

THE PHILOSOPHICAL AND EPISTEMOLOGICAL COLLAPSE OF THE MARKETPLACE OF IDEAS

*By Ananya Vinayak Kini & Likitha Venkatesh
Christ (Deemed to be University), Bengaluru*

The foundational concept of modern free-speech theory, the marketplace of ideas, was developed by Justice Oliver Wendell Holmes Jr. in his dissenting opinion in *Abrams v. United States* (1919).¹ The judge argued that “the best test of truth is the power to get itself accepted in the competition of the market,” and the “greatest good will come from the most liberty.” The marketplace of ideas and classical liberal thought that formed around it, notably Milton’s *Areopagitica*² and Mill’s *On Liberty*,³ presume the existence of an impartial judge who considers the strengths and weaknesses of each side of the argument. Truth and proper moral conduct inevitably triumph over lies and vice.

However, as we have moved our activities into the digital space, the postulates of classical liberalism have ceased to hold. Not due to direct censorship by the government or increased self-regulation by citizens, but because the digital space allows interference in the very mechanisms of information exchange and public discourse. Namely, the introduction of a for-profit intermediary that controls the flow of information between interlocutors.⁴ The principles of the classical marketplace of ideas no longer hold due to the interference of private and commercial interests: ideas no longer compete for mass audiences; rather, audiences shop for ideas.

There are three ways in which the introduction of a digital broker undermines the sanctity of the traditional marketplace of ideas. Firstly, classical liberal thought assumes the presence of an impartial judge capable of weighing the pros and cons of each side of the argument. It is implied in the *Areopagitica* and *On Liberty* that truth and goodness will always prevail over falsity and evil. That may not be the case when the ideas compete in a digital space. It is reasonable to assume that the algorithms of search engines and social media platforms

¹*Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

²JOHN MILTON, *Areopagitica* (1644).

³JOHN STUART MILL, *On Liberty* (1859).

⁴SHOSHANA ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* 189–95 (2019).

prioritise engagement and circulation.⁵ The natural advantage of easily graspable lies and conspiracy theories over truth and nuanced information will result in a higher prevalence of these ideas. This state of affairs contradicts the principles of a free and fair digital marketplace of ideas.

Secondly, the classical liberal notion of a speech implies the utterance of an idea, not its distribution. Thus, the author of the utterance has little control over how their idea will be received and disseminated after release. By contrast, proponents of the new orthodoxy believe that the distribution chain has become so important that it determines the value and relevance of ideas. When a large social media company uses algorithms to select which opinions to distribute to the masses, it engages in favouritism toward some ideas and opinions.⁶

Thirdly, the information asymmetry between the speakers and the algorithm broker makes the digital marketplace of ideas neither free nor fair.⁷ Classical liberal thinkers would argue that if the principles of free markets applied to the exchange of ideas, then the currency of choice would be information. Truth and nuance would have a natural advantage over lies and opinions. The algorithm broker, however, operates under undisclosed rules and uses information that speakers do not have access to to decide what content to distribute and amplify.⁸

In India, despite the tendency of American jurisprudence to encourage such developments, things work differently. The Indian Constitution recognises the right to freedom of speech and expression under Article 19(1)(a),⁹ but also states that Parliament may impose “reasonable restrictions” on it for the sake of “security of the State”, “friendly relations with foreign States”, “public order”, etc. under Article 19(2).¹⁰ Moreover, India’s Supreme Court delivered a landmark decision in the *Shreya Singhal v. Union of India* case in

⁵ William J. Brady et al., *Emotion Shapes the Diffusion of Moralized Content in Social Networks*, 114 PROC. NAT’L ACAD. SCI. 7313, 7313–14 (2017).

⁶ TARLETON GILLESPIE, *Custodians of the Internet: Platforms, Content Moderation, and the Cultural Politics of Online Platforms* 45–52 (2018).

⁷ George A. Akerlof, *The Market for “Lemons”: Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488 (1970).

⁸ FRANK PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and Information* 45–50 (2015).

⁹ INDIA CONST. art. 19, cl. 1(a).

¹⁰ INDIA CONST. art. 19, cl. 2.

2015,¹¹ recognising the Internet's potential to enable everyone to share their views, but also stating that speech on the Internet is not exempt from the restrictions mentioned above.

Indian jurisprudence on free speech and social media bears striking similarities to American jurisprudence. As the platforms have evolved from a simple reverse-chronological feed to a complex neural network, the nature of social media has shifted from speech to distribution, and, with it, the legal status of these platforms regarding freedom of speech has changed. It is reasonable to state that while an individual possesses a fundamental right to publish an opinion of any kind under Article 19(1)(a), they do not have the right to demand a social media platform to amplify their voice. If companies abuse their content-moderation power by using the predictive nature of their algorithmic feeds to favour certain opinions, viewpoints, and ideas, they engage in anti-competitive practices and monopolistic suppression of speech. The laws currently in place, notably Section 79 of the Information Technology Act, 2000,¹² which prohibit platforms from being held liable for content distributed by their users, are problematic.

To retrace the fall of the original, glorious promise of the digital space and trace the manner in which it violated the principles of the free and fair exchange of ideas, we need to turn from jurisprudence to economics. We will examine the technical side of a social media platform and the nature of algorithmic amplification.

THE TECHNICAL ARCHITECTURE OF ATTENTION: NEURAL OPTIMIZATION, MORAL CONTAGION, AND VERNACULAR HARMS

To understand why the marketplace of ideas fails online, we have to understand the automated mathematical intermediaries between content creators and the reading public. The early social network was a reverse-chronological bulletin board,¹³ where posts appeared in the order in which they were created. The switch to predictive feeds involved deep neural collaborative filtering networks,¹⁴ fine-tuned to minimise mathematical loss functions focused exclusively on user retention.¹⁵

¹¹ *Shreya Singhal v. Union of India*, (2015) 5 SCC 1, 12–18 (India).

¹² INFORMATION TECHNOLOGY ACT, 2000, § 79, No. 21, Acts of Parliament, 2000 (India).

¹³ Gillespie, *supra* note 6.

¹⁴ Xiangnan He et al., *Neural Collaborative Filtering*, in PROCEEDINGS OF THE 26TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB 173 (2017).

¹⁵ Paul Covington et al., *Deep Neural Networks for YouTube Recommendations*, in PROCEEDINGS OF THE 10TH ACM CONFERENCE ON RECOMMENDER SYSTEMS 191, 191–94 (2016).

A standard recommendation algorithm operates on user feeds using a binary cross-entropy loss function. This mathematical operation quantifies the statistical distance between observed user behaviour and the predicted probability distribution. The algorithm then performs gradient descent on the weights to find the function that minimises the loss. In essence, it updates itself based on information about what keeps users on the platform (watch time, click-through, re-shares, comments). It does not care about truth, democracy, or logical consistency – only about minimising drop-off.

This raises the question of the algorithm's compatibility with natural human behaviour. Human psychology is highly susceptible to emotional arousal, and a significant proportion of that arousal stems from tribalistic moral judgments. In the experiment published in the Proceedings of the National Academy of Sciences by William J. Brady et al., inserting moral-emotional language into political messages increased the rate of circulation within a network by approximately 20% per word.¹⁶ The recommendation algorithm notices this increase in user excitement. It interprets it as an indicator of quality. The system begins to recommend polarising content more frequently, as it improves user retention. It is essential to note that the system itself has no political leanings. It only reacts to observed behaviour, creating a self-reinforcing feedback loop: provocative or polarising posts prompt users to react rapidly, and the algorithm interprets this flood of activity as high value, triggering automated amplification to thousands or millions of additional feeds.¹⁷

The documents leaked by Facebook whistleblower Frances Haugen, known as the Facebook Papers,¹⁸ indicate that engagement-based ranking could create incentives to increase the visibility of divisive, sensational, or otherwise controversial content because such material generates stronger user interactions. In several key entries, the company's internal documentation states that the anger emoji reaction was counted as five times as important as a standard reaction, as in the recommendation algorithm. It confirms that the company actively encouraged political parties to post "offensive content" to retain audiences. His

¹⁶ William J. Brady et al., *Emotion Shapes the Diffusion of Moralized Content in Social Networks*, 114 PROC. NAT'L ACAD. SCI. 7313 (2017).

¹⁷ Zeynep Tufekci, YouTube, the Great Radicalizer, N.Y. TIMES (Mar. 10, 2018), <https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html>

¹⁸ See Frances Haugen, Internal Facebook Documents (The Facebook Papers): *Facebook's News Feed and Democracy* (2021) <https://www.wired.com/story/facebook-papers-internal-documents/> .

paper, “Facebook’s News Feed and Democracy,”¹⁹ even directly addresses the use of recommendation algorithms to promote divisive content in the electoral sphere: “India experiments show that people in India are exposed to posts containing controversial and unproven claims about various topics (mostly on Facebook), and it was observed that after signing up, test users were shown similar posts that continued spreading within a day.” This means that Facebook recommendation algorithms were promoting divisive content, fake news, conspiracy theories, and xenophobic discourse to improve user retention.

One of the key manifestations of the abovementioned algorithmic “wickedness” is the proliferation of Synthetically Generated Information and AI deepfakes.²⁰ In the closing months of the electoral cycle in India, political parties and independent agents began using AI-generated voices, deepfake video imagery, and audiovisual synthesis of the speech of an assassinated political leader. Since recommendation algorithms privilege high engagement at the launch, the initial burst propels the fake news article or video to the top of the feed. The higher the engagement, the more users see it. The higher they are, the more engagement they get. Synthetic information goes viral within minutes. By the time the Election Commission of India,²¹ the Fact Check Units of mainstream media outlets, or the country’s cybersecurity agencies, react, the article has already had ten million impressions: enough to sway several hundred thousand voters with a carefully placed message. For example, the Indian authorities can block misinformation under Section 69A of the Information Technology Act,²² which permits the government to take down any material that could threaten national security. The problem here is that the damage is already done: a few hours or days after the launch of a fake news story or a deepfake video, the state’s reaction is too slow.

In an Indian context, the abovementioned technical environment creates a vernacular void. India is a multilingual country, with over 800 million internet users who speak twenty-two officially recognised languages and hundreds of unofficial languages and dialects.²³ The

¹⁹*Id.*

²⁰INFORMATION TECHNOLOGY (INTERMEDIARY GUIDELINES AND DIGITAL MEDIA ETHICS CODE) 2026, Gazette of India, pt. II sec. 3(i), G.S.R. 120(E) (Feb. 10, 2026).

²¹ Election Comm’n of India, *Advisory on the Use of Artificial Intelligence and Deepfakes in Election Campaigning*, Doc. No. 437/6/INST/ECI/FUN/MCC/2024 (May 6, 2024), Election Comm’n of India, *Advisory on the Use of Artificial Intelligence and Deepfakes in Election Campaigning*, Doc. No. 437/6/INST/ECI/FUN/MCC/2024 (May 6, 2024), <https://elections24.eci.gov.in/docs/2eJLyv9x2w.pdf>.

²²INFORMATION TECHNOLOGY ACT, 2000, § 69A, No. 21, Acts of Parliament, 2000 (India).

²³ Kiran Garimella et al., *Political Discourse on WhatsApp in India: Algorithmic Amplification and Vernacular Misinformation*, INPROCEEDINGS OF THE 2021 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS 1, 4–8 (2021).

safety classifiers for Natural Language Processing technologies and systems are trained only on English data, resulting in structural disparities for low-resource Indic languages.²⁴ This leaves automated moderation systems blind to the peculiarities of Indic regional scripts, colloquialisms, and code-switching habits.

In essence, the recommendation algorithm operates on a language-agnostic basis, using broad-strokes measures of popular content (in this case, weighted towards articles and opinions on politics and current events) and pushing similar items to users to satisfy the platform's retention requirements. Meanwhile, the systems used by Facebook and other platforms for safety and moderation purposes operate on a language-specific basis, making them much less accurate in determining the context of regional content. One thing to note is that the recommendation system favours high-engagement content, and engagement is typically highest with incendiary statements and divisive rumours. This means that an algorithm can easily radicalise an unprotected vernacular regional feed.

As a result, vernacular feeds often contain highly volatile and ethically suspect content. The problem manifests particularly strongly in north Indian regional languages, which have a significant number of users. Indian law does not permit the use of Section 69A to block or take down digital content, except when it is deemed a threat to national security, the sovereignty and integrity of India, the defence of India, friendly relations with foreign States, public order, or for preventing the incitement of cognizable offences related to these grounds. As of now, existing legal and regulatory mechanisms may struggle to respond to the speed, scale, linguistic diversity, and technical characteristics of algorithmically amplified synthetic content. In the case of *Tehseen S. Poonawalla v. Union of India*,²⁵ the Supreme Court of India took note of the phenomena of mob lynchings and mob violence fueled by unverified rumours and digitally disseminated information and took measures to penalise and prevent them. It is an extreme manifestation of the abovementioned algorithmic volatility that affects India's constitutional order and the rule of law.

Having established why platform algorithms are programmed to fuel the worst aspects of human nature, it is necessary to turn to the discussion of economic incentives that create said programming.

²⁴ Pratik Joshi et al., *The State and Fate of Linguistic Diversity and Inclusion in the NLP World*, in PROCEEDINGS OF THE 58TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS 6282 (2020).

²⁵ *Tehseen S. Poonawalla v. Union of India*, (2018) 9 S.C.C. 501.

THE POLITICAL ECONOMY OF ALGORITHMIC GATEKEEPING: SURVEILLANCE CAPITALISM, MICROECONOMIC EXTERNALITIES, AND PUBLIC SPHERE EROSION

Evaluating the breakdown of the digital marketplace requires a combination of constitutional law and law-and-economics analysis. At present, digital platforms operate as two-sided marketplaces.²⁶ for platform services within the framework of Zuboff's surveillance capitalism.²⁷ Users gain free access to networking services in exchange for the collection of behavioural data from them. In turn, the platform sells behavioural data to advertisers by auctioning targeted advertisement slots to commercial and political advertisers.

Such a digital business model is profoundly different from that of the traditional print media. Unlike traditional media, Section 79 of the Information Technology Act provides that the platform publisher shall be granted safe harbour protection against legal liability. Whereas human editors decide what to publish in traditional print media, digital platforms use loss function algorithms to automate curation. Traditional print media are distributed on a one-to-many basis and are standardised for all the readers. Its revenues are derived from subscriptions or advertisements and are purchased by individual consumers. Finally, traditional print media is responsible for ensuring that published content does not violate any law; i.e., publishers bear the legal liabilities of their content, whereas on digital platforms, the algorithm sells an individual's eyeballs to advertisers for a price. In the platform economy, the consumer is merely a means to an end, and the advertiser is the real customer. Since consumer attention is a limited resource, it is in the platform's interest to manipulate its recommendation algorithms to encourage emotional engagement and maximise advertising revenue.

Your One Stop Legal Destination

From a law-and-economics perspective, the economic rationale for algorithmic amplification creates negative externalities for society. A negative externality is a phenomenon in which a third party suffers the negative effects of a private transaction.²⁸ In the case of a social media company promoting incendiary content, the short-term private gain from such a transaction is increased advertising revenue. However, the social costs are borne by an unspecified third party in the form of societal divisions, communal disharmony, religious polarisation, institutional breakdown, and public order issues under Article 19(2). As such, the platforms

²⁶ Jean-Charles Rochet & Jean Tirole, *Two-Sided Markets: A Progress Report*, 35 RAND J. ECON. 645 (2006).

²⁷ ZUBOFF, *SUPRA* NOTE 4.

²⁸ R.H. Coase, *The Problem of Social Cost*, 3 J. L. & ECON. 1, 1–5 (1960).

do not factor in the social costs of amplifying such content; therefore, it is not economically rational for them to refrain from amplifying it.

The imperfection in the market for digital platforms is remedied through ordinary competition law, which is unavailable due to two key market failures- information asymmetry and network effects.

Significant information asymmetry exists within digital marketplaces.²⁹ The platforms' ranking and weighting methodologies, as well as other underlying recommendation algorithms, are trade secrets not available to either consumers or regulators. As such, the consumer cannot determine whether an algorithm amplifies or suppresses any particular content. Similarly, a change to the recommendation algorithm would affect content visibility without the consumer being aware of why the content was suppressed.

The network effects of the digital platforms present another barrier to competition.³⁰ Digital platforms possess both direct and indirect network effects, where the value of the platform increases with the number of its users. The high switching costs for consumers further reinforce such network effects. As such, consumers are locked into a digital platform that suppresses content that contributes to public discourse, favouring recommendation algorithms that amplify divisive content.

The failure of the digital platform economy to foster competition has adverse repercussions for the democratic public sphere. Habermas's concept of a public sphere is a social arrangement in which private individuals form a public body to deliberate and reach consensus on matters of public interest through rational-critical communication.³¹ However, the personalised content curation of digital platforms does not allow the formation of a public sphere but replaces it with information cocoons, echo chambers, and epistemic bubbles.³² Information cocoons are individualised news feeds in which one's views are only reinforced, without encountering contradictory opinions. Digital platforms create echo chambers by training users to distrust external information, so all information within the echo chamber is

²⁹ George A. Akerlof, *The Market for "Lemons": Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488, 489–90 (1970).

³⁰ Michael L. Katz & Carl Shapiro, *Network Externalities, Competition, and Compatibility*, 75 AM. ECON. REV. 424, 424–28 (1985).

³¹ JURGEN HABERMAS, *The Structural Transformation of the Public Sphere and Deliberative Politics* (Ciaran Cronin trans., Polity 2023).

³² C. Thi Nguyen, *Echo Chambers and Epistemic Bubbles*, 17 EPISTEME 141, 142–45 (2020).

viewed sceptically.³³ Notwithstanding the right to freedom of speech and expression under Article 19(1)(a), the platforms engage in the practice of shadow banning- a covert method of deplatforming, suppressing, or altogether denying visibility to the online content of a user based on predetermined rules without disclosing such rules to the user.³⁴ Thus, shadow banning operates as a form of state censorship, in which the state is not held accountable for restricting citizens' right to freedom of speech and expression.

Having evaluated the economic rationale for the failure of digital platforms to maximise competition, thereby restricting the democratic public sphere, the paper shall analyse how Indian laws address this issue.

ALGORITHMIC GATEKEEPING, MARKET FAILURE AND THE DEMOCRATIC PUBLIC SPHERE

Evaluating the breakdown of the classical marketplace of ideas requires combining constitutional principles with insights from law and economics.³⁵ Modern digital platforms operate on two-sided platform economics, monetising human attention through targeted advertising in a system often described as surveillance capitalism.³⁶ In this framework, ordinary users are not the paying customers; rather, they serve as sources of behavioural data, while commercial advertisers act as the true paying clientele.³⁷ A platform's primary goal is to maximise advertising revenue by driving user retention and keeping people logged in for as long as possible.³⁸ Because human attention is a limited resource, recommendation engines intentionally favour high-arousal, divisive content, which reliably generates higher engagement, expands advertising inventory, and feeds continuous data collection.³⁹

Through the lens of microeconomic theory, this algorithmic amplification creates significant negative externalities, which occur when a private business transaction imposes

³³CASS R. SUNSTEIN, *Republic: Divided Democracy in the Age of Social Media* (2017).

³⁴ Kokil Jaidka, Subhayan Mukerjee & Yphtach Lelkes, *Silenced on Social Media: The Gatekeeping Functions of Shadowbans in the American Twittersverse*, 73 J. COMM. 163, 163–64 (2023).

³⁵ Tim Wu, *Blind Spot: The Attention Economy and the Law*, 82 ANTITRUST L. J. 771, 773–78 (2019).

³⁶ SHOSHANA ZUBOFF, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* 189–95 (2019).

³⁷ Jean-Charles Rochet & Jean Tirole, *Platform Competition in Two-Sided Markets*, 1 J. EUR. ECON. ASS'N 990, 992–95 (2003).

³⁸ DAVID S. EVANS & RICHARD SCHMALENESEE, *Matchmakers: The New Economics of Multisided Platforms* 35–42 (2016).

³⁹ JEAN TIROLE, *Economics for the Common Good* 378–85 (Steven Rendall trans., Princeton Univ. Press 2017).

uncompensated costs on third parties.⁴⁰ When platforms amplify outrage to boost ad revenue, and users consume this captivating content, the broader public suffers the consequences, including heightened communal tension, institutional distrust, and threats to public order under Article 19(2) of the Indian Constitution.⁴¹ This dynamic creates a dangerous platform externality loop: platforms and users optimise for engagement to maximise revenue, thereby imposing unpriced social costs, such as polarisation, on the general public.⁴²

This market cannot self-correct due to two structural failures. The first is information asymmetry, where the internal workings of recommendation algorithms remain completely opaque black boxes. Neither users nor government regulators can inspect how content is ranked, suppressed, or boosted, which prevents informed decision-making and destroys market transparency.⁴³ The second failure involves network effects and switching costs, in which a platform becomes exponentially more valuable as more people join it. This dynamic creates strong user lock-in and high barriers to entry for competitors, preventing users from switching to healthier, higher-quality platforms.⁴⁴ Consequently, the traditional assumption that the marketplace of ideas will naturally self-correct fails, thereby establishing a clear law-and-economics justification for constitutional and statutory regulation.

These market failures are not merely economic concerns; they have direct consequences for the functioning of the democratic public sphere. Jürgen Habermas's concept of the public sphere is built on the idea of citizens encountering a multitude of viewpoints and deliberating on various matters of shared concern.⁴⁵ However, personalised feeds hamper this space by exposing individuals to different sets of information based on predicted engagement, interests and behaviour. This is a concern that requires considerable attention as it leads to the creation of 'information cocoons' wherein individuals are presented with a limited number of previously encountered perspectives that do not challenge their existing beliefs.⁴⁶ This problem of homogenization is intensified when algorithms reward content that evokes

⁴⁰ Coase, *supra* note 28 at 1–5

⁴¹ *Kaushal Kishor v. State of Uttar Pradesh*, (2023) 4 SCC 400, 422–26 (India).

⁴² Tirole, *supra* note 30, at 382–84.

⁴³ *X Corp. v. Union of India*, 2023 SCC OnLine Kar 32, ¶¶ 45–50.

⁴⁴ Michael L. Katz & Carl Shapiro, *Network Externalities, Competition, and Compatibility*, 75 AM. ECON. REV. 424, 424–28 (1985).

⁴⁵ JURGEN HABERMAS, *The Structural Transformation of the Public Sphere and Deliberative Politics* (Ciaran Cronin trans., Polity 2023).

⁴⁶ Cass R. Sunstein, *Republic.com 2.0* 9–10 (Princeton Univ. Press 2007).

powerful reactions, posts that generate fear, outrage and division.⁴⁷ This results in fringe ideas reaching larger audiences, not because of support gained from independent research and reasoned debate, but through algorithmic manipulation.⁴⁸ This challenges the marketplace of ideas, where competing views are expected to be freely challenged.⁴⁹

Algorithms further silence dissent through ‘shadow banning’, unlike traditional censorship, where the posts are removed or banned, ‘shadow banning’ restricts the ability of individuals to view the content without openly removing it.⁵⁰ This creates a stark difference between being allowed to speak and being heard. When select ideas are constantly promoted while others are suppressed, the marketplace of ideas is no longer a level playing field. The issue is not simply that certain opinions are dismissed, but rather that some citizens are unaware of their existence. This limits exposure to disagreement, reducing opportunities for dissent, and repeatedly reinforcing previously held beliefs. Algorithms begin to influence what people see and think, and, consequently, which opinions gain prominence in the public sphere.⁵¹ This fragmentation, therefore, raises concerns not only about the functioning of the democratic public sphere but also about the constitutional framework that governs speech and its regulation. This implication of algorithmic gatekeeping must therefore be assessed within the Indian legal framework.

The “concentric circles” approach in *Dr Ram Manohar Lohia v. State of Bihar*,⁵² *Madhu Limaye v. Sub-Divisional Magistrate*,⁵³ and *Ramji Lal Modi v. State of U.P.*⁵⁴ establishes a sufficient legal nexus between speech and the grounds of restriction that can be imposed under Article 19(2). *Kaushal Kishor v. State of Uttar Pradesh* further clarifies the exhaustive nature of Article 19(2) while highlighting the State’s role as a protective mechanism against harm caused by private actors.⁵⁵

Your One Stop Legal Destination

⁴⁷Mohammed Mosleh et al., *Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media*, 122 PROC. NAT’L. ACAD. SCI. e2412846122 (2025).

⁴⁸Ferenc Huszar et al., *Algorithmic Amplification of Politics on Twitter*, 119 PROC. NAT’L. ACAD. SCI. e2025334119 (2022).

⁴⁹John Stuart Mill, *On Liberty* 21–22 (Elizabeth Rapaport ed., Hackett Publ’g Co. 1978) (1859).

⁵⁰Kokil Jaidka, Subhayan Mukerjee & Yphtach Lelkes, *Silenced on Social Media: The Gatekeeping Functions of Shadowbans in the American Twittersphere*, 73 J. COMM’N 163, 163–78 (2023).

⁵¹Efrat Nechushtai & Seth C. Lewis, *What Kind of News Gatekeepers Do We Want Machines to Be? Filter Bubbles, Fragmentation, and the Normative Dimensions of Algorithmic Recommendations*, 90 COMPUTERS HUM. BEHAV. 298, 298–307 (2019).

⁵²*Dr. Ram Manohar Lohia v. State of Bihar*, (1966) 1 S.C.R. 709, 746-47

⁵³*Madhu Limaye v. Sub-Divisional Magistrate*, (1971) 2 S.C.C. 293, 298-99

⁵⁴*Ramji Lal Modi v. State of U.P.*, A.I.R. 1957 S.C. 620, 625-26.

⁵⁵*Kaushal Kishor v. State of Uttar Pradesh*, (2023) 4 S.C.C. 1, 20-21, 43-44

Any regulation imposed must also satisfy the proportionality requirement. A four-stage test was established in *K.S. Puttaswamy v. Union of India*.⁵⁶ and reiterated in the case of *Anuradha Bhasin v. Union of India*,⁵⁷ it requires a legitimate objective, rational connection, necessity and balancing. Further, Section 79 of the IT Act distinguishes between passive intermediaries and platforms that engage in active participation, raising questions about algorithmic curation and safe harbour.⁵⁸

Finally, in the case of *X Corp. v. Union of India*⁵⁹ illustrates disputes over executive content-blocking mechanisms, including the Sahyog Portal, which raise questions about transparency and procedural safeguards in the removal of online speech. *Kunal Kamra v. Union of India* demonstrates judicial scrutiny of the 2023 Fact-check unit amendments to the IT rules. The Bombay High Court held that the amendment to Rule 3(1)(b)(v) under Articles 14, 19(1)(a) and 19(1)(g), finding the terms “fake” and “false” vague and overbroad, holding that these provisions failed the test of proportionality. The Court also recognised the potential chilling effect on intermediaries, whose safe-harbour protection could be affected by non-compliance. Together, these cases highlight the need to regulate digital platforms while preserving essential elements such as transparency, procedural safeguards, and the constitutional protection of online speech.

THE LEGAL REGULATION OF ALGORITHMIC GATEKEEPING

The rise of algorithmic gatekeeping presents a unique legal question: how can Governments regulate public platforms without curtailing the freedom of speech? In the United States, this issue can be attributed to the First Amendment, which was considered in the case of *Moody v. NetChoice*,⁶⁰ and *NetChoice, LLC v. Paxton*,⁶¹ involving laws in Florida and Texas controlling social media platforms, particularly on matters regarding removing and regulating users' content. These social media platforms argued that deciding what content appears on their services is part of their freedom, recognised and protected under the First Amendment. The Court observed that selecting and presenting content can be protected to some extent, but did not decide the matter of the specific laws, leaving it to the lower courts for further

⁵⁶*K.S. Puttaswamy v. Union of India*, (2017) 10 S.C.C. 1, 340-41.

⁵⁷*Anuradha Bhasin v. Union of India*, (2020) 3 S.C.C. 637, 709-10.

⁵⁸*Information Technology Act*, No. 21 of 2000, §79.

⁵⁹*X Corp.*, supra note 43.

⁶⁰*Moody v. NetChoice, LLC*, 603 U.S. 707, 719-45 (2024).

⁶¹*NetChoice, LLC v. Paxton*, 603 U.S. 707, 719-45 (2024).

consideration. This creates a dilemma for algorithmic regulation, if such regulation is considered as a discretion under editorial provisions,⁶² the government attempts to control what the platforms promote which directly contradicts the platform's freedom of expression. But treating every algorithmic choice as a protected one could give private companies heavy influence over public deliberations, with little opportunity for accountability. The challenge revolves around distinguishing between legitimate editorial decisions and visibility generated by algorithmic systems through engagement.⁶³

A further development can be seen in the case of *Murthy v. Missouri*, in which the Supreme Court considered allegations that government officials had pressured or encouraged social media platforms to remove or suppress certain content. Although the Court decided the case on standing grounds, it did not resolve the underlying First Amendment issue. This case is significant for algorithmic gatekeeping because it demonstrates the constitutional concern that may arise when the government indirectly influences platform moderation. This creates an additional issue identified by this case: regulation may not take the form of direct statutory control but rather may also arise through governmental pressure or influence over private platforms.⁶⁴

Similarly, this problem exists under Section 230 of the Communications Decency Act,⁶⁵ which provides online platforms with the perception that they are not the original publisher or speaker of content created by their users. The debate is whether this specific protection should be applied the same way when platforms display content and when they actively recommend or amplify it through their algorithms. In the case of *Gonzalez v. Google LLC*,⁶⁶ the Supreme Court was presented with the question of whether Section 230 protects the platform of YouTube, to recommend content, but the Court ultimately refrained from answering this question.

The uncertainty surrounding Section 230 is also evident in *Twitter, Inc. v. Taamneh*. While *Gonzalez v. Google LLC* concerned the potential application of Section 230 to algorithmic recommendations, this case addressed whether a platform could be held liable under federal law for allegedly assisting terrorist activity through its services. The Court rejected this

⁶²*Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp. of Bos., Inc.*, 515 U.S. 557, 568–72 (1995).

⁶³Alec Peters, *Machine Manipulation: Why an AI Editor Does Not Serve First Amendment Values*, 95 U. COLO. L. REV. 307 (2024).

⁶⁴*Murthy v. Missouri*, 603 U.S. 43, 49–52 (2024).

⁶⁵47 U.S.C. §230 (c)(1).

⁶⁶*Gonzalez v. Google LLC*, 598 U.S. 617, 621–23 (2023).

particular claim, holding that platform liability depends on whether the platform's own conduct satisfies an independent legal standard, rather than merely on the presence of harmful third-party content. This difference is important when algorithms recommend, organise or amplify content rather than merely making it available.⁶⁷

This leaves a significant legal gap. The recommendation system is not wholly passive because it does, in fact, decide what content users can see and are more likely to see. One possible solution to this legal question is presented in the form of the product- defect theory, which is concerned with the design and manufacture of the platform rather than the media generated by its users and in the judgment of *Lemmon v. Snap Inc.*, the Ninth Circuit allowed a product-liability claim against the app called Snapchat to proceed because the claim was entirely based on the design of Snapchat's product rather than the app's role in publishing and providing a platform for third-party content. This approach represents a step forward, allowing courts to hold platforms accountable for harmful algorithmic design rather than for every piece of content generated by their users.⁶⁸

In *Lemmon v. Snap. Inc.*, the Ninth Circuit distinguished liability based on product design from liability based on the publication of third-party content. The court held that the alleged negligent design of Snapchat could proceed because the duty to design a reasonably safe product was not correlated with the app's role as a publisher or speaker. This difference provides a possible basis for addressing the harmful platform design without imposing liability for every individual piece of user-generated content.⁶⁹

The European Union has deviated from this path, particularly through the Digital Services Act (DSA). Instead of focusing primarily on whether the platform has a constitutional right to control speech, the DSA places greater importance on transparency, risk reduction and maintaining accountability. Article 34,⁷⁰ of this act mandates that very influential online platforms and search engines assess systemic risks created by their users and third parties, while Article 35,⁷¹ This act requires them to take additional measures to reduce these risks. The DSA also provides users with greater freedom, particularly in terms of heavier or more immersive control over the recommendation systems.

⁶⁷*Twitter, Inc. v. Taamneh*, 598 U.S. 471, 478-81, 503-04 (2023).

⁶⁸*Anderson v. TikTok, Inc.*, 116 F.4th 180, 184-91 (3rd Cir. 2024).

⁶⁹*Lemmon v. Snap, Inc.*, 995 F.3d 1085, 1091-94 (9th Cir. 2021).

⁷⁰*Digital Services Act*, art. 34, 2022 O.J. (L 277) 1, 57-58 (Eur. Union).

⁷¹ *Digital Services Act*, art. 35, 2022 O.J. (L 277) 1, 58 (Eur. Union).

The DSA therefore creates a broader governance framework for algorithmic gatekeeping. Article 34 requires large platforms and large search engines to assess systemic risks, including risks arising from the designs and functioning of recommender systems. This creates an ongoing risk assessment and mitigation framework rather than relying only on liability after the harm has occurred.

Under Article 27 of the same Act, platforms must expand on the main factors used by their recommendation systems. Under Article 38,⁷² very influential platforms must provide at least one option that does not rely on profiling. Article 40 also allows qualified researchers to gain access to platform data, making it easier to understand and explain how these systems function in practice. The difference between the two approaches of the United States and the European Union is stark: the former places greater emphasis on free speech and platform rights, whereas the latter emphasises transparency and reducing systemic risks.⁷³ Neither approach provides a comprehensive answer to the question or problem at hand, but presents a way forward.

Articles 27, 38 and 40 further strengthen this framework. Article 27 requires transparency about the main parameters of the systems and allows users to modify or influence them. Article 38 requires large online platforms and search engines to provide at least one recommender system option that is not based on profiling. Article 40 facilitates access to relevant platform data for researchers, supporting independent scrutiny of systemic risks and algorithmic functioning.

RETHINKING ALGORITHMIC GOVERNANCE – A STEP TOWARD DEMOCRATIC DIGITAL PUBLIC SPHERE

The issues created by algorithmic gatekeeping cannot be resolved simply by removing or moderating content, or giving governments greater control over online speech. The problem is multidisciplinary in nature. One feasible solution is the middleware architecture, a proposal associated with Francis Fukuyama.⁷⁴ and several other scholars, potentially separating social media platforms from the algorithms that determine what users see. Instead of permitting one

⁷² Digital Services Act, art. 38, 2022 O.J. (L 277) 1, 60 (Eur. Union).

⁷³ Uta Kohl, *Toxic Recommender Algorithms: Immunities, Liabilities and the Regulated Self-Regulation of the Digital Services Act and the Online Safety Act*, 16 J. MEDIA L. 301, 301–35 (2024).

⁷⁴ See Francis Fukuyama et al., *Middleware for Democracy: Can Technology Help Us Rebuild the Public Sphere?* (2020).

company to control both the platform and the recommendation system, which is the traditional method, users could choose independent algorithms to tailor their feed.⁷⁵

This would provide individuals with greater independence over the information they receive, while reducing the concentration of power.⁷⁶ within a small number of technology firms. A second solution to this problem is the development of bridging algorithms, which are designed not just to highlight engagement but to expose users to diverse opinions and viewpoints, capturing different areas of agreement. Systems such as ‘Polis’⁷⁷ depict how technology can be created around consensus rather than disputes. However, these systems also raise significant questions about what counts as consensus and how they can avoid reproducing existing biases.

A further structural solution would be to incorporate these principles into a statutory framework for India. A proposed Digital India Algorithmic framework could address these problems. In particular, Section 79 of the safe-harbour protection could be split into two parts so that platforms retain immunity for hosting third-party content. In contrast, algorithmic systems that actively recommend, rank, or amplify content are subject to separate product-design and safety obligations.⁷⁸ Such an approach would prevent treating all platform activity as equivalent while creating greater accountability for the design choices that shape information exposure.

The framework should also introduce an Open Curation Protocol through India’s Digital Public Infrastructure, rather than mandating users to remain dependent on the recommendation system developed by a single platform; an open protocol could allow independent and interoperable recommendation systems to operate across platforms.⁷⁹ This would help users gain greater control over how information is curated while reducing the concentration of algorithmic power within individual technology companies. Such

⁷⁵ Francis Fukuyama et al., *Middleware for Democracy: A Proposed Solution to Social Media’s Algorithmic Harms* (2020).

⁷⁶ See Francis Fukuyama, Barak Richman & Ashish Goel, *How to Save Democracy from Technology: Ending Big Tech’s Information Monopoly*, 119 FOREIGN AFFS. 98, 98-109 (2020).

⁷⁷ Colin Megill, Katie McCulloch & Christopher Small, *Polis: Scaling Deliberation by Combining Human and Machine Intelligence* (2017).

⁷⁸ Information Technology Act, No. 21 of 2000, § 79, India Code (2000).

⁷⁹ Francis Fukuyama et al., *Middleware for Democracy: A Proposed Solution to Social Media’s Algorithmic Harms* (2020).

interoperability would complement the middleware model by creating a technical infrastructure through which alternative systems could compete and develop.⁸⁰

Independent statutory oversight would be crucial to ensure that the platforms themselves do not determine these obligations. A proposed Digital Regulatory Authority of India (DRAI) could oversee these functions. Its role would be limited to supervising the processes through which algorithmic systems operate and affect public discourse. Taken together, middleware, consensus-oriented algorithmic and independent statutory oversight would shift the focus of regulation away from controlling individual speech and towards regulating the structural conditions under which digital speech is curated.⁸¹

Finally, greater transparency and public oversight are necessary. Social media platforms must provide valuable information about how their recommendation systems operate and participate in independent audits.⁸² where algorithms face significant effects on public discourse, increasing transparency would streamline the process, making it easier for researchers, regulators and the general public to identify harmful repeated behaviour without requiring the government to determine which speech should be promoted or suppressed. Broadly, democratic societies should reconsider their values regarding whether essential spaces for public discussion should depend on advertising-driven business models that reward attention above all else. The goal must not be to eliminate or abolish algorithmic curation, but rather to change the objective of such curation. Instead of emphasizing maximising engagement, digital platforms should be encouraged to promote user autonomy and exposure to a multitude of perspectives to support meaningful democratic participation.⁸³ This would help reinvent technology as a tool for strengthening the marketplace of ideas rather than taking its place.

Your One Stop Legal Destination

CONCLUSION

The traditional marketplace of ideas assumes that every citizen encounters a multitude of competition viewpoints on relatively equal terms, giving space for truthful and reasoned arguments to emerge through public debate. Algorithmic gatekeeping challenges this assumption by making visibility

⁸⁰Francis Fukuyama, *Making the Internet Safe for Democracy*, 32 J. DEMOCRACY 37, 37-44 (2021).

⁸¹See Digital Services Act arts. 34-40, 2022 O.J. (L 277) 1, 57-63 (Eur. Union).

⁸² Digital Services Act, art. 37, 2022 O.J. (L 277) 1, 60-63 (Eur. Union).

⁸³ Efrat Nechushtai & Seth C. Lewis, *What Kind of News Gatekeepers Do We Want Machines to Be? Filter Bubbles, Fragmentation, and the Normative Dimensions of Algorithmic Recommendations*, 90 COMPUTERS HUM. BEHAV. 298, 298-307 (2019).

dependent on engagement, emotional reaction and commercial incentives rather than on the material of ideas. As this article has demonstrated, the resulting of such algorithmic manipulation can weaken the democratic sphere. The solution however, cannot be equated to greater governmental control over speech. A more sustainable approach requires transparency, independent oversight and user autonomy. The ultimate goal should not be to eliminate algorithmic curation but to redesign it so that its objectives, supporting exposure to diverse perspectives and meaningful democratic participation rather than focusing on maximizing attention.



Your One Stop Legal Destination