

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

Prof. Deepak Kem

Dr. K. R. Narayanan Centre for Dalit and Minorities Studies, Jamia Millia Islamia New Delhi-110025

ABSTRACT: The rapid expansion of social media has transformed the production, circulation and consumption of information and has fundamentally altered the nature of public discourse in India. Social media platforms are no longer merely passive intermediaries that host user-generated content. Through recommender systems, ranking mechanisms, automated content moderation, profiling and targeted advertising, they increasingly influence what users encounter, what information gains visibility and what forms of speech are amplified or marginalised. This transformation raises a regulatory question that extends beyond conventional content moderation: to what extent should platforms be accountable for the design and consequences of algorithmic systems that structure digital communication and public discourse?

India's regulatory framework, principally comprising the Information Technology Act, 2000 and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, as subsequently amended, has progressively expanded intermediary due-diligence, content moderation and grievance-redressal obligations. However, the regulatory framework remains substantially oriented towards intermediary liability and content governance rather than the systemic risks associated with algorithmic curation, recommendation and amplification. In particular, comprehensive and enforceable requirements relating to algorithmic transparency, systemic-risk assessment, independent auditing, recommender-system governance, researcher access and algorithmic impact assessment remain underdeveloped.

This paper examines the transition from content-centred regulation towards algorithmic accountability in India. It adopts a qualitative, doctrinal and comparative methodology, analysing legislation, subordinate legislation, constitutional principles, judicial decisions, policy documents and scholarly literature. India's emerging framework is comparatively examined against the European Union's Digital Services Act and the comparatively fragmented and platform-led regulatory model of the United States. The paper argues that India's principal regulatory challenge is not simply a deficiency of content moderation but a regulatory mismatch between conventional intermediary regulation and the increasingly systemic power exercised by algorithmic infrastructures. It proposes a rights-based, risk-sensitive and proportionate framework centred on meaningful transparency, algorithmic impact assessment, independent auditing, systemic-risk mitigation, procedural safeguards, researcher access, effective user redress and independent institutional oversight. Such a framework should regulate demonstrable risks rather than technological innovation itself, thereby maintaining an appropriate balance between platform accountability, democratic freedom, privacy, innovation and freedom of expression.

KEYWORDS: Algorithmic Accountability; Platform Governance; Social Media Regulation; Intermediary Liability; Content Moderation; Recommender Systems; Digital Services Act; Artificial Intelligence; Digital Rights; Transparency; India

INTRODUCTION

Social media platforms have become central infrastructures of contemporary communication, information exchange and public discourse. In India, their significance extends well beyond interpersonal communication. Social media increasingly influences political communication, electoral mobilisation, news consumption, public opinion, commercial advertising, social movements and the formation of digital communities. The transformation is therefore not merely quantitative, reflected in the growing number of users and volume of content, but also qualitative. The architecture through which information reaches users has itself become algorithmically mediated.

Contemporary social media platforms do not simply provide a technological space in which users communicate. They determine, to varying degrees, which content becomes visible, the order in which it appears, the users to whom it is recommended and the extent to which it is amplified or deprioritised. Recommender systems, ranking algorithms, personalised feeds, automated

moderation tools and advertising systems have therefore become important mechanisms through which platforms exercise social and communicative power.

This development creates a significant challenge for conventional approaches to intermediary regulation. Traditional intermediary regulation is primarily concerned with questions such as whether a platform should be legally responsible for unlawful third-party content, whether it has complied with prescribed due-diligence obligations and whether it has appropriately responded to complaints or takedown requests. Algorithmic accountability raises a broader question: **how should platforms be held responsible for the technological systems through which they determine the visibility, ranking, recommendation, amplification or suppression of information?**

The distinction is important because the influence of platforms is not confined to content that they remove. A platform may comply with content-removal requirements while its recommender system simultaneously gives disproportionate visibility to particular categories of information. Similarly, content may remain technically available but receive substantially reduced visibility because of ranking or recommendation decisions. Consequently, regulation that focuses exclusively on the legality of individual items of content may overlook the systemic effects produced by the architecture through which information is distributed.

India has developed a substantial legal and regulatory framework governing intermediaries. Section 79 of the Information Technology Act, 2000 provides conditional protection to intermediaries from liability for third-party information, subject to specified statutory conditions. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 subsequently expanded the due-diligence and compliance framework applicable to intermediaries and significant social media intermediaries. The framework has continued to evolve through subsequent amendments and policy initiatives.

These developments demonstrate that India does not lack regulation. The more significant concern is that the regulatory framework has historically been organised primarily around **intermediary liability, unlawful content, due diligence and procedural compliance**, while comparatively less attention has been devoted to the systemic risks generated by algorithmic recommendation, ranking and amplification.

The regulatory challenge is particularly important in the Indian context. India's digital public sphere operates within a highly diverse social environment characterised by linguistic plurality, regional differences, socio-economic inequality, caste and gender hierarchies, political contestation and varying levels of digital literacy. Algorithmic systems designed around engagement, personalisation and behavioural prediction may interact with these structural inequalities in ways that are difficult to detect through conventional content-moderation mechanisms.

Algorithmic accountability therefore has both a technological and a constitutional dimension. It concerns transparency and explainability, but also equality, freedom of expression, privacy, procedural fairness and democratic participation. The question is not whether algorithms are inherently harmful. Algorithms can improve information discovery, accessibility and user experience. The regulatory concern arises when their design, operation or consequences create significant and foreseeable risks to individuals or society without adequate mechanisms of scrutiny and redress.

The European Union's Digital Services Act (DSA) represents an important development in this regard. Rather than focusing exclusively on individual items of unlawful content, the DSA adopts a risk-based approach for the largest platforms and search engines, including obligations concerning systemic-risk assessment, mitigation, independent auditing, transparency and recommender systems. The European framework therefore provides a useful comparative reference for examining how platform responsibility can move beyond conventional intermediary liability.

The United States presents a substantially different model. Its regulatory environment combines strong constitutional protection for speech, intermediary liability protections, sectoral regulation, litigation and extensive platform self-governance. This model provides an important counterpoint because it highlights the risks of excessive state intervention and the importance of preserving private editorial autonomy, while also illustrating the limitations of fragmented regulation.

Against this background, this paper examines the following central question:

To what extent does India's existing social media regulatory framework provide meaningful algorithmic accountability, and what legal and institutional reforms are required to ensure responsible platform governance while protecting innovation and fundamental rights?

The paper argues that India should move from a predominantly **content-moderation paradigm towards a risk-based algorithmic-accountability paradigm**. This does not require direct governmental control over algorithms or compulsory disclosure of proprietary source code. Rather, it requires legally enforceable mechanisms through which platforms with substantial systemic impact can be required to identify, assess, mitigate and demonstrate accountability for algorithmic risks.

Understanding Algorithmic Accountability

Meaning and Scope

Algorithmic accountability refers broadly to the mechanisms through which organisations deploying consequential algorithmic systems can be required to explain, justify, monitor and take responsibility for the operation and consequences of those systems.

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

The concept should not, however, be reduced to transparency. Transparency is an important component of accountability, but disclosure by itself does not guarantee fairness, intelligibility or effective oversight. An algorithm may be disclosed without being understandable to ordinary users or independently assessable by regulators and researchers.

Algorithmic accountability therefore incorporates at least five interconnected dimensions:

1. **Transparency:** disclosure of meaningful information concerning the purpose, operation and governance of algorithmic systems.
2. **Explainability:** the capacity to provide intelligible reasons for significant automated decisions or outcomes.
3. **Auditability:** the possibility of independent assessment of relevant systems, processes and outcomes.
4. **Fairness and non-discrimination:** identification and mitigation of discriminatory or systematically unequal outcomes.
5. **Remedial accountability:** effective mechanisms through which affected persons can challenge decisions and obtain appropriate remedies.

In the context of social media, these dimensions become particularly important because platforms exercise power not only through removal but also through **distribution**.

A conventional content-moderation framework asks whether content should be permitted or removed. An algorithmic-accountability framework asks an additional set of questions:

- Why was particular content recommended?
- Why was it shown to one group of users and not another?
- What factors determine its ranking?
- What incentives are embedded in the optimisation process?
- What risks arise from repeated amplification?
- Can users understand and challenge consequential decisions?
- Can independent actors verify the platform's claims?

These questions shift the regulatory focus from individual pieces of content towards the technological architecture of information distribution.

Algorithmic Governance and Governance of Algorithms

A distinction should also be made between **algorithmic governance** and **governance of algorithms**.

Algorithmic governance refers to the use of computational systems to organise information, make predictions, classify users and assist or determine decisions.

Governance of algorithms refers to the legal, institutional and ethical mechanisms through which such systems are regulated.

This paper is primarily concerned with the latter. Its objective is not to determine how platforms should technically construct their algorithms but to examine what accountability obligations should apply when algorithmic systems have significant implications for rights and public interests.

Platform Governance and the Changing Nature of Intermediary Power

The traditional concept of an intermediary assumes a relatively passive technological actor connecting users to information. Contemporary platforms complicate this assumption.

Platforms establish community standards, determine content visibility, design interfaces, curate feeds, recommend content, personalise advertisements, suspend accounts and employ automated systems to enforce their rules. They therefore perform functions that resemble a combination of technological infrastructure, private regulation and media curation.

This development has generated the concept of **platform governance**. Platforms govern not only through explicit rules but also through technical design. Code, ranking systems, default settings, recommendation mechanisms and interface architecture can influence user behaviour as effectively as formal rules.

Algorithmic systems are consequently part of the institutional structure through which platforms exercise power.

The significance of this development becomes clearer when the distinction between **hosting** and **curation** is considered.

A platform that merely stores content occupies a different regulatory position from a platform that actively selects, ranks and recommends content to particular users. Recommendation involves an additional layer of decision-making. The platform is not merely transmitting information; it is constructing an informational environment.

The more extensive this role becomes, the stronger the case for corresponding accountability obligations.

This does not mean that platforms should automatically lose intermediary protections whenever they use algorithms. Such an approach would risk creating uncertainty and could undermine legitimate technological innovation. Instead, the regulatory framework should distinguish between different categories of platform activity and impose obligations proportionate to systemic risk.

REVIEW OF LITERATURE

Platform Governance

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

The literature on platform governance demonstrates that digital platforms should not be understood merely as neutral intermediaries. Gorwa's work conceptualises platform governance as an interaction between private platform rules, state regulation and broader social norms.

This literature is important because it establishes that platforms exercise normative authority. Their community standards and enforcement systems determine the conditions under which speech can circulate.

The implication for India is that regulatory analysis cannot stop at the question of intermediary liability. It must also consider how private platform rules and technical architectures interact with public values.

Algorithmic Accountability

The literature on algorithmic accountability challenges the assumption that algorithmic decision-making is inherently neutral or objective. Algorithms embody choices concerning datasets, optimisation objectives, classifications and institutional priorities.

Binns has emphasised the relationship between algorithmic accountability and public reason, while Ananny and Crawford caution against treating transparency as an adequate solution to algorithmic opacity. Their work suggests that meaningful accountability requires more than technical disclosure. It requires institutions capable of interpreting information, evaluating consequences and providing remedies.

This insight is particularly relevant to social media governance. A platform's publication of a general description of its recommender system may satisfy a limited conception of transparency without enabling meaningful scrutiny of systemic harms.

Automated Content Moderation

Research on algorithmic content moderation highlights the tension between the scale of online communication and the limitations of automated decision-making. Platforms process enormous quantities of content, making complete human moderation impractical. Automation consequently becomes operationally attractive.

However, automated moderation can struggle with context-dependent expression, satire, political speech, cultural references and multilingual communication. Errors can disproportionately affect marginalised communities whose languages, experiences and modes of expression may be inadequately represented in training datasets.

The problem therefore has an important equality dimension. Algorithmic accountability must address not only whether platforms make mistakes but whether such errors are distributed unevenly across social groups.

Digital Rights and Freedom of Expression

The literature on online speech demonstrates that platform regulation presents a difficult balance between addressing harmful content and protecting freedom of expression.

David Kaye's work emphasises the importance of human-rights principles in regulating online platforms. Napoli similarly argues for reconsidering social media governance in the context of disinformation and public-interest communication.

This scholarship is particularly relevant to India because regulation of social media operates within a constitutional framework protecting freedom of speech and expression while permitting reasonable restrictions.

The challenge is therefore not to maximise regulation but to design **proportionate and procedurally fair regulation**.

Surveillance, Data and Platform Power

Zuboff's analysis of surveillance capitalism provides another dimension to the debate. Platforms increasingly rely on large-scale data collection, behavioural prediction and personalised advertising. Algorithmic governance is therefore closely connected to data governance.

The Digital Personal Data Protection Act, 2023 adds an important layer to India's regulatory architecture by establishing statutory principles concerning the processing of digital personal data. However, data protection and algorithmic accountability should not be treated as identical. A platform may comply with data-processing requirements while its recommendation system creates other forms of systemic harm.

Research Gap

The existing scholarship provides substantial conceptual foundations for understanding platform governance, algorithmic accountability and digital rights. However, an important gap remains in connecting these debates to India's evolving intermediary regulatory framework.

Indian regulatory discussions have predominantly concentrated on intermediary liability, content removal, due diligence and grievance redressal. Comparatively less attention has been given to the systems through which platforms determine the distribution, ranking and amplification of content.

The present study addresses this gap by conceptualising algorithmic accountability as a distinct regulatory layer and examining how India can move from content-centred regulation towards systems-level platform governance.

METHODOLOGY

Research Design

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

The study adopts a **qualitative, doctrinal, analytical and comparative research design**.

The doctrinal component examines relevant legislation, subordinate legislation, constitutional provisions and judicial decisions. The analytical component identifies regulatory gaps, conceptual tensions and institutional weaknesses. The comparative component examines the European Union and United States approaches to identify principles that may inform Indian regulatory development.

The study is primarily normative. It does not seek to establish statistical causality between algorithmic opacity and misinformation. Rather, it evaluates the adequacy of existing legal mechanisms for addressing algorithmic risks and develops normative proposals for strengthening platform governance.

Sources of Data

Primary sources include:

- the Information Technology Act, 2000;
- the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 and subsequent amendments;
- the Digital Personal Data Protection Act, 2023;
- relevant constitutional provisions;
- Supreme Court judgments;
- governmental policy documents; and
- relevant regulatory and institutional materials.

Comparative primary materials include the EU Digital Services Act and relevant US statutory, regulatory and judicial materials. Secondary sources include peer-reviewed journal articles, books, research reports and scholarship on platform governance, algorithmic accountability, artificial intelligence, content moderation, digital rights and disinformation.

Analytical Framework

The study assesses regulatory frameworks across six dimensions:

Regulatory Dimension	Central Question
Transparency	What meaningful information must platforms disclose?
Explainability	Can consequential algorithmic decisions be understood?
Auditability	Can independent actors examine relevant systems?
Risk Assessment	Must platforms identify and mitigate systemic risks?
Procedural Justice	Do affected users receive notice, reasons and review?
Institutional Oversight	Who independently supervises platform compliance?

This framework is used to compare India, the EU and the US without assuming that regulatory provisions can be transplanted mechanically across different constitutional and institutional contexts.

India's Existing Legal Framework

Information Technology Act, 2000

The Information Technology Act, 2000 remains the principal statutory foundation of India's digital regulatory framework.

Section 79 provides conditional protection to intermediaries from liability for third-party information, data or communication links made available through their systems, subject to specified conditions and due-diligence requirements.

The provision is significant because intermediary protection has historically been central to India's approach to platform regulation. It establishes an important balance between protecting intermediaries from automatic liability and requiring them to meet specified statutory conditions.

However, Section 79 is not an algorithmic-accountability provision. It does not establish a comprehensive framework for assessing recommender systems, algorithmic amplification, systemic risks or algorithmic discrimination.

This distinction is fundamental to the argument of the present paper.

IT Rules, 2021

The IT Rules, 2021 substantially expanded the regulatory obligations imposed upon intermediaries, particularly significant social media intermediaries.

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

The Rules introduced or strengthened mechanisms concerning:

- due diligence;
- grievance redressal;
- compliance officers;
- information preservation;
- reporting requirements;
- content-related obligations; and
- additional obligations for significant social media intermediaries.

These provisions represent an important evolution in Indian platform governance.

Nevertheless, the principal regulatory logic remains centred on **platform compliance with prescribed procedures and content-related obligations**.

The Rules therefore provide a foundation upon which algorithmic accountability could be developed, but they do not themselves constitute a comprehensive algorithmic-governance framework.

Subsequent Regulatory Developments

The regulatory framework has continued to evolve through amendments and policy initiatives. Consequently, the 2021 Rules should not be treated as the final expression of India's digital governance policy.

The continuing evolution of India's regulatory framework demonstrates both the significance and difficulty of governing rapidly changing digital technologies. It also reinforces the need for a framework that is technologically neutral and capable of addressing future algorithmic systems without requiring constant legislative reconstruction.

Constitutional Framework

Algorithmic accountability must also be situated within India's constitutional framework.

Three fundamental rights are particularly relevant.

Article 14: Equality and Non-Arbitrariness

Algorithmic systems may produce discriminatory outcomes even where discrimination is not intentional. The principle of equality therefore raises questions concerning the fairness and non-arbitrariness of high-impact algorithmic systems.

Article 19(1)(a): Freedom of Speech and Expression

Algorithmic ranking and content moderation directly affect the visibility and circulation of speech. Both state regulation and private platform decisions can therefore influence the effective exercise of expressive freedom.

Article 21: Privacy and Personal Liberty

Personalised recommendation systems often depend upon extensive data collection and behavioural profiling. Algorithmic accountability must therefore be considered alongside privacy and data-protection principles.

The constitutional framework consequently requires a careful balance. Regulation should address genuine systemic harms while avoiding unnecessary interference with lawful expression, privacy and technological innovation.

From Content Moderation to Algorithmic Accountability

The principal weakness in the existing regulatory approach is best understood through the distinction between **content moderation** and **algorithmic governance**.

Content moderation concerns decisions about whether particular content violates applicable rules.

Algorithmic governance concerns the broader systems that determine:

- what content users encounter;
- how content is ranked;
- which content is recommended;
- what becomes viral;
- which accounts receive visibility;
- how advertisements are targeted; and
- how users are categorised.

The difference can be represented as follows:

Content Moderation Paradigm	Algorithmic Accountability Paradigm
Focuses on individual content	Focuses on systems and outcomes
Primarily reactive	Primarily preventive

Content Moderation Paradigm	Algorithmic Accountability Paradigm
Removal and restriction	Risk assessment and mitigation
Platform compliance	Platform responsibility
Internal review	Independent oversight
Individual complaints	Individual and systemic redress
Transparency reports	Risk-relevant transparency
Content legality	Rights and systemic impact

The transition proposed in this paper is therefore not a replacement of content moderation. Content moderation remains necessary. The argument is that it must be supplemented by a second regulatory layer concerned with algorithmic systems.

Comparative Analysis

European Union: Risk-Based Platform Governance

The European Union's Digital Services Act represents one of the most significant contemporary attempts to regulate large digital platforms through a systemic-risk framework.

The DSA recognises that very large online platforms and search engines can generate risks extending beyond individual pieces of unlawful content. It therefore requires these entities to assess and mitigate systemic risks, including risks relating to fundamental rights, civic discourse and certain forms of manipulation.

The framework also incorporates independent auditing, enhanced transparency obligations, researcher access and user choice concerning recommender systems.

The central lesson for India is not that the DSA should be copied wholesale. The more relevant lesson is the principle that **platform obligations can be calibrated according to systemic importance and risk**.

This approach moves beyond the assumption that all intermediaries should be regulated in essentially the same way.

United States: Fragmented and Platform-Led Governance

The US model differs substantially from the EU approach.

Its regulatory architecture is shaped by:

- intermediary liability protections;
- First Amendment principles;
- sector-specific regulation;
- federal and state enforcement;
- litigation; and
- private platform governance.

The principal strength of this approach is the strong protection it affords private editorial and expressive autonomy. Its limitation is fragmentation and the absence of a comprehensive federal systemic-risk framework equivalent to the DSA.

The US experience therefore demonstrates that algorithmic accountability must be designed carefully to avoid converting legitimate regulation into government control over lawful speech.

India in Comparative Perspective

The comparative analysis suggests that India currently occupies an intermediate position.

India has moved beyond a purely laissez-faire approach by imposing substantial intermediary obligations. However, its regulatory framework remains less developed than the EU's systemic-risk architecture with respect to algorithmic accountability.

At the same time, India should not simply reproduce the European approach. India's constitutional structure, administrative capacity, social diversity, market conditions and institutional environment require a context-sensitive model.

The appropriate Indian framework should therefore combine:

the EU's emphasis on systemic risk and independent accountability + the US emphasis on expressive freedom and limits on state control + India's own constitutional commitment to equality, liberty, privacy and due process.

Doctrinal Regulatory Gap Analysis

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

Regulatory Dimension	India	European Union	United States	Principal Indian Gap
Intermediary liability	Established	Established	Strong statutory protection	Limited
Content moderation	Extensive obligations	Extensive obligations	Predominantly platform-led	Moderate
Algorithmic transparency	Limited	More developed for systemic platforms	Fragmented	Significant
Systemic-risk assessment	Limited	Explicit for largest platforms	Fragmented	Significant
Independent algorithmic audit	No comprehensive equivalent	Required for relevant large platforms	Fragmented	Significant
Recommender-system governance	Limited	Specific obligations	Limited	Significant
Researcher access	Limited	Structured mechanisms	Fragmented	Significant
User redress	Developed	Developed	Primarily platform-dependent	Moderate
Independent oversight	Fragmented	Institutionalised	Fragmented	Significant
Algorithmic impact assessment	Underdeveloped	Risk-based mechanisms	Sectoral/fragmented	Significant

The comparison indicates that India's principal regulatory gap lies not in the absence of content regulation but in the absence of a comprehensive **systems-level accountability architecture**.

Algorithmic Bias and the Indian Context

Algorithmic bias is often discussed as a purely technical problem. Such an approach is inadequate.

Bias may originate in:

- training data;
- classification categories;
- optimisation objectives;
- proxy variables;
- linguistic limitations;
- platform policies;
- user behaviour; and
- historical inequalities reproduced through data.

In India, the problem has additional complexity because digital communication operates across multiple languages, regions and social groups.

A moderation or recommendation system that performs adequately in English may operate differently in Indian languages. Similarly, an algorithm trained on globally dominant datasets may fail to understand local political, cultural and social contexts. Algorithmic bias should therefore be treated as a potential **equality and representation problem**, not merely as an engineering error.

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

This has particular implications for marginalised communities. If algorithmic systems systematically under-recognise, over-moderate, under-recommend or misclassify particular forms of expression, technological design may reproduce existing inequalities within the digital public sphere.

Algorithmic accountability must consequently include mechanisms for assessing differential impacts across relevant linguistic and social contexts.

Algorithmic Amplification and the Democratic Public Sphere

One of the most important consequences of recommender systems is the transformation of visibility into a technologically mediated resource.

In a conventional media environment, editorial institutions determine what information receives prominence. In social media environments, this function is increasingly distributed through algorithmic systems.

This creates a new form of power: **algorithmic gatekeeping**.

The central concern is not that algorithms necessarily promote false or polarising information. Rather, optimisation for engagement may create incentives under which content capable of generating strong emotional reactions receives disproportionate visibility. The democratic significance of this process is considerable.

If algorithmic systems influence which political arguments reach large audiences, then recommender systems become part of the infrastructure of democratic communication.

This does not justify governmental determination of what users should see. Such an approach would create serious risks to freedom of expression.

The appropriate regulatory response is therefore to make the **process of amplification accountable rather than the content predetermined**.

In other words, the state should regulate the conditions under which high-impact algorithmic systems operate rather than becoming the arbiter of permissible political opinion.

Procedural Justice and User Rights

Algorithmic accountability also requires stronger procedural safeguards.

Where platforms make significant decisions concerning content, accounts or visibility, users should, subject to legitimate exceptions, have access to:

1. meaningful notice;
2. intelligible reasons;
3. an opportunity to seek review;
4. timely appeal;
5. human intervention in appropriate cases; and
6. an effective escalation mechanism.

The purpose of these safeguards is not merely administrative fairness. They recognise that platform decisions may have significant consequences for political participation, reputation, livelihood and access to information.

Procedural justice is therefore an essential bridge between algorithmic accountability and fundamental rights.

Towards an Indian Framework of Algorithmic Accountability

The analysis suggests the need for a layered and risk-sensitive framework rather than a single universal regulatory model.

Risk-Based Classification

Regulatory obligations should correspond to the scale, reach and potential impact of platforms.

A small platform serving a limited user base should not face the same compliance burden as a platform capable of influencing national political discourse.

The regulatory framework should therefore identify **systemically significant platforms** and impose enhanced obligations upon them.

Algorithmic Impact Assessments

Platforms with substantial systemic reach should be required to conduct periodic algorithmic impact assessments for high-risk systems.

Assessments should consider:

- freedom of expression;
- equality and non-discrimination;
- privacy;
- misinformation and manipulation;
- risks to children;
- political and electoral communication;
- linguistic and cultural bias; and

From Content Moderation to Algorithmic Accountability: Rethinking Social Media Platform Governance in India

- effects on vulnerable communities.

The objective should be preventive rather than punitive.

Independent Auditing

Independent auditing should be introduced for high-impact algorithmic systems.

Audits should assess:

- governance procedures;
- documented objectives;
- foreseeable risks;
- bias;
- moderation accuracy;
- recommender-system effects;
- mitigation measures; and
- compliance with applicable legal requirements.

Auditors must possess technical competence and sufficient independence from the platforms being assessed.

Meaningful Transparency

India should move beyond generic transparency reporting towards **structured and risk-relevant transparency**.

Platforms should disclose, subject to legitimate confidentiality protections:

- the principal categories of recommender systems;
- the objectives and significant parameters governing ranking;
- material changes to recommendation systems;
- risk-assessment findings;
- mitigation measures;
- moderation error rates;
- appeal outcomes; and
- information necessary for qualified independent research.

Transparency should be meaningful rather than merely formal.

Researcher and Civil Society Access

Independent research is essential for identifying systemic harms that regulators and platforms may overlook.

Qualified researchers should therefore be provided controlled access to relevant platform data, subject to appropriate safeguards concerning privacy, security and commercial confidentiality.

Civil society organisations should also have opportunities to contribute evidence concerning the effects of algorithmic systems on vulnerable and marginalised communities.

Stronger User Redress

Users should be able to challenge significant platform decisions through accessible and timely procedures.

The redress mechanism should distinguish between routine moderation decisions and high-impact decisions involving political expression, public-interest communication, account suspension or systematic deprioritisation.

Independent Oversight

India should consider establishing or designating a specialised institutional mechanism for algorithmic and platform accountability.

Such an institution should be independent in its oversight functions and should possess expertise in law, technology, human rights and data governance.

Its function should not be to determine which political views or opinions are acceptable.

Its role should instead be to supervise whether systemically important platforms comply with legally prescribed requirements concerning risk assessment, transparency, auditing, procedural fairness and user rights.

Protection of Innovation

Algorithmic regulation should be guided by proportionality and technological neutrality.

The objective should not be mandatory disclosure of proprietary source code or prior approval of every algorithmic innovation.

Rather, regulation should focus on **demonstrable and foreseeable risks**.

This approach would enable India to protect technological innovation while establishing meaningful accountability for systems capable of producing significant social consequences.

DISCUSSION

The central finding of this doctrinal and comparative analysis is that India's principal challenge is not an absence of social media regulation but a **mismatch between the object of regulation and the technological organisation of contemporary platforms**. The existing framework performs important functions. Intermediary due diligence, grievance redressal, content moderation and mechanisms addressing unlawful content remain essential components of digital governance.

However, these mechanisms largely operate at the level of individual content and individual complaints.

Algorithmic governance operates at a different level.

It concerns the systems through which millions of individual decisions are made concerning visibility, ranking, recommendation, profiling and amplification.

This distinction can be expressed as:

Content creation → Algorithmic selection → Amplification → User exposure → Social consequences

Traditional intermediary regulation focuses heavily on content and platform responses to content.

Algorithmic accountability requires regulatory attention to the entire chain.

The significance of this approach is particularly apparent in relation to democratic communication. A platform can comply with individual takedown requirements while its algorithmic architecture continues to influence political polarisation, misinformation, social exclusion or unequal visibility.

The paper therefore proposes a conceptual shift:

from regulating content alone to regulating the conditions under which algorithmic systems distribute content.

This shift also changes the meaning of platform responsibility.

Platform responsibility should not mean that platforms become guarantors of every consequence produced by their users. Such a standard would be unrealistic and could undermine freedom of expression.

Rather, responsibility should be proportionate to:

- the platform's systemic reach;
- the foreseeability of the risk;
- the degree of control exercised by the platform;
- the potential severity of harm; and
- the feasibility of mitigation.

This produces a more defensible principle of **proportionate algorithmic responsibility**.

CONCLUSION

Social media regulation in India is entering a qualitatively different phase. The central regulatory question can no longer be confined to whether platforms remove unlawful content or comply with intermediary due-diligence requirements. The deeper question concerns the algorithmic systems through which platforms organise visibility, recommendation, ranking, amplification and user interaction.

India has developed an increasingly elaborate framework of intermediary governance through the Information Technology Act, 2000 and the IT Rules, 2021, together with subsequent amendments and broader developments in digital and data governance. However, the technological transformation of platforms requires a corresponding transformation in the regulatory framework.

The central argument of this paper is that India should move from a predominantly **content-moderation paradigm towards a risk-based algorithmic-accountability paradigm**.

Such a transition does not require unrestricted governmental access to proprietary algorithms, direct state control over recommendation systems or excessive regulation of lawful speech. Instead, it requires a framework through which platforms with substantial systemic impact are required to identify, assess, mitigate and disclose relevant algorithmic risks.

The comparative experience of the European Union demonstrates the value of systemic-risk assessment, independent auditing, enhanced transparency, researcher access and user choice in recommender systems. The United States, in contrast, demonstrates the importance of protecting expressive autonomy and limiting governmental interference while also revealing the limitations of fragmented and predominantly platform-led governance.

India therefore needs a distinct regulatory model that draws selectively from both approaches while remaining grounded in its own constitutional and institutional context.

Such a framework should rest on six principles:

1. Meaningful transparency
2. Risk-based algorithmic impact assessment
3. Independent auditing
4. Procedural fairness and user redress
5. Independent institutional oversight

6. Protection of fundamental rights and innovation

The objective should not be to regulate algorithms simply because they are algorithms. Regulation should instead focus on systems whose design or operation creates significant and foreseeable risks to fundamental rights, democratic discourse or public welfare. The transition from content moderation to algorithmic accountability is therefore more than a technical modification to India's social media laws. It represents a fundamental reconceptualisation of platform power.

When platforms function as important infrastructures of public communication, the technological systems through which they determine visibility and participation cannot remain outside meaningful structures of accountability.

The future of social media regulation in India should consequently be based not on choosing between innovation and regulation, but on establishing the institutional conditions under which **innovation can coexist with transparency, accountability, constitutional rights and democratic responsibility**.

REFERENCES

- 1) Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989.
- 2) Binns, R. (2018). Algorithmic accountability and public reason. *Philosophy & Technology*, 31, 543–556.
- 3) Bradshaw, S., & Howard, P. N. (2021). *The Global Disinformation Order: 2021 Global Inventory of Organised Social Media Manipulation*. Oxford Internet Institute.
- 4) European Union. (2022). *Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act)*. Official Journal of the European Union.
- 5) Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center Research Publication.
- 6) Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, 22(6), 854–871.
- 7) Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1).
- 8) Government of India. (2000). *Information Technology Act, 2000*. New Delhi: Government of India.
- 9) Government of India. (2023). *Digital Personal Data Protection Act, 2023*. New Delhi: Government of India.
- 10) Kaye, D. (2019). *Speech Police: The Global Struggle to Govern the Internet*. Columbia Global Reports.
- 11) Napoli, P. M. (2019). *Social Media and the Public Interest: Media Regulation in the Disinformation Age*. Columbia University Press.
- 12) Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.
- 13) Supreme Court of India. *Shreya Singhal v. Union of India*, (2015) 5 SCC 1.
- 14) Ministry of Electronics and Information Technology. (2021, as amended). *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*. Government of India.
- 15) Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.



There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non-Commercial 4.0 International (CC BY-NC 4.0)

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.