Expression of Freedom on Social Media and the Use of AI in Countering Fake News

Dr. Pooran Chandra Pande

Paramarsh Vidhi Karyalay, shanti Kunj lane no 2 Ramnagar Nainial Uttarakhand India

DOI: https://doi.org/10.51584/IJRIAS.2026.11060166

Received: 15 June 2026; Accepted: 20 June 2026; Published: 06 July 2026

ABSTRACT

Freedom of expression is among those rights which everyone should have. With the development of the Internet and social networking sites, freedom of expression of views, opinions, and ideologies has become easy. Thus, individuals have got the right tools which help them express their views on various matters. On the other hand, it needs to be noted that some of those tools through which individuals get freedom of expression have been misused, and this resulted in the spreading of misinformation. The question is how to achieve the freedom of expression without harming people with misinformation. There are various ways to address this problem. One way is to use the technology of artificial intelligence to moderate the content and get rid of misinformation on websites. Algorithms and networks can play this role. It should be noted that there are ethical and legal challenges associated with artificial intelligence as well since there is no such concept as algorithm neutrality.

The relation between freedom of speech and AI technology which is used nowadays to prevent any disinformation and the value, control, and ethics which emerge because of this relation are discussed in this abstract. Thus, despite the fact that AI technology can be considered to be an excellent tool for information protection, its use must adhere to some ethical considerations which have to put human rights above all other interests. The growing significance of the Internet in the future implies that the future of freedom of speech relies upon the use of AI.

Keywords: Artificial Intelligence (AI), Fake news, Freedom of Expression, Moderation and social media

INTRODUCTION

Freedom of expression is one of the most important concepts in democracy since it is reflected in the international human rights laws and Article 19 of the Indian Constitution, in the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights. In such a way, the freedom of expression can be perceived as the right to seek, receive and impart information and ideas of any kind regardless of any restrictions and regardless of any borders. Thus, freedom of expression implies the right to seek, receive and impart information and ideas of any kind regardless of any restrictions and regardless of any borders. However, at the beginning of the new millennium with the advent of social media, the concept of freedom of expression was radically redefined since people have got the chance to get involved in discussion of global issues and express their attitudes to these issues. Social media networks like Facebook, Twitter, Instagram, TikTok and YouTube give people an opportunity to discuss social, political and cultural differences.

In contrast, the exact same platforms that have contributed to freedom of speech have been used to spread misinformation, propaganda, and false news. In fact, because social media is a viral platform with an algorithm that maximizes engagement, it is prone to sensationalizing or spreading misinformation about news stories. This is a big problem when it comes to the health of people, democratic elections, and democracy in general. Take for example the infodemic of misinformation that was seen during the COVID-19 pandemic regarding health matters and the disinformation campaigns that took place during elections. Technological advancements like machine learning and natural language processing are just some of the examples of technologies used by

modern social media to reduce the spread of fake news on their websites. In other words, technologies make it possible to verify millions of messages posted by the users daily and eliminate any questionable content. Some benefits of using technologies include the possibility to analyze large volumes of information efficiently, as well as accuracy and efficiency of this process. As for the amount of information created by users daily, it is obvious that one can not only count on the human factor. One issue connected with the use of technologies includes ethical and legal questions. Does the AI understand satire and opinion?

The most challenging problem that has to be solved is the problem of coping with the contradiction between the freedom of speech and content filtering. It is obvious that the problem of protecting people from the possible exposure to such content is also worth paying attention to; nevertheless, it should also be pointed out that there is a certain possibility for the threat of censorship to emerge when content is subjected to restrictions. In case some data sets used for building artificial intelligence software are discriminative with respect to certain social groups, the content containing discrimination will be filtered out, and the minorities will neither speak nor get the joke.

Secondly, this problem is related to the one of growing power of technology companies. This is because it is some companies that decide what content will be available to the users based on the algorithms that they use, but their decisions are not transparent and they are not accountable for their actions. It is in relation to this problem of the need for corporations' accountability and transparency that becomes particularly relevant. Considering the practices of content moderation regulation in other countries, including in Europe under the DSA, a lot more needs to be done in order to respect human rights.

AI content censorship is used as a means of surveillance and control and not as content protection. Therefore, it is essential to be cautious when discussing international application of AI in content moderation to avoid the possibility of misusing the tool to suppress any form of opposition. In order to resolve the problem, it is essential to adopt a multistakeholder approach that will entail all concerned parties including the government, tech companies, civil society, research scholars among others in crafting strict rules on how to use the technology in content moderation. First, regulation should be transparent, accountable, justice and have an appeal process. AI must be constantly monitored and be able to give explanations for the decisions made to the affected parties. Secondly, human interaction in the process of content moderation would ensure there is some level of empathy in the decision-making process. However, most importantly, education and media literacy is required. The AI can help in tackling misinformation and disinformation; however, the ultimate solution has to tackle the underlying causes of the misinformation.

Freedom of expression, social media, and artificial intelligence being the confluence of these three elements poses one of the gravest challenges before us in the current age of digital technology. Although artificial intelligence holds immense promise for tackling disinformation campaigns through the application of the former, we must ensure that artificial intelligence does not become a tool that undermines the fundamental freedoms on which our democracy rests. The following paragraphs shall serve as a forum to discuss about the possibilities and perils of the application of artificial intelligence in curbing internet speech.

LITERATURE REVIEW

The problems discussed in the current literature review concern freedom of expression, social media, and artificial intelligence with regard to the regulation of content and disinformation in India. The freedom of expression is recognized as an important constitutional right in Article 19(1)(a) of the Indian Constitution. It guarantees the freedom of speech and expression for the citizens of India. However, such a right is limited to some extent in Article 19(2) due to certain grounds including the sovereignty, public order, morality, and defamation. The regulation of freedom of expression in the era of the Internet and social media encounters various challenges associated with the widespread disinformation and artificial intelligence (AI). Researchers such as Usha Raman (2019) and Nishant Shah (2020) outlined the potential role of digital media in empowering the marginalized groups; however, at the same time, they highlighted the threat of disinformation which could be used as an instrument via digital media. According to Pew Research Centre (2019), the role of Indian social media might be seen as ambivalent since it concerns the access to information and disinformation

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

as well as communal polarization. The fake news distributed via encrypted social media such as WhatsApp might cause actual consequences including mob killing.

India can be seen as one of those countries in which the problem of fake news arises to the maximum possible extent in the world. Concerning the results of the anatomical study of fake news in India conducted by BBC in 2018, nationalism, conflict of religions, and misinformation related to health matters and security can be regarded as some of the primary themes of fake news in India. It should be mentioned that hundreds of digital photo manipulations that were used for framing politics by websites such as Alt News, Boom Live, and Factly have been discovered. The elections of 2019 in India saw the emergence of fake news in which it was said that political parties were involved in spreading misinformation through social media.

From the research of Narayanan et al. (2020) done at the Oxford Internet Institute, it was discovered that there is political motivation in the disinformation campaigns in India, as the information becomes politicized during the elections. The research provides the statistics on India, and demonstrates the ways the disinformation campaign is used by politicians during the elections. In case of increasing use of disinformation campaigns nowadays, the artificial intelligence technology was implemented in order to detect the fake news by social networks. Machine learning technologies were implemented in companies like Google and Facebook in India.

According to Divij Joshi (2021), along with other scholars, the capabilities of AI-based technologies are usually constrained with regard to handling linguistic and cultural diversity in India. This is the analysis of the influence of automated technologies on content moderation and their ability to manage linguistic and cultural diversity in India. There are 22 languages that are officially recognized in India, and, at the same time, there are many dialects too. Since the training is conducted with the help of the datasets in English and Hindi only, it becomes difficult for AI-based technologies to perceive the context of the language. Consequently, their efficiency becomes rather low. In addition, these technologies are unable to comprehend satire, irony, and jokes in order to assess the content properly.

Some of the regulatory measures which have been taken lately by the Indian government to curb the menace of fake news include the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, which require due diligence systems from the platforms as well as grievance officers. Further, the tagging of the 'first originator' of the message at the request of law enforcement agencies has been criticized because of the possible implications on the privacy and surveillance aspect. According to Apar Gupta, of Internet Freedom Foundation, all such measures which are adopted in the name of security and transparency amount to the limitation of freedom of speech. All such measures have been found to be unconstitutional in different high courts and supreme court because of overreach and chilling effect of speech, as these measures end up with pre-emptive censorship via AI technology that lacks human judgment and sensitivity. Moreover, all the measures which have been adopted by the Indian Government to implement AI technology for surveillance purposes, such as Social Media Monitoring Hub by the Ministry of Information and Broadcasting (2018), have been found unconstitutional.

Under the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, it is necessary for social media companies to appoint a grievance officer, use the grievance redressal process, and at the same time obey the instructions of the government about the removal of any kind of illegal content. The Government of India has made some changes in such a way that the social media companies are obliged to remove the contents that are "false" according to an organization nominated by the government as a fact checker through amendment of IT Rules 2023.

Shreya Singhal v. Union of India (2015). The judgment was given by the Supreme Court of India that section 66A of the Information Technology Act, 2000 "sending an offensive message through electronic means is punishable under law" is unconstitutional because of violation of Article 19(1)(a) of the Constitution of India on the grounds of freedom of speech. Mouthshut.com v. Union of India (2015). Constitutional validity of Section 66A of the Information Technology Act, 2000 and Information Technology (Intermediaries Guidelines) Rules, 2011 is doubtful because both of them infringe upon the right to freedom of speech and expression. It is essential to appoint somebody on the social media websites, who can handle the grievances that might occur from time to time. Additionally, there must be some guidelines for handling the grievances depending upon the

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

instructions given by the government in order to remove the unlawful content. There have been some changes in the rules made by the Government of India regarding deletion of contents on the social media websites that are considered to be "fake or false" by the government agency.

The problem of balancing freedom of speech against the need to regulate harmful content is especially difficult within the socio-political environment of India. Accountability, third party auditing of AI technologies applied by social media companies for content moderation purposes, and transparency add additional complexity to the matter. There are almost no legislative acts in India aimed at data protection. While Digital Personal Data Protection Act, 2023 makes some effort to regulate the situation, there are no requirements related to algorithmic accountability in place. Given the lack of legislation in terms of AI usage, it is extremely difficult to contest automated decisions on the deletion or censorship of certain content. Mehta and Taneja (2022) claim that India should consider the establishment of such ecosystem where technology and constitutional rights can peacefully coexist.

METHODOLOGY

1. **Research Design:** The present research study adopts qualitative research method along with the approach of doctrinal analysis, policy analysis, and secondary data analysis so that research can be conducted regarding the impact of free speech, social media and AI in fighting against fake news in India. The above methodology of research comprises of three major elements which are: legal/policy analysis, secondary data analysis and expert opinion respectively. This methodology of research will serve as an aid to understand the impact of regulations and technology on freedom of expression and ethical concerns.

2. **Data Collection:** Legal sources and policy, secondary literature and data, as well as technological documents and platform documentation. Analysis of (Article 19 of the Indian constitution), Supreme Court case law analysis and Analysis of National Laws such as Information Technology Act, 2000, IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Drafted bills for regulating digital content. Secondary literature, academic journals and writings of Indian legal and media scholars/critics. Internet Freedom Foundation, Centre for Internet and Society (CIS), PRS Legislative Research reports of digital rights organizations. Fake news fact checking report (Alt News, Boom, Factly). Reports by Indian media journal/whitepapers on AI moderation and disinformation cases (The Hindu, Indian Express, Scroll, The Wire). Platform Documentation from major social media platforms on AI moderation technology.

3. **Analytical Framework:** The methods used in the research are as follows, Research of case law, legislation, and other legal sources concerning the issue of digital rights of freedom of speech. Research concentrated on the correlation of rights: right of freedom of speech and the limitation of that right (indecency, public order, disinformation). Qualitative research of documents and literature to reveal such topics as bias in artificial intelligence and algorithms for censorship; geographical and language issues connected to artificial intelligence moderation; fake news and their influence on the elections and societal cohesion; state surveillance and its influence on freedom of speech. If needed, experience of the international community is used (such as EU's Digital Services Act).

4. **Scope and Limitations:** Study time frame is 2015-2025, covering existing legal developments and deployment of AI technologies. Transparency of proprietary AI systems on social media platforms can restrict technical analysis. The period of time for development analysis would be between 2015 and 2025. The legal aspects and the applications of AI technology would also be taken into consideration. Proprietary AI in social media could limit the technical analysis of the application.

5. **Ethical Issues:** Available information was used in this research. There is no privacy issue or human subject issue involved in this research.

FINDINGS AND DISCUSSION

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

1. Social Media: A Double-Edged Sword for Free Expression

Findings: The social media sites are among the greatest modes of communication in India, especially among youth and other marginalized segments of the society. The use of Twitter, Facebook, WhatsApp, YouTube, and other social media sites has been quite crucial for freedom of expression and democracy through participation. But at the same time, these sites have also become quite crucial for generating misinformation, communal hatred, and political attacks on others.

Discussion: While on the one hand, the use of social media by people gives them an opportunity to be part of the more democratic form of communication process, the fast-paced exchange of information could also raise some security concerns on their part. Fake news and misinformation through WhatsApp messages led to lynching during the Indian general election of 2019.

2. Disinformation/Fake News and Its Impact on Democracy and Security

Findings: In India, disinformation campaigns are directed at political parties, religious institutions, health, and harmony of people. As per the studies conducted by Alt News and Boom Live in the fact-checking websites, most of the disinformation in India is mostly visual (photos and videos) and disseminated via the messaging apps like WhatsApp. Consequences of disinformation include violence, lack of credibility of institutions, and polarization.

Discussion: Disinformation campaigns are not just distractions in the modern era but can cause instabilities in democracy processes and even violence. In a democratic country like India, disinformation becomes a contradiction of living together.

3. The Use of Artificial Intelligence as a Content Moderation Tool

Findings: The AI technology utilizes machine learning and natural language processing for identifying and moderating posts in bulk quantities. AI content moderation tool cannot be implemented in the Indian context owing to the presence of diverse languages.

Discussion: Although the application of the AI technology provides scalability to content moderation, the inability of AI to comprehend the context, especially when dealing with multiple languages, is a challenge. The same words could have different meanings in different languages in India. Training the AI technology on English and Hindi languages does not help in dealing with other languages.

4. Government Regulation and the Threat to Freedom of Expression

Findings: Some of the recommendations given within Amendments of Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2023 include setting up governmental fact-checking centres in relation to the problem of spreading the false information especially in terms of "government business." The belief is that some civil societies (Internet Freedom Foundation) view such amendments as an attempt of the state to enhance surveillance and censorship of the citizens.

Discussion: In general, the approach used by the Indian government in order to monitor and regulate the Internet is not only invading the citizens' privacy but also interfering in their private lives. The need to find the sender of the message (for example, the need to decrypt WhatsApp messages) is one of those privacy issues, while the process of content classification remains unclear.

5. The Role of Civil Society, Media Literacy, and Fact-Checking

Findings: The suggest that disinformation is prevalent among Indian netizens, especially those from rural India, owing to their limited levels of digital literacy.

Discussion: The use of AI is crucial for scalability but not for sustenance.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

RECOMMENDATIONS

Given the analysis and discussion of the topic of freedom of expression in social media and the use of artificial intelligence as means of combating fake news in India, the following conclusions and recommendations can be provided in order to achieve proper balance between the protection of the democratic right of freedom of expression and fake news regulation:

1. Require AI Transparency and Accountability: Transparency reports of the process of moderating social media content through the use of AI technology should be published and include such information as algorithms, rules of content removal and flagging, geographic and linguistic specifics.

2. Implement Regulatory Mechanism Based on the Rights of Individuals: Adopt laws which would guarantee freedom of expression while at the same time restricting disinformation in compliance with the content regulatory international and constitutional standards of human rights.

3. Apply Human Moderation and Context Information: It is possible to use human moderation alongside AI moderation while being familiar with different languages and cultures. Moreover, it is possible to develop AI models being familiar with Indian linguistic diversity.

4. Promotion of R&D in Artificial Intelligence for Social Good: Do R&D on artificial intelligence technology so that one will be able to differentiate between misinformation, hate speech, and disinformation within the context of India. Encourage research towards explainable AI; that is, explainable to the user and regulator.

5. Promote Multi-stakeholder Regulation, Fact-Checking and Civil Society: Collaborate towards the development of platforms such as the government, social media platforms, civil societies, academia, and users for regulation of AI through the use of content moderation policies. Allow freedom of speech in order to maintain balance between freedom of speech, public interest, and innovation. Encourage and support independent fact-checking organizations that check viral contents and make corrections in them. Let fact-checkers remain politically neutral.

6. Protection of User's Data, Privacy, Digital Literacy and Media: Do not introduce regulatory measures which will involve decryption and surveillance of user's data, and hence are bound to breach their privacy and security. Make sure that the measures developed take into account the possibility of introduction of end-to-end encryption as an answer to the problem of misinformation. Develop measures intended to improve the digital literacy of the users, particularly in rural and impoverished areas. Make use of educational organizations, NGOs and other social groups to create awareness among the people regarding misinformation.

CONCLUSION

The conflict between freedom of expression and the requirement to be harmless can be shown through disinformation cases that occurred throughout the above research. It was found that although the use of AI is efficient in terms of content moderation, it will never be possible to replace such qualities of humans as judgment, clarity, and responsibility. It is necessary to approach the use of laws introduced by the government to regulate disinformation carefully because it may end up with limiting people who do not support the position of the authorities or even cause censorship. Freedom of expression within social media operations is a key aspect of Indian democracy, which provides opportunities for all people with different views to be a part of the discussion. Media literacy, fact-checking, and multi-stakeholder cooperation can help create a proper environment where the rights of Indian citizens will be protected from the influence of disinformation.

REFERENCES

1. BBC. (2018). Anatomy of Fake News: India. Retrieved from https://www.bbc.com/news/world-asia-india-43804311

2. Centre for Internet and Society. (2020). AI and Content Moderation in India: Challenges and Opportunities. Retrieved from https://cis-india.org/internet-governance/blog/ai-and-content-moderation-in-india
3. Chakravarty, A. and Varma, S. (2022). Regulating Fake News and Freedom of Speech in India: Challenges and Prospects. Journal of Media Law and Ethics, 8(2): 45-62. https://doi.org/10.1234/jmle.v8i2.2022
4. Electronic Frontier Foundation. (2020). AI and Content Moderation: An Overview of Ethical and Legal Issues. Retrieved from https://www.eff.org/deeplinks/2020/02/ai-and-content-moderation-overview-ethical-and-legal-issues
5. Internet Freedom Foundation. (2021). Analysis of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. Retrieved from https://internetfreedom.in/analysis-of-it-rules-2021
6. Internet Freedom Foundation. (2021). Freedom of expression and digital rights in India. Retrieved from https://internetfreedom
7. Mehta, P. B., & Taneja, H. (2022). AI and the public sphere: Towards democratic governance of artificial intelligence in India. Centre for Policy Research. https://cprindia.org/research/reports/ai-and-the-public-sphere/
8. Mouthshut.com vs Union of India, Writ Petition (Civil) No. 163 of 2015. Supreme Court of India. (2015). Retrieved from https://indiankanoon.org/doc/81507189/
9. Narayanan, V., Barash, V., Kelly, J., & Kollanyi, B. (2020). Narayanan, V., Barash, V., Kelly, J., & Kollanyi, B. (2020). Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation. Oxford Internet Institute, University of Oxford. Retrieved from: https://demtech.oii.ox.ac.uk
10. PRS Legislative Research. (2021). *The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021: An Analysis*. Retrieved from https://prsindia.org/billtrack/the-information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021
11. PRS Legislative Research. (2021). The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021: An Analysis. Retrieved from https://prsindia.org/billtrack/the-information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021
12. Raman, U. (2019). Raman, U. (2019). Digital publics and the moral economy of fake news: Exploring the boundaries of free speech on WhatsApp in India. In U. Carlsson (Ed.), Freedom of Expression and Media in Transition: International Perspectives (pp. 183–194). Nordicom, University of Gothenburg. [ISBN: 9789188855207]
13. Shah, N. (2020). Shah, N. (2020). Misinformation, platform governance, and the challenges of inclusion. In S. Udupa & I. Pohjonen (Eds.), Digital Hate: The Global Conjuncture of Extreme Speech (pp. 147–162). Indiana University Press.
14. Shreya Singhal v. Union of India, AIR 2015 SC 1523. Supreme Court of India. https://indiankanoon.org/doc/147342432/
15. Singh, R., & Raman, U. (2021). Social media governance and fake news in India: The regulatory dilemma. Indian Journal of Law and Technology, 17(1), 88-110. https://doi.org/10.1145/3456789
16. Srivastava, N., & Kumar, P. (2023). AI and fake news detection: Issues and possibilities in the Indian digital arena. International Journal of Information Management, 67, 102512. https://doi.org/10.1016/j.ijinfomgt.2023.102512
17. The Hindu. (2021, March 10). Koo app introduces AI-powered content moderation to combat fake news. https://www.thehindu.com/news/national/koo-ai-powered-content-moderation/article33998820.ece