world where public trust in scientific and democratic institutions is decreasing, the ability to access credible, verifiable, and reliable information is vital. Some suggest, however, that a new world emerged instead: a "post-truth" world in which ideologically driven opinions take precedence over expertise and evidence (Lewandowsky, Ecker, & Cook, 2017). Within this landscape of the current digital age, false information of various forms, modalities, and functions has revolutionized the way individuals engage with the content they consume. Whereas misinformation is a type of false information that can be spread with harmless or unknown intentions, disinformation is the kind that is specifically created to mislead and cause harm (European Commission, 2018; Wardle, 2018) and the latter can vary on dimensions of motive and facticity (Kapantai, Christopoulou, Berberidis, & Peristeras, 2021).

Aqsa Farooq: a.farooq@uva.nl
Claes de Vreese: c.h.devreese@uva.nl
Date submitted: 2025–11–09

For a while, artificial intelligence (AI) was regarded as a possible solution to the problem of misinformation and disinformation, and indeed, the automation, scalability, and potential objectivity offered by AI-driven technologies has had promise (Saeidnia et al., 2025). However, the worldwide rollout of accessible generative AI (or GenAI) tools in late 2022 unleashed a new, unprecedented application of AI that would allow individuals to generate outputs using simple text prompts within a matter of seconds (Combs, Moyer, & Bihl, 2024). Large language models (LLMs; e.g., OpenAI's ChatGPT) and image generator models (e.g., Stability AI's Stable Diffusion) have since become rapidly adopted and frequently used in various sectors of our everyday lives (Stokel-Walker & Van Noorden, 2023). These tools, when misused, have also further exacerbated the problem of disinformation. For instance, visual disinformation such as deepfakes have become far more sophisticated (Shoaib, Wang, Ahvanooey, & Zhao, 2023), whereas AI-generated content that circulates as disinformation online has caused widespread alarm (Burgess, 2023), particularly due to people's inability to detect their authenticity (Cooke, Edwards, Barkoff, & Kelly, 2024; Farooq & de Vreese, 2025).

That said, it is not always the user's deliberate *misuse* of the GenAI tool that results in misleading output; inherent properties of these tools can also contribute to the spread of disinformation. GenAI models are trained on pools of data that can contain prejudices or harmful stereotypes, leading to the generation of content that may reflect and reinforce these biases (Zhou, Abhishek, Derdenger, Kim, & Srinivasan, 2024). Furthermore, functional limitations can result in models yielding outputs that contain outright fabrications, also known as "confabulations" (Farquhar, Kossen, Kuhn, & Gal, 2024)—which, if not corrected, can also lead to the circulation of misleading information. Scholars, journalists, and AI experts have warned that GenAI may blur the boundaries between truth and fabrication in unprecedented ways, eroding trust in media and democratic institutions at a time of rapidly declining public confidence in traditional authorities (Bontcheva et al., 2024; Flynn, 2023). It is perhaps for these reasons that AI-generated misinformation and disinformation has topped the World Economic Forum's (2024, 2025) list of perceived short-term societal risks for two years in a row. Such risks, and their proximity, suggest a need for a more comprehensive understanding of how these technologies may be contributing to further maligning our information ecosystem, as well as establishing countermeasures and analyzing current and future policies. The European Union's (EU) AI Act is one such example of legislative interventions to ensure that AI-driven tools are safer and more responsible—nonetheless, there is still a long way to go before these technologies, and the companies that own them, launch products that are fully transparent, ethical, and accountable.

Another EU initiative, The European Digital Media Observatory (EDMO), is an interdisciplinary, multinational network of researchers, policy experts, media literacy practitioners, fact-checkers, and more, united by the common cause of tackling disinformation. In February 2024, the EDMO Conference on Disinformation took place, convening over 120 researchers from various fields of expertise to critically examine the evolving dynamics of disinformation (Farooq & de Vreese, 2024). Among the myriad topics of the panels and discussions, the prominence of GenAI was undeniable, both in its role as an aggravator and as a potential mitigator of disinformation. A dedicated panel on "Disinformation in an AI Age" showcased a growing appetite for research interrogating the capacities of AI in a post-truth world. The contributions featured here aim to build on that momentum, offering empirically rich and theoretically grounded insights into AI's role in the production, spread, and potential mitigation of disinformation. This Special Section

therefore aims to provide the latest investigations into this phenomenon from multiple perspectives and also highlight underexplored avenues of research that require scholarly attention in the upcoming future.

The GenAI and Disinformation Research Space

This Special Section of the *International Journal of Communication* is situated in the broader research space exploring the link between (Generative) AI and disinformation, which has unsurprisingly seen a recent uptick in contributions over the last few years alone. From a literature search of published,[1] English-language works focusing on GenAI and misinformation/disinformation, we found that the number of works published in 2024 was more than double the number of works published in the preceding years combined (see Figure 1).

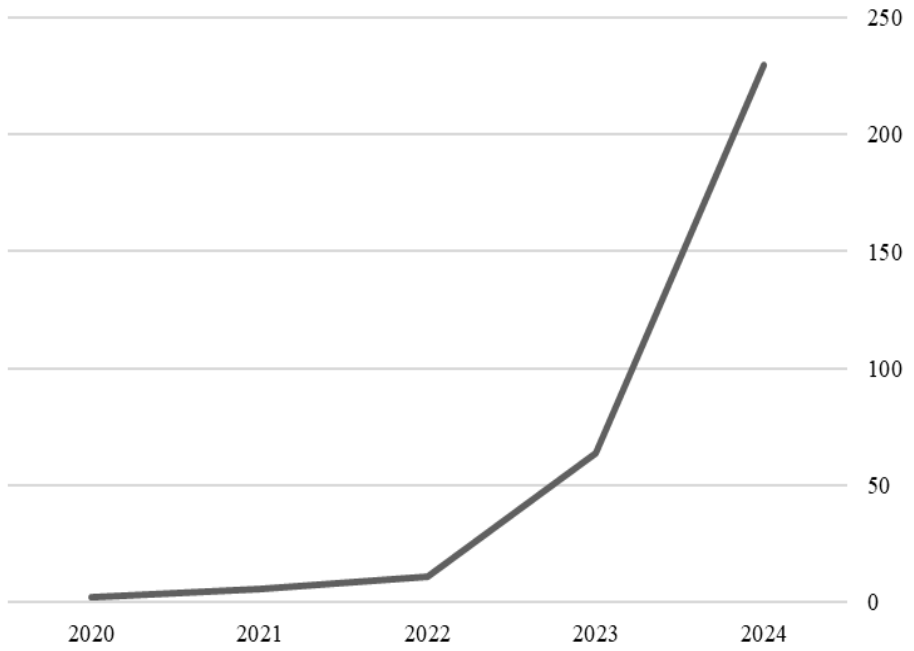

Figure 1. Number of works on GenAI and misinformation/disinformation by year of publication.

Researchers based in institutions within the United States were best represented in terms of authoring research on this topic (see Figure 2).

[1] We used Scopus to perform a literature search of published research in the English language. We searched for research with misinformation, disinformation, or "fake news" and also "generative AI" or AI-generated in either the title, abstract, or keywords. Our search was limited to conference papers, articles, reviews, and book chapters published in 2024 and earlier. This search yielded 314 documents, with the earliest publication from 2020.

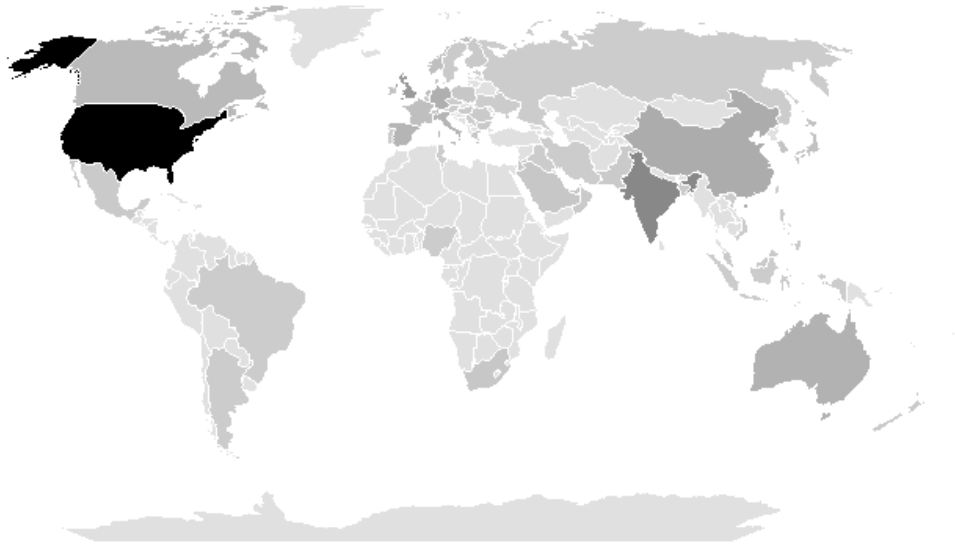

Figure 2. Geographical representation of the institutions of authors who have published works on GenAI and misinformation/disinformation.

Institutions in India, the United Kingdom, and China followed, with a total of 59 countries represented, indicating a burgeoning global interest on this topic. Most of the research within our search had been published in the subject area of computer science, followed by social sciences, and then engineering (see Figure 3).

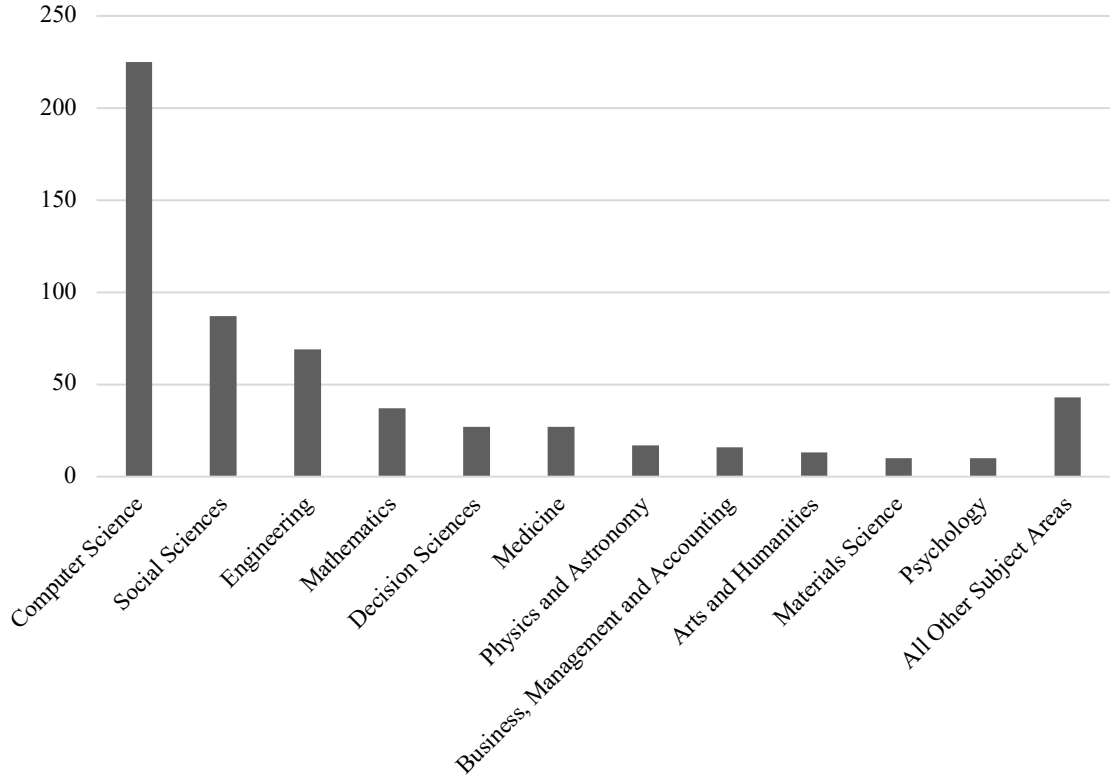

Figure 3. Number of works on GenAI and misinformation/disinformation by subject area. One publication could fall into more than one subject area.

Observing the main thematic focus of each publication, four categories could be identified as being the most prevalent (see Table 1).

Table 1. The Four Most Prominent Themes Explored in Research on GenAI and Misinformation/Disinformation.

Thematic Focus	Proportion of Works	Global South/non-Western focus	Multimodal (text, image, audio, video)	Real-world Implementation/ Field Validation
1. Impacts of Misinformation/ Disinformation	40%	Rare outside Western cases	Usually text or political deepfakes	Mostly experimental setups
2. Ethical, Societal and Regulatory Implications	31%	Predominantly Global North	Generally high-level discussions about GenAI broadly	Some empirical work, mostly case studies and surveys
3. Detection of Synthetic Content	26%	Geographical focus is rare	Mostly images, some video/deepfakes	Mostly lab tests/benchmarks
4. AI-Powered Solutions	4%	Nonspecific	Focused on countering multiple forms of disinformation	Some intervention testing, lacking scalability

Most commonly, research on this topic has focused on the *effects and consequences* of AI-generated misinformation/disinformation, usually through experimental research. Following this was research exploring the *ethical concerns*, *policy challenges, and societal implications* associated with generative AI and its potential to create disinformation. The third most prominent thematic focus was the *detection of synthetic content*, usually via the developing and testing of new techniques or methodologies for identifying AI-generated content. Research focusing on harnessing AI to *tackle disinformation* (both AI-generated and non–AI-generated) made up a small number of the publications in our search, suggesting a need for further research along the lines of this highly necessary yet underrepresented theme.

How GenAI Coverage Affects Citizens

When it comes to tackling disinformation, there used to be widespread agreement in the scholarship that one of the best ways to protect citizens from disinformation is by focusing on empowering them to make their own, informed conclusions (Bulger & Davison, 2018). Encouraging citizens to make educated and informed decisions based on evidence that they have assessed themselves has several benefits for the purpose of progressing toward an independently informed citizenry. In this sense, journalism and media and information literacy initiatives have become vital public communication and educational resources in the battle against disinformation (Anstead, Edwards, Livingstone, & Stoilova, 2025; Frau-Meigs, 2022). However, informing citizens about AI brings about a set of unique challenges. AI can be seen as a complex, opaque technology (Carabantes, 2020) and citizens have various preexisting attitudes toward AI, including who can be trusted to use it and for what purpose (Araujo, Helberger, Kruikemeier, & de Vreese, 2020;

Fletcher & Nielsen, 2024). Media coverage of AI, such as in recent years, has often focused on the negative outcomes of AI-driven technologies and AI-generated outputs—particularly in relation to amplifying misinformation and disinformation (see, for examples, "A.I.-Generated Content Discovered," 2023; BBC News, 2023; "Could ChatGPT Become," 2023; Klepper & Swenson, 2023; Menon & Khare, 2023). It has also been argued that the fears about GenAI's impact on the problem of misinformation have been greatly exaggerated by, among others, journalists (Simon, Altay, & Mercier, 2023). Against this backdrop, understanding the impact of negative news coverage of GenAI on citizens is paramount.

Dobber, Hameleers, Starke, and van der Meer set out to explore whether news articles covering the influence of GenAI on disinformation would increase reliance on journalism, depending on the framing of the coverage. From their experimental findings with a sample of Dutch citizens, they found that regardless of the framing, exposure to news coverage about GenAI did not increase trust in or reliance on journalism. Against their predictions, they instead found that being exposed to negatively framed news articles on the threat of AI-generated disinformation actually decreased participants' trust in journalism. Given the relatively high level of media trust in The Netherlands (Humprecht, Esser, & Van Aelst, 2020), this is a particularly significant finding that indicates the consequences of negatively framing public communication about GenAI and its role in the information ecosystem.

In line with the way the news media has covered GenAI by focusing on its negative potential, Tulin, Pantazi, Starke, Sivolap, and Dobber investigated the effects of preemptively warning citizens about AI-generated disinformation in the form of a GenAI literacy intervention. A representative sample of Canadian adults were exposed to a short explainer video about how GenAI tools can be used to create political disinformation. While watching the video did indeed result in a decrease in participants' perceptions of the credibility of disinformation, it also decreased their credibility perceptions of real news. As with Dobber and colleagues, these studies reveal the negative consequences of communicating the dangers of GenAI to the public, particularly in relation to journalism. For the purpose of maintaining the importance of journalism and literacy initiatives in an information landscape marred with disinformation, the findings of these studies provide much-needed insight into how citizens may be influenced by the nature of AI-generated disinformation coverage in the news.

AI in the Political Sphere

The conversation about AI in recent years has been dominated by concerns around the possibility of AI-generated content proliferating the political landscape. Ahead of the 2024 election year, where 64 countries held elections in which around half of the world's population could vote, speculation about the potential for a flood of AI-generated political content influencing voters and compromising democracies was rife (de Vreese, 2024; Molęda-Zdziech et al., 2024; Simon & Altay, 2025). Across the 50 countries holding national elections in 2024, analyses showed that a vast majority (80%) had cases of election-related multimodal GenAI content (Trauthig, Valenzuela, & Howard, 2025). Of these cases, almost three-fourths (69%) was described as having a negative impact on elections, such as by propagating harmful or misleading narratives. However, positive uses of GenAI in electoral campaigns were also noted, though such examples are harder to come by (OECD, 2025). Therefore, ongoing global and balanced assessments of the role of AI-driven disinformation are needed in order to judge their impact on electoral processes and democracy

more broadly. In their systematic review, Grub and Humprecht collated scientific articles that have investigated the role of AI in relation to online disinformation in political contexts. They identified key trends and highlighted gaps in the literature, for instance by spotlighting how bot networks have been used repeatedly across political campaigns, and yet remain understudied in relation to their impact on political behavior. This review, thus, provides a much-needed overview of the current academic knowledge in this space, laying the groundwork for future work.

One of the most prevalent and effective forms of AI-generated political disinformation is in the form of a deepfake. By combining both visual and audio modalities, these synthetic videos are created using machine learning algorithms, with particularly alarming levels of convincingness (Lee & Shin, 2022). Visual information—such as in the form of a deepfake—tends to be processed with far more ease and is subsequently better able to be retrieved from memory (Dan et al., 2021; Stenberg, 2006). As such, due to the availability of copious visual material of public figures, creators and disseminators of deepfakes have certainly exploited political actors to a notable extent, indeed contributing to the declining trust and uncertainty among the public with regards to accessing accurate information (Vaccari & Chadwick, 2020). In experimental settings, political deepfakes can have marginal effects on individuals' perceptions of politicians (Dobber, Metoui, Trilling, Helberger, & de Vreese, 2021) and research has also implicated politically interested individuals to be those most likely to also share political deepfakes (Ahmed, 2021). In recent high-stakes elections, deepfakes have been a clear weapon of choice, used both to mislead voters about political candidates (Scott, 2024) and used *by* political candidates to mislead voters (Galer, 2024). Deepfakes are therefore one of the forms of AI-enabled disinformation that provoke widespread fears about their impact on democracy (Łabuz & Nehring, 2024).

In their study, Wack and Parry investigated the influence of the ethnic makeup of targeted deepfakes on participants' support for the Cape independence movement in South Africa. The findings of their experiment revealed that despite the independence movement being predominantly relevant to the white population of the Western Cape, when a message of support via a deepfake was delivered by a non-White synthetic avatar, participants indicated increased support for the movement—albeit to little effect. A more pronounced effect on support for the movement was found when Black participants were shown a Black synthetic avatar in the deepfake message. This study provides much-needed insight into the sociopolitical dynamics of a region vastly underrepresented in the disinformation literature (Broda & Strömbäck, 2024) while also highlighting how political movements may be able to utilize GenAI to target individuals and garner support in unprecedented ways.

When it comes to mitigation strategies against AI-generated political disinformation, the need for AI literacy has been emphasized (Chu-Ke & Dong, 2024) and has also been evaluated in this collection of works. In Tulin and associates, their GenAI literacy intervention was more effective in inducing skepticism among citizens regarding the veracity of news when it involved information about generating political disinformation rather than information about generating art. In Wack and Parry, AI literacy was a better predictor of citizens' ability to identify AI-generated deepfakes compared to digital literacy. However, when the participants of their study shared an identity characteristic with the avatar depicted in the deepfake, AI literacy was linked to being less likely to identify the deepfake as being AI-generated. These works suggest that in order to address AI-generated political disinformation, the efficacy of strategies such as increasing

AI literacy may also come about with caveats, and further evidence is needed to understand how best to successfully apply such strategies.

Exploring AI as the Antidote to Disinformation

Harnessing AI as a potential tool against disinformation has indeed been widely discussed and attempted. A systematic review by Saeidnia and colleagues analyzed studies over a ten-year period from 2014–2024 in which AI-driven technologies were enlisted to "detect, analyze, and mitigate the impact of misinformation" (Saeidnia et al., 2025, p. 3139). Their review of 76 studies revealed a wide range of ways in which the abilities of AI systems could be leveraged to identify and tackle misinformation and disinformation at scale and across many social media platforms. On the one hand, using AI systems to moderate copious amounts of false online content, including disturbing and illicit content, can present the ideal solution to having humans endure this process (Gillespie, 2020). On the other hand, the biases in the training data of the AI systems may simply result in the regurgitation of harmful representations, even for such moderating or classifying tasks (Binns, Veale, Van Kleek, & Shadbolt, 2017; Gorwa, Binns, & Katzenbach, 2020). There is therefore a crucial need to advance scholarship around how AI systems can be used to tackle disinformation to better establish the potentials and shortcomings of such technologies.

In Larraz, Salaverría, and Serrano-Puche's investigation, journalists in collaboration with computer scientists enlisted AI to automate the analysis of disinformation narratives in relation to a specific case in Europe. When compared, LLM-generated narrative analyses aligned well with manual narrative analyses, suggesting a complementary methodology for unpacking disinformation narratives at scale. Though the authors highlight the limitations of the automated analyses, such as the lack of specificity and the propensity to be too literal in interpretations, their conclusions suggest a space for AI-assisted journalistic practices as a means for tackling disinformation on a large scale.

In another study comparing the efficiency of humans to AI-driven operations, Tang, Burston, Sun, Vraga, and Bode experimentally explored the effects of presenting U.S.-based parents with vaccine misinformation corrections from either a bot or a user on the platform Reddit. Their findings indicated that in the first round of corrections, a bot was effective in correcting misperceptions, but a user's correction was not. However, in a second round of misinformation exposure, bots were no longer effective in correcting misinformed beliefs, suggesting their inability to correct misinformation after a certain point. This research has implications for the efficacy of AI-assisted corrections of misinformed beliefs, and in particular, the shortcomings of AI as a tool against misinformation when false beliefs are already entrenched—such as for topics like vaccines. For the tackling of disinformation using interactive and conversational forms of AI-driven technologies, this research makes a vital contribution to the current state of knowledge.

Journalism: Under Fire and on the Frontlines

Amid the battle against disinformation, journalism stands out as one of the key democratic institutions on the frontlines. Nonetheless, faltering trust in news organizations across the world (Park et al., 2024), aided by populist politicians' attacks on journalists accused of being part of the very elite responsible for disseminating disinformation (Egelhofer, 2023), has left citizens disillusioned by the

disinformation problem and seeking out alternative media outlets for their news (Hameleers, Brosius, & de Vreese, 2022). With the dawn of accessible AI and the advanced generative capabilities of AI systems, journalism and journalists have had to reckon with the challenge of reporting on an unprecedented scale of misleading content, including during major political events (Stock, 2024). Simultaneously, journalists around the world have also started adopting AI technologies within their practices (Adjin-Tettey, Muringa, Danso, & Zondi, 2024; Brigham, Gao, Kohno, Roesner, & Mireshghallah, 2024; Dodds, Zamith, & Lewis, 2025) such as for fact-checking (Dierickx, Lindén, & Opdahl, 2023). Citizens, however, remain unconvinced about the adoption of GenAI in newsrooms—largely, trust in news organizations' responsible use of GenAI was below average across six different democracies, though human journalists were still preferred to AI as sources of news (Fletcher & Nielsen, 2024). This marks a pivotal moment for journalism and its role in protecting citizens against disinformation.

In Mattis and de Vreese's literature review on the impact of GenAI on journalism, they reflect on the possible risks and benefits of GenAI adoption in newsrooms for the purpose of addressing disinformation. While automation may facilitate the rate of certain practices, such as fact-checking, other practices, such as news gathering using GenAI, may result in the propagation of harmful and false narratives. This review is therefore a timely exploration of the intersection of GenAI and journalism and what it means for disinformation.

There may, however, be a need for journalists to adapt in different ways. In Larraz and colleagues' investigation, the collaboration between journalists and computer scientists is spotlighted as being key to the successful implementation of AI-assisted disinformation narrative analysis. For citizens to return to journalism in the age of AI and AI-enabled information threats, Dobber and associates' study has significant implications, suggesting that the negative framing of the threat to citizens' information ecosystem may not be the most conducive to engendering trust in journalism. Rather, bringing attention to journalism's role in *protecting* citizens in the AI age may be a more valuable strategy.

Safeguarding Citizens Through Scalable Solutions

For citizens navigating an environment that they already deem to be marred with disinformation, the prospect of AI-enabled technologies capable of producing hyperrealistic content at a large scale is more daunting than ever. Such sentiment has been captured in surveys of citizens from eight countries (Argentina, Brazil, Germany, Japan, South Korea, Spain, the United Kingdom, and the United States), a vast majority of which (87%) claimed their top concern was online misinformation, followed by AI-generated false content (Ejaz, Fletcher, Nielsen, & McGregor, 2024). As for platforms, their efforts to moderate false content at scale with support from AI systems has been criticized as being more likely to worsen existing issues rather than find solutions for them (Gorwa et al., 2020). Within this landscape, the need for scalable solutions to the problem of disinformation has never been more apparent.

One such solution may be through regulation. Moleón presents an analysis of the policy documents from European institutions and platforms involved in the EU's Strengthened Code of Best Practices on Disinformation (CoP). Using sociotechnical imaginaries as a theoretical framework, Moleón concludes that the European Union's strategy prioritizes scalability and efficiency using AI over a more thoughtful and

deliberate consideration of strategies that take into consideration the rigidity and unpredictability of such technologies, as well as the ever-changing political landscape.

Experimental explorations are also much needed for understanding the efficacy of scalable solutions to the problem of online disinformation. For instance, in Tang and cohorts' experiment involving AI-powered misinformation correction, bots are investigated as a tool for countering false narratives online, and their potential to evolve into conversational partners for the purpose of alleviating misperceptions is also discussed. It is, nonetheless, altogether clear that long-term, scalable solutions require much more scholarly attention and scrutiny, and with citizens' best interests always at the center.

The Dual Role of AI

Finally, this Special Section is also enriched by contributions which provide overarching analyses of the role of AI in disinformation research and disinformation governance. As AI technologies become more sophisticated and their capabilities harnessed into tools accessible to wider sections of society, it is of utmost importance to gather the evidence so far and identify patterns as well as gaps. In Grub and Humprecht's systematic literature review of the disinformation research, Mattis and de Vreese's (202X) literature review, and in Moleón's examination of the EU's Strengthened Code of Best Practices on Disinformation (CoP), the role of AI has been described as twofold: as a contributor as well as a solution. This, of course, reflects the nature of this Special Section as well. Grub and Humprecht highlight the need for more synergies between computer and social sciences, particularly in relation to strengthening and building upon theoretical frameworks. Mattis and de Vreese stress the importance for journalists to uphold their position as a key tenet of democracy, particularly in the wake of widespread platform dependence. Lastly, Moleón concluded that policy decisions around disinformation and AI must integrate both democratic principles and technological innovation. Thus, all three of these analyses identify areas of their domains that require more collaboration and unity for the purpose of promoting long-term resilience in the fight against disinformation.

The Contribution of This Special Section

This Special Section brings together the latest scholarship on one of the most urgent societal issues of our time. Overall, this collection of articles underscores both the negative potential of AI-driven technologies, such as by opportunistically targeting citizens for political campaigns (Wack and Parry), while also highlighting their promise in assisting the work of those tackling disinformation at scale (Larraz et al.; Tang et al.). Although, informing citizens about the misuse of GenAI may in turn lead to the reduced efficacy of literacy interventions and journalism (Dobber et al.; Tulin et al.). Finally, analyses of the overarching role of AI in political contexts, journalism and governance reveals both the dual nature of AI in the fight against disinformation, and the need for more collaborative efforts (Grub and Humprecht; Mattis and de Vreese; Moleón). See Table 2 for a summary of the articles.

Table 2. Articles in the Special Section.

Article	Methodology	Outcomes Explored
Dobber et al.	Experiment	1. Trust in journalism 2. Intended use of journalistic sources 3. Importance of journalistic principles
Grub and Humprecht	Systematic Review	Key coding variables: 1. Focus: dissemination and generation (looking at target groups, disseminators, strategic use, impact, effect, and countermeasures) 2. Focus: detection (stakeholders, models, practical applications, reliability, language use, platforms, and limitations)
Larraz et al.	Case Study	Fact checks about the farmer's protest in Europe
Mattis and de Vreese	Literature Review	1. Public perceptions of GenAI in Journalism 2. Current adoption, risks and opportunities of GenAI in journalism according to news media professionals 3. Practical applications of GenAI in Journalism 4. Broader discourse and debates around GenAI in journalism
Moleón	Qualitative Documentary Analysis	Eight analytical dimensions: 1. Type of framing (AI as a solution, problem or dual approach) 2. Negative perception (scalability, erosion of trust, etc.) 3. Function attributed to AI (moderation, verification, tracking, etc.) 4. Projected image (explainable and fair AI, inevitable solution, etc.) 5. Level of automation (total, with human verification, complementary) 6. Narrative justification (efficiency, reliability, protection, etc.) 7. Responsible actor (platforms, EU institutions, verifiers) 8. Adopted tone (institutional, metaphorical)
Tang et al.	Experiment	1. Belief accuracy in false claims 2. Attitudes toward the HPV vaccines
Tulin et al.	Experiment	Truth ratings
Wack and Parry	Experiment	Party support measures

This Special Section, while paving the way for future explorations, also brings to light underexplored areas of academic inquiry that require attention. For instance, many online spaces (e.g., user comment

sections online, search engines, emerging social media platforms) that may harbor disinformation have become either infiltrated by or integrated with AI technologies (Memon & West, 2024). What effect could this rapid inclusion of AI have on the level of disinformation in these spaces? With social media platforms even enlisting LLMs to generate notes that provide context to misleading social media posts (Li et al., 2025), including those generated with AI, there is an urgent need to explore the impact of using LLMs as a content moderation tool while issues inherent with LLMs' lack of transparency and accountability persist (Gorwa et al., 2020). Furthermore, as individuals move to LLMs for information retrieval (Gillespie, Lockey, Ward, Macdade, & Hassed, 2025), how do the limitations within these models affect the most vulnerable populations in society, and what implications does this have for efforts to counter disinformation? AI agents, such as chatbots and virtual assistants, may be used to verify information about politically salient topics, though performances can vary according to the LLM (Kuznetsova et al., 2025). This raises questions about how individuals perceive the trustworthiness of such LLM-driven tools and how they establish which ones are more trustworthy than others. These identified research gaps in the current literature present a host of opportunities for future research within this space. Overall, understanding how new and evolving uses of AI-driven and AI-integrated technologies may be influencing the spread of disinformation is a particularly relevant avenue of future research at this stage.

References

Adjin-Tettey, T. D., Muringa, T., Danso, S., & Zondi, S. (2024). The role of artificial intelligence in contemporary journalism practice in two African countries. *Journalism and Media*, *5*(3), Article 3. doi:10.3390/journalmedia5030054

Ahmed, S. (2021). Who inadvertently shares deepfakes? Analyzing the role of political interest, cognitive ability, and social network size. *Telematics and Informatics*, *57*. doi:10.1016/j.tele.2020.101508

A.I.-generated content discovered on news sites, content farms and product reviews. (2023, May 19). *The New York Times*. Retrieved from https://www.nytimes.com/2023/05/19/technology/ai-generated-content-discovered-on-news-sites-content-farms-and-product-reviews.html

Anstead, N., Edwards, L., Livingstone, S., & Stoilova, M. (2025). The potential for media literacy to combat misinformation: Results of a rapid evidence assessment. *International Journal of Communication*, *19,* 2129–2151.

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. doi:10.1007/s00146-019-00931-w

BBC News. (2023, July 7). *Martin Lewis felt "sick" seeing deepfake scam ad on Facebook.* Retrieved from https://www.bbc.com/news/uk-66130785

Binns, R., Veale, M., Van Kleek, M., & Shadbolt, N. (2017). Like trainer, like bot? Inheritance of bias in algorithmic content moderation. In G. L. Ciampaglia, A. Mashhadi, & T. Yasseri (Eds.), *Social informatics* (Vol. 10540, pp. 405–415). Oxford, UK: Springer International. doi:10.1007/978-3-319-67256-4_32

Bontcheva, K., Papadopoulous, S., Tsalakanidou, F., Gallotti, R., Dutkiewicz, L., Krack, N., . . . Srba, I. (2024). *Generative AI and disinformation: Recent advances, challenges, and opportunities*. Retrieved from https://lirias.kuleuven.be/retrieve/758830

Brigham, N. G., Gao, C., Kohno, T., Roesner, F., & Mireshghallah, N. (2024). *Developing story: Case studies of generative AI's use in journalism*. doi:10.48550/arXiv.2406.13706

Broda, E., & Strömbäck, J. (2024). Misinformation, disinformation, and fake news: Lessons from an interdisciplinary, systematic literature review. *Annals of the International Communication Association*, *48*(2), 139–166. Doi:10.1080/23808985.2024.2323736

Bulger, M., & Davison, P. (2018). The promises, challenges, and futures of media literacy. *Journal of Media Literacy Education*, *10*(1), 1–21. doi:10.23860/jmle-2018-10-1-1

Burgess, M. (2023, May 21). AI: Can you tell that these "photographs" are artificial? *BBC News*. Retrieved from https://www.bbc.com/news/uk-wales-65562658

Carabantes, M. (2020). Black-box artificial intelligence: An epistemological and critical analysis. *AI & SOCIETY*, *35*(2), 309–317. doi:10.1007/s00146-019-00888-w

Chu-Ke, C., & Dong, Y. (2024). Misinformation and literacies in the era of generative artificial intelligence: A brief overview and a call for future research. *Emerging Media*, *2*(1), 70–85. doi:10.1177/27523543241240285

Combs, K., Moyer, A., & Bihl, T. J. (2024). Uncertainty in visual generative AI. *Algorithms*, *17*(4), 136–156.

Cooke, D., Edwards, A., Barkoff, S., & Kelly, K. (2024). *As good as a coin toss: Human detection of AI-generated images, videos, audio, and audiovisual stimuli.* Retrieved from http://arxiv.org/abs/2403.16760

Could ChatGPT become a monster misinformation superspreader? (2023). *NewsGuard*. Retrieved from https://www.newsguardtech.com/misinformation-monitor/jan-2023

Dan, V., Paris, B., Donovan, J., Hameleers, M., Roozenbeek, J., van der Linden, S., & von Sikorski, C. (2021). Visual mis- and disinformation, social media, and democracy. *Journalism & Mass Communication Quarterly*, *98*(3), 641–664. doi:10.1177/10776990211035395

de Vreese, C. H. (2024, May). *Keynote to EDMO Annual Conference, Brussels.* Retrieved from https://edmo.eu/edmo-news/2024-edmo-annual-conference-report-and-video/

Dierickx, L., Lindén, C.-G., & Opdahl, A. L. (2023). Automated fact-checking to support professional practices: Systematic literature review and meta-analysis. *International Journal of Communication*, *17*, 5170–5190.

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *International Journal of Press/Politics*, *26*(1), 69–91. doi:10.1177/1940161220944364

Dodds, T., Zamith, R., & Lewis, S. C. (2025). The AI turn in journalism: Disruption, adaptation, and democratic futures. *Journalism*, 1–15. Advance online publication. doi:10.1177/14648849251343518

Egelhofer, J. L. (2023). How politicians' attacks on science communication influence public perceptions of journalists and scientists. *Media and Communication*, *11*(1), 361–373. doi:10.17645/mac.v11i1.6098

Ejaz, W., Fletcher, R., Nielsen, R. K., & McGregor, S. (2024). *What do people want? Views on platforms and the digital public sphere in eight countries*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/what-do-people-want-views-platforms-and-digital-public-sphere-eight-countries

European Commission: Publications Office of the European Union. (2018). *A multi-dimensional approach to disinformation: Report of the independent high-level group (HLEG) on fake news and online disinformation.* Retrieved from https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation

Farooq, A., & de Vreese, C. H. (2024, May 30). *Reflecting on the success of the EDMO Scientific Conference on disinformation.* Retrieved from https://edmo.eu/blog/reflecting-on-the-success-of-the-edmo-scientific-conference-on-disinformation/

Farooq, A., & de Vreese, C. H. (2025). Deciphering authenticity in the age of AI: How AI-generated disinformation images and AI detection tools influence judgements of authenticity. *AI & SOCIETY*. Advance online publication. doi:10.1007/s00146-025-02416-5

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, *630*(8017), 625–630.

Fletcher, R., & Nielsen, R. (2024). *What does the public in six countries think of generative AI in news?* Retrieved from https://ora.ox.ac.uk/objects/uuid:bad671a6-860d-4513-951b-ef9cb2ac454b

Flynn, D. J. (2023). *The role of an informed public in democratic systems.* Retrieved from https://www.ie.edu/cgc/research/tech4democracy/

Frau-Meigs, D. (2022). How disinformation reshaped the relationship between journalism and media and information literacy (MIL): Old and new perspectives revisited. *Digital Journalism*, *10*(5), 912–922. doi:10.1080/21670811.2022.2081863

Galer, S. S. (2024, August 22). If you are outraged by Trump's use of AI and deepfakes, don't be—that's exactly what he wants. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/article/2024/aug/22/outraged-donald-trump-artificial-intelligence-deepfakes-taylor-swift

Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025*. Retrieved from https://mbs.edu/faculty-and-research/trust-and-ai

Gillespie, T. (2020). Content moderation, AI, and the question of scale. *Big Data & Society*, *7*(2). doi:10.1177/2053951720943234

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1). doi:10.1177/2053951719897945

Hameleers, M., Brosius, A., & de Vreese, C. H. (2022). Whom to trust? Media exposure patterns of citizens with perceptions of misinformation and disinformation related to the news media. *European Journal of Communication*, *37*(3), 237–268. doi:10.1177/02673231211072667

Humprecht, E., Esser, F., & Van Aelst, P. (2020). Resilience to online disinformation: A framework for cross-national comparative research. *International Journal of Press/Politics*, *25*(3), 493–516. doi:10.1177/1940161219900126

Kapantai, E., Christopoulou, A., Berberidis, C., & Peristeras, V. (2021). A systematic literature review on disinformation: Toward a unified taxonomical framework. *New Media & Society*, *23*(5), 1301–1326. doi:10.1177/1461444820959296

Klepper, D., & Swenson, A. (2023, May 14). AI-generated disinformation poses threat of misleading voters in 2024 election. *PBS NewsHour.* Retrieved from https://www.pbs.org/newshour/politics/ai-generated-disinformation-poses-threat-of-misleading-voters-in-2024-election

Kuznetsova, E., Makhortykh, M., Vziatysheva, V., Stolze, M., Baghumyan, A., & Urman, A. (2025). In generative AI we trust: Can chatbots effectively verify political information? *Journal of Computational Social Science*, *8*(15), 1–31. doi:10.1007/s42001-024-00338-8

Łabuz, M., & Nehring, C. (2024). On the way to deep fake democracy? Deep fakes in election campaigns in 2023. *European Political Science*, *23*(4), 454–473. doi:10.1057/s41304-024-00482-9

Lee, J., & Shin, S. Y. (2022). Something that they never said: Multimodal disinformation and source vividness in understanding the power of AI-enabled deepfake news. *Media Psychology*, *25*(4), 531–546. doi:10.1080/15213269.2021.2007489

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. doi:10.1016/J.JARMAC.2017.07.008

Li, H., De, S., Revel, M., Haupt, A., Miller, B., Coleman, K., . . . & Bakker, M. A. (2025). *Scaling human judgment in community notes with LLMs*. doi:10.48550/arXiv.2506.24118

Memon, S. A., & West, J. D. (2024). *Search engines post-ChatGPT: How generative artificial intelligence could make search less reliable*. doi:10.48550/arXiv.2402.11707

Menon, S., & Khare, V. (2023, May 31). Wrestlers' protest: The fake smiles of India's detained sporting stars. *BBC News*. Retrieved from https://www.bbc.com/news/world-asia-india-65757400

Molęda-Zdziech, M., Kiljan, K., Łabuz, M., Wójtowicz, A., Szymik, J., Nawrotkiewicz, J., & Szkoła, P. (2024). *Things to watch for in 2024 European Parliament elections*. Retrieved from https://ceeddw.org/wp-content/uploads/2024/04/ThingsToWatchFor.pdf

OECD. (2025). Governing with artificial intelligence: The state of play and way forward in core government functions. Retrieved from https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/06/governing-with-artificial-intelligence_398fa287/795de142-en.pdf

Park, S., Fisher, C., Tandoc, E., Dulleck, U., Yao, S. P., & Lukamto, W. (2024). The relationship between news trust, mistrust and audience disengagement. *Journalism*. Advance online publication. doi:10.1177/14648849241299775

Saeidnia, H. R., Hosseini, E., Lund, B., Tehrani, M. A., Zaker, S., & Molaei, S. (2025). Artificial intelligence in the battle against disinformation and misinformation: A systematic review of challenges and approaches. *Knowledge and Information Systems*, *67*(4), 3139–3158. doi:10.1007/s10115-024-02337-7

Scott, M. (2024, April 16). Deepfakes, distrust and disinformation: Welcome to the AI election. *POLITICO*. Retrieved from https://www.politico.eu/article/deepfakes-distrust-disinformation-welcome-ai-election-2024/

Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023, November). Deepfakes, misinformation, and disinformation in the era of frontier AI, generative AI, and large AI models. *2023 International Conference on Computer and Applications (ICCA)*, 1–7. doi:10.1109/ICCA59364.2023.10401723

Simon, F. M., & Altay, S. (2025, July 7). *Don't panic (yet): Assessing the evidence and discourse around generative AI and elections*. Retrieved from http://knightcolumbia.org/content/dont-panic-yet-assessing-the-evidence-and-discourse-around-generative-ai-and-elections

Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. doi:10.37016/mr-2020-127

Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology*, *18*(6), 813–847. doi:10.1080/09541440500412361

Stock, E. (2024, May 22). How to cover elections in 2024? Journalists weigh in on AI, disinfo and more. *International Center for Journalists*. Retrieved from https://www.icfj.org/news/how-cover-elections-2024-journalists-weigh-ai-disinfo-and-more

Stokel-Walker, C., & Van Noorden, R. (2023). What ChatGPT and generative AI mean for science. *Nature*, *614*(7947), 214–216. doi:10.1038/d41586-023-00340-6

Trauthig, I., Valenzuela, S., & Howard, P. (2025). *Generative AI in electoral campaigns: Mapping global patterns*. International Panel on the Information Environment (IPIE). Retrieved from https://www.ipie.info/research/sfp2025-1

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1), 1–13. doi:10.1177/2056305120903408

Wardle, C. (2018). *Information disorder: The essential glossary*. Harvard, MA: Shorenstein Center on Media, Politics, and Public Policy, Harvard Kennedy School.

World Economic Forum. (2024). *Global risks report 2024*. Retrieved from https://www.weforum.org/publications/global-risks-report-2024/

World Economic Forum. (2025). *Global risks report 2025*. Retrieved from https://www.weforum.org/publications/global-risks-report-2025/

Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). *Bias in generative AI*. doi:10.48550/arXiv.2403.02726