SAFEGUARDING INFORMATION INTEGRITY IN THE AGE OF AI: PERSPECTIVES FROM INDIA FOR THE GLOBAL SOUTH

RANJINI RAGHAVENDRA

REFLECTIVE PIECE

SAFEGUARDING INFORMATION INTEGRITY IN THE AGE OF AI: PERSPECTIVES FROM INDIA FOR THE GLOBAL SOUTH

Ranjini Raghavendra

Ranjini Canchi Raghavendra is an advisor at the Public Affairs Centre (PAC) and a Professor of Public Policy at R V University, India. She is a scholar with over 20 years of experience in academia, research, and journalism. She holds a PhD in Health Informatics from Lancaster University.

Abstract

Information integrity refers to the accuracy, consistency, and reliability of information (United Nations, 2023). It is threatened by information pollution, which takes various forms, including misinformation, disinformation, malinformation, fake news, and hate speech. While information pollution has a long history, generative artificial intelligence (GAI) has added a new dimension to the problem. GAI is unregulated and freely available, making it easy to create and spread false information. This creates an urgent need to improve monitoring and accountability mechanisms aimed at safeguarding information integrity. This reflective piece analyses the consequences of information pollution in the spheres of politics and health in India and discusses key lessons for the Global South. Based on a literature review and key informant interviews, the paper recommends practical solutions: multi-stakeholder engagement, oversight mechanisms, and accountability guidelines for intermediaries. The reflective piece also argues for the creation of a Global South consortium to combat information pollution and calls for enhanced global collaboration on prebunking, media and information literacy, and fact-checking mechanisms.

Keywords

Misinformation; disinformation; information pollution; India; Global South; artificial intelligence; social media.

Evidence for decision-making

1. Misinformation, disinformation, malinformation, fake news, and hate speech (all examples of information pollution) are a global risk, hindering human progress and need to be addressed immediately across the Global South.

2. Strategies to safeguard information integrity in the face of new technologies, such as generative artificial intelligence (GAI), are very limited in the Global South. Efforts to strengthen these strategies should focus particularly on ensuring information integrity for vulnerable groups such as minorities, migrants, low-literacy, non-English-speaking, and Indigenous groups, especially given rising majoritarianism.

3. There is a lack of international cooperation to address the issue. The reflective piece proposes a high-level Global South consortium to fund research and advocacy on information pollution and develop context-specific guidelines and policies. Users may be empowered by improved media and information literacy through the development of sophisticated tools and initiatives to address information pollution.

Introduction

In the run-up to the 2024 parliamentary elections in India, an inaccurate message misinforming voters that they could cast their vote even if their name was not on the voter list went viral. The Election Commission of India later clarified that there is no such 'challenge vote' and that anyone not on the voter list cannot vote (First Post, 2024).

In May- June of 2018, misinformation about child abductors was shared on WhatsApp and circulated in various Indian states, including in vernacular languages, with gory images and videos of alleged child kidnappers being assaulted by the public. This fake news resulted in the death of more than 20 innocent people by lynching (Sinha et al., 2019).

It is claimed that drinking lemon juice mixed with baking soda or aspirin will cure COVID-19. This is false (National Academies, 2020). Misleading messages such as these have become an increasingly familiar sight on social media and messaging platforms. These are examples of information pollution and its possible consequences for individuals and society more widely. Reliable and accurate information is crucial for human progress. Information pollution damages the foundations of human development and directly hinders the achievement of the 2030 Agenda. For instance, during the COVID-19 pandemic, an 'infodemic'[1] of mis- and disinformation undermined public health measures and vaccination drives (World Health Organisation [WHO], 2020).

Propaganda and political polarisation, driven by the pursuit of political advantage, are among the key catalysts of information pollution in the political sphere.

1. This refers to excess information (including false or misleading information) that spreads in digital and physical environments during a public health emergency (Wilhelm et al., 2023).

Information pollution has a long history and takes various forms, including misinformation,[2] disinformation,[3] malinformation,[4] fake news,[5] and hate speech.[6] However, generative artificial intelligence (GAI) is fundamentally changing how information is created, distributed, and consumed. GAI systems can generate text (such as Google's Gemini, Meta's LLaMa, or OpenAI's ChatGPT), visuals (such as Stable Diffusion or OpenAI's DALL-E), or audio (such as Microsoft's VALL-E) by applying machine learning to large quantities of training data (Simon et al., 2023). The output is generated with great speed and ease and is now so sophisticated that it is often almost impossible to tell if content is human- or AI-generated (Groh et al., 2022). The inability of the public to discern which information is accurate and trustworthy becomes a pressing challenge.

As new technologies are making it easier than ever to create and disseminate content, misinformation and disinformation have been identified as the biggest short-term risks facing the world (World Economic Forum, 2024). Experts surveyed for the 2024 Global Risk Report chose misinformation and disinformation as the number one risk in India, ahead of infectious diseases, illicit economic activity, and labour shortages (World Economic Forum, 2024). In a recent report, Blair et al. (2023) note that there is an acute need for research on information pollution in the Global South, where it has proven to be a serious challenge with devastating consequences. Further, because of the lack of studies on countering information pollution in the Global South, there is a risk that findings from the Global North will be applied to the Global South, or that Global South countries will be generalised. This gap in scientific knowledge is alarming, especially given that strategies and interventions that work in the Global North might not be effective in Global South contexts.

---

2. Misinformation is the unintentional spread of inaccurate information shared in good faith by those unaware that they are passing on falsehoods (United Nations [UN], 2023).
3. Disinformation refers to false or misleading content that can cause harm, irrespective of motivations, awareness or behaviours (Frau-Meigs, 2024).
4. Malinformation refers to information that is based on real facts but is deliberately manipulated, presented out of context, or shared with the intent to cause harm. Unlike misinformation (false but shared without harmful intent) or disinformation (deliberately false and meant to deceive), malinformation is rooted in truth but is weaponised to damage reputations, incite violence, or undermine trust in institutions. Examples include leaking private information to harm individuals, selectively presenting truthful data to mislead, and using real events to fuel hate speech or conflict (Wardle & Derakhshan, 2017).
5. Fake news is false or misinformation presented as news, including news satire, parody, fabrication, manipulation, advertising, and propaganda (Tandoc Jr et al., 2020).
6. Abusive or threatening speech or writing that expresses prejudice based on ethnicity, religion, sexual orientation, or similar grounds.

The Summit of the Future (UN, 2024) outlines a vision for an open, free, and secure digital future for all, anchored in global cooperation for human development and SDG progress. In the Pact for the Future (2024), its outcome document, UN member states reaffirmed their commitment to integrity in public information to achieve an information ecosystem—particularly online—that is inclusive and safe for all. To achieve this vision, it is crucial to understand the incidence and consequences of information pollution in the Global South, as well as to identify measures to strengthen information integrity, defined as the accuracy, consistency, and reliability of information (UN, 2023). This reflective piece seeks to analyse the drivers and consequences of information pollution in India and to propose strategies to safeguard information integrity.

Methods

This piece employed a mixed-methods approach, conducted in five phases, to understand the factors that contribute to information pollution in India and its consequences and to explore strategies to safeguard information integrity in the Global South.

The first phase included the analysis of two systematic reviews (Blair et al., 2023; Muhammed & Matthew, 2022) and one book (Sinha et al., 2019), providing a comprehensive overview of scholarship on countering information pollution. Blair et al. (2023) synthesise evidence from 176 intervention tests reported in 155 unique studies conducted in both Global North and Global South countries. Muhammed and Mathew (2022) employ a structured approach based on Webster's guidelines to identify relevant literature on the spread of misinformation focused on politics, health and disaster. Two of these three themes—health and politics—emerged frequently in the analysis and the interviews, and so are the main focus of this paper. For historical documentation of the prevalence of misinformation in India, the book India Misinformed (Sinha et al., 2019) was included in the literature review. The author of this book was later interviewed at length.

In the second phase, drawing on the UNDP's strategic guidance conceptual framework (UNDP, 2023), a questionnaire was developed to determine drivers and purveyors of information pollution in India in the following areas: (1) socio-political; (2) media and information; and (3) health communication see Appendix 1).

In the third phase, 50 instances of misinformation collected from WhatsApp, Boom Live, and the Alt News website—well-known social media platforms—

were examined to explore the nature and diversity of misinformation in India. Additionally, 12 monthly and two annual reports from Boom were analysed to gain a deeper understanding of the range of misinformation and the responses of fact-checkers.

In the fourth phase, key informant interviews were conducted with technology experts and people actively involved in combating information pollution through fact-checking and advocacy. Twenty organisations in India were identified, and from 18 responses, six informants were selected for semi-structured interviews based on their expertise, research, and advocacy work (see Appendix 2). Interview transcripts were manually coded for thematic analysis.

Finally, in the fifth phase, analysis was further narrowed to focus on emerging regulatory frameworks in India relating to information on politics and health specifically, and measures adopted by governments to enhance information integrity.

Results

A number of key themes emerge from the findings relating to information pollution and its consequences, specifically: polarisation, health infodemic, technology-mediated tools, human factors and governance. These themes are discussed in greater detail in the following section.

Polarisation

Key informant interviews found political polarisation, which is on the rise across the world (Kubin & Sikorski, 2021), to be one of the main drivers of information pollution in India, propagated by political parties. From around 2014, the ruling Bharatiya Janata Party (BJP) was the first political party in the country to leverage social media (Carney, 2024). Carney (2024) describes in detail how WhatsApp has become a powerful tool for political campaigning in India, with the BJP being a pioneer in leveraging the platform for electioneering. The party established an extensive network of WhatsApp groups, supported by tens of thousands of "IT cell" volunteers responsible for disseminating campaign content (Murgia et al., 2019). This strategy was integral to the BJP's digital outreach, with the party's social media head famously referring to the 2019 general election as a "WhatsApp election," a phrase that gained traction in both national and international media (Perrigo, 2019). Inspired by the BJP's success, other political parties have adopted similar tactics, further entrenching WhatsApp as a key platform for political

communication. The use of WhatsApp for election campaigns has since expanded globally, particularly in developing democracies, enabling political parties to engage with remote voters (Renno, 2019).

Both the government and opposition parties are failing to prevent information pollution. In fact, they are causing what is known as 'demographic anxiety'—a fear of specific demographic groups. This anxiety is mainly caused by misinformation campaigns driven by nationalism.

An expert in politicians' use of social media and misinformation in India explained in an interview that dangerous online speech and propaganda pose three main challenges: sophistication, believability, and virality. He reported that due to high levels of polarisation and distrust of mainstream media, some citizens quickly believe negative information about groups they view as opposing their interests, making believability particularly strong. Additionally, he noted that various groups, including politicians, have established networks that can rapidly spread content, enhancing virality. AI will grow in sophistication, allowing political parties to use data to refine propaganda that can effectively influence voters.

Health infodemic

The World Health Organisation (WHO, 2020) defines the term 'infodemic' as the spread of too much information, including false or misleading information, in digital and physical environments during a disease outbreak. Infodemics cause confusion and risk-taking behaviours that can be harmful. It also leads to mistrust in the health authorities and undermines the public health response. An infodemic can intensify or lengthen outbreaks when people are unsure about what to do to protect their health. It can quickly fill information voids and amplify harmful messages (WHO, 2020). For example, during the COVID-19 pandemic, there was an information void in the state of Karnataka in India relating to the gap between the first and second vaccination doses (Ranjini, 2021). There was also hesitancy toward the second dose among those who had developed fever, swelling, vomiting, nausea, headache and other mild ailments when they took the first dose. There was no mechanism to resolve queries and doubts about the vaccine, side effects, personal medical conditions, or even about the interval between the two doses. While many doctors and health workers recommended eight to ten weeks between doses, as per WHO guidelines, the Government of India facilitated vaccination with a four-week gap. This created confusion among the population. While there was information overload on some issues, there was an information void on certain other issues. This led to the spread of

misinformation. Thus, the main problems in the vaccination drive were related to information and two-way communication between the authorities and the population (Ranjini, 2021).

Technology-mediated tools

GAI can deliberately generate factually incorrect content and articulate it in reasonably good language, mostly English, with a high degree of persuasiveness. It has the potential to create misinformation with ease and spread it at scale. AI is trained on massive volumes of unfiltered or minimally filtered data from the internet, which means there is a potential for biases in these datasets to perpetuate existing systems of segregation, social dominance, and inequality, as noted in an interview by an AI scholar. Experts interviewed for this reflective piece noted that these tools can generate misleading content through voice clones, deepfakes or robocall messages, making it harder to tell if the content originates from a human or a machine. This has led some to declare that GAI is the ultimate disinformation amplifier (DW Akademie, 2024). Popular messaging platforms such as Twitter and WhatsApp have also facilitated the spread of information pollution since these are anonymous, end-to-end encrypted, and, as a result, dangerously unrestrained.

Human factors

The Healthy Indian Project (THIP) was established as a health literacy platform dedicated to empowering Indians with the knowledge and skills needed to make informed healthcare decisions. By providing accessible, culturally sensitive education and resources, THIP bridges the gap between medical information and public understanding. In an interview, the founder of THIP outlined key human factors contributing to information pollution, drawing from extensive hands-on experience and valuable insights from the field (Table 1). This list categorises how and why various individuals or groups generate, amplify, or fall victim to misinformation, disinformation, and other forms of digital content distortion.

Table 1. **Types of drivers of information pollution**

Category	Human drivers of information pollution
1	Those for whom sharing is a way of keeping in touch with their friends and family. They share and spread misinformation without understanding the consequences or without even reading the messages completely.
2	Those sharing information with a genuine intention to help others, however, have low media and health literacy and thus cannot discern what is true or false.
3	People who share information after it has worked for them. These people have good digital literacy and a medium level of health literacy. They were the most problematic group during the COVID-19 pandemic since they shared misinformation stating that some alternative medicine worked for them.
4	Those with certain cultural and medicinal biases confidently share information, saying that they are experts on the topic.
5	Conspiracy theory peddlers, who believe in their stance, support dubious claims and spread misinformation.
6	Those who have a high degree of digital and health literacy who deliberately create and spread misinformation to gain viewership/likes/leads. This could also be big media houses, social media influencers and others who knowingly share misinformation.

Note. Data from expert interview

Governance

India does not currently have specific laws or statutory rules to regulate AI. However, various frameworks, advisories and guidelines have been adopted,[7] and the Indian Government made amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 for an open, trusted and accountable internet. As per this regulation, foreign technology companies such as WhatsApp, Facebook, Twitter, Netflix, Amazon, etc., are required to deploy technology-based regulation measures, have a physical presence in India, maintain appropriate human oversight, and periodically review automated tools.

In a recent development, the Artificial Intelligence (AI) Action Summit, convened in Paris in February 2025, was co-chaired by French President Emmanuel Macron and Indian Prime Minister Narendra Modi. At this summit, approximately 60 countries, including France, China, and India, endorsed a joint declaration titled the "Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet". In his opening statement, Modi said that "Governance is also about ensuring access to all, especially in the

7. For a detailed discussion see Harmon et al. (2024); for comparison with other countries see Baldota (2024).

Global South. It is where the capacities are most lacking—be it computer power, talent, data, or the financial resources," also adding that "we must address concerns related to cyber security, disinformation, and deep fakes." Some of the principles emphasized in the declaration are: guaranteeing equitable access to AI technologies while addressing disparities in digital access; encouraging open, ethical AI systems that are secure, reliable, and accountable[8]; and enhancing global collaboration and establishing comprehensive governance frameworks for AI[9].

Media and information literacy educates the public to discern credible information in the context of evolving generative AI.

In summary, the study reveals that propaganda and political polarisation, driven by the pursuit of political advantage, are among the key catalysts of information pollution in the political sphere. This conclusion is drawn from extensive research and in-depth analysis conducted by the author. Conversely, despite the spread of health misinformation, there is no evidence of malicious intent in health-related information pollution. In addition, information pollution disproportionately affects minorities, migrants, and low-literacy communities, making them more vulnerable to manipulation. Furthermore, tools like Deepfakes amplify the scale and reach of false information. Overall, the significant challenge of AI-generated misinformation remains unregulated, and the study finds that Global South governments and institutions are ill-equipped to safeguard information integrity in the face of rapidly evolving technologies, leading to an urgent need for stronger policies and technological cooperation.

Government responses

During the 2024 parliamentary elections, the Election Commission of India issued guidelines on the responsible and ethical use of social media platforms and avoidance of wrongful use by political parties and their representatives during general and by-elections. It warned political parties, their representatives, and key campaigners against using deepfakes, AI-generated distorted content that spreads fake information, mis- and

8. See website: https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet
9. See website: https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2101947

disinformation, and against distorting facts which lowers the standards of electioneering. Most importantly, it directed that "whenever such deep fake audios/videos are posted, they should be taken down immediately within three hours of being ordered to take down the content" (Election Commission of India, 2024).

Institutional interventions can dramatically improve information integrity. For example, the Government of Karnataka in southern India sponsored a trial intervention called the "Information Disorder Tackling Unit", which was operational for 90 days before the general elections. The authorities reviewed around 64,000 internet articles daily. A total of 84,47,361 posts were scanned during this period, in which 182,450 threats were identified. A total of 537 fact checks were conducted, and 39 were escalated for legal review, of which 18 first information reports were filed relating to disinformation. Politics and elections accounted for 54% of all fact-checking (Joshi, 2024). The government published these misinformation reports online. While this initiative might be driven by political motives from a specific party, it serves as an example of strategies that policymakers and governments can adopt.

The findings reported here point to three main tools necessary to safeguard information integrity in India: strengthening institutions at different levels, technological innovation, and collaborative policy formulation/regulations.

Strategies to safeguard information integrity

The literature review and key informant interviews suggested the following strategies to safeguard information integrity:

- Inoculation (prebunking), involves preparing individuals to recognise misinformation by forewarning them, making them more resistant to it later (Compton et al., 2021).
- Debunking, which corrects misinformation after it spreads, using fact-checking, individual rebuttals, or platform-based algorithmic corrections.
- Credibility labels/tags, which are stamps that provide a quick true/false check on misinformation, are marked by fact-checkers without detailed explanations.
- Contextual labels/provenance cues offer background information to help users understand the origins and context of information but not its truthfulness.
- Media and information literacy, which educates the public to discern credible information in the context of evolving GAI.

Evidence on interventions and recommendations

Information pollution disproportionately affects marginalised, minority, migrant, and low-literacy groups, who are more vulnerable. There is also an urgent need to develop better monitoring capabilities and accountability mechanisms in different areas, including technology companies and platforms, government, and media, among others. Collaboration between different groups and actors is essential to facilitate fact-checking and safeguarding information integrity. Collaboration among various stakeholders plays a crucial role in verifying facts and protecting information integrity.

International cooperation

The research agenda on information integrity is relatively new and has focused mainly on US and European contexts. The reflective piece showed that there are few fact-checking organisations and credible verification bodies. There is a need to promote context-specific research and the mapping of the information ecosystem in the Global South and to develop a Global South-specific network, similar to the International Fact-Checking Network or the Vaccine Safety Net. On the technical front, there is a need for more projects like the Coalition for Content Provenance and Authenticity, which is building a system to provide provenance and history for digital media, providing tools for creators to claim authorship as well as empowering consumers to make informed decisions about what to trust.

Collaboration between different groups and actors is essential to facilitate fact-checking and safeguarding information integrity.

The findings also point to the following recommendations in terms of information governance:

- Convene a high-level group to strengthen the capacity of various institutions in the Global South to counter threats to information integrity. Specifically, it recommends establishing a Global South consortium on information pollution, supporting research and advocacy efforts. Such a consortium would develop guidelines and policies specific to the Global South context to mitigate this problem. Given the fast-paced evolution of GAI, the group should frame (self)

regulations collaboratively with technology companies, platforms, and governments.

- Fund fact-checkers, which are organisations dedicated to debunking misinformation across media platforms and verifying claims from political figures and authorities.
- Enforce platform alterations: compel social media platforms (e.g., Facebook, Twitter, YouTube, WhatsApp) to address issues such as fake news, fake user accounts, originator of messages, and monitor illegal content, specifically to modify interfaces and algorithms to limit misinformation spread. For example, WhatsApp could reduce its message-forwarding limit from over 200 to five recipients to help curb false information.

To conclude, an informed citizenry is key to human progress and democracy. GAI introduces a new layer to the issue of information pollution, intensifying the need for improved monitoring systems and accountability measures. The unregulated and easily accessible nature of GAI allows anyone to generate and spread misinformation at scale. Analysis revealed that propaganda and polarisation motivated by political gain are the main drivers of information pollution. While information pollution in health is often unintentional, without malicious intent, it can have serious consequences and also needs to be addressed.

This reflective piece argues for concrete strategies to enhance information integrity in the Global South, including practical interventions, multi-stakeholder collaboration, oversight frameworks, and accountability standards. It emphasises pre-bunking, promoting media and information literacy, and fact-checking information through government and other entities. Lastly, it proposes creating a Global South consortium to tackle information pollution and support research and advocacy efforts in the Global South.

Limitations of the study

Analysis of hate speech, although important, is beyond the scope of this reflective piece since it requires a different technical and methodological approach. The reflective piece primarily focuses on India, with insights applicable to the broader Global South. This specific geographical focus may limit the generalisability of the findings to other regions within the Global South.

References

Baldota, K. (2024, May 13). India's evolving approach to AI governance [Blog entry]. https://artha.global/blog/indias-evolving-approach-to-ai-governance/

Carney, K. (2024). *The effect of social media on voters: experimental evidence from an Indian election.* Paper presented at Stanford University Department of Economics seminar. https://economics.stanford.edu/events/effect-social-media-voters-experimental-evidence-indian-election

Compton, J., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. *Social and Personality Psychology Compass,* 15(6), e12602. https://doi.org/10.1111/spc3.12602

Election Commission of India. (2024, May 6) Responsible and ethical use of social media platforms and strict avoidance of any wrongful use by political parties and their representatives during the MCC period in General Elections and by-elections-regd. [Government of India communication] https://elections24.eci.gov.in/docs/2eJLyv9x2w.pdf

Élysée. (2025, 11 February). *Statement on inclusive and sustainable artificial intelligence for people and the planet.* Retrieved from https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet

DW Akademie.(2024, March 26). Generative AI is the ultimate disinformation amplifier [Online article] https://akademie.dw.com/en/generative-ai-is-the-ultimate-disinformation-amplifier/a-68593890

First Post. (2024, April 11). Fact-check: The truth behind the 'challenge vote' in 2024 Lok Sabha elections [Online article] https://www.firstpost.com/explainers/fact-check-truth-challenge-vote-2024-lok-sabha-elections-13758473.html

Frau-Meigs, D. (2024). *User empowerment through media and information literacy responses to the evolution of generative artificial intelligence (GAI): Summary.* (UNESCO Policy Brief). https://unesdoc.unesco.org/ark:/48223/pf0000388547

Groh, M., Sankaranarayanan, A., Lippman, A., & Picard, R. (2022). Human detection of political deepfakes across transcripts, audio, and video. *arXiv,* 2202.12883. https://doi.org/10.48550/arXiv.2202.12883

Harmon, S., Wilsmann, M., Joshi, G., Ballesteros, A., & Baitinger, P. (2024). Decoding India's AI governance strategy and its implications for the U.S.-India bilateral relationship. *Indian Public Policy Review,* 5(4), 51-82. https://doi.org/10.55763/ippr.2024.05.04.003

Joshi, B. (2024, July 18). 537 fact-checks, 18 FIRs in a 90-day trial of fake news tacking unit. *Deccan Herald.* https://www.deccanherald.com/india/karnataka/537-fact-checks-18-firs-in- 90-day-trial-of-fake-news-tackling-unit-3109912

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association, 45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Muhammed T., S., & Mathew S. K. (2022). The disaster of misinformation: A review of research in social media. *International Journal of Data Science and Analytics,* 13(4), 271-285.

National Academies. (2020, April 22). Lemon juice does not cure COVID-19 [Online article]. https://www.nationalacademies.org/based-on-science/lemon-juice-does-not-cure-covid-19

Murgia, M., Findlay, S. & Schipani, A. (2019, May 4). India: The WhatsApp Election. *Financial Times.* https://www.ft.com/content/9fe88fba-6c0d-11e9-a9a5-351eeaef6d84

Perrigo, B. (2019. January 25). How volunteers for India's ruling party are using WhatsApp to fuel fake news ahead of elections. *Time Magazine.* https://time.com/5512032/whatsapp-india-election-2019/?xid=time_socialflow_twitter&utm_source=twitter.com&utm_campaign=time&utm_medium=socialflowtw

Press Information Bureau. (2025). *Press Release ID: 2101947.* Retrieved 13 February 2024, from https://pib.gov.in/PressReleaseIframePage.aspx?PRID=2101947

Ranjini, C. R. (2021, February 18). Toolkit to amplify vaccine messaging. *Deccan Herald.* https://www.deccanherald.com/opinion/a-toolkit-to-amplify-vaccine-messaging-952821.html

Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown.

Harvard Kennedy School Misinformation Review 4(5). https://doi.org/10.37016/mr-2020-127

Sinha, P., Shaikh, S., & Sidharth, A. (Eds.). (2019). *India misinformed: The true story.* HarperCollins Publishers India.

Tandoc Jr, E. C., Lim, D., & Ling, R. (2020). Diffusion of disinformation: How social media users respond to fake news and why. *Journalism, 21*(3), 381-398.

United Nations. (2023, June). *Information integrity on digital platforms* (Our Common Agenda Policy Brief 8). https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf

United Nations. (2024). *Pact for the Future. Global Digital Compact and Declaration on Future Generations* (Summit of the Future outcome document) https://www.un.org/sites/un2.un.org/files/sotf-pact_for_the_future_adopted.pdf

UNDP. (2022). *Information integrity: Forging a pathway to truth, resilience and trust.* (UNDP Strategic guidance note). https://www.undp.org/publications/information-integrity-forging-pathway-truth-resilience-and-trust

Wardle, C., & Derakshan, H. (2017). I*nformation disorder: Toward an interdisciplinary framework for research and policy-making.* (Council of Europe report). https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

Wilhelm, E., Ballalai, I., Belanger, M. E., Benjamin, P., Bertrand-Ferrandis, C., Bezbaruah, S., ... & Purnat, T. D. (2023). Measuring the burden of infodemics: Summary of the methods and results of the Fifth WHO Infodemic Management Conference. *JMIR Infodemiology,* 3(1), e44207.

World Economic Forum (2024) *The Global Risks Report 2024,* 19th Edition, Insight Report. https://www.weforum.org/publications/global-risks-report-2024

World Health Organization. (2020). Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. [Online joint statement] https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic

Appendix 1

Interview guide

This research focuses on misinformation and fake news in two important domains: politics and health. Some of the questions that I would like to ask you are as follows:

1. Misinformation: Who is producing it?
2. Who is consuming it?
3. How is it being spread?
4. What is being spread? Different types?
5. What is the overall prevalence? How big is this problem?
6. What is the impact of misinformation and fake news on these two domains - Politics and health?
7. What are the current models for mitigating/fact-checking?
8. What are the strategies to control the spread of misinformation? What can be done to stop it? In India and in the Global South?
9. Any initiatives that you are part of?
10. Do you think international attention and cooperation is required, particularly in the Global South, to address this issue? If so, how?
11. What are your recommendations for the future?

Appendix 2

List of key informants interviewed

1. Amrita Sengupta was interviewed on May 7, 2024.
2. Joyojeet Pal, interviewed on May 8, 2024
3. Padmini Murray was interviewed on May 3, 2024.
4. Pratik Sinha, was interviewed on May 4, 2024.
5. Suditpa Sengupta, was interviewed on May 7, 2024.
6. Tarunima Prabhakar, was interviewed on May 3, 2024.