

The Right to Privacy vis-à-vis Freedom of Expression on Social Media: A Constitutional and Statutory Analysis of India's Governance Framework

Srinivas Krishnan¹, Dr. Sameer Kumar Dwivedi²

¹PhD Scholar, Department of Law, Shri Venkateshwara University, Gajraula, Amroha, Uttar Pradesh

²Professor, Department of Law, Shri Venkateshwara University, Gajraula, Amroha, Uttar Pradesh

Abstract

The unprecedented growth of social media in India has generated an acute conflict between two constitutionally guaranteed fundamental rights — the right to privacy under Article 21 and the freedom of expression under Article 19(1)(a). With over 462 million active social media users in India, the digital public sphere has become both the principal arena of democratic participation and the principal site of systematic privacy violation. This paper examines the legal, constitutional, and philosophical dimensions of this conflict, analysing the adequacy of India's existing governance framework — comprising the Information Technology Act 2000, the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, and the Digital Personal Data Protection Act 2023 — in addressing the harms that social media generates. Drawing on the foundational theories of Warren and Brandeis, Alan Westin, Ruth Gavison, and Daniel Solove, the Indian Supreme Court's landmark judgment in Justice K.S. Puttaswamy (Retd.) v. Union of India (2017), and comparative analysis of the European Court of Human Rights jurisprudence, the EU Digital Services Act 2022, and the UK Online Safety Act 2023, the paper identifies five principal categories of privacy-expression conflict that social media produces and evaluates the adequacy of India's legal response to each. It further analyses the constitutional challenges generated by the Kunal Kamra fact-checking amendment litigation, the Tanul Thakur blocking regime case, the Sahyog Portal controversy, and the Rajeshwari Khandasami non-consensual intimate images judgment. The paper concludes by proposing a comprehensive reform agenda — including a Social Media Privacy Rights Framework, an Independent Digital Rights Authority, mandatory algorithmic impact assessments, a seven-factor judicial balancing test, and a phased legislative roadmap — that provides a constitutionally grounded pathway toward a governance framework adequate to the challenges of the digital age.

Keywords: Right to Privacy, Freedom of Expression, Social Media, IT Act 2000, IT Rules 2021, DPDPA 2023, Puttaswamy, Algorithmic Amplification, Blocking Orders, Digital Rights, Sensationalisation, Independent Digital Rights Authority

I. Introduction

"Social media as a modern amphitheater of ideas cannot be left in a state of anarchic freedom." Justice Nagaprasanna of the Karnataka High Court's observation captures the central dilemma that social media poses for constitutional law in India. On the one side stands the freedom of expression — the oxygen of democracy, the precondition of accountable government, and the mechanism through which hundreds of millions of Indian citizens participate in the civic life of the nation. On the other stands the right to privacy — the sanctum sanctorum of human personality, the condition of individual autonomy, and the foundation of the dignity that the Constitution requires the state to secure for every person.

The conflict between these two rights is not new. Courts and scholars have grappled with it in the context of print journalism, broadcast media, and telephone surveillance for decades. What is new is the scale, speed, and structural character of the conflict on social media. A single social media post can reach tens of millions of people within hours. Algorithmic systems designed to maximise engagement systematically amplify the most emotionally arousing and sensational content — including content that violates individual privacy — at the expense of accurate, proportionate, and respectful discourse. Private information shared in one context can, within seconds, be stripped of its original context and disseminated to hostile audiences in a different context. Fabricated images and synthetic media can create false realities that are indistinguishable from genuine recordings. And the permanent accessibility of digital information ensures that momentary violations of privacy can follow individuals indefinitely, destroying reputations and opportunities long after the original violation.

The legal framework that governs this conflict in India is substantial but imperfect. The Information Technology Act 2000, as amended by the 2008 Amendment and supplemented by the IT Rules 2021, provides a framework of platform obligations, blocking powers, and grievance mechanisms that reflects a genuine legislative engagement with the governance challenges of social media. The DPDPA 2023 gives statutory content to the constitutional right to informational privacy recognised in Puttaswamy. The Supreme Court and High Courts have developed a body of jurisprudence on the privacy-expression conflict that is growing in sophistication and doctrinal coherence. But the framework contains structural gaps — in its treatment of peer-to-peer privacy violations, in its response to algorithmic amplification, in the independence of its oversight mechanisms, and in the adequacy of its remedies for victims of the most serious social media harms — that prevent it from adequately protecting either right.

This paper proceeds in five parts. Part II examines the philosophical foundations of both rights, establishing that they share a common root in the value of human autonomy and that neither can be exercised in a manner that destroys the conditions of the other's exercise. Part III analyses the constitutional and statutory framework for social media governance in India, with particular attention to the Section 69A blocking regime, the IT Rules 2021, the DPDPA 2023, and the major recent constitutional challenges. Part IV develops a typology of privacy-expression conflicts on social media and analyses the adequacy of India's legal response to each category through the lens of Indian and comparative jurisprudence. Part V proposes a comprehensive reform agenda addressing the most significant failures of the existing framework. Part VI offers conclusions that synthesise the paper's principal findings.

II. Philosophical Foundations: Privacy, Expression and the Common Root of Human Autonomy

A. The Right to Privacy: From Warren and Brandeis to Puttaswamy

The foundational text of modern privacy law — Samuel D. Warren and Louis D. Brandeis's 1890 article in the Harvard Law Review — identified the right to privacy as the right of the individual to protect their personality, their thoughts, sentiments, and emotions from the unwanted gaze of others. Warren & Brandeis, 'The Right to Privacy', 4 Harvard Law Review 193 (1890). Three aspects of this foundational account are directly relevant to contemporary social media law. First, truth is not a defence to a privacy violation: the injury caused by the unauthorised exposure of private information does not depend on the information's accuracy. Second, public interest constitutes a genuine qualification but not an unlimited one: the existence of a genuine public interest in some aspects of a person's life does not license the exposure of all aspects. Third, the primary injury of a privacy violation is dignitary, not material: it damages personality — the individual's capacity to present themselves to others on their own terms.

Alan Westin's foundational recharacterisation of privacy in *Privacy and Freedom* (1967) shifted the conceptual centre of gravity from the negative liberty of non-interference to the positive liberty of informational self-determination. Westin defined privacy as the individual's claim to determine for themselves when, how, and to what extent information about them is communicated to others. Westin, *Privacy and Freedom* (Atheneum, 1967). His identification of four states of privacy — solitude, intimacy, anonymity, and reserve — provides an analytical framework for understanding the specific ways in which social media erodes privacy: platforms structurally eliminate anonymity through real-name policies; they undermine reserve through the compulsory disclosure of personal information as a condition of access; and they destroy intimacy by enabling the mass redistribution of content originally shared in contexts of selective trust. Westin's most profound insight was his recognition that privacy is a precondition of authentic expression: where individuals know that their communications are being surveilled and their profiles harvested, they inevitably self-censor, producing a chilling effect on the freedom of expression that operates without any formal legal prohibition.

Ruth Gavison's access-based framework — defining privacy as limited access to the person through secrecy, anonymity, and solitude — Gavison, 'Privacy and the Limits of Law', 89 Yale Law Journal 421 (1980) — provides the analytical vocabulary for distinguishing between the distinct categories of social media privacy violation and the distinct legal responses each requires. Daniel Solove's taxonomy of privacy harms — encompassing information collection, information processing, information dissemination, and invasion — Solove, 'A Taxonomy of Privacy', 154 University of Pennsylvania Law Review 477 (2006) — identifies the aggregation problem as the most distinctively digital privacy threat: the combination of individually innocuous pieces of information into profiles that pose a specific and serious risk to individuals' safety and autonomy.

The Indian constitutional development of the right to privacy culminated in the unanimous judgment of the nine-judge bench in *Justice K.S. Puttaswamy (Retd.) v. Union of India* (2017) 10

SCC 1, which overruled the earlier cases that had denied a constitutional right to privacy and established that privacy is a fundamental right under Article 21 of the Constitution, with dimensions of personal liberty, informational self-determination, and human dignity. Justice Chandrachud's opinion in that case — the most philosophically comprehensive of the six opinions written by the bench — located privacy at the intersection of dignity, autonomy, and liberty, and established the proportionality framework — drawing on Aharon Barak's systematic analysis in *Proportionality: Constitutional Rights and Their Limitations* (Cambridge University Press, 2012) — as the constitutional standard for evaluating any restriction on the right.

B. Freedom of Expression: Mill, Habermas and the Democratic Public Sphere

John Stuart Mill's foundational defence of freedom of expression in *On Liberty* (1859) rests on two distinct but related arguments. The epistemological argument — that the free exchange of ideas is the mechanism through which truth is discovered and error corrected — generates the 'marketplace of ideas' metaphor that has been enormously influential in free speech doctrine. The autonomy argument — that the freedom to form and express opinions is constitutive of rational self-authorship — grounds expression as an intrinsic rather than merely instrumental right. Both arguments are relevant to the social media context, but both face specific challenges. The marketplace of ideas argument presupposes conditions of open, equal access and the ability to distinguish truth from falsehood — conditions that social media, with its algorithmic amplification of engagement-maximising content, systematically fails to create. The autonomy argument presupposes a public sphere in which individuals can genuinely form their own views through engagement with diverse perspectives — a public sphere that is distorted when algorithms curate information to confirm existing beliefs and provoke emotional reactions.

Jürgen Habermas's concept of the public sphere — *The Structural Transformation of the Public Sphere* (MIT Press, 1989) — provides the normative standard against which social media's performance as a forum for democratic deliberation must be evaluated. The public sphere, for Habermas, is constituted by openness, equality of access, and the orientation of discourse toward the common good rather than private interests. Social media has both realised and betrayed the promise of a democratised public sphere: it has removed the barriers of geography, class, and institutional access that restricted traditional public discourse, while simultaneously subjecting that discourse to the colonising logic of engagement-maximisation that treats attention as a commodity rather than a democratic resource.

Ronald Dworkin's account of rights as trumps — *Taking Rights Seriously* (Harvard University Press, 1977) — provides the framework for understanding the constitutional status of freedom of expression and its relationship to other values. If freedom of expression is a right in Dworkin's sense — a principled constraint on the pursuit of collective goals — then the state can restrict expression only where there is a genuine and specific risk of harm to identifiable others, not merely where a majority finds the speech offensive or inconvenient. This principle has direct application to the evaluation of blocking orders under Section 69A of the IT Act, which must be justified not by vague assertions of public interest but by identification of specific harms that the restriction is designed to prevent.

The common root of both rights lies in the constitutional value of human autonomy: both exist to protect individuals from forms of domination — privacy from the intrusive gaze of the state, the corporation, and the crowd; expression from the silencing power of authority. When either right is exercised without regard to the other, the result is not freedom but a different form of domination. The governance framework must protect both rights in their constitutional integrity, ensuring that neither can be used to

destroy the conditions of the other's exercise.

III. India's Governance Framework: Architecture, Constitutional Challenges and Recent Developments

A. The Information Technology Act 2000: From Neutrality to Accountability

The IT Act 2000 was enacted on the premise of network neutrality: platforms were conceived as passive conduits for information, bearing no responsibility for the content they transmitted. This model, drawn from Section 230 of the United States Communications Decency Act 1996 and the European Electronic Commerce Directive 2000, was adequate for the internet of 2000 but is fundamentally inadequate for the social media of 2024. The Act's three principal governance mechanisms — the criminal and civil liability framework for cyber offences (Sections 66C, 66E, 67, 67A, 67B), the safe harbour for intermediaries under Section 79, and the regulatory power to make rules under Section 87 — were not designed with the governance of social media as a mass public communication medium in mind.

The IT (Amendment) Act 2008 introduced the most significant expansions to the Act's governance framework, including Section 43A (data security obligations), Section 66A (criminalising offensive online speech — subsequently struck down in *Shreya Singhal*), and Section 69A (content blocking powers). It also reframed the safe harbour under Section 79 to condition intermediary immunity on compliance with due diligence obligations — a shift from pure neutrality toward conditional accountability that the IT Rules 2021 have further developed. By the mid-2010s, the limitations of this framework had become apparent: social media had grown from a niche communication service to the primary medium of Indian public discourse, with all the governance challenges that this transformation entailed.

B. Section 69A: The Blocking Power, Procedural Failure and Constitutional Correction

Section 69A empowers the Central Government to direct the blocking of online content on grounds confined to those specified in Article 19(2) of the Constitution. The Supreme Court upheld the provision in *Shreya Singhal v. Union of India* (2015) 5 SCC 1, specifically because it incorporated procedural safeguards — senior official approval, reasoned written orders, and the maintenance of records susceptible to judicial review — that distinguished it from the unchecked censorship power of the struck-down Section 66A.

The Delhi High Court's judgment in *Tanul Thakur v. Union of India* (W.P.(C) 13037/2019, decided 2022) exposed a systematic failure of those safeguards in practice. The petitioner, a journalist whose satirical website had been blocked, discovered that he had received no notice of the blocking order and could not obtain a copy of it because Rule 16 of the Blocking Rules 2009 classified all orders as 'strictly confidential.' The Court held that this interpretation was constitutionally impermissible: the natural justice principle *audi alteram partem* — that no one should be condemned without being heard — applies to blocking orders with the same force as to any adverse administrative action affecting fundamental rights. The confidentiality rule could be read consistently with natural justice by maintaining confidentiality against the general public while still providing notification to the specific individual whose expression is being suppressed. The *Tanul Thakur* judgment is significant beyond its specific facts: it establishes that the Puttaswamy proportionality framework applies to blocking orders and that procedural compliance with the constitutional requirements articulated in *Shreya Singhal* is not optional.

The Karnataka High Court's engagement with the blocking power in *X Corp v. Union of India* (W.P. No. 13710/2022) raised the additional question of the enforcement of IT Act obligations against foreign-incorporated platforms. X Corp challenged certain blocking orders on the ground that they had not complied with the procedural requirements of the Blocking Rules. The Court, while upholding the orders, confirmed that proportionality was subject to judicial scrutiny and acknowledged the existence of significant procedural concerns. The X Corp litigation also crystallised a gap in the IT Certifying Authorities Rules 2000 that does not define 'foreign company,' creating uncertainty about the legislative basis for requiring foreign-incorporated platforms to comply with Indian regulatory orders — a structural weakness that requires legislative remediation.

C. The IT Rules 2021: Accountability Framework and Constitutional Contestation

The IT (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021, notified on 25 February 2021, represent the most ambitious attempt yet made by the Indian legislature to impose accountability on social media platforms. Three legislative concerns drove the Rules: the proliferation of disinformation associated with episodes of mob violence; the tension between the government and platforms over content moderation decisions; and the use of encrypted messaging services to coordinate criminal activity. The Rules created a two-tier framework of obligations — general due diligence requirements for all intermediaries under Rule 3, and enhanced obligations for Significant Social Media Intermediaries (SSMIs, defined as those with over five million Indian users) under Rule 4.

Rule 3(1)(b) enumerates content categories that intermediaries must not host, including content that is 'invasive of another's privacy' — a provision that, if enforced, creates a positive obligation on platforms to prevent privacy violations rather than merely respond to them after the fact. However, the absence of a statutory definition of privacy-invasive content limits this provision's practical utility. Rule 4(1) requires SSMIs to appoint resident compliance officers and publish monthly compliance reports — a significant accountability mechanism that enables monitoring of platforms' content moderation activities. Rule 4(2), requiring messaging SSMIs to enable identification of the 'first originator' of messages, has been challenged on the ground that compliance necessarily requires the breaking of end-to-end encryption, exposing all users' private communications to potential interception — a restriction that fails the proportionality test of *Puttaswamy* at the stage of minimal restriction, given that *PUC v. Union of India* (1997) 1 SCC 301 established that private communications are protected by both Articles 19 and 21.

The most constitutionally significant challenge to the Rules arose from the April 2023 amendment to Rule 3(1)(b)(v), which required SSMIs to act against content designated as 'fake, false or misleading' by government fact-checking units, on pain of losing safe harbour protection under Section 79. In *Kunal Kamra v. Union of India* (Bombay High Court, September 2024), the third judge of a divided bench struck down the amendment as unconstitutional. Applying the *Puttaswamy* proportionality framework, the Court held that while preventing misinformation is a legitimate aim, the means employed — designating the government itself as arbiter of truth about its own conduct — were not rationally connected to that aim; the provision was not the least restrictive means; and the cost to free speech was grossly disproportionate. The judgment establishes a constitutional principle of wide import: the state cannot use the threat of private law consequences — including withdrawal of safe harbour protection — to achieve content regulation that it could not achieve through direct prohibition under Section 69A.

D. The Digital Media Ethics Code: Technological Neutrality and its Complications

Part III of the IT Rules 2021 establishes a three-tier Digital Media Ethics Code for online news publishers, applying the Press Council's Norms of Journalistic Conduct 2010 to digital media on the principle of technological neutrality. The first tier is publisher self-regulation; the second is industry self-regulatory bodies; and the third is Ministry of Information and Broadcasting oversight. The Code extends to digital media the privacy standards that the Press Council has developed for print journalism — standards that prohibit the identification of victims of sensitive crimes, require respect for the private life of non-public figures, and demand genuine public interest justification for any privacy-invasive publication.

The extension of these standards to online news is both appropriate and overdue. But complications arise. The Press Council's norms were designed for organised institutional journalism; applying them to individual bloggers, citizen journalists, and social media commentators risks a chilling effect on democratic participation. The third tier's Ministry oversight power — empowering the government to issue binding content takedown orders — is constitutionally contested: it amounts to a form of executive oversight of news content that sits uneasily with the constitutional prohibition on prior restraint, and it lacks the independence safeguards required when the content being ordered down may criticise the very officials exercising the power. The Madras High Court has *prima facie* assessed these concerns as raising serious constitutional questions.

E. The Sahyog Portal and the Architecture of Informal Governance

The Sahyog Portal, launched by the Ministry of Home Affairs in 2023, centralises law enforcement agencies' requests to social media platforms for content removal, account suspension, and user information disclosure. The Karnataka High Court in Justice Krishna S. Dikshit's order of 30 June 2023 upheld the Portal as a lawful procedural mechanism, noting that it creates no new legal powers and merely provides a more efficient channel for the exercise of existing statutory authority.

The constitutional concern, however, lies not in the Portal's formal legal status but in the informal governance dynamics it creates. A platform receiving a large volume of requests through a government-operated portal, dependent on government goodwill for regulatory clearances, and knowing that non-compliance is visible to senior officials, faces overwhelming incentives to over-comply with requests that may exceed strict legal requirements. The result is a system of informal censorship that is difficult to challenge through judicial review precisely because it produces no formal written orders. The Portal also enables the submission of user information disclosure requests — for identity details, IP addresses, and communication content

— that may not be backed by the court orders that the constitutional right to informational privacy under Article 21 requires. Any such disclosure without proper legal authority fails the proportionality requirements of *Puttaswamy*.

IV. Typology of Conflicts and Adequacy of Legal Response

A. Five Categories of Privacy-Expression Conflict on Social Media

The privacy-expression conflicts that social media generates are not a homogeneous category. Five distinct types of conflict can be identified, each requiring a different legal analysis and a different regulatory response.

The disclosure conflict arises when expression takes the form of the disclosure of true information about another person that the person regards as private — medical diagnoses, sexual orientation, financial

difficulties, private communications, or past conduct the individual has moved beyond. The defining characteristic is that the information is accurate; the issue is not its truth but the unauthorised appropriation of information that the subject reasonably expected would remain private. The accountability conflict arises when persons exercising public power invoke privacy to shield their public conduct from scrutiny — a conflict in which the key principle is that the public figure's privacy claim is weakest where information relates to the exercise of public power and strongest where it concerns genuinely private matters unconnected to the public role, as established in *R. Rajagopal v. State of Tamil Nadu* (1994) 6 SCC 632. The harassment conflict arises when expression is directed at a specific individual with the purpose of harming, intimidating, or humiliating — doxxing, coordinated harassment campaigns, trolling, and the extension of domestic abuse through digital channels. This category of expression makes no genuine contribution to the marketplace of ideas or the public sphere and is unambiguously restrictable under Mill's harm principle. The misrepresentation conflict encompasses defamation, deepfakes, and AI-generated synthetic content depicting real persons in conduct they did not engage in — combining the falsity harm of defamation with the intimacy violation of non-consensual image sharing. The aggregation conflict — identified by Solove — arises when the combination of individually innocuous pieces of information creates a profile that poses a specific risk to the subject's safety or autonomy, even though no individual item constitutes a privacy violation.

B. Media Trials, Algorithmic Amplification and the Irreversibility Problem

The phenomenon of trial by media — the parallel process of investigation, accusation, and judgment conducted by news organisations and social media users alongside formal legal proceedings — represents the most persistent and damaging systemic form of the privacy-expression conflict. The Law Commission's 200th Report (August 2006) documented the problem for traditional media; social media has transformed it qualitatively. The Sushant Singh Rajput coverage of 2020 — in which private WhatsApp messages were read on national television, intimate details of mental health and personal relationships were speculated about publicly, and individuals who were never charged with any offence were subjected to sustained vilification and threats — illustrates the scale of the harm and the total inadequacy of the existing legal framework to address it in real time.

The algorithmic amplification of privacy-invasive content is the structural mechanism through which social media transforms individual acts of privacy violation into mass harms. Platforms' engagement-maximisation algorithms do not distinguish between content that is engaging because it is informative and content that is engaging because it is sensational, humiliating, or invasive. Research by Metzler and Garcia demonstrates that content evoking moral outrage and negative emotional responses spreads significantly faster and further than neutral content, and that algorithmic recommendation systems amplify this differential — Metzler and Garcia, 'Social Drivers and Algorithmic Mechanisms on Digital Media', 19(5) *Perspectives on Psychological Science* 735 (2024). The victim of a privacy violation on social media is harmed not only by the original disclosure but by the cascade of algorithmic amplification that multiplies the harm before any legal remedy can be obtained. The notice-and-takedown model of Section 79 is structurally inadequate to address this amplification problem: by the time content is notified and removed, the irreversible harm of mass exposure has already been inflicted.

C. The Right to be Forgotten and the Permanence Problem

The destruction of practical obscurity — the pre-digital condition in which information that had entered the public domain was nonetheless practically inaccessible to most people because of the physical

constraints on its retrieval — is one of social media's most significant privacy consequences. The CJEU's recognition of the right to be forgotten in *Google Spain SL v. Agencia Española de Protección de Datos* (Case C-131/12, 2014) — requiring search engines to de-index links to personal information that is outdated, irrelevant, or disproportionate to any current legitimate purpose — provides the model for an Indian statutory framework that the Puttaswamy judgment acknowledged as constitutionally grounded but that legislation has not yet fully implemented. The DPDPA 2023's limited erasure right under Section 13, confined to Data Fiduciary-held data, does not address the broader category of social media content published by private individuals, nor does it address the search engine de-indexation dimension that has the greatest practical importance.

D. Non-Consensual Intimate Images, Deepfakes and the New Frontiers of Harm

The emergence of deepfake technology — generative AI capable of creating convincing synthetic audiovisual content depicting real persons in conduct they did not engage in — represents a qualitatively new category of privacy-expression conflict for which existing law is demonstrably inadequate. The Bharat Nyaya Sanhita 2023's provisions on voyeurism (Section 77), stalking (Section 78), fraud (Section 318), and impersonation (Section 340) provide a partial criminal law framework but do not specifically address deepfakes or the non-consensual sharing of images originally taken with consent. The Karnataka High Court's judgment in *Rajeshwari Khandasami v. State of Karnataka* (2023) — granting emergency injunctive relief against the non-consensual posting of intimate images as a violation of the constitutional right to privacy under Article 21 — provides the constitutional basis for urgent judicial relief, but a comprehensive statutory framework for NCII and deepfakes providing civil remedies, rapid content removal orders, and platform proactive detection obligations remains to be enacted.

E. Comparative Perspectives: ECtHR, US and UK Frameworks

The ECtHR's multi-factor balancing framework for privacy-expression conflicts, developed in *Von Hannover v. Germany* (No. 2) (2012) 55 EHRR 1 — assessing the contribution to public interest debate, the public or private status of the subject, the prior conduct of the person, and the content and consequences of the publication — provides the most practically applicable model for the principled resolution of specific privacy-expression conflicts, and closely parallels the Puttaswamy proportionality framework. The ECtHR's platform liability jurisprudence in *Delfi AS v. Estonia* (2015) ECHR 586 establishes that platforms have positive obligations to take reasonable preventive steps against manifestly illegal content causing serious harm to identifiable individuals — obligations that go beyond India's purely reactive Section 79 notice-and-takedown model.

The US approach — combining the actual malice standard of *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964) with the sweeping platform immunity of Section 230 of the Communications Decency Act 1996 — provides very strong protection for political speech but minimal accountability for platform design choices that amplify harmful content. The inadequacy of this model is increasingly recognised in the US itself. The UK Online Safety Act 2023 adopts a duty of care framework that is more relevant to India's governance tradition, imposing positive obligations on platforms to assess and mitigate risks to users' safety — including risks to privacy and dignity — associated with their algorithmic systems. The EU Digital Services Act 2022 goes furthest in imposing structural regulatory obligations on very large platforms, requiring systematic algorithmic risk assessments, the provision of non-profiling-based content options for users, and transparency about content moderation for researchers and regulators.

V. Reform Proposals: A Comprehensive Agenda for Constitutional Adequacy

A. Legislative Priorities

Six legislative reforms are required to address the most urgent failures of the existing framework. First, dedicated NCII and Deepfake legislation should create a specific criminal offence of non-consensual image sharing covering both genuine images and synthetic deepfake content, a civil cause of action without the need to prove a criminal offence, an emergency 24-hour injunctive remedy for content removal, and a proactive detection obligation on SSIMs with prescribed performance standards. Second, the DPDPA 2023 should be amended to extend its remedial framework to peer-to-peer privacy violations by private individuals on social media — the most significant legislative gap in the current framework, leaving victims of doxxing, non-consensual disclosure, and coordinated data exposure without an adequate statutory remedy. Third, a comprehensive Right to be Forgotten Act should establish statutory rights to de-indexation and content removal with clear criteria, accessible mechanisms, and appeal to an independent tribunal, with a public interest exception narrowly defined to prevent its misuse as a shield for ongoing privacy violations.

Fourth, the IT Act should be amended to introduce a statutory duty of care framework for platform design — requiring SSIMs to conduct annual algorithmic risk assessments identifying specific risks to users' fundamental rights, to implement proportionate mitigation measures, and to offer users meaningful choice over algorithmic curation. Fifth, the IT Rules 2021 should be amended to replace the Ministry oversight third tier of the Digital Media Ethics Code with an independent regulatory body; to provide a constitutionally precise definition of privacy-invasive content in Rule 3(1)(b); to limit Rule 4(2) traceability obligations to serious criminal investigations with court order requirements that prohibit the general weakening of encryption; and to strengthen the Grievance Appellate Committee's independence through a Collegium appointment process. Sixth, a Social Media and Fair Trial Act should address the media trial problem specifically, creating a specific offence of content likely to prejudge pending criminal proceedings and providing trial courts with an expedited takedown power.

B. A Social Media Privacy Rights Framework

Beyond specific legislative amendments, India requires a comprehensive Social Media Privacy Rights Framework — a coherent, principles-based statutory instrument establishing users' rights and platforms' obligations. The Framework should be organised around four foundational principles: informational self-determination (the right to control what personal information about oneself is published on social media); purpose limitation (information shared for one purpose cannot be appropriated for an incompatible purpose); proportionate public interest (privacy can be overridden only by a genuine, specific, and proportionate public interest, not by mere public curiosity); and structural accountability (platforms bear responsibility for the privacy harms generated by their design choices, not only for specific items of harmful content).

The Framework should translate these principles into six specific user rights: the right to informational privacy; the right to correction; the right to removal; the right to de-indexation; the right to explanation of algorithmic amplification; and the right to effective redress. These rights should be enforceable against both platforms (as Data Fiduciaries) and individuals, with a hierarchy of remedies from platform-level removal through civil damages to criminal sanctions for the most serious violations.

C. An Independent Digital Rights Authority

The most significant institutional gap in India's social media governance is the absence of an

independent, expert regulatory authority. The proposal for an Independent Digital Rights Authority (IDRA) — established by primary legislation as a body corporate, independent of government, with its composition determined by a Collegium including representatives of the judiciary, legal profession, civil society, and academic experts — addresses this gap. The IDRA should have four principal functions: regulatory oversight (including algorithmic audits, compliance directions, and financial penalties); adjudication of user-platform disputes as a first-tier tribunal; standard-setting through codes of practice and technical guidance; and transparency reporting through annual assessments of social media governance in India. Drawing on the model of TRAI, SEBI, and the Competition Commission, the IDRA should be given sufficient financial resources and statutory independence to conduct meaningful oversight of platforms whose technical and legal resources vastly exceed those of most national regulators.

D. A Seven-Factor Judicial Balancing Test

Courts resolving privacy-expression conflicts in the social media context should apply a seven-factor multi-factor balancing test grounded in the Puttaswamy proportionality framework. The seven factors are: (1) the nature and severity of the privacy violation — more severe violations of more fundamental privacy interests require stronger justification; (2) the public or private status of the subject — public figures have reduced expectations in relation to their public roles, per R. Rajagopal (1994) 6 SCC 632; (3) the public interest served by the expression — genuine accountability journalism and political commentary receive stronger protection than mere entertainment or voyeurism; (4) the accuracy and legitimacy of the source — information obtained through hacking, surveillance, or breach of confidence receives weaker protection; (5) the availability of less privacy-invasive means of serving the same public interest; (6) the scale and permanence of the disclosure — mass audiences and permanent digital records require stronger justification than limited or ephemeral communications; and (7) the conduct of the parties — including the public figure's prior voluntary disclosures and the publisher's observance of professional standards.

These seven factors are not a mechanical checklist but a structured framework for the principled resolution of specific cases, consistently with the constitutional commitment to both rights as expressions of the fundamental value of human dignity. Applying the test to illustrative cases:

political satire about a minister's public policies clearly favours expression; the non-consensual posting of intimate images clearly favours privacy; accountability journalism about a public official's undisclosed financial interests falls toward expression; the aggregated doxxing profile of a private journalist falls decisively toward privacy; and the viral amplification of an old social media post about a private individual's personal crisis triggers the right to be forgotten.

E. A Phased Implementation Roadmap

The reform agenda should be implemented in three phases. In the immediate term (one to two years), priority should be given to: amendment of the Blocking Rules 2009 to provide for notification of affected persons and mandatory periodic review; enactment of NCII and deepfake-specific legislation; amendment of the IT Rules 2021 to revise the third-tier Ministry oversight mechanism; and strengthening the Grievance Appellate Committee's independence. In the medium term (two to five years): enactment of the Right to be Forgotten Act; amendment of the DPDPA 2023 to cover peer-to-peer privacy violations; enactment of the Social Media and Fair Trial Act; establishment of the IDRA; and constitution of Data Tribunals with emergency jurisdiction. In the long term (five to ten years):

mandatory algorithmic impact assessments and the statutory duty of care framework; enactment of the comprehensive Social Media Privacy Rights Framework; development of international cooperation agreements; and evaluation of the DPDPA 2023 in the light of operational experience and comparative developments.

VI. Conclusion

This paper has argued that the conflict between the right to privacy and the freedom of expression on social media in India is a conflict about what kind of digital public sphere India wants to be. A public sphere that treats privacy as a dispensable luxury — in which every person's most intimate vulnerabilities are commercially exploited and algorithmically amplified for mass entertainment — is incompatible with the Constitution's commitment to human dignity. A public sphere in which expression is subject to opaque blocking orders, informal government pressure on platforms, and government-operated fact-checking units is equally incompatible with the Constitution's guarantee of free speech as the foundation of democratic governance.

The analysis has established three principal findings. First, the existing legal framework — while more substantial than its critics sometimes acknowledge — contains structural gaps and implementation failures of sufficient seriousness to prevent it from adequately protecting either right. The reactive notice-and-takedown model is inadequate to the speed and scale of social media harm; the absence of dedicated NCII and deepfake legislation leaves victims without effective remedies; the right to be forgotten lacks a comprehensive statutory framework; and the governance of algorithmic amplification is almost entirely outside the legal framework. Second, the most significant recent developments — the Puttaswamy proportionality framework, the Tanul Thakur judgment on natural justice in blocking proceedings, the Kunal Kamra decision on government fact-checking, and the Rajeshwari Khandasami constitutional framing of NCII — provide a growing body of doctrine from which a more adequate framework can be developed. Third, the reform agenda required is comprehensive but achievable: it requires targeted legislative amendments in the short term, institutional creation in the medium term, and structural platform regulation in the long term, all grounded in the constitutional values of dignity, proportionality, and effective rights protection.

The right to privacy and the freedom of expression are not enemies. They are, as this paper has argued, complementary expressions of the same constitutional value — the autonomy of the individual to live a life of their own choosing, to participate in democratic discourse on their own terms, and to be treated by the state and their fellow citizens as a person of inherent dignity and worth. The governance of social media must be guided by this understanding. Creating the legislative, institutional, and judicial conditions in which both rights can flourish together — rather than trading one off against the other — is the central challenge of India's digital constitutionalism in the years ahead.

References

1. Warren, S.D. and Brandeis, L.D., 'The Right to Privacy', Harvard Law Review, Vol. 4, No. 5 (1890), pp. 193–220.
2. Westin, A.F., Privacy and Freedom, Atheneum, New York (1967).
3. Gavison, R., 'Privacy and the Limits of Law', Yale Law Journal, Vol. 89, No. 3 (1980), pp. 421–471.

4. Solove, D.J., 'A Taxonomy of Privacy', University of Pennsylvania Law Review, Vol. 154 (2006),
5. p. 477.
6. Solove, D.J., Understanding Privacy, Harvard University Press, Cambridge MA (2008).
7. Habermas, J., The Structural Transformation of the Public Sphere, trans. T. Burger, MIT Press, Cambridge MA (1989).
8. Dworkin, R., Taking Rights Seriously, Harvard University Press, Cambridge MA (1977).
9. Barak, A., Proportionality: Constitutional Rights and Their Limitations, Cambridge University Press, Cambridge (2012).
10. Zuboff, S., The Age of Surveillance Capitalism, PublicAffairs, New York (2019).
11. Metzler, H. and Garcia, D., 'Social Drivers and Algorithmic Mechanisms on Digital Media', Perspectives on Psychological Science, Vol. 19, No. 5 (2024), pp. 735–748.
12. Justice K.S. Puttaswamy (Retd.) v. Union of India (2017) 10 SCC 1. Shreya Singhal v. Union of India (2015) 5 SCC 1.
13. R. Rajagopal v. State of Tamil Nadu (1994) 6 SCC 632.
14. People's Union for Civil Liberties v. Union of India (1997) 1 SCC 301. Anuradha Bhasin v. Union of India AIR 2020 SC 1308.
15. Nipun Saxena v. Union of India (2019) 12 SCC 753.
16. Tanul Thakur v. Union of India, W.P.(C) 13037/2019 (Delhi HC, decided 2022). X Corp v. Union of India, W.P. No. 13710/2022 (Karnataka HC).
17. Kunal Kamra v. Union of India (Bombay HC, September 2024). Rajeshwari Khandasami v. State of Karnataka (Karnataka HC, 2023). Kinnori Ghosh v. Union of India (Supreme Court, 2024).
18. Google Spain SL v. Agencia Española de Protección de Datos, Case C-131/12 (CJEU, 2014). Von Hannover v. Germany (No. 2) (2012) 55 EHRR 1.
19. Delfi AS v. Estonia (2015) ECHR 586 (Grand Chamber). New York Times Co. v. Sullivan, 376 U.S. 254 (1964).
20. Constitution of India, 1950, Articles 19(1)(a), 19(2), 21.
21. Information Technology Act, 2000 (as amended), Sections 43A, 66C, 66E, 67, 67A, 67B, 69A, 79, 87.
22. IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Rules 3, 4, 9.
23. IT (Procedure and Safeguards for Blocking for Access of Information by Public) Rules, 2009, Rules 8, 9, 15, 16.
24. Digital Personal Data Protection Act, 2023, Sections 6, 11, 13.
25. Bharat Nyaya Sanhita, 2023, Sections 77, 78, 318, 340. EU Digital Services Act, 2022.
26. UK Online Safety Act, 2023.
27. Law Commission of India, 200th Report on 'Trial by Media: Free Speech and Fair Trial under Criminal Procedure Code, 1973' (August 2006).
28. Press Council of India, Norms of Journalistic Conduct, 2010.
29. Council of Europe, Guidelines on Safeguarding Privacy in the Media, October 2018. DataReportal, Digital 2024: India — Global Digital Insights, January 2024.
30. Bhatia, G., 'The Supreme Court's Right to Privacy Judgment', Indian Constitutional Law and Philosophy (2017).
31. Dhawan, S., 'Karnataka High Court Ruling on Content Blocking: A Setback for User Rights', IJCL, NALSAR (2025).

32. Allahrakha, N., 'Constitutional Safeguards for Digital Rights and Privacy', International Journal of Law and Policy, Vol. 2(4) (2024), pp. 31–43.
33. Mathew, M., 'Freedom of Information, Right to Express and Social Media in India', Interactive Entertainment Law Review, Vol. 3(2) (2020), pp. 94–104.